TECHNISCHE UNIVERSITÄT
CHEMNITZ

Fakultät für Informatik

CSR-24-06

Design and Development of a Diagnostic Learning Analytics System

Saddaf Afrin Khan · Ummay Ubaida Shegupta · Wolfram Hardt

August 2024

Chemnitzer Informatik-Berichte

TECHNISCHE UNIVERSITÄT
CHEMNITZ

Design and Development of a Diagnostic Learning Analytics System

Master Thesis

Submitted in Fulfilment of the
Requirements for the Academic Degree
M.Sc.

Dept. of Computer Science
Chair of Computer Engineering

Submitted by: Saddaf Afrin Khan
Student ID: 623861
Date: 02.08.2023

Supervising tutor: Prof. Dr. W. Hardt
Ummay Ubaida Shegupta
René Schmidt

Acknowledgement

I am utterly humbled and grateful to everyone who has assisted me in turning my ideas into something real and substantial that goes much beyond the level of the merely conceptual.

I want to express my gratitude to Ummay Ubaida Shegupta, who oversaw my thesis and functioned as both my mentor and my counselor. I was able to successfully complete my thesis because of her compassion, perseverance, and ambition. She also gave me writing tips that I used when I was putting my report together. René Schmidt and André Böhle provided me with a lot of helpful feeedback about the implementation, so I also want to thank them for that.

In addition, it gives me great pleasure to thank my academic advisors, Prof. Dr. Dr. h. c. Wolfram Hardt and Dr. Ariane Heller. Dr. Ariane Heller's advice kept me motivated throughout our periodic meetings, and her kind criticism of my report pointed me in the proper path.

My genuine gratitude goes out to my friends and coworkers for their constant support and patience when I was under academic pressure. This voyage was made more bearable and delightful by their company and support. I want to sincerely thank my family for their everlasting support and affection. My desire to learn has been motivated by their confidence in my talents and ongoing support.

The ultimate source of my wisdom and enlightenment is none other than my Creator, Allah, who granted me the ability to remain focused. I'm grateful that You have helped me down this academic path and have given me the opportunity to add, however little a way, to the huge field of human understanding. I appreciate You being my constant companion and the source of all my motivation as I pursue knowledge.

Abstract

In today's modern educational systems, teachers and institutions always try to improve academic performance and the student learning experience. The topic of diagnostic learning analytics is quickly developing at the nexus of technology, data science, and education. By utilizing the enormous volumes of educational data produced by various learning management systems, online platforms, and educational technology, diagnostic learning analytics provides a data-driven method for achieving these objectives.

The basic goal of diagnostic learning analytics is to collect, analyze, and interpret learner-related data in order to acquire insights about their learning patterns, strengths, limitations, and overall development. Educators may detect individual student requirements, learning styles, and knowledge gaps using modern data mining approaches. This enables teachers to modify teaching tactics and curricula to match the individual needs of each student, establishing a personalized and adaptable learning environment.

Furthermore, diagnostic learning analytics supports not only individual learners but also educators and institutions in making data-driven decisions. It enables them to optimize resource allocation, devise focused interventions, and track the long-term performance of various instructional techniques.

In this proposed diagnostic learning analytics, the student test data has been analyzed with drill-down and correlation for figuring out the problem laying under the dataset. The test data has been provided by some APIs, where along with test data the details of the student information can be found. These data have been analyzed by using Python and visualized those analyses in a dashboard by using ReactJS and concluded with a satisfying result which has been evaluated with the existing final grade score. The evaluation shows that this proposed analytical system gives enough accurate results to detect the student who is having trouble with the course and the tests.

Keywords: Learning Analytics, Diagnostic Learning Analytics, Data Analytics, Dashboard, Data Visualization

Table of Contents

List of Figures

List of Tables

List of Abbreviations

LA Learning Analytics

SOLAR Society for Learning Analytics Research

LAK Learning Analytics and Knowledge

KPI Key Performance Indicator

PCA Principal Component Analysis

LOF Local Outliner Factor

HR Human Resource

ML Machine Learning

SQL Structured Query Language

SPSS Statistical Package for the Social Sciences

DA Data Analytics

NC Numeral Controls

SM Smart Manufacturing

MPFM Multi Phase Flow Meter

PTA Programmable Time Accumulator

GOR Genomic Ordered Relational

LOS Length Of Stay

LWBS Left without Being Seen

LBCT Left Before Complete Treatment

VCE Virtual Classroom Environment

MOOC Massive Open Online Courses

LSTM Long Short-Term Memory

ANN Artificial Neural Networks

SVM Super Vector Machine

CAROL Center for Advanced Research Through Online Learning

BPTT Backpropagation Through Time

T-MON Traces Monitor

GLA Game Learning Analytics

KNN K-Nearest Neighbors

IDE Interactive Development Environment

ARS Audience Response System

OPAL Online Platform for Academic Learners

1 Introduction

The phrase "Learning Analytics" is regarded as an emerging area that would be implemented in many educational contexts since the first LAK conference in 2011 [17] [18]. According to SoLAR, "*Learning Analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs*"[1]. Nowadays, it refers to a distinct discipline and is utilized in connection with the use of analytics in e-learning environments. Academics, researchers, and administrators have recently been interested in LA. This interest stems from a desire to comprehend "intelligent content," personalisation, and adaption better as well as teaching and learning [19]. Utilizing LA to understand students' learning paths has been made possible by the usage of LMSs, which make it simple to access the activity of the students. This allows their improvement during the learning process [20].

There is currently a wide range of educational settings available, including virtual environments, MOOC platforms, and learning management systems (LMS). This "Big Data" of students is stored in these educational information systems, which are vast data banks [18]. Learning analytics has grown in popularity along with the development of these increasingly massive data sets. A variety of fields has influenced the methodology employed in EDM (Educational Data Mining) and LA (Learning Analytics). Still, the two main ones that have had the biggest impact on the field are those of data mining and general analytics [21]. Sometimes researchers use "Educational data Mining" (EDM) and "learning analytics" interchangeably. However, there are significant disparities between these two study communities. Assume that EDM prioritizes automated discovery while learning analytics stresses human judgment. EDM models provide automatic adaptation, whereas learning analytics empowers teachers and learners. EDM evaluates individual components, whereas learning analytics comprehends systems as a whole. Basically, educational analytics figures out the hidden pattern of big data related with education and learning analytics uses these pattern to optimize the learning environments [22].

Diagnostic analytics analyzes data to determine why something occurred. Diagnostic analytics is used in education to determine why a student did well or poorly. It everything comes down to relationships: X occurred as a result of Y. Jina did well because she paid greater attention to detail. Shopie fared poorly because she did not see a tutor when she should have. Educators will go deep into the tactics to see which ones helped children and which ones might have done more. Understanding why a certain instructional method succeeded or did not work is essential for making

[1] https://www.solaresearch.org/

changes.

The main goal of diagnostic learning analytics is to collect, examine, and evaluate learner-related data in order to acquire an understanding of learners' learning patterns, aptitudes, and general development. The demands, learning preferences, and knowledge gaps of each individual student may be determined by instructors using cutting-edge data mining tools. This enables teachers to modify teaching tactics and curricula to fit the unique needs of each student, promoting a customized and adaptable learning environment.

The data processing cycle can be seen in the Figure 1.1 for a learning analytics system. According to [1], the input data would come from educators, learners or may be from the institution. These data would be classified into academic, cognitive, or psychological the other three metric (progression, performance and compliance). After classifying the data it would be used by learning analytics which can be descriptive, diagnostic, predictive or prescriptive.

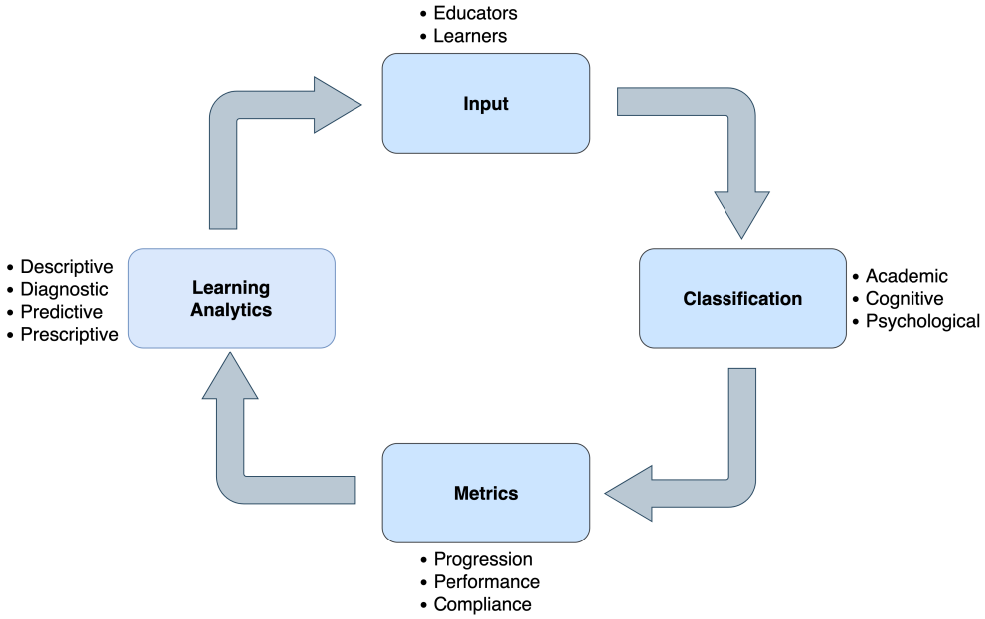


Figure 1.1: Data processing in learning analytics [1]

Any analytics process consists of basic five steps - data gathering, pre-processing, analysis, post-processing and interpreting the result. When it comes to automatic analytical process then these steps are supposed to be automatically done by a system, except the interpreting the result step. This step needs human interaction for better decision-making. In Table 1.1 some of the very famous learning analytics models are listed. These are all based on the basic analytics steps. This section will give some short description of all of the listed models.

Tanya Elias processed a learning analytics [2] model which is an improved version of three-phase cycle learning analytics process, proposed by Dron and Anderson in

Learning Analytics Models	**Year**	**Focused Criterion**
Elias's LAM	2011	Four types of technology resources.
Chatti's learning analytics reference model	2012	Based on four dimension.
Greller & Drachsler Framework	2012	Six dimensions are focused.
Siemens LAM	2013	Focused on seven components.
LAVA Model	2020	Four dimensions are focused like Chatti, but considered human perspective

Table 1.1: List of Learning Analytics models

the year 2009. Elias focused on four types of technology resources (Computer, People, Theory and Organization) which were basically the challenges for the previous LA model. Computer comes first in the mind when the resource is about technology. Most of the higher education systems are based on distance education. Therefore, software and hardware plays a major role in learning analytics. Computer is useful that time when it has sound knowledge about the solution. At this stage, theory plays the role as the knowledge. People always plays an important role for any kind of anlytics. Though the modern applications are build up in such way so that human effort becomes lower day by day, but human knowledge and ideas are always act as a fuel in any technical process. Last but not least is, organization. It is very important to know the targeted organization and their need. Figure 1.2 is showing the learning analytics model which is proposed by Elias.

In Figure 1.3, the reference model for learning analytics proposed by M.A. Chatti can be seen. This model is focused on four dimensions. The first concern of this model is about What? This means, what kind of data will be used for the further analysis. The data source can be LMS, student attendance system etc. Next concern is Who? - the stakeholders. The analysis needs to be specified for the targeted users. For a good analysis, it is very important to know the objective, which means - Why? The main base of the learning analytics should stand upon this dimension. The forth dimension is How?. Which techniques are going to use to develop the analytics.

Figure 1.4 depicts the proportions of this structure. Each dimension has a collection of instantiations. The list of occurrences in the figure is not intended to be complete and can be expanded. Greller and Drachsler consider these dimensions crucial since each one must have at least one occurrence in a completely LA design.

According to George Siemens, under the suggested Learning analytics model [5], a systematic approach guarantees that all support resources are systematized so

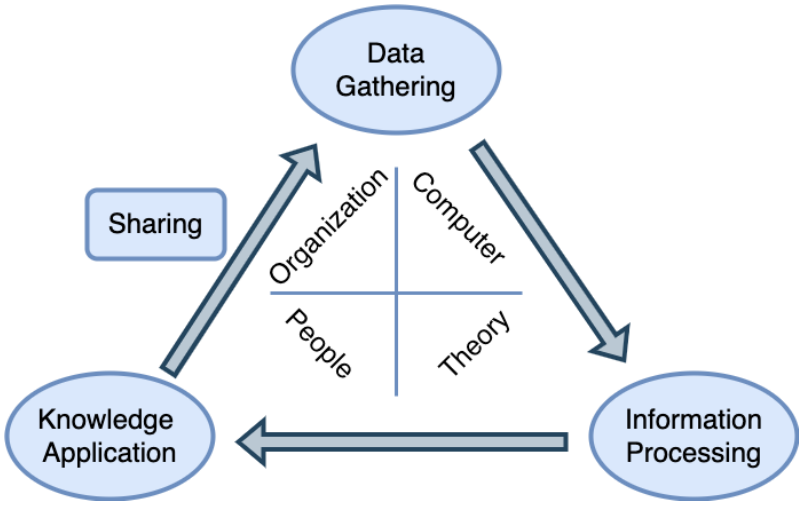


Figure 1.2: Elias learning analytics model [2]

that interventions or the building of predictive models are only feasible with the complete assistance of the whole education institution. Collection, storage, data cleansing, integration, analysis, representation and visualization, and action are the seven components of this approach. In Figure 1.5, this model can be seen with all of the components. From this figure it is also clear about the data team role in this model. There are different roles of data team can play for this learning analytics model.

In year 2020, M.A. Chatti with A. Muslim proposed an updated version of previous Chatti's LA. This version of model is human centered learning analytics [6]. In this model a visual analytics (VA) is integrated in a leaning analytics, so that the users can control the learning analytics process through the VA. The model has been enhanced from the last version by incorporating the human point of view in Who? and research in How? dimension. Figure 1.6 illustrates the eight stages of this model. This model starts with *"Learning Activites"* and ends at the *"Visualization"* stage. But in LAVA, human plays very important role. An user can change the whole analysis path by giving his/her own knowledge in the *"Preception"* stage. Either this knowledge will be used by *"Exploration"* then the *"Analysis"* stage will change the result and visualize it via visualization or the whole concept of the analysis can be changed via *"Action"* stage.

Learning analytics has very old history. This concept is not recently developed. The first idea of learning analytics started in the year 1927 with Pressey's intelligent tutoring systems (ITS). In 1956, the Self-Adaptive Keyboard Instructor (SAKI) was introduced which served as the first step in how technology can help in a student's learning journey. The use of technology mainly started with computer-assisted instruction (CAI) system and Programmed Logic for Automatic Teaching Operations (PLATO) in 1960s, though their usage was very limited until 1970s. In between late 1970s and early 1980s artificial intelligence (AI) was introduced in learning training. In very limited way, AI was able to present and generate different options

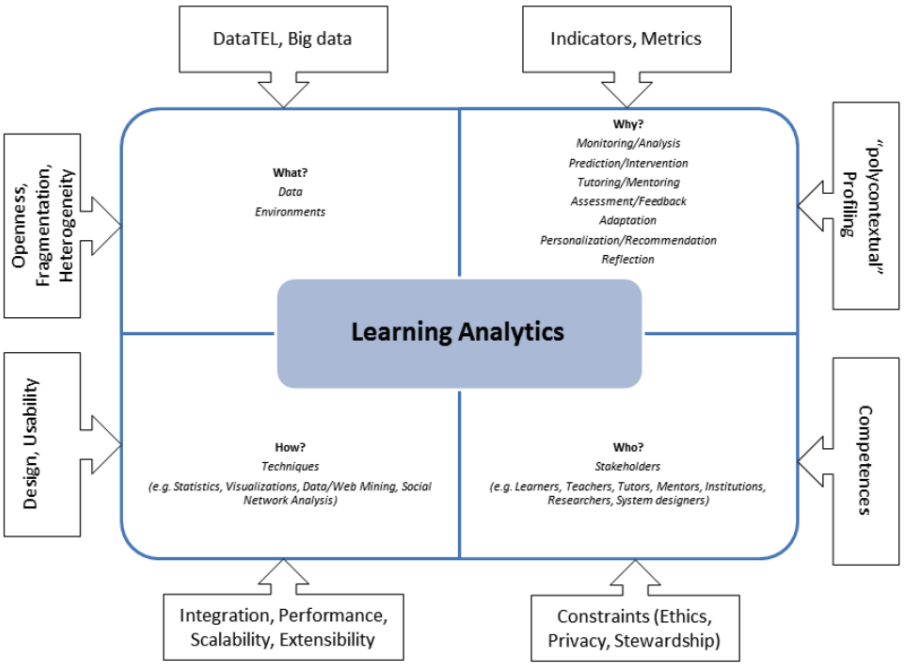


Figure 1.3: Chatti's learning analytics reference model [3]

about based on student's performance. With the development of World-Wide Web (WWW) in late 1990, the modern online learning evolved. "Interactive Learning Network" (1997) and Desire2Learn (1999) were become widely used LMS (learning management system) that time. In 2002, Moodle (Modular Object-Oriented Dynamic Learning Environment) was introduced and this can be called the first open source learning management system. Massive Open Online Courses (MOOC) was introduced in 2008 as development in the distance education. This was massive step for the beginning of analysis the learning data. 2010s was the biggest evaluation era for learning analytics. From this a drastic grow started happening in the learning sector. In 2011, International Educational Data Mining Society developed. Same year the first time Learning Analytics Knowledge conference occurred. As the number the reasearchers were growing, in 2013 Society for Learning Analytics Research (SoLAR) established. The first LA handbook published in 2017 [23], to meet the needs of a new and growing field of learning analytics. The whole timeline can be seen in Figure 1.7, which has been adapted from [7].

Learning analytics is a subset of data analytics that is especially applied to the subject of education. Data analytics is a larger domain that incorporates the act of analyzing massive amounts of data to identify patterns, trends, and insights. Learning analytics is using data analytics techniques in educational data to acquire insights and enhance learning results. In learning analytics, data analytics techniques such as statistical analysis, data mining, machine learning, and visualization are used for educational data gathered. These strategies aid in discovering patterns, trends, correlations, and anomalies in data. The study attempts to offer useful insights that

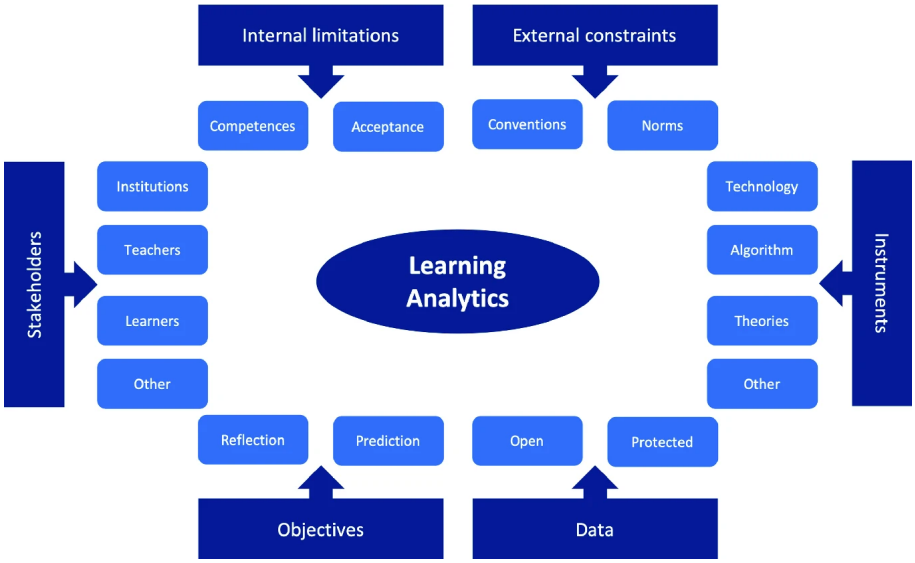


Figure 1.4: Greller & Drachsler learning analytics framework [4]

may be used to influence decision-making and enhance educational practices.

1.1 Data Analytics

Early in the 2000s, the phrase "Data Analytics" first appeared [24]. Data analytics refers to examining raw data to extract important, useful insights that can then be applied to guide and make wise decisions. The term "data analytics" can be used in multi-disciplined and a huge range of analytical techniques are covered by it. When it is time to gain insight into any kind of information, different data analytics techniques can be applied to make the data human readable. Techniques for data analytics can make trends and indicators visible that might otherwise be lost in the sea of data. The efficiency of a firm or system can then be improved by using this knowledge to optimize procedures.

What is the difference between analysis and analytics? [25]. Although the phrases analytics and analysis looks similar and are sometimes used interchangeably, in real case they are not the same and the concept of these two factors are totally different. If the basic definition is considered, then analysis is a process where a large problem dissected into small components to dive deep into the problem. It is frequently used for complicated systems that must be made simpler by being broken down into their more elucidating/understandable components because analyzing the system as a whole is neither viable nor practical. A procedure is known as synthesis is used to reassemble the entire system (either a conceptual or physical system) once the changes at the simplistic level have been realized and the analysis of the pieces has been completed.

On the other hand, analytics refers to a wide range of techniques, technologies, and related instruments for generating fresh information or insight to address challenging

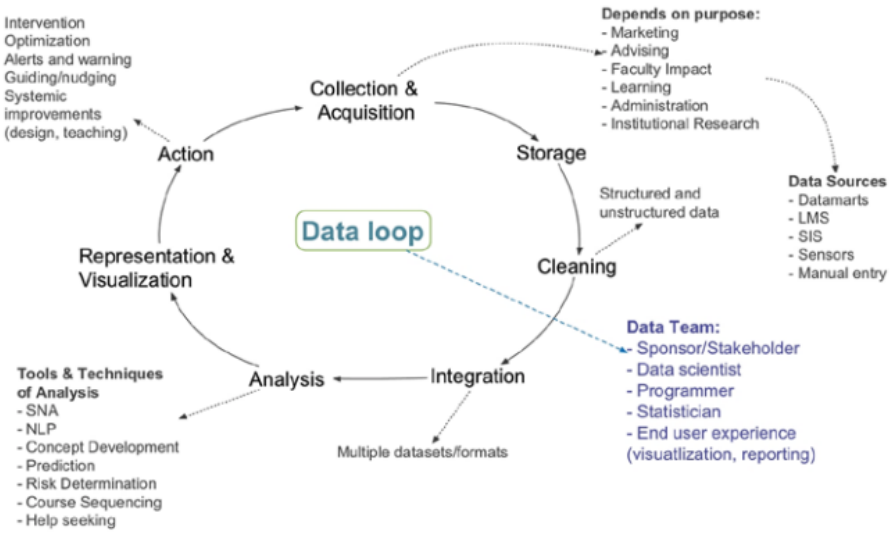


Figure 1.5: Siemens Learning Analytics model [5]

issues. Analytics is essentially a complex method to comprehending and dealing with complex circumstances. In order to make sense of an increasingly complex reality, analytics makes use of data and mathematical models. Analytics is more than simply analysis; it also includes synthesis and later execution. Analytics does encompass the act of analysis at various phases of the discovery process. It is primarily a methodology that includes a wide range of techniques and procedures.

Data analytics is called multidisciplinary research field. It combines ideas with big data from different scientific fields, such as - statistics, machine learning, artificial intelligence etc. [26]. A corporation can transform data into knowledge to improve its capacity for strategic decision-making. That firms' efforts to launch new ventures and line extensions are receiving a new lease of life thanks to developments in analytics [27]. Data is pervasive nowadays and is expanding at an exponential rate. It is becoming more and more crucial to various industries as data volume and complexity increase.

Technology is taking the lead in the modern world. Every aspect of life, such as communication, work environment, healthcare, education, etc., has been revolutionized by it. In modern life, it is possible to develop a smart city or track health issues without staying in a healthcare institution or create a smart home where it would interact as a human being by using the Internet of Things (IoT). Big data serves as fuel for this massive evaluation. Data is becoming a valuable asset in every industry. As a result, it becomes a tremendous obligation to handle and evaluate ethical issues of any form of data in a sophisticated manner.

Now a question arises, *What is big data?* Samuel Madden, Computing Distinguished Professor of Computing at MIT, described big data in [28] as *"data that's too big, too fast, or too hard for existing tools to process"*. For explaining these three terms, he said, too big means to get petabyte of data after one single action, too

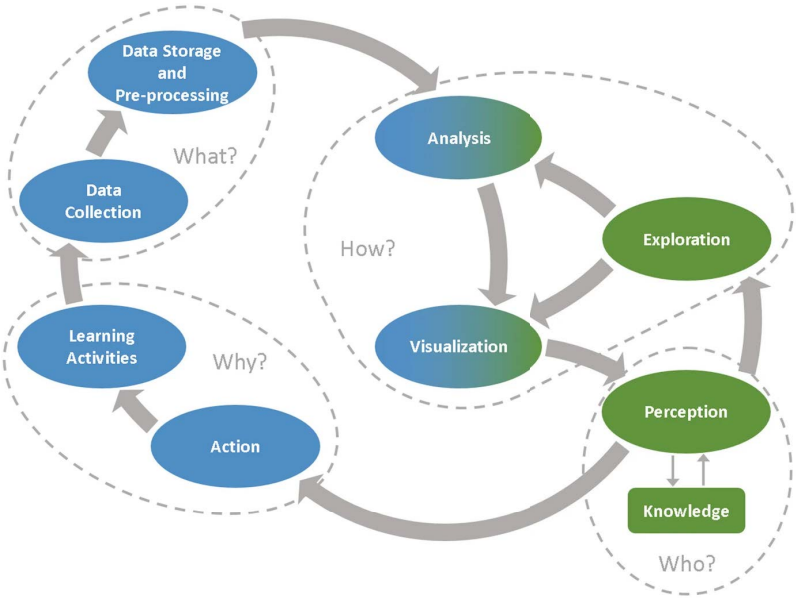


Figure 1.6: The LAVA Model [6]

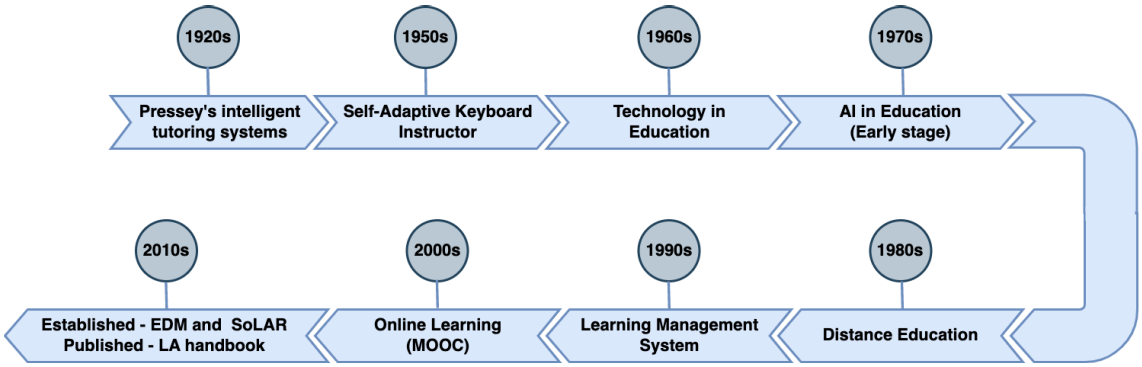


Figure 1.7: History of learning analytics [7]

fast means processing the data so quickly so that no ethical hamper can be done with that big scale of data and too hard means existing tools can not analysis the catchall data rapidly. Gartner[2] defined big data with three Vs (volume, velocity, and variety) - which are complex versions of too big, too fast and too hard.

The three characteristics mentioned above of big data can be managed and handled by using data analysis. The goal of big data analysis is to predict the structure of the hidden features of the whole data, while the sample data is small and another goal is extracting essential common traits across numerous subgroups when there are huge individual variances [29]. While analyzing big data, analysts need to face issues of heterogeneity, experimental variances, and statistical biases. These are the reasons for necessitating the development of more adaptable and resilient techniques. At this point, big data analytics takes place.

[2]https://www.gartner.com/

Statistics play a significant role in the data analytic world. In Oxford English Dictionary[3], statistics has been defined as - *"The systematic collection and arrangement of numerical facts or data of any kind"*. This definition is clearly saying that statistics is all about numbers and arrange them in such a way that the outcome gives a good understanding about some of the scattered numbers. When applying statistics to a scientific, industrial, or social problem, it is normal to start with a statistical population or a statistical model to be researched. Populations can be varied groupings of people or things, such as "all people living in a country" or "every atom composing a crystal." Statistics deals with all aspects of data, including data collection planning in the form of survey and experiment design.

Although statistics is all about numbers it can be called a mathematical field of science that deals with the gathering, analysis, interpretation or explanation, and presentation of different types of numbers. Some believe statistics to be a separate mathematical discipline rather than a part of mathematics [30]. While much scientific research makes use of information, statistical analysis has to do with the use of data in the context of uncertainty and deciding in the face of uncertainty [31]. When applying statistics to an issue, it is usual practice to do research on a common statistical topic, such as population. The reason to choose a common topic is to understand how the statistics method would work with the new dataset.

Statistics will summarize the data quantitatively. This summary can describe the features of a collection of different kind of information. This kind of statistics is called descriptive statistics [32]. There is another type of statistics, which give a deeper meaning to the information. This statistics can do the hypotheses or derive the estimation. This statistics is called statistical inference. In data science world the exploratory data analysis (EDA) is used. EDA is a process to summarize their main features of a dataset. Sometimes it also used visualization. With statistical method EDA identify the meaning of data beyond any kind of formal modeling or testing methods.

The role played by statistics in the data analytics field can not be misjudged. Statistics is kind of ground field for data analytics. It provides many tools and techniques which very necessary to extract the deeper meaning from the data. Data analytics works with big data where it is very necessary to understand the patterns. Statistics can recognize this pattern from any data by sampling it from a large set of it. Statistics will also help to predict what kind of data can come and where the change can happen. As mentioned above, descriptive statistics would summarize the main characteristics of data by using mean, median, mode, standard deviation. Then sampling is most significant player in data analytics, as it would provide methods to select small samples from the large dataset and further investigation can be done on that sample dataset. After getting this sample, inferential statistics can draw a conclusion or can perform some future predictions on it. Hypothesis testing, regression analysis can fall into this type of statistics. There is another theory most frequently used in data analytics, that is probability. It calculates the chance of

[3]https://www.oed.com/

events occuring based on the attributes of a given dataset. Theories like - Normal distribution, Binomial distribution- are common in data analytics for probability. In data analytics it very important to know the relationship between different variables. For this feature, Statistics also provide a theory called correlation analysis. This will help to understand how the variables are interconnected with each other. In every scientific fields, experiment is very important to take a better and innovative step. Statistics would also help any analytical system to determine the sample size, defining control groups and effect of the variables in the controlled sample dataset. Mathematical model present the real world scenario from the existing data. Linear regression, time series analysis - are some statistical mathematical model which can be used the data analytics to develop predictive models and can be used in data driven decision making.

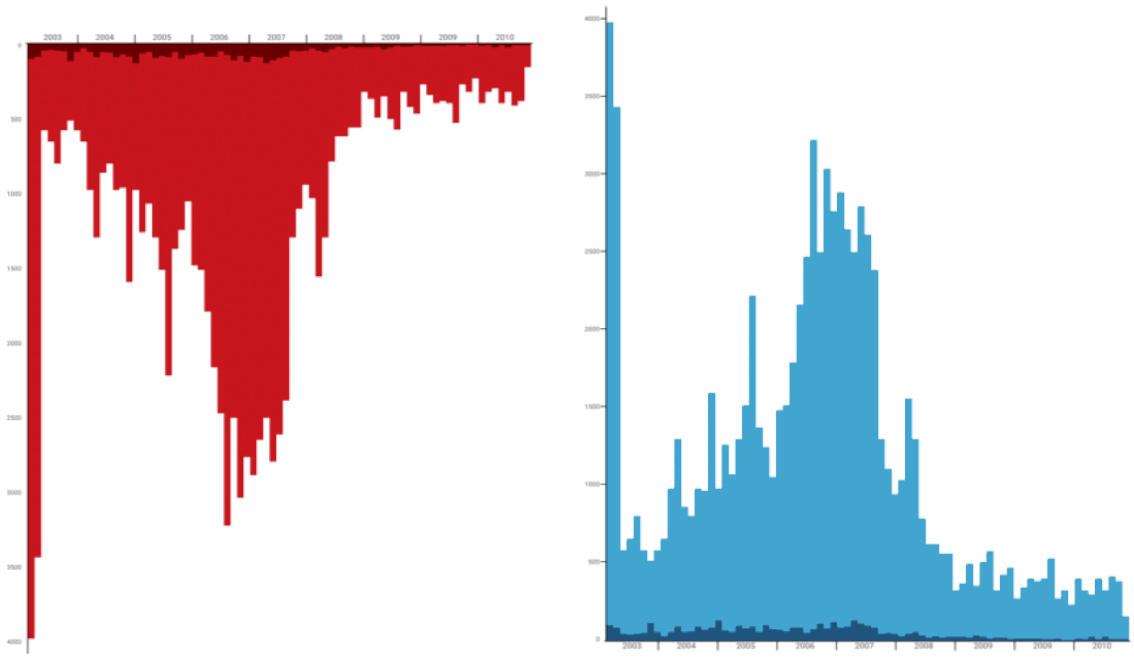
After performing the statistics on the big data, it is very important to present the result in a way that any human can easily understand it. Here visualization takes this responsibility. Visualization is a basic approach for creating a vivid mental image of an event. In general, visualization is to gather all of the scattered thoughts in one place and create an image of it. The main benefit of visualization is it can present a messy thing in a beautiful ordered way, which can be understandable and readable very easily. At first glance, the visualization can give the main context of all the thoughts which are not even in order.

In data science world, this visualization becomes data visualization. According to Tableau[4], *"Data visualization is the graphical representation of information and data."* There are some popular and common graphics techniques that can be used for presenting the data. Data visualization makes it very easy to understand the data trends, outliers and pattern. With data visualization, it is also possible to present, how the complex data relationship communicates and the insight of the relationship. A successful data visualization recognizes the importance of the target audience's wants and queries, as well as their degree in understanding, and purposefully guides them to the appropriate conclusion [33]. So it is very important to choose the graphical items or techniques so that the meaning of a complex dataset becomes very appealing and non-distracting.

In data visualization, color plays a very important role. Robert Simmon, Senior Data Visualization Engineer of NASAEarth, mentioned color as a crucial tool to covey a piece of quantitative information to the audience in the documentation of the *NASA Earth Observatory*[5]. Colors cause a chemical reaction in human brain that results in an emotional response. They elicit ideas, memories, and associations with places, people, and events. Colors with long wavelengths, such as red, produce a faster identification reaction in the brain. Colors with shorter wavelengths, suppose blue, are more relaxing and yellow is a hue with a medium wavelength that draws attention. In Figure 1.8, two graphs can be seen. They are using same data points but there are two differences. The first version (Figure 1.8(a)) is using the opposite

[4]https://www.tableau.com/
[5]https://earthobservatory.nasa.gov/

direction for the point arrangement and used red as color. At first glance this image seems like an alert image and giving kind of creepy feelings as it looks like blood is dripping from the top. On the other hand, the second version (Figure 1.8(b)) is quite smoothing and relaxing. From this image it can be clearly seen that data points are reducing than middle of the graph. This is happening because each culture symbolizes a color for a specific event. The famous data-journalist David McCandless creates an color graph for each culture. He named it *Colours in Culture*[6]. This chart can help to select different colors for different events for different cultures.



(a) First version (b) Second Version

Figure 1.8: Mirror images with different color and arrangement[7].

Most of the time it is seen that the data visualization designer would use limited numbers of colors in a dashboard. Too many colors can cause confusion and it will oppositely affect the audience. If the dashboard is representing different dimensional then the number of colors could increase, but it is ideal to keep the number under 10. There are so many solutions can be rethinked to limit the color numbers. For this different chart types can be selected, or may be the data could be categorised in similar groups or may be important data points could be selected etc.

There are many data visualization types. But it is not feasible or lucrative to use all of them in one dashboard. So it is very important to know about the data type

[6]https://www.informationisbeautiful.net/visualizations/colours-in-cultures/
[7]https://towardsdatascience.com

which are needed to visualize. Some of the visualtype will be described below.

- **Bar Chart:** If the data changes over time and need to show the difference between data then bar chart is the best option. There are different kinds of bar charts that can be seen. The vertical bar chart is used for consecutive data. The stacked bar chart is for illustrates the comparison between different parts of data or the comparison between a segment and full data. The horizontal bar chart is similar to the vertical bar chart. But this bar chart is suitable when the text of layout is too much large. More details has been given later.

- **Pie Chart:** When the data category is small pie chart is best for showing the comparison between the data part. It can visualize both discrete or continuous data. When multiple pie charts needed to show the comparison then it is not best idea to use it. Ideally, it is said that if the category is less than six then pie chart can be used for showing the comparison. Ring pie chart is another version of pie chart. The most important object would be placed in the middle.

- **Line Chart:** For time series data line chart is the perfect visual tool. It is very easy to understand the trend of data from a line chart. If the number of sample is huge then it is not feasible to use line chart. A good visual board would not show more than five lines together. It is suggested to use bold lines for this chart. Sometimes dotted lines are used in dashboard. But dotted lines are very distracting and the whole dashboard loses the importance of the information.

- **Box Plot:** When it is necessary to show the distribution of data in those cases box plot works as magic. This will show the ranges between variables and measures. For showing outliers, box plot can be used. The most use cases of box plot is comparing distributions between members of a category of the existing data.

- **Scatter Plot:** This can be used when it is need to show a relationship between data points. Scatter plot is best for showing the correlation between two variables.

- **Heatmap:** Heat maps may show areas of interest or data lists by employing a strong feeling of color contrast to convey categorized data. Some more details has been given later about it.

- **Table:** While the emphasis is solely on reading numbers, the material may be overlaid on the table to make it easier to understand. Table will show the same type of information in row and column format. For making it more easire to digest coloring could be applied for categorized the data.

From the above discussion it is clearly understandable that data analytics consists of different fields. It is all about dig up useful knowledge from some scatter

information and show them into clear and user friendly way. Data analytics can be used in different areas such as - marketing , finance, customer service etc. This can be used to understand the trend of the data like, behavior and preferences of different consumers, condition of the market trends, evaluate the pricing strategies, customer service improvement and so on. Any business can gain deeper insight into their operations and make decisions by using data analytics, which will lead to get efficiency and profitability in a great way. This can be used for recognizing the growth by uncovering new markets, understanding customer needs, and developing more effective marketing strategies. Businesses can identify customer segments and target them with customized offers by analyzing their data. This analysis can lead to escalating in sales and profits. Furthermore, data analytics can identify and address problems before they become costly.

For improving the operational efficiency, data analytics can be used. Businesses can identify areas of waste and inefficiency, and develop strategies to reduce costs and increase productivity by using the analysis of different sources of data. Production processes also use data analytics to identify the bottlenecks and can take steps to solve those processes by developing streamline. Predictive models are getting more popular day by day. These models identify the risks and opportunities. Predictive models also use data analytics for doing their further work. By analyzing historical data, businesses can create models to anticipate future trends and events and make decisions based on those predictions. This can help businesses stay ahead of their competitors and create opportunities for growth.

Customer service uses data analytics to improve their service for a better customer experience. Business needs to identify customer needs and develop a solution to meet them, data analytics plays a great role. Data analytics can also identify customer complaints so that those can be solved to increase customer satisfaction and loyalty. Data analytics is an invaluable tool for businesses to gain a competitive edge, and increase efficiency. Data analytics will become increasingly important for businesses to get an incredible advantage as the amount of data is increasing day by day.

So it is clear that data analytics can give a deep insight of data and how it can help different types of businesses. According to the characteristics, there are four types of data analytics. In Figure1.9 four of them can be seen. Two of them work with past data and others with future data. The more analytics go upwards the complexity of the analysis level goes high too. Diagnostic and predictive analytics are in the medium complex region though their data type is not same.

Descriptive Analytics - What happened? As the name suggest, descriptive analytics will describe the existing data to answer "What happened?". This analytics can be called as the foundation of all of the other analytics. This analytics will transform the raw data into readable information to all of the audience of the analytics. This information would explain what already happened or what is happening in the organization. The description of descriptive analytics clearly states that, it is a manual process to examine the raw data. This analysis will identify the data pattern and its meaning by summarizing them.

There are two techniques basically, descriptive analytics used to discover the ex-

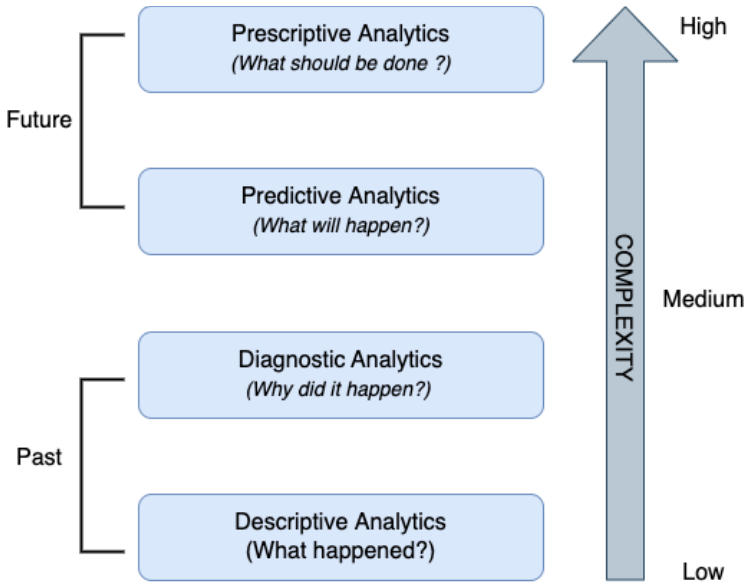


Figure 1.9: Data analytics types

isting historical data. Data aggregation will sort and combine data. So the data becomes more manageable to analysis. Another technique is data mining. This technique will explain the futher analysis. Data mining normally identifies the meaning and pattern of the data by searching. After doing so, the identified pattern will be analyzed to discover the main data content and then this outcome must be visualized through some common visual tools and techniques. Most common visual techniques used for descriptive analytics is different types of chart, table or maps. This visual techinques will clearly show the audience the trends of the data. A good example of descriptive analytics is - Google Analytics[8]. In Figure 1.10, a sample of the dashboard of Google Analytics can be seen. This sample dashboard mainly used Area and Pie chart. Area chart has been used for describing the session number, active user number, bounce rate etc. and pie is used to compare the new visitor and the returning visitors.

The usage of this analytics can differ from organization to organization. Some organizations can use it for comparing the performance at year-end to understand their place in the market. Descriptive analytics also can be used to understand financial trends. If the organization is related to learning material then it will also reflect the information about learning data. Descriptive learning analytics can identify the engagement and performance of learners. This analytics will help to detect the learner's participation level in a specific course. In Figure 1.11, a descriptive learning analytics from IntelliBoard[9] can be seen. This dashboard showing the an overview of different users. Like - how many users (teachers or learners) are online, When learners visit most frequently. Learners' overview, like - their progress and

[8]https://analytics.google.com
[9]https://intelliboard.net/

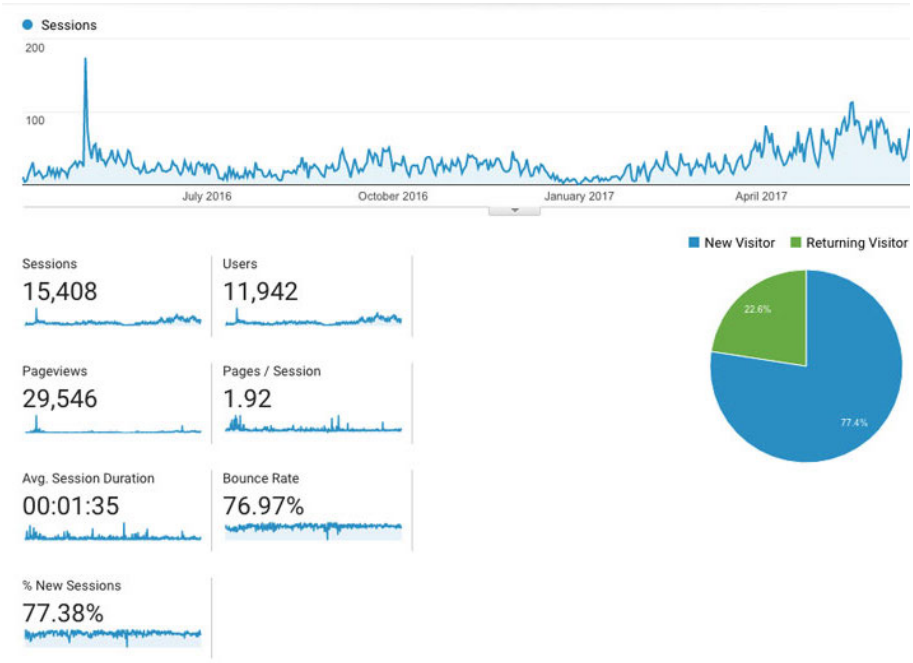


Figure 1.10: A descriptive analytics example - Google Analytics

score, and visiting time, also can be seen.

Diagnostic Analytics - Why happened? Metrics vary as a reflection of how organization is impacted by external factors and how users behave. To make better decisions, organizations need to understand what motivates KPIs and why they are changing. Only by understanding diagnostic analytics — which offers thorough insights about metrics changes at the rate of progress this be achieved. Diagnostic analytics to transform big data into informative data. It delves deeply into the "why" of the outcomes as opposed to just looking at the results. So, it is possible to say that diagnostic analytics sits between the goal and the data. After the process of descriptive analysis, diagnostic analytics is often seen as the obvious next logical step in data analysis. In learning analytics this analysis can be used by students to understand what might explain their success or failure; by teachers to explain actual learning paths and compare them to the a priori scheme; or by the institution to assess the impact of specific actions on student outcomes, such as extending library hours or developing blended learning [34].

Predictive Analytics - What can happen? Predictive analytics is used for guessing what will happen in the future. After getting all of the trends and patterns of data by doing descriptive analysis, a diagnostic analysis will define why this kind of situation is happening. This diagnosis will help to predict what can be happened in the future. As example, the diagnosis says that in winter the number of patients with flu increased dramatically, but in the hospital, they did not have enough measurements to take action to manage this kind of situation. In the future, this analysis will help them to predict what kind of measurement they need to take to manage flu patients

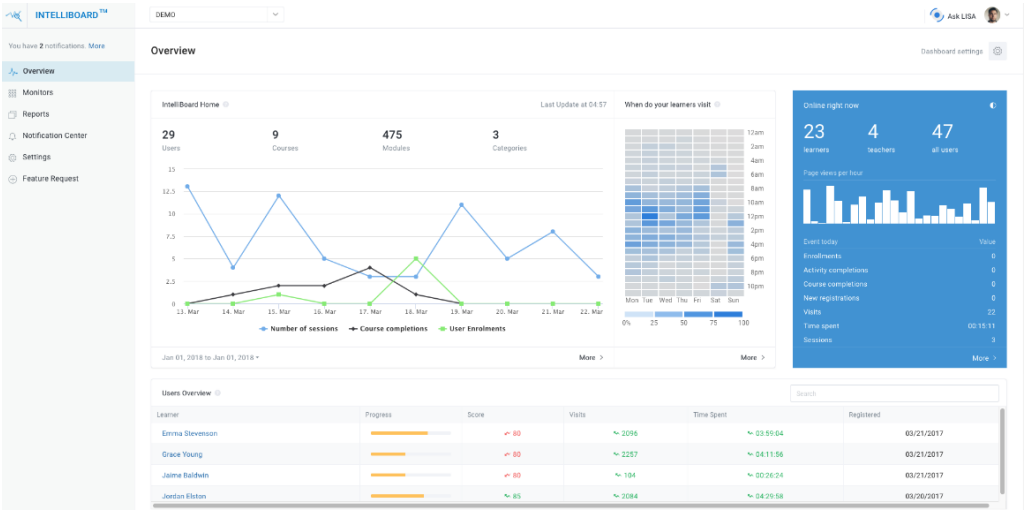


Figure 1.11: IntelliBoard's learning analytics dashboard

properly.

Predictive analytics can be done manually or automatically. For building up this automatic predictive analytics some predictive models are used. Most of these models use the relationship among different variables to guess what can happen in the near or distant future. These relationship can be figure out with regression analysis. Single linear regression will determine the relationship between two variables and for more variables multiple regression can be used. But before developing a predictive model it is very important to remeber that these predictions can be only estimated, and the accuracy of these models are highly depend on the data quality and how stable the situation is. So, for best result the data analysis should be done very carefully.

In Figure 1.12, a dashboard can be seen for online food delivery purposes from BoldBI[10]. The red marks of the figure are actually prediction about the revenue by month and busiest month and the week by total food order. With the revenue forecast the dashboard users can guarantee that resources are allocated correctly during busy months. Total order by week and month, heatmap is indecating the busiest days of week for food delivery service. This can help the organization to manage enough staff for delivering food on those specific days of the week.

There are some common techniques used for predicitive analytics. Among them - Forecast model, Outliers model, Classification model, Time series model, Regression model, Clustering model - are very popular. Forecast model estimates value of new data from past historical information. When there is no information about the existing data this model tries to use historical data for generating some numbers for them. Outliers model identifies the anomalies data from the normal. This model can help to detect fraud. When the analysis needs to be done on time related data then time series model is used. How data are changing over time this model will

[10]https://www.boldbi.com/

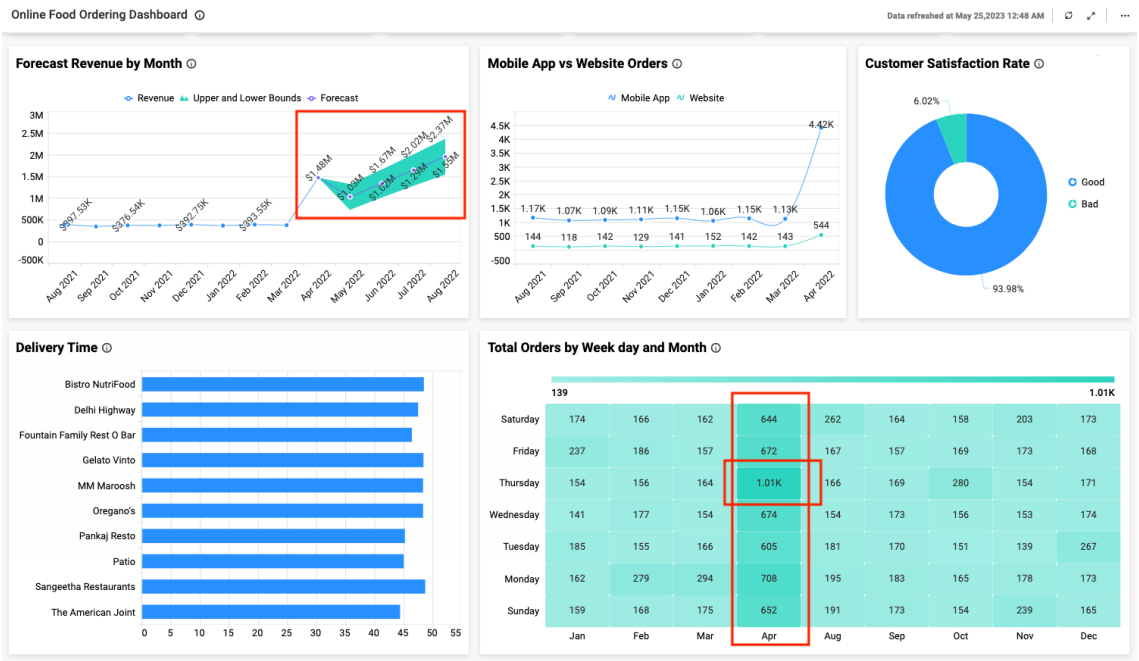


Figure 1.12: Predictive analytics dashboard for online food delivery

detect it. It is very important to know how strongly variables are related with each other. Estimating this strength of relationship can be done by regression model. The other models also do their parts in a predictive analytics.

In learning analytics, predictive analytics also can be used. This analytics can be used for identify possible struggles the learners are facing during their learning process. If these difficulties can be identified before the event happens then it is possible for the organization to develop opportunities for early support. In learning domain predictive analytics can also help to improve the learner's engagement. In Figure 1.13 a learning dashboard can be seen, where both predictive and prescriptive model used to find the course completion risk for a student. This will be discussed briefly in the next section.

Prescriptive Analytics - What can be done? The goal of this analytics is to suggest an action to reach to the targeted goal. Prescriptive analytics considers all conceivable aspects in a circumstance and recommends actionable takeaways. When making data-driven judgments, this form of analytics may be extremely valuable. This analytics goes beyond descriptive and predictive analytics by offering potential outcomes. Prescriptive analytics is typically utilized in major corporations seeking advise on issues such as inventory or supply chain management. These are frequently phrased as optimization and simulation issues in which a firm or manager attempts to maximize (or decrease) some objective (e.g. profit, efficiency, cost, employee satisfaction, etc.) while working within a set of resource, contractual, or other constraints.

Guessing the future is always tough and generate suggestions to avoid the prdicted

future is more complicated. In this sense prescriptive analytis is the most complex analytics to build up. Different kinds of machine learning algorithms, statistical methods, and computational modeling are involved in this kind of analysis. Prescriptive analytics evaluates all of the possible decisions an organization can consider to solve an issue. This will allow to check all of the combinations of decisions that can affect the future. Then as a human being it becomes their responsibility to take the best action to avoid any losing situation.

Prescriptive analytics are typically utilized in supply chain, routing, and operations when the number of decisions is too large for a person to manage efficiently - however they have been applied successfully in many other business domains. Like other sectors, prescriptive analytics also plays a very important role in the learning analytics. Researchers in [8], tried to design a learning analytics dashboard, where descriptive predictive and prescriptive models have used to give a brief overview of a student. Figure 1.13, the red marked boxes are the result of prediction model after the descriptive analysis (non marked charts). The lower red box of second column (marked as 1) is showing the estimated score the student will get in upcoming assigments and final exam based on the historial data of previous students. Then in the second red box, the model is showing the course completion risk and the key factors for predicting the result. The blue box is basically showing advise what can the student do for alter the predicted outcomes.

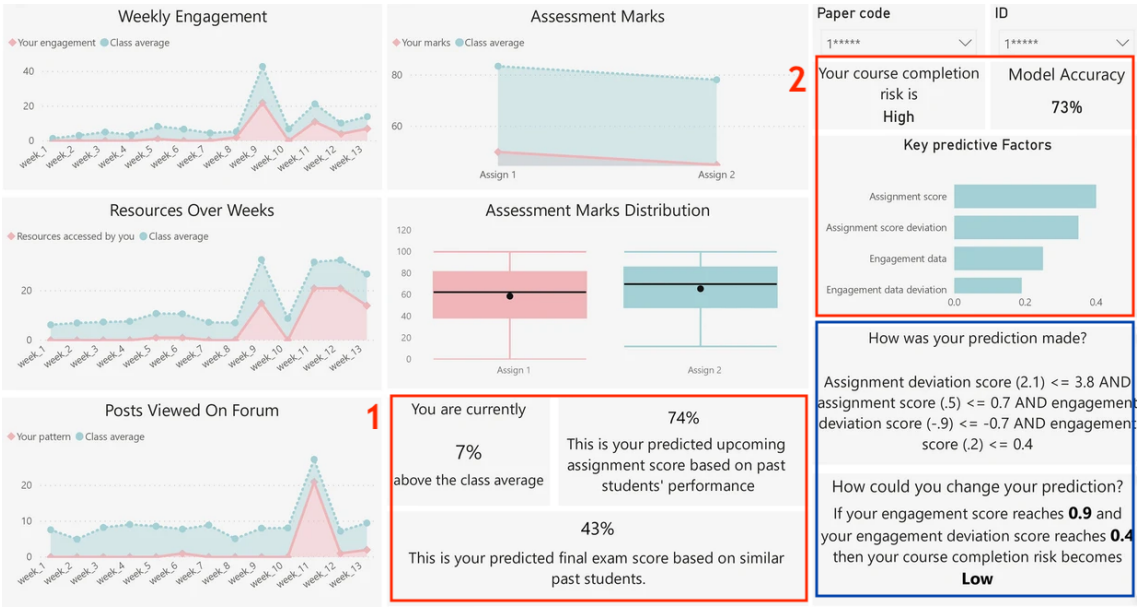


Figure 1.13: Learning analytics dashboard using Predictive and Prescriptive models [8]

1.2 Current State of Diagnostic Analytics

The diagnostic analytics procedure was carried out manually in the past. Presently, without the aid of technology, it would be nearly impossible for an individual to complete all the job, corporate performance, or financial analyses. The use of diagnostic analytics is widespread in a variety of sectors, including retail, manufacturing, finance, healthcare, education etc. By the use of relevant insights and visualizations that anybody can readily comprehend and utilize, this sort of analytics enables business executives to extract crucial information from their data. Some recent work on diagnostic analytics has been discussed below.

In order to comprehend operations and offer recommendations, railway asset management analytics makes use of cutting-edge methodologies and technologies for gaining deeper insights, more pertinent foresight, and relevant hindsight. These cutting-edge approaches include data mining, data discovery, text mining, categorization, forecasting, various machine learning algorithms, cluster analysis, visualization, and graph analysis, according to the "Gartner Glossary." Maintenance analytics plays a significant part in railway infrastructure decision support systems [35]. In this study, a diagnostic analytic system has been proposed by using historical rail profile data gathered between 2007 and 2016 for nine sharp curves on the heavy freight lines in Sweden. This analytical system has used anomaly detection methods to identify the root cause of unexpected rail wear behavior. For anomaly detection approaches this paper selects PCA (a spectra based approach) and LOF (a density based approach). The degree of an observation being an anomalous case is calculated in addition to the conventional techniques of PCA, by utilizing two applicable statistics. They are Hotelling's T^2 and Q. For the visualization of each approaches, boxplot has been used.

For many years, smart manufacturers have utilized modeling and simulation to evaluate their processes and offer decision assistance. Ever since the idea of simulation was first introduced, data analytics has been a crucial component of simulation. In the context of simulation, DA encompasses both input and output data analysis, both of which typically call for the processing of significant volumes of data. Applications for data analysis (DA) also enable simulation analysis by calibrating data, estimating unknown input parameters for simulation, and validating simulation outcomes. The use of simulation has therefore been promoted by DA, but DA may also benefit from simulation in a number of ways. The two primary categories of simulation's uses for DA in manufacturing are its direct use as a data analytics tool and its usage to assist other DA applications. [36] introduces a machining technique that addresses the issue by employing process plan data as input information rather than an NC program. Using object-oriented working steps, STEP-NC explicitly illustrates process sequencing and parameter selection. On the other hand, data from machine monitoring may be utilized to determine effects. Fundamental data are provided for measuring machining performance through MTConnect-based data contents. Hence, a paired set of cause and effect data may be provided by the STEP2M simulator for diagnostic analyses.

Data-driven diagnostic analytics are used in the current processes to assess well performance and expedite the process of finding under performing wells and inefficiencies. These data-driven diagnostic analytics were implemented on a digital oilfield workflow platform where data is aggregated from various data sources including continuous real-time sensor and model-generated data, non-real-time well data, well events, well test history, MPFM, interpreted PTA, reservoir simulation, well integrity, and tie-in data. Workflows and asset hierarchies are mapped with the analytics. Based on historical data, the linear regression approach is used to predict trends for water cut and GOR [37]. Based on well models, well-level allocation analytics enable comparisons between the in-flight export meter and terminal values at the same timestamp. Wellhead pressure estimation from the most recent working model was added for model calibration. Based on models and actual data, well monitoring and management diagnostics examine wells that are running in a critical or sub-critical state and seeing an increase in water cuts. It is possible to confirm the accuracy of the simulation data and lessen the uncertainty in the well models by combining reservoir simulation data, PTA, bottom-hole surveys, and estimated data from well models. Asset operators can respond more quickly to problems with reservoir performance management thanks to compartment- and reservoir-specific VRR diagnostics. The conventional model-based automated procedures used to identify wells for improvement were supplemented by these analytics. In order to execute diagnostic analytics in the first phase and to create a roadmap for its migration to a next-generation data-driven platform with better predictive capabilities, a digital oilfield solution platform has been used.

Nowadays, it takes a team to manage an organization's workforce. HR analytical tools may be used to manage personnel and analyze their performance online as a result of changing company needs and technological advancements. Most organizations needs to be concerned with a question like *"Why do employees should stay with our company?"* [38]. To answer this question, HR analytics can be used to determine the need and lacking of employees. HR analytics helped some big companies like - Google, Experian, Walmart, Johnson & Johnson, etc. to manage employee performance, understand the behavior of the employees and understand the retention and turnover rate of employees. Each company used a different HR analytic system to find out the root cause of their specific problem. Most of these analytical systems used data mining and data drilling. Some of them use different data visualization techniques to identify the weak area. The analytical team research the unorganized data then forecast the decision and present them in a prototype or visualize them in the different chart (pie chart, bar chart, etc.). When the visualization and prototype are developed the organization can take a good approach by testing it before launching it in the organization.

Talent analytics is a relatively new yet interesting and expanding topic in HR practices in Nigeria [39]. It provides a game-changing potential for businesses that are into human resources management as an organization, department, course of study, and discipline. This analytics may assist the management in keeping top performers on staff, comprehending their reasons for staying with the company, ap-

preciating their job contentment, and getting a sneak peak into their motivations. Determine remedial steps to solve retention concerns in the business with the use of talent analytics. A properly assessed performance aids in the selection of top candidates, boosts job happiness, and increases employee loyalty to the company. To help firms avoid the mistakes that led to unfavorable results in the past, diagnostic analytics use historical data to discover various anomalies. For diagnostic analysis drill-down, data discovery, data mining, and correlations are very common techniques to find out the anomalies of the data. By examining the possible causes of the detected abnormalities, hidden connections are revealed. Filtering, regression analysis, time-series data analytics, and probability theory can all be helpful for locating hidden flaws in the data.

The other booming sector for data analytics is Health Analytics. The second important health analytics technique is diagnostic analytics. With this strategy, healthcare organizations are analyzed in order to make them completely data-driven businesses that provide competitive benefits. In health sector diagnostic analytics answers question like *"why did a patient get worse even with the best care?"* [40]. For diagnostic health analytics most of the time data discovery and data exploration is used. Data discovery entails gathering and analyzing data from many sources in order to find hidden patterns, trends, and outliers. This could be called the first step of health analytics. This paper classified health diagnostic analytics into two, qualitative and quantitative diagnostic analytics. The data categorization is utilized as a machine learning algorithm with the goal of deciphering the reasons behind actions and behaviors. For data analysis, SPSS, Matlab, SQL, Java, Weka, Rapidminer, R analytics suite, and Python Scikit-learn can be used as analysis tools.

The examination of several healthcare platforms/frameworks that have been employed to date for the detection, diagnosis, and treatment of various chronic illnesses including cancer, heart disease, diabetes, and kidney disease is presented in this study [41]. The examination of several healthcare platforms/frameworks that have been employed to date for the detection, diagnosis, and treatment of various chronic illnesses including cancer, heart disease, diabetes, and kidney disease is presented in this study [41]. Previous research in the healthcare field have demonstrated that there are several associated diseases that afflict populations of people; for instance, heart valve disease has been identified utilizing cluster methodologies. Another illustration is the detection of breast cancer utilizing cluster approaches that identify patterns in both malignancies and benign tumors within tutor characteristics. In order to identify different characteristics of patients with heart disease, the K-means method is also utilized to cluster a set of patient records. healthcare sector uses a variety of data mining approaches, mostly rule-based, artificial neural networks, and decision trees, to classify a wide range of disorders. The tasks of class description, classification, association, prediction, clustering, and time series analysis are completed through data mining. One method for processing data in a multi-dimensional capacity is online analytical processing (OLAP).

The most prevalent and persistent issue in the global healthcare system is the overcrowding of emergency rooms (ER), which often results in significant consequences.

Performance and quality in healthcare may be determined by a number of factors and quantifiable characteristics, including equity, accessibility, availability, and timeliness. To identify problem areas and make recommendations for prospective ER performance enhancements, King Faisal Specialized Hospital and Research Center (KFSH&RC) used health analytics techniques by using two main KPIs, the ER LOS for ER patients and the percentage of patients who leave the ER unattended [42].

In the first phase of this study, eight variables have been identified, among them, Patient Acuity Level has significant statistical effects on the admission rate of ER patients. The analysis has been prepared in a tabular format with the percentages. As the summary of the analysis has been seen, when the acuity level (less serious conditions) goes down, the rate of admission becomes less. After getting this summary an investigation has been done. The explanation of the study came like this, many qualified patients visit the ER rather than their clinic visits because they may have trouble getting access to primary care or have to wait a long time for an outpatient appointment. Depending on this clarification, the executive manager of the hospital remodel a portion of the ER into a Fast-Track section with 20% of the ER bed capacity and limited patient admissions to those with at least acuity levels in this area.

Tools for diagnosing diseases at the point of care (POC) are crucial in the prevention of infectious diseases (like HIV, TB, Malaria etc.) as well as other long-term and acute illnesses. When patient demographic information is combined with test results data (produced by POC), a full dataset that may be effectively utilized to extract fine-grained surveillance information at both the individual and population levels is created. Almost any kind of analytics may be carried out when a large dataset is available. The data may be gathered, saved, and analysed in batch and real-time modes using Internet of Things (IoT) enabled POC devices and the big data analytics system to offer a full image of a healthcare system as well as to detect high-risk groups and their locations. The collecting, distribution, and use of data for customized healthcare have recently undergone a revolution thanks to various forms of real-time health monitoring devices. The notion of mHealth was made possible by the availability of a large dataset, improvements in analytical techniques, and emerging technology, particularly the usage of mobile devices.

The POC produced data, including test results, test duration, device location, warning and error, and quality control parameter, would never be recognizable, ensuring the rights and confidentiality of the person while still enabling significant population-level evidence to be acquired, which is a highly critical aspect of this form of knowledge extraction. In order to identify relevant information in extremely valuable POC produced data, this study [9] explores the possibilities of applying big data analytics in the healthcare area. This paper mainly used visual analytics and spatial analytics as advanced analytics to explore the knowledge.

A common data structure has been used for POC generated data (CD4 t-cell count). a hierarchy data about test (id, date and time, test type, count of t-cell), device (id, longitude latitude, cartridge id), test quality (warning and error) etc. This study proposed a diagnostic analytics to answer *"Why is a specific type of POC*

error absent?" or *"Why isn't a particular device performing to its full capacity?"* and *"Why do all test results that show CD4 levels below 500 originate from a single device?"*

Certain concerns and solutions are needed for the management and analysis of the data generated by POCs. Generally speaking, the system must be able to handle various forms and evaluate a sizable amount of data in both batch mode and real-time mode. When a complex system is seen holistically, batch processing is employed, meaning that the processing takes place after almost all of the data items have been fed into the system within a given time frame. As soon as new data items are fed into the system, processing and output production of the data begin in real-time mode. Because there would be enormous amounts of data to manage, the Lambda architecture is the best option.

This study used mainly two analysis. Location (positioning) and spatial connections (such as distance, direction, and connectivity between locations of the devices) were employed in spatio temporal data analysis for grouping and clustering the data, however, time of the test is primarily used in temporal data analysis for the same purpose. For diagnosis, this paper used pie chart and tabular view to answer to explain the root cause of the events.

Site	Total Tests
Lab102	383
Lab105	360
Lab139	359
Lab114	355
Lab127	354
Lab120	350
Lab107	350
Lab121	343
Hospital104	343
Hospital101	343

Figure 1.14: List of location of damaged devices [9]

The table in Figure 1.14, shows the list of location where the devices are counting the cell number less than 500 (CD4 $<$ 500). With this list it becomes easy to get the device location (as the location can be found in the dataset along with the device id) and can do further observation for improving the performance of devices in these locations.

Lately, the novel coronavirus pandemic (COVID-19) epidemic has posed a major danger to human health, life, productivity, social connections, and international relations. Big data and IoT technology have been crucial in this case in battling the epidemic. A few examples of the applications include the quick collecting of huge data, visualization of epidemic data, breakdown of pandemic risk, monitoring of cases reported, tracking of preventive levels, and proper assessment of COVID-19

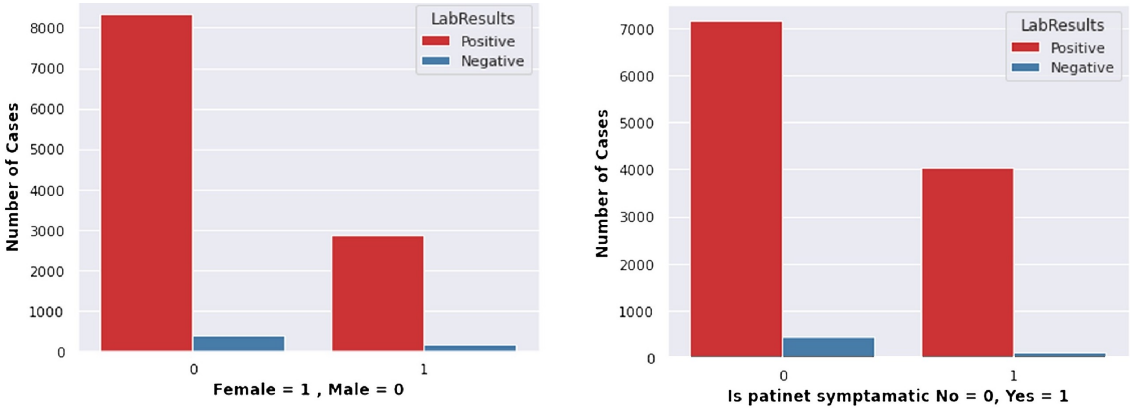
prevention and control. [10] presents a framework for health monitoring that analyzes and forecasts COVID-19. The framework makes use of IoT and big data analytics. Data visualization techniques are being utilized for descriptive and diagnostic analyses to offer insight into the various pandemic symptoms.

A structured data collection from several Khyber-Pakhtunkhwa, Pakistan, hospitals was used in this work. It has over 26000 patient records, both male and female, representing a range of age ranges. For study, prediction, and detection of a pandemic, many characteristics or symptoms are employed. The Lab Findings and the Patients' Survival following the Virus Diagnosis are the target attributes of the data collection. The majority of the dataset's properties are typical sickness symptoms including the flu, fever, sore throat, cough, etc. The majority of the features in the data set are categorical; for example, 1 for positive lab test results and 0 for negative findings for lab test results. The age element is numerical, while the category characteristic is the ultimate patient status.

Diagnostic analytics are used in the healthcare industry to examine data and establish relationships utilizing various characteristic data. For instance, it could reveal that the same viral agent is to blame for all of the patient's symptoms, including a high temperature, a dry cough, the flu, and weariness. The diagnostic analysis examines the illnesses' signs and underlying causes. Data discovery, data mining, and correlation approaches can be used for diagnostic analysis. Different visualization methods and multivariate analysis method have been used to point out the relation between lab results and different attributes (gender, age, different illness symptoms, etc.). In Figure 1.15 the multi variate analysis can be seen. Figure 1.15(a) shows the relation between lab results, number of cases, and gender. The relationship between "lab results," the number of instances, and various illness symptoms is depicted in the remaining example figures. Also correlation matrix (Figure 1.16) is used for defining the strength or volume of variables and shows the patterns and variation of each characteristic in the data collection. When the correlation is positive then it represents the attributes are positively related. That means when both of the atrributes are increasing in the same direction. In the correlation matrix figure, there are some negative values, these values are representing that when one value is increasing another one is decreasing. This study concluded that the pandemic sickness primarily affects persons between the ages of 30 and 60.

1.3 Motivation

Despite being debated for a decade, learning analytics in educational institutions has had minimal organization-wide adoption. Higher education institutions have experimented with student and teacher dashboards, but more widespread worldwide adoption models and regulations are required to mainstream learning analytics [43]. Currently, some research is going on with learning analytics in Australia, UK and US and those research results are proving the value of learning analytics to indetify the students who are at risk and are facing issues to complete their studies [44].



(a) Gender & Lab Results

(b) Patient symptomatic & Lab Results

Figure 1.15: Multi variate analysis for lab results and different patients attributes [10]

Learning analytics at the mega-level incorporates data from all levels of the framework, allowing for the detection of trends across institutions and improving educational decisions. The macro-level is concerned with institutional-wide analytics in order to streamline processes, allocate resources, and increase retention and success rates. The meso-level informs curriculum design and learning materials, improving course quality and alignment with intended results. The micro-level provides tailored recommendations and adaptive scaffolding inside the digital learning environment, while also taking non-educational variables such as emotional states into account for a thorough analysis. Table 1.2 showing the list of benefits of learnig analytics facilitators and learners for three perspectives, summative - learning phase completion data, realtime - ongoing learning phase and predictive - forecasting the learning phase [15].

Stakeholders	**Summative**	**Real-Time**
Facilitator	Comparing and analysing students, cohorts and different courses to increase the teaching quality	Monitoring the learning progression so that interaction can be increased
Learner	Understanding and analysing the learning habits and outcome and track the progress achieving the goal	Self-assessments with feedback

Table 1.2: Learning analytics benefits [15]

Though learning analytics has many benefits, still there are some challenges.

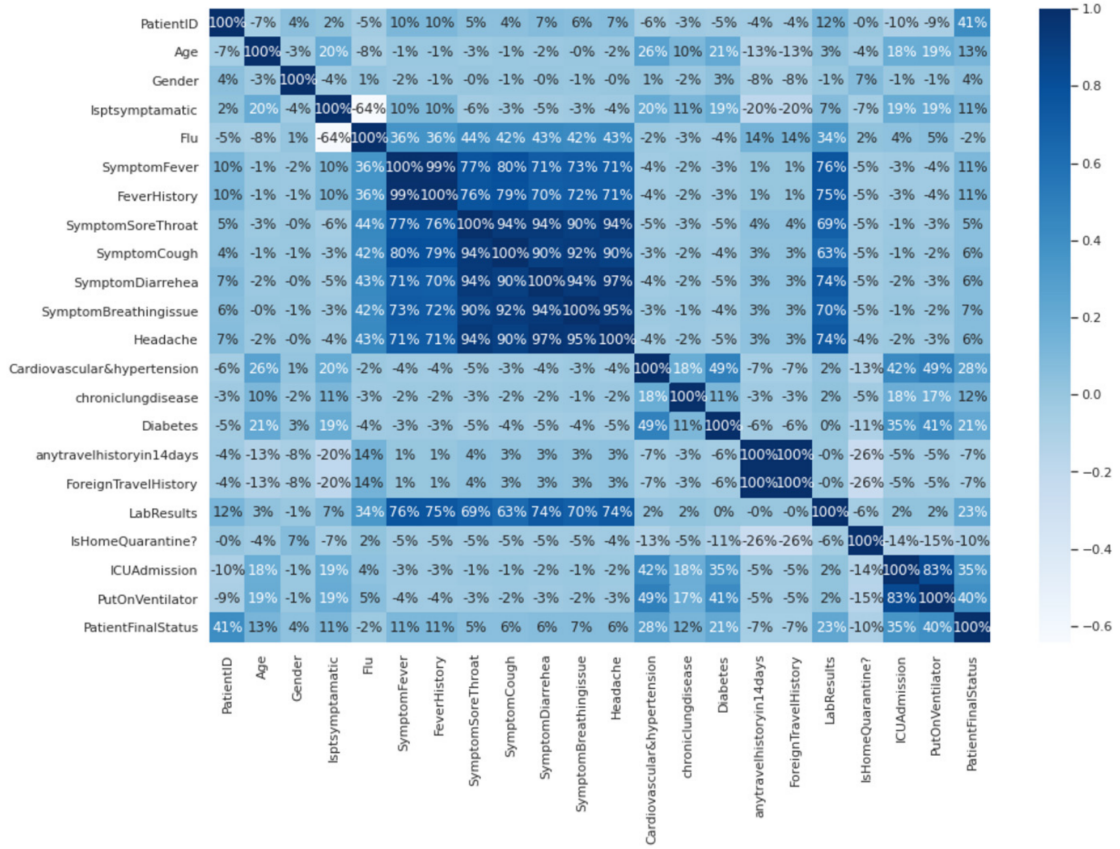


Figure 1.16: Correlation between different attribute of patients [10]

Among all of the challenges of learning analytics is the cycle of a learning analytics has to be action-oriented. Learning analytics is simply a feedback mechanism in which data must be fed up to desired objectives, fed back on where students are currently at, and fed forward on measures to improve. As a result, learning analytics should never be limited to providing feedback. When giving learning analytics to encourage self-regulated learning, teachers should aim to offer strategies for students to improve. Students, on the other hand, must cultivate reflective abilities that enable them to transform facts into action.

In traditional classrooms teachers face the pervasive but difficult issue to understand students' learning and thinking progress. Misconceptions and mistakes have the ability to provide a wealth of information about pupils' thinking and reasoning processes [45]. One of the most crucial elements in developing and managing classroom lessons is teachers' diagnostic practice with students who have difficulty. According [46], teachers do not undertake an in-depth analysis of their pupils' challenges with learning. Teachers only concentrate on subject matter and other difficulties, rather than on pupils' cognitive processes. Nowadays online learning is becoming popular day by day. In this learning method face to face mentoring is decreasing. But online communication is again making it easy by using E-mentoring

and blended mentoring [47]. This kind of mentoring tool helps students to stay motivated in their learning journey. Audience Response System (ARS), a blending mentoring technology, with pedagogical approach a qualitative exploration has been made to find out how the final grade has been affected by using this blending technology [TUC1]. An ARS can also increase the motivation, which is a vital feature, of the learners learning progress by implementing it as a mentoring tool. It helps the student to maintain self-regulated learning and motivated them to further learning [TUC2]. Even though an ARS can encourage student to stay motivated, it is important to understand the students' learning difficulties. A diagnostics analytics can help to figure out the reason behind these difficulties. The authors of *Learning Analytics Based Smart Pedagogy* [48] mentioned that, a learning analytics should able to diagnosis the - (1) Student's motivation (ex. Assignments submission time - first or always "last minute" submitter), (2) Students who are at risk, (3) The obstacles for the student success and (4) Students who need teachers help, While *SoLAR*, was describing the diagnostic analytics, it includes some features which can improve the learning environment. They are -

- Data analysis to inform and improve key performance metrics throughout the enterprise.
- Pattern analysis is used to create acceptable metrics.
- Reporting on equity access and analyzing successful student assistance programs.
- Metrics from learning management systems to boost student engagement

The research aims to develop a diagnostic learning analytics, Avenseguim[11], including (1) a visual dashboard, (2) an automated analytics system with diagnostic analysis methods, and (3) different course evaluation data. This research project used result data from ARS test, where about two hundreds students participate. Each test has three parts. In the first part, the students are taught techniques of the research. After that, the students can choose a research topic and present it on their understanding. The student can discuss tips to improve their understanding with the teacher after the presentation. The last part is, writing a scientific report on the topic they chose and the understanding they presented [TUC2]. Each part of the test has been scored separately and these scores have been used as the input data of this research implementation.

This thesis report has been divided into several chapters, covering everything from the fundamental of data analytics to a learning diagnostic analytics. The chapter introduction sought to provide the background of the overall project and explain how this thesis project fits into it. Also covered the project motivation as well as this

[11] A spell from the movie *"Fantastic Beasts: The Crimes of Grindelwald"* which turn an object into a tracking device.

thesis motivation. The goal of each chapter of this work has been briefly described in below -

Chapter 2 - State of The Art includes a review of the literature and study fields with a focus on the issues relating to this thesis work. This chapter discussed different learning issues and how the researchers solved them by using different technologies. After that, Chapter 3 - State of Techniques, discusses different technical tools and framework which have been used to develop this project.

This thesis project has been developed based on some methodes. Chapter 4 - Methodology will give a brief view of all of the methodology steps which has been taken to build this project. Chapter 5 - Implementation explains the implementation steps to develop diagnostic learning analytics.

The output obtained after the implementation are evaluated and analysed in Chapter 6 - Results and Evaluation. How accurately this analytics will work for the diagnosis which students need teacher consultation and attention.

Finally, Chapter 7 - Conclusion wraps up this thesis with the summary of the overall work along with the explanation of the possible future scopes based on the result from the implementation and what else could be added to make it more suitable for teachers.

2 State of The Art

Diagnostic learning analytics goes past standard appraisal strategies by leveraging the control of information to discover covered up designs, patterns, and conceivable ranges for development in a learner's travel. The state of the art in demonstrative learning analytics is continuously creating as propels innovation and information mining proceed to modify the teach. Learning analytics frameworks and innovations are developing more progressed, empowering for information examination and significant experiences. In this respect, this investigation points to examining the display state of the art in symptomatic learning analytics. This consideration looks to get the guarantee and pitfalls of this cutting-edge approach to instruction by investigating the foremost later investigate, unused innovation, and effective case considers.

2.1 Learning Analytics

Monitoring online one-to-one tutoring quality

The conventional face-to-face approach of education can be replaced in some cases with online tutoring. However, there is solid proof to show that just providing instructors and students with online engagement possibilities falls short of delivering the desired learning results. However, it presents a substantial barrier to keep track of the caliber of such instruction. [49] offers a method for assessing the effectiveness of online one-on-one tutoring sessions. With the use of a tagging interface, the suggested method creates data from online tutoring behaviors initially. Then, in order to discover developing pattern frequencies, these data are examined using a sequential pattern algorithm (CM-SPAM). Finally, decision trees are constructed to identify effective and ineffective tutoring sessions using the emerging sequential patterns and their frequency values. A total of 2,250 minutes of online instruction from 44 tutors were recorded. 26 male and 18 female tutors participated in the study, providing online instruction to students from eight different schools. Each tutor's performance was assessed based on one tutoring session, which lasted, on average, 50 minutes.

For tagging interface, seven behavior signs have been chosen after some research. They are - initiating self-correction, giving hints, planning proactively, clarifying and monitoring the questions, pausing after an appropriate time, writing on VCE and asking context-based examples/questions. While tutors were instructing students via the virtual classroom interface, these behaviors were noticed as they developed.

The interactive whiteboard and other elements that make up the online tutoring platform are used to display the course goals, subject questions, and plenary questions. In addition to using the pointer and whiteboard tools to write, draw, and erase, the learner and the teacher converse verbally. There is no video of either the learner or the instructor available, but if the audio is interrupted, the learner and tutor can speak through the messaging box that is shown on the side of the screen. By awarding effort points, sending an emoticon, or showing a photo, tutors may commend students for their hard work. A local SQL database was used to compile all tags and associated timestamps. Logs included the timestamp, tag name, human observer ID, and video ID. After that, the SQL database's log files were exported for use in modeling and data analysis.

With four metrics, the tutors are evaluated in this study. The first metric depends on the clarification or comforts the student's rating for the sessions. Then, students' outcome from the session for a specific teacher is computed. The third one is, three human evaluators will randomly assess and score selected sample of sessions of a tutor. Finally, the tutor's score on the subject matter is counted. An overall assessment score for each teacher was determined by evenly weighting these four parameters. Based on this totaled score for tutors, those who scored more than the average for all tutors were classified as effective tutors, while those who scored lower than average were classified as less effective instructors.

This study divided each session into time bins based on the mentioned organized learning architecture of the online tutoring sessions so that it could be seen at which point in a tutoring session sequential behavior patterns indicate changes.Technical checks up to two minutes, warm-up up to five to ten minutes, lesson goals up to ten minutes, topic questions up to twenty to twenty-five minutes, and lesson reflection up to five to ten minutes were the five sections of a typical session. To find the effective and less effective behaviors of the tutors SPMF, an open source java pattern mining library, is used. For faster vertical mining of sequential patterns that employ co-occurrence data, specially the CM-SPAM technique has been adopted in this study.

Calculate the frequency with which each sequential pattern emerged in each session and normalize these values in accordance with the total number of patterns in the session in order to construct a classification model that would assist in automatically classifying tutors into effective and less effective tutor classes based on their behavioral patterns. In 44 sessions, there were 44 total patterns, ranging in length from 83 to 212. In order to verify the effectiveness of each of these models, 10-fold cross-validation is used with the WEKA program to build a number of classifiers. By counting the frequency of the tutor behavior in their session it has been determined that tutors who are less effective do not exhibit acceptable hint-providing and proactive planning behaviors as frequently.

The sequence pattern analysis done by chi-square test to compare the differences among three most significant time bins (warm up questions, lesson object and topic question stages). The results shows that effective tutors have monitoring behaviors with appropriate pause. But the less effective tutors allows no pause and time for the tutees. The less effective tutors also shows six monitoring action in a same row. But

Sessions with excellent teachers do not involve such protracted monitoring action sequences. Another behavior can be seen for effective teachers, is self-correction which is not presented in the less effective tutors behavior.

Understanding learning progress by Digital Game-Based Learning

Learning analytics may give teachers or students useful data that aids in evaluating game quality as well as learning development, engagement, and enjoyment. To assess or forecast student learning results, digital game-based learning (DGBL) offers learning analytics [50]. In DGBL, players frequently have to use their subject knowledge or skills to make the proper option during gaming; their choices might show how well-versed they are in the knowledge and skills they are learning. Many people consider DGBL to be an effective substitute for conventional classroom mathematics training. Abstract mathematical topics may be illustrated and shown using DGBL. Science and math learning results, engagement, and motivation can all be significantly enhanced by DGBL [51, 52]. In STEM education, it can also boost self-esteem and lessen study anxiety. Many of the skills employed in DGBL, such as goal-oriented decision making, spatial navigation, and sequential thinking abilities, are mathematical in nature, according to researchers.

As students engage meaningfully with the symbolic depiction of abstract mathematical concepts, DGBL can foster higher-level mathematics competency. The focus of DGBL research has been automated assessment. Automatic assessment is used in educational game research as a covert evaluation that discreetly gauges student performance and fuels adaptive learning support [53]. By forecasting students' current skill mastery as shown by gaming activities, automatic evaluation in educational games tries to prevent disruptions to the flow state brought on by external measurements. Current research has had mixed success in providing educators with transparent, helpful assessments in DGBL, in contrast to developing research on learning analytics that has looked for methods to integrate automated assessments into DGBL [54]. A game learning analytics system employing the continuous conjunctive model (CCM), a particular kind of CDM, was created and tested to fill this need. To monitor pupils' progress, a DGBL application called "The Nomads" was implemented [55]. It develops adaptive expertise in rational number arithmetic. The study estimated learner skill mastery profiles by fitting the game log data to the CCM.

The gathering, examination, and visualization of player interactions with serious games is referred to as game learning analytics. The primary goal of serious games is to increase public awareness of social concerns, although learning is the common goal. Information regarding player behaviors in the games is frequently provided to various stakeholders through visual analytics and dashboards. These analytics' data may be used to enhance serious games [56], better comprehend player behavior and tactics, and enhance player evaluation. [57] presents a two-pronged approach for better understanding of the player learning patterns, 1) a visual dashboard and 2) an assessment approach for player interaction data.

The serious game utilized in this study is the First Aid Game, which is designed to teach step-by-step techniques in three emergency situations - unconsciousness, choking, and chest pain. This work used data of 112 players of age group of twelve to sixteen. The game lasted for 50 minutes and the session for the game placed in year 2017. The xAPI-SG format is used to capture interaction data. The actor that carried out the action, the verb capturing the action, and the object that receives the action are the three key fields found in xAPI-SG format. An additional field can be added to the dataset is traces, which represent the timestamp for capturing the exact time when the action occurred. xAPI-SG Profile defines completables, alternatives, accessibles and game objects as key concepts which are connected to the verbs.

The gathered xAPI-SG traces are examined using the visualization tool T-MON to provide hypotheses about player outcomes, behaviors, and how they affect player learning, as well as GLA variables for further investigation. The raw xAPI-SG traces gathered from player interactions with the game are fed into T-MON as a JSON file. Following such xAPI-SG analysis, T-MON creates a collection of game state metrics for each player that is updated with the player's subsequent statements. T-MON shows a default set of visuals that summarize the received data once all the data has been evaluated. The xAPI-SG statements supplied by the game may be processed to produce the output visualizations featured in T-MON without the need for any additional game-dependent data. Information on game and level progress and completion is visualized, together with scores, time, accessible used and skipped, choices made (including successes), and interactions with game objects. Three visualization techniques has been used - line chart (Player progression throughout time), bar chart (the maximum and minimum times for per completeble), heatmap (the number of interactions of players with the different game items) and stacked bar chart (Correct and wrong replies per alternative).

Accept or reject the provided hypotheses using an evidence-based evaluation technique that uses the GLA for prediction models whose outcomes show the impact of player actions on learning. The evidence-based assessment technique expands on the gaming interaction data that has been gathered to produce pertinent factors for player evaluation. To construct GLA variables that include the data gathered from player interactions, the collected xAPI-SG statements are evaluated. To retrieve the target variable for the prediction models, several surveys are also analyzed.

White-box models, which are more conventional basic models that offer information about the importance of the variables in the findings, are among the prediction models that were examined. Other, more sophisticated models were also tried (black-box models). Support vector regression (SVR), linear regression, regression trees, Bayesian ridge regression, k-nearest neighbors (kNN), and neural networks are among the prediction models that have been put to the test. Python was used for all of the analysis for the evidence-based assessment method. Scikit-learn was used for all machine learning models, while Pandas and Numpy were used for data processing and mathematical operations. The mean absolute error discovered for each type's tested most accurate prediction model. With the lowest mean absolute error

for predictions, a neural network was the prediction model that produced the best results. An SVR prediction model had the second-best performance and a prediction error that was acceptable given the range of forecasts. The study's findings can help educators create and deploy serious games in the classroom more effectively.

Additionally, GLA may be utilized to offer a perceptive look at how players move about the game and how their movement patterns could impact how well they play. Making sense of the data processing is difficult due to the enormous bulk of the game data. In order to make educated judgments about game design, interventions, and using games for teaching and learning, researchers are particularly interested in employing such analytics to assess the success of the games and develop a deeper knowledge of student learning processes [58].

For GBL settings, advanced data analytics approaches offer analysis techniques to better understand learner behaviors. The methods may be divided into three groups: (a) supervised models (such as decision trees, logistic and linear regression), (b) unsupervised models (such as correlation and clustering), and (c) visualization approaches. (e.g., display of gameplay pathways). In order to evaluate the association between student behaviors in GBL and learning performance—either in-game learning performance or external outcome measures—a number of research have been carried out. In order to help students in the seventh grade better comprehend biological evolution, [59] created an educational game and looked at the relationships between game performance, concept learning, and in-game behaviors. The findings of the correlation study showed a strong association between student performance and their in-game behavior. The higher the game score, which indicates a superior learning performance, the more frequently and thoroughly you examine the pertinent material.

The purpose of [60] was to comprehend the learning routes taken by middle school students when they utilized a digital learning game based on their behavioral patterns and the correlation between performance levels. It made use of the log information obtained from the digital science teaching game Alien Rescue. As an open environment, Alien Rescue is developed. As they choose how to go on with their problem-solving process, it promotes students' independent discovery, exploration, and knowledge acquisition. Setting their own learning objectives, cooperating with others, and managing their own learning are all challenges for students. Nine multimedia-rich tools are offered in the game to help students solve problems. While the Communication Center serves as the hub, each integrated tool has a distinct purpose to aid students in their overall problem-solving.

Three weeks of this online game served as the sixth-grade scientific curriculum for almost 4,000 students. These pupils' mouse movements are recorded, amounting to more than 3 million lines of unprocessed log data. These log files provide a priceless glimpse into how students learn by reflecting their in-game behavior in real-time. These entries with the timestamps and dates show how each student applied the nine tools to aid in problem-solving. The frequency and duration of each of the nine tools were determined using the cleaned and aggregated raw log data. These nine resources are used based on how frequently students visit them (frequency) and how

long they remain in each one (duration in minutes). The analyses utilize the average frequency and duration of tool usage.

Both statistical and visual analysis methods were used to examine student activity logs and performance ratings. All 4,115 students participated in a descriptive study of tool use, which looked at frequency and duration of tool use to gain a general idea of how students utilized the game's tools. After that, Spearman's correlation analysis was carried out to look for any connections between the major performance indicators of tool frequency, length of usage, solution success rate, and solution justification score. For students that submit at least one solution, this analytic approach has also been utilized to examine the solution success rate as well as the frequency and length of tool usage. It was also looked at how tool utilization related to the solution justification score. The reason a student offers for their answer is measured by the solution justification score.

The sample was split into three groups based on the students' solution performance, including solution success rates and solution justification scores, and then on their probe success rates in order to account for performance-related behavioral changes. These groups were labeled as high, medium, and low performers. Students who scored above the 75th percentile on each of the three performance measures were classified as belonging to the high-performing group; those who scored below the 25th percentile were classified as members of the low-performing group. The medium-performing group consisted of students whose success rates on both probing and solution attempts fell between the 25th and 75th percentiles. Afterward, Mann-Whitney U tests were used to examine the differences between these groups. The result shows that the high solution rate group employed a lot more tools than the low group in terms of frequency. In addition, the middle group used all nine tools substantially more frequently than the low group. The frequency of tool use between the high and medium groups did not differ significantly.

It was determined if certain characteristics may predict student performance and the link between probe performance and other metrics using Spearman's correlation and regression analyses, which were carried out as part of the investigation into student success with sending probes. In order to determine how students interacted with the game, route analyses were lastly performed. The paths were shown as a network of graphs using the network analysis application Gephi [61]. For edge ranking size, a weighted degree metric was utilized, and a directed graph was used to display the movement directions from one tool to another. Students who played the game for at least five or more days participated in pathway analysis.

The research's findings revealed a strong positive correlation between various tool usage and performance metrics, as well as a range of tool use patterns by high- and low-performing students at various stages of problem-solving. Importantly, these results showed that, as the high-performance group demonstrated, students were more likely to achieve when they used tools effectively and sensibly. A surprising discrepancy between the two performance indicators was also shown by thorough multiple analyses.

Eighty-four fourth through sixth graders from a public school district in West Al-

abama are participating in [55] where numbers of males are 35 and females are 49. The Nomads' game challenges need a fundamental grasp of mathematical fundamentals related to whole numbers, which is in line with the Common Core Mathematics Standards for fourth graders. Two stages of data gathering were used. Students participated in the second session alone. (45 minutes). The only game log data that was gathered and examined was that from the second session. Every time a player completes a game level, their replies are recorded in the game log data. Each log comprises six factors, including the number of tasks, the timestamp at when they were performed, how long it took to accomplish each job, the abilities required to do so, the accuracy of the response, and any specialized tools that were utilized. The Nomads also possess six math skills (in the analysis it is known as attributes), such as the capacity to solve mathematical problems using four different approaches (A1), flexibility in addressing math problems (A2), able to identify the "nice number" (A3) algebraic abilities (A4), ability of using ratios and logic to solve difficulties (A5), and understandings of area, volume (A6). The tasks to be done in this game are divided into nine categories, such as gathering berries, creating various weapons, hunting buffalo, etc.

It is crucial to determine which qualities are monitored by which game activity in order to match the CCM to our gaming log data and draw conclusions about learner attributes. Subject matter experts choose the Q-matrix that details the talents measured in The Nomads. The frequency of each task, the average time spent on each task, the equipment usage rate, and the accuracy rate have all been used in a descriptive analysis. This analysis shows that the most common task in mining for gems and minerals. Both internally and externally, the value of CCM has been assessed. Model-data fit is evaluated as part of the internal evaluation. For each student, the difference between the observed and implied total scores is compared. a histogram used to view total scores as well as the density plot of total scores suggested by the model. Since the observed and indicated total scores are comparable, the model may be able to match the data.

In order to determine if playing the game had an effect on learning outcomes, the participant's replies were split into two sets. One set of students is unable to accurately predict an answer to a question, while another can. Both the first and second halves of the data have been subjected to the weighted and unweighted MAD. All of the MAD values, which range from 0 to 1, were less than .25, indicating that the model can effectively fit the majority of the data. The participant's distribution patterns of five qualities in the first and second half of the game have been illustrated using a non-parametric curve by density and level of mastery.

The quantiles of the distributions of the level of mastery for each attribute estimated from the first and second half data were compared by using the method D2, which is a reliable substitute for the paired-sample t-test that uses a percentile bootstrap approach in conjunction with the Harrell-Davis estimator. Equipment's effect on the degree of mastery has also been looked at. All five qualities have correlations assessed between equipment use and competence level. At a.05 nominal level, the usage of scaffolding is substantially connected with Attributes 1, 3, and 4, but not

with Attributes 2 and 5.

With the help of the entirely new game The Nomads, this study investigated stealth evaluation in DGBL. Players' interactions with the pre-specified events were coded and recorded in the game log database using an event-based method that was used in the game to transmit data to the server. The continuous conjunctive model has been employed in the game system since the numbers in the game tasks are created at random and players' paths are arbitrary. The CCM predicts learner skill competency using continuous variables, providing educators and researchers with more accessible and diagnostic performance evaluations. The CCM does away with item-level characteristics and uses continuous variables to indicate participant skill mastery. This study explored the model-data fit and examined how playing the game affected participants' skill mastery profiles by fitting the game log data with the CCM.

Predict learners' learning progrss in MOOCs

Since the majority of students watch the same online lecture videos, analyzing the learning behavior of Massive open online course (MOOC) aficionados has emerged as a problem in the field of learning analytics. It is beneficial to undertake a thorough analysis of these behaviors, investigate different learning styles for students, and forecast their success using MOOC course videos. In order to forecast students' weekly performance and give teachers the tools to set up effective interventions, [62] uses a deep neural network, Long Short-Term Memory (LSTM), on a collection of implicit variables that were derived from video clickstream data. According to the findings, the suggested model's accuracy rate ranges from 82% to 93% throughout the course of the twelve weeks. In datasets from real-world courses, the proposed LSTM model performs 93% more accurately than the baseline analytical models, like - ANNs, SVM, and Logistic Regression.

The data for this study was taken from two independent MOOCs that the University of Stanford created and made available. The University of Stanford's (CAROL) team gathered the information. Online resources provide table formats and access methods for the CAROL-shared datasets. There are three tables in the table schema for each course. The video interaction and activity grade tables are taken into consideration in this investigation. Each log entry in the video interaction table includes temporal information on how students interacted with video events, such as clickstream events (such as load, play, pause, and speed change), learner/video identity data, and the course. The marks for the homework assessment are included in the activity grade table along with the answer choice and date.

This work divided the solution in three major steps. First one, data cleaning. To guarantee the correctness of the model output, this step is crucial. In this phase, data validity and integrity must be confirmed. The initial stage in data cleaning is to exclude any entries without a name or a video code, after which these records are mapped to each learner's predefined unique ID. Then it's time to get rid of the distracting, empty columns.

The second step is feature extraction where the resultant clickstream data were cleaned to extract implicit characteristics which are the primary events of play, pause, rate-change, seek ahead or backward, and stop, learners' interactions with the video player were recorded. Each time one of these events is triggered, a log entry is made containing the learner and video IDs, the event kind, the event's present time in relation to the video time, and a UNIX timestamp. The implicit characteristics of video-clickstream behavior that were gleaned from the raw dataset are the subject of this work. After the video-clickstream data was extracted, the records of the students were kept week by week. The video-clickstream activity from the prior week was added to the data for each week. The machine learning algorithm was given the transformed function, which created a vector of extracted characteristics for each video.

The last one is, a predictive model by using LSTM, which has ability to recognize long-term dependencies in time series data and is a specialty of deep artificial neural networks. A recursive loop seen in LSTMs enables the model to take into account both the current input and any previous ones. The repeating module, however, is structured differently. Three gates—the input, output, and forget gates—replace the single neural network layer. Constant Error Carousel, a fourth unit that is the memory cell, is also mentioned. BPTT is used to learn the LSTM parameters.

Based on how students engage with video clickstream data, the LSTM model is used to assess learners' performance. Each week, the clickstream data from the videos are piled one on top of the other to forecast how well students will perform in a certain week or course. As a result, the model layers get the weekly data stack of learners and process the data for each learner. The two final features are connected to the weekly quiz, which is taken as the probability of true values, and the features for timestamp t (X_t) input data are related to how the learners are engaged with video clickstream features. The first "attempt" shows if the quiz was taken; the other is the weekly quizzes, which each covered the information in videos (1,..., m) to create a performance measure at "time" t.

The data set obtained from the feature extraction procedure was converted into useful input data for the model, and padded vectors with a consistent shape were built. These vectors were then mask before being supplied to the model layers. Each course dataset was divided into training, testing, and validation groups of 60/30/10 each. Scikit-Learn, Keras, and TensorFlow libraries with Python have been used.

The shortest and most frequently connected with quizzes videos are those that were most likely to be seen, it is expected. More specifically, certain students' click-events showed they had varied objectives for seeking out or getting any information. In order to enhance the predictability of learners' performance, this research focused on the (LSTM) model to simulate learners' behavior in video clickstreams. With the help of this study, the suggested model's accuracy was improved when compared to baseline models, both in terms of its ability to forecast student performance accurately and to detect students at risk of quitting in the first few weeks. In this study, a weekly prediction was made using the extracted feature from each learner's video-clickstream data from week ith. Instructors can step in at this moment and

take the proper action. Overall, this distinguishes this technique from the other way more clearly, not just in terms of implicit feature extraction from video-clickstream data, but also in terms of predicted accuracy.

Diagnosis cognitive load in E-Learning system

Within learning systems, a learner's cognitive load is closely related to their academic success. In order to achieve the best possible learning experience, diagnostic data regarding a learner's cognitive load is helpful since it enables the learner to regulate and control their cognitive load in the e-learning environment. This [63] study used the Bayesian Network (BN) as a learning analytic tool to examine a customized diagnostic assessment for a learner's cognitive load in an e-learning system. To measure the three components of a learner's cognitive load - germane cognitive burden, intrinsic cognitive load, and extraneous cognitive load - seven hundred students's data from Cyber University has been used.

In order to assist an effective representation, the BN combines a graph model in conjunction with a graphical illustration. The graphic simply illustrates the idea of a finite acyclic directed graph. (DAG). There are nodes and edges in the DAG. The nodes are variables that can be observed or unobserved. Edges are the connections between different variables. A graph is defined as the pair G = (A, E), where A is a collection of nodes (variables), and E is a collection of edges, where an edge is a connection between two vertices.

Participating in this study are a total of 700 pupils. The "Introduction to Statistics Science Class" e-learning sample course consisted of fifteen classes on fundamental statistics. The ultimate academic achievements of the students are also evaluated depending on the outcomes of the midterm and final exams for the course. The ultimate academic achievement might be calculated by adding the results of the midterm and final exams to the standardized test scores. Based on educational level (which year), student age, gender, and employment status, a descriptive analysis has been conducted.

The measures, the three cognitive load components and Academic achievements, in the descriptive statistics are computed using mean, standard deviation, skewness, and kurtosis. These four variables were also the subject of a correlational analysis, which revealed that germane cognitive load was not statistically significantly related to academic achievement in this study, while extraneous cognitive load and intrinsic cognitive load were both negatively correlated with academic achievement. The conditional and marginal probability parameters for the BN representation were calculated using these data. *Netica*, a piece of software from Norsys Software Corporation, uses its *Learning EM* function to estimate the probabilities of the network. In this study, the BN was utilized to estimate diagnostic data on a learner's pattern of cognitive load with reference to their academic performance. This system also predict a learner's academic success based on their pattern of cognitive load. The results of this study indicate that the BN can gather data to determine a learner's specific cognitive load pattern and may forecast that learner's academic progress

based on that pattern. More precisely, depending on a learner's response to the cognitive load questions, the BN calculates the learner's extrinsic, intrinsic, and relevant cognitive load levels.

Among all of these learning analytics researches, it was hardly found about diagnostic learning analytics in an automated way which can help to identify some specific problems and help the organization to improve it. But some researchers have explained how a diagnostic learning analytics can help the whole organizational system (including teacher, student, institution) to recognize different issues. Those evidence has been explained in the following section.

2.2 Empirical Evidences on Diagnostic Learning Analytics

A diagnostic analytics reveal the root cause of an event that has already happened. Diagnostic learning analytics seeks to identify the underlying reason for incidents that occur in learning scenarios. In the contemporary world, online learning has surpassed traditional face-to-face instruction in popularity. Identification of students' learning and perseverance as well as success prediction, much like in a traditional educational system, are crucial in this type of system. LA is currently a very common instrument in many educational institutions. Each educational institution has unique goals and methods for modifying this instrument [64].

According to [65], a LA must diagnose student motivation (ex. assignment submitting pattern), the students who are at risk, obstacles to student success and if any student needs help from their teachers. Learning analytics should also be able to identify, whether a student is struggling with a particular lesson so that they can be supported by the teachers in time. Social Networks Adapting Pedagogical Practice (SNAPP) has been used as a diagnostic learning tool to identify isolated students by The University of Wollongong [66].

A diagnostic model for in-class peer evaluation that was suggested in [67] by utilizing social networking principles offered teachers a teacher oversight panel enabling task solution tracking and filtering of gathered information according to their diagnostic or educational interest. StoryTec is a software program used in this research for in-class learning. Some paper-based math tasks are transformed into math task scenarios using this application. A student then requests to evaluate his peer's response by confirming its accuracy and thoroughness and providing helpful comments. The recipient student sees the qualitative comments on a comparable screen arrangement.

When an instructor enters into the scenario, a special management panel has been added to the player program StoryPlay. It offers a filter-based search interface so users can view database responses by job or student, with or without comments. From the selection, the instructor can choose a specific answer. It is presented in the same manner as pupil feedback. The instructor has the option to review the

submitted feedback and offer personal comments to particular pupils. To avoid overlapping problems brought on by screen size limitations, the control panel can move up and down. A crucial component of classroom learning that should also be taken into account in digital settings is the incorporation of social networks. This research examined important design issues and the advantages and potential of such learning tools for instruction and learning.

The "Teachers' Diagnostic Support System" (TDSS) presents a task for teachers in their everyday job to address heterogeneity in the classroom by tailoring the curriculum to the requirements of the students. Teachers must evaluate each student's unique characteristics, including learning requirements, learning needs, learning experiences, and learning progress, as well as the characteristics of the learning setting, in order to provide adaptive teaching in the most flexible manner possible. TDSS is a client-server based mobile application, created by researchers at University Hohenheim in Stuttgart, helps instructors adapt their instructional strategies to the diverse requirements of their students [68]. With the help of this application, instructors can access and analyze data before, during, and after class to improve their methods of instruction and plan future lessons and learning resources. .NET Core 2.0 was utilized as the foundation for web services in order to build this app. Microsoft Azure serves as the server, and Angular was used for the front end. Using C#, application development interfaces were developed. MSSQL was used to build the database, and Chart.js was used to create the charts for the graphical data analytics [69].

In the area of learning analytics, many educational models are tacit and do not openly influence student behavior. Experiential learning, joint learning, and the learning analytics process paradigm are all included in the Team and Self Diagnostic Learning (TSDL) [70]. The TSDL framework seeks to promote congruence between learning metrics and the learning design while theoretically clarifying the learning process. The four phases of the TSDL architecture are shown in Figure 2.1. The learners must pose and respond to reflective queries in the stage "self and team reflection and sense-making" in order to diagnose their learning style. Regarding the visual analysis, there were three components: a radar map that compared oneself, peers, and overall; data displayed in a table; and a rating for general similarity. The radar chart served as a potent visual comparison tool for students to compare their own, peers', and general ratings. To see the real figures, the table was also helpful.

The No Child Left Behind Act of 2001's demand for more formative assessments in educational institutions has in large part contributed to the high level of interest in cognitive diagnostic models (CDM). A discussion of an Educational Companion App (ECA) from ACTNext can be found in [71]. There are six useful sections in this program. The third module of this software is a CDM-based diagnostic model that identifies students' areas of vulnerability using data from the app's data storage, the Learning Analytics Platform (LEAP). In a diagnostic situation, digital games provide a fun atmosphere for evaluating pupil proficiency, including skills and misconceptions. An instructional video game can be used to observe pupil performance using a dynamic Bayesian network modeling method [72]

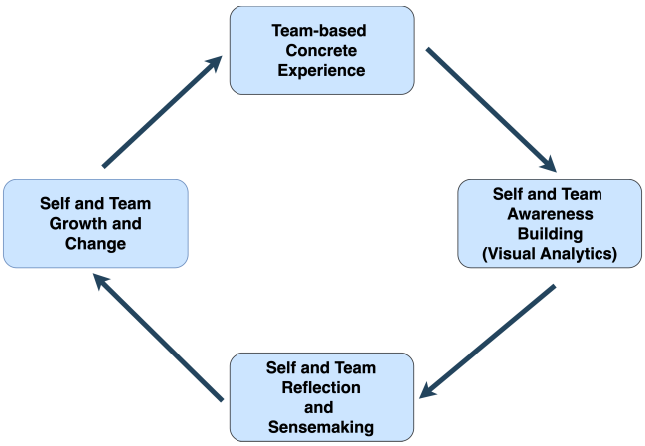


Figure 2.1: Team and Self Diagnostic Learning Framework

In the past 20 years, there has been a significant rise in the number of college students who have been diagnosed with Learning Disabilities (LD). [73] tested various diagnostic models (such as the IQ-Achievement discrepancy model, the DSM-IV model, and the model developed by Dombrowski et al.) to identify students who have LD and came to the conclusion that an LD diagnosis is unlikely to disclose a student's academic skills. Data from 336 individuals were used in this study. Means, standard deviations, and exam score categories have all been used in the data analysis as the input of the diagnostic models.

3 State of Techniques

In this chapter, some of the technologies have been discussed which are used in this study. This chapter has been divided into three parts. In each part, the state-of-art of each technology has been illustrated.

3.1 Backend

The portions of a computer application or program's code known as the backend (or "server-side") enable it to function but are inaccessible to users. The backend of a computer system is where most data and operational syntax are kept and accessible. Programming languages often make up one or more of the lines of code. Any functionality that has to be accessed and navigated to digitally is included in the backend, which is also known as the data access layer of software or hardware.

Basically, the backend is in charge of data storage, data organization, and making sure that the client-side functionality is effective. When the frontend and backend are in communication, data is sent and received for a web page to be displayed. The backend also includes tasks like writing APIs [74], building libraries, and interacting with system elements devoid of user interfaces or even systems of scientific programming. Developers can connect apps to cloud storage via the backend [75].

Requests from clients must be received by the backend. It processes the incoming request and ensures that the proper data is fetched and linked to the right user. The customer receives a reply after that. With frontend code written by a frontend developer, the required user data is shown to the appropriate user who has access to it in a pleasing visual manner. The server, a program that watches and waits for incoming requests from the client, makes up the backend. It sends database queries to interact with the database. The server retrieves the appropriate data that is required when the database provides information. The requests are processed by backend server-side scripts created in a backend scripting programming language. The database, which is sometimes referred to as the brain of an application, receives a response from the server.

There are several languages used for developing backend. Each language has it's own advantages. Depending on the project type the backend language is selected for developing a complete solution. But most commonly there are two types of programming concept, Object Oriented Programming (OOP) [76] and Functional Programming [77], which helps to decide backend programming language. Common examples of OOP backend programming include PHP, Java, Ruby, and Python.

3.1.1 Python

A high-level interpreted programming language is Python. It is one of the most widely used programming languages that supports OOP, structured programming, and functional programming. Python may be used for high-level programming tasks including web applications, machine learning, and data analytics. Python is a robust, slick, and simple to read and comprehend programming language. It assembles information and operations into collections called objects. Additionally, it is free software with a single standard implementation [78].

Python promotes clean code design, high-level structure, and packaging of various components, all of which provide flexibility, consistency, and a quicker development time as projects broaden in scale. When an application has to develop and extend, Python's pluggable and modular design enables the project to thrive while maintaining manageability. Python offers the fundamental building blocks on which an application may be built [79].

Python's versatility and readability make it useful across a variety of fields. Python may be used to highlight the key concepts of biological computing [80]. In health informatics, doctors also utilize Python to create and manage big clinically important datasets [81]. A web-based automated spine segmentation approach for diagnosing back pain has been created utilizing a deep learning model that makes use of Python, Keras, and Tensorflow. The project also makes use of the Python-based Flask web framework [82].

Additionally, Python includes modules and packages for distributed parallel computation. An all-purpose Python package called MPI for Python offers bindings for the Message Passing Interface (MPI) standard. The Portable, Extensible Toolkit for Scientific Computation (PETSc) libraries are accessible through PETSc for Python [83]. Since the middle of the 1990s, Python has gained popularity among mathematicians. The three most well-known projects for mathematical and symbolic computing are SymPy, mpmath, and Sage [84]. Python is an interpreted language, thus instead of converting a program or script into machine code, it may be written and executed immediately. Python is growing in popularity for Rasberry Pi as a result [85].

Python offers free pre-packaged installs for the majority of the popular platforms. Being a free and open source language, the software created with it may scale to thousands of computers without incurring any additional costs. Python is growing in popularity in the field of image processing as a result of these factors [86]. Scikit-image is an image processing package that offers tools and methods for usage in academic, professional, and commercial settings [87]. The mixture of low-level libraries with clear high-level APIs makes Python the most popular language for scientific computing, data science, and machine learning. This increases performance and productivity [88]. Evidence-based storytelling is at the heart of data science, and this sort of procedure calls for the right equipment. The Python data science toolbox is one of the most advanced settings for performing data science [89].

Due to the recent hoopla around data science, data analysis has grown in pop-

ularity. Data analysis is the process of obtaining information from data. Natural language processing, machine learning, and statistics can all be employed for this extraction [90]. Information extraction is made simple and meaningful with the help of certain free and practical Python programs. NumPy, SciPy, matplotlib, pandas, and other well-known Python data analysis packages are only a few examples. Data analysis may also be done using machine learning. Scikit-learn is a Python package that is frequently used for machine learning techniques under uniform data and modeling process rules, making it a useful toolset for statisticians studying education and human behavior [91].

3.1.2 Flask

A Python-based "micro web framework" is called Flask. The term "microframework" was coined since it does not require any particular equipment, programs, packages, or libraries. It fits into only one Python file. A database abstraction layer, form validation, or any other feature where alternative libraries currently exist that can handle it is not included in Flask. To add such functionality to the application as if it were built into Flask itself, Flask enables extensions. Despite Flask's "micro" size, it may be used in production for a multitude of purposes.

Flask differs from other frameworks in that it gives developers complete creative power over their apps. The phrase "fighting the framework" has a solution in flask [92]. Flask supports relational databases, NoSQL databases, and custom databases. A Model-View-Controller (MVC) framework may be created using Flask, which might make it simpler for users to create new projects and have a quicker completely loaded time [93]. Comparing Flask to Django, another Python-based web framework, it can be shown that Flask offers more speed, flexibility, fine-grained control, and simplicity [94].

The web is the most popular and quickly adopted networking tool that fits the needs of all types of users and offers a solution to every issue. A successful web page or application may readily draw visitors, which contributes to the success of various sorts of projects, like hand gestures detection. This type of web development project can use Flask to meet its technological requirements [95]. Python is a powerful language with many applications outside of script creation. RESTful APIs, which may serve as an application's backend, can be created with Flask [96]. A web service based on Flask may be used to create a data analysis platform. Flask will retrieve the data from a data source (such as MySQL), and the front end will utilize a separate route to present the analysis [97]. A Python-based AI back end program may be used to develop a facial recognition-based student attendance system. They are accessible over the web thanks to Flask and PHP [98].

Many machine learning algorithms require error-free data transmission from the user interface to the database. Because Flask is so effective, it can reduce the workload on any type of machine learning model (even KNN) [99]. The identification of malicious URLs is becoming more and more prevalent. Applications that employ ML and NLP to determine whether a URL is dangerous may be created using Flask

[100]. Additionally, healthcare services can leverage Flask API. The significance of the proposed model in [101] is to construct a flask API to consume the model and automatically rebuild the model anytime model performance degrades, at which point the model will automatically rebuild to choose the best model for analysis. Only a few minutes are required for the analysis of the Surface Plasmon Resonance (SPR) data. Therefore, the data transfer component of this type of analytical system must be quick. The ideal option for these kinds of time-consuming apps is Flask [102].

3.1.3 Pandas

Pandas is one of the key components that makes Python a influential tool for data analytic settings. This name was inspired by panel data, a phrase used in econometrics to describe multidimensional structural datasets [103]. For working with structured data sets used in many different domains, including statistics, finance, the social sciences, and many more, Python has a large collection of data structures and tools which is called Pandas. On such data sets, the library offers integrated, simple methods for typical data operations and analysis. It intends to serve as the basic framework for Python's statistical computing in the future. In addition to adding and expanding the sorts of data manipulation features present in other statistical programming languages like R, it acts as a good complement to the current scientific Python stack [104].

Python's Pandas package offers a high-performance data structure called a Data Frame, which is comparable to a table in a relational database[1]. Additionally, it may be used to read and write data to and from in-memory data structures and a variety of formats, including CSV and text files, Microsoft Excel, SQL databases, and the quick HDF5 format. A data structure may occasionally require a form change. Pandas' adaptability makes it particularly good in reshaping and pivoting a dataset. It is ideal to have a piece of a large dataset to do various statistical analyses since maintaining a large dataset takes a lot of work. Pandas' sophisticated label-based slicing, fancy indexing, and subsetting of big datasets may do this. Similar to a relational database, Pandas can merge and connect two datasets without the need for an additional library. The Pandas are very well-liked in a broad range of academic and commercial sectors (such as finance, neuroscience, statistics, web analytics) due to these kinds of remarkable features.

Data preparation libraries are required since everything in the field of data science is dependent on data. A wide range of features for input/output data formats, including Excel, csv, Python/NumPy, HTML, SQL, and more, make pandas the finest and most used Python library in this area at the moment [105]. GIS (Geographic Information System) data is extremely complicated in the big data space due to the high data volume and diversity, multi-dimensional information, and various data structures as the information may originate from many data sources [106].

[1] https://pandas.pydata.org

Pandas is hence the data scientists' favorite option for creating this sort of complicated dataset [107]. Pandas is the ideal option for creating the dataset needed for the visualization of marine geological data [108] Pandas is used for reading and manipulating tables from CSV format file.

Data shipping, query shipping, and hybrid shipping are three extremely common concepts in distributed data processing systems. These models determine whether data should be taken to one place or the query or both of them should exchange their place [109]. The preferred data shipping tool among data scientists is Pandas. Pandas make it easy to access the external data and user-specific functions can be defined by it [110]. The data structure provided by Pandas is called Dataframe, which looks like exactly relational database table with rows and columns. Pandas perform all of the relational database functions such as joining different datasets, managing and grouping categorical variables. But in addition Pandas also perform data alignment, handle the missing value and datetime comparison and last but not least implement different statistical models very efficiently [111].

3.1.4 NumPy

The processing of data using arrays or matrices is emphasized in the programming paradigm known as "array programming." When programming with arrays, operations are carried out on entire arrays or subsets of arrays rather than on single components. In particular when working with big data sets, this can lead to significantly quicker and more effective programming. For accessing, manipulating, and working on data in vectors, matrices, and higher-dimensional arrays, array programming offers a robust, condensed, and expressive syntax [112]. The main Python library for array programming is called NumPy. NumPy, or Numerical Python, is the abbreviation for the fundamental package for scientific computing in Python [103].

The N-dimensional array object, or ndarray, in NumPy is one of its standout features. It provides a quick, adaptable container for big data sets in Python. With arrays, you may conduct mathematical operations on large blocks of data using syntax that is comparable to that of operations between scalar items [103]. The form of an array, which is a tuple of N positive integers that specifies the size for each dimension, determines the number of the dimensions and objects in the array. Axes are used to specify the dimensions, and the quantity of axes is known as rank [113]. NumPy makes very easy to read and write normal text files and binary files. NumPy has it's own math tools, by using them the user can solve different mathematical problems like trigonometric, linear algebra, statistics etc.

MATLAB is a multiple programming paradigm used for parallel process application development ans scientific computing. But it is not feasible to use MATLAB sometimes due to its high cost. However, the trio of Python libraries—NumPy, Scipy, and Matplotlib—can readily take the place of MATLAB for tasks requiring modest to medium-sized numerical computations [114]. Current academics prefer Python and its various libraries, because of its many useful features, over MATLAB. Python is quickly taking over as the preferred language for computational

science [115].

The flexibility of NumPy makes very easy to use for scientific computing and data analysis [116]. Astrophysical simulations need quick, precise, and repeatable analysis and visualization. NumPy is the program that scholars choose to use for this sort of study [117]. Measurements made using magnetoencephalography and electroencephalography (M/EEG) are made from the weak electromagnetic signals produced by brain activity. It is a difficulty to use these signals to identify and pinpoint neural activity in the brain and calls for knowledge of physics, signal processing, statistics, and numerical approaches. This type of data analysis requires a highly quick and effective instrument. The module NumPy is widely used for M/EEG data analysis too [118]. A Python package called CuPy can do matrix calculations on NVIDIA GPUs. The NumPy compatible interface served as the foundation for the development of this library by Python programmers [119]. SciKit-Learn, a Python library for machine learning that was also built on NumPy, is another one that is quite well-liked [120].

3.2 Primary Analysis

A Jupyter notebook is web-based application for interactive computing [121]. This program, a server-client program, enables editing and executing notebook pages over a web browser. The Jupyter Notebook App may be run locally on a desktop without a connection to the internet or it can be set up on a distant server and viewed online[2]. According to "*Kaggle Survey 2022*"[3] Jupyter notebook is the most frequently used by data scientists and analists. Lorena Barba, a mechanical and aeronautical engineer at George Washington University in Washington, DC, claims that Jupyter notebook became the de facto industry standard due to its interactive interface, which allows researchers to run and edit their code. It appears as though the researchers and their data can have a meaningful conversation in their own language [122]. In the Figure 3.1 it is clearly seen that more than 80% people chose Jupyter notebook. The main advantage of Jupyter notebook is almost all python libraries are pre-installed in it and the user just need to import and use the libraries. A research of 1.4 million Jupyter notebooks from GitHub was conducted to better understand reproducibility. The results showed that 24% of the notebooks could be successfully executed [123], which is consistent with results from a prior reproducibility study [124]. Any scientific effort should be reproduced since it can be expanded. But simply extending is insufficient; the work must also be verified. Since the Jupyter notebook allows for dynamic updating and mistake correction, scientific work can be published by using this [125].

For big data programming classes, Jupyter notebook has been utilized as an interactive, hands-on training tool, free of charge to the students and without the need for an additional textbook [126]. It is challenging to convey the research-related

[2]https://jupyter-notebook-beginner-guide.readthedocs.io

[3]https://www.kaggle.com/kaggle-survey-2022

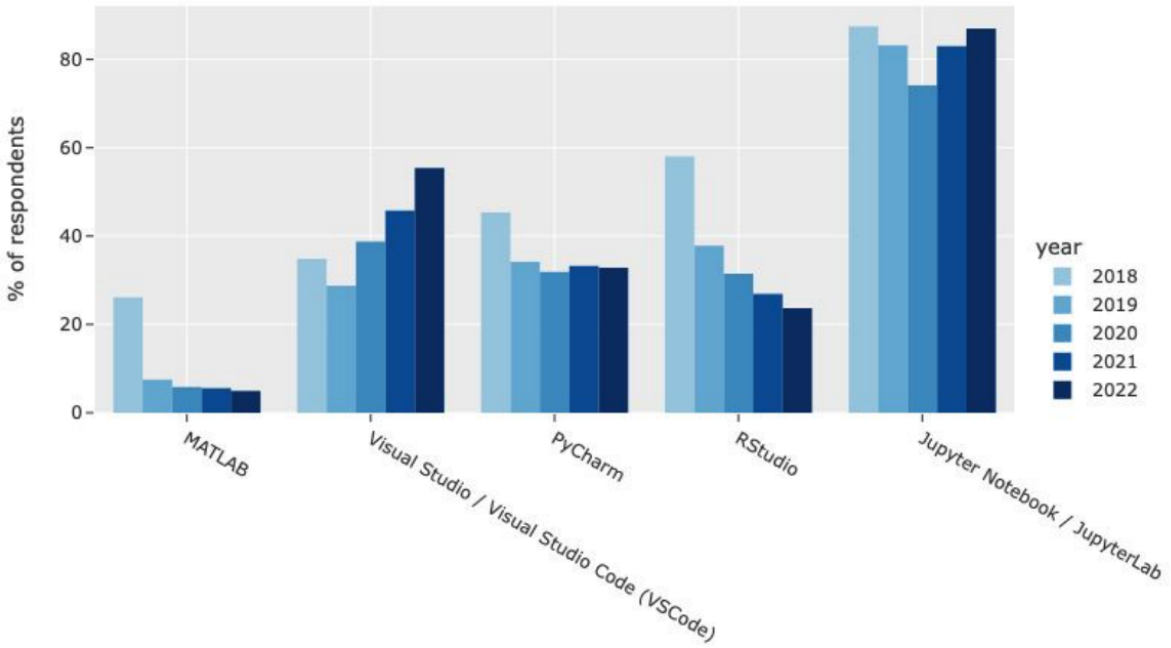


Figure 3.1: Survey of different data science IDE [11]

work to other users since Geographic Information Science and Systems (GIS) require hundreds of tools and databases. GISandbox has been designed on the Jupyter notebook to facilitate quick access to the GIS data-related work to the many users [127]. Jupyter notebook has been used by the Minnesota Supercomputing Institute to develop an all-purpose interactive tool for HPC (High Performance Computing) services for data exploration and visualization, training, and prototyping [128]. It is crucial for chemists to be able to process, analyze, and visualize data. Although spreadsheets may be used for this, handling huge datasets, multivariate analysis, and image processing is not feasible. In order to educate undergraduate chemistry students how to utilize a computer model to solve chemistry issues, nine notebook series have been constructed using Jupyter notebook [129]. The partial equation solution using neural network may also be explained using Jupyter notebook [130].

3.3 Frontend

The frontend is the presentation layer, where users may engage with the whole program. Frontend is another term for "client-side". A frontend can be assumed as the intersection of frontend development and web designing. In order to make the application more user-friendly and engaging, frontend developers incorporate concepts from web designers into the application. The main goals of frontend applications are two. The first is responsiveness, followed by performance. The front-end developer must make sure that the program is cross-platform compatible and that it behaves appropriately regardless of screen size. An excellent frontend illustration would be

a graphical user interface (GUI).

The images, typography, and visual layouts for websites or web apps are often created by a web designer using Photoshop and other tools. HTML, CSS, and JavaScript are used by front-end developers to make these designs interactive for users. They'll create dynamic features like contact forms, drop-down menus, buttons, transitions, and sliders for the user interface. Developers are restricted to the scripting and markup languages that web browsers like Google Chrome, Firefox, and Safari support because consumers interact with the front-end of online applications using web browsers.

Businesses that want to compete via innovation need to handle the front-end activities of new product development effectively. A comprehensive strategy that connects company strategy, product strategy, and decision-specific decisions may lead to success, and businesses that are adept at front-end tasks are more likely to produce ground-breaking goods that spur business expansion [131]. A useful technique for raising healthcare quality and reducing costs is clinical decision support (CDS). In order to evaluate commercial support and electronic health records, a taxonomy of frontend CDS tools is being developed [132]. In the integrated circuit board business, a study was conducted with 15 high-tech companies. This study focuses on important phases of product evaluation, defining product features, and front-end development using fresh concepts [133].

The foundational languages for front-end programming are HTML, CSS, and JavaScript. A web application may be created fairly easily thanks to the abundance of JavaScript frameworks that are available for front-end development and which also support other languages. The most popular Javascript frameworks utilized at the industrial level are Angular, React, and Vue. The frontend retrieves data from the backend and displays it in a user interface (UI) that is human understandable. APIs are the most popular approach for handling backend and frontend interactions. REST APIs or GraphQL can manage the APIs. The format after data retrieval might be in JSON or XML. Therefore, all of these technologies and approaches must be the emphasis of a frontend developer. The technology chosen may differ from project to project. Diverse project requirements call for diverse technology selections.

3.3.1 ReactJS

ReactJS is a front-end Javascript library that is free to use and open source [4]. It has been employed to create concepts for interface-based components. Applications with a single page and server rendering can be created using ReactJS. Making interactive user interfaces is simple. Create straightforward views for each application state, and React will quickly update and render the appropriate components as the data changes. ReactJS creates self-contained, encapsulated components, which are subsequently assembled to create intricate user interfaces. Rich data can be easily sent

[4] https://legacy.reactjs.org/

through the app, and state is kept out of the DOM since component functionality is defined in JavaScript rather than templates.

ReactJS is a new web development tool that is gaining popularity every day. Reusable UI components may be deployed with ReactJs. ReactJS serves as the view in the Model View Controller (MVC) architectural paradigm. The components don't need to communicate with the main DOM because of the effective and lightweight document object model. The virtual DOM functionality makes ReactJS a very effective performer. This virtual DOM is displayed first if any component is changed [134]. Sometimes additional libraries are used with React apps for state management and routing. Facebook uses a type of engineering called Flux when it collaborates with React [135]. There are many different components utilized in a large and sophisticated program. React advises moving the state to the component at the top and passing it to the nested component via properties. It aids to some amount, but as the number of components rises, it gets more difficult. Redux is an open-source JavaScript library for controlling and centralizing application state as the React application becomes complex [136]. React Hooks were introduced with React 16.8's release. Classes were the only way to add states to a component before then. However, as a result of the switch from classes to functional components, hooks made it possible for the functional components to be used inside the state and other features [137].

A web application utilizing ReactJS as the frontend framework has been created in Indonesia for Covid-19 vaccine booking and locating the closest vaccination site [138]. Even in the presence of recombinant sequences, Datamonkey offers a user-friendly online interface to a vast array of cutting-edge statistical algorithms for calculating synonymous (dS) and non-synonymous substitutions (dN) and detecting codons and lineages under selection. ReactJS was utilized in the second iteration of this web application to combine different reusable, encapsulated page components into a single, seamless document [139]. A billing method has been created in India for small business owners who lack the financial means and technological sophistication to purchase technologies like OCR. ReactJS and Flask have been used to create the system's GUI [140].

3.3.2 CSS

The display of a page created in HTML or XML is described using Cascading Style Sheets (CSS), a stylesheet language. When used in other medium, such as voice or paper, CSS specifies how items should be shown [5]. Users with limited vision can design their own large-font, high-contrast style sheets and instruct the browser to replace the default style sheets on websites with the ones that best suit their requirements and aesthetic preferences. For writing CSS file, there is no need to have any extra IDE, it can be written on any textfile. Some styling sheet can be downloaded from different website too.

[5]https://developer.mozilla.org/

A style sheet is a compilation of stylistic instructions that specify how an HTML page should be displayed to users in a browser. The developer may set styles using CSS, including the size, color, and text spacing, as well as the positioning of text and pictures on the website. The ability of style sheets to cascade is an important component of CSS. In other words, a document might have numerous different style sheets connected to it, and each of them can affect how the document looks. In this method, a reader may attach a personal style sheet to modify the appearance of the page to account for technological or human restrictions, while an author can construct a style sheet to indicate how the page should look [141].

Additional CSS analysis is being done, such as finding duplicate [142] CSS files or detecting errors [143], which is greatly enhancing the style industry. Personalized dynamic CSS can be for designing different web application for mobile devices [144].

3.3.3 Material-UI

Material UI is a very popular React UI package that helps developers to create user interfaces fast and effortlessly which are attractive, responsive, and accessible. This package has been created on the Google's Material Design [145]. It offers a high collection of reuseable user interface components and style sheets[6]. Google created Material Design for giving the same user interface experience across all platforms and devices such as computers, tablets, and mobile devices. Material Design employs a visual language that uses grid-based layouts, strong typography, and vibrant colors. Material UI is developed for ReactJS. It includes pre-defined components (such as buttons, forms, dropdown, textboxes etc.) that may be reused across the program as React is popular for its component based paradigm.

One of the primary advantages of utilizing Material UI is that it gives the application with a uniform and unified design language. This implies that the user interface will appear and feel the same across all displays and devices, which may improve the user experience and minimize the cognitive burden. Material UI also includes a suite of responsive design utilities that enable developers to easily create interfaces that adjust to various screen sizes and devices. Another advantage of Material UI is that it is extremely configurable. Developers may easily alter Material UI's default styles and themes to fit the appearance and feel of their brand or application. In Material UI, there is a feature called "Theming", by using this developers can customize the application with their own design [146].

Material UI is a powerful and adjustable UI package that allows developers to create attractive, responsive, and accessible user interfaces fast and effortlessly. It includes a set of pre-built components that are similar to Material Design principles, which might assist to improve the uniformity and usability of the user interface. Developers may quickly adapt the appearance and feel of their brand or application because of its flexibility.

[6] https://mui.com

3.3.4 Axios

In web applications, it is very common to make HTTP requests. For doing this, it is very important to choose a powerful HTTP client for the application. Axios is a very powerful Javascript library that is a promise-based HTTP client. Among developers, Axios is becoming popular day by day because of its powerful features and easy usage [7]. With Axios, developers can send and receive data by using API. Axios works with all current browsers and has a simple and beautiful request syntax. It may be used in both browsers and server applications.

Axios has capabilities such as interceptor support, automated conversion of response data to JSON, support for HTTP cookies, request cancellation, and much more. Axios has capabilities such as interceptor support, automated conversion of response data to JSON, support for HTTP cookies, request cancellation, and much more Axios also supports, request cancellation. Sometimes because of network connection, there are possibilities to not get data for a long time like 3 to 5 seconds. That time this request needs to be canceled otherwise the application would work incorrectly [147]

3.3.5 PlotlyJS

Plotly is a popular data visualization python framework that allows users to create interactive dashboards with different types of charts, graphs and other visual components. This framework can be used with a wide range of programming knowledge of users. This flexible feature makes this framework work with not only python but also Javascript and R. Plotly.js[8] is a robust JavaScript framework that allows the user to create dynamic and aesthetically attractive data visualizations. This framework has various types of visual components, with those the researchers and the data analysts can create very interactive and user-friendly charts and graphs. Plotly.js has gained appeal among developers and data scientists for constructing rich and dynamic data-driven apps due to its wide capabilities and adaptability.

This visual library is accordant with a wide range of data sources so that the user can work with different formats of data. Whether the data format is in CSV, JSON, or a database, Plotly.js provides simple APIs to load and show the data. This framework does not only work with the static dataset it also supports real-time updates, so that the visual components can be dynamically updated with the newly available data. Another important component of Plotly.js is customization. The library provides a plethora of choices for customizing the look of charts. To fit the desired visual style, the user may modify many variables such as colors, fonts, axes, legends, and annotations. Plotly.js also supports responsive design, which allows charts to adjust to multiple screen sizes and orientations.

[7] https://axios-http.com/
[8] https://plotly.com/javascript/

4 Methodology

The main goal of this thesis project was to develop a diagnostic learning analytic system. The plan for this project has been done in two steps. Figure 4.1 is illustrated the first phase of the methodology. On this phase the plan was the raw data will pass through a data preparation process where the data would be extracted, cleaned, transformed etc. In this step the raw data files would be transfromed into JSON format. The backend environment will use this JSON data and through API it would be saved into a database (MySQL). A Frontend environment would call these data through the backend environment via API. In the frontend environment a data pattern analysis would be performed with the data. In the data pattern analysis these data would go through exploration and discovery then an appropriate analysis would take place. After that this analysis result would be visualised in a dashboard. If additional analysis needed then the frontend will call the backend through API to fetch proper data from database to perform same steps for the next analysis.

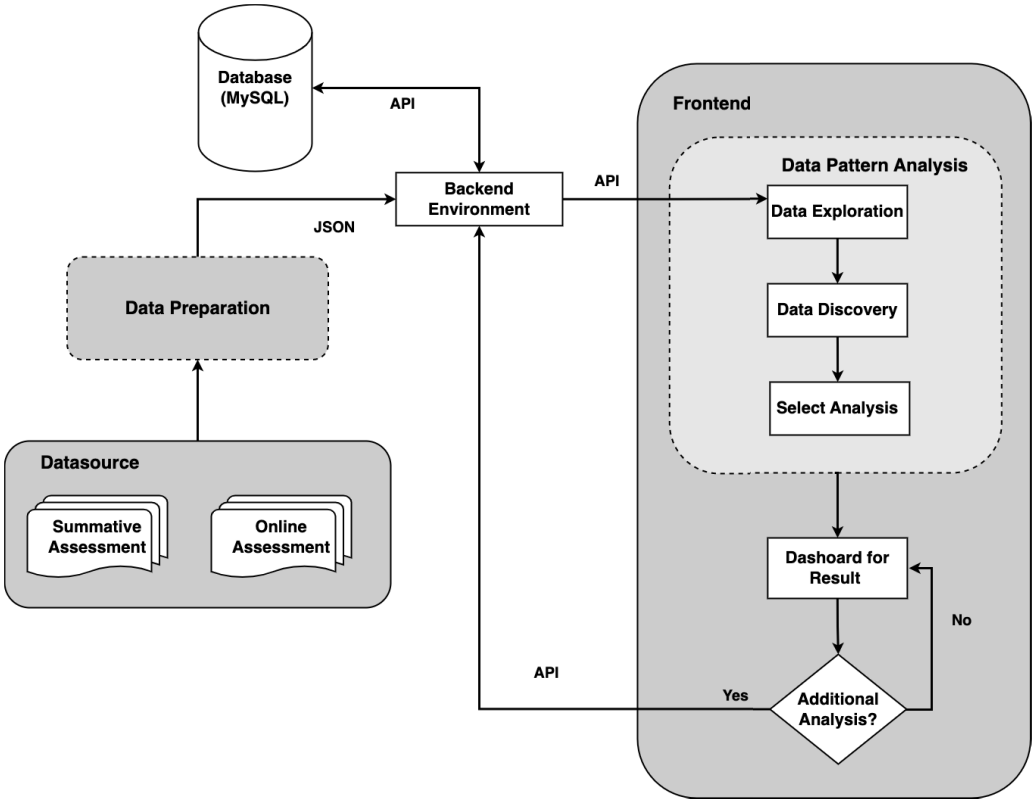


Figure 4.1: Block diagram for methodology (Step 1)

In Figure 4.2 the final phase (which has been implemented) has been presented. In this methodology, all of the data will come from some provided API from TUC server ina backend environment. In the data fetch all the data will be fetched from the API and a pattern analysis will be performed. In the data pttern analysis the data will be explored and then data process will take place. After processing the data some proper analysis will be done by using the data. An frontend environment will call these analysis results through some APIs and will be illustrated in a dashboard. Like the first step if further analysis needed then the frontend will call the backend through API and the result will be shown it the dashboard like before.

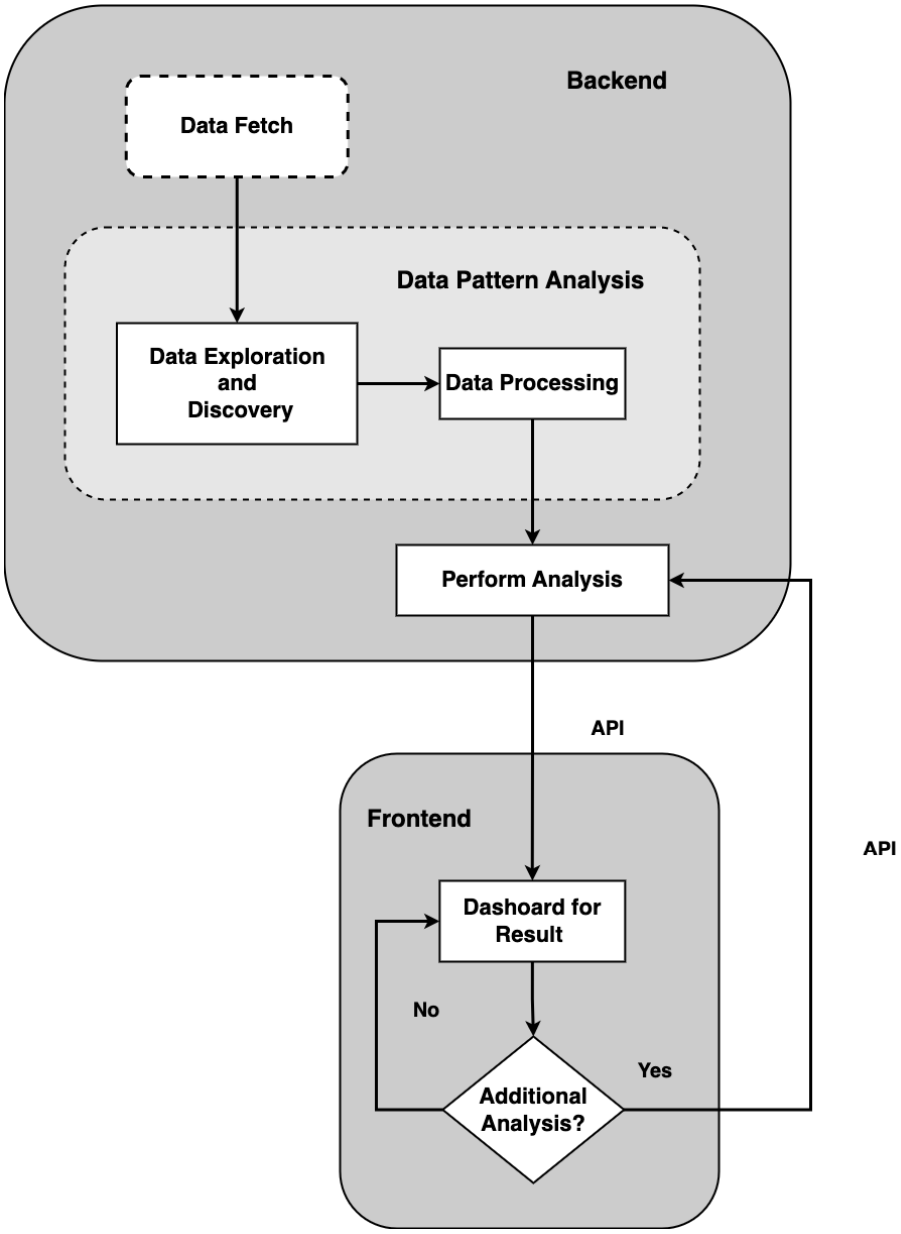


Figure 4.2: Block diagram for methodology (Finalized)

4.1 Use Case Description

In 2022, Technische Universität Chemnitz proposed a digital mentoring system *along with Universität Leipzig, Technische Universität Dresden and Hochschule für Technik, Wirtschaft und Kultur Leipzig Fakultät Informatik und Medien*, where a seminar will be held for introducing the different steps of the scientific research work. To complete this course, the students need to prepare a presentation and a research paper. This monitoring system has been integrated in OPAL, the students will get the self-tests for all lessons. These self-tests help students to figure out their weaknesses in any lessons of this seminar [TUC3].

The participants of this seminar are the students of Bachelor's and Master's programs but majority participants (greater than 90%) [TUC1] are from Master's programs and they belong to different cultures. The proposed formative assessment has two forms. First one is ARS test and second test is a standard Self-Test. ARS test is used for motivate the students during the lecture and train the participants with the learning contents to increase their motivation in this seminar course. In this test form the mentor starts a discussion about the problems of the topic and discusses them during the lecture time to clarify the mistakes by the mentee. After the lecture, it is necessary to train the students about the learning contents. The Self-Test has been provided to fulfill this purpose. The students take part in this self test, where the system would give them the feedback about their participation.

Each of these two test forms has four different tests. The first test is about reviewing or searching the trustworthiness about different works of literature and their sources. In this test, the students would be given different literature sources and they have to find out the reliability of these sources. After this, the second test is about the scientific topic presentation. Two different types of presentation video will be there and the participants need to figure out which presentation techniques is more accurate with the scientific presentation. The third topic is a discussion about scientific topics. The students need to assess the given typical questions by answering them to figure out how those answers are appropriate for scientific research fields. The last test focus on scientific writing skills about plagiarism. Some examples of quotations (both direct and indirect) would be provided to the students for checking the correctness of them.

4.2 Data Fetch

In web development, data fetching is a very crucial and vital feature. This feature could be called the heart of a web application. If anything goes wrong with this section, there is a high chance of falling down of the whole project. This feature enables to development of an interactive and dynamic project and the user can get up-to-date information. Most of the time this "Data fetch" term refers to retrieving data from an external source. The fetched data can be in JSON, XML, HTML etc formats. Data fetching is an important component of modern web development,

and there are several ways and packages for doing this task. The approach used will be determined by the project's individual demands and goals. Data fetching best practices include reducing data transmission, caching and compression, and error handling. That is the reason, data fetching is an important stage in data-driven operations. It guarantees that the appropriate data is available at the appropriate moment, supporting data analysis, decision-making, and a wide range of applications across businesses and disciplines.

In this study, the data fetch is done in the "*Backend*" project, which is developed by Python. Python provides different techniques and modules for data retrieval, but each of them has its own advantages and disadvantages. The approach chosen will be determined by the project's individual demands and constraints, but developers can be confident that Python provides a strong and versatile ecosystem for data retrieval and processing.

urllib[1] is a Python library for performing HTTP queries, which is part of the standard library. It has a lower degree of abstraction than the requests library, but it is also more versatile and customizable. The urllib library can manage authentication and cookies and supports multiple HTTP protocols, including HTTP, HTTPS, and FTP. It also provides common and basic error handling features, like 404 (Not Found) or 500 (Internal Server Error). But sometimes this library will cause confusion for 4XX URL errors. It can be said it is a small drawback of this library.

Figure 4.3: Block diagram for Data fetch in Avenseguim

A total of eight APIs are provided for this study. All of them are GET requests. That means these APIs would fetch information from the TUC server but it would not send any data to the server. Each API returns a Python dictionary in JSON format. As the response is a Python dictionary so, every entries or values for each keys are at same position and the key names represents the column names of the table. In Figure 4.3, the block diagram of data fetching "*Avenseguim*" has been illustrated, where it can be seen that TUC server is only sending the data to the Analytics system. But there is no HTTP response is going back to TUC server. A sample of response data can be seen in the code snippet below. From this snippet, it can be understand that, this table has columns (Semester_ID and Semester_Name) and it has two rows. The semester id 3 belong to semester name "SS_2020" and 2 belong to "SS_2021". In Figure 4.4, the database schema of the API can be seen. As this study used only four APIs, the schema is only showing the relationship

[1]https://docs.python.org/3/library/urllib.html

among the database table of the APIs.

```
{
    "Semester_ID": [
            3,
            2
        ],
    "Semester_Name": [
            "SS_2020",
            "SS_2021"
        ]
}
```

Listing 4.1 Sample of API response

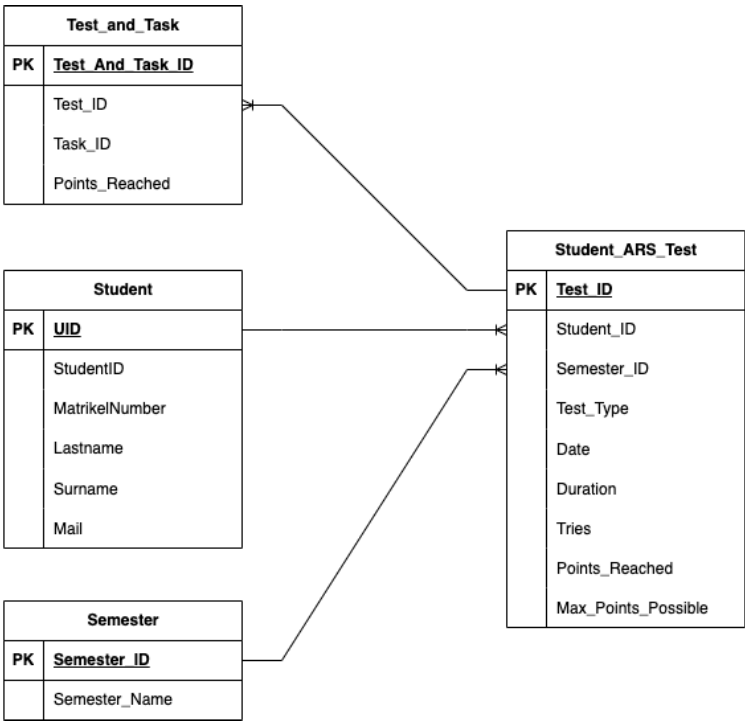


Figure 4.4: Database Schema

4.3 Data Pattern Analysis

After importing the data in the system it is important to analyze the data pattern. This data pattern analysis will help to find the number of missing or duplicate values. Sometimes it needs to identify that which values are not depended on each

other or maybe deepened with each other. This analysis would also figure out this information. How much an attribute can be used in further analysis, also can be figured out from this analysis. Data pattern analysis makes it very easy to find out the target dataset and its columns. Except these, data pattern analysis would identify the column type (like numerical or categorical). Data pattern analysis can be done in many ways. This study, followed basically two steps to do this analysis. These steps have been explained in the next section. At the end of these steps, two data frames achevied like Figure 4.5

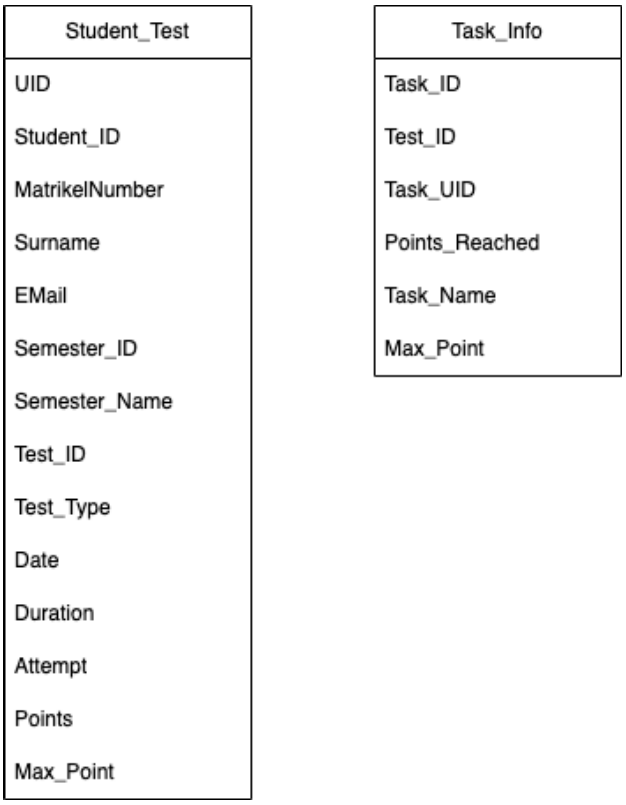


Figure 4.5: Dataframe Schema

4.3.1 Data Exploration and Discovery

In the data science world data exploration and data discovery they look similar, but there is a slight difference between them. The data exploration is which part of the data will reveal the answer to the target problem. This also can help to explore different hypotheses. Sometimes data exploration can refer to as a data refinement process. On the other hand, data discovery comes after data exploration. When the part of the data has been found out, data discovery helps to dig deep into the data and gives more insight about the dataset. Basically in the data discovery phase, the data pattern gets started.

Some steps could be followed for performing these two phases. First of all, the target columns need to identify which can be used for solving the target problem. Then basic metrics that means the total column and row numbers, and their data type need to be checked. After that variant analysis needs to be performed. This analysis could be graphical or maybe non-graphical. For graphical analysis, histogram, box plot or count plot can be used for univariate. For analyzing bivariate, scatter plots, Linear Correlation, and Regression Analysis can be used. After the analysis, the column needs to identified whose value needs to be transformed from categorical to numerical or vice versa. Last step is variable interaction. How the variables are interacting with ecah this need identifed for the further steps.

4.3.2 Data Processing

Data processing is critical for organizations and researchers to acquire insights into consumer behavior, market trends, and other vital elements affecting their operations. Data processing is the most important step for further decision-making steps. After gathering the raw data, it is needed to be converted into meaningful information. In the data world, this step is very sophisticated because if this does not have done appropriately the output would not work properly or maybe some information could be lost. Most of the data model depends on this step. If the data is not processed properly then the whole model could fall down.

Data processing starts after exploring and discovering the data. It can be done by various tools and software. Among them, Excel, and Python are very popular. Python is commonly used for data analysis and can handle larger datasets. This could be done in some steps. These steps could be different for different projects or sets of data. In this study, data processing has been done in six steps. In the Figure 4.6 these steps have been given.

Data Cleaning

After exploring and discovering the data, it needs to be cleaned. The process of discovering and deleting flaws or refurbishing the inconsistencies, and inaccuracies in raw data is known as data cleaning. It is an important phase in data processing since it guarantees that the data is correct, dependable, and fit for analysis. A variety of circumstances can cause the raw data to fluctuate, be inaccurate, or be inaccurate. The circumstances include human-made mistakes (like typing mistake, or manually entering data), machine errors or sometimes system glitches. When erroneous data is sent across systems, system faults can occur, resulting in data corruption, loss, or duplication. So it becomes very difficult to use them further. This is the reason data cleaning is needed to polish them into a good format so that the data can be used in further decision-making. Data cleaning involves different techniques and methods, depending on the type of data. These techniques include removing duplicate records, correcting spelling and formatting errors, and dealing with missing or null values. Sometimes data cleaning needs some difficult techniques

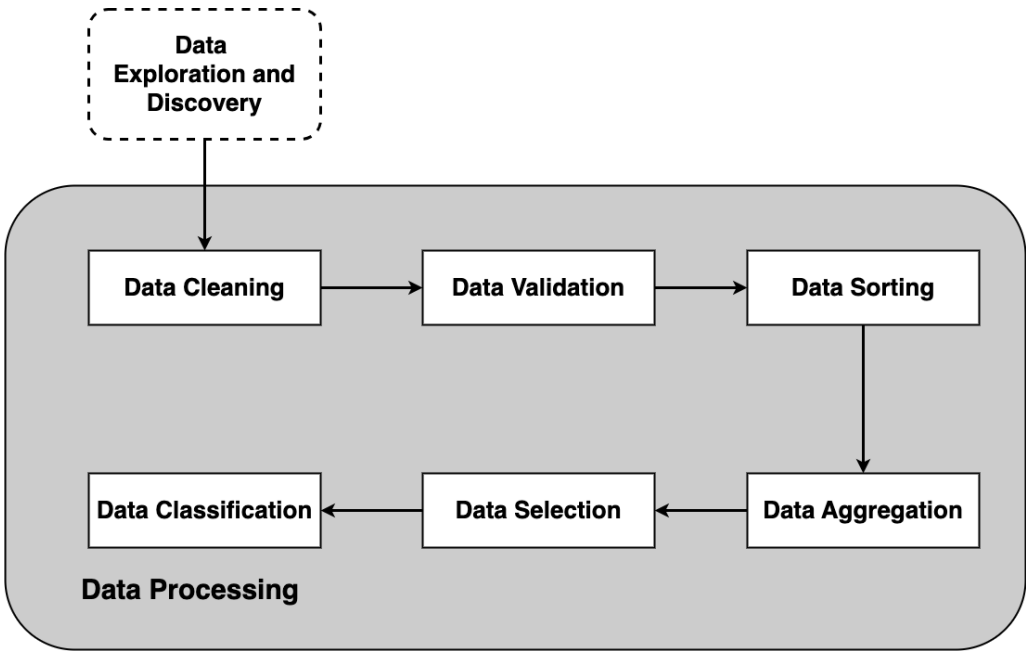


Figure 4.6: Data Processing Steps

like standardizing variables, transforming variables, or removing outliers to make them suitable for analysis.

Some critical columns need correct data during the aggregate of various datasets. Those columns must contain legitimate data, or if they have a trash value, they must be eliminated from the whole dataset. Those datasets may occasionally be populated with accurate data. It guarantees that the data utilized for analysis is trustworthy, accurate, and appropriate for the task at hand, resulting in more robust and relevant insights.

Data Validation

After cleaning the data, validation takes part. Data validation is the process of verifying that data has been cleansed to guarantee that it is both correct and valuable. It is accomplished by including multiple checks into a system or reports to assure the logical consistency of input and stored data. Data is input into automated systems with little or no human oversight. So it is critical to guarantee that the data entering the system is valid and satisfies the specified quality criteria. If the data is not input correctly, it will be useless and may cause larger downstream reporting concerns. Unstructured data, even if entered correctly, will incur expenses for cleaning, converting, and storing. There are some common data validation techniques have been followed. It is not necessary that every project would follow the same steps.

First one is data type validation. This step ensures that the data is entered in this field its type is same as it should be. Sometimes it can be seen that the data field is

integer but the entered data type is string. Then range check comes, here the data should in a certain range. But in the erroneous data it can happen that some data are out of the range. Data format checking is another validation that needs to be done with the existing data. In this step, mainly the date or time format is checked. Consistency validation is the fourth phase. Here the data needs to be consistent with the real world.

In this study, the data type has been checked and the data has been converted into the required type. Some data has found which are not consistent like the test starting date and time is overlapping with the next test start date. This consistency has been maintained in this study. Data format check also has been done in Anvenseguim. The date time fields were in string type.

Data Sorting

Data sorting is a procedure that includes putting data in some meaningful order to make it simpler to interpret, or display. When dealing with research data, sorting is a frequent strategy for presenting data in a way that makes it simpler to understand the story the data is saying. There are two ways to do it. First one is ordering where the data were arranged in a sequence of number. Another one is categorizing, here data organized in a certain same group. In this study, both ways have been adapted. some has been sorted by the SemesterID which is a integer data type. And some data has been organized by Test Type.

Data Aggregation

Data aggregation, in its most basic form, is the act of gathering often enormous volumes of information from a particular database and arranging it into a more usable and complete medium. Data aggregation may be used to summarize information and draw conclusions based on data-rich results at every size, from pivot tables to data lakes. Because of the increasing availability of information and the relevance of customization metrics across the company, data aggregation has become highly essential. Data acquired by businesses is critical for making better decisions, understanding customer behavior, improving process efficiency, and, lastly, understanding performance, whether of the firm or its products. To that goal, just obtaining high-quality, dependable data is insufficient. Consistent discoveries, consistent evolution, and data usability all play important roles. Business decisions including activities such as strategy planning, pricing, and marketing campaigns rely significantly on information derived from aggregated data.

In this study, data aggregation has been done into one data source with different data tables. After sorting the data it has been seen that some columns have such values which are not easy to read (hash code or system generate numbers). To make them easily accessible other related data table has been joined with the main data table.

Data Selection

Data selection is a process where a small chunk of a dataset would be selected from a large dataset. This data set selection would be based on certain criteria. This process makes the dataset more manageable and makes easily to perform the next analysis steps. Another reason doing this step is very necessary, it reduces the volume of the data set, which is very essential for the data world. It is possible to decrease noise and remove irrelevant data that may cause erroneous results or misleading patterns and trends in the data by selecting a small subset of data that is more relevant to a particular task.

Data Classification

Data classification is a process where the whole data set is categorized into multiple groups or clusters that behave the same. The purpose of this step is to organize the scattered data into clearly understandable and searchable. Another reason for classification is, to manage the whole data set based on the similarities and differences. Data classification increases the effectiveness of the data quality. Effectiveness means that the data would be more accurate than before. The data would not contain random or garbage values. The data would behave the same as the group it belongs to. Robustness can be achieved by data classification. The data set should be predictable by the group name which makes the dataset easily searchable and accessible. Classification can be accomplished with different techniques. These techniques depend on the data structure and pattern. Most common techniques are clustering and decision tree. Clustering groups the data based on the similarities of each other and the decision tree uses a set of rules to classify the dataset. This study uses clustering technique to classify the whole data set into different test types. Each test could occured in different groups. The dataset has made different clusters for the different test groups also.

4.4 Data Analysis

The main purpose of this study is to build a learning analytic system. Specially a diagnostic learning analytic system, that all the data and the information are about learners and goal of this system would be to find out the reason behind any occurrence. As described before diagnostic analytic is an out of box process. After presenting the available data then the main steps for diagnostic starts. Descriptive analytics is used to present the existing data and based on that analysis the diagnostic analysis has been performed. This thesis project has performed both analyses. In the next section, the procedure and the steps taken for performing these analyses have been discussed.

4.4.1 Descriptive Analysis

This study aims to create diagnostic learning analytics. For any diagnostic analytics, the root analysis is descriptive analysis. As descriptive analysis is a key part of data analysis that focuses on summarizing and explaining the primary features, patterns, and trends in a dataset. It entails arranging and displaying data in an understandable manner in order to obtain insights, analyze distributions, and convey crucial results. To explain the data in descriptive analysis, several statistical measures and techniques are used, such as mean, median, standard deviation etc. These measurements give information on the data's usual values, distribution, and variability.

Descriptive analysis can be performed on a single variable or may be multiple variables. The analysis performed on a single variable is called univariate analysis. For central tendency, this analysis uses mean, median, mode. Measures of dispersion also performed in this analysis by using range and standard deviation. This study uses this analysis to find out the central tendency by performing mean of different test points.

Mean is the most common and frequently used method to find out the average of the given items. The formula 4.1 for mean very simple like below. Though it is very simple, but this can help the information and help to predict the what can happen in the future.

$$mean = \frac{\text{Sum of items}}{\text{Total numbers of items}} \tag{4.1}$$

Mean can be useful for analyzing the historical information of an organization. By doing mean of the historical data of the organization it could be figured out that how the data is changing over time. Another usage of the mean is comparing values of different departments of an organization. This could the first step to answer further prediction about the organization. A good visualization would convey the best result to the audience. That is the reason not only performing the mean calculation is enough, the visualization is also important. In the latter sector how the result of descriptive analysis illustrated has been discussed.

4.4.2 Data Drill-Down

In diagnostic analysis drill down plays a great role. Drill down is an analytics function that allows users to immediately switch from an overview of data to a more comprehensive and granular view inside the same dataset they are examining by clicking on a measure in a dashboard or report. It allows everyone to investigate particular information in a report from various perspectives by descending from one level of a specified data structure to the next. Drilling down into a dataset can offer more comprehensive information about which components of the data are causing the observed patterns. Suppose, the user may deep dig into national sales data to evaluate whether certain areas, consumers, or retail channels are responsible for

higher sales growth. Drilldown is a technique in online analytical processing (OLAP) and information retrieval in which the many features of current information items are employed to gradually narrow the analysis or search for them. In OLAP, the attributes are referred to as dimensions. Facets, on the other hand, are used in information retrieval. Drilldown allows for "zooming in," or seeing existing data at various degrees of detail.

In [148], the researchers introduced a new technique, smart drill down, for exploring the whole relational dataset in efficient way. With this the researchers, found out the set of groups which are represented by a common rule. As example, the meaning of (a, b, *) is with first column value a and second column value b there is a possibility that in the third column there could be thousand of values. So, this smart drill down presents some set of rule where the analyst can understand the structure of the dataset. Data drill can also be used in learning analytics dashboard to visualise the learning progress of students. According [12] the data drill-down process can lead the users to very interesting details after applying a predictive model. In Figure 4.7 it can be seen that the international student with high engagement, the model proposes the high percentage rate of coverage of the course module path.

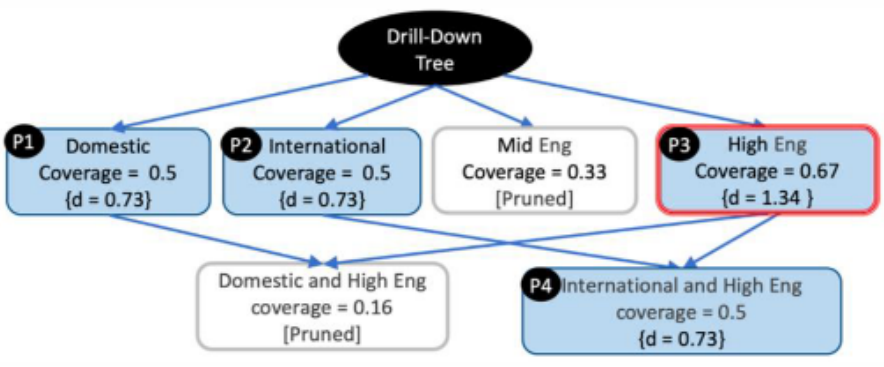


Figure 4.7: Drill down tree for recommendation on student performance [12]

4.5 Correlation

According to Webster's online dictionary, correlation means *"a relation existing between phenomena or things"*[2]. Most of the time the term correlation is used in statistics for expressing the intensity of association between two variables. In short, it can be called bivariant analysis. This relation intensity is expressed in a number which is called the correlation coefficient. This coefficient value ranges between -1 to 1. The negative coefficient represents that the variables are moving in opposite directions. In the same way positive coefficient means, the two variables are moving the same direction. If the variables are not related to each at all then the coefficient

[2]https://www.merriam-webster.com/

would be zero. In diagnostic analysis and different ML algorithms, it is very important to know which variables are correlated with other. This information can help the analysts to prepare data to fulfill the expectation of those algorithms and analysis, whose performances are dependent on these inter-dependencies. Correlation also indicated the predictability of relationship between two variables. As a classic example, the correlation between systolic blood pressures (SBP) and diastolic blood pressures (DBP) (in Figure 4.8) for genders. From the figure it is clear to understand that, SBP and DBP are highly correlated as they are increasing or decreasing together.

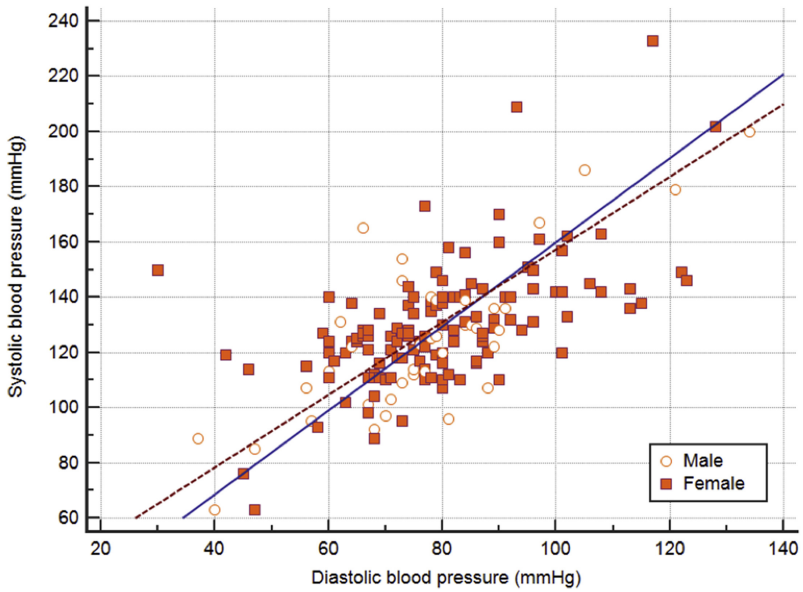


Figure 4.8: Correlation of SBP and DBP according to gender [13]

There are several methods to perform correlation. The most common and frequently used methods are Pearson correlation and Spearman correlation. In this study, the Pearson correlation has been used to figure out the relationship between different test results. This correlation used for two continuous and quantitative variables. The Pearson correlation used for normally distributed data and when their relationship is linear [149]. To calculate the Pearson correlation the formula 4.2 given below is used -

$$r = \frac{n \times (\Sigma(X,Y) - (\Sigma(X) \times \Sigma(Y)))}{\sqrt{(n \times \Sigma(X^2) - \Sigma(X)^2) \times (n \times \Sigma(Y^2) - \Sigma(Y)^2)}} \tag{4.2}$$

where,
r = Correlation Coefficient between X and Y
n = Number of observations

Pearson correlation can be used in different sectors. A study is being conducted in the healthcare industry to find out the relationship between Post Traumatic Growth

(PTG) and Post Traumatic Stress Disorder (PTSD) symptoms by using Pearson Correlation. The result comes out that, there is a positive correlation between these two symptoms for different variables like - age, trauma type, and time since trauma [150]. Noise reduction research focuses on estimating clean speech from noisy observations. Optimal filters traditionally use the mean-square error criterion, but a new approach based on the Pearson correlation coefficient offers an improved analysis of noise-reduction performance, making it a more suitable criterion for optimizing filters [151]. The relationship strength can vary for different sectors. Table 4.1 represents the thumb rule of Pearson correlation.

Coefficient value (r)	Description
±0.90 to ±1.00	Strongly high positive (negative) correlated
±0.70 to ±0.90	Moderately high positive (negative) correlated
±0.50 to ±0.70	Moderately positive (negative) correlated
±0.30 to ±0.50	Low positive (negative) correlated
±0.00 to ±0.30	Negligible positive (negative) correlated

Table 4.1: General thumb rules for correlation strength [16]

4.6 Clustering

Clustering is a key approach in data analysis in which comparable data points are grouped into separate clusters based on their commonalities. Clustering seeks to uncover patterns, structures, or natural groups in a collection without the use of explicit labels or preset categories. All clustering algorithms are designed on unsupervised learning principles, that the reason those algorithms can work with any kind of unlabeled data for discovering the underlying meaning of it. Clustering may be used in a variety of disciplines, including image segmentation, customer segmentation, anomaly detection, document categorization, and others.

The process of classifying involves putting a fresh observation into a set of established categories despite obstacles including overlapping classes and uncertainty. Probability estimate, which uses statistics, aids in choosing the group that is most likely, but taking into account imbalances or different classification costs may require more complex criteria [152]. Classification algorithms can be divided into nine different categories depending on the dataset type and the problem categories [153]. According to the problem category and dataset pattern, this thesis study needs to partition the dataset in some categories. To solve this kind of problem partition based clustering algorithm is best way to use. Among all of the partition based algorithm, K-Mean [154] is very popular and most frequently used algorithm.

In [155], the K-Mean algorithm has been refereed as *"An Ageless Algorithm"*. The K-Mean method may be described as a straightforward partitional algorithm that seeks to identify K numbers of non-overlapping clusters, where each cluster is determined by the centroids of various data points. The algorithmic steps for

K-mean algorithm has been pointed out by using Euclidean distance metric in [156] like following -

1. Define the number of clusters, K
2. Initialized the cluster centroids. At the beginning, these points would be unknown. So random point would be selected at starting point of the algorithm. The number of centroids are same as the number of the clusters.
3. Next step is to allocate the data points to their nearest cluster centroid. This allocation would depend on the distance of the data point and the centroid. If this distance is minimum the data point would be assign to that centroid. This distance is calculated by using Euclidean Distance equation 4.3

$$d\left(p,q\right)=\sqrt{\sum_{i=1}^{n}\left(q_i-p_i\right)^2} \tag{4.3}$$

   where,

   $d(p,q)$ is the distance between p and q

   pi, qi are the initial origin for the p and q

   n is the total number of the datapoints

4. After the initial allocation of centroid, the centroids need to be re-initialized by using the following formula 4.4

$$V_i=\left(\frac{1}{C_i}\right)\sum_{1}^{cl}p_i \tag{4.4}$$

   where,

   Ci is the data point numbers in the ith iteration.

5. Until no data points are reallocated Step 3 and 4 needs to be repeated.

Clustering analysis can be used in different domains for classifying different data point in the same category. In Indonesia, Bengkulu Province is an earthquake-prone area so the characteristics of this area needs to be studied. To find out the earthquake epicenter [14] used K-Mean analysis. In the Figure 4.9 the clusters can be seen. The reasearchers have found the fourteen clusters for the earthquake-prone area.

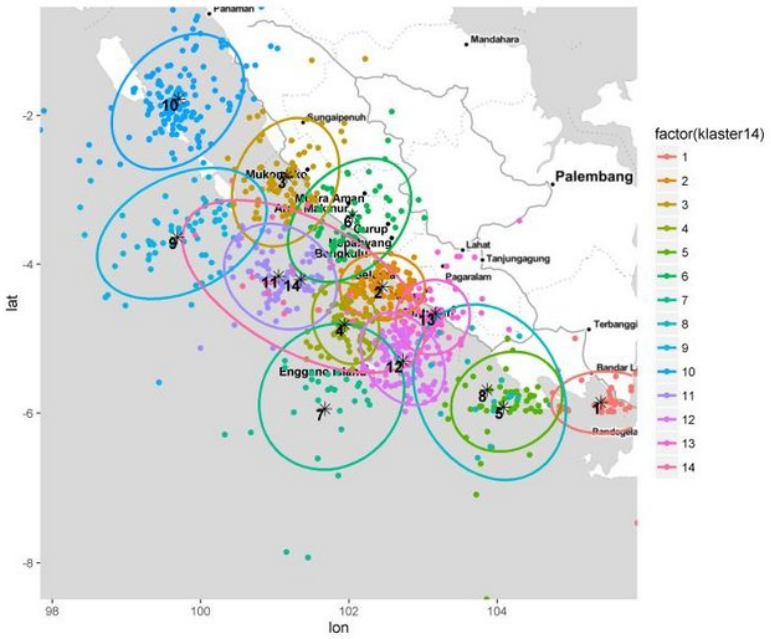


Figure 4.9: K-Mean cluster analysis for earthquake [14]

4.7 Visualization Techniques

Visualization techniques are all about generating a mental picture that helps to understand the information in a detailed way. It should be understandable by anyone after watching it. This is the reason the visualization technique needs to be chosen very proper way to present the insight of information. This thesis project used basically three techniques for visualization which have been discussed below.

Table

The first visualization technique this project adapt is table. Reading line after line does not give a human a very good and clear picture of the data. But when the same person sees the data or the information in a tabular format, for that person it becomes more clear and understandable. At first glance, the person can have a good understanding of the data pattern and the quality of it from the tabular data. Table data visualization is an excellent approach to help people understand data. People can easily spot patterns and trends because perception is easier. This assists them in identifying outliers and anomalies.

A table visualization displays data from a metric set in a tabular format. A table, also known as a data grid or data table, is the default style of visualization that is utilized when selecting data for the first time. Tables are designed to be read, therefore they are perfect for presenting data that cannot be simply displayed visually, or when the data demands more specialized attention. Furthermore, tables emphasize precision, allowing the analyst to go deeper into the statistics and study

specific values rather than relying on approximations or representations. Finally, tables aid in the evaluation of data sets with various dimensions and values without the need for elaborate displays.

A table can act as a symbol of a sector, which can be accepted as a constant world wide. A classic example of this kind of table is a periodic table of chemistry. This table saved so much time to remember the elements of chemistry and their behavior according to their atomic numbers. The most important point of this table is a person can see the trends in the properties of the chemical elements. The organization of this table has been divided into three parts - periods, groups, and blocks, which are known to all worldwide chemists and scientists who are working with this table.

In data science, before doing any kind of analysis the scientist presents the data in a table so that they can have a good picture of the data orientation. For example in [157] the researchers tried to survey of best visualization of network security. But before doing the survey they generate a table with the network data source to understand about them and take further steps with this kind of information. This thesis report has used this kind of table for showing the overall student information which is interactive with each action.

Bar Chart

Visualization raises learners' awareness of the learning process [158]. Aside from that, if given in meaningful ways, visual displays play a key role in sense-making [159]. In order to assist learners grasp and analyze data, visualization should be familiar and appealing to them [160]. Various visualization techniques can be used to develop a learning analytics dashboard. In [161] the researchers compare some very common visual techniques for dashboard and the result came out that line chart and bar chart are the most popular techniques to show the student's performance. According to this [162] study, it is seen that bar chart is best visual technique to show the students different types of scores and compare them to each other.

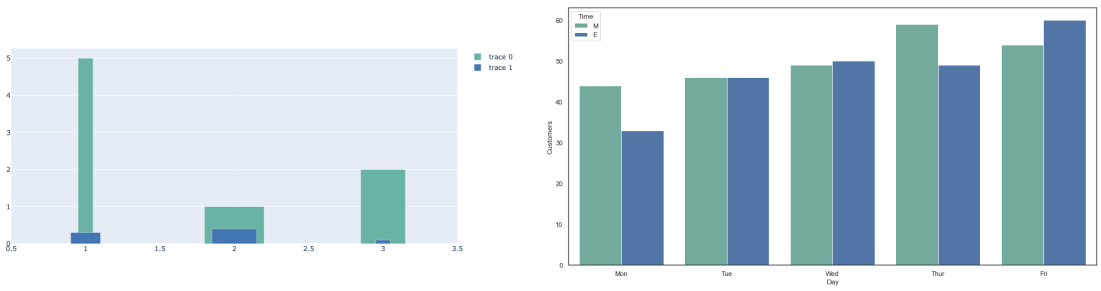
A bar chart depicts categorical data by using rectangular bars with lengths or heights that match to the value of each data point. Volume is used in bar charts to show changes between each bar. As a result, bar charts should always begin at zero. When bar charts do not begin at zero, consumers may misinterpret the difference between data values. The bar chart displays discrete, numerical comparisons across categories by using either horizontal or vertical bars. The chart's one axis depicts the precise categories being compared, while the other axis provides a discrete value scale. Bar charts differ from histograms in that they do not show continuous changes throughout an interval. Instead, the discrete data in a Bar Chart is categorical, and it answers the question of "how many in each category?"

According to "Think Design"[3] bar should be used in some special cases when other visualization can not represent the information readable way. When com-

[3]https://think.design/

paring discrete categories, bar graphs/charts can be used to graphically illustrate comparisons. Categorical data is data that is divided into discrete groups, such as months of the year, age groups etc. Think design suggests showing this kind of data in vertical alignment since the text of labels can be presented in a wider space. The second scenario is comparing categories and subcategories. In this case a bar chart with bars grouped (Figure 4.10(b)) in groups of more than one to represent the values of more than one measured variable. Stacked bar charts work well here to show how much each sub-group contributes to the sum in its category. It may be used to compare how entities perform against one another and how much each sub-group contributes to the overall performance.

The overlapping bar charts (Figure 4.10(a)) can be used to compare similar data sets, which is the third case, on the same chart by using different width bars. It can serve a dual purpose by comparing categories on one axis and representing a discrete value on the other. A single data set can be seen or a comparison of two data sets can be performed concurrently by preferentially choosing or deselecting legend labels. Sometimes it is possible to have some negative data in the dataset. It is important to consider this kind of data in an analysis to get insights of the deviation of data. So for comparing numbers that tend to fall into negatives as well as positives column charts can be used to understand the deviation of the dataset.



(a) Example of overlap bar chart

(b) Example of grouped bar chart

Figure 4.10: Example of different bar charts

Heatmap

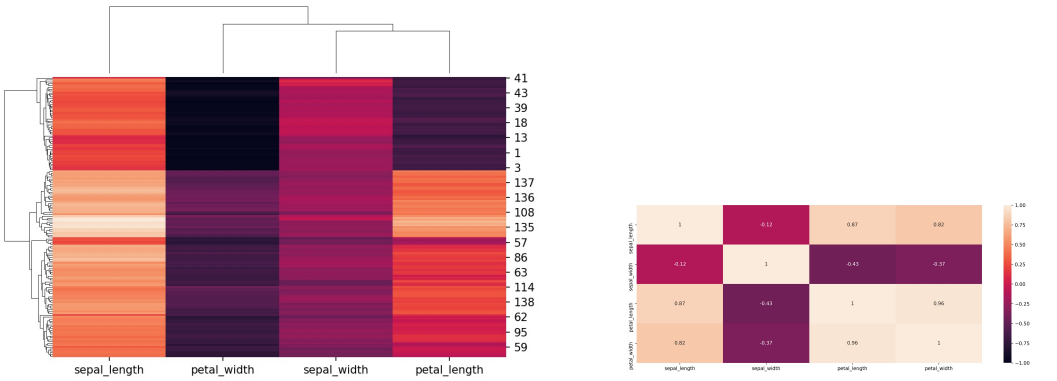
A heatmap is a grid of colored squares that represents values for a primary variable of interest over two axis variables. Like a bar chart or histogram, the axis variables are separated into ranges, and the color of each cell represents the value of the primary variable in the corresponding cell range. Heatmaps are also far more visually striking than regular analytics reports, making them easier to grasp at a glance. This makes them easier to interact with, especially to those who are unfamiliar with processing big volumes of data.

Heatmaps are used to demonstrate relationships between two variables, one on each axis. It is possible to see if there are any patterns in value for one or both

variables by monitoring how cell colors vary across each axis. The variables shown on each axis might be of any type, including category labels and numeric values. In the latter situation, the numeric value must be binned, as in a histogram, to produce the grid cells that will depict the colors associated with the main variable of interest.

In data analytic, there are two types of heatmap used most frequently. The first one is cluster heatmap and second one is correlation heatmap. Instead than having the horizontal axis reflect levels or values of a single variable, it is usual to have it represent measurements of many variables or metrics. If we set the vertical axis to individual observations, we get something that looks like a conventional data table, with each row representing an observation and the columns representing the entity's value on each measured variable. This form of heatmap is also known as a clustered or clustering heatmap since the purpose of this style of chart is to create relationships between data points and their attributes. In Figure 4.11(a) a simple clustered heatmap can be seen, Each column in the figure represents an individual floral specimen, and each row is a measurement from that item.

The second type of heatmap is the correlation heatmap, which has been used in this paper. A correlation heatmap is a heatmap version that substitutes each of the variables on the two axes with a list of numeric variables from the dataset. Each cell illustrates the connection between the crossing variables, such as a linear correlation. These basic correlations are sometimes replaced by more complicated representations of relationships, such as scatter plots. Correlation heatmaps are frequently used in an exploratory function, assisting analysts in understanding correlations between variables in the service of developing descriptive or predictive statistical models. In Figure 4.11(b) a correlation heat has been illustrated. It is very seen to understand from the figure that petal length is strongly connected to petal width and sepal length, whereas sepal length is adversely associated to the other three factors.



(a) Example of clustered heatmap

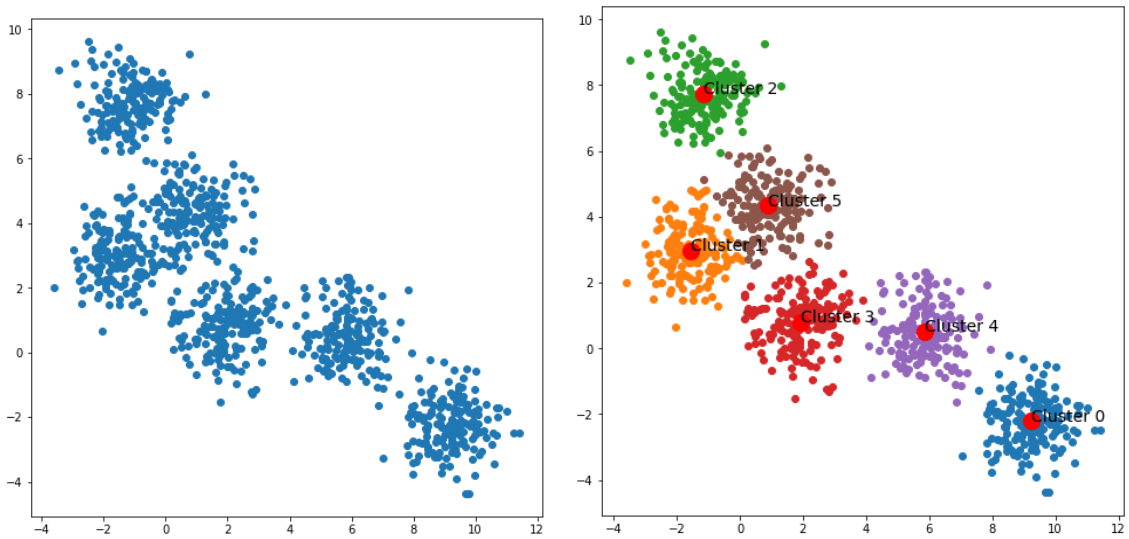
(b) Example of correlation heatmap

Figure 4.11: Example of different heatmaps

Scatter Plot

When displaying the relationship between two continuous variables, a scatter plot is a particular sort of graphical representation. This visualization is very helpful when the connection between two variable or distribution of data points need to visualize. An illustration of the distribution and grouping of data points according to their characteristics or qualities is called a scatter plot, and it is frequently used in clustering. Finding organic clusters or groups within the data is useful. Using k-means, one may locate clusters in a scatter plot and see what each cluster's centroids (means) are like.

The effectiveness of the K-means algorithm's division of the data points into discrete clusters may be clearly seen in scatter plots. The borders between the clusters are often easily visible since distinct clusters are typically represented by various colors or symbols. On the scatter plot, it is simple to see each cluster's centroid, which reflects the mean of the data points within that cluster. This makes it possible for us to see where each cluster's center is located in the feature space. Even for audiences that are not technically savvy, scatter plots are simple to understand and comprehend. Effectively communicating the clustering findings requires the data points to be shown in a scatter plot with various colors denoting distinct groupings. Even while scatter plots are useful for displaying K-means research, it's important to be aware of their limits, particularly when working with high-dimensional data. Direct visualization of the clusters gets difficult in higher dimensions.



(a) Input dataset in Scatter plot

(b) Output dataset in Scatter plot after applying K-Mean

Figure 4.12: Example of Scatter Plot[4].

[4]www.towardsai.net

Figure 4.12(a) is showing the raw data points, where some of the cluster can be assumed. But in the second Figure 4.12(b) the clusters become more recognizable after applying the K-Mean analysis in the raw dataset.

5 Implementation

This chapter describes the implementation details as well as the functioning procedures of the proposed diagnostic learning analytics, Avenseguim. These include the used technologies, procedures and structure of the implementation process. Additionally, how the datasets have been fetched and processed for the further analysis steps. Finally, the technologies and procedures to visualize the analysis also have been briefly discussed.

5.1 Project Setup

In the *Methodology* chapter, the finalized block diagram shows that the project has main two projects, Frontend and Backend. In Backend project, all of the data related work has been done, such as data fetching, processing, API for analysis creation etc. And in the Frontend project, a dashboard has been developed, where the analysis of data has been visualized and the further analysis has been narrowed down.

5.1.1 Backend

Backend project has been developed by Python. The version of python is 3.8. As IDE for developing this project has been used PyCharm[1]. PyCharm is a robust IDE created exclusively for python development. It is a popular product of JetBrains that offers a complete range of tools and capabilities to help python developers increase productivity and streamline the development process. PyCharm's intelligent code editor is one of its most notable features. It provides extensive code completion, syntax highlighting, and error detection, all of which substantially assist in producing clean and mistake-free code. The editor also allows code refactoring, allowing developers to quickly adjust and improve their code structure without sacrificing dependability.

For maintaining the good robustness of a project, its' structure is very important. The structure of the backend project has been discussed below. In the root, a folder named *src* can be found. This folder has two more folders *Data* and *API*. The *Data* folder contains all of the data computation work. First, in the *fetch.py* file all the data would be fetched with the provided API from the TUC server (Figure 4.3). In the same file the fetched dataset has been converted to panda *DataFrame* for further data processing. The next file is *processing.py*. In this file, all data has been processed for further analyzing them easily. The other steps (Figure 4.6) of the data

[1]https://www.jetbrains.com/pycharm/

processing have been done in this file. The last file of this directory is *analysis.py*. In this file, all of the queries for the different analyses (both descriptive and diagnostic) has been placed.

The second folder of the root directory contains the logic for different APIs creation. These APIs are sending the analysis data to the frontend project. In *utils.py* file the analytics has been fetched from the *analysis.py* file and changed the response data to API-supportive data format (such as JSON) through some functions. After that, these functions have been used in the *main.py* file for creating different APIs. There is another folder named Notebooks, which can be found in the backend project. This folder has some jupyter python notebooks. In this notebooks some intermediate analysis tests have been performed before sending those analyses to the frontend project. Each step of backend project has been tested in these notebooks. For each step of the backend project there three notebooks have been created to check how the functionality is working.

5.1.2 Frontend

The frontend project has been developed by ReactJS and the latest version 18.2.0 has been used. For developing this frontend project, PhpStorm[2] has been used. Like PyCharm, PhpStorm is another very well-known IDE of JetBrains for frontend developers. Its powerful features support different web development languages, like HTML, CSS, PHP, JavaScript etc. The wide range of tools of PhpStorm, helps to enhance the productivity of the development process. The most notable feature is it's Version Control System (VCS). It can support different types of VCS (such as GitHub, SVN etc.). A tree view has been integrated into it for version control. Any kind of changes (commit, push, pull, merge, stash) can be understood in this tree view very easily. This makes it easy for the developers to manage their code and collaborate with their teammates.

Like the *Backend* project, the *Frontend* project has been structured too. In the *api* folder the *utils.js* file has all the API call functions by using *axios* framework. The route of the whole application starts from the *App.js* file of *root* folder. The *view* folder contains all of the main and child components for the dashboard. This front-end project has been developed with the single-page application concept. Different components would be visible according to its need. The visibility has been determined by the different requirements and API calls.

The lifecycle of the frontend components has been illustrated in the Figure 5.1. This application starts with the *Landing View*. With the users' choice this view can be split into two other views, Teacher and Student View. These views can communicate with their parent component, *Landing View*. The Teacher View consists of *Heatmap*, *Table view* and *Barchart*. Among these three child components only *Table View* can interact with its parent component, Teacher view. The Student View has only one child component, *Barchart*. This component can throw action to itself to

[2]https://www.jetbrains.com/phpstorm/

change the data to show another bar chart.

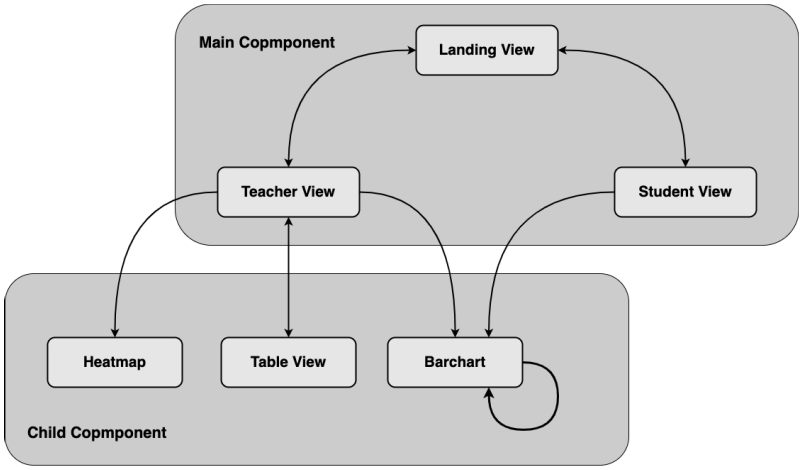


Figure 5.1: Frontend components lifecycle

5.2 Dataset Preparation

Data preparation is an important stage in data analysis. It is an iterative process, and the stages followed may differ based on the individual problem and the dataset's features. Data preparation for this project is detailed in the chapter *Methodology*. Cleaning, converting, and arranging retrieved data into a manner suitable for further analysis is required. Proper data preparation is critical for every data-driven project's quality, dependability, and effectiveness. Data preparation is a crucial aspect of every data science effort since a well-prepared dataset may dramatically increase the performance and generalizability of machine learning models.

In this project after collecting the data from the API, the data preparation step started. Data preparation started with the exploration and discover the data pattern. This step has been done manually after loading the data in the system. In the pattern analysis the first it has been checked that which columns are more important in the further analysis. After check the column, the type of the column determined. Suppose the Point column must be a numerical column. This could be either an integer or a decimal. The existing data shows that the values should be decimal. But the existing data were in string format. The another important analysis was the readable data. The dataset should contain readable data so that it can be more flexible to the developers. As example, each student has their unique matriculation number and an unique id. But this unique id was a hash id, which is 32 character long and which becomes very hard to read. So in the student dataset the matriculation id can act as its' unique id. So this UID column can be removed, but in the test dataset the student id came as the student UID. So after the aggregation step this UID column can be avoided from the final dataset.

After the manual data exploration and discovery step data cleaning can be started. The next steps all has been performed in the backend project with python. The collected dataset where it has been seen that the identity (ID) columns are null or empty, those columns have been removed from that specific datasets. The reason behind is, the ID columns are needed to aggregate one dataset with another. There are other data fields found empty, they are student score fields. Those fields have been filled with zeros. Otherwise, it would cause an issue with further calculation.

The next step is data validation to check the required column type is same as its' value. In this study, the column data type has been checked and the value of those column has been converted into the required type. Some data has found which are not consistent like the Test score column should have decimal type of data. But in the dataset it is found that some of the value are in string format. Another scenario found in the dataset that, the decimal numbers had some string character (like comma(,)) those characters have been replaced by full stop (.) character to do further calculations easily.

Next comes, data sorting. The test type has its' own order. So this test names should be in the same order as the real world. As the test name are string it was difficult to sort them with their names. To have a consistent result a new integer type of column has been introduced where for each test name. As example Search test (for both ARS and Self Test) has given 0, then Presentation test (again for both ARS and Self Test) has given 1 and so on. With this new column the test type order could be maintain in the whole project.

In the aggregation step, three datasets has been aggregated together here. In the first step, the semester and test data has been joined together. Then in the second round this new dataset has been joined to the student dataset. In the final dataset, all students would have their test and semester data. To aggragte the dataset only common data needed among these three different dataset. That is the reason inner join has been performed based on Student ID, Semester ID. The advantage of this step is, for the further analysis there is no need to get these three data information from three different datasets. These information would be in one single place. So it makes very easy to access either one specific semester test data or student test data or both for both test data. After aggregation dataset, it is wise to select only those column which are important for the further analysis and with an understandable name. This is the reason, renaming some of the column before selecting the vital columns.

The last step of data preparation is data classification. This study used two types of test data, Self test and ARS test. But ARS test can be performed into two groups. So the whole dataset has been clustered into three group. One for self test, second one for ARS test Group 1 and last cluster is ARS test Group 2. A new column has been introduced in the final dataset where these three groups have been inserted. So that it becomes easy to identify which test data is coming from which test group. In the further analysis it was needed to get these three group data for finding out the comparison among test point.

5.3 Perform Analysis

This study aims to develop a diagnostic analytics system with learning domain data. The first step of this type of analytics system is descriptive analysis. Descriptive analysis will give an overview of the whole dataset. It becomes very easy to find the fault, which would be the diagnosis, from this dataset overview.

The first descriptive analysis has been done by using tabular method. Tabular Methods are used to break down data in the form of tables. It is a method of organizing information in a grid row and columnar layout. Frequency table and cross-tabulation are the most widely employed tabular formats for data summaries. Tabular methods are useful for displaying data in a compact manner while also allowing comparisons and insights. They are frequently utilized in a variety of sectors, including as research, business, and data research. The second analysis has been done by graphical methods. This study has used two types of visualization. First one is bar chart, as different test results has been presented here so a group chart is used to present a different type of test result. The second one is a heatmap, to show the correlation between different test types and test groups, which is a part of diagnostic analysis. The graphical methods will be discussed in the latter section.

At begin, all of the semester has been selected from the Semester API. But there is a possibility that, the course have not been offered in every semester. So, it would be a overwork if all of the semester has been selected. To reduce this overwork, only those Semesters have been selected on which the course has been offered. This is the reason, in data preparation the data aggregation is needed. From the aggregated dataset only those semesters could be found which has the offered tests. That means the course has been offered only on those semesters.

After selecting the semester, then comes to select the student according to the semester. In general, the best scenario is one student should belong to only one semester. In real world, one same student can not be in multiple semesters. But in the provided dataset, it has been investigated that one student could be in different semesters. In one case, the same student is in three different semesters. The reason could be like this, the student may be enrolled the course, but could not pass in one attempt. Then in the another semester, s/he again enrolls to the course. In this semester the student took the ARS test and self-test, but at the end of the semester, s/he did not submit the report physically, as the Report submission is the last step of course completion. As one student can take the course maximum three times. And same way the scenario could be repeated again for the last course attempt time. At this time, the same student, took the self-test and tried to improve his/her skill in the specific test task. After doing all this analysis this student list has been presented in a tabular format according to the semester, where only the basic information of the student can be seen. This is just a summary table, where only student summaries can be observed.

After this student summary analysis, next comes the overall test score analysis. In the data preparation, it has been observed that the tests could be classified in three categories. First one is self-test, and other two are ARS test group one and

group two. In this study, analysis has done for each test category. These test overall score analysis has been done by the semester id. There is a possibility that any semester could have this three categories of tests. Some of the semesters could have only one category of test. Each test category, has four different test types, Search, Presentation, Discussion and Report (which is the most important test for completion of the course.). For each test type, the percentage has been calculated against its own maximum points. In the same way test point percentage has been calculated for student wise. Suppose, the maximum point of the Search Type test is 12 and the student got 5 in that test. So the percentage would be 41.66%. In the same way, the for all test type percentage has been calculated.

As this study is about developing a diagnostic analytical system, the analysis needs to be done at the molecular level. Each test have task. To diagnose the problem in a test, the task score for each test has to be analysed too. The task score also analysed in the same way as the test. A test can be attempted in multiple times. So the task data can be found for each attempt. This study compares the highest test attempt task score and latest attempt task score. The reason behind considering the highest score attempt is, assuming that the student score in every task of that test and got the highest score among all of the attempt except the last one. The last attempt task score has been considered because it is assuming that, the student did well in other test task in other attempt but s/he was not doing well in any specific test task which s/he did well in the last attempt. These are reason behind to consider the highest score attempt and latest attempt in the comparison.

For both test and test task score data has been presented in a group bar chart. But before sending this analysis data in bar chart needs to be formatted as the required data format as the visualization framework. There is a possibility that each test may not contain all of the test types. As example, for winter semester 2022, a student may be attempt for ARS search test, ARS Presentation and finally ARS Report. For this student ARS Dicussion is missing. On the other hand this student has attempted all of the test type of Self-test. If this two dataset send to the visual framework, it will act very weirdly. That means as the ARS Discussion data is missing here, but for Self-test discussion data is presented in the bar chart in the discussion section for ARS test it would show the value of ARS Report test data, which is totally wrong data presentation. To solve this issue, if there is any missing test type for test dataset, that test value has been inserted with zero score. After this data manupulation in the test data, when the visual framework tried to show the overview of the test result, the ARS Report data would show in the Report section not in the Discussion section.

In the overall test analysis, another analysis has been done to diagnose how the self test are effecting the students task improvement. This analysis has been done by using spearmen correlation. After applying the spearmen correlation alogrithm in the whole dataset a correlation matrix comes. This correlation matrix has been presented in a heatmap, has been discussed in the next section. Another diagnosis has been done with clustering. This study has used K-mean algorithm for clustering. The total dataset has been divided into three clusters. These clusters are, Good

students, Average students and Going to fail students. With this analysis the tutor can understand that which students need help to get a good score in the course.

5.4 Visualize Analysis

In this study, for visualization, a total of four types of techniques have been used and these visualizations have been shown in the dashboard. The work of this analysis has been done in the frontend project of this study. The first element of any visulization technique is color coding. To get a consistent and easy understanding dashboard it is preferable to keep the color coding in limited number. The number of color code depend on the number of data points. As example, in this study while showing the barchart maximum three colors needed as in the barchart three types of test data are showing. But on the other hand in heatmap will have a range of data point from -1 to 1. So for this visualization a custom color range, which is Blue to Yellow, has been defined, like Figure 5.2. The first visualization technique is, a summary table of student basic information. The row of this table is clickable. If the user click on any row of this table the application will show the details of that student. There is no color code has been used for this table.

Figure 5.2: The used color range

The second type of visualization is, barchart. In this study, the analysis have been done to compare different test score. Group bar chart is the best way to show this comparison. Before showing the comparison data in the barchart some data needed to manipulate. This data manipulation process has been explained in the previous section. Another barchart has been used in this study to show the comparison between different task score of a test. The task comparison has been made between two test attempts. First one is in the attempt where the student got highest score in the test and second attempt is the latest attempt of the test. The reason behind to show the comparison between these two attempts is, to check whether the student did gain the skill in the highest score attempt on the specific task or in the last attempt. It is possible that, may be at the last attempt the student was not concern about his/her other task performance but only for the task where s/he was not doing well in the whole semester. With this comparison it would be easy to understand that is that student improve in the task at the end of the semester or not.

The third visualization technique is a heatmap, which is representing the correlation between different test score. As mentioned before the data points are in range of -1 to +1. When the correlation is high the datapoint is +1, this datapoints represented by the deep blue color and at lowest datapoint is -1, which has been represented by white. There is another possibility of the datapoint is no change at all between test results. Those datapoints are zero, and they are represented by

combination of blue and green. In the Figure 5.2 the right side is representing the color for +1 datapoints and it ends in the left side with -1 datapoints.

Last visualization is a scatter plot. From this plot the three different student clusters can be observed, which can help the tutor to point out the students who needs help to improve their skills. As the scatter plot was showing three clusters here only three colors were used from the pre-defined custom color range. Deep blue is representing the top students, light blue is for the average students and the below average students are represented by the blue and green color combination.

6 Results and Evaluation

In this chapter, the result and the evalution of this analytical system has been discussed in this chapter. At first, the findings will be discussed. With this analytical system what kind of parameter can be dignosed so that the both student and teacher can work on that specific test or task. Then this result will be evaluated with the actual grade to verify how well this analytical system will diagnose the problem with the student improvement and how good to identify the problem with any specific task of any test.

6.1 Findings with Diagnostic Learning Analytics

The finding of this analytical system can be categorized in three. One is the summary table, the second one is overall test result and last two is a correlation between tests and clustering the students.

In the Figure 6.1, the student summary table can be seen. This table has been fetched by semester wise. If the user changes the semester name from the dropdown list, the data of this table will be changed accordingly. For each semester the student information (row) will be different. But the column information would be same. Initially, this table is showing eight columns about the students basic information along with the total achieved score in both ARS and Self test. There could be two types of values in these test columns. First type is a valid number in between zero to thirty seven. This value meant, the specific student attempt the test and got that score in that test. Another type of value is empty, that means the cell will have no score (not even zero) it will represent that. the student did not attempt this test.

Basically, this table is showing the student list who enrolled in the selected semester. This table has its' own functionality. In the Figure 6.1 some parts of it has been pointed out with red boxes. With the number one box, this table can be filtered by every column value. As example, with filter option the user can filter out the table with matrikel number or students' name. After filtering option there is another functionality that has been included in the table is a navigation bar, marked as three. This bar will help the user to check the next page of this table. Each page will load ten rows from the total student list.

This table rows are not clickable. If the user wants to check the details about one student, s/he has to click on the "eye" button, marked as two, under the column named "Details". When this details button is clicked, this analytical system will show another view where the user can see the details of the selected student from the table. In the student details view student's basic information can be seen along with

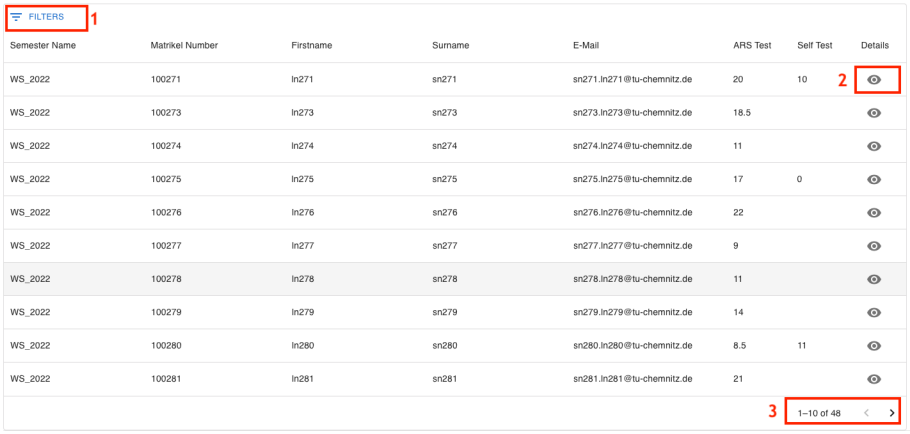


Figure 6.1: Student summary table (Semester wise)

the selected students' all test information in a group bar chart. In this chart, the comparison of test result between different test can be checked. For Self-test, there could more than one attempt. This analytical system, consider only the highest score test attempt in the overall analysis. Task score of the each self-test type also can be seen. If the user click on the any type of self-test in a new pop-up the task score comparison will be seen. In the task, group bar chart, the highest score achieved attempt and last attempt comparison has been shown. The reason behind to select this two attempt is, assuming the student has attempt all of task and got the highest score in each task. And the last attempt is considered by assuming, that the selected student did not attempt all of the task of the test but that specific task where s/he needs to be improved s/he got the good score and improve that task.

At the beginning of the dashboard, overall test information and student count in each test can be seen in a group bar chart. For both charts each group is test type. The ARS tests are divided into two groups. Overall test information also have been shown for those groups. The ARS test groups also have been considered as a group of the group bar chart. In the Figure 6.2, this group bar chart has been illustrated. From this figure, it is clear that the selected semester has four type of test for each test.

A correlation matrix has been used in the dashboard, where how the tests are co-related for the selected semester is illustrated in a heatmap. In Figure 6.3, the correlation between different test types for the selected semester can be seen. When the correlation score is positive it will represent that the test score are positively related, but when the correlation score is negative then it is presenting that the tests are negatively related. That means is one is increasing another one is decresing. From the high correlation score it can be interpreted that, the performance of the student is good from the lecture to final report submission. This student does not need any diagnosis. On the other hand, if the correlation score for ARS test is low but self test is high, which indicates a good sign. The student is also doing well in the course. But the worst case would be when, correlation score for ARS test is high

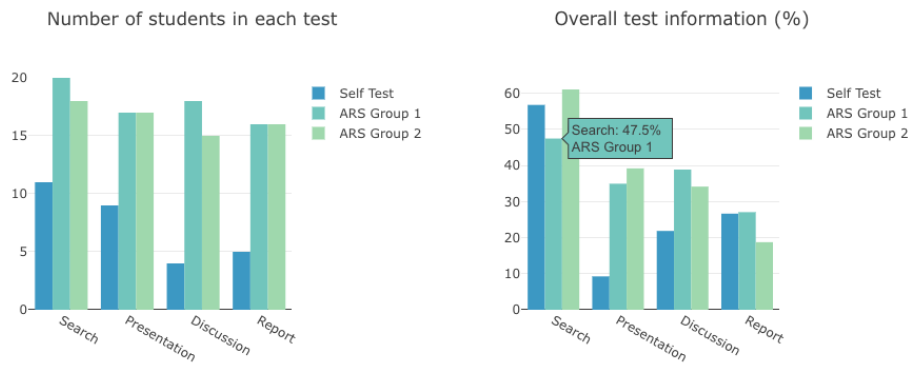


Figure 6.2: Semester wise number of student and overall test analysis in group bar chart

but self test is low. This case needs diagnosis. There could be different scenarios for this case. One is, may be the student understood the topic during the lecture (as ARS test took place that time) and after that, during the self-test, s/he forget it. The second case could be during the self-test the student is only concern about the task in which s/he has weakness. So that student participate only in that specific task of that spefic test type. As example, the student had weakness in the *Task 1* of *Self-test*, so s/he only concern about Task 1. Due to this reason, s/he only took test of that Task and improve it at the end of the semester. The third cased could be, the attend the lecture (thats' why s/he has the ARS test score) and then s/he took part in the Self-test, but finally s/he drop the course for that semester because in the halfway of the semester s/he felt s/he is not going to get his or her expected result in the course. Clustering analysis also has been done with the existing data. But with the existing data no meaningful explanation was found out from that analysis.

6.2 Evaluation

The result of this study has been evaluated with the actual grade of the semester for each student. This evaluation will show how well this analytical system will diagnosis the on going students score in their test and the task.

For performing the evaluation, only the common student has been considered in the both dataset, the final test dataset which has been used in this analytical system and the actual grade dataset, which has been provided from the university. The grade dataset looks like the Figure 6.4. For this evaluation only the yellow marked fields. In the *Type* field, there are two types of test can be found. First on is *Presentation* and second one is *Report*. *Presentation* evaluated by two tutors. So in the dataset for *Presentation* both tutors' score can be found. To calculate the score for the *Presentation*, the formula 6.1 has been used,

$$Presentation\_Score = \frac{\sum (\text{Presentation points})}{\text{Numbers of Tutors}} \tag{6.1}$$

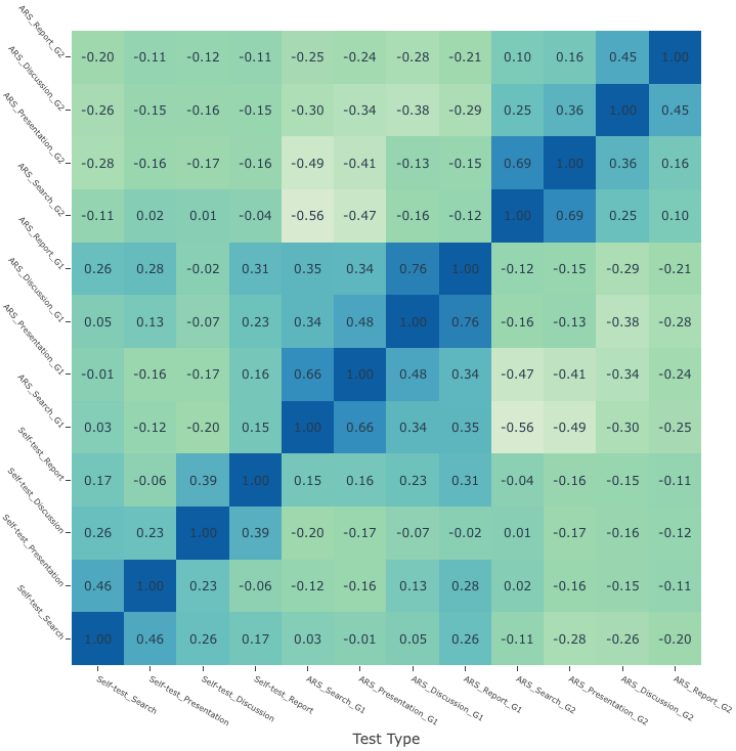


Figure 6.3: Correlation of different tests using heatmap

Here, Numbers of Tutors = 2

And the *Report* score has been calculated by using Formula 6.2. Then these two score need to sum together to get the actual grade of the student in a specific semester. For the actual grade 6.3 formula has been used.

$$Report\_Score = \sum (\text{Report Points}) \tag{6.2}$$

$$Actual\_Grade = Presentation\_Score + Report\_Score \tag{6.3}$$

Now, the analysis score needs to be calculated. For this calculation, the formula 6.4 has been used.

$$Analysis\_Score = \frac{(\text{ARS Test Score} + \text{Self Test Score})}{2} \tag{6.4}$$

Here, both score is in percentage(%)

For the optimal case the above equation 6.4 would work. But according to the existed data, there could be seen two cases. In first case, some students have only attended the ARS test. But they did not took Self-test. In this case, for analysis score only ARS test score has been considered. The second scenario is some students

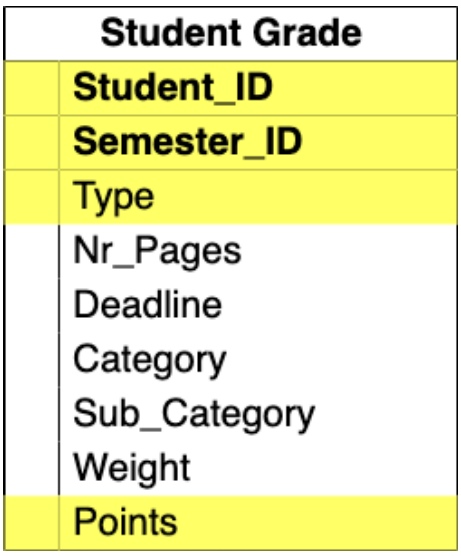


Figure 6.4: Student actual grade dataset

took only Self-test, not ARS-test. For these students, the analysis score has been calculated by using only the Self-test score. This non-optimal scenario score can be measured with formula 6.5

$$Analysis\_Score = \text{Existing Test Score}(\%) \tag{6.5}$$

Both Scores has been calculated semester and student wise. Then a comparison has been made from these two score like Table 6.1. From this comparison table it can be understand that the analysis score is almost similar to the *Actual Grade*. There could be some technical glitch because of that ±5 difference can be seen in the *Analysis Score*. With this comparison table it can be said that, this analytical system is giving a good diagnosis of the student progress in their test and according to the tasks. In this table, the comparison has been done with the case one and two. There was no optimal scenario data found.

Semester ID	Matrikel No	Analysis Score (%)	Actual Grade(%)
WS_2021	100055	32.43	29.5
SS_2020	100110	37.83	37.5
SS_2022	100333	32.43	37.5
SS_2022	100336	44.59	36.0

Table 6.1: Actual grade and Analysis score comparison

7 Conclusion

In order to improve students' learning outcomes, diagnostic learning analytics is a potent and promising method to education. Diagnostic learning analytics may offer educators, administrators, and students themselves useful information and feedback by gathering and analyzing data from many sources, such as online platforms, learning management systems, and other educational technologies.

7.1 Summary of Thesis

In this study a diagnostic learning analytics has been developed to detect the students who are struggling in the test, so that the tutor can point out the problem of the course structure and can improve them later. Educators and institutions are always looking for novel approaches to improve academic performance and student learning experiences in the context of contemporary educational environments. By utilizing the enormous volumes of educational data produced by various learning management systems, online platforms, and educational technology, diagnostic learning analytics provides a data-driven method to achieving these objectives.

In the *Introduction* chapter, the current state of diagnostic analytics in different domains has been discussed. How diagnostic analytics can help to identify the domain-specific problem and help to solve that problem in the next phase. A diagnostic analytical system plays a great role in medical science and in the business world to detect any kind of problem which can cause a big issue. Along with these two fields, a diagnostic can play a great role in other area like human resource area to point out the problem with employee of the company.

The current state of the art discussed different learning analytical systems which helps teacher, student or the whole institute to make the learning system easy and friendly. For developing the analytical system different types of techniques has been discussed. Some of the analytical systems are for analyzing the teacher's teaching patterns and others are for increasing the cognitive load of the student for achieving their success. This analysis has been done by using Bayesian Network. Different steps also have been discussed to build up a learning analytics. In learning domain learning analytics is totally new chapter. It was very rare to find out any developed diagnostic learning previously. But there are many theoritical research paper found where it was mentioned how a diagnostic learning analytics can help the whole learning system by answering *"Why this is happening?"*

The following chapter *State of Techniques*, discussed the techniques which have been used in this study. This study has two different parts. For both part, the most

frequent and famous techniques have been used. The backend has been developed by Python and with its framework. The API calling and creation has been done this backend project. The jupyter notebook also has been used to check inititally the analysis before send the analysis to the frontend. The frontend has been developed with JavaScript and some of its' very widely used frameworks. Basically the dashboards have been developed in the frontend project.

The *Methodology* chapter gives a glimpse of the steps of the whole process for developing this analytical system. This chapter has discussed how the backend and frontend will communicate with each other and what is the main functionality of the different projects in this study. The use case also has been detiled out in this chapter, where it can be seen that what kind of test data are available in the provided APIs. The descriptive analytics is the base for a diagnostic analytics. In this study two decriptive analytic system has been used. Both of them has been discussed in this chapter. After that the main two diagnostic analysis has been described in detail. This study basically used four types of visual techniques. The reasons for choosing these techniques also have been illustrated here.

After that, *Implementation* chapter states, software tools and the project structure of the both frontend and backend. This chapter also detailed out all of the methodology steps in the technical way. After doing so it states the visual techniques. In the *Result and Evalution* chapter conclude with how accurate this analytic system would perform. The formula for evaluation and a comparison table also has been presented. From the last chapter it has been seen that how well and accurately this analytical system would work. The critical conclusion from the chapter six, this thesis has some advantages and shortcomings too. Those are are explained below

Advantages

This diagnostic learning analytics' capacity to offer individualized learning experiences is one of its most important benefits. Learning outcomes may be enhanced when teachers customize their instruction and assistance to each student's requirements by taking into account each one's strengths, weaknesses, learning preferences, and progress. In order to act quickly and offer additional help before problems worsen, instructors may use these analytics to identify troubled children early on. By intervening early, teachers can increase kids' prospects of academic achievement while preventing them from falling behind. Educational institutions may allocate resources more wisely and build curriculums and instructional practices that are more effective with the use of diagnostic learning analytics. Administrators and educators may improve the learning environment and available resources for better overall results by doing so rather of depending solely on intuition. Immediate feedback on a student's performance and development is advantageous. Diagnostic learning analytics may give individuals in-the-moment insights about their areas of strength and weakness, enabling them to modify their study routines and instructional strategies as necessary.

Shortcomings

Diagnostic analytics mainly rely on past data; therefore, if the underlying assumptions change, the insights may no longer be valid or useful. In fast changing sectors or during unusual occurrences, this might be particularly difficult. Diagnostic analytics can show correlations between variables, but it may not always come up with a complete understanding of the underlying context or causal relationships. It can illustrated what occurred, but it might not accurately depict all of the ways that various components interacted. Diagnostic analytics' correctness and dependability are highly dependent on the caliber of the data being examined. Incomplete, inconsistent, or erroneous data might result in false insights and deductions.

7.2 Future Work

The suggested approach had positive results, but there is constant scope for improvement. Diagnostic analytics' precision and efficacy can be improved by integrating AI and machine learning techniques. Large datasets may be analyzed by AI to find trends that will help instructors and students make suggestions that are more exact. The queries also can be more optimised for faster performance. Another improvement can be introduced by implementing an automatic scheduler so that at a certain time this system can fetch the data from the TUC server so that the whole system can be more faster. The correlation analysis needs more research on the negative values. The summary table can be more user friendly if the row colors are different depending on the cluster analysis. It would be simpler to put data-driven decisions into practice if diagnostic analytics were seamlessly integrated with LMS systems, giving instructors immediate access to actionable information inside their instructional settings. Systems for diagnostic learning analytics in the future may give instructors and students feedback in real-time while they are learning. As a result, prompt support and help would be promoted when a student is having trouble or is not engaged in the learning process. It is also very important to introduce ethical considerations and data privacy policies in diagnostic analytics as it is becoming more and more popular.

Bibliography

[1] A. Christopoulos, N. Pellas, and M.-J. Laakso, "A learning analytics theoretical framework for stem education virtual reality applications," *Education Sciences*, vol. 10, no. 11, p. 317, 2020.

[2] T. Elias, "Learning analytics: Definitions, processes and potential," 2011.

[3] M. A. Chatti, A. L. Dyckhoff, U. Schroeder, and H. Thüs, "A reference model for learning analytics," *International Journal of Technology Enhanced Learning*, vol. 4, no. 5-6, pp. 318–331, 2012.

[4] W. Greller and H. Drachsler, "Translating learning into numbers: A generic framework for learning analytics," *Journal of Educational Technology & Society*, vol. 15, no. 3, pp. 42–57, 2012.

[5] G. Siemens, "Learning analytics: The emergence of a discipline," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1380–1400, 2013.

[6] M. A. Chatti, A. Muslim, M. Guliani, and M. Guesmi, "The lava model: Learning analytics meets visual analytics," *Adoption of data analytics in higher education learning and teaching*, pp. 71–93, 2020.

[7] S. Joksimović, V. Kovanović, and S. Dawson, "The journey of learning analytics," *HERDSA Review of Higher Education*, vol. 6, pp. 27–63, 2019.

[8] T. Susnjak, G. S. Ramaswami, and A. Mathrani, "Learning analytics dashboard: a tool for providing actionable insights to learners," *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, p. 12, 2022.

[9] P. Amirian, F. v. Loggerenberg, T. Lang, A. Thomas, R. Peeling, A. Basiri, and S. N. Goodman, "Using big data analytics to extract disease surveillance information from point of care diagnostic machines," *Pervasive and Mobile Computing*, vol. 42, p. 470–486, Dec 2017.

[10] I. Ahmed, M. Ahmad, G. Jeon, and F. Piccialli, "A framework for pandemic prediction using big data analytics," *Big Data Research*, vol. 25, p. 100190, Jul 2021.

[11] [Online]. Available: https://www.kaggle.com/kaggle-survey-2022

[12] H. Khosravi, S. Shabaninejad, A. Bakharia, S. Sadiq, M. Indulska, and D. Gašević, "Intelligent learning analytics dashboards: Automated drill-down recommendations to support teacher data exploration," *Journal of Learning Analytics*, vol. 8, no. 33, p. 133–154, Nov 2021.

[13] H. Akoglu, "User's guide to correlation coefficients," *Turkish journal of emergency medicine*, vol. 18, no. 3, pp. 91–93, 2018.

[14] P. Novianti, D. Setyorini, and U. Rafflesia, "K-means cluster analysis in earthquake epicenter clustering," *International Journal of Advances in Intelligent Informatics*, vol. 3, no. 2, pp. 81–89, 2017.

[15] D. Ifenthaler and C. Widanapathirana, "Development and validation of a learning analytics framework: Two case studies using support vector machines," *Technology, Knowledge and Learning*, vol. 19, pp. 221–240, 2014.

[16] D. E. Hinkle, W. Wiersma, and S. G. Jurs, *Applied statistics for the behavioral sciences*. Houghton Mifflin college division, 2003, vol. 663.

[17] *LAK '11: Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. New York, NY, USA: Association for Computing Machinery, 2011.

[18] M. Khalil and M. Ebner, "What is learning analytics about? a survey of different methods used in 2013-2015."

[19] G. Siemens, "Learning analytics: The emergence of a discipline," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1380–1400, 2013. [Online]. Available: https://doi.org/10.1177/0002764213498851

[20] A. Wilson, C. Watson, T. L. Thompson, V. Drew, and S. Doyle, "Learning analytics: challenges and limitations," *Teaching in Higher Education*, vol. 22, no. 8, p. 991–1007, 2017.

[21] R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," *Foundations of Learning and Instructional Design Technology*, 2018. [Online]. Available: https://edtechbooks.org/lidtfoundations/educational_data_mining_and_learning_analytics

[22] M. ŞahİN and H. Yurdugül, "Educational data mining and learning analytics: past, present and future," *Bartın University Journal of Faculty of Education*, vol. 9, no. 1, pp. 121–131, 2020.

[23] C. Lang, G. Siemens, A. Wise, and D. Gasevic, *Handbook of learning analytics*. SOLAR, Society for Learning Analytics and Research New York, 2017.

[24] R. Kohavi, N. J. Rothleder, and E. Simoudis, "Emerging trends in business analytics," *Communications of the ACM*, vol. 45, no. 8, p. 45–48, Aug 2002.

[25] D. Delen and S. Ram, "Research challenges and opportunities in business analytics," *Journal of Business Analytics*, vol. 1, no. 1, p. 2–12, Jan 2018.

[26] T. A. Runkler, *Introduction.* Wiesbaden: Springer Fachmedien, 2020, p. 1–4. [Online]. Available: https://doi.org/10.1007/978-3-658-29779-4_1

[27] S. Tyagi, "Using data analytics for greater profits," *Journal of Business Strategy*, vol. 24, no. 3, p. 12–14, Jan 2003.

[28] S. Madden, "From databases to big data," *IEEE Internet Computing*, vol. 16, no. 3, pp. 4–6, 2012.

[29] J. Fan, F. Han, and H. Liu, "Challenges of big data analysis," *National science review*, vol. 1, no. 2, pp. 293–314, 2014.

[30] W. L. Hays, *Statistics for the social sciences.* Holt Rinehart and Winston, 1978.

[31] B. L. Chance and A. J. Rossman, *Investigating statistical concepts, applications and methods.* Duxbury Harrisonburg, 2006.

[32] P. S. Mann, *Introductory statistics.* John Wiley & Sons, 2007.

[33] C. N. Knaflic, *Storytelling with data: A data visualization guide for business professionals.* John Wiley & Sons, 2015.

[34] K. Hamdane, A. El Mhouti, and M. Massar, "How can learning analytics techniques improve the learning process? an overview," in *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET).* IEEE, 2022, pp. 1–5.

[35] S. M. Famurewa, L. Zhang, and M. Asplund, "Maintenance analytics for railway infrastructure decision support," *Journal of Quality in Maintenance Engineering*, vol. 23, no. 3, p. 310–325, Jan 2017.

[36] G. Shao, S.-J. Shin, and S. Jain, "Data analytics using simulation for smart manufacturing," in *Proceedings of the Winter Simulation Conference 2014*, Dec 2014, p. 2192–2203.

[37] N. Reddicharla, M. A. Ali, R. Cornwall, A. Shah, S. Soni, J. Isambertt, and S. Sabat, "Next-generation data-driven analytics- leveraging diagnostic analytics in model based production workflows." OnePetro, Mar 2019. [Online]. Available: https://onepetro.org/SPEMEOS/proceedings/19MEOS/3-19MEOS/D031S032R003/218555

[38] H. Kale and N. Anute, *HR Analytics and its Impact on Organizations Performance*, Aug 2022.

[39] C. Biriowu and N. Kalio, "Talent analytics and employee retention in nigeria organizations," *Journal of Human Resources*, vol. 9, p. 1–12, Apr 2020.

[40] A. Alghamdi, T. Alsubait, A. Baz, and H. Alhakami, "Healthcare analytics: A comprehensive review," *Engineering, Technology Applied Science Research*, vol. 11, no. 1, p. 6650–6655, Feb 2021.

[41] S. Shafqat, S. Kishwer, R. U. Rasool, J. Qadir, T. Amjad, and H. F. Ahmad, "Big data analytics enhanced healthcare systems: a review," *The Journal of Supercomputing*, vol. 76, no. 3, p. 1754–1799, Mar 2020.

[42] M. Khalifa and I. Zabani, "Utilizing health analytics in improving the performance of healthcare services: A case study on a tertiary care hospital," *Journal of Infection and Public Health*, vol. 9, no. 6, p. 757–765, Nov 2016.

[43] D. Ifenthaler, "Learning analytics for school and system management," *OECD Digital Education Outlook 2021 Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, p. 161, 2021.

[44] N. Sclater and J. Mullan, "Learning analytics and student success–assessing the evidence," *Joint Information of Systems Committee (JISC). CC by*, vol. 4, 2017.

[45] M. Larrain and G. Kaiser, "Interpretation of students' errors as part of the diagnostic competence of pre-service primary school teachers," *Journal für Mathematik-Didaktik*, vol. 43, no. 1, pp. 39–66, 2022.

[46] A. Wijaya, H. Retnawati, W. Setyaningrum, K. Aoyama *et al.*, "Diagnosing students' learning difficulties in the eyes of indonesian mathematics teachers." *Journal on Mathematics Education*, vol. 10, no. 3, pp. 357–364, 2019.

[47] W. M. Murphy, "From e-mentoring to blended mentoring: increasing students' developmental initiation and mentors' satisfaction," *Academy of Management Learning & Education*, vol. 10, no. 4, pp. 606–622, 2011.

[48] V. L. Uskov, J. P. Bakken, L. Aluri, N. Rayala, M. Uskova, K. Sharma, and R. Rachakonda, "Learning analytics based smart pedagogy: student feedback," in *Smart Education and e-Learning 2018 5*. Springer, 2019, pp. 117–131.

[49] M. Cukurova, M. Khan-Galaria, E. Millán, and R. Luckin, "A learning analytics approach to monitoring the quality of online one-to-one tutoring," *Journal of Learning Analytics*, vol. 9, no. 22, p. 105–120, May 2022.

[50] P. Kosmas, A. Ioannou, and S. Retalis, "Moving bodies to moving minds: A study of the use of motion-based games in special education," *TechTrends*, vol. 62, no. 6, p. 594–601, Nov 2018.

[51] E. A. Alrehaili and H. Al Osman, "A virtual reality role-playing serious game for experiential learning," *Interactive Learning Environments*, vol. 30, no. 5, p. 922–935, May 2022.

[52] C.-H. Chen, G.-Z. Liu, and G.-J. Hwang, "Interaction between gaming and multistage guiding strategies on students' field trip mobile learning performance and motivation," *British Journal of Educational Technology*, vol. 47, no. 6, p. 1032–1050, 2016.

[53] V. Shute, S. Rahimi, G. Smith, F. Ke, R. Almond, C.-P. Dai, R. Kuba, Z. Liu, X. Yang, and C. Sun, "Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games," *Journal of Computer Assisted Learning*, vol. 37, no. 1, p. 127–141, 2021.

[54] Serrano-Laguna, J. Torrente, P. Moreno-Ger, and B. Fernández-Manjón, "Application of learning analytics in educational videogames," *Entertainment Computing*, vol. 5, no. 4, p. 313–322, Dec 2014.

[55] J. Yu, W. Ma, J. Moon, and A. Denham, "Developing a stealth assessment system using a continuous conjunctive model," *Journal of Learning Analytics*, vol. 9, no. 33, p. 11–31, Dec 2022.

[56] C. S. Loh, Y. Sheng, and D. Ifenthaler, *Serious Games Analytics: Theoretical Framework*, ser. Advances in Game-Based Learning. Cham: Springer International Publishing, 2015, p. 3–29. [Online]. Available: https://doi.org/10.1007/978-3-319-05834-4_1

[57] C. Alonso-Fernández, A. Calvo-Morata, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Game learning analytics:: Blending visual and data mining techniques to improve serious games and to better understand player learning," *Journal of Learning Analytics*, vol. 9, no. 33, p. 32–49, Dec 2022.

[58] Y. J. Kim and D. Ifenthaler, *Game-Based Assessment: The Past Ten Years and Moving Forward*, ser. Advances in Game-Based Learning. Cham: Springer International Publishing, 2019, p. 3–11. [Online]. Available: https://doi.org/10.1007/978-3-030-15569-8_1

[59] M.-T. Cheng, Y.-W. Lin, and H.-C. She, "Learning through playing virtual age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters," *Computers Education*, vol. 86, p. 18–29, Aug 2015.

[60] M. Liu, Y. Cai, S. Han, and P. Shao, "Understanding student navigation patterns in game-based learning," *Journal of Learning Analytics*, vol. 9, no. 33, p. 50–74, Dec 2022.

[61] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 3, no. 11, p. 361–362, Mar 2009.

[62] A. A. Mubarak, H. Cao, and S. A. Ahmed, "Predictive learning analytics using deep learning model in moocs' courses videos," *Education and Information Technologies*, vol. 26, no. 1, p. 371–392, Jan 2021.

[63] Y. Choi and J. Kim, "Learning analytics for diagnosing cognitive load in e-learning using bayesian network analysis," *Sustainability*, vol. 13, no. 1818, p. 10149, Jan 2021.

[64] B. T. M. Wong, "Learning analytics in higher education: an analysis of case studies," *Asian Association of Open Universities Journal*, vol. 12, no. 1, p. 21–40, Jan 2017.

[65] V. L. Uskov, J. P. Bakken, L. Aluri, N. Rayala, M. Uskova, K. Sharma, and R. Rachakonda, "Learning analytics based smart pedagogy: Student feedback," in *Smart Education and e-Learning 2018*, ser. Smart Innovation, Systems and Technologies, V. L. Uskov, R. J. Howlett, L. C. Jain, and L. Vlacic, Eds. Cham: Springer International Publishing, 2019, p. 117–131.

[66] N. Sclater, A. Peasgood, and J. Mullan, "Learning analytics in higher education," *London: Jisc. Accessed February*, vol. 8, no. 2017, p. 176, 2016.

[67] J. Konert, K. Richter, F. Mehm, S. Göbel, R. Bruder, and R. Steinmetz, "Pedale – a peer education diagnostic and learning environment," *Journal of Educational Technology Society*, vol. 15, no. 4, p. 27–38, 2012.

[68] OECD, *OECD Digital Education Outlook 2021 Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*. OECD Publishing, Jun 2021, google-Books-ID: Yj8yEAAAQBAJ.

[69] T. Kärner, J. Warwas, and S. Schumann, "A learning analytics approach to address heterogeneity in the classroom: The teachers' diagnostic support system," *Technology, Knowledge and Learning*, vol. 26, no. 1, p. 31–52, Mar 2021.

[70] E. Koh, A. Shibani, J. P.-L. Tan, and H. Hong, "A pedagogical framework for learning analytics in collaborative inquiry tasks: an example from a teamwork competency awareness program," in *Proceedings of the Sixth International Conference on Learning Analytics Knowledge*, ser. LAK '16. New York, NY, USA: Association for Computing Machinery, Apr 2016, p. 74–83. [Online]. Available: https://dl.acm.org/doi/10.1145/2883851.2883914

[71] B. Deonovic, P. Chopade, M. Yudelson, J. de la Torre, and A. A. von Davier, *Application of Cognitive Diagnostic Models to Learning and Assessment*

Systems, ser. Methodology of Educational Measurement and Assessment. Cham: Springer International Publishing, 2019, p. 437–460. [Online]. Available: https://doi.org/10.1007/978-3-030-05584-4_21

[72] R. Levy, "Dynamic bayesian network modeling of game-based diagnostic assessments," *Multivariate Behavioral Research*, vol. 54, no. 6, p. 771–794, Nov 2019.

[73] R. L. Sparks and B. J. Lovett, "Applying objective diagnostic criteria to students in a college support program for learning disabilities," *Learning Disability Quarterly*, vol. 36, no. 4, p. 231–241, Nov 2013.

[74] H. Chen and P. Mohapatra, "Using service brokers for accessing backend servers for web applications," *Journal of Network and Computer Applications*, vol. 28, no. 1, p. 57–74, Jan 2005.

[75] I. Costa, J. Araujo, J. Dantas, E. Campos, F. A. Silva, and P. Maciel, "Availability evaluation and sensitivity analysis of a mobile backend-as-a-service platform," *Quality and Reliability Engineering International*, vol. 32, no. 7, p. 2191–2205, 2016.

[76] T. A. Budd, "An introduction to object-oriented programming 3rd ed."

[77] S. Thompson, *Haskell: the craft of functional programming*, 3rd ed. Harlow, England; New York: Addison Wesley, 2011.

[78] W. Python, "Python," *Python Releases Wind*, vol. 24, 2021.

[79] W. Chun, "Core python programming," vol. 1, 2001.

[80] T. J. Stevens and W. Boucher, *Python programming for biology.* Cambridge University Press, 2015.

[81] D. B. Fridsma, "Health informatics: a required skill for 21st century clinicians," 2018.

[82] Y. J. Kim, B. Ganbold, and K. G. Kim, "Web-based spine segmentation using deep learning in computed tomography images," *Healthcare informatics research*, vol. 26, no. 1, pp. 61–67, 2020.

[83] L. D. Dalcin, R. R. Paz, P. A. Kler, and A. Cosimo, "Parallel distributed computing using python," *Advances in Water Resources*, vol. 34, no. 9, p. 1124–1139, Sep 2011.

[84] K. J. Millman and M. Aivazis, "Python for scientists and engineers," *Computing in Science Engineering*, vol. 13, no. 2, p. 9–12, Mar 2011.

[85] M. Richardson and S. Wallace, *Getting started with raspberry PI.* " O'Reilly Media, Inc.", 2012.

[86] R. Chityala and S. Pudipeddi, *Image processing and acquisition using Python*. CRC Press, 2020.

[87] S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in python," *PeerJ*, vol. 2, p. e453, 2014.

[88] S. Raschka, J. Patterson, and C. Nolet, "Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Information*, vol. 11, no. 4, 2020. [Online]. Available: https://www.mdpi.com/2078-2489/11/4/193

[89] L. Igual, S. Seguí, L. Igual, and S. Seguí, *Introduction to data science*. Springer, 2017.

[90] I. Idris, *Python data analysis*. Packt Publishing Ltd, 2014.

[91] J. Hao and T. K. Ho, "Machine learning made easy: a review of scikit-learn package in python programming language," *Journal of Educational and Behavioral Statistics*, vol. 44, no. 3, pp. 348–361, 2019.

[92] M. Grinberg, *Flask web development: developing web applications with python*. " O'Reilly Media, Inc.", 2018.

[93] M. R. Mufid, A. Basofi, M. U. H. Al Rasyid, I. F. Rochimansyah *et al.*, "Design an mvc model using python for flask framework development," in *2019 International Electronics Symposium (IES)*. IEEE, 2019, pp. 214–219.

[94] D. Ghimire, "Comparative study on python web frameworks: Flask and django," 2020.

[95] P. Lokhande, F. Aslam, N. Hawa, J. Munir, and M. Gulamgaus, "Efficient way of web development using python and flask," 2015.

[96] J. Chan, R. Chung, and J. Huang, *Python API Development Fundamentals: Develop a full-stack web application with Python and Flask*. Packt Publishing Ltd, 2019.

[97] Z. Liang, Z. Liang, Y. Zheng, B. Liang, and L. Zheng, "Data analysis and visualization platform design for batteries using flask-based python web service," *World Electric Vehicle Journal*, vol. 12, no. 4, p. 187, 2021.

[98] C. A. Trianti, B. Kristianto *et al.*, "Integration of flask and python on the face recognition based attendance system," in *2021 2nd International Conference on Innovative and Creative Information Technology (ICITech)*. IEEE, 2021, pp. 164–168.

[99] D. A. Anggoro and N. C. Aziz, "Implementation of k-nearest neighbors algorithm for predicting heart disease using python flask," *Iraqi Journal of Science*, pp. 3196–3219, 2021.

[100] A. Lakshmanarao, M. R. Babu, and M. M. Bala Krishna, "Malicious url detection using nlp, machine learning and flask," in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 2021, pp. 1–4.

[101] A. Yaganteeswarudu and P. Dasari, "Diabetes analysis and risk calculation–auto rebuild model by using flask api," in *Image Processing and Capsule Networks: ICIPCN 2020*. Springer, 2021, pp. 299–308.

[102] I. Anshori, S. Harimurti, M. B. Rama, R. E. Langelo, L. P. Yulianti, G. Gumilar, M. Yusuf, S. Prastriyanti, B. Yuliarto, H. Nugrahapraja *et al.*, "Web-based surface plasmon resonance signal processing system for fast analyte analysis," *SoftwareX*, vol. 18, p. 101057, 2022.

[103] W. McKinney, *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012.

[104] W. McKinney *et al.*, "pandas: a foundational python library for data analysis and statistics," *Python for high performance and scientific computing*, vol. 14, no. 9, pp. 1–9, 2011.

[105] I. Stančin and A. Jović, "An overview and comparison of free python libraries for data mining and big data analysis," in *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2019, pp. 977–982.

[106] M. J. De Smith, M. F. Goodchild, and P. Longley, *Geospatial analysis: a comprehensive guide to principles, techniques and software tools*. Troubador publishing ltd, 2007.

[107] P. Lemenkova, "Processing oceanographic data by python libraries numpy, scipy and pandas," *Aquatic Research*, vol. 2, no. 2, pp. 73–91, 2019.

[108] ——, "Python libraries matplotlib, seaborn and pandas for visualization geospatial datasets generated by qgis," *Analele stiintifice ale Universitatii" Alexandru Ioan Cuza" din Iasi-seria Geografie*, vol. 64, no. 1, pp. 13–32, 2020.

[109] D. Kossmann, "The state of the art in distributed query processing," *ACM Computing Surveys (CSUR)*, vol. 32, no. 4, pp. 422–469, 2000.

[110] S. Hagedorn, S. Kläbe, and K.-U. Sattler, "Putting pandas in a box," in *CIDR*, 2021.

[111] W. McKinney *et al.*, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, vol. 445, no. 1. Austin, TX, 2010, pp. 51–56.

[112] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, "Array programming with numpy," *Nature*, vol. 585, no. 7825, pp. 357–362, 2020.

[113] F. Nelli and F. Nelli, "The numpy library," *Python Data Analytics: With Pandas, NumPy, and Matplotlib*, pp. 49–85, 2018.

[114] J. Ranjani, A. Sheela, and K. P. Meena, "Combination of numpy, scipy and matplotlib/pylab-a good alternative methodology to matlab-a comparative analysis," in *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*. IEEE, 2019, pp. 1–5.

[115] R. Kumar, "Future for scientific computing using python," *International Journal of Engineering Technologies and Management Research*, vol. 2, no. 1, pp. 30–41, 2015.

[116] A. Sapre and S. Vartak, "Scientific computing and data analysis using numpy and pandas," 2020.

[117] M. Turk, "Analysis and visualization of multi-scale astrophysical simulations using python and numpy," SLAC National Accelerator Lab., Menlo Park, CA (United States), Tech. Rep., 2008.

[118] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in neuroscience*, p. 267, 2013.

[119] R. Nishino and S. H. C. Loomis, "Cupy: A numpy-compatible library for nvidia gpu calculations," *31st confernce on neural information processing systems*, vol. 151, no. 7, 2017.

[120] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, "Scikit-learn: Machine learning without learning the machinery," *GetMobile: Mobile Computing and Communications*, vol. 19, no. 1, pp. 29–33, 2015.

[121] B. M. Randles, I. V. Pasquetto, M. S. Golshan, and C. L. Borgman, "Using the jupyter notebook as a tool for open science: An empirical study," in *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2017, pp. 1–2.

[122] J. M. Perkel, "Why jupyter is data scientists' computational notebook of choice," *Nature*, vol. 563, no. 7732, pp. 145–147, 2018.

[123] J. F. Pimentel, L. Murta, V. Braganholo, and J. Freire, "A large-scale study about quality and reproducibility of jupyter notebooks," in *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*, 2019, pp. 507–517.

[124] G. Moraila, A. Shankaran, Z. Shi, and A. M. Warren, "Measuring reproducibility in computer systems research," Technical report, University of Arizona, Tech. Rep., 2014.

[125] N. Schaduangrat, S. Lampa, S. Simeon, M. P. Gleeson, O. Spjuth, and C. Nantasenamat, "Towards reproducible computational drug discovery," *Journal of cheminformatics*, vol. 12, pp. 1–30, 2020.

[126] R. DePratti, "Using jupyter notebooks in a big data programming course," *Journal of Computing Sciences in Colleges*, vol. 34, no. 6, pp. 157–159, 2019.

[127] E. Shook, D. D. Vento, A. Zonca, and J. Wang, "Gisandbox: A science gateway for geospatial computing," in *Proceedings of the Practice and Experience on Advanced Research Computing*, ser. PEARC '18. New York, NY, USA: Association for Computing Machinery, Jul 2018, p. 1–7. [Online]. Available: https://dl.acm.org/doi/10.1145/3219104.3219150

[128] M. B. Milligan, "Jupyter as common technology platform for interactive hpc services," in *Proceedings of the Practice and Experience on Advanced Research Computing*, ser. PEARC '18. New York, NY, USA: Association for Computing Machinery, Jul 2018, p. 1–6. [Online]. Available: https://dl.acm.org/doi/10.1145/3219104.3219162

[129] E. J. Menke, "Series of jupyter notebooks using python for an analytical chemistry course," 2020.

[130] J. Blechschmidt and O. G. Ernst, "Three ways to solve partial differential equations with neural networks—a review," *GAMM-Mitteilungen*, vol. 44, no. 2, p. e202100006, 2021.

[131] A. Khurana and S. R. Rosenthal, "Towards holistic "front ends" in new product development," *Journal of Product Innovation Management: An international publication of the product development & management association*, vol. 15, no. 1, pp. 57–74, 1998.

[132] A. Wright, D. F. Sittig, J. S. Ash, J. Feblowitz, S. Meltzer, C. McMullen, K. Guappone, J. Carpenter, J. Richardson, L. Simonaitis *et al.*, "Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems," *Journal of the American Medical Informatics Association*, vol. 18, no. 3, pp. 232–242, 2011.

[133] S. A. Murphy and V. Kumar, "The front end of new product development: a canadian survey," *R&D Management*, vol. 27, no. 1, pp. 5–15, 1997.

[134] A. Bhalla, S. Garg, and P. Singh, "Present day web-development using reactjs," *International Research Journal of Engineering and Technology*, vol. 7, no. 05, 2020.

[135] A. Boduch, *Flux architecture*. Packt Publishing Ltd, 2016.

[136] A. Banks and E. Porcello, *Learning React: functional web development with React and Redux*. " O'Reilly Media, Inc.", 2017.

[137] D. Bugl, *Learn React Hooks: Build and refactor modern React. js applications using Hooks*. Packt Publishing Ltd, 2019.

[138] A. S. Sari and R. Hidayat, "Designing website vaccine booking system using golang programming language and framework react js," *JISICOM (Journal of Information System, Informatics and Computing)*, vol. 6, no. 1, pp. 22–39, 2022.

[139] S. Weaver, S. D. Shank, S. J. Spielman, M. Li, S. V. Muse, and S. L. Kosakovsky Pond, "Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes," *Molecular biology and evolution*, vol. 35, no. 3, pp. 773–777, 2018.

[140] A. B. Bhardwaj and R. Sharada, "Product billing scheme employing image processing and edge computing using reactjs & ibm cloud."

[141] H. W. Lie and B. Bos, *Cascading style sheets: Designing for the web, Portable Documents*. Addison-Wesley Professional, 2005.

[142] D. Mazinanian, N. Tsantalis, and A. Mesbah, "Discovering refactoring opportunities in cascading style sheets," in *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, 2014, pp. 496–506.

[143] P. Geneves, N. Layaïda, and V. Quint, "On the analysis of cascading style sheets," in *Proceedings of the 21st international conference on World Wide Web*, 2012, pp. 809–818.

[144] H. Stormer, "Personalized websites for mobile devices using dynamic cascading style sheets," *International Journal of Web Information Systems*, vol. 1, no. 2, pp. 83–88, 2005.

[145] L. Guo, "Best ui experience: Material design in action," in *The First Line of Code: Android Programming with Kotlin*. Springer, 2022, pp. 519–574.

[146] A. Boduch, *React Material-UI Cookbook: Build captivating user experiences using React and Material-UI*. Packt Publishing Ltd, 2019.

[147] S. Xiong, X. Wang, and Z. Lan, "Model research of visual report components," *Procedia Computer Science*, vol. 208, pp. 478–485, 2022.

[148] M. Joglekar, H. Garcia-Molina, and A. Parameswaran, "Interactive data exploration with smart drill-down," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 1, pp. 46–60, 2019.

[149] M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal: The Journal of Medical Association of Malawi*, vol. 24, no. 3, p. 69–71, Sep 2012.

[150] A.-N. Liu, L.-L. Wang, H.-P. Li, J. Gong, and X.-H. Liu, "Correlation between posttraumatic growth and posttraumatic stress disorder symptoms based on pearson correlation coefficient: A meta-analysis," *The Journal of nervous and mental disease*, vol. 205, no. 5, pp. 380–389, 2017.

[151] J. Benesty, J. Chen, and Y. Huang, "On the importance of the pearson correlation coefficient in noise reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 757–765, 2008.

[152] M. R. Anderberg, *Cluster analysis for applications: probability and mathematical statistics: a series of monographs and textbooks*. Academic press, 2014, vol. 19.

[153] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Annals of Data Science*, vol. 2, pp. 165–193, 2015.

[154] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14. Oakland, CA, USA, 1967, pp. 281–297.

[155] J. Wu and J. Wu, "Cluster analysis and k-means clustering: an introduction," *Advances in K-Means clustering: A data mining thinking*, pp. 1–16, 2012.

[156] A. Singh, A. Yadav, and A. Rana, "K-means with three different distance metrics," *International Journal of Computer Applications*, vol. 67, no. 10, 2013.

[157] H. Shiravi, A. Shiravi, and A. A. Ghorbani, "A survey of visualization systems for network security," *IEEE Transactions on visualization and computer graphics*, vol. 18, no. 8, pp. 1313–1329, 2011.

[158] I. Khan and A. Pardo, "Data2u: Scalable real time student feedback in active learning environments," in *Proceedings of the sixth international conference on learning analytics & knowledge*, 2016, pp. 249–253.

[159] S. Shemwell, "Futuristic decision-making," *Executive Briefing Business Value from*, 2005.

[160] K. Kuosa, D. Distante, A. Tervakari, L. Cerulo, A. Fernández, J. Koro, and M. Kailanto, "Interactive visualization tools to improve learning and teaching in online learning environments," *International journal of distance education technologies (IJDET)*, vol. 14, no. 1, pp. 1–21, 2016.

[161] M. Sahin and D. Ifenthaler, "Visualizations and dashboards for learning analytics: A systematic literature review," *Visualizations and dashboards for learning analytics*, pp. 3–22, 2021.

[162] W. Humphries, J. Gawrilow, S. Turner, M. Perez-Quinones, and S. Edwards, "Helping students visualize their grade performance," in *2006 Annual Conference & Exposition*, 2006, pp. 11–683.

References from Professorship of Computer Engineering

[TUC1] U. U. Shegupta, R. Schmidt, M. Springwald, and W. Hardt, "Audience response system - an inclusion of blended mentoring technology in computer engineering education," in *2020 IEEE Frontiers in Education Conference (FIE)*, 2020, pp. 1–5.

[TUC2] R. Schmidt, U. U. Shegupta, and W. Hardt, "Implementing audience response system in structured mentoring processes to increase learning motivation," in *2022 IEEE Frontiers in Education Conference (FIE)*, 2022, pp. 1–9.

[TUC3] E. Moser, U. U. Shegupta, K. Ihsberner, O. Jalilov, R. Schmidt, and W. Hardt, "Designing digital self-assessment and feedback tools as mentoring interventions in higher education," 2022.

Chemnitzer Informatik-Berichte

In der Reihe der Chemnitzer Informatik-Berichte sind folgende Berichte erschienen:

CSR-21-01 Marco Stephan, Batbayar Battseren, Wolfram Hardt, UAV Flight using a Monocular Camera, März 2021, Chemnitz

CSR-21-02 Hasan Aljzaere, Owes Khan, Wolfram Hardt, Adaptive User Interface for Automotive Demonstrator, Juli 2021, Chemnitz

CSR-21-03 Chibundu Ogbonnia, René Bergelt, Wolfram Hardt, Embedded System Optimization of Radar Post-processing in an ARM CPU Core, Dezember 2021, Chemnitz

CSR-21-04 Julius Lochbaum, René Bergelt, Wolfram Hardt, Entwicklung und Bewertung von Algorithmen zur Umfeldmodellierung mithilfe von Radarsensoren im Automotive Umfeld, Dezember 2021, Chemnitz

CSR-22-01 Henrik Zant, Reda Harradi, Wolfram Hardt, Expert System-based Embedded Software Module and Ruleset for Adaptive Flight Missions, September 2022, Chemnitz

CSR-23-01 Stephan Lede, René Schmidt, Wolfram Hardt, Analyse des Ressourcenverbrauchs von Deep Learning Methoden zur Einschlagslokalisierung auf eingebetteten Systemen, Januar 2023, Chemnitz

CSR-23-02 André Böhle, René Schmidt, Wolfram Hardt, Schnittstelle zur Datenakquise von Daten des Lernmanagementsystems unter Berücksichtigung bestehender Datenschutzrichtlinien, Januar 2023, Chemnitz

CSR-23-03 Falk Zaumseil, Sabrina, Bräuer, Thomas L. Milani, Guido Brunnett, Gender Dissimilarities in Body Gait Kinematics at Different Speeds, März 2023, Chemnitz

CSR-23-04 Tom Uhlmann, Sabrina Bräuer, Falk Zaumseil, Guido Brunnett, A Novel Inexpensive Camera-based Photoelectric Barrier System for Accurate Flying Sprint Time Measurement, März 2023, Chemnitz

CSR-23-05 Samer Salamah, Guido Brunnett, Sabrina Bräuer, Tom Uhlmann, Oliver Rehren, Katharina Jahn, Thomas L. Milani, Güunter Daniel Rey, NaturalWalk: An Anatomy-based Synthesizer for Human Walking Motions, März 2023, Chemnitz

CSR-24-01 Seyhmus Akaslan, Ariane Heller, Wolfram Hardt, Hardware-Supported Test Environment Analysis for CAN Message Communication, Juni 2024, Chemnitz

Chemnitzer Informatik-Berichte

CSR-24-02 S. M. Rizwanur Rahman, Wolfram Hardt, Image Classification for Drone Propeller Inspection using Deep Learning, August 2024, Chemnitz

CSR-24-03 Sebastian Pettke, Wolfram Hardt, Ariane Heller, Comparison of maximum weight clique algorithms, August 2024, Chemnitz

CSR-24-04 Md Shoriful Islam, Ummay Ubaida Shegupta, Wolfram Hardt, Design and Development of a Predictive Learning Analytics System, August 2024, Chemnitz

CSR-24-05 Sopuluchukwu Divine Obi, Ummay Ubaida Shegupta, Wolfram Hardt, Development of a Frontend for Agents in a Virtual Tutoring System, August 2024, Chemnitz

CSR-24-06 Saddaf Afrin Khan, Ummay Ubaida Shegupta, Wolfram Hardt, Design and Development of a Diagnostic Learning Analytics System, August 2024, Chemnitz

Chemnitzer Informatik-Berichte

ISSN 0947-5125

Herausgeber: Fakultät für Informatik, TU Chemnitz
Straße der Nationen 62, D-09111 Chemnitz