



Utilization of Big Data Technology in the Analysis of Academic Data for Students of the Faculty of Computer Science IBI Kosgoro 1957 for Decision Making

Boy Firmansyah^{1*}, Nuraini Purwandari²

IBI Kosgoro 1957, Jakarta

Corresponding Author: Boy Firmansyah boy@ibi-k57.ac.id

ARTICLE INFO

Keywords: Big Data, Decision-Making, Predictive Model, Academic Data, Student Performance

Received : 3 January

Revised : 18 January

Accepted: 20 February

©2024 Firmansyah, Purwandari: This is an open-access article distributed under the terms of the [Creative Commons Atribusi 4.0 Internasional](https://creativecommons.org/licenses/by/4.0/).



ABSTRACT

This research discusses the use of Big Data technology in analyzing student academic data at the Ibi Kosgoro 1957 Faculty of Computer Science with the main aim of optimizing the decision-making process. The main focus of the article is to create a predictive model that can predict student academic success based on extensive data analysis. The research steps involve collecting and processing academic data, including grades, number of courses taken, and other variables that may influence student performance. The collected data is then used to train predictive models using machine learning techniques. The predictive model that is built aims to provide decision-making recommendations to academics. By utilizing Big Data, this article explores deep insights into academic patterns that may be difficult to detect with conventional methods. It is hoped that the research results can make a positive contribution in increasing the efficiency of academic management and help related parties in designing more targeted intervention strategies. In addition, it is hoped that the implementation of this predictive model can support efforts to increase student academic success at the Ibi Kosgoro 1957 Faculty of Computer Science

INTRODUCTION

The IBI Kosgoro 1957 Faculty of Computer Science has a large volume of data, including data on student grades, attendance, academic activities, and much more. This data, if analyzed properly, can provide valuable insights to help make better academic decisions. However, so far, there are still several data processing functions in the campus information system that are less efficient in managing and analyzing the data.

In recent years, Big Data technology has reached a level of maturity that allows educational institutions to exploit its enormous potential. Big Data includes the concept of collecting, storing and analyzing data in very large and varied amounts. In the context of higher education, this includes data from various sources such as learning management systems (LMS), campus administration systems (SIKAD), student satisfaction surveys, and even data from social media. By combining all this data, institutions can analyze student behavior, academic trends, and the factors that influence their academic success.

This research aims to explore the potential use of Big Data technology in the context of higher education, with a particular focus on the analysis of student academic data. By collecting, integrating, and analyzing this data, this research will attempt to achieve several goals, namely

1. Understand how Big Data technology can be used to manage and analyze student academic data.
2. Identify patterns and trends in academic data that can provide valuable insights for academic decision making.
3. Creating a predictive model that can help predict student academic success.
4. Provide recommendations and guidelines for the Faculty of Computer Science to adopt Big Data technology in academic decision making.

LITERATURE REVIEW

In the following stage, the researcher carried out a comparison with previous research carried out by Chintia Putri Demasari et al in March 2023 with the title Implementation of Big Data Analysis to Predict Student Graduation at the Faculty of Engineering, Langlangbuana University (Demasari et al., 2023). This research discusses a number of students who did not graduate on time . Based on existing observations and data, there are still many students who do not graduate on time, which causes a decrease in scores in the accreditation assessment criteria. Graduation of students on time is the criteria for accreditation assessment. In order to get good results from accreditation, predictions of graduation rates will be needed using big data analysis methods, so that anticipating delays in student graduation can be used as reference material for predicting student graduation. With the digitization of data which has resulted in a very rapid surge in data, Big Data Analytics is needed.

Further research was conducted by Irika Widiasanti et al in June 2023 with the title Implementation of the Use of Big Data in Analyzing Factors that Influence Student Performance in Exam Results (Widiasanti et al., 2023). The research aims to study the variables that influence student performance in exam scores. Data sets that appear very large and complex are known as big data, which are analyzed to find useful patterns. The data collected and analyzed in

this research came from various sources, such as national and international journals, academic publications, reports and books.

METHODOLOGY

Understanding Big Data

A lot of big data comes from continuous transaction processes, interactions, and data observations, which makes database software unable to manage and analyze large amounts of data (Liliana et al., 2019).

Big Data is a concept that includes the collection, storage, processing and analysis of large amounts of complex and varied data, to generate valuable insights and support decision making. Big Data theory includes several interrelated and influencing dimensions, which form the foundation for understanding and applying this concept in various contexts.

First of all, its large volume is the main characteristic of Big Data. The amount of data being produced at any moment through digital interactions, sensors, and other online resources is reaching an unprecedented scale. This large amount of data requires a strong technological infrastructure for management and storage. Therefore, the Big Data concept relies on the system's ability to handle and store data on a large scale.

Apart from volume, data variety is also an important focus. Big Data consists of various types of data, including structured, semi-structured, and unstructured data. This data can come from different sources such as social media, IoT sensors, text, images, and videos. The ability to integrate and analyze this varied data is a challenge, but also a potential source of deeper insights.

Speed in data processing is also an important element in Big Data theory. Today's data sources produce information at high speed, and organizations need to be able to process data in real-time to gain a competitive advantage. With technology and algorithms capable of handling this speed, Big Data provides the ability to take action or make decisions quickly.

Additionally, the value of Big Data lies in its ability to provide relevant and in-depth insights. Through advanced data analysis techniques, such as machine learning and data mining, organizations can discover patterns, trends, and correlations that may not be visible with traditional methods. This helps organizations in smarter decision making and enables better innovation.

Ethical and privacy challenges are also an integral part of Big Data theory. With greater access to personal data, there are concerns about how this data is used and protected. There needs to be special attention to privacy policies and ethics in the collection, storage and use of Big Data to ensure that the benefits of data analysis do not come at the expense of individual privacy rights.

Overall, Big Data theory creates a conceptual foundation involving volume, variety, velocity, value, as well as ethical and privacy aspects. By understanding these dimensions, organizations can take the necessary steps to manage and exploit the enormous potential of Big Data, provide competitive advantage, and generate positive impact in various industries and sectors.

Understanding Decision Making Systems

The decision making system is the development of an advanced level of computerized management information that is designed in an interactive way for the user. And its interactive nature aims to encourage integration of various elements of the decision-making process, such as analysis techniques, then policies, procedures, as well as insight and management experience, in order to form a decision-making framework that is truly supportive (Firmansyah & Wihandar, 2020).

Decision Making System Theory (Decision Support System - DSS) includes the concepts and principles that underlie the development, implementation and use of systems that support the decision making process in an organization. These systems are designed to provide relevant information and analytical tools to decision makers, assisting them in formulating better and more informed decisions. Several key elements in Decision Making Systems theory include the structure, function, and benefits of DSS.

First, the DSS structure consists of three main components, namely the database, model and user interface. Databases store data necessary for decision analysis, models provide a mathematical or statistical framework for the evaluation of alternatives, and user interfaces enable interaction between the user and the system. This structure provides the foundation for DSS to provide accurate information and effective analytical tools.

DSS functions involve the ability to collect, store, process, and present information. This system can integrate data from various sources, perform complex analysis, and produce output that is easy to understand. This function aims to provide maximum support to decision makers in understanding the context, risks and implications of possible decision alternatives.

The main benefit of Decision Making Systems theory is improving the quality of decisions. DSS help reduce uncertainty by providing more complete and relevant information, speeding up the decision-making process, and enabling decision makers to evaluate various scenarios. This increase in efficiency and effectiveness contributes to the overall performance of the organization.

Furthermore, this theory emphasizes the interaction between humans and technology. A user-friendly and easy-to-understand user interface is the key to optimizing the contribution of decision makers in using DSS. Psychological and sociological factors also play an important role in the design and implementation of systems that are acceptable and adopted by users.

Additionally, Decision Making Systems theory highlights the important role of planning and implementation. A careful planning process and a good understanding of decision makers' needs are necessary to ensure a DSS can be integrated well within the organizational environment. Successful implementation involves full engagement of users, effective training, and regular evaluation to ensure system suitability and effectiveness.

Lastly, Decision Making Systems theory also recognizes the evolution of technology and the dynamics of the business environment. The system must be able to adapt to changing needs, new technology, and market conditions.

Flexibility and scalability are necessary principles to maintain DSS relevance in the long term.

Overall, Decision Making Systems theory provides a holistic view of how a system can strengthen the decision making process by providing the right information at the right time.

Basic Website Concepts

A web is a hypertext platform that allows data elements such as text, sound, images, multimedia and animation to interact with each other. A website displays information that can be accessed via the internet. One type of website that contains personal content, such as photos, videos and articles, is called a blog (Rafi'i, 2008).

Website theory involves understanding the concept, structure, and purpose of an online platform designed to provide information, interaction, and services to users. As an integral part of the digital era, websites have a crucial role in providing access to information, facilitating communication, and supporting various goals, both for individuals and organizations.

Flow of Research

At this stage, the flow of research carried out sequentially is discussed (Purwandari & Fauzi, 2022). The flow diagram of the research can be seen in Figure 1.

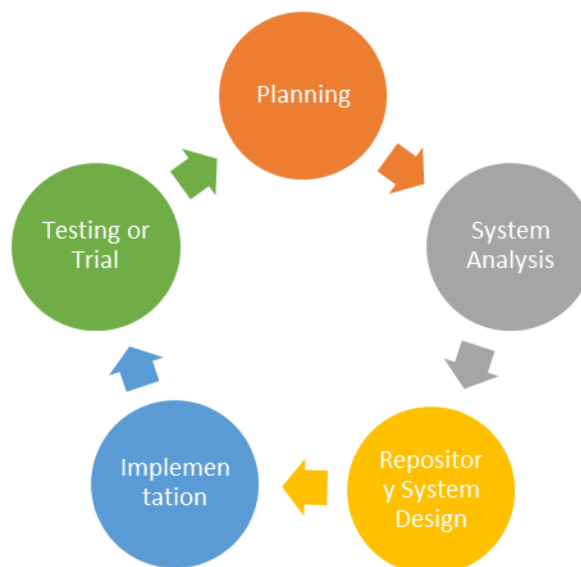


Figure 1. Flow Diagram of Research Steps

Attached below is an explanation of the stages in the flow diagram/research steps as shown in Figure 1.

1. Planning

As a stage of determining the problem, study programs will experience difficulties in documenting documents that support accreditation and finding the

required documents (Silvanie, Astried; Subekti, 2022). As a result, big data technology is needed to store these supporting documents.

The aim of the research is to design and build Big Data Technology in the Analysis of Academic Data for Students of the IBI Kosgoro 1957 Faculty of Computer Science for Decision Making.

Apart from that, the researcher also carried out data collection as follows:

2. Literature Study

One of the goals of this stage is to collect relevant articles, theories and materials about the basic concepts of big data technology. In addition, this website helps researchers understand the research subject.

Big Data Analysis

The method used in this research is big data analysis. Big data analysis is a lot of data, the processing is very fast, and the data is very diverse. With these large data sizes, different processing technologies than traditional storage are often required. A larger data volume is beneficial for higher education lecturers, because the greater the data volume, the more output the higher education lecturers can obtain from the data extraction process (Machdum & Ardhianto, 2020). Big data is data that exceeds the processing capacity of current database systems. With this type of database architectural structure, this data is very large and very fast. To extract value from this data, one must choose another way to process it.

Refers to the amount of big data that is generated every second, meaning collecting large amounts and volumes of sometimes unstructured data. For example Twitter, Instagram, Whatsapp status and chat text data, user clickstream from websites. This data flow can be up to thousands of Terrabytes (TB) per second. Data can be accessed at a very fast speed so that it can be used immediately at that second (more real time). One proof of this is the online operating system based on Microsoft Silverlight, web-based office applications such as Office365, cloud storage in other applications such as Dropbox and GDrive.

Big data has weaknesses in terms of accuracy and validity, so in-depth analysis of big data is needed to make the right decisions. The truth mark indicates the accuracy and reliability of the data (Permana & Silvanie, 2021). Some of the technologies used in Big Data Analytics applications are as follows:

1. Apache Hadoop

Apache Hadoop is open-source software that is used to store data in a cluster or single entity and run applications. Hadoop can connect different computers to work together and network while using it. Then Hadoop also stores and processes large data in a distributed manner using the MapReduce programming model. The storage can also be carried out in parallel in groups with hundreds of servers because it consists of thousands of machines.



Figure 2. Apache Hadoop Map

2. MongoDB

MongoDB is a trusted next-generation database that helps drive business change using NoSQL concepts. MongoDB is popular among developers, as it has advantages over Apache Hadoop in terms of performance and flexibility and uses structured documents such as JSON. Additionally, MongoDB has high performance, auto-scaling, and high availability.

3. Cassandra

Apache Cassandra is a database management product implemented by Apache that is scalable and designed to manage large data across multiple servers. Cassandra is suitable for hybrid and multi-cloud environments. Apart from that, Cassandra's ability to achieve faster performance. Several large companies have used Cassandra such as Facebook, IBM, Digg, Reddit, Apple, Twitter, and others



Figure 3. IT Infrastructure for Big Data

The technology in Big Data Analytics allows universities to get an accurate picture of student profiles efficiently. For example, universities can use Big Data Analytics to measure past performance.

Learning Analytics

One of the useful aspects of Big Data Analytics is its usability, which can be analyzed directly from student activities or in real-time experiences of students, such as: Lectures, payments, online activities, courses, martial arts studies Processed data analysis. Then it is analyzed in real time so that decisions can be made that are used to predict whether students will graduate on time. With big data analysis carried out at the IBIK57 Faculty of Computer Science, there may be findings that have the potential to drop out, so that testing can be implemented in the form of support to improve student achievement (Angellia, 2020).

Academic Analytics

Subjects that can be analyzed in academic analysis are students who have excelled. With academic analytics, real-time data analysis, as a metric, can be carried out for students who graduate on time, so that students who are successful and who are not successful can be seen compared to other students (Purwandari et al., 2021).

Analytical Process

Real-time analysis in business processes in Institutions is the use of Big Data Analytics. Aggregate data from academic data or event data obtained from students, faculty and a number of stakeholders related to all processes and activities carried out at IBIK57, is used to progress process analyzes that exploit processes and obtain process models for new initiatives (Subekti et al., 2023). Like the IBIK57 Faculty of Computer Science, it graduated a total of 169 students between 2016-2023. But process analysis is not only limited to defining business processes, it also enables compliance verification, error detection, delay prediction, decision support, and process design recommendations.

The design includes 3 main processes, namely Learning Analytics, Academic Analytics, and Process Analytics, which can be seen in Figure 3. The proposed design includes learning analytics, Learning Analytics, Academic Analytics, and Process Analytics. For learning analysis with various data sources to help make decisions, such as student information system data, class management, e-learning, student assessments and financial data. Where this data can help with learning analysis (analytics), then you can use results assessment data, course assessments, staff reviews, faculty reviews, and financial evaluation data for learning analysis. Here you can use data sources from Student Information System activity log data, lesson management log data, online learning activity log data, student assessment entry data and financial activity log data for process analysis. Data from these sources is then stored and then analyzed and processed using real-time forecasts (Syah & Angellia, 2020). In addition, the results of the analysis and predictions will be displayed as an analytical presentation on the dashboard which is expected to improve decision making at the IBIK57 Faculty of Computer Science in order to increase student graduation rates and improve the performance of the Faculty of Computer Science.



Figure 4. Proposed Plan for Implementing Big Data Analysis

Needs Analysis

Requirements Analysis is the stage of collecting the data needed to be used as a basis for information system development. The needs analysis carried out by researchers took the form of field studies (observations), collecting material sources (literature studies) and searching for relevant research (Syamsu Hidayat Rino Subekti, 2022). Relevant research is used as a benchmark for writing and integration between material sources .

Information Systems Design

The main target users of this information system are all academic components of the IBIK57 Faculty of Computer Science. The expected information system design is:

- 1) Administrators (Admin) can process academic data accurately and quickly.
- 2) The Head of the Study Program can exercise academic control over students and lecturers.
- 3) Lecturers can see the classes taught, teaching schedules, and input student grades.

- 4) Students can view class data, course schedules, and daily and exam grades.

RESULTS AND DISCUSSION

A. Decision Tree

Decision trees are predictive models that can be used for classification and regression tasks. How a decision tree works can be explained in several stages:

Feature Selection

Decision trees start by selecting the features that best separate data based on target values (class or output variables). This feature selection is carried out based on criteria such as Gini impurity (for classification) or variance reduction (for regression).

1. Node Creation

After the best feature is selected, the decision tree creates a node at that point. This node represents the testing condition on the selected feature. Each node has two or more branches that represent the results of testing that condition.

2. Data Splitting

Data is divided into subsets based on the results of testing conditions on nodes. Each branch leads to a child node corresponding to the condition value that the data sample satisfies.

3. Recursion (Recursion)

The above process is repeated recursively for each new child node. The decision tree continues to select the best features and create new nodes until the stopping condition is met. The stopping condition can be reaching the maximum depth level, reaching the minimum number of samples on a node, or when there is no significant increase in class separation or regression values.

4. Label Determination (Label Assignment)

When it reaches the final node or leaf node, the decision tree assigns a class label or regression value to that leaf. This becomes the prediction result for the data sample that passes through the tree and reaches the leaf.

5. Prediction

To make a prediction for a new sample, the sample passes through the decision tree from the root node until it reaches one of the leaves. The label on the leaf becomes a prediction for that sample.

Suppose we have a customer dataset that includes two features: age and income, and the target variable is whether the customer bought the product or not (binary classification). The decision tree can select the features that best separate purchasing and non-purchasing customers based on Gini impurity. At the first level, the decision tree might select an age feature and create a node with the condition "Age $\leq x$ ". The data is divided into two branches: one for ages less than or equal to x , and one for ages greater than x . At each branch, the decision tree may select a revenue feature and create a new node for each branch. This process is repeated until leaves form. Once leaves are formed, the decision tree determines a label (e.g., "Buy" or "Not Buy") for each leaf based on the majority of classes of samples that reach that leaf. Predictions for new samples can then be made by traversing the tree from the root node until reaching the leaves that correspond to the sample feature values. The advantages of decision trees include good interpretability, the ability to handle categorical data, and a tendency to

handle non-linearities in the data. However, decision trees are also susceptible to overfitting, which can be overcome by pruning techniques or the use of ensemble methods such as Random Forest.

B. Binary Classification Model

In this case, we want to use a decision tree to create a binary classification model that can predict whether a student has taken or has not taken a course. The dataset consists of two features, namely semester and course, and the target variable is whether the student has taken the course or not.

The following are the steps for using a decision tree for this case:

1. Selection of Separation Criteria

The decision tree will choose the best splitting criteria to separate the data based on the target. In this case, we might use Gini impurity or entropy as criteria to evaluate how well the features are separated.

2. First Feature Selection

The decision tree will choose the first feature that is best for separation. For example, if Gini impurity is used, the decision tree will look for the most effective separation based on the Gini impurity value for each feature selection.

3. First Node Creation

After selecting the first feature, the decision tree will create a first node (root) that represents the split based on that feature. This node may have two branches, one for students who have feature values below a certain threshold and one for students who have feature values above that threshold.

4. Feature Selection in Next Branches

This process will be repeated in each branch. The decision tree will select the best features in each branch to separate the data.

5. Node Creation in Each Branch

Each branch will have a new node representing a split based on the selected feature. This process will be repeated until it reaches the specified maximum leaf or depth level.

6. Determination of Labels on Leaves

The leaves in the decision tree will have labels or classes that indicate whether the student has taken the course or not. For example, "Taking" or "Not Taking".

7. Predictions for New Students

To make a prediction for a new student, we traverse the decision tree from the root node until we reach the leaf that corresponds to the feature value of the new student. The label on the leaf is a prediction of whether the student will take the course or not.

It is important to remember that interpreting decision tree results will provide insight into the features that are most influential in the decision of whether a student will take a course or not. These features can provide information about what factors most influenced the decision.

C. Data Set

To predict or see who has taken how many courses, the dataset will have a structure that reflects the relevant features and target variables. In this case, we can have a dataset with several features that describe student characteristics and

a target variable that reflects the number of courses the student has taken. Here is the dataset structure:

Table 1. Student Characteristics Dataset

Student Name	Age	Major	Semester	Course taken
Student1	20	Informatic	5	10
Student2	22	Economy	6	12
Student3	19	Psychology	4	8
Student4	21	Law	7	15
Student5	20	Commucation	5	11
...

Information:

Student Name : The name of the individual student.

Age : The student's age when the data was taken.

Major : The major taken by the student.

Semester : The semester the student is currently undertaking at the time of data collection.

Courses Taken : The number of courses taken by students in a particular semester.

The target variable or feature that you want to predict is "Courses Taken", while other features such as "Age", "Major", and "Semester" can be used as predictor features.

With this dataset, we can use a regression algorithm (such as linear regression) if we want to predict the number of courses taken numerically. If we want to classify based on certain categories (for example, students taking "many" or "few" courses), then we can use classification algorithms such as decision trees or other classification algorithms.

In the analysis process, we can identify patterns and relationships between features such as age, major, and semester with the number of courses taken by students. This process can provide insight into what factors most influence students' decisions in taking courses.

D. Regression Model

To build a regression model that predicts the number of courses a student will take based on variables such as age, major, and semester, we can use a linear regression algorithm. This algorithm tries to find a linear relationship between predictor variables (features) and target variables (number of courses taken).

Following are the steps to build a linear regression model:

1. Prepare Dataset

Make sure the dataset is prepared with relevant variables, including the target variable "Courses Taken" and predictor variables such as "Age", "Major", and "Semester".

2. Data Separation

Split the dataset into two parts: one for model training and one for model testing.

3. Initialization of Linear Regression Model

Initialize the linear regression model that we will train with the training data.

4. Train Model

Train the model using the training data. The model will try to fit the regression line that best fits the data, so that it can predict "Courses Taken" based on predictor variables.

5. Model Evaluation

Model evaluation uses test data to ensure that the model can generalize well on never-before-seen data. Commonly used evaluation metrics for regression include Mean Squared Error (MSE), Mean Absolute Error (MAE), and others.

6. Predictions for New Students

Once the model is assessed, you can use it to make predictions about the number of courses new students might take by inputting their age, major, and semester values into the model.

Here is an implementation using Python and the scikit-learn library:

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
# Split the dataset into training and testing data
X_train,
# Initialize the linear regression model
regression_model = LinearRegression()
# Train the model using training data
regression_model.fit(X_train, y_train)
# Prediction using test data
predictions = regression_model.predict(X_test)
# Model evaluation
mse = mean_squared_error(y_test, predictions)
print(f'Mean Squared Error: {mse}')
# Predictions for new students
new_student_data = [[20, 5]] # Example 20 years old and currently in semester 5
predicted_courses = regression_model.predict(new_student_data)
print(f'Predicted Number of Courses: {predicted_courses}')
```

E. PHP Implementation for Linear Regression Models and Making Predictions

```
<?php
// Function for calculating linear regression
function linearRegression($x, $y) {
    $n = count($x);
    $sumX = array_sum($x);
    $sumY = array_sum($y);
    $sumXY = 0;
    $sumXSquare = 0;
    for ($i = 0; $i < $n; $i++) {
        $sumXY += ($x[$i] * $y[$i]);
```

```
$sumXSquare += ($x[$i] * $x[$i]);
}
$slope = ($n * $sumXY - $sumX * $sumY) / ($n * $sumXSquare - $sumX *
$sumX);
$intercept = ($sumY - $slope * $sumX) / $n;
return [$slope, $intercept];
}
// Function to make predictions
function predict($x, $slope, $intercept) {
    return $slope * $x + $intercept;
}
// Data
$usia = [20, 22, 19, 21, 20];
$semester = [5, 6, 4, 7, 5];
$mataKuliahTerambil = [10, 12, 8, 15, 11];
// Building a linear regression model
list($slope, $intercept) = linearRegression($semester, $mataKuliahTerambil);
// Predictions for new students
$newStudentSemester = 8;
$predictedCourses = predict($newStudentSemester, $slope, $intercept);
echo "Prediksi Jumlah Mata Kuliah: " . round($predictedCourses, 2);
?>
```

The PHP program above shows how to calculate a linear regression model and make predictions for new students based on the semester input. It should be noted that this implementation is simple and does not include model validation or further data processing that may be needed in more complex situations.

predictive model

Creating a predictive model to predict student academic success involves several stages, including data collection, data processing, feature selection, model selection, model training, and model evaluation. Here are the steps:

F. Data Collection

Collect relevant data to build models. This may include information such as academic history (grades, number of courses taken, etc.), participation in extracurricular activities, demographic data, and other factors that may influence academic success.

1. Data Processing

Perform data cleaning and processing. This involves handling missing values, normalizing the data, and converting categorical data into a form that can be used by the model.

2. Data Exploration (EDA)

Conduct data exploration to understand the characteristics and patterns in the data. Identify relationships between variables, find outliers, and gain insights that can help in feature selection.

3. Feature Selection

Select the features that are most relevant and influence academic success. This can be done based on EDA analysis, university policy, or domain understanding.

4. Model Selection

Select an appropriate machine learning model for the task of predicting academic success. Models such as logistic regression, decision trees, random forests, and neural networks can be used depending on the complexity of the problem and the amount of data available.

5. Data Sharing

Split the dataset into training set and test set. The training set is used to train the model, while the test set is used to test the model's performance on never-before-seen data.

6. Model Training

Train the model using the training set. During the training process, the model will learn patterns and relationships in the data that can be used to make predictions.

7. Model Evaluation

Evaluate the model using the test set. Use evaluation metrics appropriate for the classification problem, such as accuracy, precision, recall, and F1-score. Make adjustments as necessary to improve model performance.

8. Model Optimization

If necessary, optimize the model by adjusting parameters or using techniques such as cross-validation to improve the generalization of the model to never-before-seen data.

9. Model Implementation

Implement the model on a system or platform that can be used by related parties. This may include integration with university information systems or the development of applications that can provide predictions of academic success.

10. Monitoring and Maintenance

Continue monitoring and maintaining the model over time. Ensure the model remains relevant to changes in data or university policy. If necessary, re-train the model periodically with the latest data.

It is important to remember that at every stage, the transparency and interpretability of the model is essential, especially when used in the context of academic decisions. Students and related parties must be able to understand what factors influence predictions of academic success.

G. Find Out How Long It Takes Students to Complete Their Studies

To find out how long it will take for a student to complete their studies, we can use information from the student table, course table, and grade table. To achieve this goal, the steps are as follows:

1. Student and Course Identification

Determine the student you want to analyze, and find the courses the student has taken.

2. Calculate the Total Credits Taken

Add up the credits (semester credit units) from the courses taken by the student. It provides information about a student's academic progress to date.

3. Calculate the Total Credits Required

Determine the total credits required to complete the study program. This can be calculated based on the applicable study program or curriculum requirements.

4. Calculate the Remaining Credits to be Taken

Subtract the total credits that have been taken from the total credits required. The result is the remaining credits that must be taken to complete the study program.

5. Calculate the Number of Semesters to Take

Determine how many more semesters the student must take courses to complete the remaining credits required. This can be calculated by dividing the remaining credits by the number of credits taken by the student each semester.

6. Estimated Time of Study Remaining

Determine the estimated time required to complete the remaining credits. This can be calculated based on the number of semesters to be taken and the estimated length of one semester.

If we consider that the maximum credits that can be taken each semester is 24 and students have to retake courses they do not pass in the same next semester (odd or even), there needs to be some adjustments to the SQL program. Here's a SQL program that takes that scenario into account:

-- Calculate the total credits you have taken

SELECT

SUM(mk.sks) AS total_sks_diambil

FROM

mahasiswa m

JOIN

nilai n ON m.nomor_pokok_mahasiswa = n.nomor_pokok_mahasiswa

JOIN

mata_kuliah mk ON n.kode_mata_kuliah = mk.kode_mata_kuliah

WHERE

m.nomor_pokok_mahasiswa = 'NomorPokokMahasiswa';

-- Calculate the total credits required

SELECT

SUM(sks) AS total_sks_diperlukan

FROM

mata_kuliah

WHERE

semester <= 8;

-- Calculate the remaining credits that must be taken

SELECT

(total_sks_diperlukan - total_sks_diambil) AS sisa_sks

FROM

(SELECT

SUM(CASE WHEN n.nilai = 'E' THEN 0 ELSE mk.sks END) AS total_sks_diambil

FROM

mahasiswa m


```
JOIN
  nilai n ON m.nomor_pokok_mahasiswa = n.nomor_pokok_mahasiswa
JOIN
  mata_kuliah mk ON n.kode_mata_kuliah = mk.kode_mata_kuliah
WHERE
  m.nomor_pokok_mahasiswa = 'NomorPokokMahasiswa'
  AND mk.semester <= 8) AS sks_diambil,
(SELECT
  SUM(sks) AS total_sks_diperlukan
FROM
  mata_kuliah
WHERE
  semester <= 8) AS sks_diperlukan;
-- Calculate the number of semesters you have to take
SELECT
  CEIL(sisa_sks / 24) AS jumlah_semester
FROM
  (SELECT
    (total_sks_diperlukan - total_sks_diambil) AS sisa_sks
  FROM
    (SELECT
      SUM(CASE WHEN n.nilai = 'E' THEN 0 ELSE mk.sks END) AS
total_sks_diambil
    FROM
      mahasiswa m
    JOIN
      nilai n ON m.nomor_pokok_mahasiswa = n.nomor_pokok_mahasiswa
    JOIN
      mata_kuliah mk ON n.kode_mata_kuliah = mk.kode_mata_kuliah
    WHERE
      m.nomor_pokok_mahasiswa = 'NomorPokokMahasiswa'
      AND mk.semester <= 8) AS sks_diambil,
    (SELECT
      SUM(sks) AS total_sks_diperlukan
    FROM
      mata_kuliah
    WHERE
      semester <= 8) AS sks_diperlukan) AS sisa_sks;
-- Estimated length of time required to complete the remaining credits
SELECT
  LEAST(CEIL(jumlah_semester), 14) AS estimasi_lama_studi
FROM
  (SELECT
    CEIL(sisa_sks / 24) AS jumlah_semester
  FROM
    (SELECT
```

```
(total_sks_diperlukan - total_sks_diambil) AS sisa_sks
FROM
(SELECT
SUM(CASE WHEN n.nilai = 'E' THEN 0 ELSE mk.sks END) AS
total_sks_diambil
FROM
mahasiswa m
JOIN
nilai n ON m.nomor_pokok_mahasiswa = n.nomor_pokok_mahasiswa
JOIN
mata_kuliah mk ON n.kode_mata_kuliah = mk.kode_mata_kuliah
WHERE
m.nomor_pokok_mahasiswa = 'NomorPokokMahasiswa'
AND mk.semester <= 8) AS sks_diambil,
(SELECT
SUM(sks) AS total_sks_diperlukan
FROM
mata_kuliah
WHERE
semester <= 8) AS sks_diperlukan) AS sisa_sks) AS jumlah_semester;
```

CONCLUSIONS

From all the changes that have been discussed in SQL queries, we can conclude several main points:

Credit Calculation

The credit calculation takes into account that if a student gets an E grade, then the credits for that course are considered 0 (not taken).

Maximum credits per semester

The maximum credits that can be taken each semester is 24, according to the limits given.

Re-Taking Courses That Didn't Pass

Students are required to retake courses they do not pass in the same next semester (odd or even).

Maximum Semester

The maximum semester permitted is semester 14, according to the limits given.

Estimated Study Duration

The estimated length of study is calculated based on the remaining credits that must be taken, taking into account the scenarios described above.

It is important to note that a given SQL query must be tailored to the table structure and specific needs. In addition, in real contexts, predictive models or more sophisticated academic management systems may be more suitable for providing estimates of length of study and recommendations for taking courses based on a student's academic history.

RECOMMENDATIONS

Expansion and Diversification of Data Sources:

The article can be enriched by considering the expansion and diversification of data sources. Thinking about integrating non-academic data, such as demographic data, participation in extracurricular activities, and other information, can provide a more comprehensive picture of the factors that influence academic success.

Model Validation and Testing:

It is recommended to validate and test predictive models that have been built with different datasets or use cross-validation techniques. This can provide additional confidence in the reliability and generalizability of the predictive model.

Model Interpretability:

Added a section discussing the interpretability of the predictive models that have been created. A clear explanation of how the model makes decisions can increase understanding and acceptance from the parties involved.

FURTHER STUDY

This research still has limitations, so it is necessary to carry out further research related to the topic of Utilization of Big Data Technology in the Analysis of Academic Data for Students in order to improve this research and add insight to readers.

REFERENCES

- Angellia, F. (2020). Analisis Perancangan Optimalisasi Database Akses Pemakaian Ruang Kelas di Perguruan Tinggi Berbasis RFID. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(2), 175-187. <https://doi.org/10.35957/jatisi.v7i2.369>
- Demasari, C. P., Ariyanti, N., Muhanar, P. G., & Ilman, B. (2023). Implementasi Analisis Big Data untuk Memprediksi Kelulusan Mahasiswa Fakultas Teknik Universitas Langlangbuana sehingga membuka peluang untuk Big Data Analytic di Universitas Langlangbuana sebagai berikut: *Jurnal Accounting Information System (AIMS)*, 6(1), 48-52.
- Firmansyah, B., & Wihandar, A. M. (2020). Sistem Pendukung Keputusan Monitoring & Evaluasi Kinerja Dosen Program Studi Informatika fakultas Ilmu Komputer IBI Kosgoro 1957 Menggunakan Metode ANALYTIC HIERARCHY PROCESS (AHP). *Jurnal Nasional Informatika*, 1(2), 127-142.
- Liliana, L., Vera, D., Wijaya, A. S., & Bernanda, D. Y. (2019). Penggunaan Big Data Untuk Menganalisis Tingkat Keberhasilan Siswa Menempuh Mata Kuliah. *Prosiding Seminar Nasional Teknoka*, 4(2502), I77-I82. <https://doi.org/10.22236/teknoka.v4i0.4208>

- Machdum, D. M., & Ardhianto, E. (2020). Analisis Belajar Daring Pada Pandemi Covid-19 Di Jurusan Sistem Informasi Institut Bisnis Dan Informatika Kosgoro 1957. *Jurnal Sistem Informasi Bisnis (JUNSIBI)*, 1(2), 96–103. <https://doi.org/10.55122/junsibi.v1i2.177>
- Permana, D. S., & Silvanie, A. (2021). Prediksi Penyakit Jantung Menggunakan Support Vector Machine dan Python pada Basis Data Pasien di Cleveland. *JUNIF: Jurnal Nasional Informatika*, 2(1), 29–34.
- Purwandari, N., & Fauzi, A. (2022). Penerapan Technopreneur Dasar Bagi Guru-Guru Pendidikan Anak Usia Dini dan Sekolah Dasar Tingkat Kecamatan Cikarang. 9(1).
- Purwandari, N., Baskoro, B. H., & Mutholib, E. A. (2021). Perancangan Sistem Informasi Penggajian Pegawai Pada Perusahaan Xyz Menggunakan Java Spring Framework. *Jurnal Sistem Informasi Bisnis (JUNSIBI)*, 2(2), 48–58. <https://doi.org/10.55122/junsibi.v2i2.306>
- Rafi'i, M. (2008). *Cara Cepat Membuat Website untuk Pemula* (1st ed.). Tiara Aksa.
- Silvanie, Astried; Subekti, R. (2022). CHATBOT APPLICATION FOR ACADEMIC FAQ IN IBI-K57 WITH LSTM AND. 5(1), 19–27. <https://doi.org/10.33387/jiko>
- Subekti, R., Silvanie, A., & Firmansyah, B. (2023). Analisa Aturan Asosiasi Dalam Data Tracer Study Ibi-Kosgoro1957 Menggunakan Algoritma Apriori. *Jurnal Disprotek*, 14(1), 19–26. <https://doi.org/10.34001/jdpt>
- Syah, S., & Angellia, F. (2020). Analisa Pemanfaatan Teknologi Daring Mahasiswa Ibi Kosgoro 1957 Selama Pandemi Covid-19 Dengan Regresi Logistik Ordinal. *Junif*, 1(1), 43–56. <http://ejournal-ibik57.ac.id/index.php/junif/article/view/30>
- Syamsu Hidayat Rino Subekti, A. S. (2022). Chatbot untuk konsultasi akademik mahasiswa menggunakan natural language processing (NLP) di Institut Bisnis dan Informatika Kosgoro 1957. *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, 6(Vol 6 No 2 (2022)), 396–410. <https://doi.org/10.52362/jisamar.v6i2.780>
- Widiasanti, I., Adelia, J. T., Rosidin, L., Viola, M. F., & Daniarista, M. (2023). Implementasi Penggunaan Big Data Dalam Menganalisis Faktor Yang Mempengaruhi Kinerja Siswa Dalam Hasil Ujian. *Akademika*, 12(01), 239–250. <https://doi.org/10.34005/akademika.v12i01.2607>