Student Scholarship and Placement Analytics with Machine Learning using Power BI

Mahendra Kachchhavah[1], Moneesh Kadam[2], Meet Patel[3], Asmita Rajpoot[4], Nishva Patel[5]
Tammana Patel[6], Rupali Attarde[7]
[1,2,3,4,5,6,7]*Department of Computer Engineering*

***Abstract*—Schools and universities have access to numerous sources of information on their students, including student marks, institutions they attend, and modes of transport they use. However, such data are fragmented, and their potential is not explored due to the lack of integration analytics platforms. This research presents a holistic analytics approach designed as the result of Microsoft Power BI and Machine Learning integration. The application integrates various data sources and employs structured analytics techniques, such as data pre-processing, modeling, and visualization to extract valuable insights from the raw data. The solution consists of six different dashboards, namely Overview, Academic Performance, Placement Analysis, Scholarship Analysis, School Analysis, and Life Changes. Dashboards provide flexible data analysis capabilities and enable users to perform various transformations to the dataset via interactive visualizations, KPIs, and filters. Moreover, a predictive machine learning model based on linear regression is built in the application to predict future placement trends and scholarships allocation. The outcomes of the experiment demonstrate that the system facilitates decision-making process, improves transparency and reduces labor costs. Furthermore, the solution is scalable, self-learning and highly valuable for educational institutions as it leverages predictive analytics.**

***Index Terms*—Power BI, Machine Learning, Student Analytics, Dashboard, Predictive Analytics**

I. INTRODUCTION

Recently, there has been an increasing tendency among institutions to incorporate data analysis approaches with regard to improving academic performance. A considerable number of data is accumulated from numerous sources, namely, records concerning students' academic achievements, information on their placement activities, details on scholarships, and information related to schools [7]. At the same time, a poor systemization of these data does not help optimize the process of working with them [6],[10].

Manual analysis is one of the most widely used ways of working with information. It should be emphasized that manual data processing entails not only a high level of time consumption but also the risk of making errors. Besides, it cannot be scaled and does not provide real-time insights, making it difficult for organizations to observe any changes and track performance [6]-[8].

Such platforms as Business Intelligence (BI), for example, Microsoft Power BI, offer a great tool for converting raw data into informative dashboards [1],[2],[9]. In turn, a visual representation of data through charts, tables, and KPIs allows for better understanding of datasets [8]-[9]. Nevertheless, in most cases, the existing systems perform only descriptive analytics and lack any predictive mechanisms [3]-[4].

The proposed system tackles these drawbacks through combining Power BI with machine learning algorithms. In particular, incorporating BI with machine learning makes it possible to implement both descriptive and predictive analytics, allowing the institution to explore data trends and predict future trends [3],[4],[5]. Moreover, implementing several dashboards in the project contributes to the increased overall analytical abilities [8],[9].

The objectives of the research are: • Development of a single integrated platform for collecting all the necessary student data; • Providing dashboards enabling visual analysis of complex data; • Using machine learning algorithms for making accurate predictions based on data; • Making decisions according to the data obtained.

II. LITERATURE REVIEW

Existing systems mainly focus on descriptive analyses. Power BI dashboards provide visualization but lack predictive model capabilities [1],[2],[9]. Studies show improved decision-making but highlight limitations such as lack of machine learning integration [3],[4].

The use of Business Intelligence (BI) tools and data analytics in educational systems has gained significant attention in recent years. Various researchers have explored the application of Power BI dashboards and analytical models to improve student performance analysis, dropout prediction, and institutional decision-making [7],[9],[10].

F. M. Shodiq suggested a student dropout dashboard that uses Power BI, data cleaning methods, and visual charts to find and analyze patterns. The system makes it easy to see it and can spot trends but can't make predictions.

Similarly, P. Jayalath et al. developed a student success dashboard based on descriptive analytics using Power BI.
Their system analyzes academic and behavioral factors influencing student success; however, it does not incorporate any machine learning model for predict processing Modeling and H. Joshi visualization Business Intelligence dashboard in education using KPI-based analysis. Their approach improves institutional decision-making, but the implementation involves a complex Modelling, Visualization efficient for large-scale adoption.

J. Sluijter and M. Otten developed a personalized student dashboard using Power BI with KPI-tracking. This system pro-vides customized insights for students but requires significant time and effort for implementation.

Furthermore, V. Anvhane implemented a dropout analysis system using Python integrated with Power BI. This approach enables a deeper analysis of student data; however, it remains limited to descriptive analysis readyes not exPreprocessing modeling.
The above indicates that most existing systems focus primarily on descriptive analytics and lack the integration of machine-learning techniques for predictive analysis. Addition-ally, there is limited research on combining multiple domain inconsistencies such as demographic performance, placement data, scholarship distribution, and institutional, into a unified analytics platform.
To address these limitations, the proposed system integrates Power BI dashboards with machine learning models to provide descriptive and predictive analytics. This approach enables visualization to forecast trends such as placement outcomes and scholarship distribution, thereby enhancing decision-making capabilities.

- Research Gap: Based on the above analysis, it is evident that most existing systems lack integration of machine learning for predictive analysis. Additionally, there is limited focus on multi-domain analytics combining academic performance, placement data, scholarship distribution, and institutional impact.
- Proposed Solution: To address these limitations, the pro-posed system integrates Power BI dashboards witNor malisationrning models. This standardized descriptive and predictive 3 analytics, allowing institutions to forecast trends such as placement outcomes and scholarship standards thereby enhancing decision-making capabilities.

III. PROPOSED SYSTEM

The proposed system is a comprehensive analytics platform designed to analyze student data across multiple domains including academics, placements, scholarships, and institutional impact. The system integrates data from multiple sources and provides insights using interactive dashboards. Unlike traditional categorization, the proposed system incorporates machine learning techniques to predict future trends. The key features of the system include:

- Centralized data integration from multiple sources
- Interactive dashboards with filtering capabilities

- Multiple analysis (academic, placement, scholarship, school)
- Machine learning-based prediction
- Real-time visualization and reporting the novelty of this system lies in the integration of descriptive and predictiveve analytics within a single platform

IV. SYSTEM ARCHITECTURE

The proposed system follows a multi-layered architecture designed to efficiently process, analyze, and visualize student data while incorporating predictive capabilities through ma-chine learning. The architecture ensures a structured flow of data from raw inputs into actionable insights.

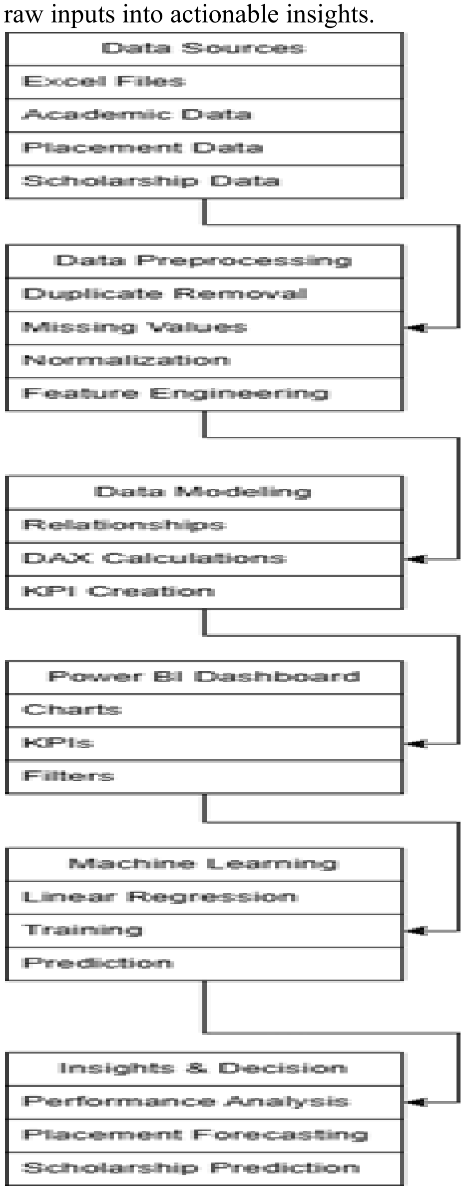

Fig. 1. System Architecture

Figure 1: Proposed System Architecture illustrating data flow from data sources, preprocessing, modeling, dashboard visualization, and machine learning to final decision-making as illustrated in Fig. 1, the system is divided into five major layers: Data Source Layer, Data preprocessing Layer, Data Modeling Layer, Visualization Layer, and Machine Learning Layer

A. Data preprocessing

Data processing is a critical phase in the proposed system, as the quality of insights and predictions directly depends on quality of the input data. Raw datasets collected from multiple sources often contain inconsistencies, missing values, duplicate records, and formatting issues. Therefore, a structured preprocessing pipeline was implemented to transform raw data into a clean and analysis-ready format. The preprocessing operations are performed using Power Query in Microsoft Power BI, which provides efficient Extract, Transform, and Load (ETL) capabilities includes:

- Data cleaning: Data cleaning is the first step in preprocessing and involves removing noise and inconsistences from the dataset.
- Duplicate Removal: Duplicate records were identified using unique attributes such as enrollment number. These duplicates may have arisen due to repeated data entry or merging of datasets. Removing duplicate entries ensured that each student record was unique and prevented biased analysis. Example: if a student record appeared multiple times with identical details, only one instance was retained.
- Handling missing values:

Missing values are common in real-world datasets and can affect analysis and model performance. The following strategies are used:

- Removing records with excessive missing values
- Filling missing numerical values using mean or median
- Filling categorical values using the most frequent value
- Normalization: Data values are standardized to maintain uniformity. For example, variations such as “CSE”, “cse”, and “Computer Engineering” are converted into a single standardized format.
- Data Type Conversion: Attributes such as CPI, year, and numerical fields are converted into

appropriate data types to enable accurate calculations and analysis
- Feature engineering: Feature engineering involves creating new attributes from existing data to enhance analytical capabilities.

CPI-Based Categorization: Students are classified into performance categories:
- High Performance (CPI ¿ 8)
- Medium Performance (6 ¡= CPI ¡= 8)
- Low Performance (CPI ¡= 6)
- Derived Metrics: Additional features such as placement status indicators and scholar ship eligibility flags are created to simplify analysis.

B. Machine Learning Model
Machine Learning is integrated into the proposed system to enhance its analytical capabilities by enabling prediction of future trends from on historical data. Unlike traditional Business Intelligence systems that rely only on descriptive analytics, the inclusion of machine learning allows the system to provide predictive insights for improved decision-making.

C. Model Selection
Linear Regression is selected as the primary machine learn-ing model due to its simplicity, interpretability, and effective-ness for predicting numerical trends. It is widely used for forecasting continuous variables such as placement rates and Scholarship distribution.
The general form of the Linear Regression model is given by:
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdot + \beta_n x_n + \epsilon$ (1)
where y is the predicted output, x_i are the input features,
β_i are the coefficients, and ϵ is the error term [?].

D. Input Features
The model uses multiple features derived from the dataset to improve prediction accuracy. These include:
- CPI (Cumulative Performance Index)
- Branch (categorical feature)
- Year (time-based feature)
- Placement status

These features are selected based on their strong correlation with placement outcomes and scholarship eligibility.

E. Prediction Output
After applying the model, the system generates predictions for:
- Future placement trends
- Scholarship distribution trends
- Student performance patterns

These predictions are visualized using line charts and scatter plots in Power BI dashboards, enabling easy interpretation by users.

F. Placement Prediction
The model predicts placement outcomes based on academic performance and other features. The analysis indicates variations in placement rates across different years and branches. Predicted trends suggest possible fluctuations in future placement opportunities.

G. Scholarship Prediction
The model also predicts scholarship distribution trends. Results indicate a gradual increase in scholarship allocation over time, highlighting potential improvements in financial support systems.

H. CPI Analysis
A scatter plot analysis is used to study the relationship between CPI and placement probability. It is observed that students with higher CPI tend to have better placement opportunities, indicating a strong positive correlation between academic performance and employability.

I. Advantages of Machine Learning Integration
The integration of machine learning in the system provides several advantages:
- Enables predictive analytics
- Improves decision-making capabilities
- Identifies patterns and trends in data
- Reduces dependency on manual analysis

Thus, the Machine Learning component transforms the system from a descriptive analytics tool into an intelligent predictive platform.

V. METHODOLOGY

A. Data Collection
Data is collected from multiple sources to ensure

comprehensive analysis of student information. The dataset includes:
- Student academic records such as CPI and branch details
- Placement data including company name, placement status, and year
- Scholarship data indicating eligibility and distribution
- School information representing student background and enrollment trends

B. Data Preprocessing

Data preprocessing is a critical step that ensures data quality and consistency. The raw data collected from different sources often contains inconsistencies, missing values, and redundant records. Therefore, preprocessing is performed through the following steps:

1) Duplicate Removal: Duplicate records are identified using unique attributes such as enrollment number and removed to avoid redundancy. This ensures that each student record is unique and prevents biased analyses.

2) Handling Missing Values: Missing values are handled by either removing incomplete records or filling them with appropriate values. Numerical values are replaced using mean or median, while categorical values are filled using the most frequent value.

3) Data Normalization: Data is standardized to maintain consistency across datasets. For example, variations such as “CSE”, “cse”, and “Computer Engineering” are converted into a uniform format.

C. Feature Engineering

Feature engineering is performed to enhance analytical capabilities by creating new attributes from existing data. Students are categorized based on their academic performance using CPI values:
- High Performance (CPI > 8)
- Medium Performance ($6 <$ CPI ≤ 8)
- Low Performance (CPI ≤ 6)

Additional derived features such as placement indicators and scholarship eligibility flags are also created to simplify analysis.

D. Data Integration

All datasets are merged into a unified structure to enable multi-domain analysis. The integration process includes:
- Combining academic, placement, scholarship, and school data
- Using enrollment number as the primary key
- Maintaining relationships between datasets

This integration enables cross-domain insights such as the relationship between academic performance and placement outcomes.

E. Data Modeling

Data modeling is performed to structure the dataset for efficient analysis. Relationships are defined between different tables such as student, placement, and scholarship datasets.

Key Performance Indicators (KPIs) are created using Data Analysis Expressions (DAX), including:
- Average Cumulative Performance Index (CPI)
- Total number of students
- Placement count
- Scholarship distribution

F. Machine Learning Integration

Machine learning techniques are incorporated to provide predictive capability. Linear Regression is fundamental technique used to analyze historical data and forecast future trends such as placement rates and scholarship distribution trends.

The model uses features such as Cumulative Performance Index (CPI), branch, year, and placement status to generate predictions. These predictions help institutions identify pat-terns and make informed decisions.

G. Data Visualization

The final stage involves visualizing the processed data using Microsoft Power BI.Number of attractive dashboards are generated to provide acumen into distinct aspects of student data, include:
- Overview Dashboard
- Educator Performance Dashboard
- Placement Analysis Dashboard
- Scholarship Analysis Dashboard
- School Analysis Dashboard
- Life transformation Dashboard

These dashboards use charts, graphs, and KPIs along with interactive filters to allow users to explore data dynamically and supports effective decision-making.

VI. DASHBOARD IMPLEMENTATION

The Overview of this Dashboard provides a top tier sum-mary of the dataset. It includes key performance indicators (KPIs) such as entire number of students, average Cumulative Performance Index (CPI), number of branches, and total schools

A. Overview Dashboard

The Overview Dashboard provides a high-level summary of the dataset. It includes key performance indicators (KPIs) such as total number of students, average CPI, number of branches, and total schools.

The dashboard includes:

- Bar chart representing admissions per year
- Line chart showing life change trends
- Tree map for evolutionary development
- Key Performance Indicator (KPI) cards for quick insights This dashboard helps users understand overall trends and Institutional performance.

B. Academic Performance Dashboard

This dashboard looks over student academic performance based on Cumulative Performance Index (CPI).

Key features include:

- Average Cumulative Performance Index (CPI) by branch
- CPI trends year over years
- Student-level comprehensive table
- Filters for branch and student selection

The analysis shows variation in academic performance across different branches and time periods.

C. Placement Analysis Dashboard

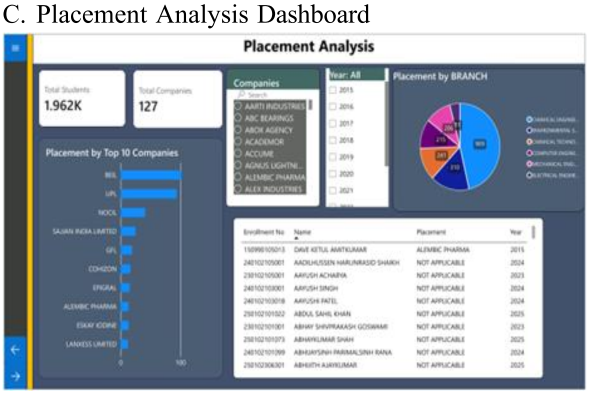

Fig. 2. Placement Analysis Dashboard showing company-wise placements, branch-wise distribution, and student placement records

As shown in Fig. 2, the Placement Analysis Dashboard provides insights into top recruiting companies, branch-wise placement distribution, and detailed student records.

The Placement Dashboard provides understanding into placement statistics and company-wise recruitment.

Key visualizations include:

- Pie chart shows the placement distribution among different branches
- Bar chart used for top enrolling companies
- Placement table

The dashboard highlights branch with higher placement rates and identified major recruiters.

D. Scholarship Analysis Dashboard

This dashboard analyzes scholarship distribution among students.

Key features:

- Scholarship trends over years
- Branch-wise scholarship allocation
- Student-level scholarship details

The analysis indicates that a limited number of students re-ceive scholarships, suggesting the need for improved financial support [8].

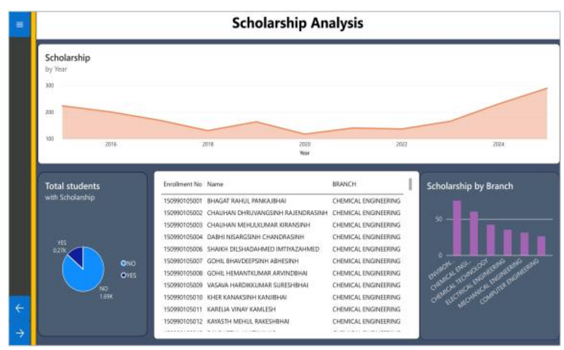

Fig. 3. Scholarship Analysis Dashboard showing yearly trends, branch-wise distribution, and student scholarship details

E. School Analysis Dashboard

The School Analysis Dashboard evaluates student performance based on your school framework.

Key visualizations:

- Schools per year
- Elite schools contributing to scholarships
- School-wise production distribution

This dashboard helps identify schools that consistently produce high-performing students [8].

F. Life Change Dashboard

The Life Change Dashboard assess the total effect of education on students by combining academic, placement, and scholarship data.

Key insights:

- Majority of students experience favorable personal development
- Increase years wise growth
- Branches wise variations

This dashboard highlights the real-world impact of institutional efforts and demonstrates the effectiveness of data visualization in understanding complex outcomes [8].

G. Key Features of Dashboard System

The dashboard system includes several advanced features:

- Attractive filters (slicers) for dynamic data exploration
- Real-time visualization of Key Performance Indicator
- Multiple-domain analysis beyond different datasets
- User-friendly interface for easy explanation

This dashboard implementation changes raw data into meaningful insights, allow efficient and data-driven decision-making [1],[8].

VII. MACHINE LEARNING IMPLEMENTATION

A. Model Training

The model is trained using historical student data to learn relationships between input features and target variables. The dataset is divided into training and testing subsets to evaluate model performance [10].

The training process involves fitting the Linear Regression model by minimizing the error between actual and predicted values using the least- squares method. The objective is to determine optimal coefficients that best represent the relationship between independent variables and the target variable [5].

$$\min \sum_{i=1} (y_i - \hat{y}_i)^2 (2)^n$$

where y_i represents the actual value and $\hat{y}_i$ represents the predicted value.

The trained model captures patterns in the data, enabling it to generalize and make accurate predictions on unseen data.

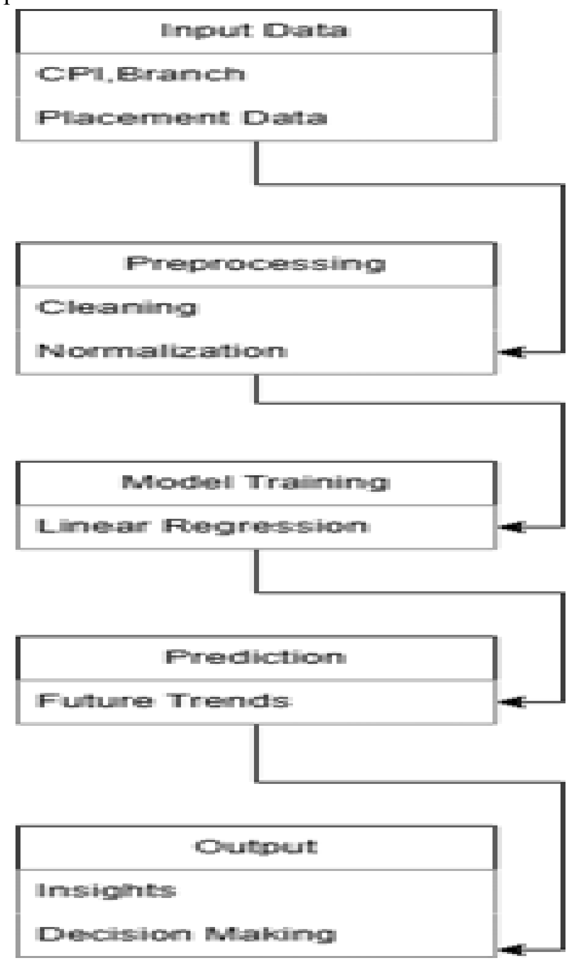

Fig. 4. Machine Learning Pipeline for Placement and Scholarship Prediction

VIII RESULTS AND DISCUSSION

The proposed system successfully establishes a robust data visualization and machine learning, generating a meaningful insight from student data. The results are obtained through interactive dashboards and predictive analysis [1],[10].

A. Key Findings

The analysis of this dashboard shows several important observations:

- Academic performance varies significantly across different branches, indicating differences in student outcomes.
- Campus recruitment and job placement trend show fluctuation with certain branches achieving higher placement rates.
- The restricted availability of scholarship highlights

- a critical need to enhance existing financial support frame-works.
- School background has a significantly impact on student performance and placement outcomes.

B. Impact of Machine Learning

The integration of machine learning enhances the system by enabling predictive analytics. The model provides insights such as:

- Forecasting future placement trends based on historical data
- Prediction of scholarship distribution patterns
- Identify the students with higher probability of placement These predictive capabilities help institutions make proactive decisions and improve planning strategies [3],[10].

C. Discussion

Our results classify that combining Power BI with machine learning provides a more comprehensive analytics solution compared with traditional systems [1],[10].

The variation in academic and placement trends sug-gests that multiple factors such as branch, performance, and background-impact student outcomes. The low scholarship distribution suggests a potential gap in sponsorship, which institutions need to address.

Additionally, the predictive insights generated by the model enable institutions to anticipate future trends and take necessary actions in advance. This reduces dependency on manual analysis, and improve overall productivity [3].

D. Comparison with Existing Systems

Compared to traditional BI dashboards, the proposed system offers:

- Both descriptive and predictive analytics
- Improved decision-making capacity
- Multi-domain data analysis in a single platform

Thus, the system provides a significant improvement over existing approaches by integrating machine learning with data visualization.

IX. ADVANTAGES AND LIMITATIONS

A. Advantages

The proposed system offers several advantages:

- Interactive Dashboards: The use of Power BI enables dynamic and responsive visualization of data, that al-lowed users to explore perceptions efficiently.
- Real-Time Data Analysis: This system provides real-time analysis through self-operating data processing and visualization.
- Predictive Analytics: Combination of machine learning allows forecasting of placement trends and scholarship distribution, improved decision-making skills [?].
- Centralized Data Management: Collect data from multiple sources is unified into a single platform, enabled comprehensive analysis.
- Multi-Domain Analysis: This system analyzes scholas-tics achievement, placement data, scholarship distribution, and school background simultaneously.
- Improved Decision-Making: Data-driven perception re-duce manual effort and help institutions make informed strategic decisions [?].

B. Limitations

Regardless of its advantages, the system has certain limitations:

- Dependency on Data Quality: The accuracy of results depends on the quality and completeness of input data. Poor data can lead to unstable predictions.
- Use of Basic Machine Learning Model: This system can currently uses Linear Regression, which not capture complex patterns compared to advanced models.
- Lack of Real-Time Data Integration: This system primarily works on stable datasets and does not support uninterrupted real-time data streaming.
- Scalability Constraints: Handling very large datasets may require additional optimization and infrastructure support.
- Limited Automation: Some preprocessing steps will require instruction intervention, which can affect proficiency.

X. FUTURE SCOPE

The proposed system can be further enhanced by incorporating advanced technologies and extended functionalities to improve its performance, scalability, and usability.

- Advanced Machine Learning Models: In future work we include the implementation of advanced algorithms such as Random Forest, Support Vector Machines, and Neural Networks to improve prophecy accuracy and trap complex patterns in data [4] .

- Real-Time Data Integration: Our system can be refined to support real-time data streaming by using cloud-based technologies, enabling uninterrupted observation and instant insights.
- Cloud Deployment: Deployment of this system on cloud platforms can improve scalability, reachability, and data security, allowance institutions to access dashboards from anywhere.
- Mobile Application Development: In mobile-based interface we provide users easy access of dashboards and analytics on smartphones.
- Integration with Government Systems: This system can be integrated with government scholarship portals to automate eligibility verification and improve clarity in scholarship dispersion.
- Automated Decision Support Systems: In future we may include AI-based recommendation systems to indicate actions such as identifying at-risk students or prescribe scholarships [3].

This development can transform the system into a fully intelligent and flexible analytics platform, capable of supporting data-driven judgment at a higher institutional level.

XI. CONCLUSION

This project is an extensive analytics solution that combines Business Intelligence software and machine learning technologies to analyze students' data across different areas. Our solution effectively converts raw data into valuable information via interactive dashboards and predictive models, thus ensuring an intuitive and comprehensible analytical process [1], [10].

The application of multiple dashboards facilitates the evaluation of academic achievements, placement patterns, scholarship allocation, and institutional influence. The visual representations of information contribute to the comprehension of complicated datasets by their users, helping them make informed decisions [8].

The adoption of machine learning algorithms adds value to our solution by predicting future trends regarding placement rates and scholarship allocation. This removes the need for manual analysis and allows organizations to make decisions ahead of time using data [3],[4].

In general, our proposed solution brings numerous ben-efits to the field of educational data analytics, providing a consistent platform for cross-domain analysis and illustrating the applicability of Business Intelligence tools combined with machine learning techniques for decision support [6], [9]. In the future, we might improve the system by using smarter machine learning methods, analyzing data as it comes in, and working with cloud services.

REFERENCES

[1] Microsoft, "Introduction to Power BI," Available: https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview

[2] M. Ferrari and A. Russo, Introducing Microsoft Power BI. Microsoft Press, 2016.

[3] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006. [Online]. Available: https://link.springer.com/book/10.1007/978-0-387-45528-0

[4] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer, 2009. [Online]. Available: https://hastie.su.domains/ElemStatLearn/

[5] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis. Wiley, 2012.

[6] R. Kimball and M. Ross, The Data Warehouse Toolkit. Wiley, 2013.

[7] G. Siemens and R. S. J. d. Baker, "Learning analytics and educational data mining," Proc. 2nd Int. Conf. Learning Analytics, 2012. Available: https://dl.acm.org/doi/10.1145/2330601.2330661

[8] S. Few, Information Dashboard Design. O'Reilly Media, 2013.

[9] A. Kumar et al., "Data Visualization using Power BI for Business Intelligence," International Journal of Scientific Research, 2021.

[10] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2011.