Enhancing Business Sustainability Through Technology-Enabled AI: Forecasting Student Data and Comparing Prediction Models for Higher Education Institutions (HEIs)

Hao Qian Gnoh[1], Kay Hooi Keoy[2], Javid Iqbal[1], Shaik Shabana Anjum[3], Sook Fern Yeo[4,5]*, Ai-Fen Lim[2], WeiLee Lim[2], Lee Yen Chaw[2]

[1]Institute of Computer Science and Digital Innovation, UCSI University, 56000 Cheras, Kuala Lumpur, Malaysia
[2]UCSI Graduate Business School, UCSI University, 56000 Cheras, Kuala Lumpur, Malaysia
[3]School of Computer Science, Faculty of Innovation and Technology, Taylor's University, 47500 Subang Jaya, Selangor, Malaysia
[4]Faculty of Business, Multimedia University, 75450 Bukit Beruang, Melaka, Malaysia
[5]Department of Business Administration, Daffodil International University, Dhaka, Bangladesh

*Corresponding author email: yeo.sook.fern@mmu.edu.my

ABSTRACT

This study aims to enhance business sustainability in the context of Higher Education Institutions (HEIs) by utilizing AI and forecasting techniques. It explores the development and comparison of prediction models, including the use of dashboard development, to support decision-making processes within HEIs. The study covers various aspects, including the background of forecasting and prediction models, the use of specific models such as the Prophet Model, Long Short-Term Memory (LSTM) Model, and Polynomial Regression Model, as well as the importance of dashboards for HEIs. The methodology section outlines the data collection and preparation process, model selection, approach, diagrams, functional and non-functional requirements, justification of tools, and libraries and models used. The implementation section delves into the system design and development of the dashboard, including the login page, homepage, forecast page, and insert data page. As for the findings, the LSTM Model has proven to be the most accurate and suitable model to be implemented for forecasting student enrolment data in this study. The dashboard's future enhancements involve adding more faculties, predictive features for resource allocation, refining the visual identity, improving user registration on the login page, and exploring better models for student enrolment predictions. Overall, the study provides valuable insights into the application of AI and forecasting techniques in HEIs, aiming to enhance business sustainability and decision-making processes. It contributes to the growing body of knowledge on the use of technology-enabled AI in higher education institutions, with a focus on forecasting student enrolment data and developing prediction models.

KEYWORDS: Business analytics, Decision making, Business sustainability, Forecasting, Quality education

Received January 12, 2024; Revised February 18, 2024; Accepted March 8, 2024
Doi: https://doi.org/10.59953/paperasia.v40i2b.86

1. INTRODUCTION

In the ever-evolving landscape of higher education, the ability to forecast and analyse student enrolment data using prediction models stands as a pivotal function for Higher Education Institutions (HEIs). The accurate prediction of student enrolment trends, coupled with a comprehensive analysis of historical and real-time data, empowers HEIs to make informed decisions, allocate resources efficiently, and provide quality education to their diverse student populations (Pepin et al., 2022). This serves as a comprehensive introduction to the concept of forecasting and comparing prediction models for student enrolment data within the context of HEIs, with a specific emphasis on the development of a dashboard as a tool for this purpose. It outlines the fundamental background, justification, and objectives of this research, providing a full overview of the subsequent chapters that delve into the intricacies of this research domain. Some of the specific challenges or issues in higher education that make forecasting student enrolment crucial are increasing competition among HEIs, changing student preferences and expectations, uncertain economic and social factors, and limited resources and budget constraints. These challenges require HEIs to adopt proactive and strategic approaches to plan and manage their enrolment activities, as well as to optimize their academic and administrative operations. The importance of the developed dashboard in addressing these challenges is that it provides a user-friendly and interactive platform for HEIs to access, visualize, and analyse student enrolment data using various prediction models. The dashboard enables HEIs to compare the performance and accuracy of different models, such as Prophet, LSTM, and Polynomial Regression, and to select the most suitable one for their specific needs and scenarios. The dashboard also facilitates the communication and presentation of the forecasting results and insights to the relevant stakeholders, such as top management, faculty, and staff, for effective decision-making and policymaking.

1.1 Background

1.1.1 Forecasting and Prediction Models

Prediction models are widely used in the modern era with all sorts of models in place and many industries are using these models to further upgrade their business or systems. Predictive modeling is a statistical technique that uses historical and current data to create, process, and validate a model that can be used to forecast future outcomes. Predictive modelling involves various applications such as assessing the value of a sales opportunity, predicting the chances of receiving spam, or determining the probability of someone clicking a link or making a purchase (Lawton et al., 2022). Forecasting is a valuable technique that uses past data to make dependable predictions about what might happen in the future. Businesses use it to figure out how much money to set aside and to plan for future costs. Forecasting is crucial for making smart business choices because it helps management see trends and shifts in important business factors like expected sales or customer behaviour, no matter how big or small the organization is (Tuovila, 2023). Given this, forecasting holds significance as it proves beneficial in various scenarios, particularly in corporate management, by enabling accurate predictions of future trends and events. Its importance lies in assessing the potential success of new business ventures, identifying financial needs, ensuring consistent operations throughout the organization, aiding managers in making sound decisions, and developing practical plans for the future.

Forecasting student enrolment entails analysing information about enrolment, demographics, geography, and land usage to predict how many students are likely to enroll in higher education institutions in the future (Vick, 2020). Forecasting is a crucial starting point for planning in any organization, especially in higher education. Among various forecasts, predicting student enrolment is particularly important as it affects institutional income, faculty needs, space requirements, budgets, and other aspects. A study has explored the most common of four of the six basic forecasting techniques—regression, time series, data mining (along with its sub-methods), and analysing historical data & trends—which are most commonly used in higher education for analysing student enrolment (Lavilles, 2012). Moreover, forecasting is commonly used in higher education for enrolment, marketing, teaching, and performance purposes. Over the last two decades, two relatively new forecasting methods, data mining and surveys, have been employed in higher education (Sinuany-Stern, 2021). There are three models that will be compared in this research. The models are Prophet Model, LSTM Model and Polynomial Regression Models. The rationale for selecting these particular models is to compare different approaches to forecasting: a statistical model (Prophet), a neural network model (LSTM), and a classical machine learning model (Polynomial Regression). By comparing their performance on the same dataset, we can evaluate their strengths and weaknesses and choose the best model for our problem.

1.1.2 Prophet Model

Prophet is a time-series forecasting tool available in both R and Python created by Facebook. It's an open-source algorithm designed to build time-series models, combining traditional concepts with innovative approaches (Krieger, 2021). The Prophet model is a significant tool in time-series forecasting due to its unique approach and capabilities. Prophet is known for its speed and ability to generate automated forecasts that can be fine-tuned manually based on the data. Prophet is based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. This allows it to capture complex patterns and changes in trends over time, which is often challenging with traditional time series models (Goyal, 2020). One of the key strengths of the Prophet model is its ease of use. It's designed to be user-friendly and completely automatic. You can point it at a time series and get a forecast. It also handles trend and seasonality features natively, making it a good baseline model if the time series follows business cycles. The model's decisions are easy to interpret thanks to its native breakdown of the forecast into time components. Furthermore, Prophet can model complex seasonality and general trends in time series data (DiBattista, 2022).

1.1.3 Long Short-Term Memory (LSTM)

Model Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) explicitly designed for processing sequential data like time series, speech, and text (Brownlee, 2017). LSTM was created by (Hochreiter & Schmidhuber, 1997) to address the challenge of RNNs not effectively handling long-term dependencies. In traditional RNNs, there's a limitation where they struggle to predict information stored in long-term memory but perform better with recent data. LSTM networks, on the other hand, excel at capturing long-term dependencies within sequential data, a feature that sets them apart from traditional RNNs. This unique capability allows them to remember information over extended periods, making them particularly well-suited for tasks that require the understanding of complex, temporal relationships. This attribute makes them particularly well-suited for tasks such as language translation, where understanding the context from the start of the sentence is crucial for accurate translation. In speech recognition, LSTM's ability to remember 'earlier heard sounds' helps in better prediction of the current sound. Furthermore, in time series forecasting, LSTM's proficiency in learning the pattern over a long duration makes it a preferred choice (Jiang, 2021).

1.1.4 Polynomial Regression Model

Polynomial regression is a form of regression analysis employed to establish the connection between a dependent variable and one or more independent variables. Its unique capability lies in its ability to capture non-linear relationships among variables, a feature that sets it apart from other regression techniques (Gavrilova, 2021). This method is a potent tool for predictive analytics, finding applications in various fields. For instance, in the field of epidemiology, polynomial regression can be used to forecast disease spread rates, providing valuable insights that can guide public health interventions. In the realm of human resources, it can help in determining fair compensation by considering multiple factors such as experience, education, and job role. Furthermore, it is instrumental in implementing advanced preventative road safety software by analyzing and predicting accident hotspots based on historical data (Kumar, 2023).

1.1.5 Dashboard For HEI

A dashboard is a central and interactive tool that enables businesses to keep track of, analyse, and gain useful insights from different sets of data in different important areas, such as student performance, enrolment, retention, graduation, and satisfaction. It displays data in a visually engaging and easy-to-understand way (Calzon, 2022). Dashboards offer a convenient way to extract insights from raw data without the need for coding. They come with several benefits, including giving a snapshot of crucial data indicators and allowing users to view all important Key Performance Indicators (KPIs) in a single place. Acting as a central hub, dashboards enable users to access and evaluate the latest information easily (Malnik, 2022). In the context of this study, the benefits that dashboard brings are being able to monitor the trends and patterns of student enrolment across different programs, campuses, and demographics. They can also compare their enrolment data with their targets, budgets, and benchmarks (Keoy & Kwek, 2012). This can help them identify the strengths and weaknesses of their recruitment strategies and make informed decisions to improve them. Moreover, dashboards can communicate the enrolment data and outcomes to various stakeholders, such as faculty, staff, students, alumni, donors, and policymakers. This can enhance the transparency, accountability, and reputation of the HEIs. The dashboard also includes filtering options to narrow down the displayed data by year and campus (Nils R, 2021). These options enhance the usability of the dashboard by allowing users to customize their view and focus on the data that is most relevant to their needs and goals. An example of this type of dashboard is provided in **Figure 1**. In order to better understand the popularity of the dashboard among Higher Education Institutions (HEIs) and the common prediction models used for student enrolment data prediction, the literature review will be conducted to understand existing dashboards used in HEIs and how the existing dashboards work to help the management of an HEI as well as finding out the common prediction models used in student enrolment data prediction. The prediction models used for student enrolment data prediction are many, however, in this paper, I will be focusing on the Prophet Model, LSTM Model, and Polynomial Regression Model to find out the best model suited for a dashboard purpose.

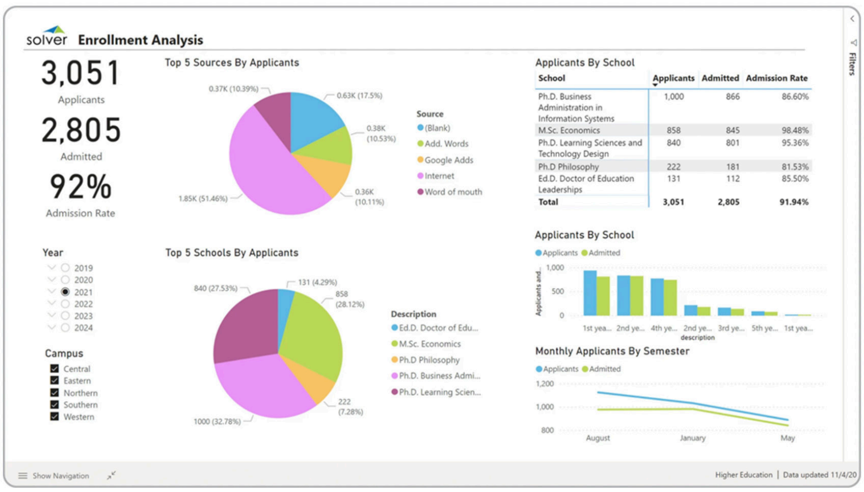

Figure 1: *Example of a Dashboard*

2.1 Popularity of Dashboard Among Higher Education Institutions

Dashboard development for student enrolment data is an important area of research in higher education. While there are several studies on the use of dashboards in higher education, there is limited research on the popularity of dashboards among higher education institutions (Kwek et al., 2022). However, some studies have explored the factors influencing the adoption of technology in higher education institutions, including dashboards. A researcher proposed the construction of a prototype of a dashboard for the BI system to be adopted as a practice of governance by results in a Brazilian public institution (Azevedo et al., 2021). The study aimed to ensure agility in accessing and processing information, support the decision-making of managers, allow efficiency and public transparency, and minimize the difficulties in access, in the management of various BI systems and the informational asymmetries existing in the Public Administration. Another study conducted in Egypt found that higher education institutions in developing countries are challenged with the high enrolment student rates, crowded classes, and inability to track the progress of each student individually, which increased the demand to find a solution that can redeem those problems (McEvoy et al., 2010; Sharma & Joshi, 2022).

2.2 Prediction Models Used for Student Enrolment Data

2.2.1 Data Mining Approach

Data mining technique is one of the techniques used in past studies to predict student enrolment data. There is a study that focuses on predicting student enrolment data using data mining techniques. It introduces a framework aimed at forecasting the university's incoming student numbers based on historical data and diverse factors like academic performance, socio-economic background, and admission criteria. To achieve this, the article employs various data mining methods, including decision trees, neural networks, and support vector machines, using data collected from a university in Malaysia. The accuracy and performance of these prediction models are thoroughly assessed and compared. In its conclusion, the article underscores the utility of data mining for enrolment management and planning, ultimately identifying the neural network as the most effective method for predicting student enrolment (Haris et al., 2016).

2.2.2 Time Series Analysis Approach

The next common approach is time series analysis, which is also the most common approach used by people. A study tackles the forecasting of enrolment for a higher education institution (HEI) located in the Philippines. The chosen method for this task is the Autoregressive Integrated Moving Average (ARIMA)(p,d,q) model, a type of time series analysis. To make these predictions, the article relies on historical enrolment data spanning from 2011 to 2019, sourced from the Office of the Registrar at Cebu Technological University-Barili Campus. Several techniques, including the autocorrelation function (ACF), partial autocorrelation function (PACF), and Akaike information criterion (AIC), are employed to determine the most suitable ARIMA(p,d,q) model for the dataset. Ultimately, the article determines that the ARIMA (0,2,1) model offers the most accurate predictions for enrolment over the next six years, from 2020 to 2025, projecting a consistent increase in enrolment during this period. The article also delves into the potential benefits and implications of these forecasts for HEI management and planning. Additionally, it suggests potential avenues for future research in this field (Dela Cruz et al., 2020).

2.2.3 Machine Learning Approach

The third common approach is machine learning. The following article summary focuses on the application of machine learning techniques to forecast student enrolment and predict attrition, aiming to assist higher education institutions in managing these aspects effectively. Real data from the Abu Dhabi School of Management (ADSM) in the UAE is utilized as the primary data source. The authors employ association rule mining, boosting, and ensemble methods to identify student characteristics influencing enrolment and attrition trends. Prediction chart of enrolment rate under ARIMA (2,1,2) model is depicted in **Figure 2**. The outcomes reveal that their models achieve high accuracy, offering valuable insights for ADSM. Additionally, they propose that these methods can be applied more broadly to benefit other higher education institutions, aiding in enhancing their overall performance and sustainability. The boosted regression tree model for predicting student enrolments achieved 89% accuracy using 10-fold cross-validation and outperformed the single regression tree model that achieved only 76% accuracy (Shilbayeh & Abonamah, 2021).

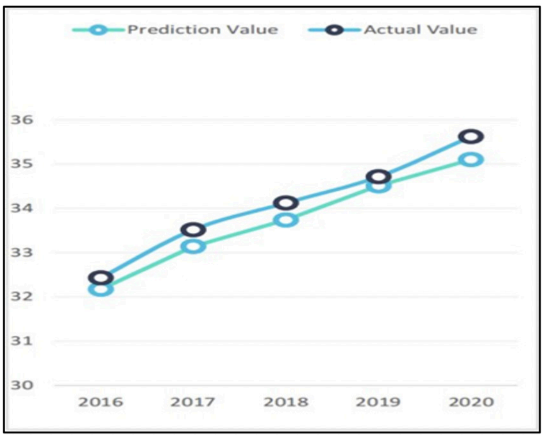

Figure 2: *Prediction chart of enrolment rate under ARIMA (2,1,2) model*

2.2.4 Deep Learning Approach

There are no studies of people using deep learning techniques to predict student enrolment data. However, other studies use deep learning techniques to focus on student learning outcomes and student performance predictions. The following paper finds that a combined CNN-LSTM (convolutional neural networks and long short-term memory) approach is the most effective for predicting student outcomes. The study focuses on the use of Learning Management Systems (LMSs) in higher education, with a particular emphasis on Blackboard, a widely used LMS. The goal is to determine how certain Key Performance Indicators (KPIs) related to student interactions with Blackboard can predict student learning outcomes. The study employs a mixed-methods approach and analyses data from seven general preparation courses. It identifies predictive KPIs from Blackboard reports and uses correlation analyses to assess their relationship with student performance (Aljaloud et al., 2022). **Table 1** depicts the summary of the literature review.

3. METHODOLOGY

3.1 Data Collection and Preparation

The mock student enrolment data is collected from the University of Cambridge Admission Office (University of Cambridge, 2023). Data has been taken from the 'Acceptances' column in the diagram below. The data taken is from 2013 to 2022, which is 10 years. However, the data provided is only based on yearly data. So, adjustments were made to accurately split the data into three intakes per year. Several factors were taken into consideration to accurately depict the specific intake in a year. Such as the popularity of the courses based on past trends of actual intakes in other universities in Malaysia. Other than that, factors like seasonal enrolment were also considered. For example, the January intake has the least number of enrolments because high school graduates in Malaysia who want to continue their studies in a university still have not gotten their results to enroll. As for the May and September intake, there has been an increase in enrolment as it is considered the prime time to enroll in a university as everyone has gotten their results. Below is the dataset that has been curated for the most accurate version of actual enrolment data in a university in Malaysia based on the enrolment data of the University of Cambridge. The dataset utilized in this study is a compilation of mock student enrolment data, which was obtained from the University of Cambridge Admission Office. The data is in numerical format, displaying the total number of students enrolled in each year for each faculty. The decision to use this source is justified by the comprehensive and detailed nature of the enrolment information provided by such a prestigious institution, ensuring the reliability and high standard of our mock data. The original data, which spanned from 2013 to 2022, was based on yearly figures. However, to accurately mirror the enrolment pattern in Malaysian universities, which have three intakes per year, adjustments were made to the data. These adjustments took into account several factors. Course popularity trends were analyzed based on historical intake data from other universities in Malaysia, aiding in the estimation of the distribution of students across different courses for each intake. The timing of high school graduation results in Malaysia also affects the enrolment pattern. For instance, the January intake typically has fewer enrolments as students are yet to receive their results. Conversely, the May and September intakes see an increase in enrolments as most students have received their results by this time. Therefore, the curated dataset provides a more accurate representation of actual enrolment trends in Malaysian universities, while still being based on reliable data from the University of Cambridge.

3.2 Model Selection

The reason why these three models, the Prophet Model, LSTM Model, and Polynomial Regression are selected as the prediction models in this development is justified. As for all three models, research and development be not done by previous authors including creating a prediction model using the above three models for HEI student enrolment data.

Table 1: *Summary of the articles*

Country	University	Prediction Model	Results/Accuracy	References
Malaysia	Not Mentioned	Decision Trees, Neural Networks, and SVM	No Results	(Haris et al., 2016)
Philippines	Cebu Technological University-Barili	ARIMA	Not Mentioned	(Dela Cruz et al., 2020)
Malaysia	Not Mentioned	Logistic Regression (LR), Decision Tree and Naïve Bayes	LR: 70.75% Decision tree: 71.36% Naïve Bayes: 70.04%	(Ghani et al., 2019)
Cambodia	Not Mentioned	ARIMA	MSE: 13% RMSE: 37 MAPE: 34%	(Meng et al., 2022)

Therefore, efforts have been made to research and implement these models in this development to test and come out with a statement on whether these models are suitable for predicting student enrolment data. Another reason why these three models were chosen is because these models are most suitable for time series analysis, in which the student enrolment data are based on time series data.

As for the Prophet model, it is one of the well-known time series models that has been used in the prediction of other data. It is a relatively new prediction model therefore new technology and faster processing speed are present in the model. Prophet can handle time series data with multiple seasonal components. It allows you to specify both yearly and weekly seasonality, making it versatile for capturing different cyclic patterns in the data. This makes it perfect for the student enrolment dataset as there is monthly seasonality. It also stands out for its automatic handling of seasonality and holidays, robustness in the face of missing data and outliers, straightforward parameter tuning, provision of prediction intervals and uncertainty estimates, and scalability for large datasets (DiBattista, 2022). As for the LSTM Model, they can model complex patterns, both short-term and long-term, in the data. This makes them effective for time series that exhibit nonlinear relationships, irregular trends, or non-obvious patterns. LSTM Models also can naturally handle irregularly spaced time series data, which is common in real-world scenarios such as the student enrolment dataset in this development. They do not require equally spaced time intervals, and this flexibility is valuable when dealing with data that is not recorded at regular intervals. LSTM are also able to incorporate multiple features, adapt to varying sequence lengths, and automatically extract relevant features (Anishnama, 2023).

As for Polynomial Regression, although it is typically used in a non-time series context for modelling, it has been selected to observe how well it deals with time series data, and whether is possible or not possible to be implemented with the student enrolment data. Despite that, one of the primary advantages of polynomial regression is its ability to capture nonlinear relationships between variables. In time series data, trends and patterns may not always follow a linear path, therefore polynomial regression can be used to model these nonlinear trends more effectively than linear models (Pant, 2019). As for why other time series models were not implemented in this study has been justified. Time series models like ARIMA which has been implemented in other studies, the model requires a stationary time series with many observations. If the student enrolment data is limited, the ARIMA model may not be suitable (Slim et al., 2018). Which in this case is true for the dataset of this development. Other than that, Vector Autoregression (VAR) models are much more complex and difficult to implement than others. The model also requires multiple time series variables and can be challenging to estimate. Without the proper expertise to implement such a complex model, it may not be a feasible option (Chen, Li & Hagedorn, 2019). Not only that, but the Seasonal Hybrid also ESD (S-H-ESD) model requires additional features or variables to make accurate predictions. It requires a set of external variables that can explain the seasonal patterns in the data. Thus, if the student enrolment data does not have enough features or variables, the S-H-ESD model may not be a suitable choice (James & Weese, 2021). Many other time series models have not been mentioned, but those are beyond the scope of this project to cover and review. Therefore, as of this study, the three most suitable models have been chosen based on the justification above.

3.3 Approach

In this paper, we will utilize the Rapid Application Development (RAD) technique as it aligns well with the project's framework, timeline, and time limitations. RAD is an iterative process that enables the rapid and cost-effective development of a high-quality system, making it well-suited for the project's schedule. The RAD method consists of four phases: requirements planning, user design, construction, and cutover. The user design and construction phases are repeated until the user confirms that the product meets all requirements. Following the initial phase, the next step is the user design stage, which marks the start of the system development process. This stage is dedicated to improving the user interface (Sharma, 2015). To guarantee that the client's requirements and expectations are met during the design phase, several presentations will be carried out. These presentations will include prototypes and encourage interactions with the top management of the higher education institution (HEI). The following stage is the construction phase, where the prototypes and beta systems created in the design phase are turned into a functional model. The project team, consisting of programmers, coders, testers, and developers, has a vital role in ensuring the system functions effectively and aligns with the client's expectations and goals (Sharma, 2015). Throughout this phase, the client can offer any additional requests or input as needed. The final stage, known as the cutover phase or execution phase, comes after multiple revisions. The dashboard management system is now ready for deployment. A final report for the project will be generated, and a presentation will be conducted to delve deeper into the report. The dashboard system will be accessible for use by the top management at any higher education institution (HEI).

3.4 Validation

Another study explores the use of data mining techniques to forecast student enrolment in a higher education institution based in Malaysia. The study follows the CRISP-DM process model, encompassing data collection, data preparation, modelling, and evaluation stages. Within this framework, the article applies three data mining methods—logistic regression, decision tree, and naïve Bayes—to the application data of three undergraduate programs: information technology, engineering, and business management. Through a thorough comparison of accuracy and performance, using 10-fold cross-validation, the article determines that the decision tree method outperforms the others across all programs. The article goes on to identify key factors that significantly influence enrolment decisions, including the intended major, parent employment status, and application type. Lastly, it discusses the potential advantages and implications of these prediction models for institutional management and planning. Since there are Information technology, engineering, and business management enrolment prediction accuracy, the study takes the average of all the 3 faculties for the accuracy models. Logistic regression is at 70.75%, Decision tree is at 71.36% and Naïve Bayes is 70.04% (Ghani et al., 2019).

4. IMPLEMENTATION

4.1 Login Page

First, users will be greeted with a login page into the dashboard. Only authorized users will be able to access the dashboard. After entering the username and password, users proceed to click enter on the keyboard to login to the dashboard as depicted in **Figure 3**.

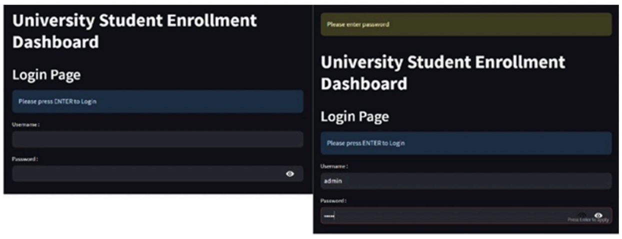

Figure 3: *Login page*

4.2 Homepage

After logging into the dashboard, users are greeted with the Homepage of the dashboard. The page is the page for users to analyse and visualize the overall current student enrolment data in the HEI. This ranges from past data to present-day data. For example, the figure below shows the current total enrolment of students in 3 different faculties from the year 2020 to current as depicted in **Figure 4**.

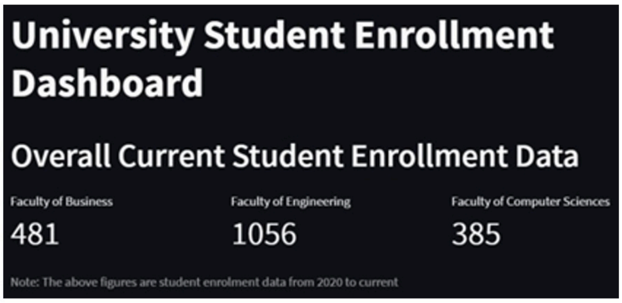

Figure 4: *Overall current student enrolment data*

4.3 Forecast Page

The forecast page is the page for users to forecast future student enrolment data in the HEI. Users are greeted with this page to forecast student enrolment. Users can select from the list of faculties on the left side of the dashboard. Followed by selecting the department in the faculty to start the forecast analysis. Users can view the current enrolment of the department to make comparisons with the forecasted values by clicking on the expandable section. The same goes for the three different models, the user can click on the expandable section for each of the models to view the graphs and values as depicted in **Figure 5**.

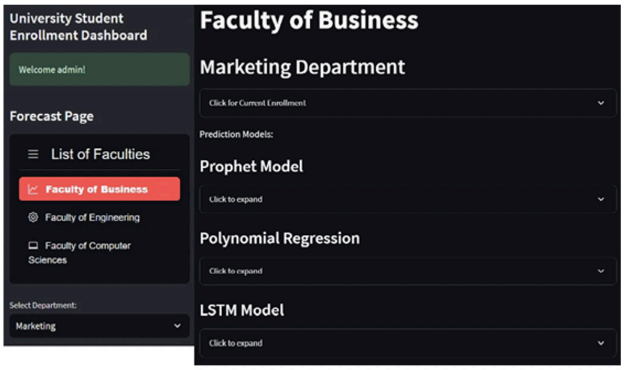

Figure 5: *View of forecast page*

5. RESULTS AND EVALUATION

5.1 Model Accuracy Results

The table below are the RMSE and MAPE results for each of the three models for each of the departments. This is to determine which model is the most accurate in terms of prediction modelling and suitable to be used in the student enrolment dashboard as a forecasting tool. The accuracy of different model is depicted in **Table 2**. Based on the RMSE and MAPE results in **Table 2**, it is shown that the LSTM model consistently outperforms the other models. The model provides more accurate predictions with lower RMSE values and precise percentage error minimization with lower MAPE values. It excels in almost all of the departments, except one department which is Electrical Engineering. The Prophet Model is also an option to consider with relatively accurate results coming in behind the LSTM Model.

Table 2: Model Accuracy Results

Prediction Models	Prophet Model		LSTM Model		Polynomial Regression	
Departments	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
Marketing	6.07	33.62%	4.69	33.11%	9.34	70.47%
Accounting	6.96	35.22%	4.07	39.04%	6.67	64.23%
Business Administration	7.92	52.90%	4.98	35.01%	8.53	65.61%
Chemical Engineering	11.00	50.41%	6.51	32.00%	12.91	65.60%
Mechanical Engineering	18.33	37.71%	15.55	36.98%	23.48	62.99%
Electrical Engineering	14.46	24.76%	10.75	37.06%	17.37	54.64%
Computer Science	6.22	39.79%	4.89	30.68%	7.46	52.18%
Data Science	5.70	26.80%	4.71	23.73%	5.33	38.76%

In fact, the Prophet Model excels in one department which is Electrical Engineering, getting a higher accuracy score result compared to the LSTM Model. In contrast, the Polynomial Regression model consistently ranks the lowest with the highest MAPE values and, in many cases, the highest RMSE values, indicating its limited forecasting effectiveness.

6. FUTURE WORK AND LIMITATIONS

The future work of this research can be to integrate the LSTM model with another artificial neural network (ANN) model so that more accurate prediction can be achieved along with corresponding decision-making and responsive analogy. This integrated model can further be validated with real-time data and help HEI to work on aspects that need improvement to increase student enrolments. Furthermore, business models can also be embedded with LSTM to create a framework and can be hypothetically tested with focus group discussions and structured interviews. The validation can be performed with technology acceptance models (TAM) and other user acceptance and usability testing methods. The combination of LSTM models with other ANN architectures provides a synergistic approach to improving predictive accuracy. Based on the analysis, LSTMs are considered the best approved for capturing temporal dependencies, in addition to the diverse strengths of additional ANN models. The results demonstrate a robust and comprehensive model, able to effectively integrate the capabilities of each component to generate more accurate predictions across various applications.

The limitations in the study can be identified as a lack of more accurate data preparation and preprocessing being done. The data collected was from a real source, however the numbers of students based on each intake were roughly estimated just by randomizing the data and factoring the previous enrolment trends according to intakes in Malaysian universities. The only data given was based on a yearly basis, so the prediction results could be biased or not reflect the accurate version of the current institution's enrolment numbers. Another limitation is the small sample size that is needed to train the model. Due to the relatively small sample size, training the model with a small data size may not be the best scenario. Prediction models usually run and predict values better with a bigger sample size provided. Other than that, the evaluation and testing of the dashboard were not done with professional users. Which could provide valuable insights into the dashboard's effectiveness based on their experience working with dashboards.

The future work to be done to the dashboard includes increasing the number of faculties and programs for a more complete view of the whole institution. Not only that, but also the dashboard could include the function to determine the number of scholarships, grants, staff, classrooms, etc. to allocate based on the predicted number of students in the future to help assist better with planning. More refinement can be done to improve the user-centricity of the dashboard by including a specific visual identity to the dashboard for the institution that is using the dashboard. The login page of the dashboard could perhaps include more functions to register new users and to have an option to forget passwords. More future work can be done by comparing other different time series prediction models and finding perhaps a better model for student enrolment prediction purposes. As time goes by, more specific features tailored for individual institutions can be added freely. In addition, external factors such as economic indicators and demographic data will also be incorporated to expand the predictive model. Such additions will improve the accuracy of the model and this increase resource efficiency.

7. CONCLUSION

In the end, if we had to choose one model to be used in the dashboard for forecasting, it would be the LSTM Model. As based on the results of the models and the prediction accuracies, LSTM Model is the most accurate model among the other models despite the slow processing time and random changes of prediction values every time the model is being ran. Even though it may be slower in processing, it produces more accurate and reliable results and as for the random values, we can counter that with the random seed function. The Prophet Model is also a valid choice of model to choose from however, despite being fairly accurate with its predictions, it may not be the best at predicting as it has a subpar predictive performance when compared to other time series models. This is especially important from a business view of perspective, which means even the slightest miscalculation of the student enrolment prediction values, means an impact on the operation of the institution. As the student enrolment predictions could help the institution with allocating the future number of staff, grants, scholarships, facilities upgrade cost and much more. Therefore, having the most accurate prediction model to predict the student enrolment number has a more accurate representation of the future enrolment number, thus allowing the management of the institution to have the ease of heart while planning of time without worrying about overpredicting or underpredicting scholarships, grants, etc. Additionally, it's essential to highlight the broader objectives of this study, focusing on the Sustainable Development Goals (SDGs) of the United Nations, specifically Goal 8 (Decent Work and Economic Growth) and Goal 4 (Quality Education) (UNDP, 2015). The dashboard's predictive accuracy can play a pivotal role in achieving these goals by enabling informed decisions related to staffing, resources, and educational quality. In conclusion, this study has successfully achieved the three main objectives written in this report and the main aim to develop a dashboard system to assist the top management of a HEI to make informed decisions based on the forecasted and visualized student enrolment numbers in the dashboard.

REFERENCES

Aljaloud, A. S., Uliyan, D. M., Alkhalil, A., Elrhman, M. A., Alogali, A. F. M., Altameemi, Y. M., Altamimi, M., & Kwan, P. (2022). A Deep Learning Model to Predict Student Learning Outcomes in LMS Using CNN and LSTM. *IEEE Access*, 10, 85255–85265. https://doi.org/10.1109/ACCESS.2022.3196784

Anishnama. (2023, April 28). *Understanding LSTM: Architecture, Pros and Cons, and Implementation. Medium*; Medium. https://medium.com/@anishnama20/understanding-lstm-architecture-pros-and-cons-andimplementation3e0cca194094

Azevedo, A., Azevedo, J., & Hayakawa, M. E. (2021). Designing and Implementing a Dashboard with Key Performance Indicators for a Higher Education Institution. *CSEDU 2021 - 13th International Conference on Computer Supported Education*, 1, 165–172. https://doi.org/10.5220/0010539501650172

Brownlee, J. (2017, May 23). *A Gentle Introduction to Long Short-Term Memory Networks by the Experts* -Machine Learning Mastery. https://machinelearningmastery.com/gentleintroduction-long-short-term-memory-networks-experts/

Calzon, B. (2022, October 18). *What Is A Data Dashboard? Definition, Meaning & Examples*. BI Blog | Data Visualization & Analytics Blog | Datapine. https://www.datapine.com/blog/data-dashboards-definition-examples-templates/

Chen, Y., Li, R., & Hagedorn, L. S. (2019). Undergraduate International Student Enrollment Forecasting Model: An Application of Time Series Analysis. *Journal of International Students*, 9(1), 242–261. https://doi.org/10.32674/jis.v9i1.266

Dela Cruz, A., Basallo, L., Bere, B., Aguilar, J., Calvo, C. K., Arroyo, J. C., & Delima, A. J. (2020). Higher Education Institution (HEI) Enrollment Forecasting Using Data Mining Technique. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(2), 2060–2064. https://doi.org/10.30534/ijatcse/2020/179922020

DiBattista, J. (2022, November 18). *Choosing a ML Time Series Model for Your Data | Towards Data Science*. Medium; Towards Data Science. https://towardsdatascience.com/choosing-the-best-ml-time-series-model-for-your-data-664a7062f418

Gavrilova, Y. (2021). *Introduction to Polynomial Regression Analysis*. Serokell Software Development Company. https://serokell.io/blog/polynomial-regression-analysis

Ghani, L. A., Cob, Z. C., Drus, S. M., & Sulaiman, H. (2019). Student Enrolment Prediction Model in Higher Education Institution: A Data Mining Approach. *Proceedings of the 3rd International Symposium of Information and Internet Technology (SYMINTECH 2018)*, 565, 43–52. https://doi.org/10.1007/978-3-030-20717-5_6

Goyal, D. (2020, February 17). *How does Prophet work? Part-1 - Analytics Vidhya - Medium*. Medium; Analytics Vidhya. https://medium.com/analytics-vidhya/how-does-prophet-work-44addaab6148

Haris, N. A., Abdullah, M., Hasim, N., & Rahman, F. A. (2016). A Study on Students Enrollment Prediction using Data Mining. In *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication*, 45, 1–5. https://doi.org/10.1145/2857546.2857592

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

James, F., & Weese, J. (2021). *Neural network-based time series forecasting of student enrollment with exponential smoothing baseline and statistical* *analysis of performance*. https://krex.k- state.edu/bitstream/handle/2097/41506/FridayJames2021.pdf?sequence=9

Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*, 184, 115537–115537. https://doi.org/10.1016/j.eswa.2021.115537

Krieger, M. (2021, February 20). *Time Series Analysis with Facebook Prophet: How it works and how to use it.* Medium; Towards Data Science. https://towardsdatascience.com/time-series-analysis-with-facebook-prophet-how-it-works-and-how-to-use-it-f15ecf2c0e3a

Keoy, K. H. & Kwek. C. L., (2012). What Can Go Wrong? Developing and Sustaining an Entrepreneurial and Entrepreneurship Eco-System within an Educational Institution. *International Conference on Innovation, Management and Technology Research*. DOI: 10.1109/ICIMTR.2012.6236454

Kwek, C. L., Yeow, K. S., Zhang, L., Keoy, K. H., & Japos, G. (2022). The Determinants of Fake News Adaptation during COVID-19 Pandemic: A Social Psychology Approach. *Recoletos Multidisciplinary Research Journal*, 10(2), 19–39. https://doi.org/10.32871/rmrj2210.02.05

Kumar, N. (2023, March 12). *Polynomial Regression with Examples*. Spark by Examples; Spark by Examples. https://sparkbyexamples.com/machine-learning/polynomial-regression-with-examples/

Lavilles, R. (2012). Enrollment Forecasting for School Management System. *International Journal of Modeling and Optimization*, 2(5), 563–566. https://doi.org/10.7763/IJMO.2012.V2.183

Lawton, G., Carew, J. M., & Burns, E. (2022, January 21). *Predictive Modeling*. Enterprise AI. https://www.techtarget.com/searchenterpriseai/definition/predictive-modeling

Malnik, J. (2022, June 15). *Why Are Dashboards So Important for Your Business? 8 Ways They Help Improve Business Performance | Databox Blog*. Databox. https://databox.com/why-are-dashboards-important

McEvoy, D., Hafeez, K. & Keoy, K. H. (2010). Special issue: ethnic minority entrepreneurship and management: introduction from special editors. *International Small Business Journal*, 28(2), 131-135. DOI.10.1177/0266242609356031

Meng, S., Cai, J., Zhang, Y., Wei, J., Zheng, X., & Wang, J. (2022). Predictive Analysis of Higher Education Enrollment Rate in Cambodia Based on ARIMA Model. *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)*, 220–223. https://doi.org/10.1109/ICPECA53709.2022.9719236

Nils R. (2021, May 12). *Student Enrollment Dashboard for Higher Education Institutions - Example, Uses. Solver*. https://www.solverglobal.com/report-budget-forecast-and-dashboard-template-glossary/student-enrollment-dashboard-for-higher-education-institutions/

Pant, A. (2019, January 13). *Introduction to Linear Regression and Polynomial Regression*. Medium; Towards Data Science. https://towardsdatascience.com/introduction-to-linear-regression-and-polynomial-regression-f8adc96f31cb

Pepin, N., Shibghatullah, A. S., Subaramaniam, K., Sulaiman, R. A., Abas, Z. A., & Sarsam, S. (2022). A Reusable Product Line Asset in Smart Mobile Application: A Systematic Literature Review. *International Journal of Advanced Computer Science and Applications*, 13(9). https://doi.org/10.14569/ijacsa.2022.0130906

Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., Mon, C. S., & Shibghatullah, A. S. (2022). Characterizing Suicide Ideation by Using Mental Disorder Features on Microblogs: A Machine Learning Perspective. *International Journal of Mental Health and Addiction*. https://doi.org/10.1007/s11469-022-00958-z

Sharma, H. S., & Joshi, H. D. (2022). Pooling Business Intelligence and Dashboard Technology for Decisions Making in Higher Education Institutions. *Towards Excellence*, 14(3), 36–48. https://doi.org/10.37867/te140306

Sharma, K. (2015, April 14). *Top 12 Software Development Methodologies*. TatvaSoft Blog. https://www.tatvasoft.com/blog/top-12-software-development-methodologies-and-its-advantages-disadvantages/

Shilbayeh, S., & Abonamah, A. A. (2021). Predicting Student Enrolments and Attrition Patterns in Higher Educational Institutions using Machine Learning. *The International Arab Journal of Information Technology*, 18(4), 562–567. https://doi.org/10.34028/18/4/8

Sinuany-Stern, Z. (2021). Forecasting Methods in Higher Education: An Overview. In *International series in management science/operations research* (pp. 131–157). https://doi.org/10.1007/978-3-030-74051-1_5

Slim, A., Hush, D., Ojah, T., & Babbitt, T. (2018). Predicting Student Enrollment Based on Student and College Characteristics. *International Educational Data Mining Society*, 383–389.

Tuovila, A. (2023). *Forecasting: What It Is, How It's Used in Business and Investing*. Investopedia. https://www.investopedia.com/terms/f/forecasting.asp#:~:text=Forecasting%20is%20a%20technique%20that,an%20upcoming%20period%20of%20time.

UNDP. (2015). *Sustainable Development Goals | United Nations Development Programme*. UNDP. https://www.undp.org/sustainabledevelopmentgoals?gclid=Cj0KCQjw98uBhCgARIsAD7eAiOq2B8amoWeChbLK3nDJAfbBykYIqQQVvHaRDvi9tZdpzijio vxmmoaAIVGEALw_wcB

University of Cambridge. (2023). *Workbook: Cambridge Undergraduate Admissions Statistics*. Cam.ac.uk. https://tableau.blue.cam.ac.uk/t/InformationHubPublic/views/CambridgeUndergraduateAdmissionsStatistics/CourseStatistics?%3Aembed=y&%3AisGuestRedirectFromVizportal=y

Vick, T. (2020, October 2). *What is Student Enrollment Forecasting? | FLO Analytics*. FLO. https://www.flo-analytics.com/news/what-is-student-enrollment-forecasting/