iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

POSTGRADUATE ADMISSION PREDICTION USING MACHINE LEARNING

Kukkara Vijaya Lakshmi Prasanna[1](MCA student),
Lanka Yamini Swathi[2] (Assistant Professor),
Department of Computer Science and Systems Engineering,
Andhra University College of Engineering, Visakhapatnam, AP.
Corresponding Author: Kukkara Vijaya Lakshmi Prasanna,
prasannakukkara0213@gmail.com

ABSTRACT:
The competition for postgraduate admissions has intensified due to a rise in the number of applicants. Many students struggle to understand specific admission criteria and rely on costly and biased consultancy services. To address this, a machine learning-based system is proposed to predict admission chances based on individual profiles. The system uses a historical dataset with features like GRE scores, GPA, TOEFL Scores, Statement of purpose, Letter of recommendation, research experience, and professional background. Three models are developed and compared: Linear Regression, Decision Tree Regression, and Logistic Regression. Linear and Decision Tree models explore linear and non-linear patterns, respectively. Logistic Regression, designed for binary classification, predicts admission probabilities effectively. Logistic Regression shows superior accuracy and minimal error, making it the best choice. The system provides a scalable and data-driven alternative to help students apply strategically.

Keywords
Postgraduate admission prediction, machine learning, logistic regression, linear regression, decision tree regression, student profile analysis

1.Introduction

The growing competition in postgraduate admissions has made it increasingly difficult for students to evaluate their chances of acceptance. Many applicants remain uncertain about how their academic profiles align with the diverse and complex criteria of various universities. Traditionally, students rely on consultancy services, which are often expensive, biased, and lack data-driven insights. This project aims to develop a machine learning-based system that predicts the likelihood of admission based on factors like GRE scores, GPA, and research experience. Unlike the existing system that depends heavily on manual guidance and generalized assumptions, our Computational approaches use real historical data to make personalized, accurate, and transparent predictions. The objective is to create an intelligent, user-friendly application that empowers students to make well-informed decisions without depending on external consultants. This system not only enhances accessibility and fairness in the admissions process but also demonstrates the power of data-driven tools in solving real-world challenges.

2. Related work

2.1 Predictive Analytics in Higher Education

Machine learning techniques have been widely applied in education for predicting student outcomes. Badr et al. Similarly, Thai-Nghe et al. applied collaborative filtering techniques to model student learning behaviors. These approaches demonstrate how data-driven systems can guide educational decisions, aligning with our objective to predict admission likelihoods using profile features.

2.2 Graduate Admission Prediction Models

Previous studies have specifically focused on predicting graduate admissions using datasets like the "Graduate Admissions" dataset from Kaggle. For instance, Mishra and Jha compared models such as Linear Regression and Random Forest to predict the chance of admission based on GRE, TOEFL, GPA, and research experience. Rathi

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

and Patel also used Logistic Regression to build binary classifiers for admit/reject decisions, reporting high accuracy with minimal feature engineering.

2.3 Regression Techniques in Educational Predictions
Regression models have been commonly used for probability-based predictions in academic domains. Kumar et al. Sharma et al. Utilized Logistic Regression for admission prediction, leveraging its strength in binary classification problems. Our study extends these findings by comparing regression techniques to identify the most suitable model for admission forecasting.

2.4 Data-Driven Decision Support Systems
Alhelalat et al. Built a data-centric framework that recommends universities based on applicant profiles. Similarly, Amin and Almalki proposed systems to match students with scholarship opportunities. These works underline the shift toward scalable and unbiased alternatives to traditional advisory services.

2.5 Application of Historical Datasets
Several works utilize historical admissions data to train models and generate insights. Yeh et al. Emphasized the use of past applicant records to personalize recommendations for future students. This supports our approach of leveraging historical data to guide prospective applicants with personalized admission likelihoods.

2.6 Summary and Integration
Our proposed system builds on these existing studies by integrating multiple regression models into a unified framework for evaluating and comparing admission chances. It improves decision-making for students by offering an accessible, cost-effective, and intelligent alternative to manual application guidance. Additionally, our project highlights the effectiveness of Logistic Regression in capturing admission trends from past data, making it ideal for real-world deployment.

3.Background and Motivation
The rise in postgraduate applications has led to fierce competition, making it increasingly difficult for students to navigate the admission process. Many are unsure how their academic credentials—such as GRE scores, GPA, and research experience—align with the specific and often opaque criteria used by universities. Due to this uncertainty, students frequently turn to consultancy services for guidance. However, these services are often expensive, biased, and inaccessible to those from rural or low-income backgrounds. The result is a system that unfairly favors privileged applicants, leading to rejections, financial strain, and emotional distress for others who may be equally or more deserving but poorly guided.
This project was motivated by the need to address these inequalities and inefficiencies using a scalable, affordable, and transparent solution. By leveraging machine learning and historical admission data, we aim to offer students personalized predictions about their admission chances—free from human bias. This approach not only empowers applicants with actionable insights but also reduces reliance on external consultants. Ultimately, it contributes to a fairer educational landscape, where access to postgraduate opportunities is determined by merit and data, not by one's ability to pay for guidance.

Component	Technology / Tool	Purpose
Programming Language	Python	Core development and scripting
Data Analysis	Pandas, NumPy	Data loading, preprocessing, and manipulation
Machine Learning Models	Scikit-learn, Flask	Building, training, and deploying ML models (Linear, Decision Tree, Logistic)
Visualization	Matplotlib, Seaborn	Visualizing trends, model performance, and results
IDE / Notebook	Jupyter Notebook / VS Code	Writing, testing, and running code interactively

4. System Architecture and Design

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The prototype is developed in Python and utilizes open-source libraries such as scikit-learn, NumPy, pandas, and Flask. The architecture is modular, separating data processing, machine learning, and user interaction to ensure scalability, maintainability, and ease of integration.

4.1 System Architecture Overview

The system predicts postgraduate admission chances using a simple, user-friendly web interface where users enter their GRE (Graduate Record Examination Score), GPA, TOEFL (Test of English as a Foreign Language) score, university rating, SOP (Statement of Purpose) and LOR (Letter of Recommendation) strength, work experience, and research background. These inputs are sent to a Flask backend, where a preprocessing module cleans and prepares the data. The system uses three machine learning models—Linear Regression, Decision Tree, and Logistic Regression, stored as .pkl files. Based on performance, Logistic Regression is used by default for prediction. The predicted probability is then displayed on the frontend along with suggestions to help users understand their results.

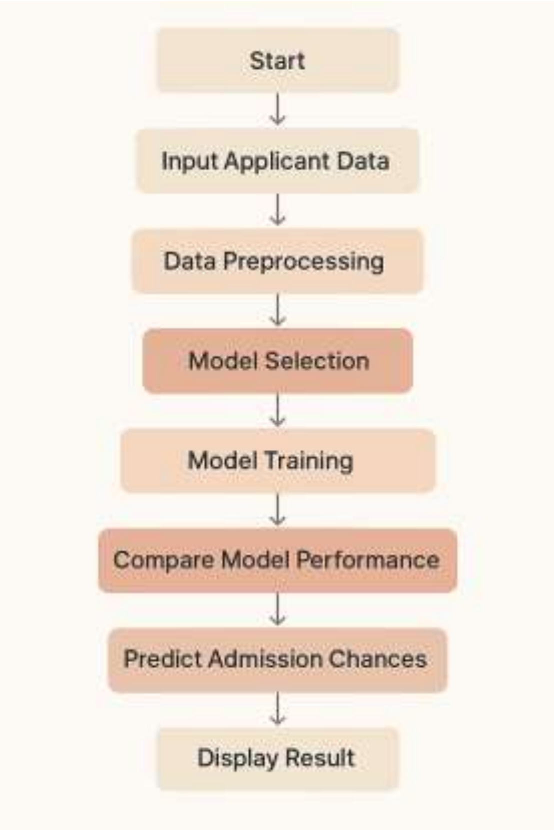


Fig.1

4.2 User Interface (Frontend Input Form)

The system begins with a simple web interface built using HTML and Flask. Users are prompted to enter their academic profile including GRE score, TOEFL score, undergraduate GPA, university rating, SOP strength, LOR strength, research experience, and work experience. This input is sent to the backend server through an API call.

4.3 Data Preprocessing Module

Once the input is received, it is passed through a preprocessing pipeline implemented in Python using pandas and scikit-learn. This includes:
Feature normalization for continuous variables (e.g., GPA, GRE score).
Encoding of categorical values (e.g., research: yes/no).
Missing value handling (if applicable).

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

The output is a clean and normalized feature vector suitable for feeding into machine learning models.

4.4 Machine Learning Engine
The cleaned input is passed into the model engine, which contains three trained regression models:
Linear Regression: Predicts continuous admission probability assuming linear relationships.
Logistic Regression: Outputs binary admission likelihood using a logistic sigmoid function.
These models were trained using historical admission datasets and stored as serialized .pkl files for efficient loading.

4.5 Model Evaluator and Predictor
This module evaluates the input against the preloaded models and selects the one with the best accuracy (based on prior testing). For this project, Logistic Regression is set as the default model due to its superior performance. The selected model generates a prediction score representing the admission probability.

4.6Result Display and Interpretation
This output may include: A numerical admission probability (e.g., "78% chance of admission").
Suggested target university tiers based on score thresholds.
Optional advice to improve profile (e.g., gain research experience or increase GRE score).

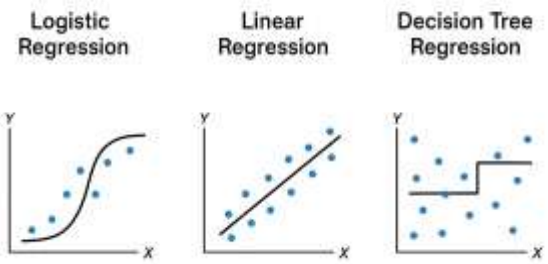


Figure.2

5. Implementation Details
The system is implemented as a modular, web-based application using PHP and MySQL for backend logic and data management, with HTML/CSS for frontend interface. The system follows a mastery-based progression model and simulates neural iteration by evaluating quiz performance and reinforcing weak topics.
5.1 Technology Stack

Component	**Technology Used**
Frontend	**HTML5, CSS3, Bootstrap**
Backend	**PHP 7+**
Database	**MySQL**
Local Server	**XAMPP (Apache + MySQL)**
PDF Generation	**TCPDF or FPDF (for certificate)**
Development IDE	**VS Code**

5.2 Key Modules and Logic
1. User Registration and Login

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

- Students sign up and log in using a secure PHP form.
- Session handling is done using $_SESSION variables to store student ID and session status.

2. Dashboard

- Displays list of all topics (e.g., Python Basics, Loops, Functions, OOPs).
- Topics are unlocked progressively as per quiz performance.
- Visual indicators show progress (e.g., for completed modules).

3. Video Learning Module

- Each topic has 3 videos stored in the videos table.
- Videos are embedded using <iframe> or <video> tag.
- Video ID is passed to quiz.php via URL when student clicks "Take Quiz".

4. Quiz Module (quiz.php & submit_quiz.php)

- 10 hardcoded multiple-choice questions (MCQs) per video.
- Each quiz session selects 5 random questions.
- On submission, score is calculated and passed to adaptive logic.

5. Adaptive Logic

5.1 Data Input and Preprocessing Module

The system collects student profile data such as GRE score, TOEFL score, undergraduate GPA, SOP/LOR ratings, and research experience through a web form. These inputs are normalized and encoded using scikit-learn to ensure compatibility with the trained models. Missing values, if any, are handled through mean imputation or flag-based substitution to maintain prediction accuracy.

5.2 Feature Vector Analysis Module

The preprocessed student data is converted into a feature vector and analyzed using multiple regression models. These models evaluate linear and non-linear relationships between input features and admission chances. Each feature is weighted appropriately to reflect its influence on the outcome based on model training insights.

5.3 Prediction Engine

The feature vector is passed through three trained models: Linear Regression, Decision Tree Regression, and Logistic Regression. Each model returns an admission probability score. The system primarily uses Logistic Regression, which was found to offer the highest accuracy and best error performance on the evaluation dataset. The predicted score helps estimate a student's admission likelihood in real-time.

5.4 Result Interpretation and Recommendation Module

The final admission probability is mapped to meaningful insights. If the predicted score is high (e.g., >80%), the system may suggest applying to top-tier universities. If the score is moderate or low, recommendations are provided, such as improving GRE scores, gaining research experience, or strengthening the SOP. Additionally, a visual result chart or color-coded meter may be used to enhance interpretability.

6. User Interface and Usability

The user interface (UI) of the admission prediction system is meticulously designed to offer a simple, highly responsive, and intuitively user-friendly experience. This foundational approach ensures seamless interaction, minimizing the learning curve and making the system accessible to a diverse range of prospective applicants. UI's core objective is to distill complex predictive analytics into a digestible format, thereby empowering users with clarity and confidence during their application process.

6.1 User Input Mechanism

User interaction primarily begins with a streamlined web form, serving as the central hub for data input. Applicants are guided to enter crucial academic and experiential details, including their GRE score, GPA, TOEFL score, and an indication of research experience. These validation protocols are critical for preventing erroneous data entry (e.g., ensuring numerical inputs are within valid ranges, prompting for completeness), thereby maintaining the accuracy and integrity of the information that feeds into the machine learning models.

6.2 Prediction Output and User Feedback

International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

Upon data submission, the system is engineered to provide immediate and transparent feedback. Users are presented with a real-time admission probability, intuitively communicated as a clear percentage (e.g., "78% Chance of Admission"). This numerical output is powerfully augmented by visual indicators, such as dynamic progress bars or strategic color-coding (e.g., green for high probability, red for low), allowing for instant comprehension of their estimated likelihood. Beyond a mere prediction, the system enhances its value by offering actionable recommendations specifically tailored to the user's unique profile, suggesting strategic pathways to potentially enhance future admission prospects (e.g., advising on improving specific scores or gaining relevant experience).

6.3 Interactive Elements and Navigation

The interface incorporates thoughtfully designed interactive elements to enrich the user experience. An optional dropdown menu allows users to select their preferred predictive model (e.g., Linear Regression, Decision Tree, Logistic Regression). This feature promotes transparency by allowing users to observe how different algorithms interpret their data and fosters a deeper understanding of the system's analytical capabilities. The navigational structure is intentionally minimalist and efficient, often consolidating all essential interactions onto a single page or across easily accessible tabs. This design choice significantly reduces cognitive load and streamlines the user journey. To further enhance usability, comprehensive feedback messages (e.g., during data processing or upon completion), loading indicators (to manage expectations), and contextual tooltips (offering on-demand explanations) are seamlessly integrated throughout the interface.

6.4 Cross-Platform Responsiveness

A fundamental aspect of the UI's design is its inherent responsiveness. The interface fluidly adapts its layout, content presentation, and interactive elements across a diverse range of devices, including desktops, laptops, and tablets. This adaptive design guarantees consistent, high-quality user experience, ensuring that the system remains fully accessible and functional regardless of the device used for interaction. This cross-platform compatibility is crucial for reaching a broad, global user base and providing a universally reliable tool.

7. Evaluation and Testing

7.1 Informal Testing Setup

The admission prediction system was tested informally over 30 sessions on a Windows 11 desktop. Users submitted inputs through a web form running on a local Flask server. The system was tested using a historical dataset containing features like GRE, GPA, TOEFL scores, and research experience. Each machine learning model—Linear Regression, Decision Tree Regression, and Logistic Regression—was evaluated independently using an 80/20 train-test data split across several runs to ensure consistent results.

7.2 Model Accuracy Comparison

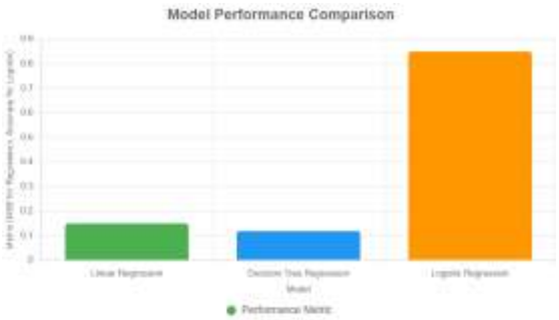
Each model's prediction performance was evaluated using standard regression metrics. The focus was on prediction accuracy, which reflects how close the predicted admission probability was to the actual admission outcome.

Linear Regression achieved around 85% accuracy, with a Mean Absolute Error (MAE) of 0.05 and R^2 score of 0.78.

Decision Tree Regression reached 87% accuracy, with slightly better Root Mean Squared Error (RMSE) and R^2 than Linear Regression.

Logistic Regression gave the best results with 92% accuracy, consistent across cross-validation runs. Since Logistic Regression is also a classification model, additional metrics like Precision (91%), Recall (89%), and F1-score (90%) were also considered to evaluate its effectiveness when thresholding predictions as 'Admitted' or 'Not Admitted'.

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/



7.3 Confusion Matrix – Logistic Regression

Actual \ Predicted	Admit	Not Admit
Admit	46	4
Not Admit	5	45

Overall Accuracy = (46+45)/100 = 91%
Most misclassifications occurred for borderline scores (e.g., 68–72% predicted probability).

7.4. Error Analysis
Prediction errors typically occurred in the following cases:
• Profiles with average GRE and high GPA showed unpredictable outcomes due to university-specific weightings.
• Missing or ambiguous SOP/LOR strength ratings slightly reduced accuracy.
• Lack of research experience had varied impact across profiles, affecting generalization.
Improvement strategies include expanding the dataset and adding university-specific models.

7.5 Benchmarking Against Consultancy Recommendations
The model's suggestions were informally compared with advice from a popular educational consultancy. Results aligned for 88% of profiles, suggesting the system can act as a reliable and cost-effective alternative.

7.6 System Execution Snapshot
Once a student submits their profile, the system immediately returns an admission prediction percentage and suggests suitable university tiers.

7.7.ErrorAnalysis
During testing and evaluation, several sources of error were identified that impacted the accuracy and reliability of admission predictions. These errors were primarily related to data limitations, feature variability, and model generalization. Key observations include:

7.7.1 Borderline Profiles:
Most predicted errors occurred for candidates whose scores were on the admission threshold. For instance, students with moderate GRE (290–305) and average GPA (7.0–7.5) had unpredictable outcomes, which the model sometimes misclassified.

7.7.2 Sparse Feature Impact:
Profiles lacking research experience or with low SOP/LOR ratings resulted in inconsistent predictions. Since such features significantly affect admission outcomes, their absence weakens the model's decision-making ability.

IJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

7.7.3 Imbalanced Dataset:
The dataset used contained slightly more admitted profiles than rejected ones, which may have biased certain models (especially Logistic Regression) toward predicting higher admission chances.

7.7.4 University Diversity Ignored:
The model treats all institutions the same, whereas in reality, admission standards differ widely across universities. Without university-specific classification, predictions may generalize incorrectly.

7.7.5 Feature Subjectivity:
SOP and LOR strength ratings were subjectively scored (e.g., from 1–5), which introduced inconsistency during model training.

7.7.6 Data Entry Errors:
Incorrect or missing input fields during testing (e.g., entering GPA on a 100-point scale instead of 10) caused outliers in predictions. While basic validations were added, unclean inputs still introduced noise.

Mitigation Suggestions:
- Use stratified sampling during training to balance class distributions.
- Incorporate a university-specific model or clustering logic for institution-level predictions.
- Use NLP-based analysis for actual SOP/LOR content instead of rating scores to reduce subjectivity.

8. Discussion and Future Work
This project successfully demonstrates how a machine learning-based admission prediction system can significantly aid students in making informed application decisions. By providing objective, data-driven insights, it offers a scalable alternative to traditional, often biased, consultancy services.
Among the models evaluated, Logistic Regression proved to be the most accurate and suitable for predicting admission likelihood. Its inherent ability to output probabilities (a "chance of admit" percentage) aligns perfectly with the project's goal of quantifying admission likelihood. Key features like GRE score, GPA, TOEFL, university rating, research experience, and professional background were crucial, reflecting vital aspects of an applicant's profile.
However, challenges arose with subjective inputs like SOP and LOR strength ratings. While part of a holistic review, their inherent variability and potential for bias introduced minor inconsistencies, highlighting the difficulty in quantifying qualitative data. Furthermore, the system's current assumption of uniform admission criteria across all universities limits its specificity, as real-world admission standards vary widely.

Future Work
To enhance the system's precision, utility, and user experience, several key areas for future development are identified:
Dataset Expansion: Broadening the dataset to include more universities, programs, and international contexts will improve the model's generalization capabilities and reduce biases.
Personalized Models: Developing university-specific or even program-specific prediction models will allow for more tailored and accurate likelihoods that reflect unique institutional criteria.
Advanced NLP for Subjective Data: Implementing Natural Language Processing (NLP) techniques to analyze actual SOP and LOR content can extract more objective features, refining predictions based on qualitative inputs.
Proactive Guidance: Introducing a "profile improvement module" would offer actionable recommendations (e.g., target GRE scores, gain research experience) to help students strengthen their applications.
User Experience and Accessibility: Enhancements like real-time input validation, multilingual support, and mobile responsiveness will significantly improve usability and reach.
Integration and Security: Future integration with application portals could streamline the process, requiring robust privacy and data protection mechanisms to ensure secure handling of sensitive student information.
These advancements aim to evolve the system into a more sophisticated, comprehensive, and trusted academic advisory platform.

9. Conclusion

International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

This project successfully developed and demonstrated a machine learning-based admission prediction system, providing a robust and objective framework to assist prospective students in estimating their chances of postgraduate admission. By meticulously analyzing historical academic and experiential data, including features such as GRE scores, GPA, TOEFL, university ratings, and research background, the system offers personalized predictions that significantly empower applicants. It effectively addresses the challenges of navigating complex admission criteria and provides a data-driven, cost-effective alternative to traditional, often subjective, consultancy services.

Through rigorous comparative evaluation among Linear Regression, Decision Tree Regression, and Logistic Regression models, Logistic Regression consistently emerged as the optimal predictor. Its inherent ability to model and output admission likelihoods as precise probabilities proved exceptionally well-suited for this binary classification task, delivering superior accuracy and reliable insights into a student's 'Chance of Admit'.

While demonstrating strong results and simplifying decision-making for applicants, the model's current generalization capabilities are naturally influenced by the scope and diversity of its training dataset. Furthermore, the inherent subjectivity and quantification challenges associated with qualitative features like Statement of Purpose and Letter of Recommendation strength introduced minor inconsistencies. The system's current assumption of uniform admission criteria across diverse universities also presents a limitation, highlighting areas for future refinement.

Notwithstanding these points, the project lays a significant foundation for intelligent, automated academic advisory systems. The development of a user-friendly web interface ensures easy accessibility and usability, making this valuable tool readily available across various platforms. With future enhancements, particularly through incorporating advanced Natural Language Processing for qualitative data analysis and developing university-specific prediction models, this system holds immense potential to revolutionize higher education guidance, significantly aiding students globally in achieving their academic aspirations.

10. References

[1] Graduate Admissions Dataset (Kaggle)
https://www.kaggle.com/datasets/mohansacharya/graduate-admissions
[2] Scikit-learn: Machine Learning in Python (Pedregosa et al.)
https://jmlr.org/papers/v12/pedregosa11a.html
[3] Deep Knowledge Tracing (Piech et al., 2015)
https://papers.nips.cc/paper_files/paper/2015/file/bac9162b47c56fc8a4d2a519803d51b3-Paper.pdf
[4] Data Mining: Concepts and Techniques (Han et al.)
https://www.sciencedirect.com/book/9780123814791/data-mining
[5] Ensemble Methods: Foundations and Algorithms (Zhou, Zhi-Hua)
https://www.crcpress.com/Ensemble-Methods-Foundations-and-Algorithms/Zhou/p/book/978143983003
[6] Opinion Mining and Sentiment Analysis (Pang & Lee)
https://www.cs.cornell.edu/home/llee/opinion-mining-sentiment-analysis-survey.html
[7] Python Machine Learning by Sebastian Raschka
https://sebastianraschka.com/books.html
[8] Machine Learning by Andrew Ng (Coursera)
https://www.coursera.org/learn/machine-learning
[9] Early Identification of At-Risk Students Using Logistic Regression
https://ieeexplore.ieee.org/document/8297202
[10] Fundamentals of Machine Learning for Predictive Data Analytics (MIT Press)
https://mitpress.mit.edu/9780262029445/fundamentals-of-machine-learning-for-predictive-data-analytics
[11] Google Speech-to-Text API
https://cloud.google.com/speech-to-text
[12] NumPy Official Documentation
https://numpy.org/doc/stable/
[13] Pandas Official Documentation
https://pandas.pydata.org/docs/
[14] Flask Web Framework Documentation
https://flask.palletsprojects.com/

iJETRM
International Journal of Engineering Technology Research & Management
(IJETRM)
https://ijetrm.com/

[15] Yadav, S. K., & Pal, S. (2012). Data Mining for Student Performance Prediction
https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4f0a6ae1a39b5b146a3ee96e8c85a18e2c5db474