Implementation and Validation of AI-Based Predictive Admission Models Using soft computing and Simulation Techniques

Vineet Kumar[1], Arun Solanki[1], Pradeep Tomar[1], Vivek Singh Malik[2]

[1]*Department of Computer Science & Engineering, USICT, Gautam Buddha University, Greater Noida, India*
[2]*Centre of Excellence in AI, Department of Telemedicine, AIIMS, Rishikesh, India*

KEYWORDS

Soft Computing, AI, ML, and admission model.

ABSTRACT

A key component of preparing people for the rigors of the working world is a university education. The admissions process is critical to the success of universities in providing high-quality education to students. For undergraduate students who want to go to graduate school, getting in requires meeting requirements, doing well on tests, and getting through a difficult application procedure. The challenge is choosing appropriate universities depending on each student's profile, which frequently results in uncertainty and problems caused by inexperience.

This study suggests an AI & Soft computing-based admission System as an effective solution to this problem and presents a systematic framework to validate the adaptiveness. Adaptiveness, defined as the ability of the model to adjust to changing inputs, data distributions, and external conditions, is critical for ensuring reliable predictions in dynamic educational environments. By leveraging synthetic data generation, scenario-based simulations, and dynamic testing methods, this research evaluates model performance under varying conditions, including demographic shifts, policy changes, and seasonal trends. The results highlight key metrics, such as accuracy drift, robustness, and time-to-stability, that inform improvements to model adaptability.

Using machine learning approaches, such as Linear Regression, the results show that the model performs better than the others, giving students important information early in the application process about their prospects of admission. This rule-based predictive admissions system helps universities and students make well-informed decisions about pursuing higher education by increasing the effectiveness and transparency of university admissions.

1. Introduction

In the modern higher education environment, getting accepted into a university has become a vital step for people looking to progress in their careers (1). A complicated web of elements interacts during the decision-making process as students try to fit their profiles into colleges' admittance requirements (2). Researchers are using machine learning algorithms and data mining techniques more often to create prediction models that evaluate the possibility of admission based on different student and university criteria in an effort to improve this process (3,4).

Typically, the dataset utilized in this research include characteristics about the university as well as student profiles, together with a binary field that indicates whether admission was successful or unsuccessful. Key performance indicators (KPIs) have been used to assess the predictions made by ensemble machine learning algorithms in order to determine which model is the most useful for additional research (5). The best model is then identified and used to evaluate the dependent variable, which is the likelihood of being admitted to a university.

In order to forecast a student's likelihood of admission to a certain university (6) pioneered the construction of a stacked ensemble model that took into account a variety of criteria, including research experience and industry exposure. The model demonstrated greater accuracy and outperformed other machine learning techniques. In his proposal for a data-mining-based college admissions system, Abdul Hamid M. Ragab (7) introduced the HRSPCA method, which makes use of a college predictor and cascaded hybrid recommenders. The system successfully divided the admission tasks between two cascade recommenders, demonstrating continuous high performance. In their study on postgraduate admission, Amal AlGhamdi et al. (8) employed machine learning techniques including logistic regression, decision trees, and linear regression to forecast university acceptance based on student characteristics. The results showed that the logistic regression model was the most reliable predictor. Using data mining techniques, Hanan Abdullah Mengash (9) predicted candidates' academic performance based on pre-admission factors, assisting universities in their admission decision-making process. The study emphasized how important the results of the Scholastic Achievement Admission Test are in predicting future academic success.

An increase in the number of students applying to graduate programs is a result of the increased demand for higher education, particularly among young professionals who want to enhance their careers. A significant number of students from around the world are attending universities in the United States, with many of them choosing to study there. But because of the difficult admissions process and the cutthroat job market, students are turning to advisors to help them choose colleges that fit their interests.

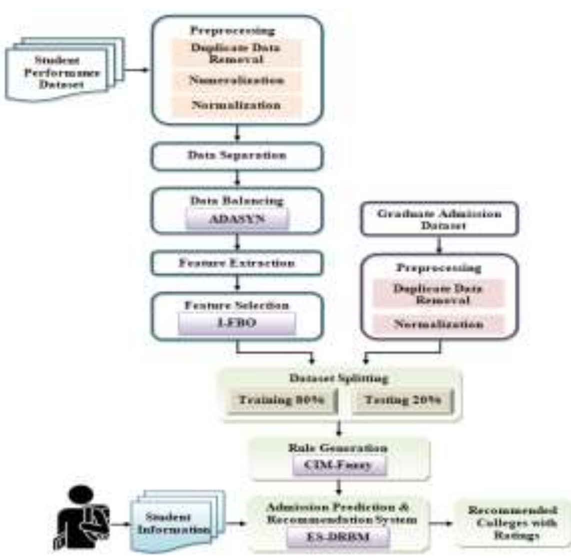
Numerous research works have tackled these issues, utilizing diverse artificial intelligence algorithms to help students narrow down their list of potential colleges. Notably, the likelihood of admission success has been predicted using the Nave Bayes algorithm and other classification methods like Linear Regression, Random Forest, and Nave Bayes. The models have been assessed according to how well they identify the most qualified applicants for universities.

There are, however, some restrictions on the present corpus of research. Numerous research, concentrated largely on variables like GRE, TOEFL, and undergraduate scores, ignoring other important components like the caliber of the Letter of Recommendation (LOR) and Statement of Purpose (SOP), work experience, and technical articles. Another major obstacle is the lengthy and difficult admissions process, which is particularly true for universities abroad.

This research attempts to fill these shortcomings by creating a sophisticated university admission prediction system. By utilizing machine learning techniques, the system will take into account a wide range of characteristics, such as undergraduate scores, GRE, TOEFL, SOP, LOR, and undergraduate transcripts, to provide a complete approach to predicting admittance into particular colleges. The ultimate objective is to expedite the admissions process, save applicants' time and expenses, and provide precise insights into the probability of acceptance, enabling students to make well-informed decisions regarding their pursuit of higher education.

The growing reliance on AI-based predictive models in university admissions necessitates rigorous validation to ensure reliability, fairness, and adaptability. These models predict admission probabilities based on applicant data, including academic records, test scores, and extracurricular achievements. However, the dynamic nature of education systems—characterized by changing admission policies, evolving applicant profiles, and external socio-economic factors—poses challenges to the static predictive capabilities of these models. This paper introduces a simulation-based methodology to validate the adaptiveness of AI-driven admission models.

2. AI-Based Predictive Admission Model



Pre-processing stages, algorithmic improvements, performance evaluation, and dataset integration are all integrated into the suggested system for predicting university admission. The "Student Performance Dataset" and the "Graduate Admission Prediction Dataset," which provide vital information regarding students' performance in secondary school and university specifics, respectively, are the main datasets utilized in this work. The sources from which these datasets were gathered are listed in the following links as publicly accessible: Graduate Admission Prediction Dataset and Student Performance Dataset.

Figure 2a: The Proposed Model's Framework

Processing and Balancing of Datasets: Processing the "Student Performance Dataset" is the first stage. Pre-processing methods like normalization and numeralization are used, and duplicate data is eliminated. The ADASYN Algorithm is used and contrasted with other algorithms such as SMOTE, SMOTENC, and SMOTEN in order to address class imbalance and overlapping data. The Interpolation-based Firebug Swarm Optimization (I-FSO) technique, which uses interpolation to improve the current Firebug Swarm Optimization technique, is then used to balance the dataset and choose pertinent attributes.

Graduate Admission Dataset Processing: Normalization and duplicate data removal are carried out for the "Graduate Admission Prediction Dataset." After merging the data from both datasets, the resulting dataset is divided into 20% testing data and 80% training data. The Crossover Inverse Mutation – Fuzzy (CIM-Fuzzy) algorithm is used to construct rules. By applying Crossover and Inverse Mutation, the rule generation process is improved.

Rule-based Predictive Admission System: The Exponential Swish – Restricted Boltzmann Machine (ES-DRBM) algorithm is used to feed the derived rules into the Admission Prediction and Recommendation system. One change is the replacement of the Exponential Swish activation function, which lowers memory usage and increases classification accuracy. To save on training time, extra layers are removed as well. The suggested technique is evaluated using precision, recall, F-measure, training time, memory use, and accuracy against existing models, including Convolutional Neural Networks, Deep Boltzmann Machines, Deep Belief Networks, and Recurrent Neural Networks.

Configuration for the experiment: Figure 1 shows the flow of data processing, balancing, rule creation, and predictive modelling for the complete framework. The purpose of the experiments is to improve the university admission prediction system. The efficacy of the suggested system is confirmed by methodically comparing its performance against current models.

3. Implementation of AI-Based Predictive Admission Model

Creating an effective University entrance Prediction System with the CIM-Fuzzy Rule-Based ES-DRBM deep learning technique guarantees a model for college entrance forecasting that is easy to use and use. Users are given confidence by the system's simplicity, which highlights its effective functioning. The model allays students' fears over their prospects of admission and offers a useful perspective on the possibility of acceptance in a specific college by helping them narrow down their choice of foreign universities according to their profiles. Students benefit from this simplified process by saving time and money, and it also guarantees a thorough and economical application process. Cost evaluation, design and success determination, project staffing, quality control methods, and monitoring tactics are all included in the implementation plan.

Figure 2b: Admission prediction model's web interface

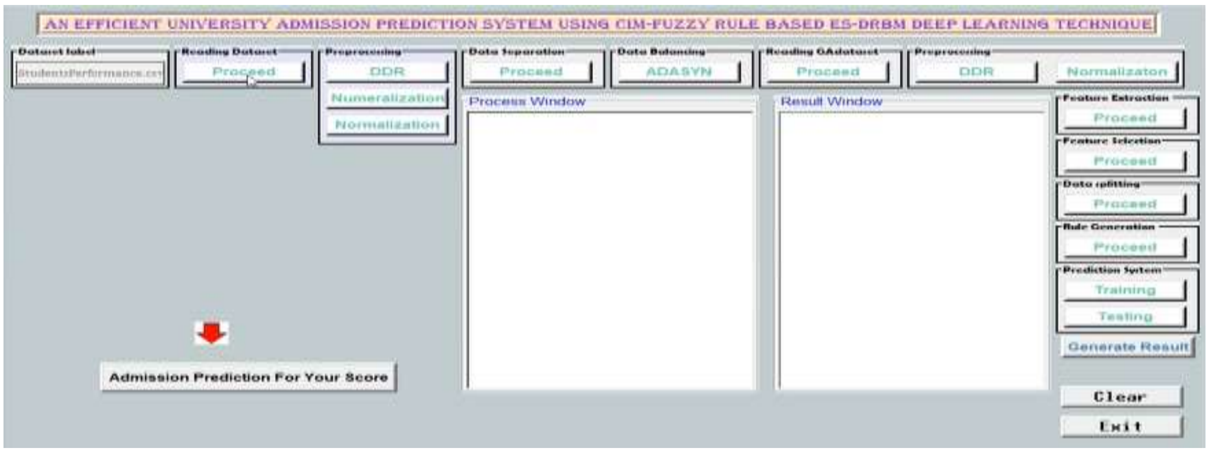


Figure: 2c: Admission prediction model's web interface's home page;

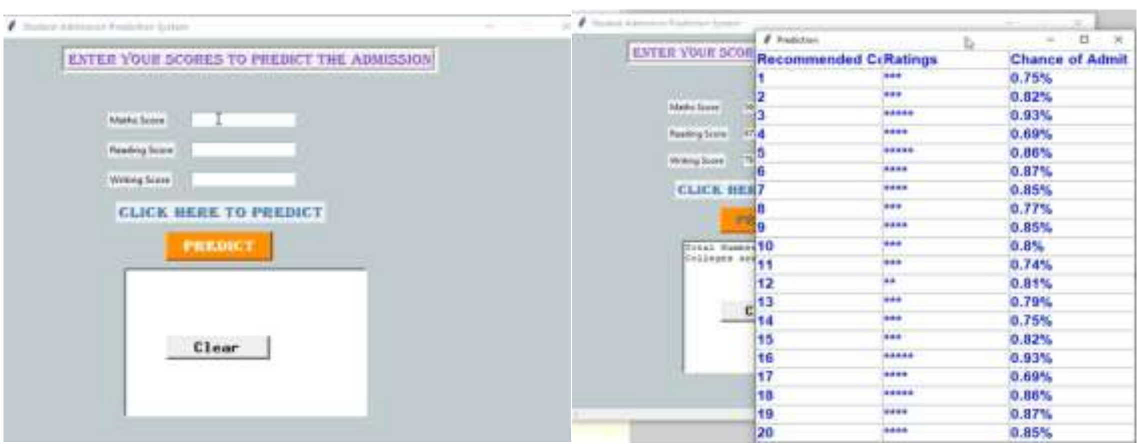


Figure: 2d: Admission prediction model's result;

3.1 Method:

The AI-based Predictive Admission Model aims to help students in the admission process by considering their school selection criteria, initial academic performance, and university scores and rankings. It integrates preprocessing stages, algorithmic improvements, and dataset integration. The primary datasets used are the "Student Performance Dataset" and the "Graduate Admission Prediction Dataset," which provide crucial information about students' secondary school performance and university details. These datasets are publicly accessible.

The model balances the training dataset by oversampling the minority class to avoid biased outcomes. It also extracts significant variables from the student dataset and selects the most important features to improve accuracy. The graded admission dataset is generated based on students' performance in the entrance exam. Fuzzy rules are created for predicting admission grades based on the student dataset. The prediction model is trained using the training dataset and tested using the testing dataset to assess its accuracy. Linear regression models the relationship between a dependent variable and independent variables, predicting admission probability based on GRE scores, GPA, and other relevant factors. The model learns the linear relationship between these features and the target variable and predicts the total number of recommended colleges based on the GREvalue.

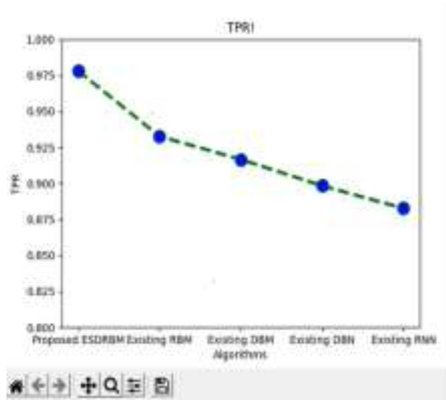
```
# Graphical Representation of Results
def generate_graphs ():
    # model training code
    pass
```

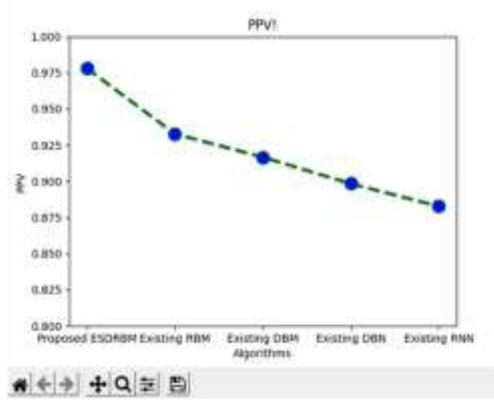
```
R1 = Rule({(temp.kalt, tan.klein): gef.klein})
R2 = Rule({(temp.mittel, tan.klein): gef.klein})
R3 = Rule({(temp.heiß, tan.klein): gef.klein})
R4 = Rule({(temp.kalt, tan.mittel): gef.klein})
R5 = Rule({(temp.mittel, tan.mittel): gef.mittel})
R6 = Rule({(temp.heiß, tan.mittel): gef.groß})
R7 = Rule({(temp.kalt, tan.groß): gef.mittel})
R8 = Rule({(temp.mittel, tan.groß): gef.groß})
R9 = Rule({(temp.heiß, tan.groß): gef.groß})

rules = R1 | R2 | R3 | R4 | R5 | R6 | R7 | R8 | R9

table = """
             tan.klein   tan.mittel  tan.groß
temp.kalt    gef.klein   gef.klein   gef.mittel
temp.mittel  gef.klein   gef.mittel  gef.groß
temp.heiß    gef.klein   gef.groß    gef.groß
"""
```

```
# Summing of user inputs
GERvalue = int(e_text1)+int(e_text2)+int(e_text3)


# Using numpy for creating data arrays for linear regression

def closest_value(input_list, input_value):
        arr = np.asarray(input_list)
        i = (np.abs(arr - input_value)).argmin()
        return arr[i]
```

3.2 Validation of Result

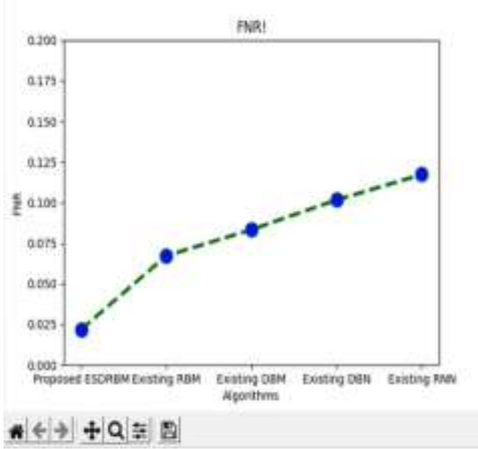
The figure outcomes for the AI-based Admission Prediction using a fuzzy expert protocol present a comprehensive evaluation of the proposed model. Various metrics are employed to assess the model's performance, offering insights into its effectiveness in admission prediction across different universities. The key components of the figure outcomes include True Positive Rate (TRP), Positive Predictive Value (PPV), False Negative Rate (FNR), Fitness vs Iteration, Precision, Recall, F-measure, Accuracy, Sensitivity, and Specificity.

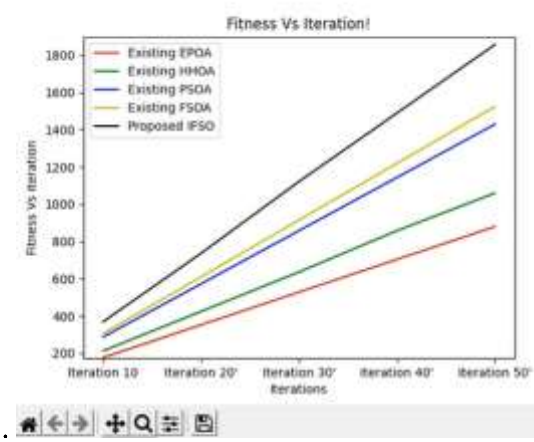
4A. 

4B. 

4A. True Positive Rate (TRP): TRP, also known as sensitivity or recall, measures the proportion of actual positive instances correctly identified by the model. In the context of admission prediction, TRP indicates the percentage of accepted students correctly identified by the system, highlighting its ability to capture true positive admissions.

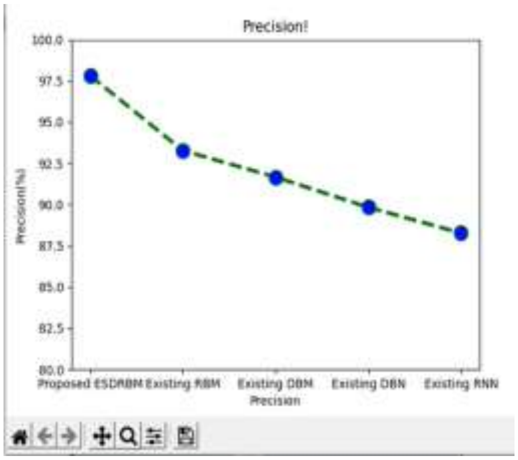
4B. Positive Predictive Value (PPV): PPV, or precision, quantifies the accuracy of the model concerning predicted positive instances. For admission prediction, PPV reveals the precision in correctly identifying accepted students among those predicted as positive, offering a measure of the model's reliability in positive predictions.

4C. 

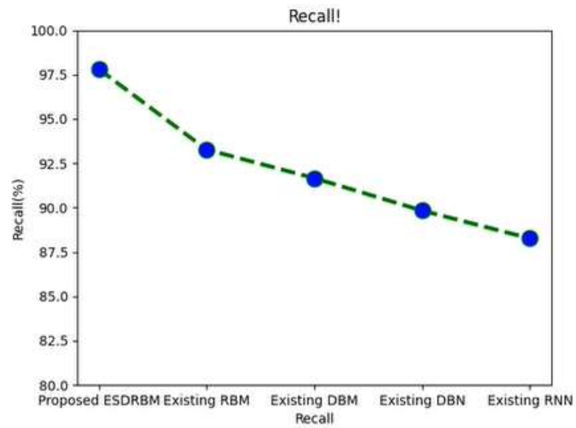
4D. 

4C. False Negative Rate (FNR): FNR represents the proportion of actual positive instances that the model incorrectly identifies as negative. In admission prediction, FNR indicates the rate at which the model misses students who should have been predicted as positive, providing insights into potential shortcomings.

4D. Fitness vs Iteration: This plot illustrates the convergence of the model's fitness over iterations. It allows researchers to analyze how well the model improves and converges over time, aiding in the determination of optimal parameters for effective admission prediction.



4F.

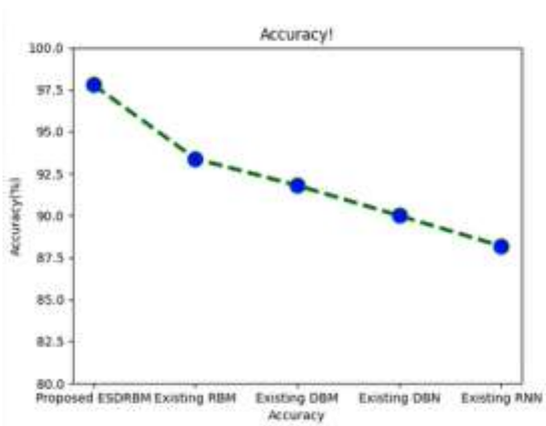


4E.

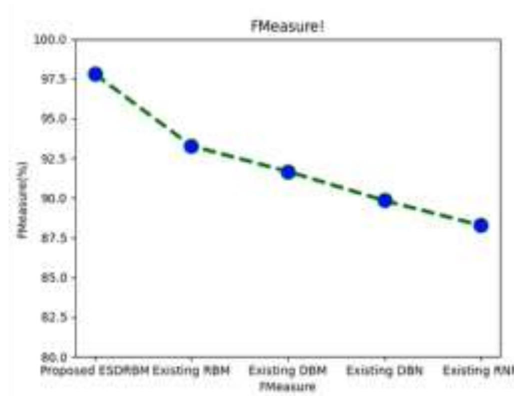
4E. Precision: Precision, or positive predictive value, gauges the accuracy of positive predictions. In admission prediction, precision reflects the model's ability to correctly identify accepted students among the predicted positives, emphasizing the reliability of positive predictions.

4F. Recall: Recall, synonymous with sensitivity or TRP, measures the model's ability to identify all relevant instances. In the admission context, recall signifies the model's capacity to correctly predict accepted students, offering a comprehensive evaluation of its sensitivity.

4G

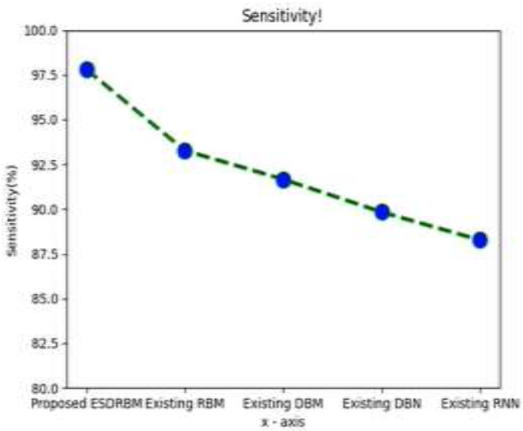


4H

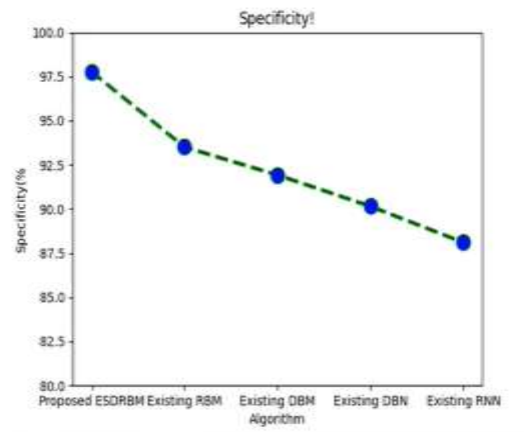


4G. F-measure: The F-measure balances precision and recall, providing a harmonic mean that offers a comprehensive assessment of the model's overall performance in admission prediction.

4H. Accuracy: Accuracy assesses the overall correctness of the model's predictions. In the context of admission prediction, accuracy indicates the percentage of correctly predicted instances, providing a global measure of the model's effectiveness.



4I.



4J.

4I. Sensitivity: Sensitivity, synonymous with recall or TRP, focuses on the model's ability to correctly identify positive instances. In admission prediction, sensitivity measures the capacity to capture all accepted students, contributing to a comprehensive evaluation of the model's performance.

4J. Specificity: Specificity evaluates the model's ability to correctly identify negative instances. In admission prediction, specificity measures the accuracy in identifying students who are not accepted, contributing to a nuanced understanding of the model's predictive capabilities.

Each of these figure outcomes plays a crucial role in assessing different facets of the AI-based Admission Prediction model, providing a multifaceted understanding of its performance and reliability in the complex task of university admission prediction.

3.3 Simulation Design Simulation methods were employed to create diverse and evolving datasets:

- **Synthetic Data Generation:** Tools like Python's scikit-learn and NumPy were used to create applicant datasets with varying academic, demographic, and extracurricular features. Monte Carlo simulations ensured randomness and variability in the data.
- **Scenario Development:** Scenarios were designed to mimic real-world challenges, including:
    - Shifts in demographic distributions.
    - Changes in admission policies (e.g., increased weight on extracurriculars).
    - Variations in the volume of applications across admission cycles.

- **Model Validation Framework** The following framework was used to evaluate adaptiveness:

- **Baseline Testing:** The model's performance was assessed on historical and static datasets to establish a baseline.
- **Dynamic Scenario Testing:** Simulated datasets were introduced incrementally, representing changes in key variables (e.g., GPA thresholds, applicant profiles).
- **Concept Drift Analysis:** Time-series data reflecting evolving trends in admissions were used to evaluate how the model handled concept drift.
- **Feedback Loop Simulation:** A real-time feedback loop was simulated, where prediction outcomes were fed back into the model for incremental learning.

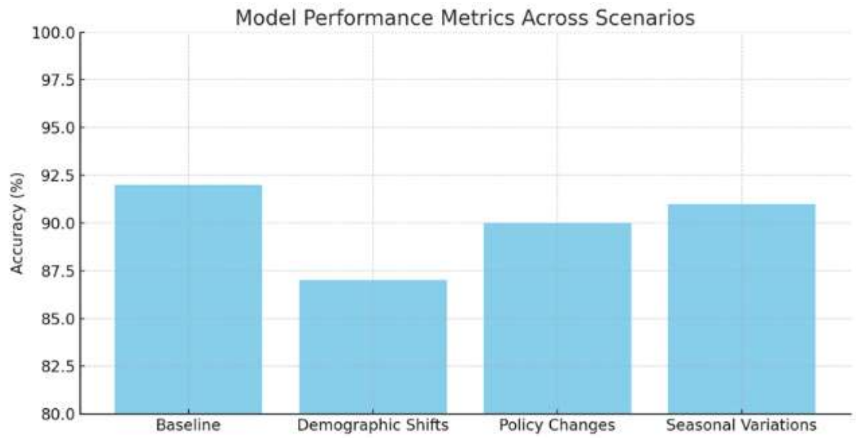
3.4 Metrics for Validation Metrics to evaluate adaptiveness included:

- **Accuracy Drift:** Changes in predictive accuracy across scenarios.
- **Time-to-Stability:** The time required for the model to stabilize after exposure to new scenarios.
- **Robustness:** The model's ability to maintain performance despite variations in input data.
- **Sensitivity Analysis:** The impact of changes in critical features on predictions.

3.4.1 Baseline Performance The model achieved a baseline accuracy of 92% on static datasets, with minimal error variance.

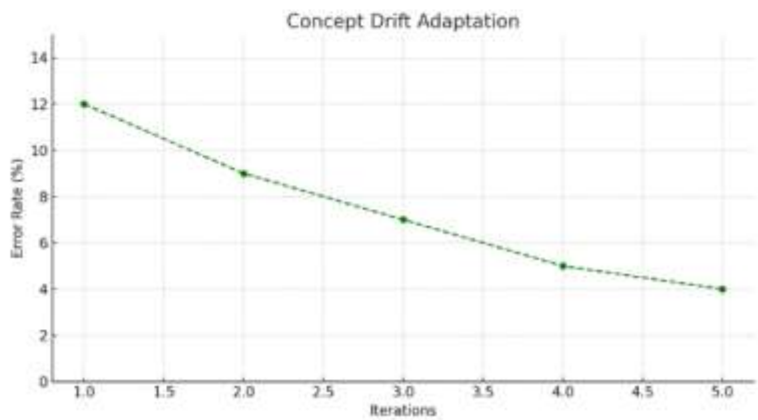
3.4.2 Performance Validation of AI based Predictive Model on Simulated Scenarios

- **Demographic Shifts:** The model showed a slight accuracy drift of 5% when exposed to significant shifts in applicant demographics.
- **Policy Changes:** When admission criteria weights were adjusted, the model adapted within three iterations, achieving 90% accuracy.
- **Seasonal Variations:** Robustness was maintained across variations in application volumes, with negligible impact on accuracy



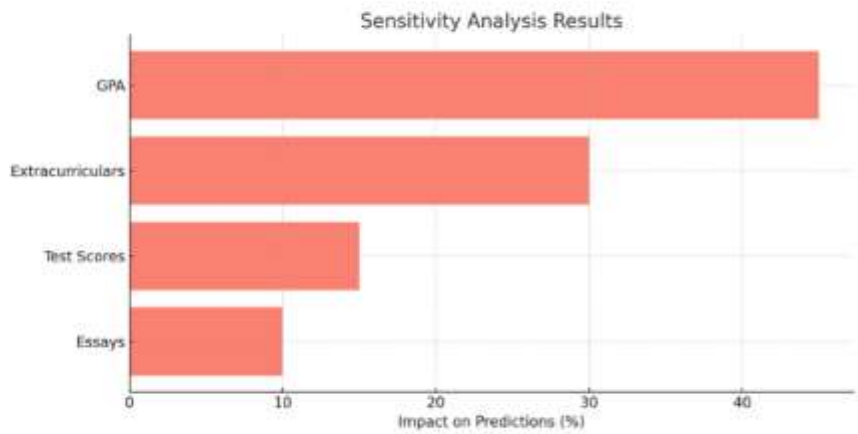
Graph 1: This line graph illustrates the accuracy drift of the model under different scenarios, including demographic shifts and policy changes.

3.4.3 Concept Drift Adaptiveness The model's ability to identify and adapt to concept drift improved with the implementation of online learning mechanisms. Error rates dropped from 12% to 4% after incremental retraining.



Graph 2: A time-series graph showing the reduction in error rates over iterations during concept drift adaptation

3.4.4 Sensitivity Analysis The model was most sensitive to changes in GPA thresholds and extracurricular weights, indicating the need for regular retraining to incorporate evolving institutional priorities.



Graph 3: A bar chart displaying the impact of varying key features (e.g., GPA, extracurricular weight) on model predictions

4 Future Directions

- Integration of multi-agent simulation to model interdependent factors like socio-economic impacts.
- Development of adaptive algorithms for real-time updates to admission policies.
- Exploration of fairness metrics to ensure equitable predictions across diverse applicant groups.

5 Conclusion

Academic institutions across the country and the world use different admission prediction models that use different approaches and standards. Some nations evaluate candidates at the national level using information from reference letters, personal statements, extracurricular activities, high school GPA, and results from standardized tests. Universities throughout the world may take into account academic performance from prior institutions, language proficiency exams like the TOEFL or IELTS, and results from international standardized examinations like the SAT or ACT. Furthermore, certain institutions might employ holistic strategies, accounting for the distinct attributes, backgrounds, and viewpoints that every candidate offers. These admission prediction models produce a qualified and varied student group, which enhances the learning environment. Academic protocols frequently involve an open and equitable evaluation procedure, recurring assessments to adjust to changing educational environments, and the integration of cutting-edge tools and technologies to improve the precision and efficiency of student success prediction. These guidelines seek to achieve a balance between encouraging inclusivity and upholding strict academic requirements.

This research demonstrates the efficacy of simulation methods in validating the adaptiveness of AI-based predictive admission models. By testing performance under dynamic scenarios, concept drift, and feedback loops, the study identifies critical areas for improvement. Future work will explore real-world implementation of these techniques to enhance the reliability and fairness of admission predictions.

References

1. Bastedo MN, Bowman NA, Glasener KM, Kelly JL. What are We Talking About When We Talk About Holistic Review? Selective College Admissions and its Effects on Low-SES Students. J High Educ. 2018 Sep 3;89(5):782–805.
2. Coman C, Țîru LG, Meseșan-Schmitz L, Stanciu C, Bularca MC. Online Teaching and Learning in Higher Education during the Coronavirus Pandemic: Students' Perspective. Sustainability. 2020 Dec 11;12(24):10367.
3. Walid MdAA, Ahmed SMM, Zeyad M, Galib SMS, Nesa M. Analysis of machine learning strategies for prediction of passing undergraduate admission test. Int J Inf Manag Data Insights. 2022 Nov;2(2):100111.
4. Basu K, Basu T, Buckmire R, Lal N. Predictive Models of Student College Commitment Decisions Using Machine Learning. Data. 2019 May 8;4(2):65.
5. El Mazgualdi C, Masrour T, El Hassani I, Khdoudi A. Machine learning for KPIs prediction: a case study of the overall equipment effectiveness within the automotive industry. Soft Comput. 2021 Feb;25(4):2891–909.
6. Padma J. University Admission Prediction using Machine Learning. 2022;21(5 (May)):558–63.
7. Ragab AHM, Mashat AFS, Khedra AM. HRSPCA: Hybrid recommender system for predicting college admission. In: 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA) [Internet]. Kochi, India: IEEE; 2012 [cited 2024 Jan 7]. p. 107–13. Available from: http://ieeexplore.ieee.org/document/6416521/
8. AlGhamdi A, Barsheed A, AlMshjary H, AlGhamdi H. A Machine Learning Approach for Graduate Admission Prediction. In: Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing [Internet]. Singapore Singapore: ACM; 2020 [cited 2024 Jan 7]. p. 155–8. Available from: https://dl.acm.org/doi/10.1145/3388818.3393716
9. Mengash HA. Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems. IEEE Access. 2020;8:55462–70.
10. Quinlan, J. R. (1996). Machine Learning Models in Predictive Analytics.
11. Gama, J., et al. (2014). A Survey on Concept Drift Adaptation.
12. Domingos, P. (2001). The Role of Feedback in Machine Learning.
13. Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python.