Admission Prediction in Undergraduate Applications: an Interpretable Deep Learning Approach

Amisha Priyadarshini
Department of Computer Science
University of California, Irvine
Irvine, CA, USA
apriyad1@uci.edu

Barbara Martinez-Neda
Department of Computer Science
University of California, Irvine
Irvine, CA, USA
barbarm@uci.edu

Sergio Gago-Masague
Department of Computer Science
University of California, Irvine
Irvine, CA, USA
sgagomas@uci.edu

***Abstract*—This article addresses the challenge of validating the admission committee's decisions for undergraduate admissions. In recent years, the traditional review process has struggled to handle the overwhelmingly large amount of applicants' data. Moreover, this traditional assessment often leads to human bias, which might result in discrimination among applicants. Although classical machine learning-based approaches exist that aim to verify the quantitative assessment made by the application reviewers, these methods lack scalability and suffer from performance issues when a large volume of data is in place. In this context, we propose deep learning-based classifiers, namely Feed-Forward and Input Convex neural networks, which overcome the challenges faced by the existing methods. Furthermore, we give additional insights into our model by incorporating an interpretability module, namely LIME. Our training and test datasets comprise applicants' data with a wide range of variables and information. Our models achieve higher accuracy compared to the best-performing traditional machine learning-based approach by a considerable margin of 3.03%. Additionally, we show the sensitivity of different features and their relative impacts on the overall admission decision using the LIME technique.**

***Index Terms*—undergraduate admissions, machine learning, deep neural networks, LIME, input convex neural networks.**

I. Introduction

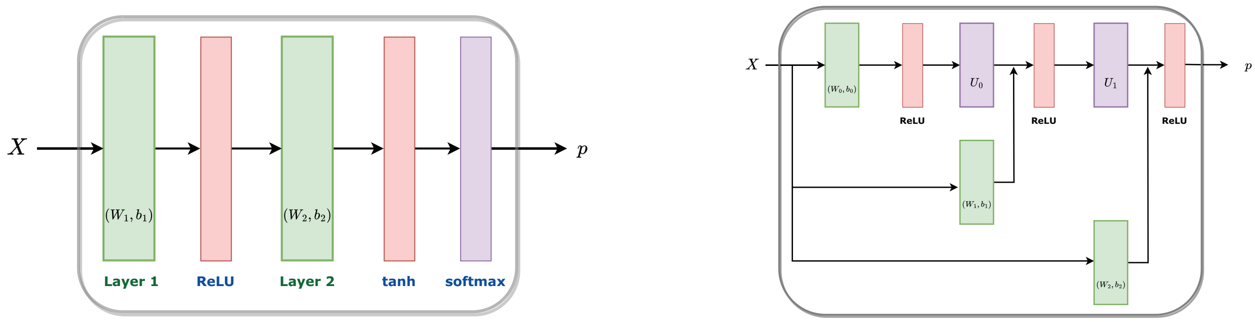
In recent decades, the undergraduate admission processes have witnessed significant transformations aimed at fostering fairness and equal opportunities for applicants. These changes are evident in the education system of the University of California (UC), which includes eliminating the consideration of race and gender, implementing the percentage plans to recognize the top-performing students, and removing standardized testing along with the holistic review approach, which considers factors beyond the academic achievements of applicants. Integrating machine learning (ML) techniques in the undergraduate admissions decision-making process could ensure fairness by eliminating any human bias and prejudices that may inadvertently influence otherwise. Automating certain aspects of the decision process using ML models could also significantly enhance overall efficiency by saving time and resources for the admissions staff. Articles [1] and [2] showcase two such software implementations of classical ML tools which could be used for the admission process and analyze the hidden patterns in the corresponding datasets. Another advantage of using ML classifiers is that they provide scalable solutions to tackle the increased workload by handling the increasing number of applications while ensuring every application's thorough and timely review. In support of our assertions, [3] explores the applicability and viability of classical ML models in the undergrad admissions process. The extensive simulation studies and findings thereof have led the way to multiple follow-up works like investigating the performance of state-of-the-art classifiers, such as Deep Neural Networks (DNN), for undergraduate admissions.

Deep learning (DL) has emerged as a powerful tool in various fields, from demonstrating their potential in understanding complex data to revolutionizing the decision-making process. While classical ML models have shown promise, their performance in admission decisions may still be limited by their inability to capture intricate relationships within high-dimensional data. In contrast, DL models, with their multi-layer architecture, can learn hierarchical representations and are hence well suited for extracting meaningful information from complex and diverse applicant data. Furthermore, the flexibility and adaptability of DNNs make them suitable for handling the dynamic nature of admissions data.

Given the success of DNN models, the investigation of transparency, fairness, and bias, concerning the domain of the admission decision process, still remains an active research area. To ensure the interpretability and explainability of the predictions obtained from the DNN models, we explore the Local Interpretable Model-agnostic Explanations (LIME) technique [4] coupled with a gradient-based approach. This could help extract invaluable information to understand an applicant's specific attributes and characteristics that heavily influence admission decisions.

For the paper, we have considered 4,442 application records of California freshman applicants for the Fall 2021 cycle to the Department of Computer Science at UC, Irvine. The dataset encompasses a range of variables, including demographics,

979-8-3503-5801-8/23/$31.00 ©2023 IEEE
DOI 10.1109/TransAI60598.2023.00040



(a) (b)

Figure 1: Visual representation of deep neural network architectures, (a) Feed-Forward neural network; (b) Input convex neural network.

academic records, high school information, and essay question responses. Using the Python-based PyTorch framework, we have trained the datasets on two different architectures of DNN models, namely Feed-Forward (FF) Neural Networks and Input Convex Neural Networks (ICNN). Additionally, we have incorporated the LIME technique to offer interpretability and explanations for the predictions made by the DNN models. By presenting our findings, we aim to contribute to the existing body of knowledge on leveraging DNN models in undergraduate admission decisions while upholding the holistic review process.

II. Related Works

The use of ML models is gaining significant attention in recent years. However, a few studies have investigated the application of ML algorithms to predict and enhance the accuracy of admission decisions. Most of the studies conducted so far have focused on the graduate admission process, emphasizing on factors such as Undergrad Cumulative Grade Point Average (CGPA), Research, and Vacancies at research groups but there is a lack of research on undergraduate admissions which requires a broader evaluation of application aspects. For instance, [5] and [6] have conducted studies using supervised learning techniques where the sole focus is on standardized testing approaches for training the models while lacking transparency in their prediction explanations. In another line of work [1] and [2], depict the development of ML tools for predicting post-graduate and graduate admissions respectively. The former was discarded because of non-updated training data and counterproductive approach towards campus diversity [7]. Another study [8] talks regarding the bias in admission predictions, which we have tried to address in our study. Apart from that, [9] and [10] conducted study using Deep neural models with the latter using three different frameworks. Both papers did not explore any interpretation techniques that could be deployed to understand the model's predictions and also, in turn, preserve transparency. Furthermore, [11] focuses on machine learning-based prediction for graduate admissions but the scarce number of training data samples make the predictions unreliable.

III. Methodology

This section provides an overview of the methodology, starting with data preprocessing, neural network training, and incorporating the LIME model for interpretability.

A. Data preprocessing

To facilitate the training of deep learning models and assess their performance in a comparison study with the classical machine learning models, we have obtained and pre-processed information on 4,442 applications to the CS department at UC, Irvine. The dataset used in this study comprises several categories of information, including the student Grade Point Average (GPA), Advanced Placement (AP) test scores, participation in educational programs, and responses to admission-related questions. Some dependent variables are demographic information, academic history, high school attended, and the responses to selected Personal Insight Questions (PIQs) [12].

Given the problem we are trying to solve is a binary classification task, the final read score is assigned as the target variable, representing the review score assigned to every applicant. For this task, the top review score is mapped to 1, and the lower scores are mapped to 0. The records missing a final read score value are excluded from the analysis to ensure the efficacy of the model training and performance evaluation. Next, to enhance the effectiveness of the model training process, any records with high school attendance outside California are excluded, considering that California applicants provide a wider range of potential information compared to out-of-state ones. Also, the records missing any numerical data are dropped from the dataset, provided its high importance for the classification task. Following Proposition 209, we have dropped features like gender and ethnicity. Features like primary major value consist of string entries, which are transformed into new binary columns after one-hot encoding, indicate the presence or absence of the respective values from the original feature. Students were allowed to choose four PIQs to respond to from a collection of eight, which are later used to extract necessary information. TextBlob [13] and textstat [14] libraries have been employed to extract the information from the PIQs.

After obtaining a tabular dataset, we handle the missing entries by the method of median imputation, where the missing values are replaced with the corresponding feature's median value. This ensures robustness to outliers or extreme values and helps preserve the variable's overall distribution. This step is followed by data normalization, which transforms the data to have zero mean and unit variance. It, in turn, prevents feature dominance and improves performance by ensuring data convergence for gradient-based optimizations performed in a neural network. Subsequently, data scaling was performed to map the normalized data to the range of [0, 1], which is beneficial for algorithms relying on specific input ranges, like the DNN models.

B. Neural Network Training

In the paper, we focus on two different neural network architectures, namely Feed-Forward (FF) and Input convex neural networks (ICNN), for predicting undergraduate admission decisions.

1) Feed-Forward Neural Network: The Feed-Forward neural network is a fundamental architecture where the information propagates within the network in a unidirectional manner, originating from the input layer and progressing towards the output layer. The network lacks any recurrent or feedback loops. We consider a three-layered architecture or a deep neural network with two hidden layers for the binary classification problem. The proposed FF architecture is illustrated in Figure 1(a). For an FF neural network, its input is a state vector and its output is a scalar value.

Definition 1 (Feed-Forward Predicted Output)**.** The predicted output of the network, p, is defined as follows,

$$p = \text{softmax}[\ \tanh(W_2.\ \text{ReLU}(W_1.x + b_1) + b_2)]$$

where W_1 and W_2 are the weight vectors and, b_1 and b_2 are the bias vectors respectively, and x is the input to the network.

We use tanh and ReLU activation functions on the first and second hidden layers, respectively. Despite using the softmax activation function on the output layer of a binary classification problem, it has shown a tremendous improvement in the model's performance along with the cross-entropy loss function.

This architecture enables the network to learn the hierarchical representations of the input data. We employ the ADAM [15] optimizer which is a first-order gradient-based optimization technique to update the weights and bias based on the calculated gradients of the loss function accordingly.

2) Input Convex Neural Network: The second type of architecture used for the study is the ICNN, a scalar-valued neural network with constraints on its parameters or weights such that the output of the network is a convex function of the inputs [16]. The architecture ensures convexity, a fundamental concept in the optimization theory that allows efficient and reliable optimization, explicitly capturing the convex relationships between the input variables. For the study, we have considered a fully connected Input Convex neural network (FICNN) consisting of several passthrough layers. The significance of including passthrough layers between the hidden layers can be beneficial to preserving the convexity of the input space while allowing the neural network to perform non-linear transformations to understand the complex patterns in data. An illustration of the said model is provided in Figure 1(b).

Definition 2 (FICNN Predicted Output)**.** Say, $f(x;\theta)$ is the scalar-valued neural network where x denotes the input to the function and θ are the parameters such that the network, f, is convex to the input x. In order to understand the predictions, the layer-wise output is calculated as,

$$z_{i+1} = g_i(U_i.z_i + W_i.x + b_i)$$

where W_i are the real-valued weights mapping from inputs to the $i+1$ layer activations; U_i are positive weights mapping previous layer activations z_i to the next layer; b_i are the real-valued bias terms; and g_i are convex, monotonically non-decreasing non-linear activation functions for every i representing the training samples. Then the predicted output of the network could be defined as follows,

$$p = g(x) \equiv z_k$$

where k signifies the number of layers in the network.

Furthermore, we incorporate the Dropout regularization technique in the architecture, enhancing the model generalization and robustness to noise and variation in the input data, simultaneously reducing model sensitivity. The network imposes constraints to preserve the convexity with respect to the input, which ensures the output has convex regions in the input space. The significance of the model lies in its ability to preserve convexity leading to more reliable solutions, and helps facilitate optimization to yield higher performance metrics.

3) Principal Component Analysis: In the next step, we incorporate the Principal Component Analysis (PCA) with both the neural network architectures discussed to check for any significant improvement in the model prediction. PCA is a dimensionality reduction technique implemented to transform a high-dimensional dataset into a low-dimensional one while preserving the important features. The technique is applied to address the challenges associated with high-dimensional datasets, potentially improving the model's performance and robustness to noise and redundancy.

C. Prediction Interpretability using LIME

Local Interpretable Model-agnostic Explanations (LIME) is a technique used to interpret the predictions of complex ML models like deep neural networks. It is particularly used when the ML model's inner workings are opaque or difficult to interpret. Being a model-agnostic method, it tries to learn the underlying behavior of the ML model by perturbing the input and observing the changes in the model predictions. As the LIME model is only capable of providing local explanations by approximating the behavior of the ML model in the vicinity

Table I: Performance metrics for deep neural network models

Neural Network Model Architecture	*Accuracy*	*Precision*	*Recall*	*F1-Score*	*AU-ROC Score*
Feed-Forward Neural Network	0.8056	0.8159	0.7883	0.8018	0.8056
Feed-Forward Neural Network with PCA	**0.8067**	0.8000	**0.8073**	0.8037	0.8068
ICNN	**0.8067**	**0.8281**	0.7961	**0.8118**	**0.8073**
ICNN with PCA	0.8056	0.8248	0.7983	0.8113	0.8060

of a data instance, we have tried to adopt a slightly different technique to improve the transparency of model predictions by generalizing them over a broader spectrum of testing samples. To cascade over the limitation, we have tried to do the feature selection using the Gradients in the FF neural network architecture, followed by applying the LIME model on the neural network trained over the selected features. The reason behind the choice of architecture is that we wanted the model to be comparatively easy to interpret and thus help us understand the predictions more efficiently. We try to compute the gradients of the neural network with respect to the input features, which is the measure of the sensitivity of the output due to any perturbation in the input.

Next, we implement the LIME model on the neural network trained on the selected features. This step ensures the dimensionality reduction necessary to eliminate redundant features for better performance and improve the scope of interpretability. In order to tackle the local explanations provided by LIME, we intend to combine the gradient-based feature selection to improve the results by evaluating the model's behavior on a diverse range of instances and perturbing a selected set of features. This helps in understanding the pattern and behavior beyond individual predictions, providing an overview of the model's decision-making process and in turn an attempt at deriving a global explanation using the LIME model.

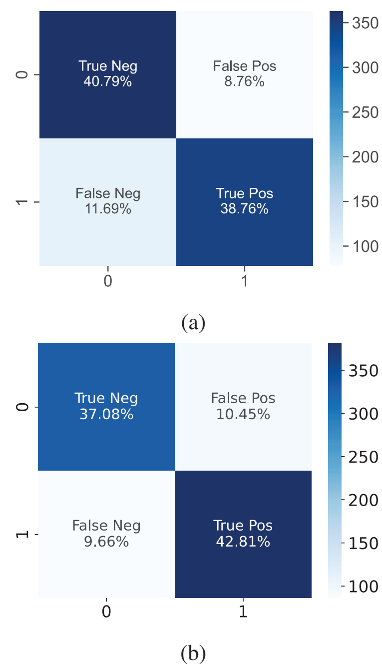


Figure 2: Confusion matrices for fully-trained deep learning models, (a) Feed-Forward neural network; (b) Input convex neural network.

IV. Simulation Results

In this section, we discuss the various performance metrics of our DNN model for student admissions and provide a comparative study with the classical ML models. We further discuss the two DL models, namely FF and ICNN, and the significance of feature importance in the admissions process.

A. *Performance Metrics*

We have implemented and trained the FF and the ICNN, deep learning models on the dataset with 80% used for training and 20% for testing. We have utilized Python-based PyTorch [17] as the DL framework and trained the DL models using the standard backpropagation procedure and included the related code[1]. To evaluate the model's performance we have employed five types of metrics, namely accuracy, precision, recall, F1-score, and AUC-ROC score. As opposed to the classical ML performance in [3] we have successfully been able to showcase significant improvements in performance using the same dataset for the student admission problem in our paper. The accuracy metric is chosen as one of the performance metrics obtained by dividing the number of correct predictions by the total number of records in the dataset, providing a percentage of correct predictions. Owing to the balanced target class, the accuracy metric makes for a suitable and reliable measure to assess model performance. Apart from that, the F1-score and the recall served as effective metrics for understanding the DL model performance. The F1-score represents the harmonic mean of precision and recall, and the recall represents the proportion of true positive predictions out of all actual positive instances.

The ICNN and the FF models both seem to achieve an accuracy of 0.8067, outperforming the classical ML models by a substantial margin. We have also been able to outperform the classical ML models by demonstrating higher precision and F1 scores. As observed from Table I, the ICNN model outperforms the FF model on four metrics including precision, recall, F1-score, and AUC-ROC score, with Accuracy remaining the same as the FF model employed with PCA. Based on the

[1]The GitHub repository link can be found at https://github.com/apriyad1/Deep-Learning-in-Admissions.

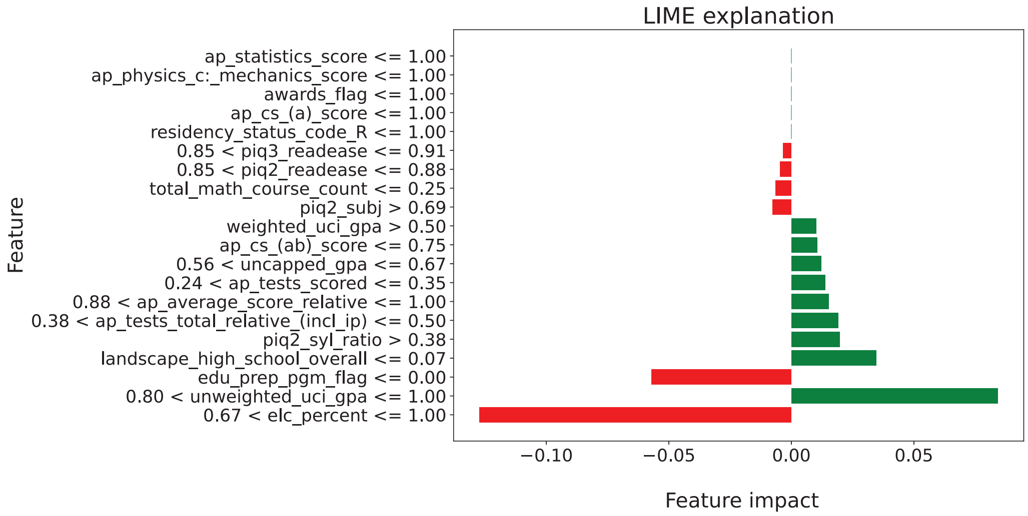


Figure 3: LIME interpretation of the Feed-Forward model predictions. The green and red horizontal bars signify the key features influencing the overall classification in positive and negative ways respectively.

overall statistics, the ICNN model shows promising results in solving the student admission decision problem. The superior performance obtained by the ICNN model can be attributed to its intrinsic convexity that results in robust and efficient non-linear decision boundaries in input space. We note that our simulation results for ICNN conform to the findings of [16] concerning classification tasks.

In Figure 2 we present the confusion matrices for both the FF neural network and ICNN models. The confusion matrix has been provided to give a visual representation of the models' relative performance on the test dataset in predicting admission decisions and thereby evaluating their metrics. As discussed previously, we have incorporated the LIME model using the gradient-based approach for understanding the feature significance of the Deep neural network model. This augmentation leverages the gradients of the Feed-forward neural network model's output with respect to the input features to assess their contribution to the predicted output. We then employ the LIME model to further understand and differentiate the impact of each selected feature on the test dataset.

B. Discussion

In this paper, we discuss the usage of a gradient-based method to initially select the top 20 highly affecting features from our trained Feed-forward neural network model for the student admissions problem. Subsequently, applying the LIME technique to the selected features, we try to analyze the positive and negative feature impacts on our decision-making process. Every class of feature is assigned a probability score based on the possibility of getting chosen. This method enhances our chance to understand the feature significance in the decision-making process over a broader aspect and hence tackle the local interpretability.

As illustrated in Figure 3, the features signifying academic performance, including GPA, namely the Unweighted and Weighted GPAs, and performance in AP tests such as the AP tests total(relative) scores, average AP scores, and AP CS A score, act as the major key features in influencing the admission decision process in a positive manner, which is marked in green. This suggests that students with higher academic achievement are more likely to be admitted as opposed to the ones with lesser achievement. The performance in the AP tests, as shown to influence positively, showcases student competence in challenging coursework and also reflects student's ability to handle rigorous academic pressure. Similarly, the high school Landscape score also accounts for heavily affecting the decision process in a positive manner. This feature helps the admission decisions to be informed and equitable by considering an applicant's accomplishments within their high school environment mitigating any kind of unfair bias. Moreover, the PIQ2 Syllable ratio also accounts for positively affecting admission decisions, providing insights into the applicant's communication skills, critical thinking ability, and writing.

At the same time, certain features are found to impact the student admission decision process negatively, which are marked in red. Out of all, the Eligibility in Local Context (ELC) percentage score and the Education Preparation Programs (EPP) involvement flag have the highest negative influence on the decision-making process. The ELC score recognizes the achievements of the top 9% of students from each high school in California. For example, students with a value of 1 are the top 1% performers in their high school, while a 9 means that students are in the top 9% of their class. Those who are not in the top 9% did not receive an ELC score and their records were imputed with a value of 10. As a result, the negative LIME score associated with the ELC feature shows that students who had a higher value, meaning that they were further away from the top-performing status, were less likely to be classified as admitted. This ensures a fair evaluation, considering the opportunities and challenges faced by the applicant, and promoting equity in the admissions process.

Continuing on, the EPP flag indicates whether the students participated in any kind of educational preparation programs or not. Given our analysis, students who participated in these programs were less likely to be admitted. Apart from that, the Total Maths score count, PIQ2 and PIQ3 Flesch-reading scores also have a negative influence on the decision-making process. Additionally, we identified a set of features that demonstrate a negligible effect on admission decisions, including the AP statistics scores, the AP Physics C score, and the awards flag. While these features did not exert significant influence individually, they still highly contribute to the overall assessment of an applicant.

Furthermore, the results obtained using the LIME interpretability model in our study are found to be mostly consistent with the feature coefficients (another way to interpret classical ML models) derived in [3]. It further validates the influence and importance of the identified features, reinforcing our model to be robust and reliable in capturing the key factors contributing to the applicant evaluation for the undergraduate admissions process. We note that by highlighting the overall impact of the features, our study emphasizes on a broad spectrum including academic aptitude and holistic qualities while evaluating the applicants for undergraduate student admissions.

V. Conclusion and Future Works

In this paper, we have demonstrated the superior performance showcased by the Deep neural network models in the undergraduate student admission decision-making process. The ICNN model outperforms the remaining baselines by a considerable margin, including the classical models. We have achieved high accuracy in our experiments with both the ICNN and FF neural network models making them suitable choices for the task, further affirming their effectiveness. Furthermore, we have leveraged a gradient-based method coupled with the LIME model to extract significant features responsible for the decision process. The approach allowed us to understand the feature importance and the various types of impact it has on the admission process.

Through our analysis, we have identified certain key features, both positive and negative, that influence the decision-making process, emphasizing the necessity of a holistic evaluation procedure that takes into consideration various factors beyond academic performance. In summary, the application of the ML prediction interpretation technique provides valuable insights that further enable a deeper understanding of the feature significance and contribute to enhancing transparency, fairness, and accountability in the admission process. In future work, we will try to explore anomaly detection to identify unusual patterns in the admission data that would contribute towards developing more robust and qualified admission systems and thereby mitigate any bias or inconsistency in the decision-making process.

References

[1] A. Waters and R. Miikkulainen, "Grade: Machine-learning support for graduate admissions," in *AI Magazine*, vol. 35, no. 1, 2014, pp. 64–75. [Online]. Available: www.aaai.org

[2] J. E. Gilbert, "Applications Quest: Computing Diversity," Auburn University, Tech. Rep., 2006.

[3] B. M. Neda and S. Gago-Masague, "Feasibility of machine learning support for holistic review of undergraduate applications," in *2022 International Conference on Applied Artificial Intelligence (ICAPAI)*. IEEE, 2022, pp. 1–6.

[4] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why should i trust you?" Explaining the predictions of any classifier," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144. [Online]. Available: http://dx.doi.org/10.1145/2939672.2939778

[5] T. Lux, R. Pittman, M. Shende, and A. Shende, "Applications of supervised learning techniques on undergraduate admissions data," *2016 ACM International Conference on Computing Frontiers - Proceedings*, pp. 412–417, may 2016. [Online]. Available: http://dx.doi.org/10.1145/2903150.2911717

[6] C. Andris, D. Cowen, and J. Wittenbach, "Support Vector Machine for Spatial Variation," *Transactions in GIS*, vol. 17, no. 1, pp. 41–61, feb 2013.

[7] "The death and life of an admissions algorithm." [Online]. Available: https://www.insidehighered.com/admissions/article/2020/12/14/u-texas-will-stop-using-controversial-algorithm-evaluate-phd

[8] B. Martinez Neda, Y. Zeng, and S. Gago-Masague, "Using machine learning in admissions: Reducing human and algorithmic bias in the selection process," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 2021, pp. 1323–1323.

[9] A. J. Alvero, N. Arthurs, A. L. Antonio, B. W. Domingue, B. Gebre-Medhin, S. Giebel, and M. L. Stevens, "AI and holistic review: Informing human reading in college admissions," in *AIES 2020 - Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 200–206. [Online]. Available: https://doi.org/10.1145/3375627.3375871

[10] T. Doleck, D. J. Lemay, R. B. Basnet, and P. Bazelais, "Predictive analytics in education: A comparison of deep learning frameworks," *Education and Information Technologies*, vol. 25, no. 3, pp. 1951–1963, may 2020. [Online]. Available: https://doi.org/10.1007/s10639-019-10068-4

[11] M. Vasani, S. Patel, and J. Kaur, "Comparative analysis of baseline models, ensemble models, and deep models for prediction of graduate admission," in *Proceedings of 3rd International Conference on Machine Learning, Advances in Computing, Renewable Energy and Communication: MARC 2021*. Springer, 2022, pp. 515–525.

[12] "Personal insight questions," https://admission.universityofcalifornia.edu/how-to-apply/applying-as-a-freshman/personal-insight-questions.html.

[13] "Simplified text processing." [Online]. Available: https://textblob.readthedocs.io/en/dev/

[14] "Textstat." [Online]. Available: https://textstat.readthedocs.io/en/latest/

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[16] B. Amos, L. Xu, and J. Z. Kolter, "Input convex neural networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 146–155.

[17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.