CSCE 590-1: Trusted AI Course Project Report

Predicting and Explaining Graduate Admissions

Vishal Pallagani[1]

[1]University of South Carolina

December 2, 2021

1 Problem Statement

The process of US graduate admissions for Indian students is usually carried out by consultancy companies, which are often biased towards a limited set of universities. Thus, the research problem being solved in this work is predicting the chance of admission to a university and also explaining the prediction. A comparison of 5 different AI methods have been carried out in this work and the predictions are explained by 2 methods - LIME and ProtoDash.

2 Related Work

In [1, 2], the authors explore the use of different machine learning approaches to predict the chance of admission for post-graduate degree in US. However, both the paper doesn't talk about providing explanations to the end user. The authors in [3] make use of a deep learning network to predict the chance of admission for US universities. Deep learning network, being a black box model, makes it even tougher to explain its predictions, thereby, leading to a lack of trust on the system. I could not find existing literature that looked into predicting the chance of admission in other countries apart from the US.

3 End User

The intended end user of this system is a student from India who is aspiring to get into a graduate program in the United States.

4 Data

The dataset used for this work is taken from [4]. The dataset initially consisted of 300 rows, but I have extended them to 500 by randomly sampling the existing data entries. Figure 1 shows the snapshot of the dataset along with the features.

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65

Figure 1: Snapshot of the dataset

5 AI Methods

The AI methods used for predicting the chance of admission given the features (GRE score, TOEFL score, university rating, SOP strength, LOR strength and prior research experience) are as follows:

- Linear Regression
- Support Vector Machine
- Gradient Boosting Regression
- Random Forest Regression
- K-Nearest Neighbors Regression

The performance comparison of these 5 difference models is shown in Figure 2. In addition to using a regression model to predict the chance of admission, two explainability methods, namely, LIME and Protodash are used to explain the model predictions.

6 Evaluation

The evaluation metric followed for this approach is mean absolute error.

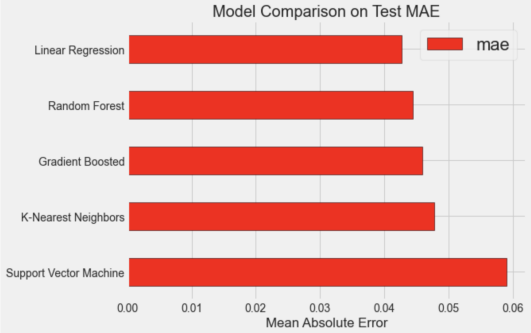

Figure 2: Comparison of model performances

Chance	1.000000
CGPA	0.882413
GRE	0.810351
TOEFL	0.792228
UniversityRating	0.690132
SOP	0.684137
LOR	0.645365
Research	0.545871

Figure 3: Correlation between features and the target variable.

7 Human-AI Issue

The trust issue is being tackled here by providing explanations regarding the prediction to the user. Figure 3 shows the correlation between different features and the target variable, thereby setting a foundation for understanding the explanations provided by LIME and Protodash. It can be seen here that CGPA, GRE and TOEFL play an important role in determining the chance of admission. Also, having a prior research experience can benefit the students by increasing their chances of admission.

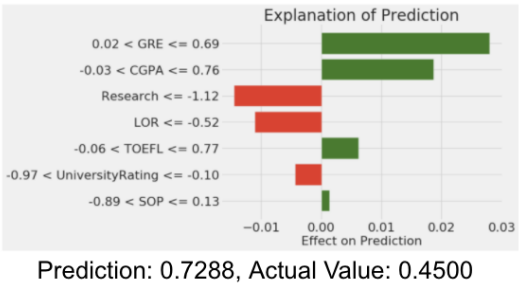

Figure 4: Explanation provided by LIME for incorrect prediction

Figure 4 shows the explanation provided by LIME for an incorrect prediction. It can be seen that the model is predicting incorrectly because the student has high GRE score and CGPA. On the other hand, Figure 5

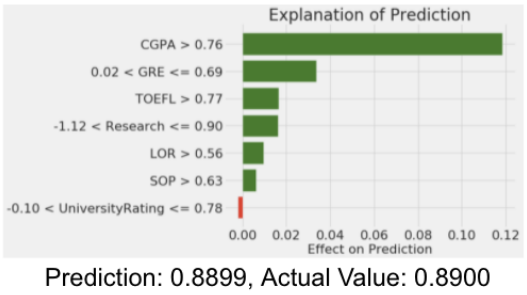

Figure 5: Explanation provided by LIME for correct prediction

shows the explanation provided by LIME for a correct prediction. The prototypes that represent the dataset provided by Protodash are shown in Figure 6, which tally with the conclusions we have arrived from the correlation scores between features and the target variable.

	Serial No.	GRE Score	TOEFL Score	University Rating	...	CGPA	Research	Chance of Admit	Weights of Prototypes
371	372	324	110	3	...	9.22	1	0.89	0.11
36	37	299	106	2	...	8.40	0	0.64	0.10
89	90	316	109	4	...	8.76	1	0.74	0.11
421	422	321	112	3	...	8.95	1	0.77	0.10
358	359	314	105	2	...	7.64	0	0.70	0.11
483	484	304	103	5	...	7.92	0	0.71	0.11
481	482	323	107	4	...	8.48	1	0.78	0.10
447	448	320	108	3	...	8.97	1	0.84	0.09
256	257	309	99	3	...	8.56	0	0.76	0.10
419	420	308	102	2	...	7.98	1	0.58	0.08

Figure 6: Top 10 prototypes provided by Protodash

8 Conclusion

In this project, I have explored and compared the use of various regression models for predicting the chance of admission to a given university. In addition to this, my major learning in this course project is hands-on experience of using explainability frameworks such as Protodash (AIX360) and LIME. The code is located at: https://github.com/VishalPallagani/CSCE-590-1/blob/main/Course%20Project/code/GraduateAdmission_Prediction.ipynb

References

[1] Amal AlGhamdi, Amal Barsheed, Hanadi AlMshjary, and Hanan AlGhamdi. A machine learning approach for graduate admission prediction. In *Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing*, pages 155–158, 2020.

[2] S Sujay. Supervised machine learning modelling & analysis for graduate admission prediction. *Int. J. Trend Res. Dev*, 7(4):5–7, 2020.

[3] Md Omaer Faruq Goni, Abdul Matin, Tonmoy Hasan, Md Abu Ismail Siddique, Oishi Jyoti, and Fahim MD Sifnatul Hasnain. Graduate admission chance prediction using deep neural network. In *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, pages 259–262. IEEE, 2020.

[4] Mohan S Acharya, Asfia Armaan, and Aneeta S Antony. A comparison of regression models for prediction of graduate admissions. In *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, pages 1–5. IEEE, 2019.