**Program: DAB-402 Capstone Project Winter 2023-003**

**Group: - 5**

**Project Title:**

**Develop AI and ML frameworks/models to automate the university application process.**

**Group members (Name, ID):**

* Tanmay Daulatwal (0798588)
* Niren Patel (0794177)
* Shashank Tripathi (0789803)
* Nitin Singh Rana (0770080)

**Problem Statement:**

* One of a university's most crucial operations is the admission of new students since they cannot function without them. Due to mistakes or excessively long response times, a poor admissions process may result in fewer students being admitted into a college.
* Universities are increasingly struggling to handle the admission process effectively due to the advancement of technology and rise in student enrolment. The traditional admission process requires a lot of manual effort and is time consuming. This is where machine learning (ML) and artificial intelligence (AI) may be used to automate the admissions process and make it more effective.
* To create an AI and ML-based system that can automate the university application process, including challenges like gathering and processing applicant information, producing individualized recommendations for colleges and programs, and managing communication with applicants and universities.
* A lot of high-quality data are needed to build an accurate ML model, yet this data may not always be available or may not be accurate. An appropriate prediction may be difficult to train as a result. When a model is overfitted to the training set of data, it might perform poorly on untested, fresh data. When creating a model for an automated admission process, this might be a concern because the model might not be able to generalize properly to new applicants.
* The model must be able to manage the increasing volume of data and make recommendations quickly as the number of applications rises. The model has to be connected with other platforms, such as learning management systems (LMS) and student information systems (SIS)

**Solving the Problem:**

* The model may be trained using high-quality data; therefore, work can be done to collect additional data from a variety of sources, clean the data, and preprocess it to remove mistakes and missing values.
* Distributed computing and cloud-based solutions may be utilized to make sure the model can manage the increasing volume of data and make conclusions quickly.

**Research Question:**

1. What data about the university must be gathered, such as the courses / programs, admission criteria, etc.
2. Which AI and ML framework can be applied to automate the university admissions process?

**Knowledge, skills, and technology needed for this solution:**

* The analyst will need to be knowledgeable about Techniques utilizing artificial intelligence (AI) and machine learning (ML), such as computer vision, natural language processing, and predictive modelling.
* Also, familiarity with programming languages including Python, Java, and R, as well as NumPy, Pandas, and Matplotlib are data science tools.
* Knowledge of the university application procedure and awareness of the requirements of the universities, applicants.
* A strong analytical and problem-solving ability, excellent teamwork and communication skills are also important.

**Evaluating Solution:**

Currently the stakeholders do not have any existing solutions or any previous attempt. So, we have to create from scratch.

**Metrics Available:**

* Classification Metrics (accuracy, precision, recall, F1-score, ROC, AUC)
* Regression Metrics (MSE, MAE)
* Ranking Metrics (MRR, DCG, NDCG)
* Statistical Metrics (Correlation, RMSE, Z-score)

**Success Measure:**

There are several metrics that can be used to measure the success of a machine learning model for an admission process. Some commonly used metrics include:

1. Accuracy: The proportion of correct predictions made by the model. It represents the proportion of accurate forecasts to all predictions.
2. Precision: The proportion of true positive predictions among all positive predictions made by the model. It is a measure of the model's ability to avoid false positives.
3. Recall: The percentage of accurate forecasts among all instances of good outcomes. It is a measure of the model's ability to identify all positive cases.
4. F1 Score: The harmonic mean of precision and recall. It is a balance between precision and recall and provides a single number that can be used to compare models.
5. AUC-ROC: AUC-ROC curve is a performance measurement for classification problem at various thresholds settings. ROC is a probability curve and AUC represents degree or measure of separability. It reveals how well the model can discriminate between classes.
6. You can also use Cross validation to measure the success of the model, this will give you a better understanding of the model's generalization error.

**Evaluating model & feasibility of Model:**

For evaluating the model or solution we need to have a great dataset which includes data from different colleges and universities. We need to have admission selection criteria in the dataset so that we can create a feature for our model. The selection criteria, such as GRE score, TOEFL, IELTS, Grades, Experience, certificates and many more.

This solution can be feasible if we get accurate data of different colleges and universities.

The Feasibility of the solution depends on the accurate features of the model and the training of the model with cleaned and relatable data.

**Impact on Stakeholders:**

The impact would be good on the stakeholders because they will be providing this solution to the colleges and universities so that the manual work and delay in the response or admission can be eliminated from the educational institute.

**Ethical Concerns:**

For right now we do not have ethical concerns.

**Data Source:**

General the Data source links are of educational institute and National Center for Education Statistics (NCES).

**References:**

[1] "Admissions Requirements," Berkeley Graduate Division, [Online]. Available: <https://grad.berkeley.edu/admissions/requirements/>

[2] I. Hatzilygeroudis, "PASS: An Expert System with Certainty Factors for Predicting Student Success," ResearchGate, 20 September 2004. [Online]. Available: [https://www.researchgate.net/publication/221017874\_PASS\_ An\_Expert\_System\_with\_Certainty\_Factors\_for\_Predicting\_ Student\_Success](https://www.researchgate.net/publication/221017874_PASS_%20An_Expert_System_with_Certainty_Factors_for_Predicting_%20Student_Success)

[3] E. Roberts, "using machine learning and predictive modeling to assess admission policies and standards," 2013. [Online]. Available: <http://sites.tntech.edu/weberle/wpcontent/uploads/sites/87/2018/06/NSSR_2013-1.pdf>.

[4] M. M. P. K. P. S. J. Heena Sabnani, "Prediction of Student Enrolment Using Data Mining Techniques," April 2018. [Online]. Available: <https://www.irjet.net/archives/V5/i4/IRJET-V5I4408.pdf>.

[5] J. Brownlee, "Linear Regression for Machine Learning," [Online]. Available: <https://machinelearningmastery.com/linear-regression-formachine-learning/>.

[6] A. Géron, Hands on Machine Learning with Scikit Learn and Tensorflow, O'Reilly Media, 2017.

[7] R. S. Brid, "Decision Trees," Greyatom, 26 Oct 2018. [Online]. Available: <https://medium.com/greyatom/decisiontrees-a-simple-way-to-visualize-a-decision-dc506a403aeb>.

[8] mohansacharya, "Graduate Admissions 2," kaggle, 28 12 2018. [Online]. Available: <https://www.kaggle.com/mohansacharya/graduate-admissions>.