GRADUATION PREDICTION MODEL

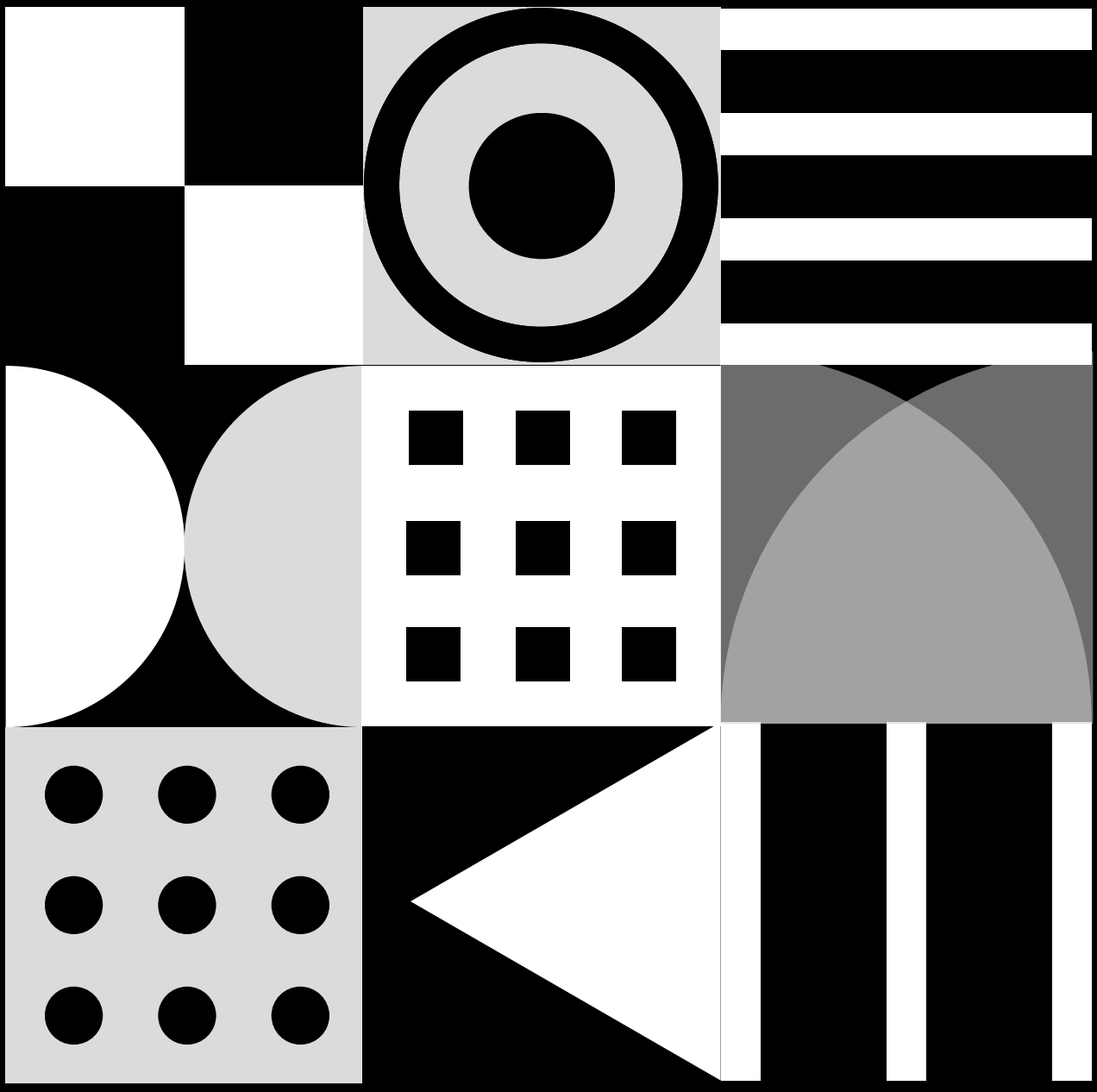


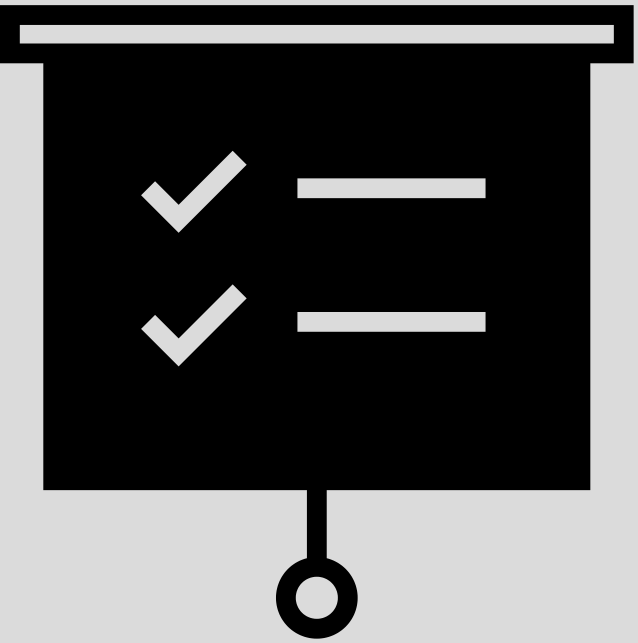
Table of Contents

Introduction

Based on Indian standards these are the parameters needed:
GRE Scores (out of 340)
TOEFL Scores (out of 120)
University Rating (out of 5)
Statement of Purpose and Letter of Recommendation Strength (out of 5)
Undergraduate GPA (out of 10)
Research Experience (either 0 or 1)
Chance of Admit (ranging from 0 to 1)

A model is needed to predict the likelihood of a student joining a university based on the parameters above.

VISION



The model is supposed to gauge on a scale between 0 and 1 (0 being no chance to join a university while 1 being most certain to join a university).

This should be on the basis of the Indian standards of studies mentioned below:

GRE Scores (out of 340)
TOEFL Scores (out of 120)
University Rating (out of 5)
Statement of Purpose and Letter of Recommendation Strength (out of 5)
Undergraduate GPA (out of 10)
Research Experience (either 0 or 1)
Chance of Admit (ranging from 0 to 1)

SOLUTION

The model used : Linear Regression. (python)

First data was preprocessed by removing all duplicates, missing data, null values. Renaming of columns was also helpful.

Outliers were of essence and therefore were not removed from the dataset. They were valid data and within the set parameters

Univariate analysis was necessary so as to evaluate individual columns, as well as scatter plots for Bivariate analysis.

This provided a visual representation of data to draw more conclusions.

- The relevant columns for independent variables (features) are selected and stored in the variable 'X.' These columns include 'GRE Score,' 'TOEFL Score,' 'University Rating,' 'SOP,' 'LOR,' 'CGPA,' and 'Research.'
- The target variable is selected and stored in the variable 'y.' In this case, it is assumed to be the 'Chance of Admit' column from the 'admit' DataFrame.
- The data is split into training and testing sets using the 'train_test_split' function from scikit-learn.
- A linear regression model is instantiated using the 'LinearRegression' class from scikit-learn and stored in the variable 'regressor.'

The model is trained on the training data using the 'fit' method of the 'regressor' object

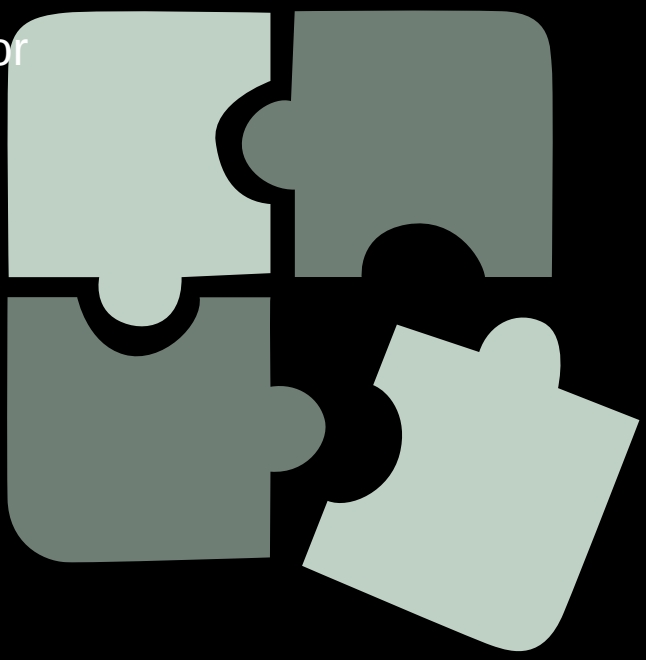
OVERVIEW

With an MSE of 0.0037046553987884 05, the model's performance can be considered relatively good.

The small MSE value suggests that, on average, the predicted chances of admission are very close to the actual values in the testing set.

CHALLENGES

1. Limited availability or quality issues with the dataset used for training and testing the model can affect the model's performance and generalization.
2. The model may suffer from overfitting, where it performs well on the training data but poorly on unseen data, or underfitting, where the model fails to capture the underlying patterns in the data.
3. Ensuring that the trained model can generalize well to new, unseen data is a crucial challenge in machine learning.
4. Linear regression models are relatively interpretable, but as the complexity of models increases, interpretability and explainability become more challenging.



FUTURE DIRECTION

The model can be further refined by exploring alternative regression algorithms or advanced machine learning techniques to potentially enhance its predictive performance.

Fine-tuning hyperparameters or investigating different regularization techniques could help optimize the model's performance and mitigate issues such as overfitting or underfitting.

Conducting a thorough feature selection process or exploring feature engineering methods may uncover additional informative features to improve the model's accuracy.

REFERENCES

Mohan S Acharya, Asfia Armaan, Aneeta S Antony : A Comparison of Regression Models for Prediction of Graduate Admissions, IEEE International Conference on Computational Intelligence in Data Science 2019