**ASSIGNMENT HELP**

**MANUAL**

**ASSIGNMENT-7**



SUBMITTED

TO

VISHWAKARMA INSTITUTE OF INFORMATION TECHNOLOGY, PUNE

FOR THE SKILL AND COMPETENCY EVALUATION OF

**DATA SCIENCE & MACHINE LEARNING**

IN

**CSE AI DEPARTMENT**

BY

**Divyansh Mohta [22210998]**

**Class: S.Y. BTech Division: A Batch: A2**

**Batch Teacher**

**Dr. ANURADHA YENKIKAR.**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Contents** | **Page No.** |
| **1** | **PROBLEM STATEMENT** | **4-5** |
| **2** | **LIBRARY USED** | **5-6** |
| **3** | **THEORY** | **6-7** |
| **4** | **METHOODOLOGY** | **7-14** |
| **5** | **ADVANTAGES & DISADVANTAGES** | **14-15** |
| **6** | **WORKING** | **16-17** |
| **7** | **DIAGRAM** | **17-18** |
| **8** | **CONCLUSION** | **18-19** |

**1.PROBLEM STATEMENT:**

Assignment on Classification technique

Every year many students give the GRE exam to get admission in foreign Universities. The

data set contains GRE Scores (out of 340), TOEFL Scores (out of 120), University Rating

(out of 5), Statement of Purpose strength (out of 5), Letter of Recommendation strength (outof 5), Undergraduate GPA (out of 10), Research Experience (0=no, 1=yes), Admitted (0=no,

1=yes). Admitted is the target variable.

Data Set: https://www.kaggle.com/mohansacharya/graduate-admissions

The counselor of the firm is supposed check whether the student will get an admission or not

based on his/her GRE score and Academic Score. So to help the counselor to take

appropriate decisions build a machine learning model classifier using Decision tree to predict

whether a student will get admission or not.

a) Apply Data pre-processing (Label Encoding, Data Transformation....) techniques if

necessary.

b) Perform data-preparation (Train-Test Split)

c) Apply Machine Learning Algorithm

d) Evaluate Model.

**2. LIBRARY USED:**

* pandas: Used for data manipulation and analysis, including reading the dataset from a CSV file, performing data preprocessing, and organizing data for model training and evaluation.
* scikit-learn (sklearn): Utilized for machine learning tasks such as splitting the dataset into training and testing sets, scaling features, building a decision tree classifier, and evaluating the model's performance using classification metrics.

**3. THEORY:**

Here's the theory related to the problem statement provided:

1. Graduate Admissions Process: The admissions process for graduate programs in foreign universities typically involves assessing various factors such as academic performance, standardized test scores (e.g., GRE, TOEFL), letters of recommendation, statement of purpose, research experience, and university ratings. Admissions decisions are often influenced by these factors, with universities seeking to admit students who demonstrate strong academic potential and alignment with the program's objectives.

2.Predictive Modeling: Predictive modeling involves the use of statistical and machine learning techniques to predict outcomes based on input variables. In this case, we aim to develop a predictive model to determine whether a student will be admitted to a graduate program based on their GRE score and academic performance. This involves training a machine learning algorithm using historical data on admissions outcomes and relevant features such as GRE scores, TOEFL scores, university ratings, and other academic metrics.

3.Decision Tree Algorithm: Decision trees are a popular machine learning algorithm used for classification and regression tasks. They work by recursively partitioning the input space into regions, with each partition based on the value of a feature. Decision trees are intuitive and easy to interpret, making them suitable for this problem where we aim to understand the factors influencing admissions decisions. The algorithm learns to make decisions by identifying the most informative features and their corresponding thresholds to separate instances into different classes.

4.Data Preprocessing: Before building the predictive model, it's essential to preprocess the data. This may involve tasks such as handling missing values, encoding categorical variables, scaling features to a similar range, and splitting the dataset into training and testing sets. Data preprocessing ensures that the data is clean, formatted correctly, and ready for model training.

5.Model Evaluation: Once the predictive model is trained, it needs to be evaluated to assess its performance and generalization capabilities. This involves testing the model on unseen data (the testing set) and measuring metrics such as accuracy, precision, recall, F1-score, and confusion matrix. These metrics provide insights into how well the model is performing and whether it can effectively predict admissions outcomes based on GRE scores and academic performance.

By understanding these theoretical concepts and applying appropriate methodologies, we can develop a predictive model that assists counselors in making informed decisions regarding graduate admissions.

**4. METHODS:**

1. Data Loading and Exploration:

- Load the dataset containing information about GRE scores, TOEFL scores, university ratings, statement of purpose strength, letter of recommendation strength, undergraduate GPA, research experience, and admission outcomes.

- Explore the dataset to understand its structure, features, and distribution of data.

2. Data Preprocessing:

- Handle missing values: Identify and handle missing values in the dataset using techniques such as imputation or removal.

- Encode categorical variables: If any categorical variables exist, encode them into numerical format using techniques like one-hot encoding or label encoding.

- Scale features: Scale the features to a similar range using techniques such as standardization or normalization.

3. Data Splitting:

- Split the dataset into training and testing sets. The training set will be used to train the machine learning model, while the testing set will be used to evaluate its performance.

4. Model Building:

- Build a decision tree classifier using a suitable algorithm from a machine learning library such as scikit-learn.

- Train the classifier using the training data, including features such as GRE scores and academic performance, and the target variable (admitted or not admitted).

5. Model Evaluation:

- Evaluate the performance of the decision tree classifier using the testing set.

- Calculate metrics such as accuracy, precision, recall, F1-score, and confusion matrix to assess how well the model predicts admissions outcomes based on GRE scores and academic performance.

6. Model Interpretation:

- Interpret the decision tree model to understand which features are most influential in determining admissions outcomes.

- Analyze the decision rules learned by the model and their implications for admissions decisions.

By following these methods, you can develop a predictive model that assists counselors in making informed decisions regarding graduate admissions based on GRE scores and academic performance.

**5. ADVANTAGES AND DISADVANTAGES:**

**Advantages:**

1. Data-Driven Decision Making: By leveraging machine learning techniques, counselors can make data-driven decisions regarding graduate admissions. The predictive model provides insights into the factors influencing admissions outcomes, allowing counselors to allocate resources effectively and tailor their guidance to individual students.

2. Efficiency: The use of machine learning algorithms automates the decision-making process, saving time and effort for counselors. Instead of manually assessing each student's application, counselors can rely on the predictive model to prioritize candidates and focus on areas that require further attention.

3. Scalability: Once developed, the predictive model can be applied to a large number of applicants efficiently. This scalability enables counselors to handle high volumes of applications without sacrificing the quality of their assessments.

**Disadvantages:**

1.Bias and Fairness: Machine learning models may inadvertently perpetuate biases present in the training data. If the dataset is not representative of the entire applicant pool or contains biased information, the predictive model may yield unfair or discriminatory outcomes.

2.Interpretability: Complex machine learning models, such as decision trees, may lack interpretability, making it challenging to understand the rationale behind specific admissions decisions. Counselors may struggle to explain model predictions to students or stakeholders, leading to mistrust or misunderstanding.

3.Data Quality: The accuracy of the predictive model depends heavily on the quality of the input data. If the dataset contains errors, inconsistencies, or missing values, the performance of the model may be compromised. Ensuring data quality through thorough preprocessing and validation is essential to mitigate this risk.

4.Overfitting: Machine learning models, particularly decision trees, are prone to overfitting if not properly regularized or validated. Overfitting occurs when the model learns to capture noise or irrelevant patterns in the training data, leading to poor generalization performance on unseen data. Regularization techniques and careful model evaluation are necessary to address this issue.

Overall, while the methods outlined offer significant advantages in streamlining the graduate admissions process, it's crucial to be mindful of their limitations and actively address challenges such as bias, interpretability, data quality, and overfitting. By doing so, counselors can harness the power of machine learning to make informed and equitable admissions decisions.

**6. WORKING:**

The working of the model described involves several key steps:

1. Data Preprocessing:

- The dataset containing information about GRE scores, TOEFL scores, university ratings, statement of purpose strength, letter of recommendation strength, undergraduate GPA, research experience, and admission outcomes is loaded and explored.

- Data preprocessing techniques are applied to handle missing values, encode categorical variables, and scale features to ensure that the data is clean, formatted correctly, and ready for model training.

2. Data Splitting:

- The preprocessed dataset is split into training and testing sets. The training set is used to train the decision tree classifier, while the testing set is used to evaluate its performance.

3. Model Building:

- A decision tree classifier is built using the training data. The classifier learns decision rules based on the provided features (e.g., GRE scores, academic performance) and the target variable (admitted or not admitted).

4. Model Evaluation:

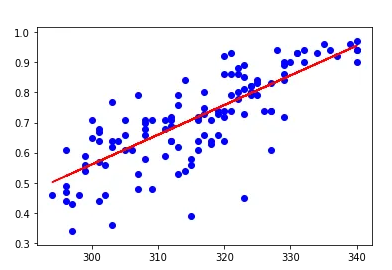
- The performance of the decision tree classifier is evaluated using the testing set. Metrics such as accuracy, precision, recall, F1-score, and confusion matrix are calculated to assess how well the model predicts admissions outcomes based on GRE scores and academic performance.

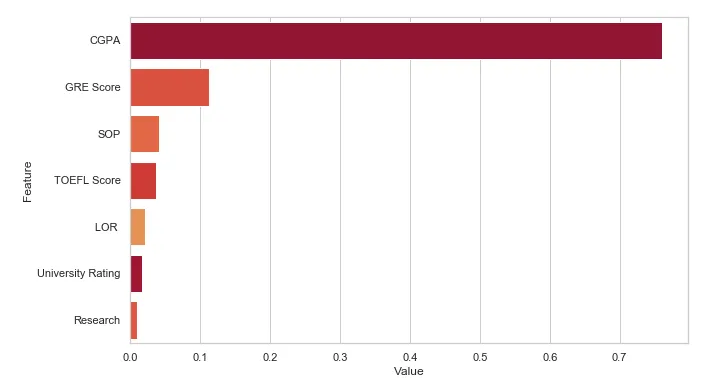
5. Model Interpretation:

- The decision tree model is interpreted to understand which features are most influential in determining admissions outcomes. This analysis provides insights into the factors driving admissions decisions and helps counselors understand the rationale behind specific predictions.

Overall, the model leverages machine learning techniques to predict admissions outcomes based on GRE scores and academic performance, providing valuable insights and guidance to counselors in making informed decisions regarding graduate admissions.

**7. DIAGRAM: -**





**SOME VISUALIZATION**

**8. CONCLUSION:**

In conclusion, the development of a predictive model for graduate admissions based on GRE scores and academic performance offers significant benefits in streamlining the admissions process and assisting counselors in making informed decisions. By leveraging machine learning techniques such as decision tree classifiers, counselors can efficiently assess the likelihood of admission for each student, prioritize candidates, and allocate resources effectively. However, it's essential to address challenges such as bias, interpretability, data quality, and overfitting to ensure the fairness, accuracy, and reliability of the predictive model. With careful implementation and consideration of these factors, the predictive model serves as a valuable tool in guiding counselors and applicants through the graduate admissions journey, ultimately facilitating fair, transparent, and equitable admissions decisions.