BrightLearn Machine Learning Camp

Read the project scenario below and answer ALL the questions that follow.

SCENARIO

A healthcare startup is interested in developing a machine learning model to predict patient readmission rates based on various patient data points collected from their hospital visits. The company aims to utilise Python's extensive libraries and machine learning frameworks to analyse the provided historical patient data and build a predictive model that can help hospitals reduce unnecessary readmissions, thus saving costs and improving patient care.

Your task is to develop a predictive model using Python to determine the likelihood of patient readmissions within 30 days of discharge based on the provided dataset. This project will involve data preprocessing, exploratory data analysis, feature selection, model building, and model evaluation.

SOURCE:

- The dataset provided includes various patient data points such as age, gender, type of admission, length of stay, number of diagnoses, blood pressure, blood sugar levels, previous admissions, and readmission status.

YOUR REPORT SHOULD FOCUS ON THE FOLLOWING:

- Data Preprocessing and Cleanup

- Exploratory Data Analysis and Insights

- Feature Selection and Justification

- Model Development and Selection

- Model Evaluation and Testing

Main Question

Develop a predictive model using Python to assess the likelihood of patient readmissions within 30 days of discharge. Your work should leverage Python's data science libraries to process and analyse the data, build a predictive model, and evaluate its effectiveness.

Sub-Questions (100 Marks Total):

1. Data Preprocessing (20 Marks): Clean the dataset by handling missing values, normalising data, and encoding categorical variables using Python libraries such as Pandas and NumPy.
2. Exploratory Data Analysis (20 Marks): Conduct an exploratory analysis to understand the distribution of key variables and their relationships using Matplotlib and Seaborn.

3. Feature Selection (20 Marks): Identify and select the most significant predictors for patient readmissions. Justify your selection of features.

4. Model Building (20 Marks): Construct a machine learning model using Scikit-learn. Evaluate and select appropriate algorithms like logistic regression, decision trees, or a random forest classifier.

5. Model Evaluation and Testing (10 Marks): Evaluate your model's performance using appropriate metrics (accuracy, precision, recall, F1 Score, ROC-AUC). Discuss any potential overfitting and strategies to mitigate it.

FORMAT

Criteria	Marks	Details
Title Page	2	Properly formatted title page with project title, student's name, and date.
Table of Contents	2	Clear and well-organized table of contents.
Introduction	2	Brief introduction of the project scope and objectives.
Data Preprocessing	20	Comprehensive cleaning and preprocessing (handling missing data, normalisation, encoding).
Exploratory Data Analysis	20	Detailed analysis with visualisations and interpretations of key variables and relationships.
Feature Selection	20	Logical selection and justification of features based on statistical tests and data insights.

Model Development	20	Application of appropriate machine learning models, including model tuning and optimisation.
Model Evaluation	10	Thorough evaluation using appropriate metrics (e.g., accuracy, ROC-AUC). Discussion of model performance and limitations.
Conclusion	2	A concise summary of findings, implications, and potential improvements.
List of References	2	Correctly formatted references and adequate sourcing of external data and literature.
Total	100	

END OF PAPER