

Naive Bayes Classifier Assignment

Problem Statement:

The objective of this assignment is to implement a Naive Bayes classifier to predict target classes based on historical data. By exploring the datasets and employing the Naive Bayes algorithm, students are expected to generate accurate predictions and evaluate the model's performance.

Guidelines:

1. Foundational Knowledge:
 - Understand the principles of classification modeling and the components of the Naive Bayes algorithm.
 - Familiarize yourself with the assumption of independence among features in Naive Bayes.
 - Recognize the importance of probability theory in Naive Bayes classification.
2. Data Exploration:
 - Analyze the dataset's structure and characteristics.
 - Explore features' distributions and relationships with the target variable.
 - Identify any potential challenges or biases in the dataset.
3. Preprocessing and Feature Engineering:
 - Handle missing values appropriately.
 - Encode categorical variables if necessary.
 - Perform feature scaling or normalization if needed.
4. Model Building and Hyper-Parameter Tuning:
 - Split the dataset into training and testing sets.
 - Initialize a Naive Bayes classifier model.
 - Perform hyper-parameter tuning.
5. Model Training and Evaluation:
 - Train the model using the training set.
 - Evaluate the model's performance using metrics such as accuracy, precision, recall and F1-score.
6. Interpretation and Conclusion:
 - Interpret the model's predictions and analyze any observed trends or patterns.
 - Discuss the strengths and limitations of the Naive Bayes classifier model for the given dataset.
 - Propose potential improvements or alternative modeling techniques if applicable.

Step-by-Step Approach to Naive Bayes Classifier:

1. Setup and Data Preparation:
 - Import necessary libraries: pandas, matplotlib, scikit-learn.
 - Load the dataset for classification analysis.
 - Preprocess the data, handle missing values, encode categorical variables, and perform feature scaling if necessary.

2. Model Training:

- Split the dataset into training and testing sets.
- Initialize a Naive Bayes classifier model.
- Train the model using the training set.

3. Model Evaluation:

- Evaluate the model's performance on the testing set using evaluation metrics (accuracy, precision, recall, F1-score).

4. Interpretation and Conclusion:

- Interpret the model's predictions in the context of the dataset.
- Discuss the implications of the findings and potential next steps for improvement.

Links to Datasets for the Assignment:

- Mushroom Dataset (Binary Classification) Dataset

[<https://www.kaggle.com/datasets/prishasawhney/mushroom-dataset/data>]

- Credit Score Classification Dataset

[<https://www.kaggle.com/datasets/sujithmandala/credit-score-classification-dataset/data>]

- Simple Rainfall Classification Dataset

[<https://www.kaggle.com/datasets/sujithmandala/simple-rainfall-classification-dataset/data>]