

ML CW PROCEDURE

1. Data Understanding and Preparation:

- Download the dataset from the provided link and explore its meta information, class distribution, attributes, and statistics.
- Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features if necessary.
- Split the dataset into training and testing sets to evaluate the models' performance.

2. Model Implementation:

a. Naïve Bayes:

- Use libraries like scikit-learn in Python to implement Naïve Bayes classifier (e.g., Gaussian Naïve Bayes or Multinomial Naïve Bayes).
- Train the Naïve Bayes classifier on the training data.
- Evaluate the model's performance using appropriate evaluation metrics (e.g., accuracy, precision, recall, F1-score, ROC-AUC).

b. Random Forest Classification:

- Utilize the RandomForestClassifier from scikit-learn to implement Random Forest Classification.
- Train the Random Forest classifier on the training data.
- Evaluate the model's performance using the same evaluation metrics as Naïve Bayes.

3. Model Comparison:

- Compare the performance of the Naïve Bayes and Random Forest classifiers based on the chosen evaluation metrics.
- Choose the model that performs better on the test dataset.

4. Experimental Results Showcase:

- Present the experimental results, including accuracy, precision, recall, F1-score, ROC-AUC, and any other relevant metrics.
- Visualize the performance metrics if necessary using plots or tables.

- Discuss any observations or insights gained from the experiments.

5. Optimization and Further Analysis (Optional):

- Experiment with hyperparameter tuning for both models to further improve their performance.
- Explore feature importance analysis for Random Forest to understand which features are most influential in predicting income levels.

6. Documentation and Reporting:

- Document the entire process, including data preprocessing steps, model implementation details, evaluation metrics, and results.
- Prepare a report summarizing the findings, methodology, and conclusions.

By following these steps, you should be able to successfully complete the task of predicting income levels using Naïve Bayes and Random Forest Classification and compare their performances. Make sure to adhere to best practices in machine learning, such as proper data preprocessing, model evaluation, and result interpretation.
