TreeLeaf.Ai

A brief Report on Analysis of the Tasks

2023

ML internship hiring exam

Name: **Bimal Yadav**
Contact No: **9813585735**
Email: **np01cp4a210010@islingtoncollege.edu.np**

Submitted to : **TreeLeaf.Ai**
https://treeleaf.ai/

Table of Contents

Task 1 : Bank loan classification

In this analysis, we addressed the task of building a machine learning model for bank loan classification based on a provided dataset. The dataset contained information about customers' attributes, such as age, gender, income, education level, and more, along with a target variable indicating whether a personal loan was accepted. Our approach involved several key steps, which led to valuable insights and findings.

Approach Overview

Data Preprocessing and EDA:

1. Loaded the dataset and conducted initial data exploration.
2. Handled missing values in columns like 'Home Ownership' and 'Gender' using appropriate imputation methods.
3. Transformed categorical variables like 'Gender' into numeric format for model compatibility.

Feature Engineering and Data Splitting:

1. Created necessary features for analysis and modeling.
2. Split the dataset into training and testing sets to ensure unbiased model evaluation.

Model Selection and Training:

1. Evaluated multiple classification algorithms, including Logistic Regression and k-Nearest Neighbors (k-NN).
2. Scaled the data to improve the performance of algorithms that are sensitive to feature scaling.
3. Observed that k-NN with scaled data outperformed other algorithms in terms of accuracy.

Model Evaluation and Findings:

1. Assessed model performance using metrics such as accuracy, precision, recall, F1-score.
2. Noted that k-NN with scaled data achieved an accuracy of 94%, making it the best-performing model for this task.

Key Insights and Observations

1. Logistic Regression provided reasonable accuracy, but k-NN demonstrated superior performance.
2. The choice of scaling the data significantly impacted the accuracy of k-NN, highlighting the importance of preprocessing steps.
3. Feature engineering and appropriate handling of missing values improved the overall quality of the dataset and the models.

Conclusion

In conclusion, our analysis showcased the importance of data preprocessing, feature engineering, and model selection in building a successful machine learning model. The k-Nearest Neighbors algorithm with the **Accuracy 94%**, when applied to scaled data, emerged as the most accurate choice for classifying personal loan acceptance. The findings of this analysis can guide future decision-making processes and contribute to more accurate predictions in the domain of bank loan classification.

Task 2 : Rectangle numbering and Alignment:

Task 1: Assigning Numbers to Rectangles Based on Length

Approach: For Task 1, the goal was to assign numbers (1 to 4) to rectangles based on their lengths inside an image. Initially, a traditional approach using edge detection and contour analysis was considered. However, due to the complexity of the images and the presence of various shapes, this method did not yield the desired results.

Insights:

1. Traditional computer vision methods can struggle with complex scenes and varied shapes.
2. Complex images require advanced techniques or machine learning-based approaches.
3. More robust solutions, like deep learning-based object detection, may provide better results.

Task 2: Aligning and Displaying Aligned Rectangles

Approach: For Task 2, the objective was to align the rectangle images and display them in separate windows. The code aimed to detect rectangles, identify their orientations, and align them properly before display.

Insights:

1. Alignment requires understanding the orientations of the rectangles.
2. Manual intervention might be necessary for accurate alignment.
3. Advanced techniques, such as deep learning models or manual annotation, could provide better results for challenging cases.

Overall Insights and Recommendations:

1. Complex tasks involving image analysis often require more advanced methods, such as deep learning-based models like YOLO or SSD.
2. Experimentation with parameters is crucial to fine-tune traditional methods.
3. Task 1 and Task 2 highlight the challenges in handling images with variations and complex scenes.

4. Real-world scenarios may require a combination of approaches, including manual intervention and algorithmic processing, to achieve desired outcomes.

Conclusion

In conclusion, both Task 1 and Task 2 demonstrated that image analysis can be complex, requiring a combination of techniques such as OpenCV, NumPy , parameter tuning, and potentially manual intervention to achieve accurate and desired results. The choice of approach depends on the complexity of the images and the level of precision required for the task.