

Census Dataset Exploration and Modeling Report

1. Introduction

This report details the exploration, preprocessing, and modeling activities performed on the 'Census.csv' dataset...

2. Dataset Overview

The 'Census.csv' dataset comprises 32,561 entries and 15 features, detailing individual demographic and employment information...

3. Data Preprocessing

Missing values in 'workclass', 'occupation', and 'native-country' were replaced with 'Unknown'...

4. Exploratory Data Analysis (EDA)

EDA revealed significant variables influencing income, including age, education, and hours per week worked...

5. Model Building and Evaluation

Initial models tested were Logistic Regression, Random Forest, and SVM. The Random Forest model showed promising results...

6. Improvement with PCA

PCA was applied to reduce dimensionality while retaining variance. The Random Forest model trained on the PCA-reduced dataset...

Census Dataset Exploration and Modeling Report

7. Results and Discussion

The final Random Forest model, after PCA, achieved an accuracy of approximately 83.56%...

8. Conclusion

This project highlighted the predictive power of demographic and employment data in determining income levels...

9. Appendices

- A. Code Snippets: Detailed Python code for preprocessing, modeling, and evaluation steps...