NULLCLASS INTERNSHIP REPORT

TASK 1

Submitted by,

SECHI BURMAN

PROBLEM STATEMENT

Create a feature to detect the number of females as well as males in a meeting room along with their age. If the person wears a white shirt, we should make their age as 23 irrespective of their age and gender as well if they wear black shirt, we should predict that person as child irrespective of their age and gender and this feature should not work if we have less than 2 people in the meeting.

INTRODUCTION

This report details the first task of my data science internship, which involves developing a feature to detect the number of females and males in a meeting room along with their age. Specific conditions include setting the age to 23 if wearing a white shirt, predicting as a child if wearing a black shirt, and the feature not working if fewer than 2 people are present.

BACKGROUND

In many organizational settings, meetings play a crucial role in decision-making processes. Knowing the demographics of meeting participants can provide valuable insights for resource planning and ensuring diversity. This project aims to develop a feature that identifies the number of males and females in a meeting room and their ages, with specific rules based on the colour of their shirts.

LEARNING OBJECTIVES

The primary learning objectives for this task include:

- Understanding and implementing synthetic data generation.
- Developing algorithms for demographic detection.
- Handling special conditions in data processing.
- Enhancing skills in Python and relevant libraries such as Pandas, Numpy, and Streamlit.

ACTIVITIES AND TASKS

- Data Collection: Instead of real data, synthetic data was generated to simulate different scenarios in a meeting room.
- Preprocessing: Created a DataFrame with synthetic data representing genders, ages, and shirt colours.
- Model Training: While not involving traditional machine learning models, the task focused on rule-based data manipulation.
- Special Condition Implementation: Logic was implemented to set the age to 23 if the person wears a white shirt and classify the person as a child if they wear a black shirt.
- Validation and Testing: The feature was tested by generating synthetic data with a varying number of participants and shirt colours to ensure all conditions were handled correctly.

SKILLS AND COMPETENCIES

- Programming: Proficiency in Python, including libraries such as Pandas for data manipulation and Streamlit for creating interactive web applications.
- Data Generation: Skills in generating synthetic data to simulate real-world scenarios.
- Rule-Based Processing: Ability to implement complex rules for data processing.
- Interactive Applications: Experience in developing and deploying an interactive web application using Streamlit.
- Data Representation: Understanding the benefits of text data for simplicity and clarity versus the complexity and richness of image data.

FEEDBACK AND EVIDENCE

Since I did not have a supervisor during this internship, the feedback and evidence were based on self-assessment and peer reviews. Key points include:

- Ensuring the synthetic data accurately represented potential real-world scenarios.
- Effectively handling the special conditions related to shirt colour.
- Making the feature robust to variations in the number of participants.
- Justifying the choice of text data for this task, considering the simplicity and clarity it offers compared to image data.

Evidence of my work includes:

- The Python code for generating synthetic data, applying rules, and displaying results.
- Screenshots and logs from the Streamlit application demonstrating the feature's functionality under different conditions.

CHALLENGES AND SOLUTIONS

Challenge 1: Generating synthetic data that accurately represents real-world scenarios.

Solution: Used Numpy and Pandas to create a diverse dataset with varied ages, genders, and shirt colours. Decided to represent data in text format to simplify processing and presentation.

Challenge 2: Implementing the special conditions for shirt colours while maintaining code clarity.

Solution: Used clear and concise logic within the apply_rules function to handle the special conditions effectively.

Challenge 3: Ensuring the feature does not work when fewer than two people are present.

Solution: Added a preliminary check in the detect_people_features function to count the number of individuals before applying gender and age detection logic.

OUTCOMES AND IMPACT

The developed feature achieved the objective of detecting the number of males and females in a meeting room and estimating their ages, considering the special conditions related to shirt colours. This tool can significantly aid in demographic analysis during meetings, providing insights for organizational planning and diversity tracking. The decision to use text data ensured the feature was straightforward to implement and understand.

CONCLUSION

This task provided valuable experience in synthetic data generation, rule-based data processing, and interactive application development. The feature developed not only met the specified requirements but also demonstrated the practical application of data science techniques in real-world scenarios. The choice of text data over image data was justified by the simplicity and clarity it offered for this specific task.