CSE4288-Introduction to Machine Learning

Term Project Proposal

Group17

150120024-Enes Haksoy Öztürk

150123842-Enes Akturan

150121025-Ahmet Abdullah Gültekin

150120005-Ayşe Gülsüm Eren

150119762-Muhammet Ali Özkul

Gender-Based Wage Disparity Analysis Across Sectors

Description

In this project, we aim to address the issue of gender-based wage disparities in various sectors. Wage disparity between genders is a persistent and globally recognized problem that reflects social and economic inequality. By leveraging machine learning techniques, we will analyze existing data to identify patterns and potential causes of wage disparities, evaluate their magnitude across sectors, and offer insights for addressing these inequalities.

Problem Statement and Objectives

Gender inequality in salary and career opportunities remains prevalent. This project explores the disparities and aims to promote awareness through data-driven insights.

Our main goal is predicting gender based on features to identify gaps. Also, prediction of salary trends to identify wage inequality pattern is our aim. We will be using machine learning algorithms to provide our points.

Dataset Overview

The dataset includes these features:

- Age
- Gender
- Education Level
- Job Title
- Years Of Experience
- Salary

Proposed Methodologies

1.Data Preprocessing:

- Handle missing values.
- Encode categorical features.
- Scale and normalize numerical data.

2. Model for Gender Prediction:

- **Algorithm:**
  - Logistic Regression
  - Naive Bayes
  - Decision Tree Classifier
- **Why these algorithms?**
  Logistic Regression was chosen for its simplicity and effectiveness in linear classification tasks, providing clear insights into the relationship between features and gender.
  Naive works well with large datasets and categorical features, offering a good baseline.
  The Decision Tree Classifier allows us to visualize how features like salary or job position influence gender prediction.
  Together, these algorithms provided a comprehensive approach to understanding gender wage disparities.

3. Evaluation Metrics:

- Gender Prediction: Accuracy, Precision, Recall.

Proposed Timeline

Phase 1: Project Initiation
Duration: Week 1

Tasks:
• Define project scope and objectives.
• Collect and understand the dataset.
• Identify necessary tools and libraries.

Phase 2: Data Preprocessing
Duration: Week 2

Tasks:
• Handle missing values in the dataset.
• Encode categorical features.
• Scale and normalize numerical features.

Phase 3: Exploratory Data Analysis (EDA) and Modal Development
Duration: Week 3-4

Tasks:
• Analyze data distribution and correlations.
• Visualize patterns and trends in gender and salary data.
• Develop and train the Gender Prediction model.
• Experiment with various algorithms for gender classification.
• Develop and train the Salary Prediction model using Polynomial Regression.

Phase 4: Model Evaluation and Optimization
Duration: Week 4

Tasks:
• Evaluate Gender Prediction model using Accuracy, Precision, and Recall metrics.
• Optimize Polynomial Regression model for salary prediction.

Phase 5: Report and Recommendations
Duration: Week 5

Tasks:
• Prepare a detailed report with findings and insights.
• Present recommendations to address gender wage disparities.