Project Proposal

Problem Statement: Implementation of Fair Machine Learning Classifiers.

Team Members:

1. Anurag Krishna Sharma :2022MCS2071
2. Nitesh Dohre :2022MCS2070

Introduction: The objective of this project is to implement fair machine learning classifiers on two datasets among the given datasets in the paper (Currently we are considering Ricci promotion dataset and Heritage health dataset). The goal is to identify any biases or discrimination present in the classifiers and take measures to ensure fairness in the classification process. We will perform single-variable and two-factor analysis on various fields to understand the relationships between the target variable and the protected variables, as well as between the protected variables and other input variables. We will then train machine learning classifiers both with and without blinding, and evaluate the classifiers based on various metrics with respect to the protected variables. Finally, we will modify the classifiers or post-process the output to ensure fairness in the classification process.

Methodology:

1. Dataset Selection: We will use the Ricci promotion dataset and Heritage health dataset for this project.
2. Data Analysis: We will perform single-variable and two-factor analysis on various fields to understand the relationships between the target variable and the protected variables, as well as between the protected variables and other input variables.
3. Training of Machine Learning Classifiers: We will train machine learning classifiers both with and without blinding, and evaluate the classifiers based on various metrics with respect to the protected variables.
4. Fairness Assessment: We will identify any biases or discrimination present in the classifiers and take measures to ensure fairness in the classification process. This may involve changing the classifiers themselves, or post-processing the output to adjust for bias or discrimination.

Expected outcomes:

1. Identification of any biases or discrimination present in the classifiers.
2. Implementation of fair machine learning classifiers that promote fairness in the classification process.
3. A better understanding of the relationships between the target variable and the protected variables, as well as between the protected variables and other input variables.

Deliverables:

1. A report on the findings of the single-variable and two-factor analysis.
2. A report on the performance of the machine learning classifiers with and without blinding, and the evaluation metrics with respect to the protected variables.
3. A report on the measures taken to ensure fairness in the classification process.
4. A codebase for the implementation of fair machine learning classifiers on the Ricci promotion dataset and Heritage health dataset.

Libraries we will use to achieve our goal:

1. Pandas: Used for data manipulation and analysis
2. NumPy: Used for numerical operations and mathematical functions
3. Scikit-learn: Used for machine learning algorithms and evaluation metrics
4. Matplotlib: Used for data visualization and plotting graphs
5. Seaborn: Used for advanced visualization and statistical graphics
6. Fairlearn: Used for implementing fair machine learning algorithms and evaluating fairness metrics
7. AIF360: Used for bias mitigation techniques and fairness evaluation in machine learning
8. Keras: Used for building and training deep learning models

Conclusion: The implementation of fair machine learning classifiers is an important step towards promoting fairness and reducing bias in the classification process. This project will provide a better understanding of the relationships between the target variable and the protected variables, as well as between the protected variables and other input variables. The outcome of this project will be beneficial in promoting fairness and reducing bias in machine learning algorithms and can be applied to various other domains as well.