"A Machine Learning Study of Fair Employment Factors Across the U.S."

Jacelyn Asafo-Adjei, Maya Ody-Ajike • 04.23.2025

Overview

The issue of unfair employment practices is not only a historical concern but a persistent challenge that has lasting implications for individuals and communities. The disparities in hiring processes have far-reaching consequences, contributing to societal inequities and hindering the full realization of diverse talents and capabilities. It is within this context that this research project sought to shed light on the multifaceted nature of prejudiced hiring practices, aiming to uncover the hidden complexities that perpetuate inequality.

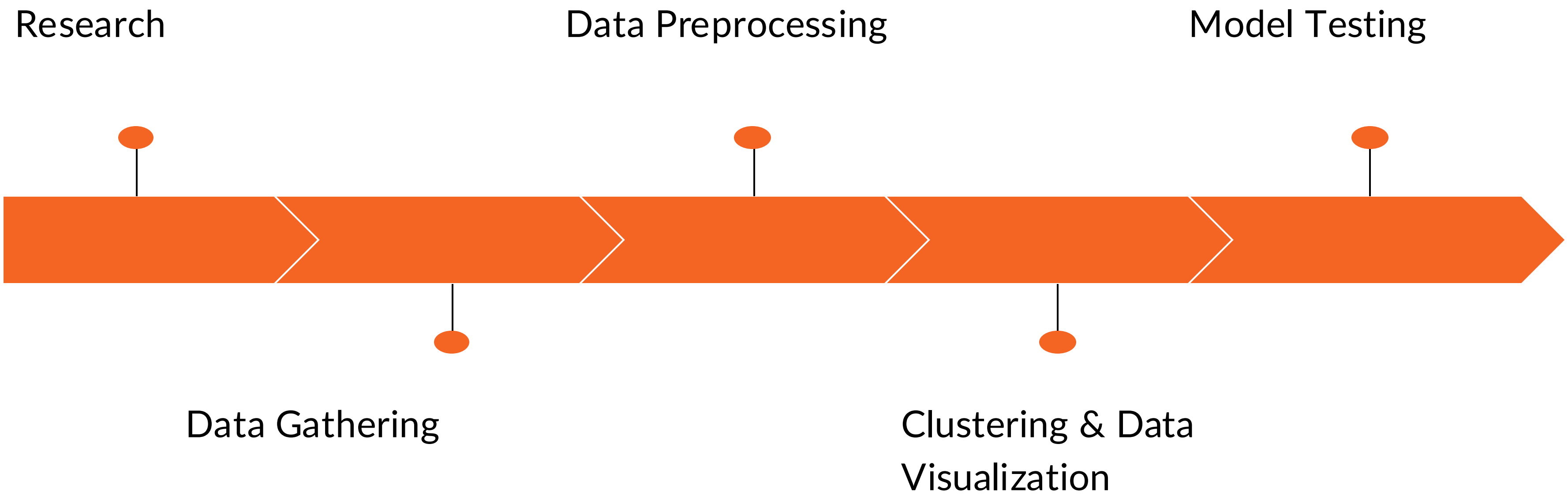
Problem Statement

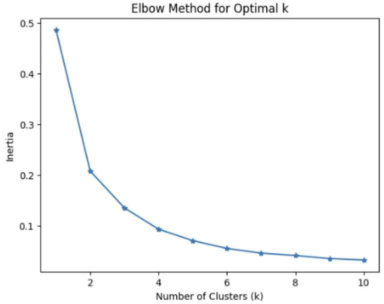
The persistent issue of unfair hiring practices has become a pressing societal concern, necessitating a comprehensive investigation into the intricate factors that contribute to hiring disparities. Despite advancements in technology and an increasing awareness of the importance of diversity and inclusion, biased hiring processes continue to perpetuate inequalities across various demographic groups. The problem at hand is twofold: first, the lack of a standardized and transparent methodology for assessing and mitigating biases in hiring models; and second, the limited understanding of the spatial and regional dynamics that influence these disparities.

Methodology

- Clustering
- Classification
- K Nearest Neighbors
- ROC Curve

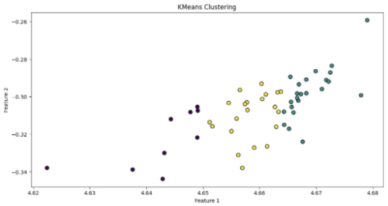
Design Process





Graph of our elbow method, showing our optimal k = 3.

A figure showing clusters in PCA space and our silhouette score of 0.4.



A figure showing our kMeans clusters.

```
accuracy: 0.75
              precision    recall  f1-score   support

        Fair       0.60      1.00      0.75         6
        Good       1.00      0.20      0.33         5
        Poor       1.00      1.00      1.00         5

    accuracy                           0.75        16
   macro avg       0.87      0.73      0.69        16
ghted avg       0.85      0.75      0.70        16
```

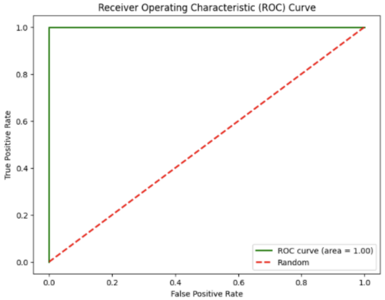
A figure showing the model's kNN accuracy score of 0.75.

	Poor	Fair	Good
Poor	6	0	0
Fair	4	1	0
Good	0	0	5

A figure showing the outcome of our confusion matrix.

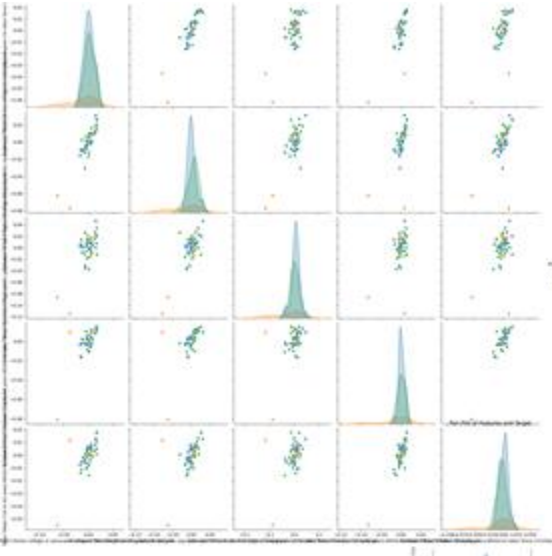
```
The cross validation accuracy scores are: 1.0 1.0 0.91 1.0 1.0
The cross validation precision scores are: 1.0 1.0 0.8 1.0 1.0
The cross validation recall scores are: 1.0 1.0 1.0 1.0 1.0
The cross validation f1 scores are: 1.0 1.0 0.89 1.0 1.0
```

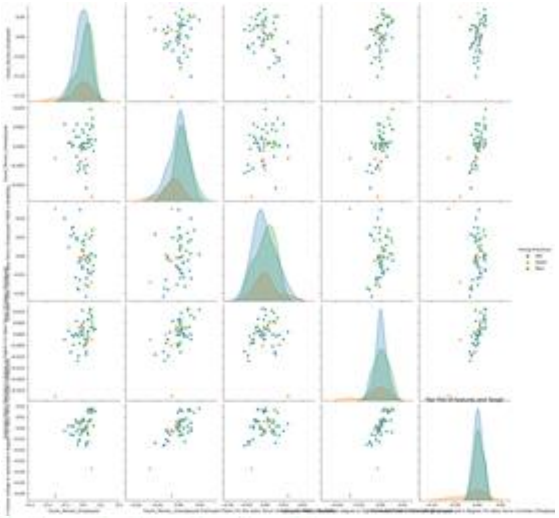
A figure showing the cross-validation results.

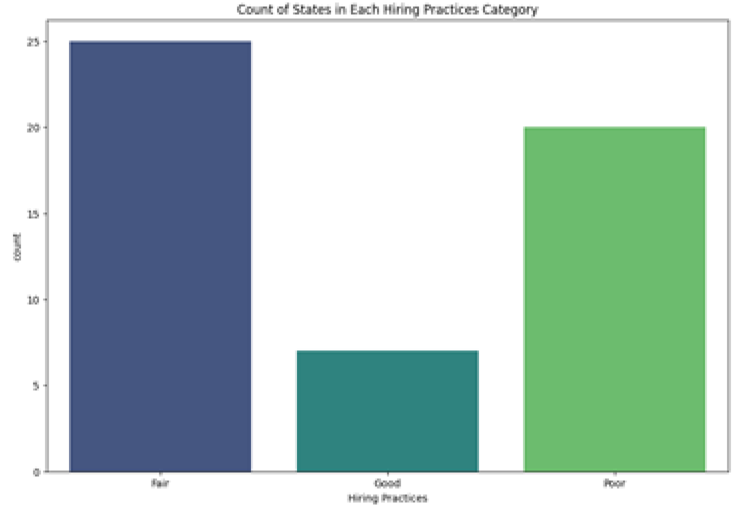


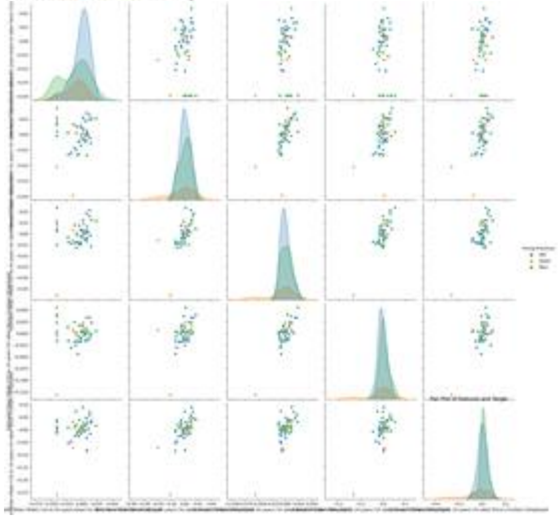
A figure showing the ROC curve visualization.

Result






Count of States in Each Hiring Practices Category
Count
0
5
10
15
20
25
Fair
Good
Poor
Hiring Practices



Future Works

1. Data found was not enough. Hiring practices are subject to legal and regulatory frameworks, and biased models may lead to legal consequences.

2. Presence of NaN values in the data gotten affected the results. Better handling of these values would provide a more accurate result

3. Regularly monitor and assess the real-world impact of the model, and be prepared to intervene and make adjustments if unintended consequences are identified.

4. Document all aspects of the model development process, including data preprocessing steps, feature selection, and model architecture. This transparency aids in identifying and addressing potential biases.

References

1. Python Software Foundation (2021). Scikit-learn: Machine Learning in Python Retrieved from https://scikit-learn.org/stable/

2. McKinney, W., & others. (2011). pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing, 14.

3. Seaborn (2021). Statistical Data Visualization. Retrieved from https://seabom.pydata.org/

4. Matplotlib Development Term. (2021). Matplotlib: Visualization with Python Retrieved