

Gender Neutrality and Inclusion Using Machine Learning

Nikhil Singh



Methods Tried

- **Data Engineering:**
 - **Feature Binning, Aggregation.**
 - **Normalization of Continuous Values**
 - **Normalization of Target**
 - **Value[Fitment Percentage]**

What Worked:

Model: The three models performed equally well but LightGBM

performed a little better.

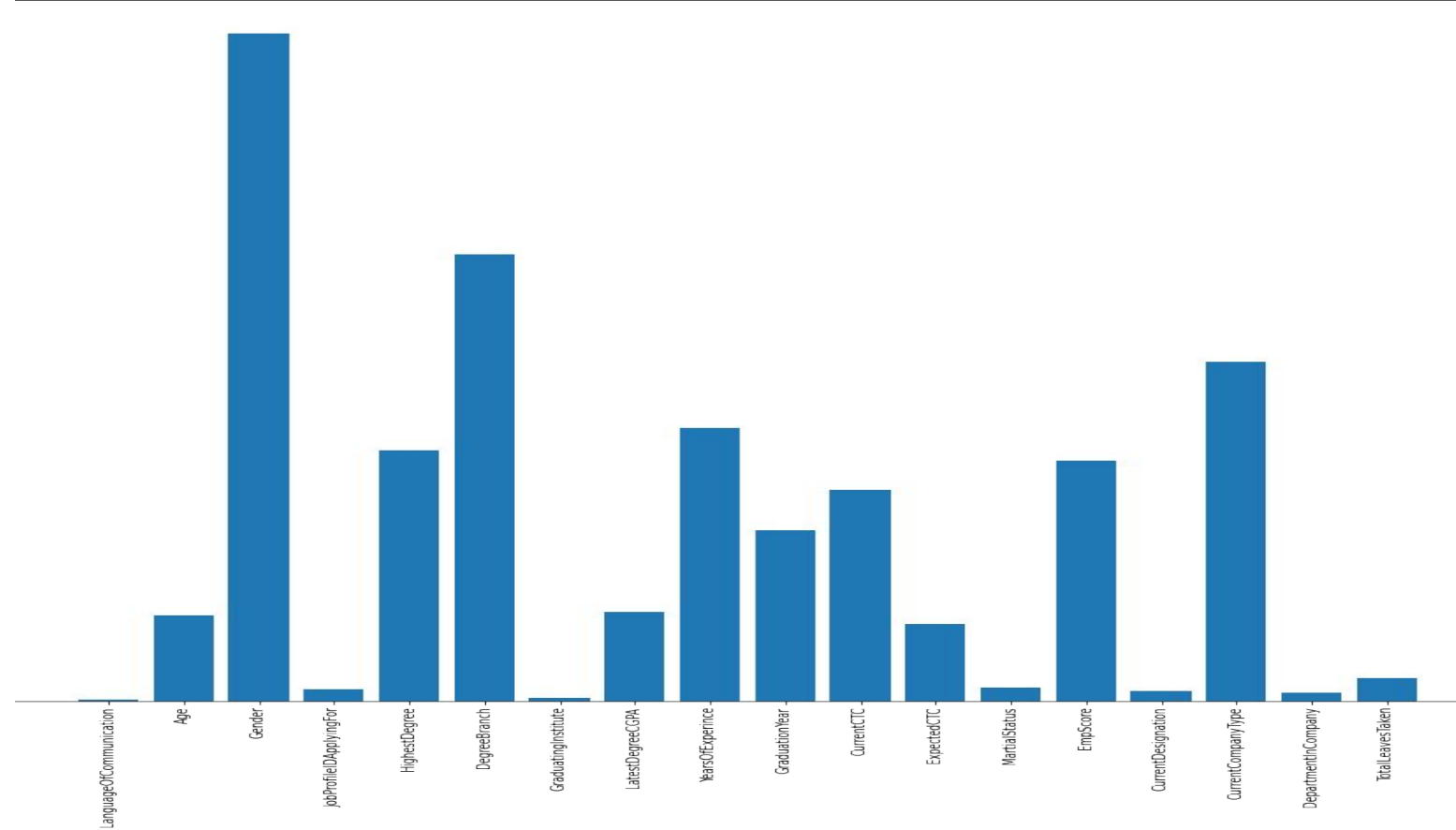
Data Handling: Binning of continuous values was performed but simple normalization performed better.

Results

Using a 5 fold CV the GBM based Bias-Influential classifier achieved an ROC of ~ 0.99 on 20% of the hold-out test data.

The GB Regressor achieved an MAE of 5.60 on the same test set.

*The NaN values were imputed using a ml model.



Final Conclusion: Gender, DegreeBranch and Current Company Type seems to be the most important in deciding the Bias.

Notebook

For More Detailed Understanding of the technique including the comparative analysis see:

https://colab.research.google.com/drive/1zK1mOFicTBJoJMmKcR_4m8qWR2uMDta7?usp=sharing

*The NaN values were imputed using a ml model.