AI Monitoring & Ethics Dashboard Report

Generated on: 2025-11-04 20:09
Author: Diego Ferreira

1. Summary of Results

Logistic Regression:
Accuracy: 0.825
F1-score: 0.557
ROC-AUC: 0.854
DP diff: 0.176
EO diff: 0.261

Random Forest:
Accuracy: 0.858
F1-score: 0.676
ROC-AUC: 0.911
DP diff: 0.170
EO diff: 0.079

2. Interpretation and Ethical Insights

The evaluation of both models shows that the **Random Forest** model outperforms the Logistic Regression in predictive capability and fairness balance. Its ROC-AUC of 0.91 indicates strong discrimination power between income categories, while the low Equalized Odds difference (0.08) suggests more equitable treatment between groups.

Ethically, this means that the Random Forest model not only performs better statistically but also makes fewer biased predictions related to sensitive attributes such as gender and race. Nevertheless, a model's fairness is not static — it can degrade over time due to data drift, population shifts, or changes in social context. Therefore, regular audits and fairness re-evaluation are essential to maintain responsible AI deployment.

Additionally, while bias mitigation was not directly applied here, results suggest that data pre-processing strategies (like reweighting) and threshold adjustments could further reduce disparities in future iterations.

3. Results Summary and Challenges

Summary of Results:
The Random Forest model achieved the best overall performance and fairness metrics, with an F1-score of 0.68 and ROC-AUC of 0.91, outperforming Logistic Regression in both predictive power and ethical balance. Fairness metrics indicated limited bias between gender and racial groups, confirming that the Random Forest model produces more

consistent and equitable outcomes.

Challenges Faced:
• Ensuring fairness required identifying and isolating sensitive attributes without compromising predictive accuracy.
• The dataset contained historical bias that influenced early model iterations, requiring careful preprocessing and evaluation.
• Achieving interpretability was a challenge due to model complexity — balancing transparency with performance was a central design decision.
• Managing and comparing two models (Logistic Regression vs Random Forest) demanded cross-validation and consistent metric tracking.

These challenges highlight the importance of combining technical optimization with ethical awareness throughout the machine learning lifecycle.

4. Recommendations for Future Work

• Implement fairness-aware algorithms such as Reweighing or ThresholdOptimizer.
• Include continuous fairness tracking as part of an MLOps monitoring pipeline.
• Expand the dataset to capture a wider range of demographics and socioeconomic patterns.
• Integrate SHAP and LIME explanations for transparency and accountability.
• Publish a formal **Model Card** documenting ethical and performance considerations.

5. Final Reflection

The project demonstrates how performance and fairness can be jointly monitored using explainable and auditable processes. Among the tested models, Random Forest shows the best trade-off between predictive power and ethical balance, making it the recommended choice for production deployment.

Ultimately, responsible AI engineering demands continuous monitoring, transparent communication of model limitations, and proactive governance to ensure models align with human values and social equity.