

D-BIAS Analysis Report

cached_dataset.csv

Generated: 18/11/2025, 02:42:34

Fairness Score: 90/100

Fairness Label: Excellent

Bias Risk: Low

Detected Biases

- Categorical Imbalance (col: transmission) – Moderate
- Categorical Imbalance (col: sales_classification) – Moderate

Recommendations

- **Prioritize Rebalancing for `sales_classification`:** Given that **69.5%** of this feature belongs to a single class, this is the most critical bias to fix. Implement **SMOTE** to oversample the minority classes ('Medium', 'High') to create a more balanced dataset for model training.
- **Apply Class Weights for `transmission`:** For the less severe imbalance in the `transmission` feature, a simpler approach like applying **class weights** during model training is recommended. This will penalize the model more for misclassifying the rarer transmission types, encouraging it to learn their patterns.
- **Adopt Robust Evaluation Metrics:** Move beyond simple accuracy. Use metrics like **F1-Score, Precision, Recall, and AUC-PR** to evaluate model performance, as they provide a much clearer picture of how well the model handles the underrepresented classes.
- **Investigate Data Collection:** Review the data sourcing process. The imbalances may reflect real-world distributions, but they could also be artifacts of how the data was collected. Understanding the source can provide context and guide future data acquisition efforts to create more balanced samples.

Conclusion

The dataset's overall "fairness health score" is **Moderate**. It contains actionable insights but is marked by significant imbalances that must be addressed before deployment in a real-world application. The biases are not critical flaws but are substantial enough to require careful preprocessing and mitigation to ensure fair and accurate outcomes.