COMPAS Privacy vs Fairness Trade-off

Team:

Ayush Oturkar(ao586)

Keshvi Gupta(kg835)

Atharva Sherekar(as4138)

Problem Statement :

Fairness and privacy are two important considerations in ethical statistical learning. They are often at odds because protecting privacy can introduce bias, and ensuring fairness can require revealing sensitive information. Therefore, in most real life applications, it is important to make a choice of models, algorithms and techniues based on the situation at hand and the specific needs of each application.
We are implementing this project to observe this notion in practice by implementing variety of statistical modelling, privacy and fairness preserving techniques.

Dataset :

- The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) recidivism dataset is a public dataset that contains information about criminal defendants in Broward County, Florida. The dataset includes information such as the defendant's age, race, gender, criminal history, and risk of recidivism.
- Owing to the trade-off between fairness and privacy in ethical statistical learning, a model that is very accurate at predicting recidivism(re-offending) may be unfair to certain groups of people, such as minorities. This is because the model may be trained on data that is biased against these groups. Additionally, a model that is very private may not be very accurate at predicting recidivism. This is because the model may not be able to access all of the data that it needs to make accurate predictions.

Evaluation :

- We will evaluating the tradeoff by simultaneously comparing Model Performance, Fairness metric(s) and Privacy Metric(s)
- Performance Metrics :
  - Model : APR/F-1 Score
  - Fairness: Statistical parity, Equal opportunity, Calibration, Predictive parity and Equalized odds
  - Privacy Metrics : Epsilon and Sensitivity

	Model Performance	**Fairness Performance**	**Privacy Performance**
Baseline model	High Performance	No Fairness	No Privacy
Baseline + Privacy implementation	Comparatively lower performance	No Fairness	More Private
Baseline + Fairness Preservation	Comparatively lower performance	Higher Fairness	No Privacy
Baseline + Privacy implementation + Fairness Preservation	Even lower performance	Somewhat Fair	Somewhat Private, but lesser private than iteration 2

Methodology :

- Data preparation:
    - cleaning the data, removing any outliers, and imputing any missing values.
    - Scale the data so that all features are on the same scale.
- Model selection:
    - Binary classification task: logistic regression, random forests, and NN.
- Model training:
    - Cross-validation split to ensure no overfitting.
- Model evaluation
- Fairness evaluation: statistical parity, disparate impact, and equal opportunity.
    - In the context of the COMPAS recidivism dataset, it is important to consider all three of these fairness metrics. Statistical parity is important because it ensures that the model is not disproportionately predicting recidivism for certain groups. Disparate impact is important because it ensures that the model is not disproportionately harming certain groups. Equal opportunity is important because it ensures that the model is giving all individuals with the same criminal history and other relevant characteristics the same chance of a positive outcome.
    - In our case Disparate impact would be more important, as a model that achieves disparate impact would not disproportionately predict recidivism for certain groups, even if it does not achieve statistical parity.
- Privacy evaluation:
    - Differential privacy and Federated Learning.
    - Sensitivity analysis

Details :

- Libraries
    - Data preparation: NumPy, Pandas, scikit-learn
    - Model selection: scikit-learn
    - Model training: scikit-learn
    - Model evaluation: scikit-learn
    - Fairness evaluation: AIF360, Fairlearn
    - Privacy evaluation: TensorFlow Privacy, PyDP