Model Card for XG Boost

Model Details

- Developed by researchers at the University of Delaware, Newark, Delaware, USA, 2022, no information on version update.
- The model is part of an extensive data preprocessing pipeline for MIMIC IV.
- The model is part of a greatly customizable pipeline to extract, clean, and preprocess the data available in the fourth version of the MIMIC dataset (MIMIC-IV).
- Innovation: A standard model presented as part of an innovative end-to-end wizard-like package for predictive model creations and evaluations.
- Paper: XG Boost

Intended Use

- Intended to extract data for risk prediction from MIMIC IV CRD.
- Particularly intended for Healthcare machine learning (HML) practitioners and hospitals.
- Not applicable for anything other than the intended use.

Experimented Factors

- The model showed varied performance across demographic groups.
- Evaluation factor is the highly correlated ethnic attribute. When the model's performance is evaluated against ethnic subgroups (White, Asian, Black, Hispanic/Latino, and others), it shows inconsistent results.
- Further factors like age, gender, insurance, and other covariates are also possible for evaluation.

Metrics

- Evaluation metrics include Receiver Operating Characteristic Area Under the Curve (ROC-AUC) and Precision-Recall Area Under the Curve (PR-AUC).
- Fairness metrics - The 'Fairness Report' provides a detailed report on HML fairness and its metrics.

Training Data

- MIMIC IV v2.0. It is available in the 'Datasheet for CRD'.

Evaluation Data

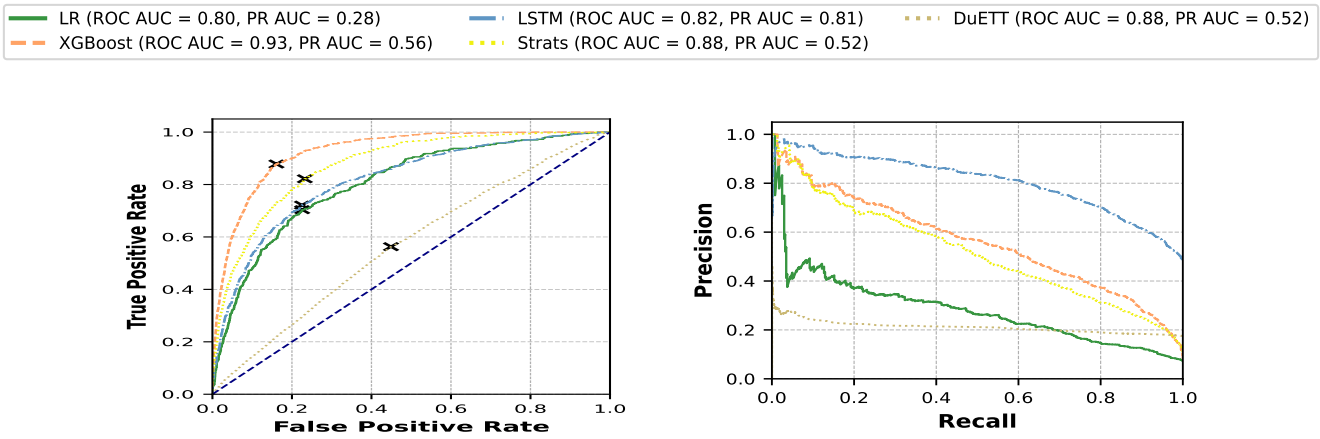
- 80:20 split for the training and test sets.
- The models were trained for up to 1000 epochs until the validation loss stopped improving for 10 continuous epochs.
- We used the same hyperparameters of the model listed in the code repository.

Ethical Considerations

- EHR data is used for evaluating the model's performance where its decision is life-altering based on the intended tasks.

Caveats and Recommendations

- Findings show that the model showed inconsistent results when tested on fairness even though it performed better in terms of prediction when compared to the SOTA DuETT and STraTS.



(a) ROC Curve

(b) PR Curve

Figure 1: Prediction performance analysis across static and time series models. Panels (a) and (b) show the ROC-AUC and PR-AUC of the models, respectively, and the operating points in (a).