WCLD: Curated Large Dataset of Criminal Cases from Wisconsin Circuit Courts

Elliott Ash, Naman Goel, Nianyun Li, Claudia Marangon, Peiyao Sun

Main Contribution

- Research-ready large criminal cases dataset, for research in algorithmic fairness and beyond.
- 1.5 million instances.
- Variables such as prior criminal counts and recidivism outcomes (including violent recidivism).
- Large number of samples from five racial groups.
- Other attributes: sex, age (at judgment and first offense), neighbourhood characteristics obtained from census data, detailed types of offense, charge severity, case decisions, sentence lengths, year of filing etc. Pseudo-identifiers for judge, county and zipcode.

Dataset Construction

- WCCA API indexes public case records and docket information from 72 county courts.
- Collected records of cases filed from 1970, through 2020.
- 11M records (2.5M criminal).
- Constructed a dataset for machine learning by using a combination of variables that were directly available in the records and calculating unavailable variables using various techniques.
- For example, created prior criminal counts and recidivism outcomes by performing database search over the records. Used GPT-4 for labelling violent crimes given the charge descriptions.

Summary of the Dataset

	Full sample	Caucasian	African American	Hispanic	Native American	Asian
Sample size	1,476,967	964,922	333,036	101,607	63,862	13,540
Sample share		65.33%	22.55%	6.88%	4.32%	0.92%
Recidivism (if observed)	42.21%	40.34%	46.43%	38.76%	56.47%	37.80%
Sex						
Male	80.40%	79.05%	83.47%	88.88%	69.65%	87.57%
Age						
Below 30	51.38%	49.45%	54.13%	56.91%	53.71%	68.60%
30 to 60	47.44%	49.09%	45.17%	42.61%	45.58%	30.85%
Case type						
Felony	32.18%	30.76%	39.98%	21.09%	29.80%	36.39%
Misdemeanor	43.04%	43.67%	43.14%	34.12%	47.55%	40.89%
Criminal Traffic	24.78%	25.57%	16.88%	44.79%	22.66%	22.73%

Strengths

Addresses several limitations of previous datasets like COMPAS/ProPublica:

- ✓Large size.
- ✓Large number of samples from five racial groups.
- ✓Data from different courts.
- ✓Data from 72 counties.
- ✓More attributes.
- ✓Less variance.
- ✓Data from a long period of time (1970-2020).

Summary of Key Observations

- More training data does not necessarily lead to a fairer model.
- Base rates and group sizes are not the only determinants of unfairness; the disparity does not necessarily decrease when we balance these between races.
- Depending on the time of training data and when the model is applied, fairness evaluation varies significantly.
- Adding race as an attribute may increase unfairness without increasing accuracy, but adding neighbourhood characteristics increases fairness in our experiments.
- For some types of offense, fairness is much worse than other types of offense.
- Training separate models for different races is not always favorable for the minority.
- Data-centric interventions often affect fairness metrics but not accuracy metrics.
- Fairness and accuracy estimates often vary significantly under distribution shift.

Temporal Factors

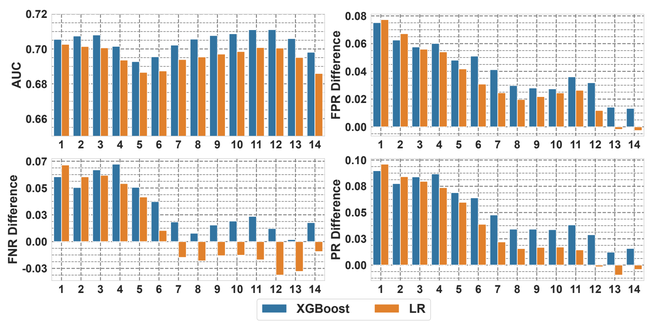


Figure 4: The Role of Time: X axis corresponds to training datasets from two consecutive years between 2000 and 2018 (e.g., "1" on x axis denotes the training data from the years 2000 & 2001, "2" denotes the training data from the years 2001 & 2002 and so on. The test data comes from the next two years after a two year gap (e.g. if training data is from 2000 & 2001, test comes from 2003 & 2004.)

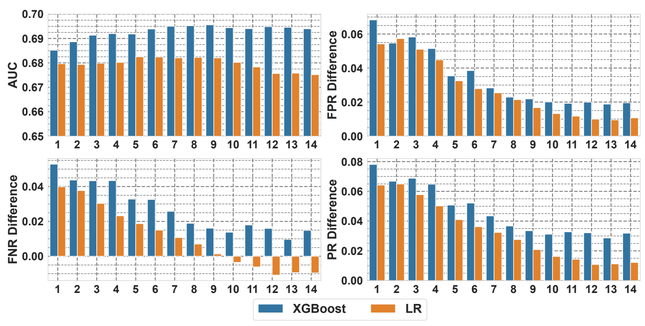
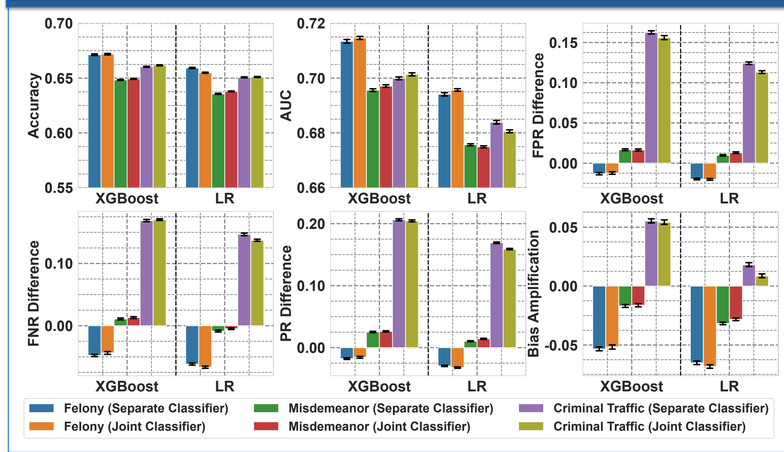


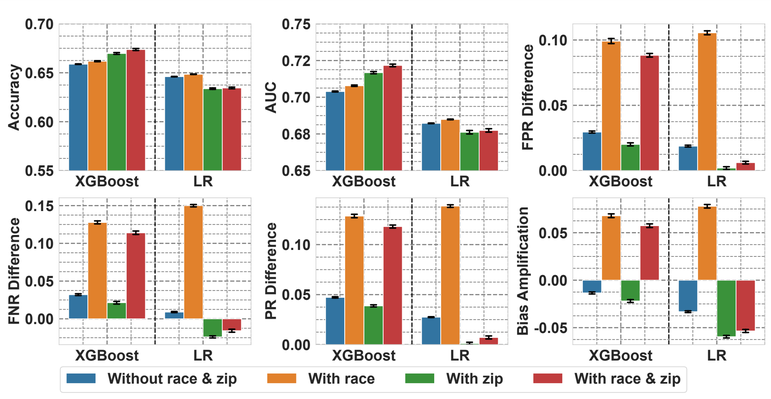
Figure 5: The Role of Time: X axis corresponds to training datasets, as described in the caption of Figure 4 caption. Main difference is that the test data in this figure is the reserved data from all the years between 2000 and 2018.

Type of Offense



A separate classifier for each offense type is trained on data from that offense type. The performance of the classifiers are then observed on respective offense types. For comparison, the performance of a joint classifier, that is trained on all the data and uses offense type as a predictor, is also shown by offense type.

Race and Zipcode Demographic Data



* Zipcode level demographic data (from census):

· Population density · Proportion who attended college · Proportion eligible for food stamp · African American population share · Hispanic population share · Proportion of male · Proportions who live in rural and urban area · Median household income

Limitations

- Biases encoded in various variables is a fundamental limitation, difficult (or perhaps impossible) to address in any dataset despite coverage and size.
- Known biases and limitations discussed in the paper in detail.
- Must be considered while using the dataset and drawing conclusions.

Dataset Availability

- Dataset is freely available at:

  http://clezdata.github.io/wcld/

- License: CC-BY-NC-SA 4.0
- Restricted to academic research use.

ETH zürich

UNIVERSITY OF OXFORD