

Model Card - Credit Prediction

Model Details

- Developed by Lehigh University DSCI 451 Student
- Random Forest
- Pretrained for loan approval prediction and validated using 10-Fold cross validation

Intended Use

- Intended to be used by financial institutions to determine if an applicant should be approved or denied for a loan
- Not intended for fully automated uses
- Not intended to make judgements about specific users

Factors

- Model performance is affected by the number of trees used in the model, which is dependent upon the computation power of device.

Metrics

- Confusion matrix
- Overall accuracy
- Precision
- Recall
- F1-Score

Training Data

- Source: state-level loan application data from the U.S. CFPB, complying with the HMDA.
- Features: applicants loan amount, income, property value, loan type, loan purpose, occupancy type, lien status, lei, credit score type, hoepa Status, area code and areas median family income.
- Excluded features: those that contain protected classes, such as age, sex, race, and ethnicity

Evaluation Data

- 10-Fold cross validation training/testing data split from source data and selected features

Ethical Considerations

- Although protected classes aren't used in training, the dataset itself is biased against blacks and hawaiian native americans.

Caveats and Recommendations

- Dataset used for training is based from Georgia, U.S., so model may not best generalize for other regions or financial institutions not present in the dataset.

Model Results

Overall accuracy: 0.9859

Precision: 0.992

Recall: 0.990

F1-Score: 0.991

Prediction	Truth	
	1(Approve)	3(Denied)
	1	3
1	205,410	1,539
3	2,472	74,991

Dataset Analysis

