Home Credit Default Risk with LightGBM

Apurva Bhargava, Eileen Cho

Nutritional Labeling

----->Performance

    ----->AUC

    ----->Accuracy

----->Diversity

----->Fairness

    ----->Subpopulation Accuracy

    ----->Statistical Parity

    ----->Disparate Impact

    ----->False Positive Rate Ratio (FPRR) & False Negative Rate Ratio (FNRR)

    ----->Positive Predictive Value (PPV) & Negative Predictive Value (NPV)

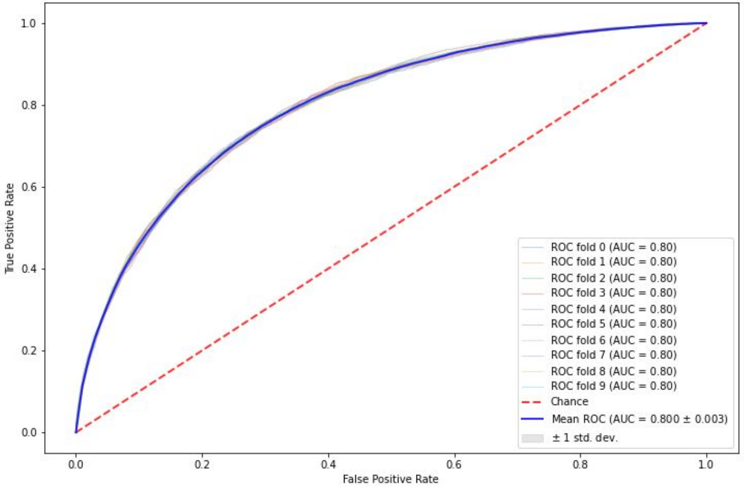
    ----->Fraction of risk/ no-risk assignment conditioned on an attribute

----->Stability

----->Interpretability

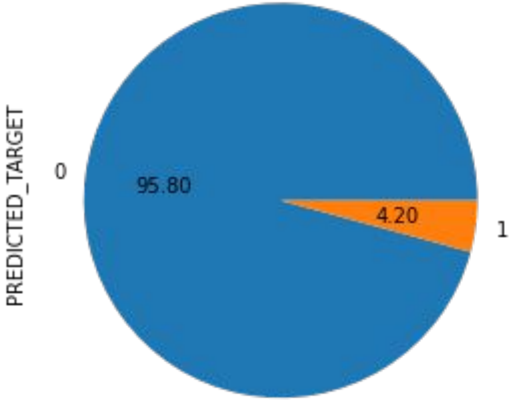
ADS Implementation Pipeline

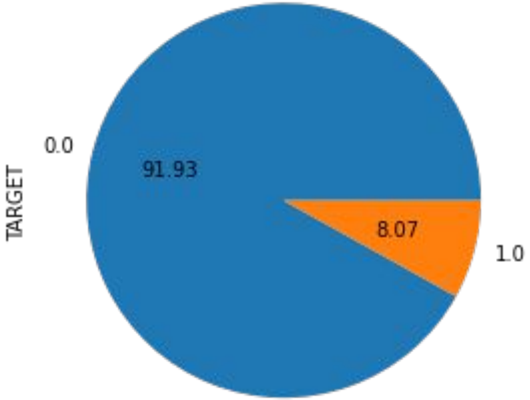
- Feature engineering - results in 660 input features
- LightGBM with GOSS - specific implementation of Gradient Boosted Tree from Microsoft
- K-fold cross-validation with k=10
- Achieves AUC score of 0.80

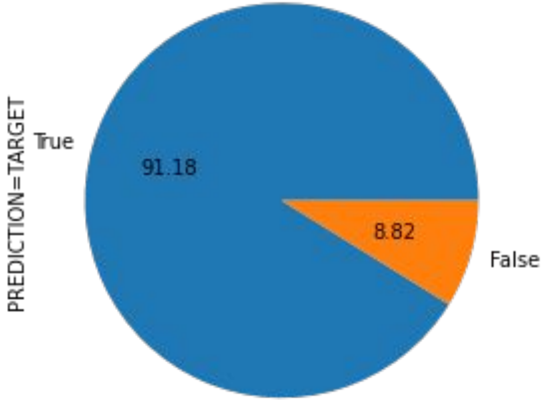


ADS Performance: 0 is low-risk and 1 is high-risk

Training (Actual Target)
(very skewed)





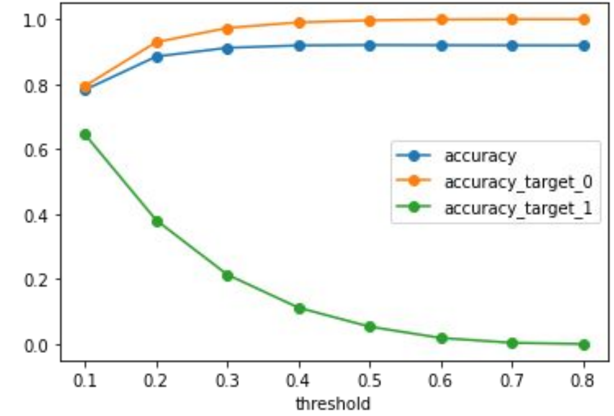


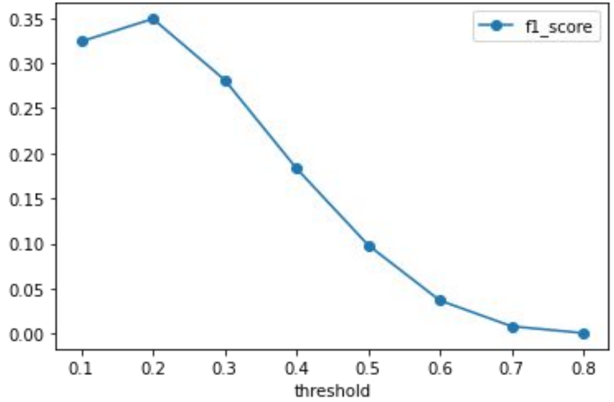
Validation (Prediction Probability
-> Predicted Target)

AUC = 0.80

Accuracy = 91.18%
at threshold=0.3

ADS Performance: Selecting a threshold





Goals	**Stakeholders**
Minimize high-risk clients	Home Credit
Maximize (low-risk) clients	Home Credit, Applicants
Maximizing overall accuracy	Home Credit
Maximizing F1 Score	Home Credit, Everyone

Suitable Thresholds: 0.2 0.3

Data Diversity

-Gender

-Ages

-Work Experience

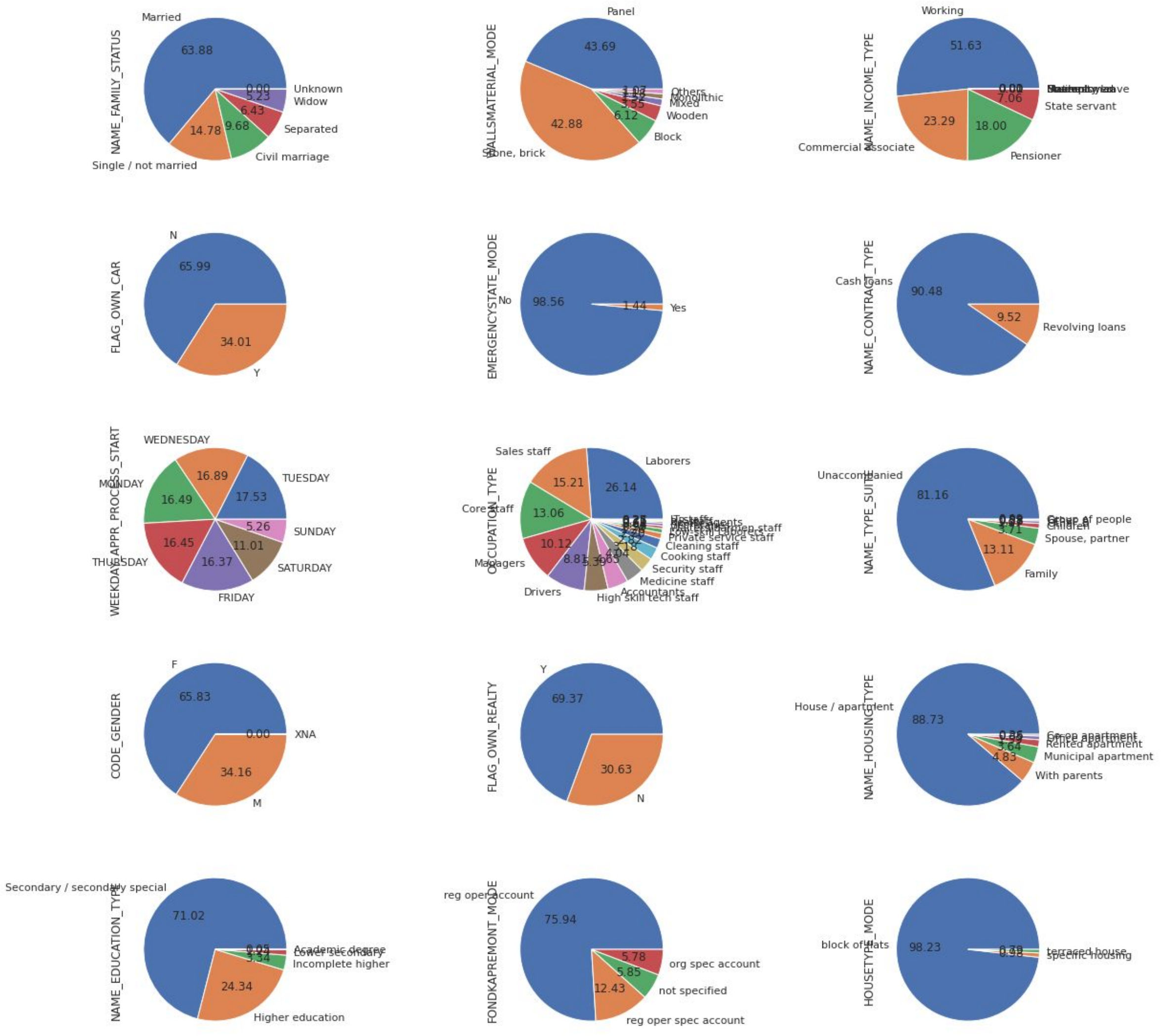
-Regions

-Occupation Types

-Income Types

-Family Status

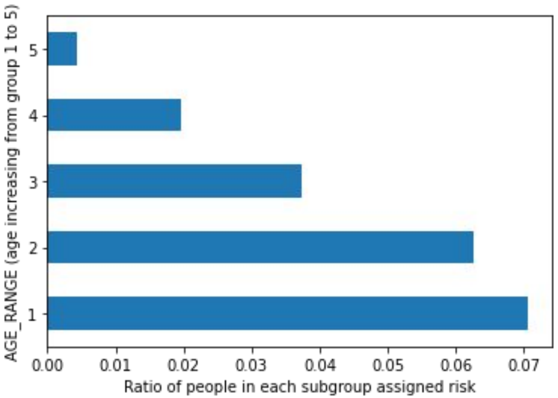
-House Types

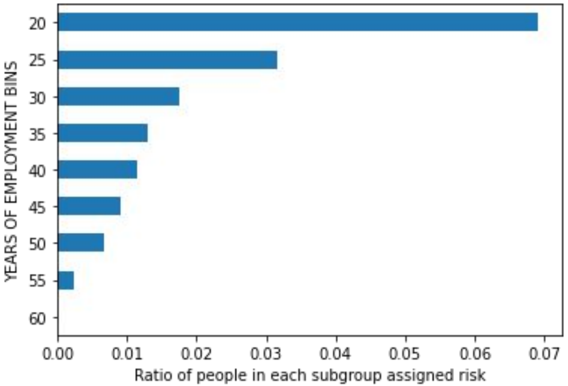


ADS Fairness: Subpopulation Accuracy and AUC

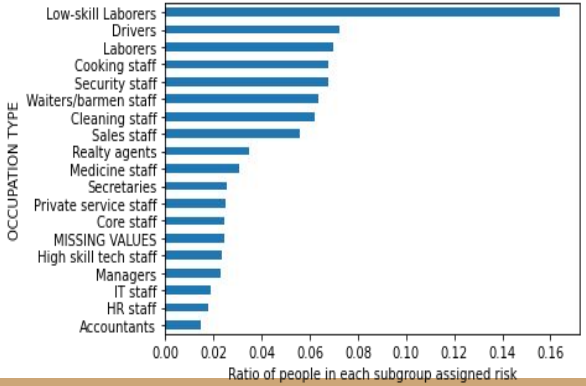
- Higher accuracy for privileged groups
- Almost constant AUC for most subpopulations (the metric used for solution validation is stable)
- For further analysis, positive means 0 or low-risk label, and negative means 1 or high-risk label

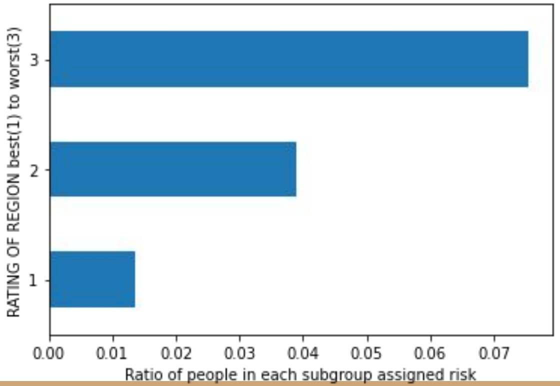
ADS Fairness: Statistical Parity

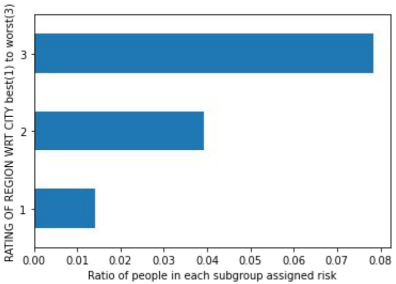


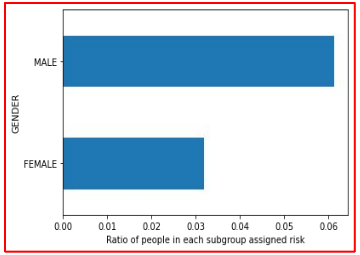


- ❏ **Statistical Parity not satisfied**
- ❏ **Less privileged groups have higher ratio of high-risk assignment**
- ❏ **Exception: Gender**









ADS Fairness: Disparate Impact, FPRR and FNRR

Attribute	Privileged Group	Unprivileged Group	Disparate Impact
Gender	Male	Female	1.031392078802175
Age	>=50	<50	0.9638316697820878
Region Rating	1	3	0.9372257744139441
Region rating	2	3	0.9618648938300324
Region rating	1	2	0.9743840121683013

Attribute	Priv. Group	Unpriv. Group	FPRR	FNRR
Gender	Male	Female	1.0273	0.7316
Age	>=50	<50	0.9734	1.4676
Region Rating	1	3	0.8836	1.9405
Region rating	2	3	0.9767	1.3054
Region rating	1	2	0.9047	1.4865

- Disparate Impact below 1 except for Gender. Very high values because of target skewness

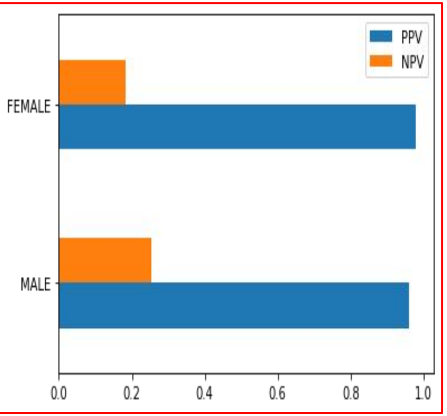
- FPR-> probability of false low-risk assignment. FPRR = FPR(unpriv)/FPR(priv)

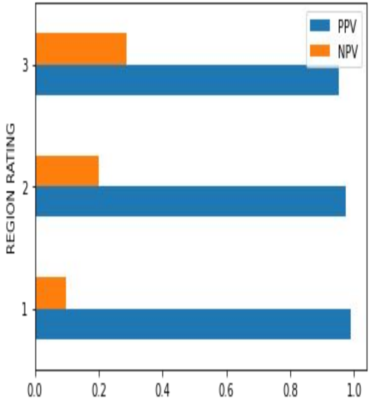
Privileged groups likely to be wrongly assigned low-risk

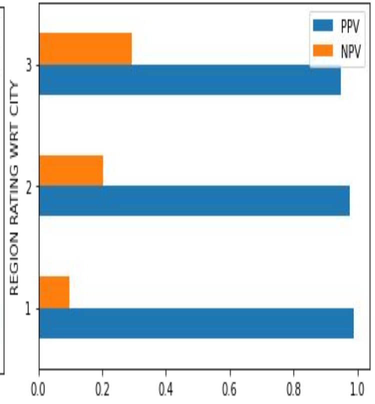
- FNR-> probability of false high-risk assignment. FNRR = FNR(unpriv)/FPR(priv)

Unprivileged groups likely to be wrongly assigned high-risk

ADS Fairness: PPV and NPV

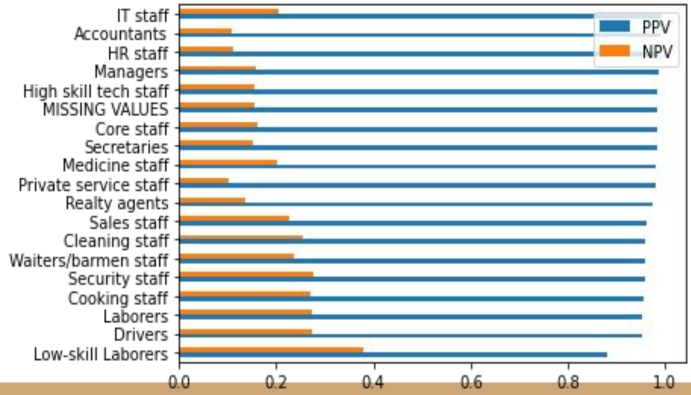


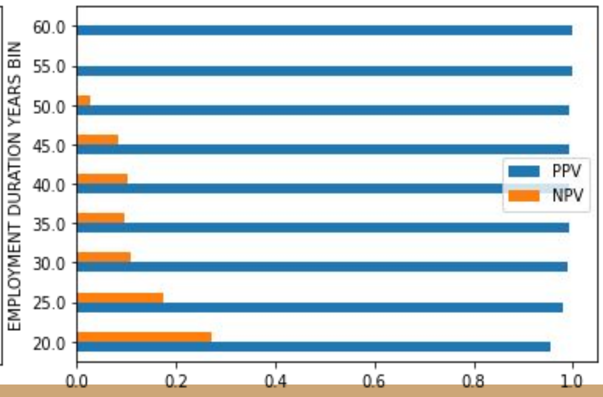


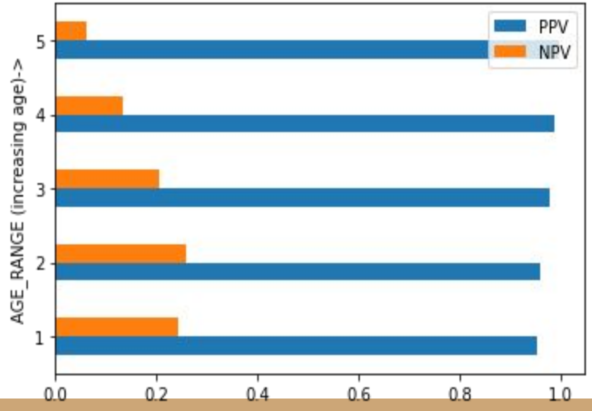


High PPV for privileged groups means low-risk individuals are more likely to be marked low-risk in privileged groups as compared to unprivileged groups.

High NPV for unprivileged groups means high-risk individuals are more likely to be marked high-risk in unprivileged groups as compared to unprivileged groups.



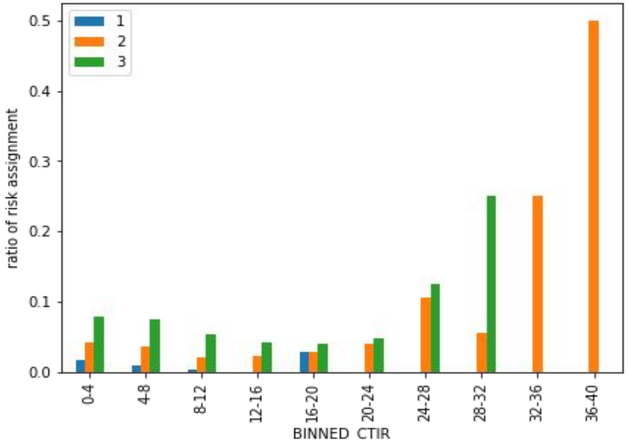


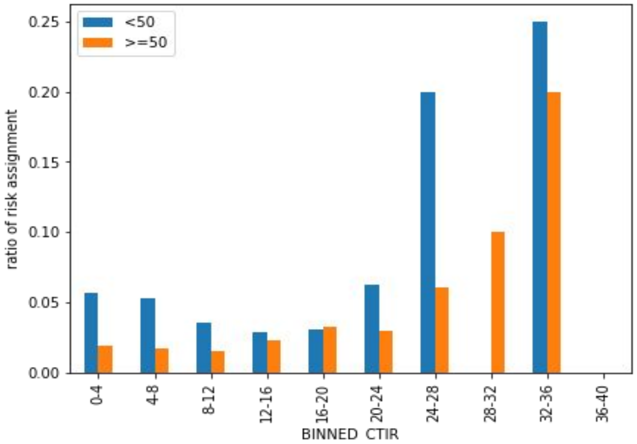


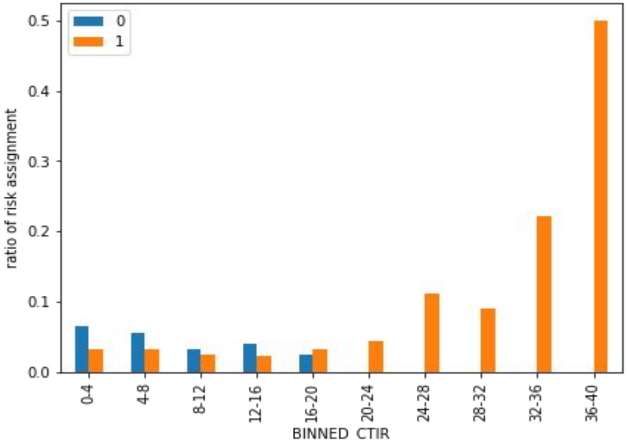
ADS Fairness: Outcome conditioned on an attribute

For the same CREDIT_TO_INCOME_RATIO bin, unprivileged subgroups are assigned higher risk.

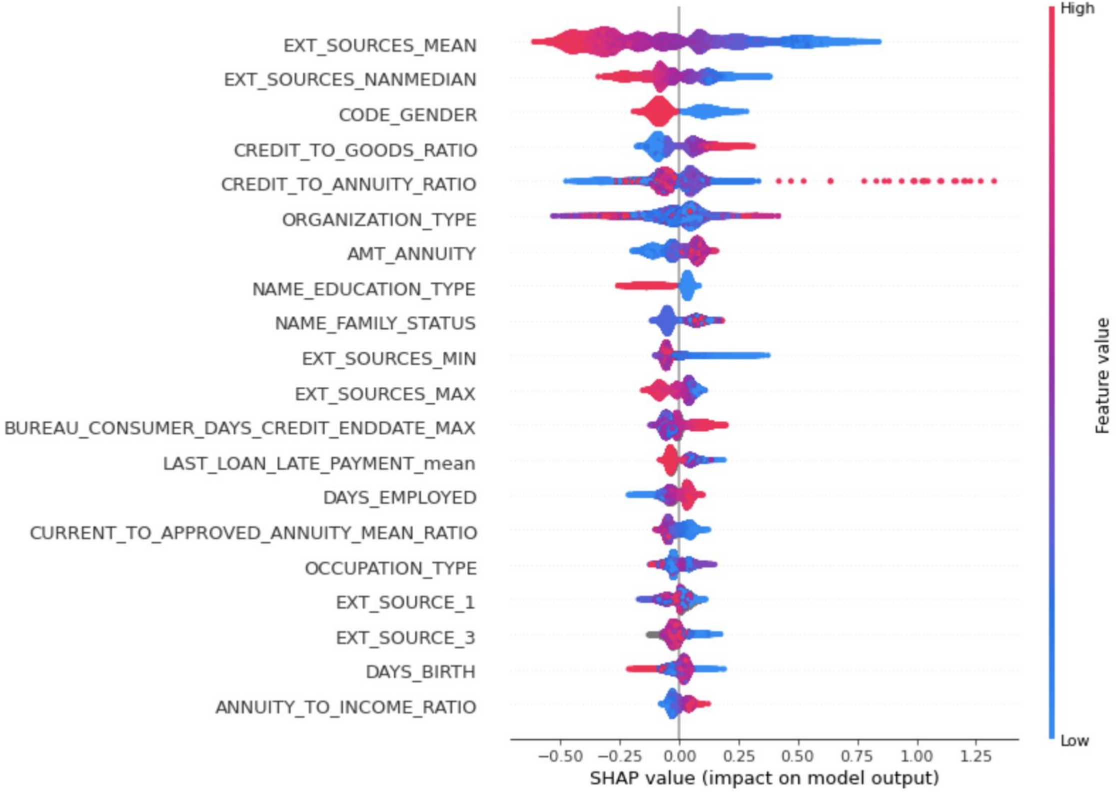
Other conditions tried: ANNUITY_TO_INCOME_RATIO, CREDIT_TO_ANNUITY_RATIO







Explaining ADS predictions



Conclusion

- ADS prediction is mainly driven by the score from external sources
- 660 features- numerical, non-numerical (encoded), aggregations, ratios, differences.
- Low accuracy for high-risk, high accuracy for low-risk
- Unbiased for gender; biased for age, employment duration, occupation type and rating of region where client lives.
- This ADS can be a helpful tool for support, but not for deployment on its own-- requires manual selection of a threshold