# CAPSTONE 2: LENDING CLUB LOAN DEFAULT PREDICTIONS

SIDNEY MCDERMOTT

# LENDING CLUB'S BUSINESS PROBLEM

**5%** of borrowers defaulted on loans between 2007 and 2017 equating to

**$8,695,172,153** in lost principal, interest, and fees
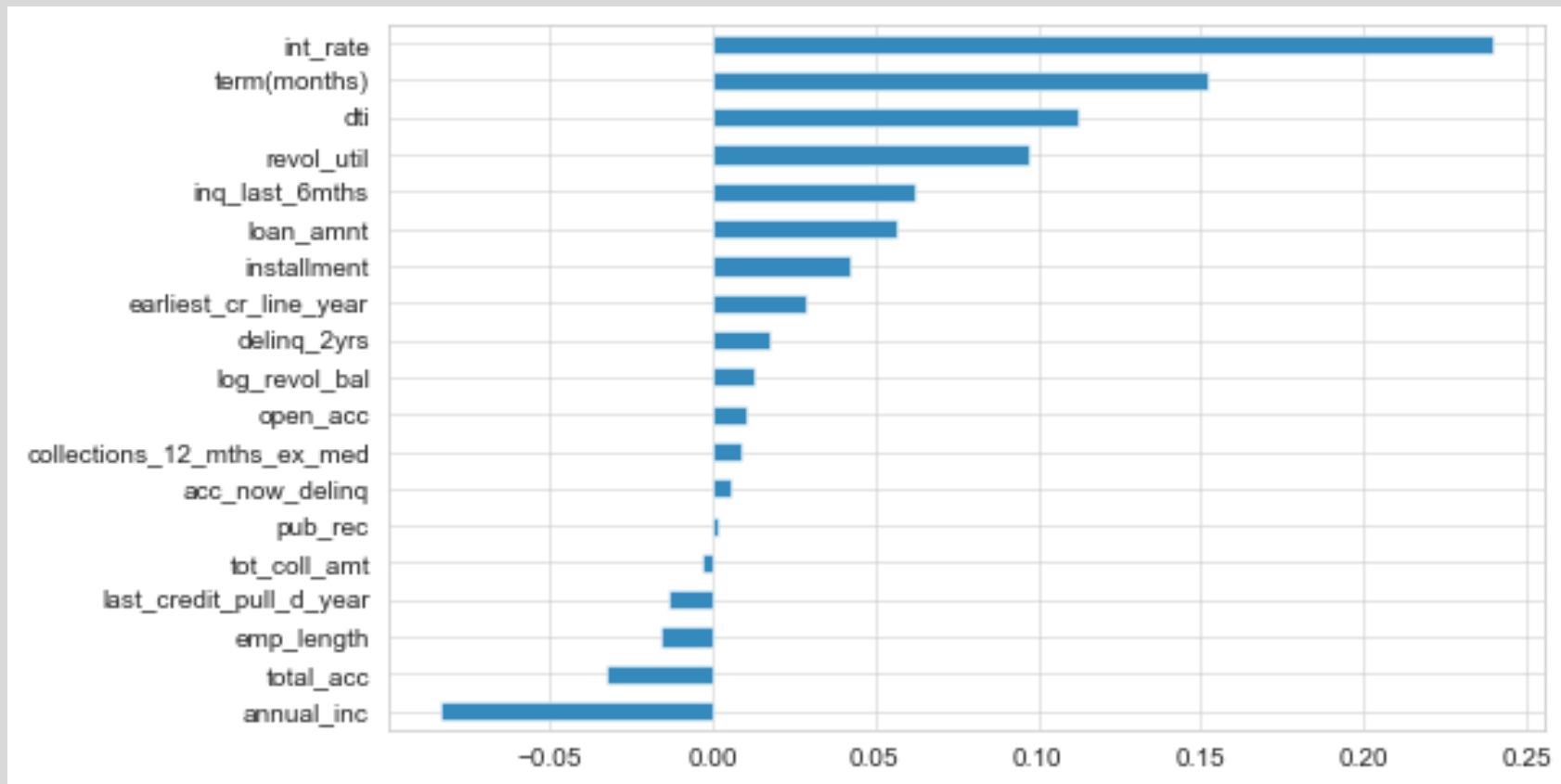
# WHO CARES ABOUT DEFAULTS?



**Lenders**

**Investors**

- Banks

- Credit Unions

# WHAT FEATURES INFLUENCE DEFAULTS?

# BEST MODEL PARAMETERS

RandomForestClassifier(

bootstrap = True,

max_depth = 40,
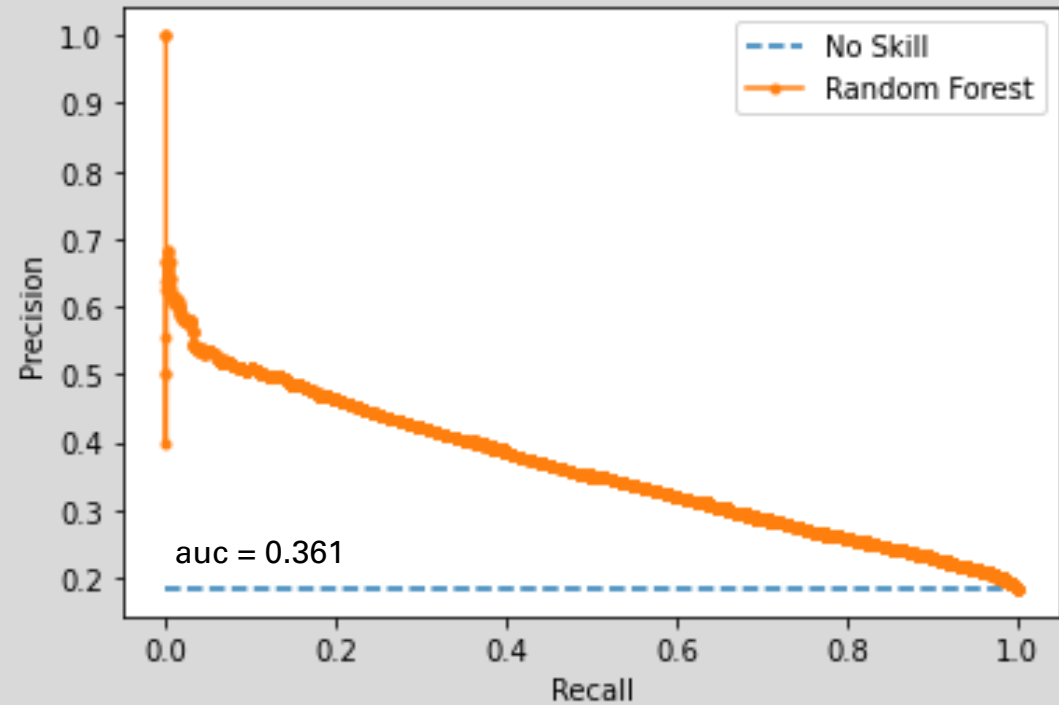
max_features = 'auto',

min_samples_split = 2,

n_estimators = 200)

### Random Forest Precision-Recall Curve on Test Set

# MODEL TRAINING RESULTS

Train Result:
================================================

CLASSIFICATION REPORT:

|  | 0 | 1 | accuracy | macro avg | weighted avg |
|---|---|---|---|---|---|
| precision | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| recall | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| f1-score | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| support | 122183 | 36655 | 1.0 | 158838 | 158838 |

_____

Confusion Matrix:
```
 [[122183     0]
 [    0  36655]]
```

Test Result:
================================================

CLASSIFICATION REPORT:

|  | 0 | 1 | accuracy | macro avg | weighted avg |
|---|---|---|---|---|---|
| precision | 0.829281 | 0.506489 | 0.816682 | 0.667885 | 0.769944 |
| recall | 0.976400 | 0.107539 | 0.816682 | 0.541970 | 0.816682 |
| f1-score | 0.896847 | 0.177410 | 0.816682 | 0.537129 | 0.764597 |
| support | 62839.00 | 14153.00 | 0.816682 | 76992.00 | 76992.00 |

_____

Confusion Matrix:
```
 [[61356  1483]
 [12631  1522]]
```

# MODELING RESULTS ON HOLDOUT DATA
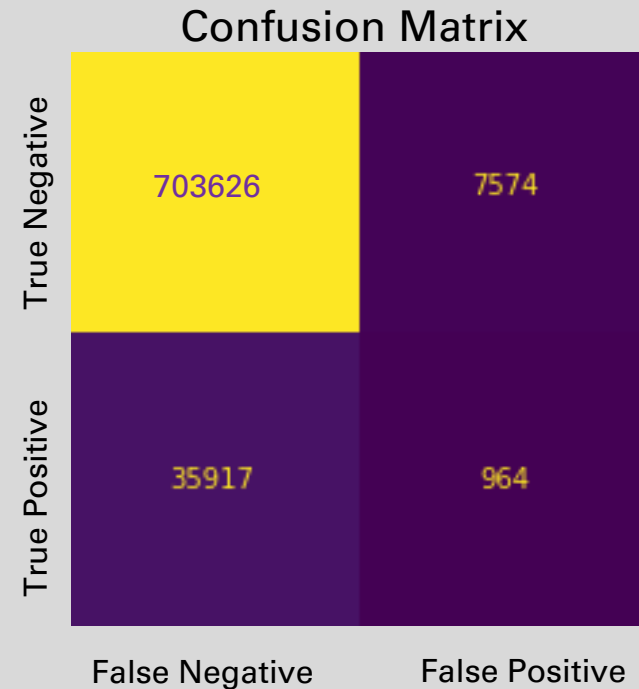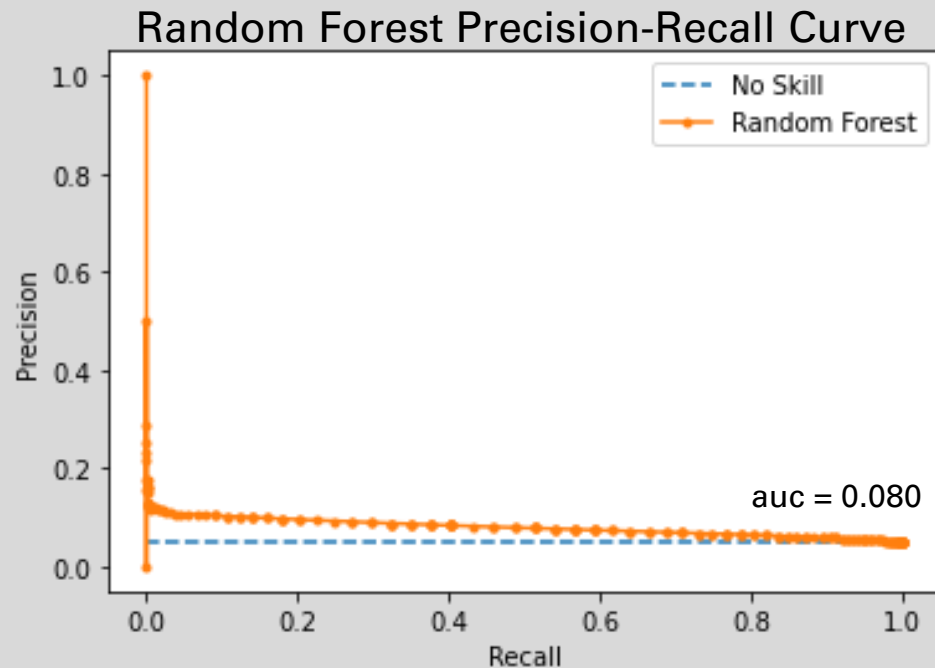


Model is only predicting 2.6% of true defaults on holdout data

# MODEL REPORT ON HOLDOUT DATA

| CLASSIFICATION REPORT | | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | accuracy | macro avg | weighted avg |
| precision | 0.951434 | 0.112907 | 0.941863 | 0.532170 | 0.910093 |
| recall | 0.989350 | 0.026138 | 0.941863 | 0.507744 | 0.941863 |
| f1-score | 0.970022 | 0.042449 | 0.941863 | 0.506235 | 0.924291 |
| support | 711200.0 | 36881.00 | 0.941863 | 748081.00 | 748081.00 |

# SUMMARY AND CONCLUSION

Model is not ready for business use

- Predicting more false defaults than true defaults leads to missed revenue by rejecting loan applications for loans that would be paid back

# FUTURE WORK TO IMPROVE MODEL PERFORMANCE

➤ Collect additional data e.g., FICO scores, or data Lending Club already uses before issuing loans

➤ Engineer better features e.g., applicant's debt to income

➤ Identify cause of the number of outlier and missing values in the data in order to start with better data