Machine Learning:

Generating Loan Interest Rates

UCB Data Analytics Final Project

LENDINGCLUB DATA

We used the data collected in Project 2 to train and test our Machine Learning models. We wanted to see if we had enough data to train our model to produce an accurate interest rate if an applicant input their personal financial information.

Maya Reiss

Bonnie Shen

Lauralynn Stubler

HYPOTHESIS

We aimed to build a tool to determine loan details for a potential applicant of LendingClub.

With the data we gathered, we thought we could build a tool to determine loan elegibility.

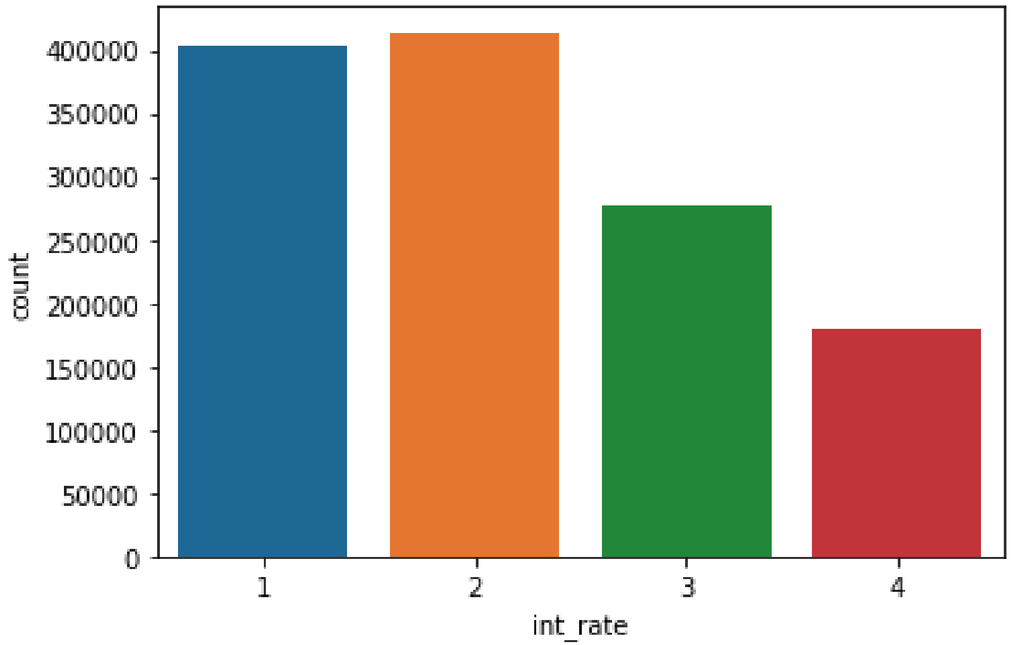
Could the applicant get a loan?

Maximum loan amount approved for?

Interest rate on loan based on applicant's financial information?

Y = INTEREST RATE

```
1  import seaborn as sns
2  sns.countplot(df["int_rate"], label="Count")
3  plt.show()
```



1 = A-grade: 0 - 9.99% interest rate

2 = B-grade: 10 - 13.99%

3 = C Grade: 14 - 17.99%

4 = D Grade: 18% and up

120 Columns, 1.5M rows

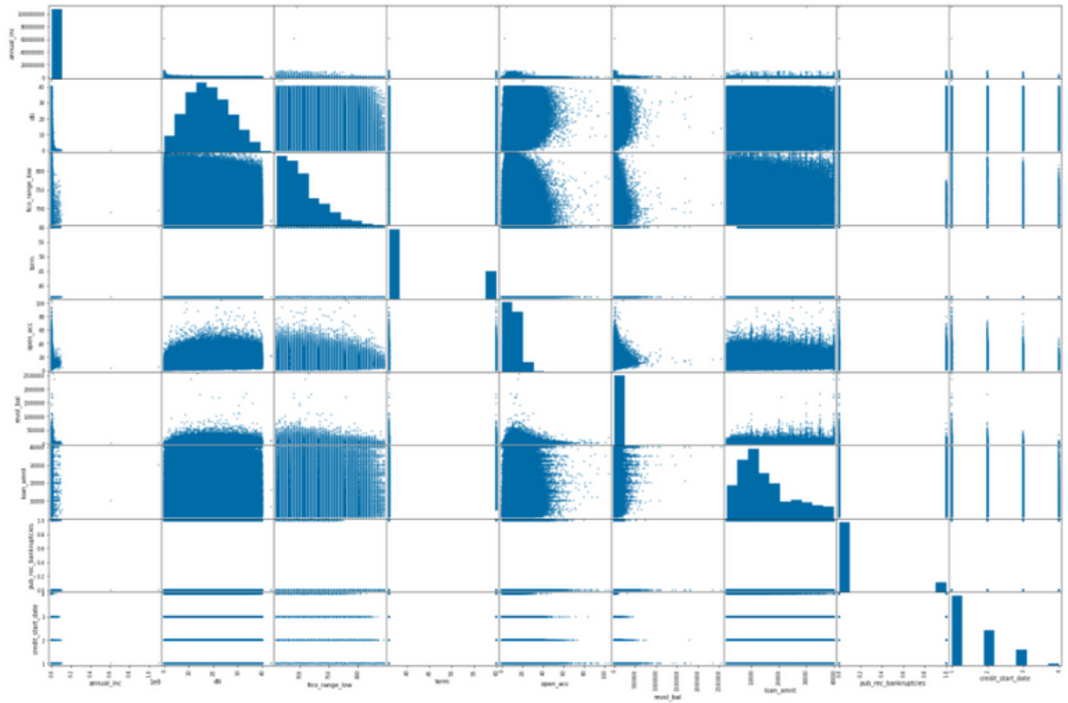
Removed joint applicants

Selected top 20 relevant columns

Cleaned data, i.e. filled in Nans
changed booleans + dates

Scatter matrix

9 features



TRAINING AND TESTING THE MODELS

We scaled our data using StandardScaler

We ran the X data through different models such as Logistic Regression, Decision Tree Classifier, SVM, Random Forest, KNeighbor Classifier and more

1.27M

Total Rows of Data

cleaned and used to train models

96+

Hours Spent

trying to run one of the longest models

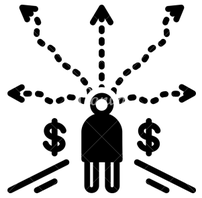
FINAL CHOSEN FEATURES

Annual Income	Term of Loan
Loan Amount	FICO Score

RESULTS

```
# Setting max decision tree depth to help avoid overfitting
from sklearn.tree import DecisionTreeClassifier
dt3 = DecisionTreeClassifier(max_depth=10)
dt3.fit(X_train_scaled, y_train)
print('Accuracy of Decision tree on training', dt3.score(X_train_scaled, y_train))
print('Accuracy of Decision tree on testing', dt3.score(X_test_scaled, y_test))
```

```
Accuracy of Decision tree on training 0.4753040667156607
Accuracy of Decision tree on testing 0.47134476109118545
```



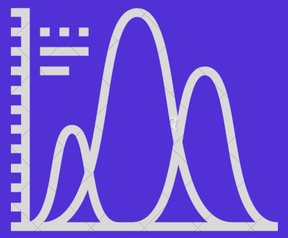
The most accurate model was a Decision Tree with a max depth of 10.

While we wish our accuracy scored about 50%, sadly this was not the case. Our only solace was that the testing data scored pretty accurate to the training model.

CHALLENGES

Practice makes progress.

While our models did not have great accuracy, we were excited to see the different accuracy scores each model produced. If we had more time to pull more data, we would eventually be able to train our model to be more accurate.



01	Data cleaning: selected columns - we had to make bin and make data binary.
02	Size of file slowed down our progress significantly - one model took over 96 hours to run.
03	Our initial goal was to get an accuracy score above 50% which we sadly could not reach.
04	Machine power limitations and ran out of time!

DEMONSTRATION

https://sheltered-citadel-81280.herokuapp.com/

QUESTIONS?

Thank you for your time!

Congratulations to everyone who completed this course and best of luck!