




LoanEvaluator.net

dn-ds

What is LoanEvaluator?

A web app that predicts the probability that a given LendingClub loan will be charged-off.

[Home](#) | [About](#)

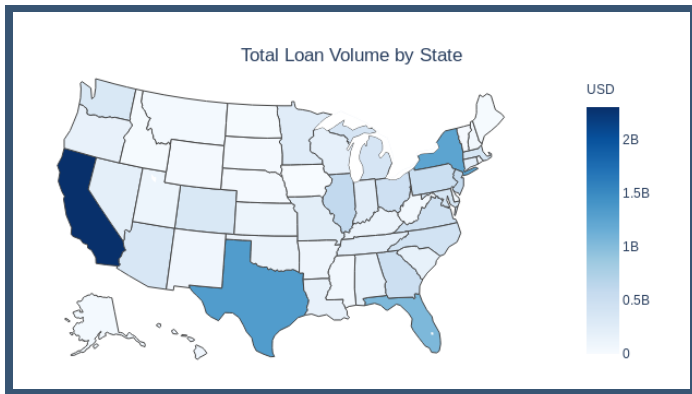
Predict the probability of charge-off of a LendingClub loan:

Loan Amount \$1000 to \$40000	Annual Income	Credit Utilization %
Terms*	Income Verification	# Credit Lines
Interest Rate*	Employment Length	# Open Credit Lines
Monthly Installment	Home Ownership	# Mortgage Accounts
Purpose	Debt Payment to Income %*	# Derogatory Records
Subgrade*	Fico Score*	# Bankruptcies
Application Type	Earliest Credit Account e.g., Sep-2020	# Tax Liens
Initial List Status	Credit Balance	State Code e.g., IL

Make Prediction

What is LendingClub?

A peer-to-peer lending company that directly matches borrowers and investors through an online platform. LendingClub claims to have issued loans totaling approximately \$60 billion, as of June 2020.



The Dataset

- Downloaded from [kaggle/wordsforthewise](#)
- Size 2.5 GB
- 2.2 million rows
- 151 features
- Target variable: loan status ('Fully Paid', 'Charged-off')

Goal: Given loan details, predict the probability of charge-off.

Project Outline

Exploring and Cleaning the Data



Examining Relationships Between Features and the Target



Feature Engineering



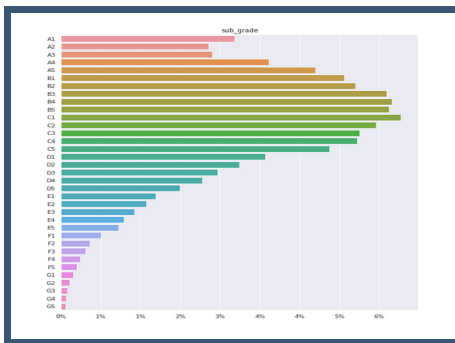
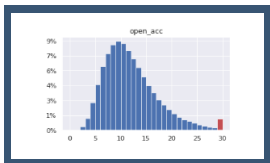
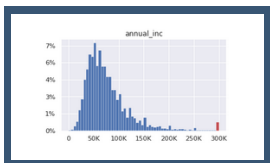
Training Machine Learning Models



Web App

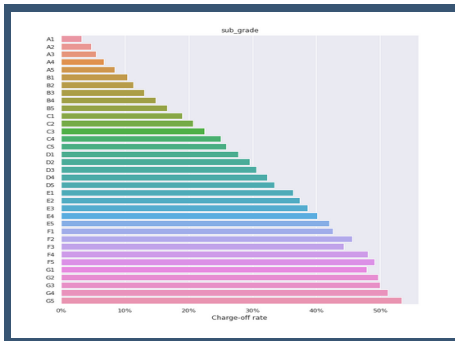
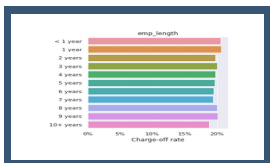
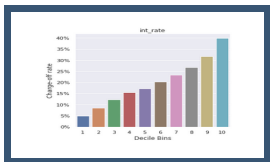
Exploring and Cleaning the Data

- Features that are unavailable to the potential investor at the time of investment are identified and dropped.
- Features that are missing more than 30% of the values are dropped.
- Numerical and categorical features are identified and studied.
- Distribution of each feature is studied.
- A test set is put aside.



Examining Relationships Between Features and the Target

- The potential usefulness of each numerical feature is determined by calculating charge-off rates for binned data, and by considering the Pearson and the Spearman correlation coefficients.
- The charge-off rate for each category of categorical features is determined. The gathered data helped determine the appropriate encoding (ordinal or one-hot) for the features.



Feature Engineering

- New features are engineered. Some perform better than some existing features.
- The most important features are determined and ranked:

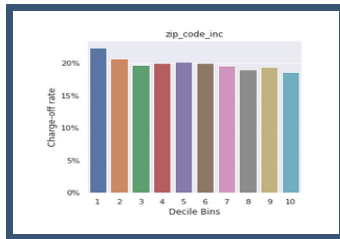
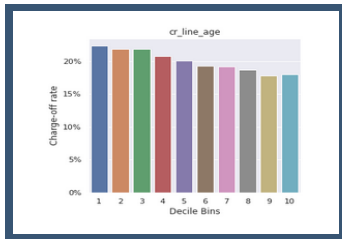
Sub grade

Interest rate

Term

Borrower's FICO score

Borrower's debt payment-to-income ratio.



Training Machine Learning Models

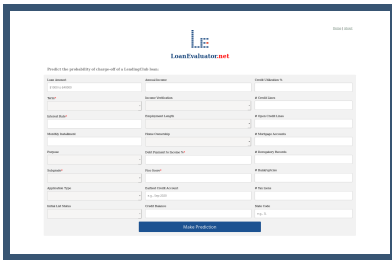
- The dataset is imbalanced: 80-20% split.
- Evaluation metrics used: **Precision-Recall AUC, ROC AUC.**
- A pipeline is created to perform the tasks of imputation, scaling, encoding categorical features, and feature engineering.
- Four models are considered:
 - Logistic Regression**
 - Random Forest**
 - Linear Discriminant Analysis**
 - K-Nearest Neighbors.**
- Overfitting is estimated using cross-validation.

Training Machine Learning Models (continued)

- Models are ranked by cross-validation score.
- Top models are selected, and their hyperparameter are tuned using a grid search.
- Final model: **Logistic Regression, with L2 regularization**.
Test set ROC AUC score: **0.71**.
- The Regression model has the added advantage that it is naturally well-calibrated in terms of output probabilities.
- Training was done on an AWS EC2 c5.9xlarge instance.

Web App

- When loan details are submitted, the information is preprocessed using jQuery and PHP, and then passed onto the machine learning model.
- The model processes the data and returns a prediction.
- The machine learning model is deployed on an AWS EC2 t2.micro instance using the Flask framework.



The screenshot shows the LoanEvaluater.net web application. At the top, there is a logo and the text "LoanEvaluater.net". Below the logo, it says "Predict the probability of charge-off of a Loan/EvalRisk Score". The form is divided into several sections with input fields and dropdown menus. The sections include: "Loan Details" (Loan Amount, Term, Interest Rate, Monthly Repayment, Frequency, Repayment Period, Application Type, Initial Loan Status), "Borrower Details" (Borrower Name, Employment Length, Education, Debt-to-Income Ratio, Credit Score, Annual Income, Credit History, Credit Score, Credit Score, Credit Score), "Loan Details" (Loan Amount, Term, Interest Rate, Monthly Repayment, Frequency, Repayment Period, Application Type, Initial Loan Status), and "Loan Details" (Loan Amount, Term, Interest Rate, Monthly Repayment, Frequency, Repayment Period, Application Type, Initial Loan Status). At the bottom, there is a "Make Prediction" button.



Main Tools and Packages Used

