Description of data

Below is a short description of each feature in the data set:

* ****credit\_policy****: 1 if the customer meets the credit underwriting criteria of LendingClub.com, and 0 otherwise.
* ****purpose****: The purpose of the loan such as: credit\_card, debt\_consolidation, etc.
* ****int\_rate****: The interest rate of the loan (proportion).
* ****installment****: The monthly installments ($) owed by the borrower if the loan is funded.
* ****log\_annual\_inc****: The natural log of the annual income of the borrower.
* ****dti****: The debt-to-income ratio of the borrower.
* ****fico****: The FICO credit score of the borrower.
* ****days\_with\_cr\_line****: The number of days the borrower has had a credit line.
* ****revol\_bal****: The borrower’s revolving balance.
* ****revol\_util****: The borrower’s revolving line utilization rate.
* ****inq\_last\_6mths****: The borrower’s number of inquiries by creditors in the last 6 months.
* ****delinq\_2yrs****: The number of times the borrower had been 30+ days past due on a payment in the past 2 years.
* ****pub\_rec****: The borrower’s number of derogatory public records.
* ****not\_fully\_paid****: indicates whether the loan was not paid back in full (the borrower either defaulted or the borrower was deemed unlikely to pay it back).

Terms

Principal component analysis (PCA)

Linear dimensionality reduction using Singular Value Decomposition of the data to project it to a lower dimensional space.

**Training and Testing Set**

We have to do this because we will use “cross-validation” to measure the accuracy of our predictive model.

We will train our classifier on the training set and test it’s accuracy on the testing set.

Intuitively, if our classifier should classify credit risks in the testing set the same as in the real world. This makes the testing set a proxy to how it would behave in production.

**X\_1: independent (target) variables for first data set**

**y\_1: dependent (outcome) variable for first data set**

**X\_2: independent (target) variables for the second data set**

**y\_2: dependent (outcome) variable for the second data set**

**Classifier**

We will use the Naive Bayes classifier for building our model.

**It is a classification technique based on Bayes’ Theorem with an assu Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other featuremption of independence among predictors.**

**Save Classifier**

With our classifier done we can save it so that we can use it a separate program

**Create Web API**

With our model created, we can now create our web service that can decide if we should give credit to someone based on certain demographic information.

Using flask framework