**Predict Loan Defaults**

SIB(Small Industries Bank) loans money to companies in exchange for the promise of repayment.  Some will default on the loans, being unable to repay them for some reason. The bank maintains insurance to reduce their risk of loss in the event of default. The insured amount may cover all or just some part of the loan amount.  SIB wants to predict which companies will default on their loans based on their financial information. They have provided you with a dataset that consists of loan related information such as loan amount, term, and state. Also, there is company information such as the number of employees, operating sector, etc.

Using machine learning, predict which companies will default on their loans and explain how different features impact the predictions.

**Files**

* train.csv
* test.csv

**Problem**

Analyze the given data and learn how different features are related to and affect *default\_status*. With the given data, build a machine learning model which can be used to predict the ‘*default\_status'*.

For each record in the test set (test.csv), you must predict the value of the *default\_status* variable. You should submit a CSV file with a header row and one row per test entry. The file (submissions.csv) should have exactly 2 columns:

* id
* default\_status

**Evaluation Metric:**  
The metric used for evaluating the performance of the model would be Accuracy:  
    Accuracy= Number of correct Predictions / Total number of predictions

**Deliverables**

* Well commented **Jupyter notebook/Python Script/R Script** named ‘**code\_yourname’** in either IPYNB/PY/R format
* For submissions NOT using Jupyter notebook, save and clearly label your visualization output and zip it in a folder named **‘visualization\_yourname.zip’**
* Test prediction named **'submissions\_yourname.csv'**

The notebook/script should contain your solution, visualizations (notebook), and thought process, including the top features that go into the model. If required, please generate new features. Make appropriate plots, annotate the notebook/script with comments/markdowns, and explain the necessary inferences. A person should be able to read your notebook/script and understand the steps are you taking and the reasoning behind them.