“Credit Risk Analysis”

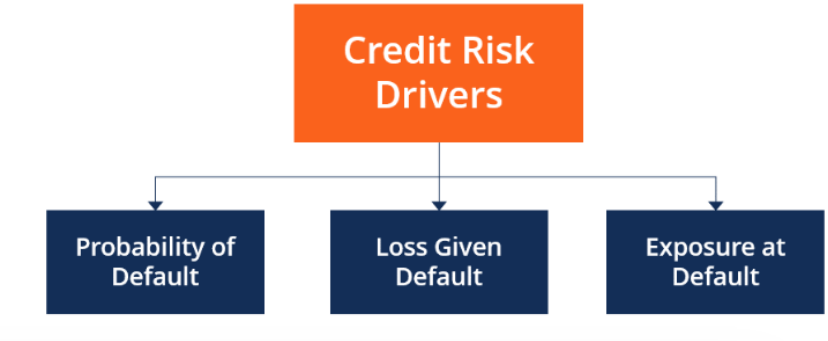
Anirudha Patil

Date: 10/01/2023

Problem Statement:

To determine A person's creditworthiness with the likelihood they may default on loans.

By balancing the costs and benefits of granting credit, lenders measure, analyze and manage risks their business is willing to accept.



Market/Customer/Business Need Assessment:

Credit officers, Audit and risk management teams, credit analysts with Investment managers, Bank sector and financial industries which lend loans on different assets.

- Data is generated on regular basis in terms of online transactions.
- Use of this data in credit risk analysis is crucial to determine who is worth giving a credit.
- Banks, reeling under a mountain of bad debt, have accordingly started to work on ways to improve the efficacy of the credit assessment process and to price loans in relation to credit risk
- It starts with the core concepts and basic principles of credit risk and goes on to cover advanced concepts and their application in complex cases
- Credit risk analysis is an inspection performed by the credit team of an organization to determine the customer's ability to repay credit.
- Banks and other financial sectors are the target which needs credit risk analysis.

Target Specifications and Characterization:

There are two datasets one is related to client details and the other is related to the details of the credit/loan details.

- In client details there are details of age, marital status, address, property, education, employment and other related details
- In credit details there are details of loan id, purpose, months loan taken for, existing emi and loans, and his/her guarantor

So using the above features and details we need to determine the client's worthy of giving credit and build a ML model that predicts the best fit Client

External Search:

The dataset is taken from the organization, and given for analysis.

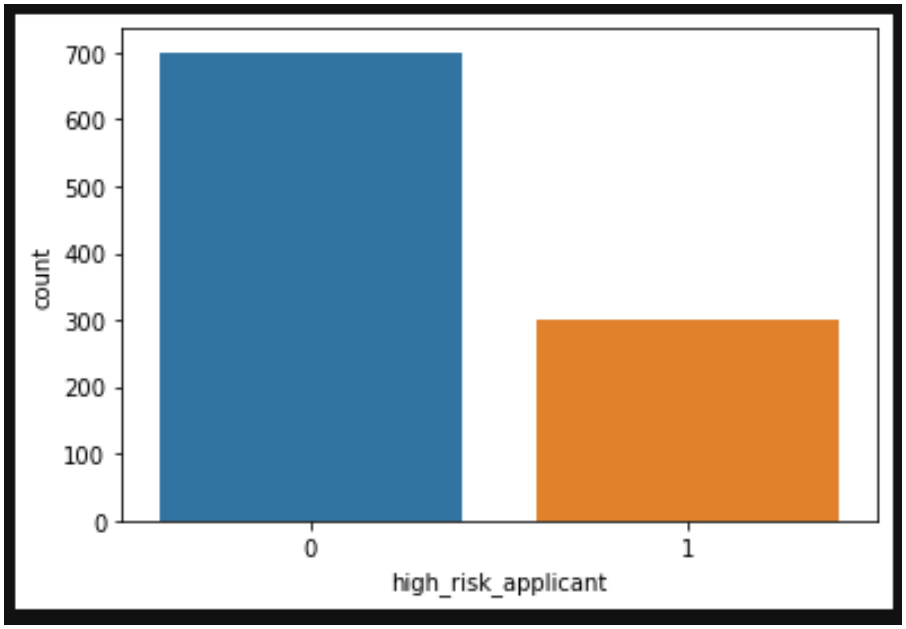
Client details:

	applicant_id	Primary_applicant_age_in_years	Gender	Marital_status	Number_of_dependents	Housing	Years_at_current_residence	Employment_status
0	1469590	67	male	single	1	own	4	skilled employee / official
1	1203873	22	female	divorced/separated/married	1	own	2	skilled employee / official
2	1432761	49	male	single	2	own	3	unskilled - resident
3	1207582	45	male	single	2	for free	4	skilled employee / official
4	1674436	53	male	single	2	for free	4	skilled employee / official

Credit/Loan details

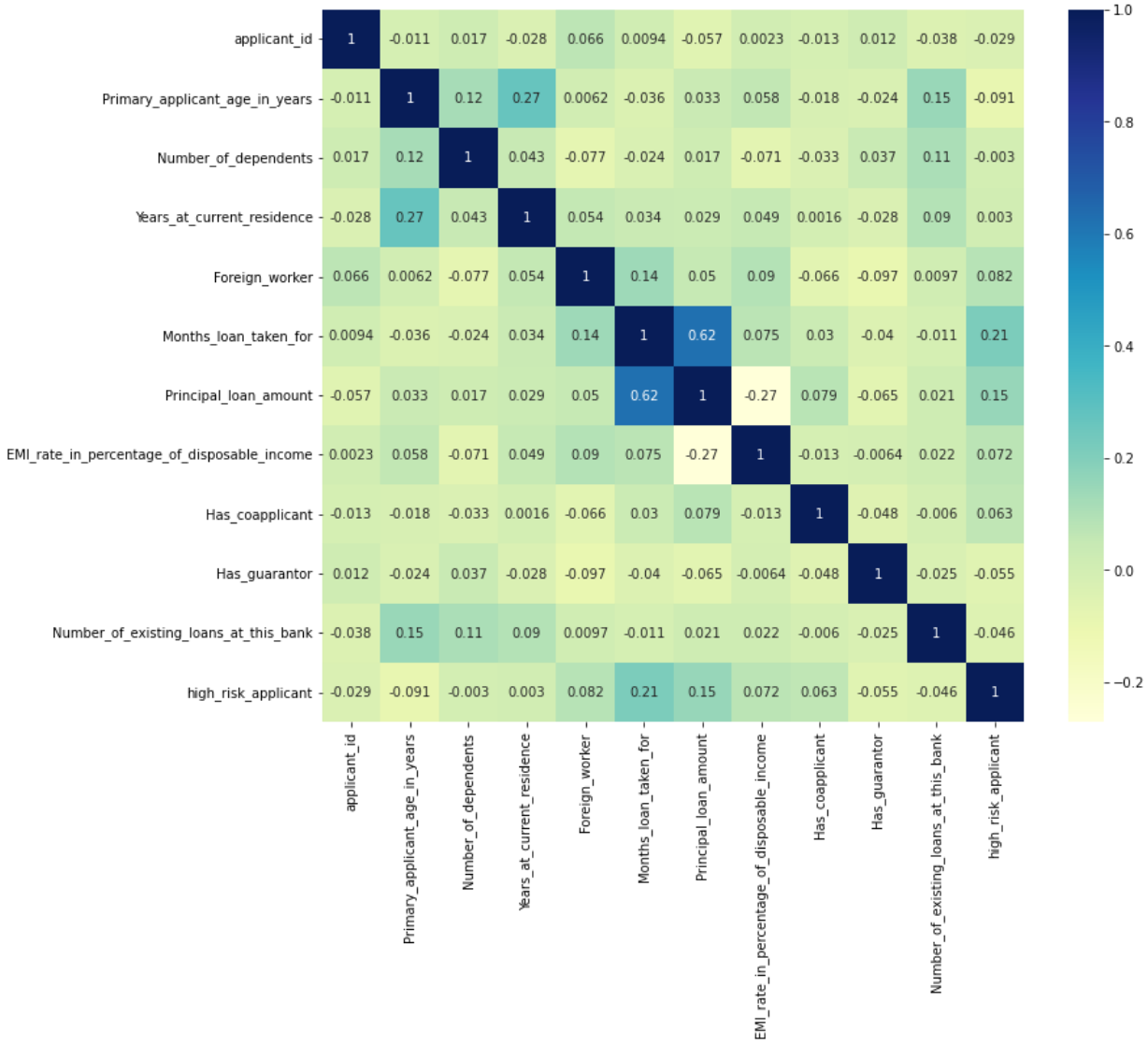
rate_in_percentage_of_disposable_income	Property	Has_coapplicant	Has_guarantor	Other_EMI_plans	Number_of_existing_loans_at_this_bank	Loan_history
4	real estate	0	0	NaN	2	critical/pending loans at other banks
2	real estate	0	0	NaN	1	existing loans paid back duly till now
2	real estate	0	0	NaN	1	critical/pending loans at other banks
2	building society savings agreement/life insurance	0	1	NaN	1	existing loans paid back duly till now
3	NaN	0	0	NaN	2	delay in paying off loans in the

Bench marking:



Usually data for credit risk analysis is always imbalanced i.e., The risk applicants are comparatively very low. So we need to build a model keeping this constraint in mind.

In above figure we can see that the data is imbalanced. In such cases the models accuracy doesn't play a good role rather we need to focus on its performance using F1 Score and Confusion matrix.



The correlation between the details of credit data can be understood from the above image.

Applicable Patents:

Method, system and program for credit risk management utilizing credit exposure Applicable Regulations with patent no/id WO2005029274A3 can be used to analyze and build our model for credit risk management.

Applicable Regulations

This is a continuation of U.S. patent application Ser. No. 10/942,185, filed on Sep. 16, 2004, entitled, “Method, System And Program For Credit Risk Management Utilizing Credit Limits”, which claims the benefit of priority under 35 U.S.C. §119(e) from U.S. Provisional Application No. 60/503,429, entitled, “Method, System and Program for Credit Risk Management Utilizing Credit Exposure,” filed on Sep. 16, 2003, and U.S. Provisional Application No. 60/503,422, entitled, “Method, System and Program for Credit Risk Management Utilizing Credit Limits,” filed on Sep. 16, 2003, which disclosures are incorporated herein by reference.

The present application is related to co-pending U.S. patent application Ser. No. 10/942,196, filed on even date herewith and assigned to the assignee hereof, and incorporated herein by reference in its entirety.

Applicable Constraints:

- Use of python, and its libraries for the exploration of the data
- Use of frameworks like Django for the deployment
- Use of cloud platform like azure and aws
- Machine learning algorithms for predictions
- Use of visualization tools like Tableau / Power Bi for report and dashboard generation

Business Opportunity

The opportunity can be seen in all the financial sectors that give credit/loan to different applicants based on some conditions. So if we build a model and help them analyze the risk free applicants for credit, definitely it would be a success.

Concept Generation

- Data is collected from the past credit applicants to analyse the relation
- This data is used for future predictions based on the features of the given current applicants
- Jupyter notebook is used for analyzing, cleaning and exploring the data
- Different python libraries like numpy, pandas, seaborn are used in the process

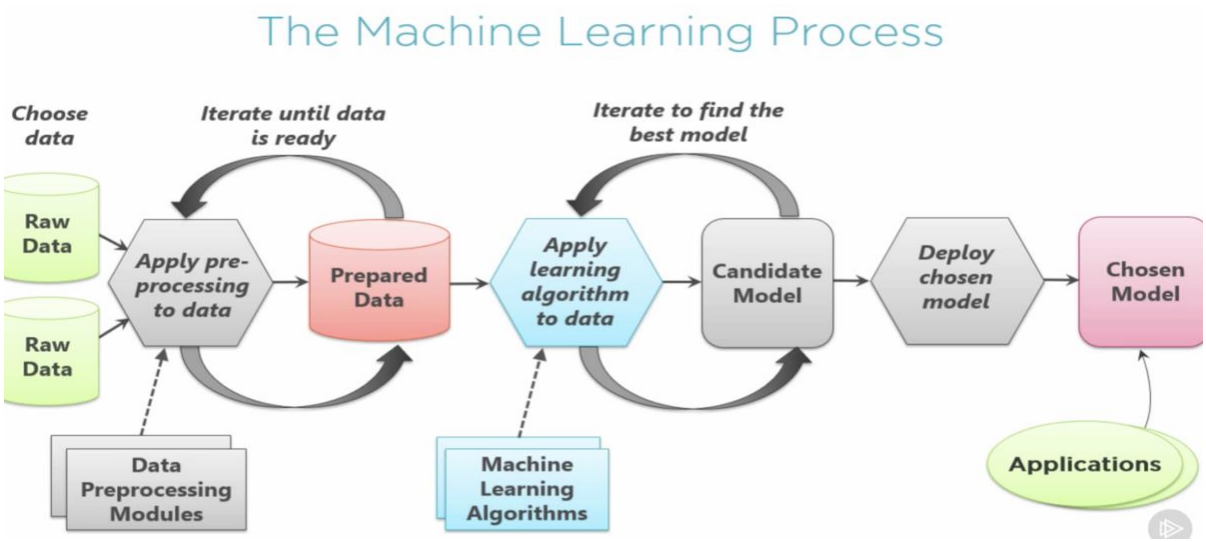
- Data is divided into train and test to build a model and check its performance
- Finally, algorithm is deployed to cloud to predict the credit risk applicants

```
def split(x,y):

    x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=.30,random_state=1)
#     print("Shape of x_train is:",x_train.shape,"; Shape of x_test is:" ,x_test.shape)
#     print("Shape of y_train is:",y_train.shape,"; Shape of y_test is:" ,y_test.shape)
    models={
    "LogisticRegression":LogisticRegression(n_jobs=10,max_iter=5,random_state=42),
    "DecisionTreeClassifier":DecisionTreeClassifier(criterion='gini',max_depth=3,random_state=3),
    "KNeighborsClassifier":KNeighborsClassifier(n_neighbors=28,p=1,n_jobs=1),
    "GaussianNB":GaussianNB(),
    "BaggingClassifier":BaggingClassifier(n_estimators=30,max_samples=30),
    "AdaBoostClassifier":AdaBoostClassifier(n_estimators=50,learning_rate=7,random_state=2),
    "GradientBoostingClassifier":GradientBoostingClassifier(n_estimators=10,max_depth=2,random_state=1),
    "RandomForestClassifier":RandomForestClassifier(n_estimators=50,random_state=1,max_samples=28),
    "Support vector classifier":SVC(),
    "XGBClassifier":XGBClassifier()
    }
    for name,model in models.items():
        model.fit(x_train,y_train)
        y_pred_train = model.predict(x_train)
        y_pred_test = model.predict(x_test)
#         confusion_matrix_of_test_data = pd.DataFrame(confusion_matrix(y_test, y_pred_test))
#         confusion_matrix_of_train_data = pd.DataFrame(confusion_matrix(y_train, y_pred_train))

        print('Accuracy score for test data is:', accuracy_score(y_test,y_pred_test))
        print(name,"test score is: ",model.score(x_test,y_test))
#         print('Confusion matrix of test data :\n',confusion_matrix_of_test_data)
        print('Accuracy score for train data is:', accuracy_score(y_train,y_pred_train))
        print(name,"train score is: ",model.score(x_train,y_train))
        print("-------------------------------------------------------------------------------")



x=trans_data
y=Final_data['high_risk_applicant']
split(x,y)
```

Different algorithms are used to determine the best fit model.

Concept Development:

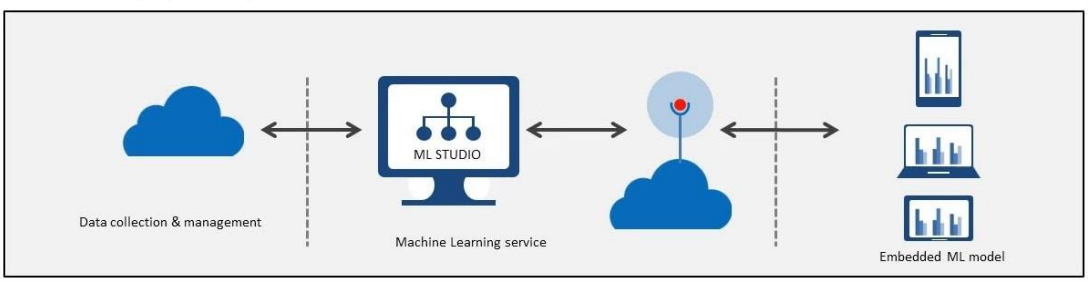
- Get the data
- Clean and process the data
- Bring in structured format
- Explore and find hidden relations
- Build the algorithm
- Test and validate the model
- Deploy and use.



Final Report Prototype:

The product can be prototyped in following way:

- The Ml model is based on the data gathered from past transactions and tested against some test cases
- After the testing, the model is validated and deployed in cloud and tested
- Again it is checked for quality score and deployed in production for real time use
- Now whenever the client gives it details the data is fed to the model and detects whether the client is risk free or not



Product details:

- After deployment, the data is fed to the ML model and prediction is done for the risk management
- Usually it is handled by the financial service providers and the handler just needs to feed the client data to check his/her eligibility for the credit

Code Implementation

The code is implemented here till the model Building and its performance optimization. Deployment is not done for the model.

GIT Link:

https://github.com/patilanirudh/Analysing_credit_risk/blob/main/Credit_Risk_Analysis.ipynb

References:

We can use different datasets available from kaggle for credit risk analysis and other ML models study in different field of study

https://www.kaggle.com/

You can get many datasets for the case study here.

Conclusion:

- Data can be used in many applications
- A good use of data can help in building ML models and solve some real world problems
- Credit risk analysis is necessary as it is crucial aspect in today's world
- Use of optimized and hypertuned ML algorithm gives good result