LOAN STATUS APPROVAL PREDICTION USING IBM WATSON

1.INTRODUCTION

1.1 Overview

Aim of the project is to predict the loan status according to the sample data given. Predicting loan defaulters is an important process of the banking system as it directly affects profitability. However, loan default data sets available are highly imbalanced which results in poor performance of the algorithms. As a result, the loan prediction machine learning model can be used to assess a customer's loan status and build strategies. Loan Prediction system is a system which provides you a interface for loan approval to the applicants application of loan. Applicants provides the system about their personal information and according to their information system gives his status of availability of loan. This model extracts and introduces the essential features of a borrower that influence the customer's loan status. Finally, it produces the planned performance (loan status).

1.2 Purpose

Using IBM Watson Studio we train the data set using Random forest Regression algorithm that help to train the model with the help of machine learning services provided by the IBM. Using the dataset which have the existing sample data of the quality determining experimented values, Machine learn and study The loan status is used for creating a credit scoring model. The credit scoring model is used for accurate analysis of credit data to find defaulters and valid customers. The objective of this paper is to create a credit scoring model for credit data.

2.LITERATURE SURVEY

2.1 Existing problem

Like dream housing finance company deals in all home loans. They have presence across all urban, semi urban and rural areas. Customer first apply for home loan after that company validates the customer eligibility for loan the Company wants to automate the loan eligibility process (real time) based on customer detail provided while filling online application form. These details are term, Income, Loan Amount, Credit History and others. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers. The above problem is a clear classification problem as we need to classify whether the Loan_Status is yes or no. So this can be solved by any of the classification techniques like

1. Logistic Regression .
2. Decision Tree Algorithm.
3. Random Forest Technique.

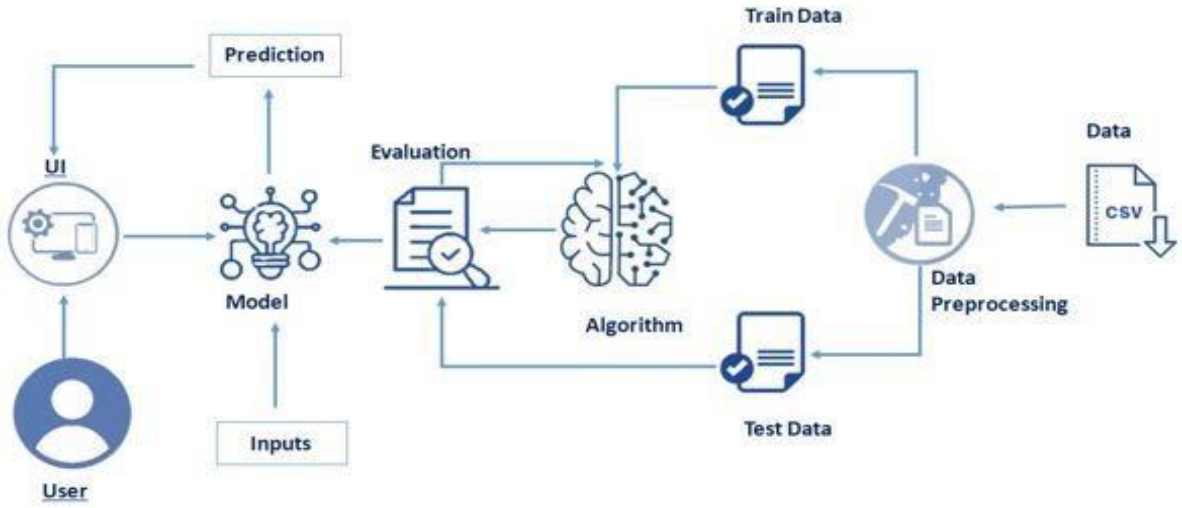
2.2 Proposed solution

To eliminate all these difficulties and problems we proposed a system that predict the quality using Machine Learning services that we train the data from the past data that helps to predict the result very fast and accurate by using the random forest regression algorithm. It is done by predicting if the loan can be given to that person on the basis of various parameters like credit score, income, amount, term etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.

As a result, this project aims to develop a Machine Learning (ML) model Loan Prediction system which provides you a interface for loan approval to the applicants application of loan. Applicants provides the system about their personal information and according to their information system gives his status of availability of loan.

3. THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software designing

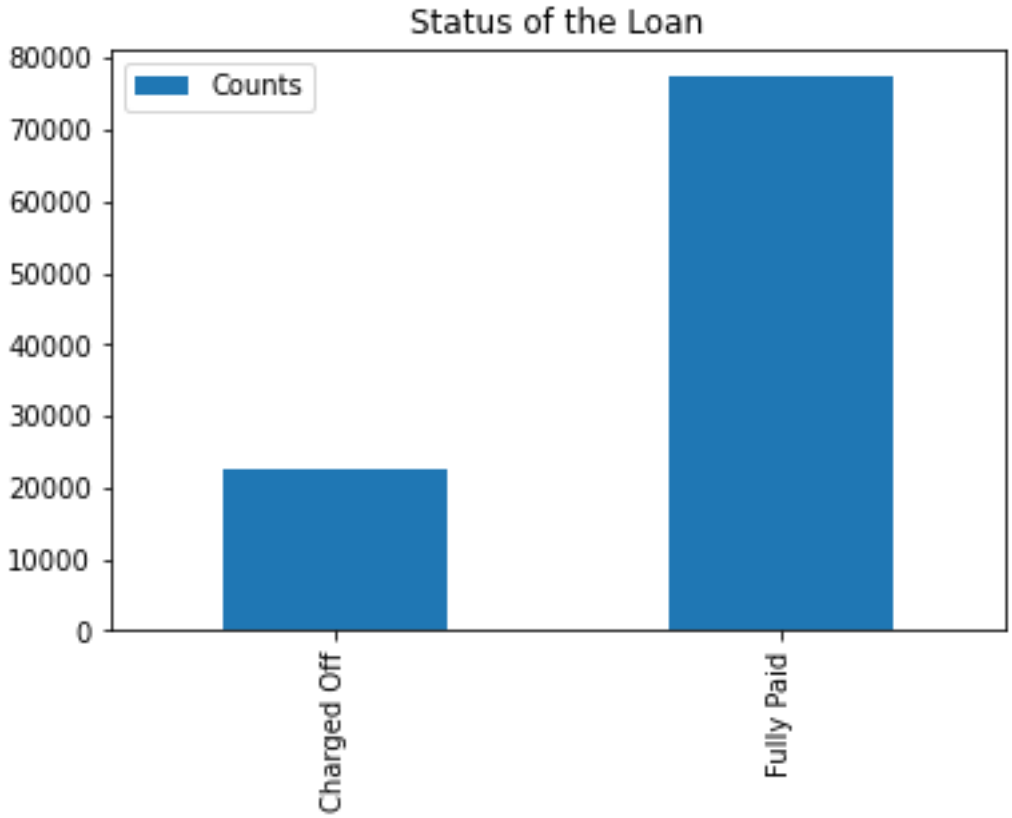
IBM Watson Studio - IBM Watson Studio helps data scientists and analysts prepare data and build models at scale across any cloud. IBM Watson Machine Learning - IBM Watson Machine Learning helps data scientists and developers accelerate AI and machine learning deployment. IBM Cloud Object Storage - IBM Cloud Object Storage makes it possible to store practically limitless amounts of data, simply and cost effectively. Machine Learning Services - Machine learning as service is an umbrella term for collection of various cloud-based platforms that use machine learning tools to provide solutions that can help ML teams with: out-of-the box predictive analysis for various use cases, data pre-processing, model training and tuning,

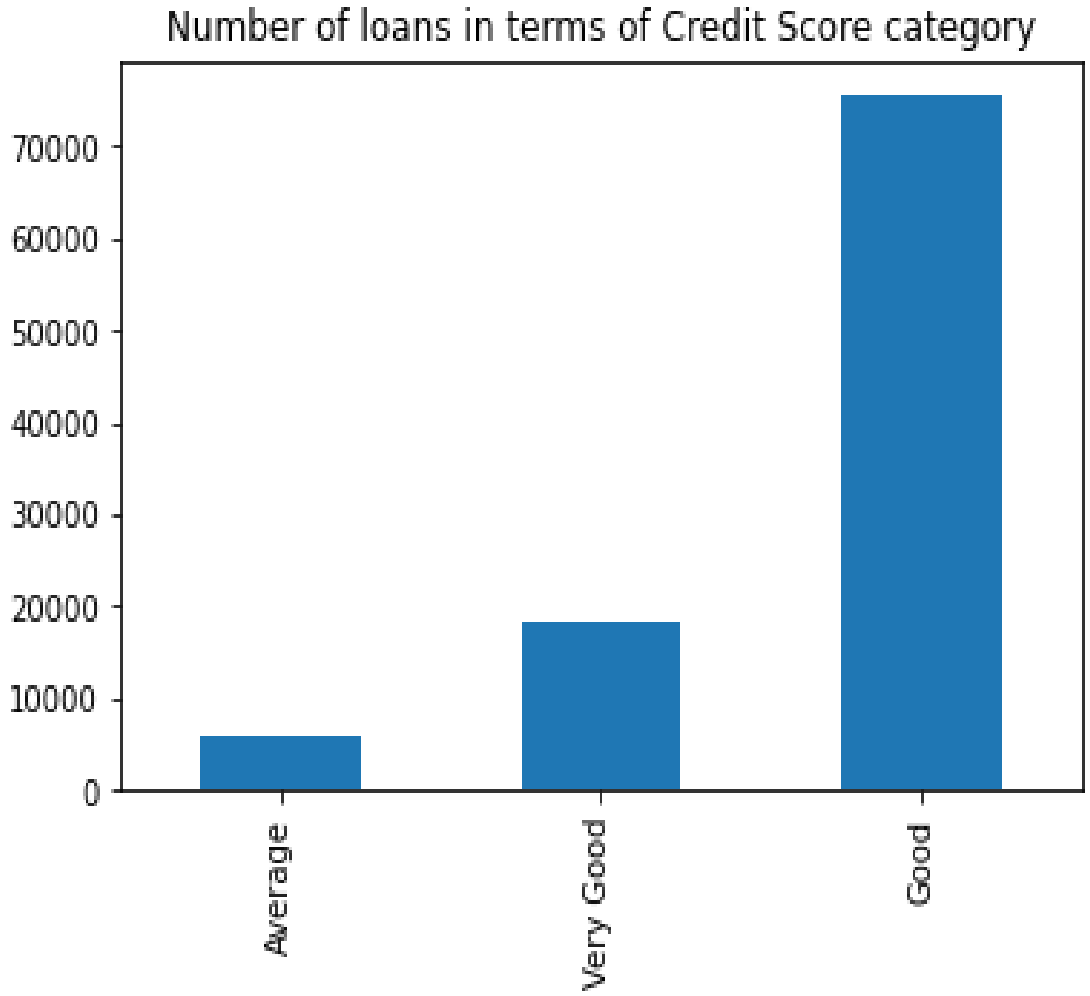
4 . EXPERIMENTAL INVESTIGATIONS

Here we are going to build a machine learning model that predicts the Status of the loan. . It is done by predicting if the loan can be given to that person on the basis of various parameters like credit score, income, amount, term etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.

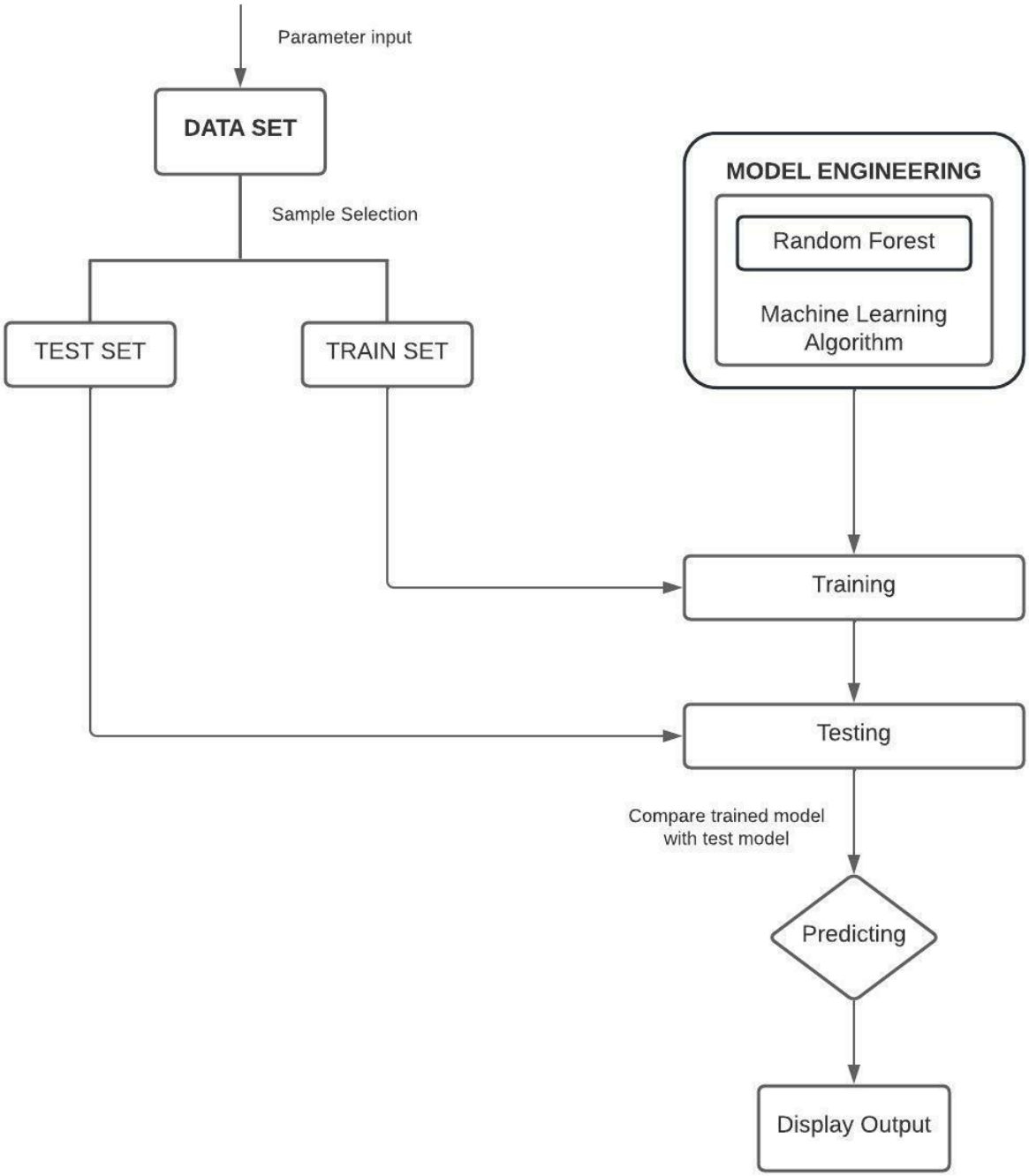
4.1 EXPERIMENTAL ANALYSIS

Visualization




Number of loans in terms of Credit Score category
70000
60000
50000
40000
30000
20000
10000
0
Average
Very Good
Good

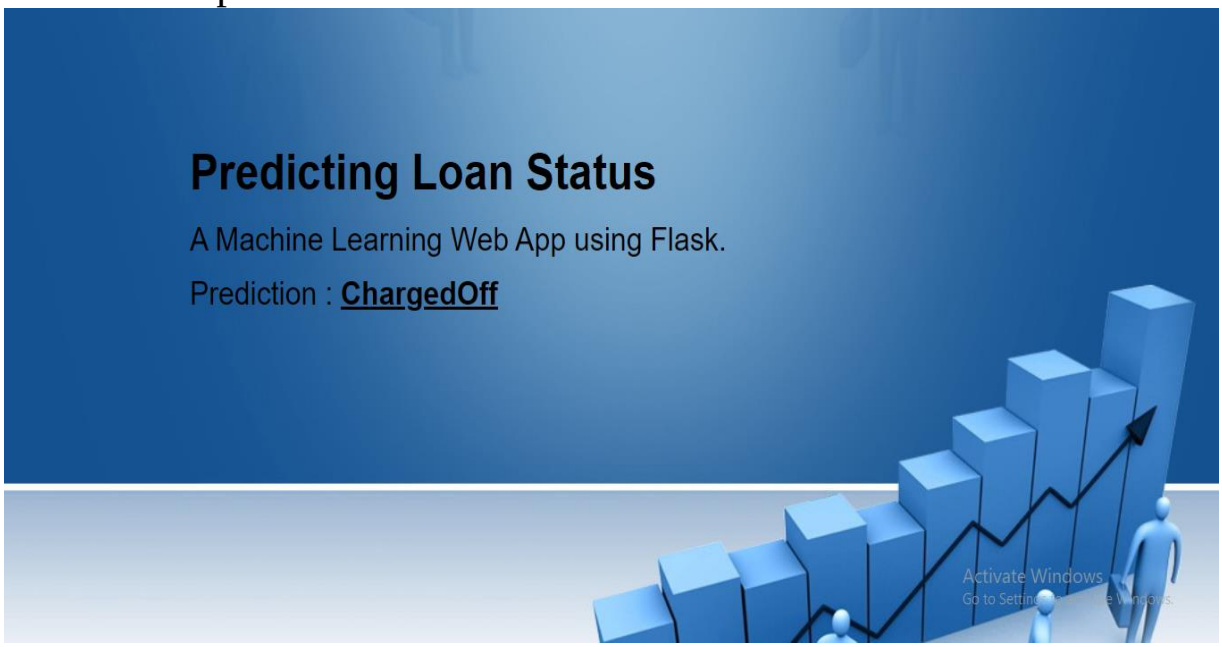
5.FLOWCHART


Parameter input
DATA SET
Sample Selection
TEST SET
TRAIN SET
MODEL ENGINEERING
Random Forest
Machine Learning Algorithm
Training
Testing
Compare trained model with test model
Predicting
Display Output

6.RESULT

When we input the values in each input fields amount, credit, income etc

Else the output will be.



Here we got the prediction of loan status.

7. ADVANTAGES & DISADVANTAGES

7.1 Advantages

Faster Claim Settlements & Save time:
We can predict the output very fast compared to other system and it save time due to machine learning so we can predict the quality of
the water using the machine learning techniques
Cost Efficiency:
Due to automation of everything therefore the cost is reduced compared to the manual process

Automation of Everything
Machine Learning is responsible for cutting the workload and time. By automating things we let the algorithm do the hard work for us. Automation is now being done almost everywhere. The reason is that it is very reliable. Also, it helps us to think more creatively.
Due to ML, we are now designing more advanced computers. These computers can handle various Machine Learning models and algorithms efficiently. Even though automation is spreading fast, we still don't completely rely on it. ML is slowly transforming the industry with its automation.

Wide Range of Applications

ML has a wide variety of applications. This means that we can apply ML on any of the major fields. ML has its role everywhere from medical, business, banking to science and tech. This helps to create more opportunities.

Scope of Improvement

Machine Learning is the type of technology that keeps on evolving. There is a lot of scope in ML to become the top technology in the future. The reason is, it has a lot of research areas in it. This helps us to improve both hardware and software.

In hardware, we have various laptops and GPUs. These have various ML and Deep Learning networks in them. These help in the faster processing power of the system. When it comes to software we have various UIs and libraries in use. These help in designing more efficient algorithms.

Efficient Handling of Data

Machine Learning has many factors that make it reliable. One of them is data handling. ML plays the biggest role when it comes to data at this time. It can handle any type of data.

Machine Learning can be multidimensional or different types of data. It can process and analyze these data that normal systems can't. Data is the most important part of any Machine Learning model. Also, studying and handling of data is a field in itself.

7.2 Disadvantages

- Possibility of High Error

In ML, we can choose the algorithms based on accurate results. For that, we have to run the results on every algorithm. The main problem occurs in the training and testing of data. The data is huge, so sometimes removing errors becomes nearly impossible. These errors can cause a headache to users. Since the data is huge, the errors take a lot of time to resolve.

- Algorithm Selection

The selection of an algorithm in Machine Learning is still a manual job. We have to run and test our data in all the algorithms. After that only

we can decide what algorithm we want. We choose them on the basis of result accuracy. The process is very much time-consuming.

Data Acquisition

In ML, we constantly work on data. We take a huge amount of data for training and testing. This process can sometimes cause data inconsistency. The reason is some data constantly keep on updating. So, we have to wait for the new data to arrive. If not, the old and new data might give different results. That is not a good sign for an algorithm.

Time and Space

Many ML algorithms might take more time than you think. Even if it's the best algorithm it might sometimes surprise you. If your data is large and advanced, the system will take time. This may sometimes cause the consumption of more CPU power. Even with GPUs alongside, it sometimes becomes hectic. Also, the data might use more than the allotted space.

8. APPLICATIONS

The credit scoring model is used for accurate analysis of credit data to find defaulters and valid customers. The objective of this paper is to create a credit scoring model for credit data. To automate this process, they have given a problem to identify the customers segments, those are eligible for loan amount so that they can specifically target these customers. The above problem is a clear classification problem as we need to classify whether the Loan_Status is yes or no. So this can be solved by any of the classification techniques like Logistic Regression ,Decision Tree Algorithm,Random Forest Technique.

9. CONCLUSION

Using this machine learning techniques and IBM Watson studio we are able to predict the loan status according to the sample data given. Predicting loan defaulters is an important process of the banking system as it directly affects profitability. However, loan default data sets available are highly imbalanced which results in poor performance of the algorithms. As a result, the loan prediction machine learning model can be used to assess a customer's loan status and build strategies. Loan Prediction system is a system which provides you a interface for loan approval to the applicants application of loan. Applicants provides the system about their personal information and according to their information system gives his status of availability of loan. This model extracts and introduces the essential features of a borrower that influence the customer's loan status

10. SCOPE

It is done by predicting if the loan can be given to that person on the basis of various parameters like credit score, income, age, marital status, gender, etc. The prediction model not only helps the applicant but also helps the bank by minimizing the risk and reducing the number of defaulters.

While machine learning artificial intelligence may be seen as a data hungry machine, the crucial aspect of a successful AI system that manages a client’s healthcare is its ability to develop efficient reasoning and intuitively read and understand trends.

So in the future we can add more parameters to predict and analyze the loan status as fast and accurate. In the future, more evolution algorithms, like differential evolution.

11.BIBILOGRAPHY

https://ieeexplore.ieee.org/document/7494106
https://www.datascience2000.in/2021/10/water-quality-prediction-usingmachine.html
https://iwaponline.com/wqrj/article/53/1/3/38171/Water-quality-predictionusing-machine-learning
https://www.researchgate.net/publication/351077205_Efficient_Water_Quality_Prediction_for_Indian_Rivers_Using_Machine_Learning
https://www.sciencedirect.com/science/article/abs/pii/S004896972200403X

APPENDIX
A. Source Code
Attach the code for the solution built.