IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11227

Loan Prediction System Using Machine Learning

Galiboina Akhil[1], Bachina sai Krishna[2], S.Srinivasa Rao[3]

Student, Computer Science and Engineering, KLUniversity, Guntur, India[1,2]

Associate Professor, Computer Science and Engineering, KLUniversity, Guntur, India[3]

Abstract: A lot of individuals are applying for bank loans but the bank has its limited assets which it's to grant to limited people only, so checking out to whom the loan is granted could be able to be a safer option for the bank is a typical process. So here we attempt to reduce this risk factor behind selecting the safe person so as to save many bank efforts and assets. This is done by mining the large Data of the previous records of the people to whom the loan was granted before and on the premise of these records/experiences the machine was trained using the machine learning model which gives the foremost accurate result. The foremost objective of this paper is to predict whether assigning the loan to a particular person is safe or not. This paper is split into four sections (i)Data Collection (ii) Comparison of machine learning models on collected data (iii) Training of system on most promising model (iv) Testing.

Keywords: Loan, Prediction, Safe Person, Machine Learning.

I.INTRODUCTION

Both Iris flower data set and Fisher's Iris data set are multivariate data sets. They are introduced by Ronald Fisher (British statistician and Biologist) in 1936. In his paper called "The use of multiple measurements in taxonomic problems as an example of linear discriminate analysis". Two of the three species are picked on the same day and measured at the same time by the same person with the same apparatus" and were collected in the Gaspé Peninsula "all from the same pasture. The three species of Iris are Iris setosa, Iris virginica, and Iris versicolor from which we collect 50 samples each to have a dataset. All the four features the length and the width of the sepals and petals were measured from each sample in centimeters. Fisher developed a linear discriminant model to distinguish the species from each other depending on the combination of these four features. The dataset contains two clusters with a separation. One of the clusters contains Iris setosa, while the other cluster contains the rest i.e., Iris virginica and Iris versicolor. It is not separable without the specific information that the fisher uses. The unsupervised procedures of nonlinear principal component analysis can separate three species of Iris (Iris setosa, Iris virginica, and Iris versicolor).

II.LITERATURE SURVEY

The related work on the application of machine learning and data approaches to studying financial data are comprehensively described below. Li et al. [6] conducted research on using attributes of customers to assess credit risk by using a weighted-selected attribute bagging method. They benchmarked their result experimentally by using two credit databases and reported outstanding performance both in terms of prediction accuracy and stability as compared with other state-of-the-art methods. A data mining approach is also proposed by Moro et al. [7] to predict the success or otherwise of a Portuguese retail bank in telemarketing. They applied various data mining models on the bank telemarketing data and reported that the neural network data mining method was the best for analyzing the data. The role of machine learning techniques in business data mining is outlined by [8]. Their work described the strengths and weaknesses of various machine learning techniques within the context of the business data mining approach. Their analysis revealed that Rule Induction Technique was the best approach in mining business data, followed by that of the neural network approach. C. Tsai and M. Chen [9] used a hybrid machine learning approach to study credit rating by comparing four different types of hybrid machine learning techniques. They showed experimentally that the 'classification + classification' hybrid model based on a combination of logistic regression and neural networks provides the highest prediction accuracy and also maximizes the profit. Bank default data was used by [10] to model bank failure predictions using a neural network approach. They compared their result with other machine learning approaches and concluded that the neural network approach is a promising method in terms of pre- dictive accuracy, adaptability, and robustness. [11] proposed a generalized switching hybrid recommendation algorithm by combining machine learning classifiers and collaborative filtering recommender systems. They experiment with the hybrid recommendation algorithms on two sets of data and reported high scalability and better performance in terms of accuracy and coverage. A hybrid online sequential extreme learning machine with the simplified hidden layer is proposed by [12]. The algorithm is a combination of the Online Sequential Extreme Learning Machine and the Minimal

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11227

Resource Allocation Network. Their experimental results showed that the algorithm has comparable performance as that of the original online sequential extreme learning machine but with a reduced number of hidden layers. Our approach used in this paper is complimentary but different in many ways. We employed diverse machine learning approaches to predict the creditworthiness of bank credit data. Comparative analysis is carried out to understand the best-fit algorithms for predicting the creditworthiness of a bank's credit data. Secondly, we extracted the 5 most important features that determine the creditworthiness of the data. Thirdly, we developed a predictive model using the ordinary linear regression approach. These approaches offer a better perspective on doing holistic analysis on bank credit data.

III.METHODOLOGY

A. EXISTING SYSTEM

Machine Learning implementation is a very complex part in terms of Data analytics. Working on the data which deals with prediction and making the code to predict the future of outcomes from the customer is a challenging part.

The different challenges in the existing system are Complexity in analyzing the data, Prediction is a challenging task working in the model, Coding is complex maintaining multiple methods and Library's support was not that much familiar.

B. TOOLS USED

We use Windows10 as Operating system. Python is used as Coding Language and PyCharm as a framework. Database related work is done using MySql. Server used is Flask.

C. PROPOSED SYSTEM

Python has a good area for data analytical which helps us in analysing the data with better models in data science. The libraries in python make the prediction for loan data and results with multiple terms considering all properties of the customer in terms of predicting.

The few Advantages of Proposed System are Libraries help to analyse the data, Statistical and prediction is very easy compared to existing technologies and Results will be accurate compared to other methodologies.

IV. CODE

```
import pandas as pd
import numpy as np
import matplotlib as plt
df = pd.read_csv(r"C:/Users/hp/Downloads/LoanPrediction/Project14 clp/data.csv")
print(df.head(10))
print(df.tail(10))
df.describe()
df['Property_Area'].value_counts()
import matplotlib.pyplot as plt
df['ApplicantIncome'].hist(bins=50)
plt.show()
df.boxplot(column='ApplicantIncome')
plt.show()
df.boxplot(column='ApplicantIncome', by = 'Education')
plt.show()
df['LoanAmount'].hist(bins=50)
plt.show()
df.boxplot(column='LoanAmount')
plt.show()
temp1 = df['Credit_History'].value_counts(ascending=True)
temp2 = df.pivot_table(values='Loan_Status',index=['Credit_History'],aggfunc=lambda x: x.map({'Y':1,'N':0}).mean())
print('Frequency Table for Credit History:')
```

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11227

```
print(temp1)
print('\nProbility of getting loan for each Credit History class:' )
print(temp2)

import matplotlib.pyplot as plt
fig = plt.figure(figsize=(8,4))
ax1 = fig.add_subplot(121)
ax1.set_xlabel('Credit_History')
ax1.set_ylabel('Count of Applicants')
ax1.set_title("Applicants by Credit_History")
temp1.plot(kind='bar')
plt.show()

ax2 = fig.add_subplot(122)
temp2.plot(kind = 'bar')
ax2.set_xlabel('Credit_History')
ax2.set_ylabel('Probability of getting loan')
ax2.set_title("Probability of getting loan by credit history/data.csv")
df.head(10)
df.describe()
df['Property_Area'].value_counts()
import matplotlib.pyplot as plt
df['ApplicantIncome'].hist(bins=50)
plt.show()
df.boxplot(column='ApplicantIncome')
plt.show()
df.boxplot(column='ApplicantIncome', by = 'Education')
plt.show()
df['LoanAmount'].hist(bins=50)
plt.show()
df.boxplot(column='LoanAmount')
plt.show()
temp1 = df['Credit_History'].value_counts(ascending=True)
temp2 = df.pivot_table(values='Loan_Status',index=['Credit_History'],aggfunc=lambda x: x.map({'Y':1,'N':0}).mean())
print('Frequency Table for Credit History:')
print(temp1)
print('\nProbility of getting loan for each Credit History class:' )
print(temp2)

import matplotlib.pyplot as plt
fig = plt.figure(figsize=(8,4))
ax1 = fig.add_subplot(121)
ax1.set_xlabel('Credit_History')
ax1.set_ylabel('Count of Applicants')
ax1.set_title("Applicants by Credit_History")
temp1.plot(kind='bar')
plt.show()
ax2 = fig.add_subplot(122)
temp2.plot(kind = 'bar')
ax2.set_xlabel('Credit_History')
ax2.set_ylabel('Probability of getting loan')
ax2.set_title("Probability of getting loan by credit history")
```

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11227

V.OUTPUT SCREENS

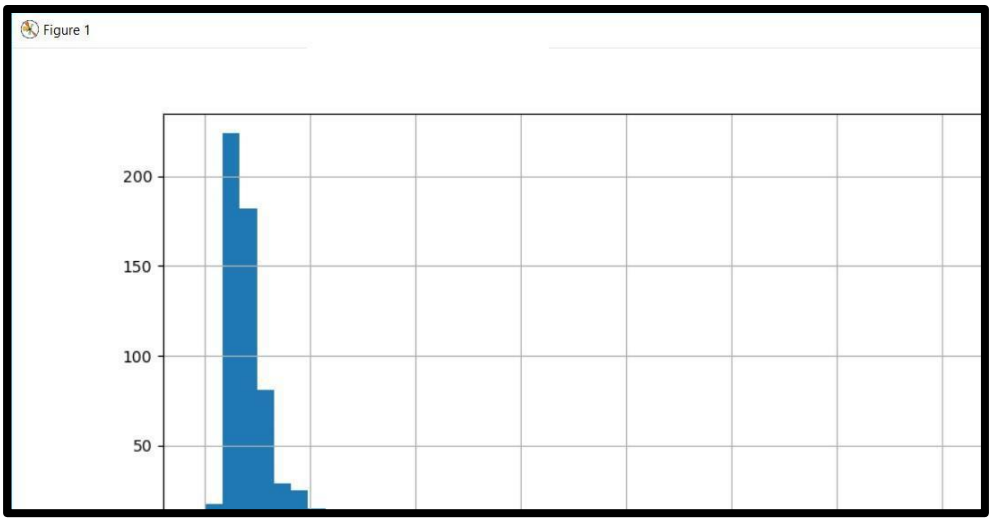

Fig. 1 Output1

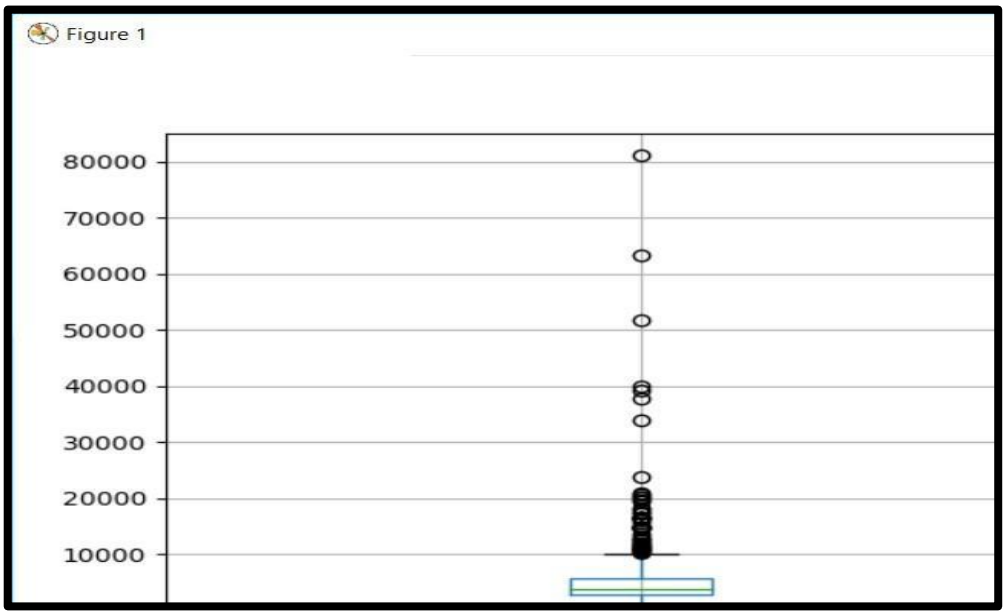

Fig. 2 Output 2

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11227

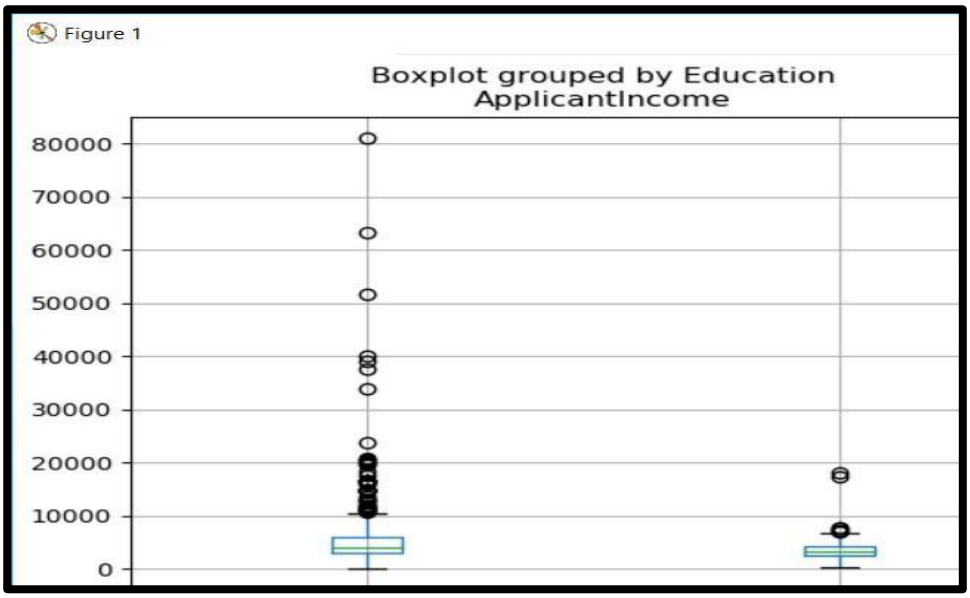

Fig. 3 Output 3

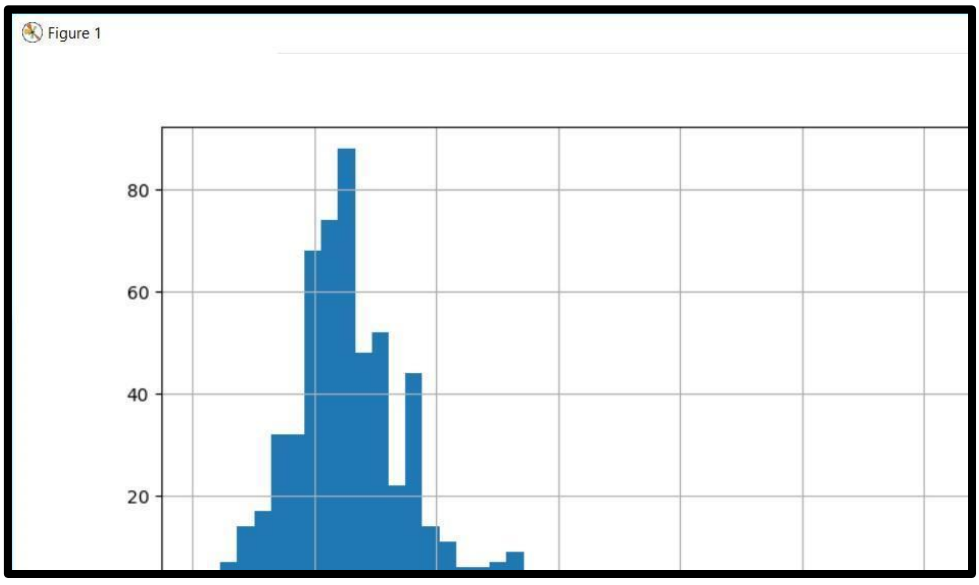

Fig. 4 Output 4

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11227

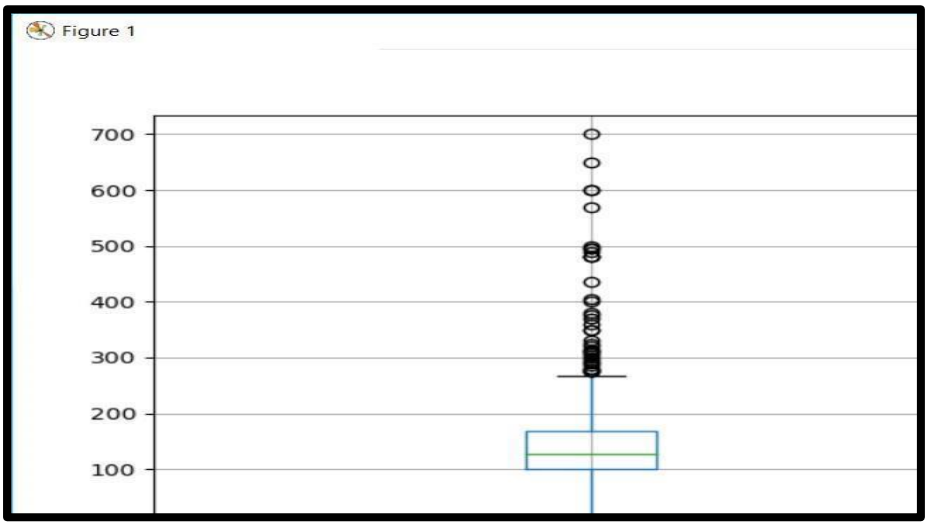

Fig. 5 Output 5

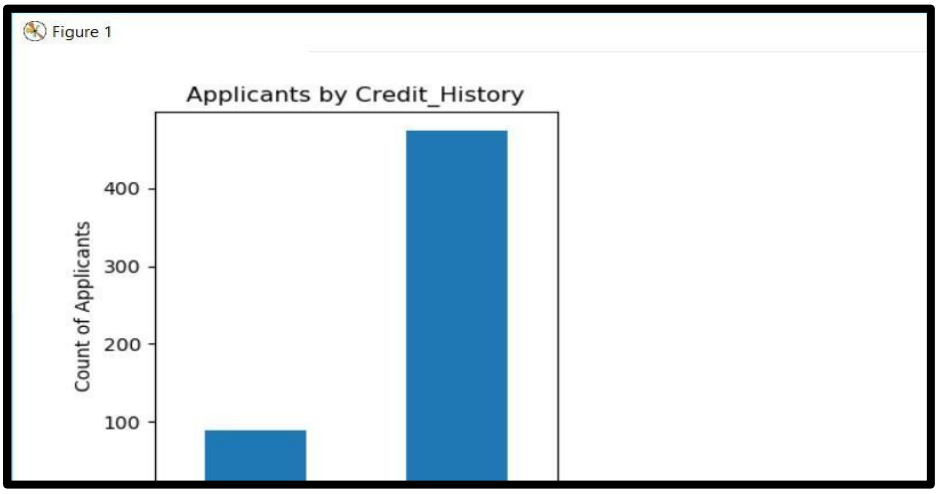

Fig. 6 Output 6

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 2, February 2022

DOI: 10.17148/IJARCCE.2022.11227

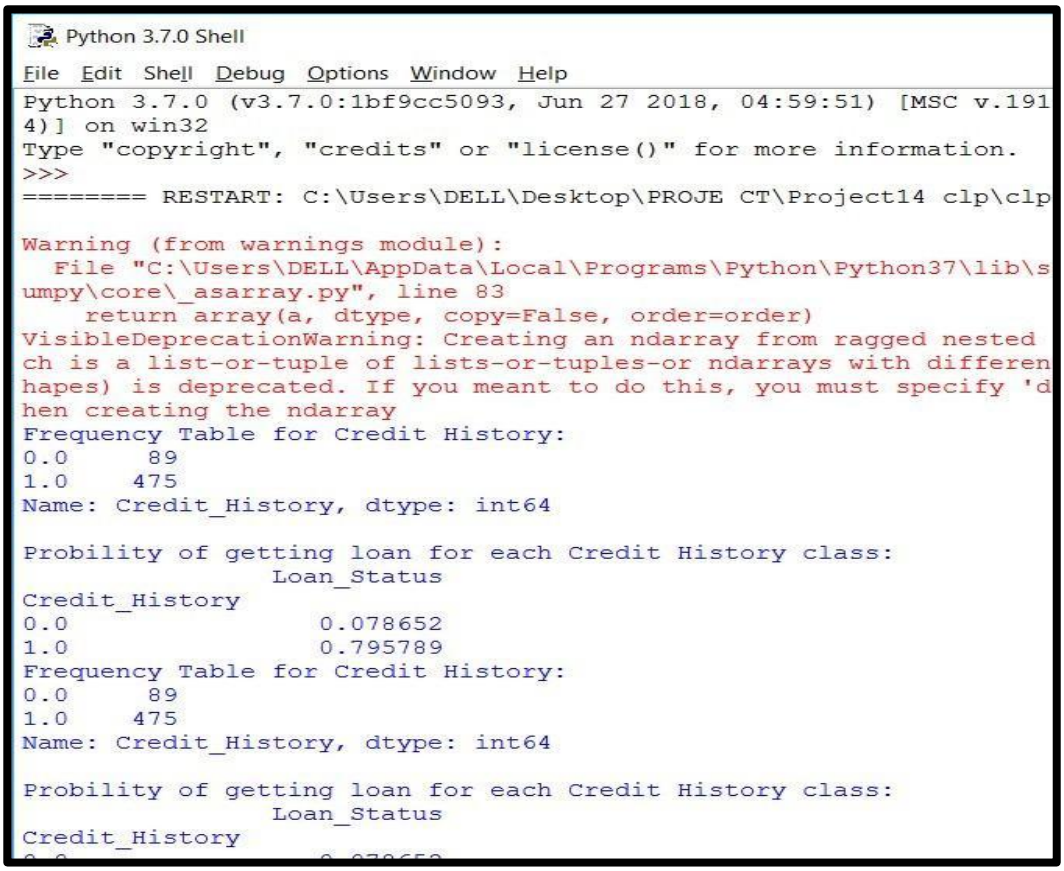

Fig. 7 Output 7

VI.CONCLUSION AND FUTURE ENHANCEMENT

In this study, a unique method for recognizing sentiment in the iris has been proposed. so as to take advantage of this fact, a replacement method that uses Keyword Spotting (KSW) to look for sentiment-bearing terms in iris has been proposed. By specializing within the terms that impact decision and ignoring non-sentiment bearing words/phrases, the ultimate system is more proof against speech recognition errors. Additionally, a clean method to form the sentiment-bearing keyword list for KWS has also been proposed. Two of the three species were gathered within the Gaspé Peninsula "all from an analogous field, and singled out that day and estimated within the meantime by the same individual with an identical mechanical assembly.

REFERENCES

[1]. G. McLachlan, K.-A. Do, and C. Ambroise, Analyzing microarray gene expression data, vol. 422. John Wiley & Sons, 2005.
[2]. E. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,” Decision Support Systems, vol. 50, no. 3, pp. 559–569, 2011.
[3]. R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification. John Wiley & Sons, 2012.
[4]. I. H. Witten and E. Frank, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.
[5]. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Van- derplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[6]. J. Li, H. Wei, and W. Hao, “Weight-selected attribute bagging for credit scoring,” Mathematical Problems in Engineering, vol. 2013, 2013.
[7]. S. Moro, P. Cortez, and P. Rita, “A data-driven approach to predictthe success of bank telemarketing,” Decision Support Systems, vol. 62, pp. 22–31, 2014.

DOI: 10.17148/IJARCCE.2022.11227

[8]. I. Bose and R. K. Mahapatra, “Business data mininga machine learning perspective,” Information & management, vol. 39, no. 3, pp. 211–225, 2001.
[9]. C.-F. Tsai and M.-L. Chen, “Credit rating by hybrid machine learning techniques,” Applied soft computing, vol. 10, no. 2, pp. 374–380, 2010.
[10]. K. Y. Tam and M. Y. Kiang, “Managerial applications of neural networks: the case of bank failure predictions,” Management science, vol. 38, no. 7, pp. 926–947, 1992.
[11]. M. Ghazanfar and A. Prugel-Bennett, “Building switching hybrid recommender system using machine learning classifiers and collaborative filtering,” IAENG International Journal of Computer Science, vol. 37, no. 3, 2010.
[12]. M. Er, L. Zhai, X. Li, and L. San, “A hybrid online sequential extreme learning machine with simplified hidden network,” IAENG International Journal of Computer Science, vol. 39, no. 1, pp. 1–9, 2012.
[13]. M. Lichman, “UCI machine learning repository,” 2013.