RESEARCH ARTICLE OPEN ACCESS

Deep Learning Anti-Fraud Model for Internet Loan

Mrs. B.Vijaya MCA M.Tech [1], A.Siva Krishnudu [2]

[1] Associate Professor, Department of Computer Applications
[2] Student, Department of Computer Applications
[1], [2] Chadalawada Ramanamma Engineering College(Autonomous)

ABSTRACT

In recent years internet fraud methods are increasing dramatically, However,bad debt has become a serious threat to internet financial companies. Recently, Internet finance is increasingly popular. However, bad debt has become a serious threat to Internet financial companies. The fraud detection models commonly used in conventional financial companies is logistic regression. Although it is interpretable, the accuracy of the logistic regression still remains to be improved. The fraud detection models commonly used in conventional financial companies is logistic regression. Although it is interpretable, the accuracy of the logistic regression still remains to be improved. This project takes a public loan dataset to explore the potential of applying deep neural network for fraud detection. Then, an XGBoost algorithm is employed to select the most discriminate features. After that, we propose to use a synthetic minority oversampling technique to deal with the sample imbalance. After Processing, we design a deep neural network for Internet loan fraud detection.

Keywords: - Deep Neural Networks, Loan, Anti-Fraud, Deep Learning.

I. INTRODUCTION

In recent years internet frauds methods are increasing dramatically. The main contribution of the project is first, analyze the real-world internet financial data for the missing data and simple imbalance. Train a deep neural network by the preprocessed data. Researchers have developed various anti-fraud prevention systems over the years.

In existing system they proposed a rule-based expert system for fraud detection. The rules of this model were manually constructed by the fraud experts from the bank. A set of financial fraud modeling languages (FFML) for better describing and combining fraud rule sets to assists fraud analysis. However, the rule-based models require sufficient and accurate expertise knowledge. While Internet loan Dataset is given, Features are Extracted From Dataset. Based on Features extraction, Experts can detect Internet loan Fraudsters Based on Knowledge Base. Internet fraud methods are increasing dramatically in recent years, together with the rapid development of Internet financial models and the Internet business used to be handled by traditional financial institutions. In this regard, Internet lending companies face an unprecedented risk of online fraud. Luckily, the rapid development of computer technology, the accumulating data, and the emerging data analysis techniques bring new opportunities to financial risk management and analysis on the big data in the financial industry. Researchers have developed various anti-fraud measures and fraud prevention systems over the years.

Leonard [1] proposed a rule-based expert system for fraud detection. The rules of this model were manually constructed by the fraud experts from the bank. Sanchez *et al.* [2] proposed to use association rules to detect fraud and help risk analysts extract more fraud rules. Edge and Sampaio [3] proposed a set of a _nancial fraud modeling language (FFML) for better describing and combining fraud rule sets to assist fraud analysis. However, the rule-based models require suf_cient and accurate expertise knowledge and can not be updated timely to new frauds. To this end, machine learning models have been introduced for fraud detection. Ghosh and Reilly [4] uses neural networks to detect credit card fraud. Kokkinaki [5] proposed decision trees and Boolean logic functions to characterize normal transaction patterns to detect fraudulent transactions. Peng *et al.* [6] compared nine machine learning models for fraud detection. The results demonstrate linear logistic and Bayesian networks are more effective. Lei and Ghorbani [7] proposed a new clustering algorithm namely improved competitive learning network (ICLN) and supervised an improved competitive learning network (SICLN). Sahin *et al.*

[8] designed a decision tree based on cost sensitivity. Halvaiee and Akbari [9] proposed to use an AIRS improved algorithm for fraud detection. However, these traditional machine learning methods heavily rely on manual subjective rules and easily lead to model risk. These methods also tend to overfit due to the imbalance training dataset with serious pollution by noises. Thus, ensemble learning methods have also been introduced to integrate different models for complicated fraud detection. Louzada and Ara [10] proposed a bagging ensemble model that integrates k-dependence probabilistic networks. The results show that the proposed ensemble model has stronger modeling capabilities. Carminati *et al.* [11] proposed a combination of semi-supervised and unsupervised fraud and anomaly detection methods, mainly using a histogram-based outlier score (HBOS) algorithm to model the user's past behavior.

II. RELATEDWORKS

Y Li proposed Relationship between Health Status and Physical Fitness of College Students from South China: An Empirical Study by Data Mining Approach. This study aims to reveal the scientific associations between the motor competence related physical fitness and the medical health status of college students from south China. Two hundred and fourteen college students, including 112 males and 102 females, from 17 provinces were administrated with the Shantou University fitness test battery twice. Y Li and X Sun proposed a Hybrid Learning Model for Short-Term Traffic Flow Forecasting**.** Accurate and reliable traffic flow forecasting is of importance for urban planning and mitigation of traffic congestion, and it is also the basis for the deployment of intelligent traffic management systems. However, constructing a reasonable and robust forecasting model is a challenging task due to the uncertainties and nonlinear characteristics of traffic flow. Aiming at the nonlinear relationship affecting traffic flow forecasting effect.

Recently, deep learning techniques have attracted a lot of academic and industrial attention that provides a new insight for financial data analysis. Fu *et al.* [12] used convolutional neural networks to effectively reduce feature redundancy. Tu *et al.* [13] design a deep feature representation technique for fraud detection. To incorporate with prior knowledge with the deep network, Greiner and Wang [14] pointed out the borrower is likely to conceal information that is not beneficial to him or even fictitious favorable information before obtaining the loan. After obtaining the loan, the borrower is likely to default unilaterally. Pope and Sydnor [15] also found it difficult to judge the risk of the personal information provided by the borrower unilaterally because the authenticity of this information cannot be verified. Freedman and Jin [16] uncovered that the borrower may commit fraudulent behavior by reporting false information, which exacerbates the information asymmetry between the two parties. Herzenstein *et al.* [17] also found that the borrowers' repayment ability and credit rating are the factors that have the greatest impact on personal credit risk. They concluded that economic strength is the determinant of judging the availability

of borrowing. At the same time, Herzenstein *et al.* [18] depicted the borrowers' spending power can also directly affect the success rate of borrowing. These methods reveal the characteristics of the borrowers would be helpful for fraud detection. Motivated by such an idea, we propose a deep learning technique to mine the fraud in a public lending dataset with 200,000 records. We analyze the customer credit rating, which can help us to identify customers' actual situations. Intuitively, the lower a customer has a credit rating, such as the E rating, the greater the likelihood of being a fraudulent user. Internet finance small loan companies set different thresholds on their customer credit rating data to build anti-fraud rules based on the true information of their customers. This paper aims to provide small financial credit companies a simple yet effective model to improve their risk control and the level of anti-fraud. Such companies often have a poor-risk control capacity with limited capacity for data engineering, modeling, and optimization.

III. PROPOSED SYSTEM ARCHITECTURE

In the proposed system deep neural networks are use for fraud detection. In this method, fill the missing values by using a random forest. Then,XBBoost algoritham is employed to select the most discrimate features. After that, we used a synthetic minority oversampling technique to deal with the sample imbalance.

The admin module has following functionalities:

- Admin enter his user id and password for login.
- View All Users who registered
- Browse For Loan Dataset
- Upload
- Training and Testing by using Classification algorithm such as Random Forest Classifier
- Predict Internet Loan Prediction
- View Anti-Fraud Model for Internet Loan
- If Required , Download Predicted Dataset

The User module has following functiooonalities

- User login with username and password
- Enter Loan Applicant Details
- Predict Loan Approved status
- View Profile.

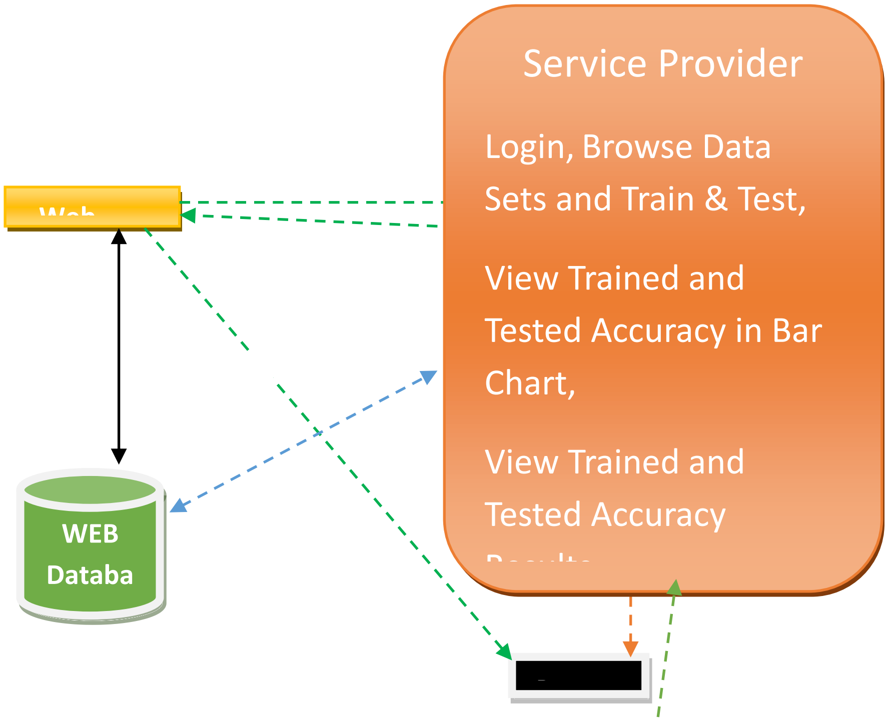
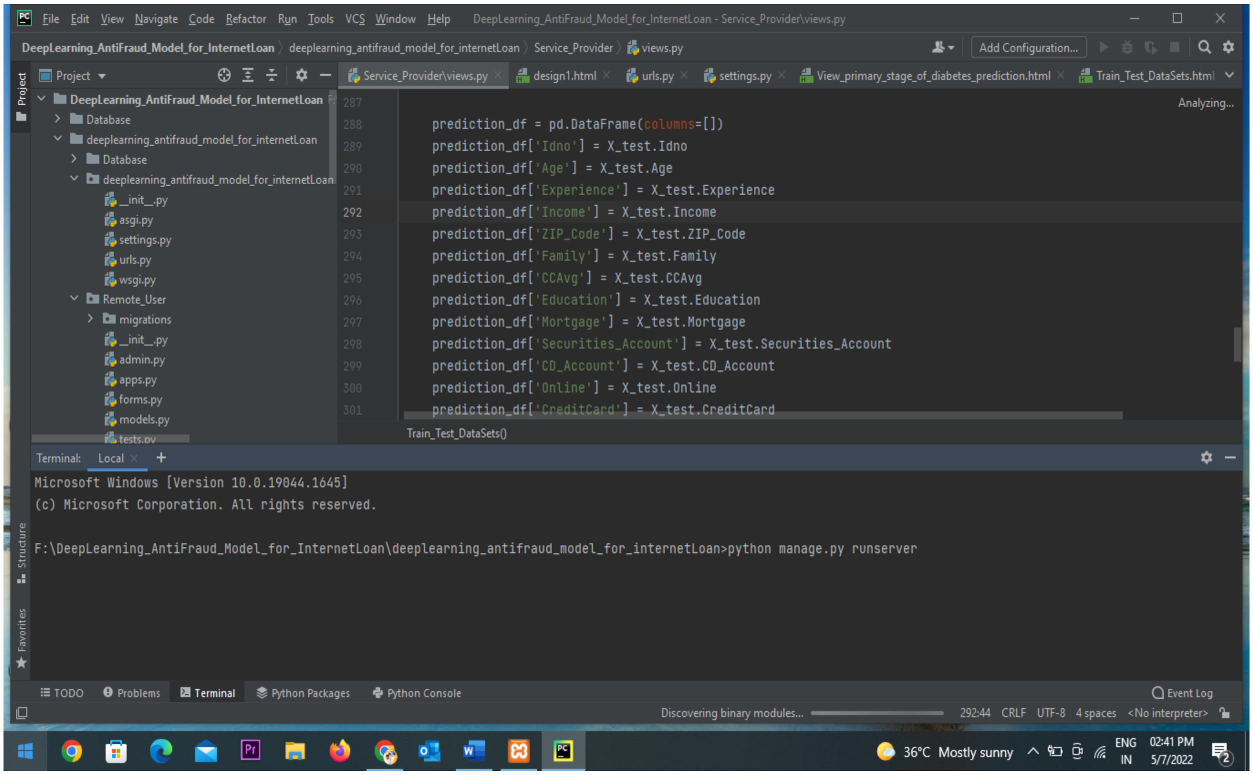

Service Provider
Login, Browse Data Sets and Train & Test,
View Trained and Tested Accuracy in Bar Chart,
View Trained and Tested Accuracy Results
Web
WEB Databa

Fig.1 Architecture of proposed system

IV. RESULTS AND DISCUSSION

The output screens obtained after running and executing the system is shown from Fig.2 to Fig.8


File Edit View Navigate Code Refactor Run Tools VCS Window Help
DeepLearning_AntiFraud_Model_for_InternetLoan - Service_Provider\views.py
DeepLearning_AntiFraud_Model_for_InternetLoan 〉 deeplearning_antifraud_model_for_internetLoan 〉 Service_Provider 〉 views.py
Add Configuration...
Project
Service_Provider\views.py
design1.html
urls.py
settings.py
View_primary_stage_of_diabetes_prediction.html
Train_Test_DataSets.html
DeepLearning_AntiFraud_Model_for_InternetLoan
Database
deeplearning_antifraud_model_for_internetLoan
Database
deeplearning_antifraud_model_for_internetLoan
__init__.py
asgi.py
settings.py
urls.py
wsgi.py
Remote_User
migrations
__init__.py
admin.py
apps.py
forms.py
models.py
tests.py
Analyzing...
Train_Test_DataSets()
Terminal: Local
TODO
Problems
Terminal
Python Packages
Python Console
Event Log
Discovering binary modules...
292:44 CRLF UTF-8 4 spaces <No interpreter>
36°C Mostly sunny
ENG IN
02:41 PM 5/7/2022

```
    prediction_df = pd.DataFrame(columns=[])
    prediction_df['Idno'] = X_test.Idno
    prediction_df['Age'] = X_test.Age
    prediction_df['Experience'] = X_test.Experience
    prediction_df['Income'] = X_test.Income
    prediction_df['ZIP_Code'] = X_test.ZIP_Code
    prediction_df['Family'] = X_test.Family
    prediction_df['CCAvg'] = X_test.CCAvg
    prediction_df['Education'] = X_test.Education
    prediction_df['Mortgage'] = X_test.Mortgage
    prediction_df['Securities_Account'] = X_test.Securities_Account
    prediction_df['CD_Account'] = X_test.CD_Account
    prediction_df['Online'] = X_test.Online
    prediction_df['CreditCard'] = X_test.CreditCard
```

```
Microsoft Windows [Version 10.0.19044.1645]
(c) Microsoft Corporation. All rights reserved.

F:\DeepLearning_AntiFraud_Model_for_InternetLoan\deeplearning_antifraud_model_for_internetLoan>python manage.py runserver
```

Fig.2 Prediction Loan

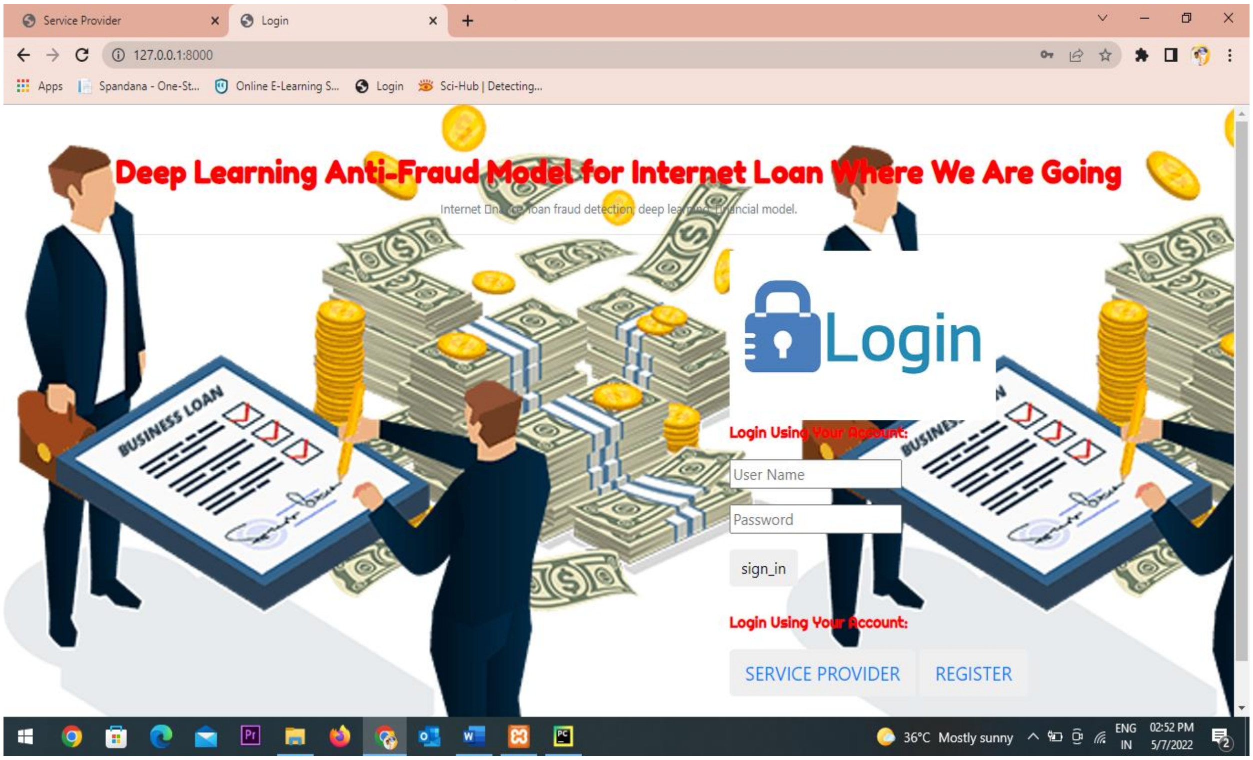
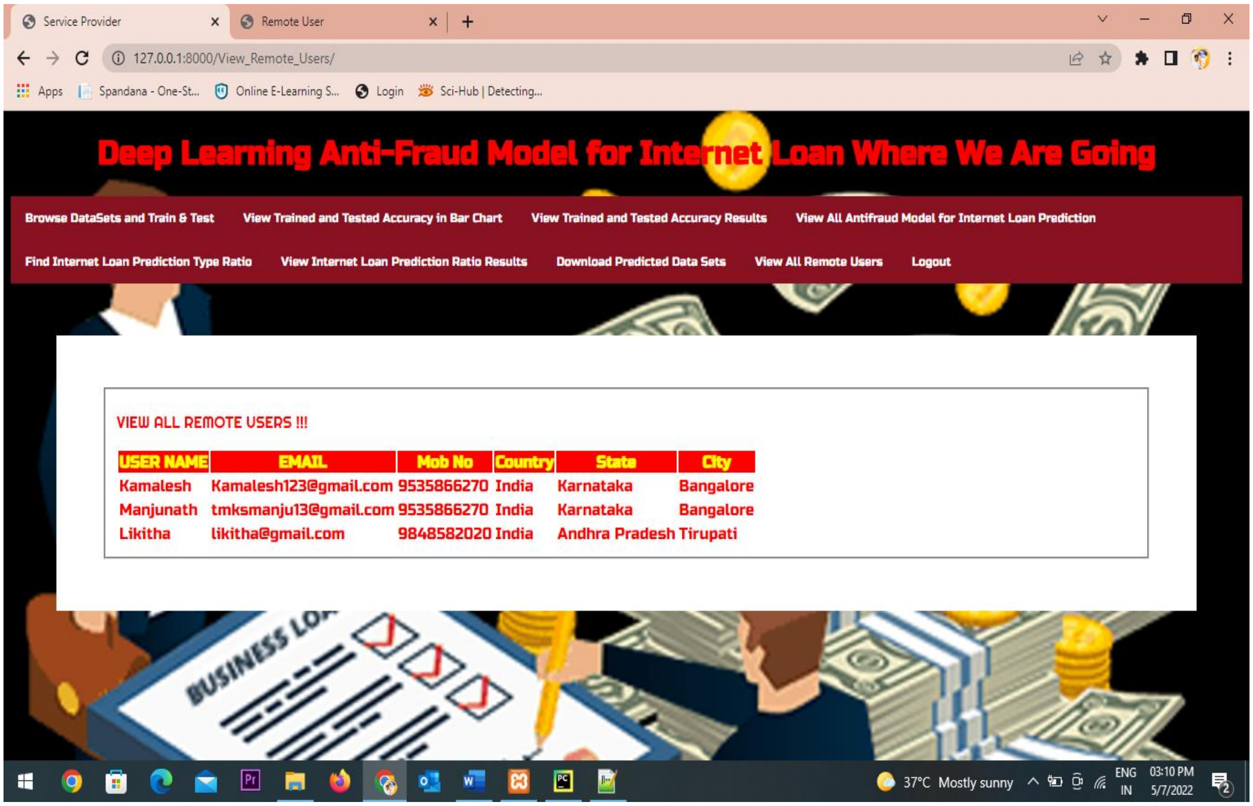


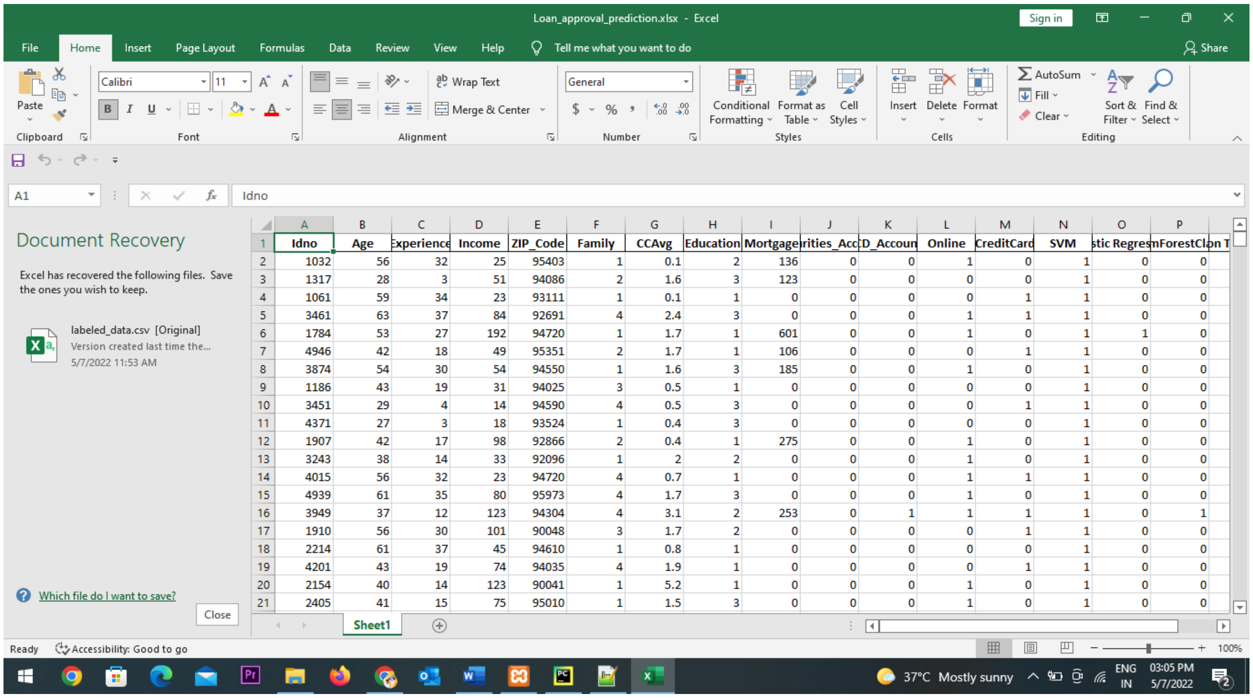
Fig.3Login Page


Service Provider
Remote User
127.0.0.1:8000/View_Remote_Users/
Apps
Spandana - One-St...
Online E-Learning S...
Login
Sci-Hub | Detecting...
Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going
Browse DataSets and Train & Test
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Antifraud Model for Internet Loan Prediction
Find Internet Loan Prediction Type Ratio
View Internet Loan Prediction Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout
BUSINESS LOA
37°C Mostly sunny
ENG IN
03:10 PM 5/7/2022

VIEW ALL REMOTE USERS !!!

USER NAME	EMAIL	Mob No	Country	State	City
Kamalesh	Kamalesh123@gmail.com	9535866270	India	Karnataka	Bangalore
Manjunath	tmksmanju13@gmail.com	9535866270	India	Karnataka	Bangalore
Likitha	likitha@gmail.com	9848582020	India	Andhra Pradesh	Tirupati

Fig.4 Users list



Document Recovery

Excel has recovered the following files. Save the ones you wish to keep.

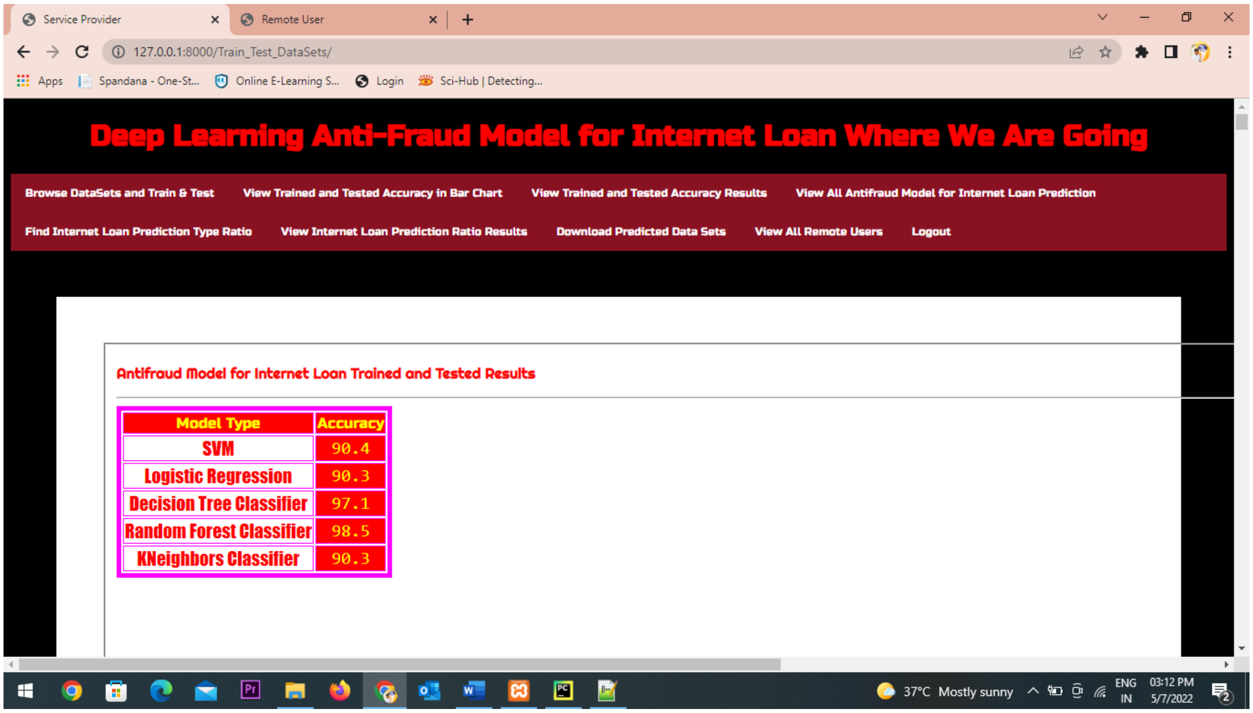
labeled_data.csv [Original]
Version created last time the...
5/7/2022 11:53 AM

Which file do I want to save?

Close

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	Idno	Age	Experience	Income	ZIP_Code	Family	CCAvg	Education	Mortgage	rities_Acc	D_Accoun	Online	CreditCard	SVM	stic Regres	nForestCl	on T
2	1032	56	32	25	95403	1	0.1	2	136	0	0	1	0	1	0	0	
3	1317	28	3	51	94086	2	1.6	3	123	0	0	0	0	1	0	0	
4	1061	59	34	23	93111	1	0.1	1	0	0	0	0	1	1	0	0	
5	3461	63	37	84	92691	4	2.4	3	0	0	0	1	1	1	0	0	
6	1784	53	27	192	94720	1	1.7	1	601	0	0	1	0	1	1	0	
7	4946	42	18	49	95351	2	1.7	1	106	0	0	0	1	1	0	0	
8	3874	54	30	54	94550	1	1.6	3	185	0	0	1	0	1	0	0	
9	1186	43	19	31	94025	3	0.5	1	0	0	0	0	0	1	0	0	
10	3451	29	4	14	94590	4	0.5	3	0	0	0	0	1	1	0	0	
11	4371	27	3	18	93524	1	0.4	3	0	0	0	0	0	1	0	0	
12	1907	42	17	98	92866	2	0.4	1	275	0	0	1	0	1	0	0	
13	3243	38	14	33	92096	1	2	2	0	0	0	1	0	1	0	0	
14	4015	56	32	23	94720	4	0.7	1	0	0	0	1	1	1	0	0	
15	4939	61	35	80	95973	4	1.7	3	0	0	0	1	0	1	0	0	
16	3949	37	12	123	94304	4	3.1	2	253	0	1	1	1	1	0	1	
17	1910	56	30	101	90048	3	1.7	2	0	0	0	0	1	1	0	0	
18	2214	61	37	45	94610	1	0.8	1	0	0	0	0	0	1	0	0	
19	4201	43	19	74	94035	4	1.9	1	0	0	0	0	1	1	0	0	
20	2154	40	14	123	90041	1	5.2	1	0	0	0	1	0	1	0	0	
21	2405	41	15	75	95010	1	1.5	3	0	0	0	1	0	1	0	0	

Fig.5 Dataset view


Service Provider
Remote User
127.0.0.1:8000/Train_Test_DataSets/
Apps
Spandana - One-St...
Online E-Learning S...
Login
Sci-Hub | Detecting...
37°C Mostly sunny
ENG IN
03:12 PM 5/7/2022

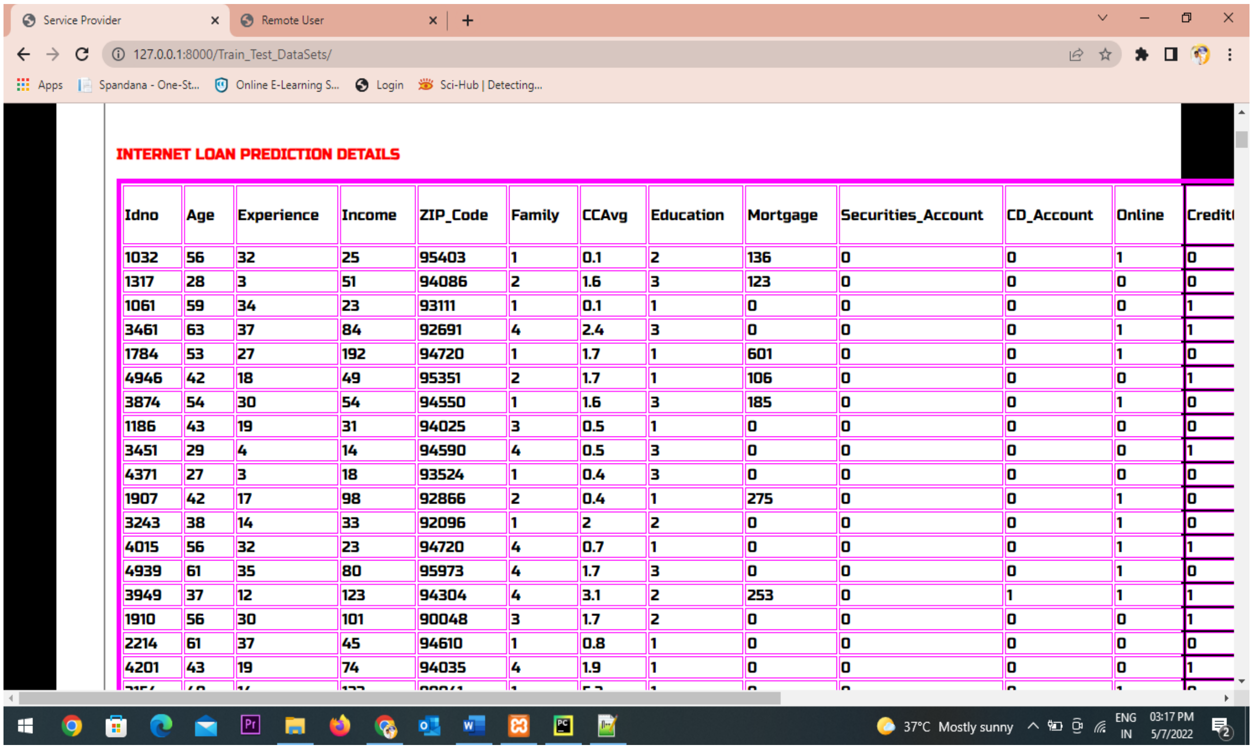
Deep Learning Anti-Fraud Model for Internet Loan Where We Are Going

Browse DataSets and Train & Test
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View All Antifraud Model for Internet Loan Prediction
Find Internet Loan Prediction Type Ratio
View Internet Loan Prediction Ratio Results
Download Predicted Data Sets
View All Remote Users
Logout

Antifraud Model for Internet Loan Trained and Tested Results

Model Type	Accuracy
SVM	90.4
Logistic Regression	90.3
Decision Tree Classifier	97.1
Random Forest Classifier	98.5
KNeighbors Classifier	90.3

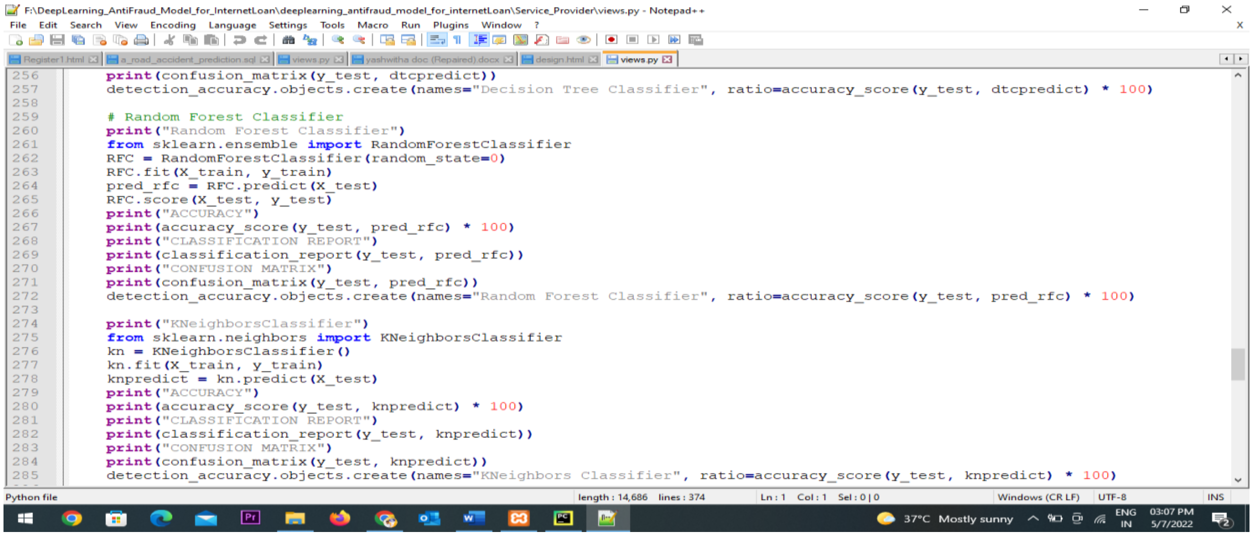
Fig.6 Anti fraud model



INTERNET LOAN PREDICTION DETAILS

Idno	Age	Experience	Income	ZIP_Code	Family	CCAvg	Education	Mortgage	Securities_Account	CD_Account	Online	Credit
1032	56	32	25	95403	1	0.1	2	136	0	0	1	0
1317	28	3	51	94086	2	1.6	3	123	0	0	0	0
1061	59	34	23	93111	1	0.1	1	0	0	0	0	1
3461	63	37	84	92691	4	2.4	3	0	0	0	1	1
1784	53	27	192	94720	1	1.7	1	601	0	0	1	0
4946	42	18	49	95351	2	1.7	1	106	0	0	0	1
3874	54	30	54	94550	1	1.6	3	185	0	0	1	0
1186	43	19	31	94025	3	0.5	1	0	0	0	0	0
3451	29	4	14	94590	4	0.5	3	0	0	0	0	1
4371	27	3	18	93524	1	0.4	3	0	0	0	0	0
1907	42	17	98	92866	2	0.4	1	275	0	0	1	0
3243	38	14	33	92096	1	2	2	0	0	0	1	0
4015	56	32	23	94720	4	0.7	1	0	0	0	1	1
4939	61	35	80	95973	4	1.7	3	0	0	0	1	0
3949	37	12	123	94304	4	3.1	2	253	0	1	1	1
1910	56	30	101	90048	3	1.7	2	0	0	0	0	1
2214	61	37	45	94610	1	0.8	1	0	0	0	0	0
4201	43	19	74	94035	4	1.9	1	0	0	0	0	1

Fig.7 Loan prediction details


F:\DeepLearning_AntiFraud_Model_for_InternetLoan\deeplearning_antifraud_model_for_internetLoan\Service_Provider\views.py - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
Register1.html
a_road_accident_prediction.sql
views.py
yashwitha doc (Repaired).docx
design.html
views.py
Python file
length : 14,686 lines : 374
Ln : 1 Col : 1 Sel : 0 | 0
Windows (CR LF)
UTF-8
INS
37°C Mostly sunny
ENG IN
03:07 PM 5/7/2022

```
        print(confusion_matrix(y_test, dtcpredict))
        detection_accuracy.objects.create(names="Decision Tree Classifier", ratio=accuracy_score(y_test, dtcpredict) * 100)

        # Random Forest Classifier
        print("Random Forest Classifier")
        from sklearn.ensemble import RandomForestClassifier
        RFC = RandomForestClassifier(random_state=0)
        RFC.fit(X_train, y_train)
        pred_rfc = RFC.predict(X_test)
        RFC.score(X_test, y_test)
        print("ACCURACY")
        print(accuracy_score(y_test, pred_rfc) * 100)
        print("CLASSIFICATION REPORT")
        print(classification_report(y_test, pred_rfc))
        print("CONFUSION MATRIX")
        print(confusion_matrix(y_test, pred_rfc))
        detection_accuracy.objects.create(names="Random Forest Classifier", ratio=accuracy_score(y_test, pred_rfc) * 100)

        print("KNeighborsClassifier")
        from sklearn.neighbors import KNeighborsClassifier
        kn = KNeighborsClassifier()
        kn.fit(X_train, y_train)
        knpredict = kn.predict(X_test)
        print("ACCURACY")
        print(accuracy_score(y_test, knpredict) * 100)
        print("CLASSIFICATION REPORT")
        print(classification_report(y_test, knpredict))
        print("CONFUSION MATRIX")
        print(confusion_matrix(y_test, knpredict))
        detection_accuracy.objects.create(names="KNeighbors Classifier", ratio=accuracy_score(y_test, knpredict) * 100)
```

Fig.8 Random forest classifier

V. FUTURE SCOPE AND CONCLUSION

In this project, take the real customer information of the public loan data set of the lending club company as a sample. Then, we build a deep learning based Internet fraud detection model. We introduce the main parameters of the model and optimizes to find the optimal parameter combination of the model. Finally, the most popular logistic regression in the financial industry as well as other comparisons are used as a baseline to evaluate the performance of the proposed model. The results reveal the deep neural network achieves better performance, which is promising to be used in the financial industry for Internet fraud detection.

REFERENCES

[1] K. J. Leonard, ''The development of a rule based expert system model for fraud alert in consumer credit,'' Eur. J. Oper. Res., vol. 80, no. 2, pp. 350–356, Jan. 1995.

[2] D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, ''Association rules applied to credit card fraud detection,'' Expert Syst. Appl., vol. 36, no. 2, pp. 3630–3640, Mar. 2009.

[3] M. E. Edge and P. R. F. Sampaio, ''The design of FFML: A rule-based policy modelling language for proactive fraud management in financial data streams,'' Expert Syst. Appl., vol. 39, no. 11, pp. 9966–9985, Sep. 2012.

[4] S. Ghosh and D. L. Reilly, Credit Card Fraud Detection With a NeuralNetwork. Wailea, HI, USA: IEEE, 1994.

[5] A. I. Kokkinaki, ``On atypical database transactions: Identi_cation of probable frauds using machine learning for user pro_ling," in *Proc. IEEE Knowl. Data Eng. Exchange Workshop*, 1997, pp. 229_238.

[6] Y. Peng, G.Wang, G.Kou, andY. Shi, ``An empirical study of classi_cation algorithm evaluation for _nancial risk prediction," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2906_2915, Mar. 2011.

[7] J. Z. Lei and A. A. Ghorbani, ``Improved competitive learning neural networks for network intrusion and fraud detection," *Neurocomputing*, vol. 75, no. 1, pp. 135_145, Jan. 2012.

[8] Y. Sahin, S. Bulkan, and E. Duman, ``A cost-sensitive decision tree approach for fraud detection," *Expert Syst. Appl.*, vol. 40, no. 15, pp. 5916_5923, Nov. 2013.

[9] N. Soltani Halvaiee and M. K. Akbari, ``A novel model for credit card fraud detection using arti_cial immune systems," *Appl. Soft Comput.*, vol. 24, pp. 40_49, Nov. 2014.

[10] F. Louzada and A. Ara, ``Bagging k-dependence probabilistic networks: An alternative powerful fraud detection tool," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11583_11592, Oct. 2012.