Detection of Fraudulence in Credit Card Transactions using Machine Learning on Azure ML

Abhishek Shivanna
Center for Distributed and Mobile Computing, EECS
University of Cincinnati
Cincinnati, USA
shivanak@mail.uc.edu

Sujan Ray
Center for Distributed and Mobile Computing, EECS
University of Cincinnati
Cincinnati, USA
raysu@mail.uc.edu

Khaldoon Alshouiliy
Center for Distributed and Mobile Computing, EECS
University of Cincinnati
Cincinnati, USA
alshoukr@mail.uc.edu

Dharma P Agrawal
Center for Distributed and Mobile Computing, EECS
University of Cincinnati
Cincinnati, USA
dpa@cs.uc.edu

Abstract— **With the advancement of mobile and cloud technologies, there is a sharp increase in online transactions. Detecting fraudulent credit card transactions on a timely basis is a very critical and challenging problem in Financial Industry. Although online transactions are very convenient, they bring the risk of fraudulence on many aspects. Some of the key challenges in detecting fraudulence in online transactions include irregular behavioral patterns, skewed dataset i.e. high normal transaction to fraudulent transaction ratio, limited availability of data and dynamically changing environment. Every year people lose millions of dollars due to credit card fraud. There is a lack of quality research in this domain. We have used a dataset comprising of European cardholders which has 284,807 transactions to model our system. In this paper, we will design and develop credit card fraudulence detection system by training and testing two ML algorithms: Decision Forest (DF) and Decision Jungle (DJ) classifiers. Our results successfully demonstrate that DJ classifier delivers higher performance compared to DF classifier.**

Keywords— Big Data; Credit Card; Finance; Machine Learning; Decision Jungle; Decision Forest; SMOTE; Online Transactions; Azure ML.

I. INTRODUCTION

The usage of credit cards for online transactions has drastically increased due to the rise of e-commerce. Credit card frauds are turning out to be one of the most common problems faced by customers and banking system worldwide. Online credit card fraudulent transactions amounted to a whopping $28 Billion in 2018. Nielson Report estimates credit card frauds to increase to nearly $36 billion over the next five years, and $41 billion in the next 10 years [1].

Credit card frauds are of two types [20]: offline frauds and online frauds. Typically, a fraudster commits offline fraud by using a stolen physical credit card. In majority of cases, the card issuing institution can disable before it can be misused [21]. Generally, online frauds are committed via cardholder-not-present, web or by phone shopping. As one doesn't need physical signature or card imprint, and only cards' details are required to complete an online transaction, this type of fraud is widespread [22]. As there is an increase in electronic commerce in today's day and age, credit card usage for online purchases has become user friendly and necessary. Every day huge number of credit card transactions are processed worldwide.

Fraud detection models can be built either by using supervised or unsupervised technique, or a combination of the two. In a supervised technique, past occurrences of legitimate and fraud transactions are made use of to train a model. New transactions are assigned a suspicion score by this model. In unsupervised technique, there are no prior information on the nature of transactions [3].

In this work, we have used credit card transactions dataset [2]. This dataset comprises of online credit card transactions for two days made by European card holders in September 2013. To protect confidentiality and privacy issues, all the columns have been PCA transformed. Features: V1, V2, V3, V4…V28 are PCA-transformed, features 'time' and 'amount' are not subjected to PCA transformation [2].

Over the last several years, the volume of online transactions has increased multifold to several petabytes [7][8]. It's very challenging to catch fraudulent transactions from such large-scale data sets as training fraud detection systems is quite hard. In this paper, we are detecting online credit card fraudulent transactions by designing and training a model on Azure ML studio which is a cloud-based platform using ML algorithms: DF and DJ. We then evaluate the performances of these two techniques.

The size of financial data is larger than ever and is exponentially increasing due to modern day data collection storage and processing techniques. It is not possible for conventional techniques to handle data of this size and scope. In order to process this data which can be in different forms: structured, semi-structured and unstructured, we need improved analytical methods.

At the beginning of this work, we have carried out exploratory data analysis to understand more about the dataset. After this step we get more insight into feature distributions and how they are related to each other and to the class outcome. After this, we have balanced the dataset using Synthetic Minority Over-sampling Technique (SMOTE). Next step is to train our ML models in Azure ML Platform. We have used various classifier evaluation metrics such as AUC, accuracy, F-1 Score, recall and precision to determine a better performing model.

The paper is organized as follows: review of literature in section II, methodology in section III, data analysis techniques that were explored in section IV, detailed model implementation in section V, results in section VI, and conclusions in Section VII.

978-1-7281-9656-5/20/$31.00 ©2020 IEEE

II. Review of Literature

There has been a lot of research in trying to leverage the power of ML systems in solving some of the challenges of financial sector. Sahin et al. [3] have developed credit card fraud detection system using DTs and SVMs, and have compared the performance of decision trees and SVMs. They have achieved maximum accuracy of 99.78% with SVM on training set and 83.02% on testing dataset. Vlasselaer et al. [4] have combined effort of analyzing inherent characteristics of incoming transactions and network-based features. They have correctly predicted 98.7% of the cases using random forest, they are getting AUC of 0.987 for their classifier. Shen et al [5] have tested the applicability of 3 classification models and have compared them for prediction accuracy. They found that neural networks and logistic regression have better performance compared to decision tree. M.J.Islam et al. [6] have reviewed Bayesian theory and tried to explore Naive Bayes classifier and KNN Classifier. By varying the values of k, they have attempted to classify the data points. They have achieved the best results for k value of 5. For this k value, the percentage error of classification is 9.45%.

III. Proposed Methodology

The methodology we have proposed to detect online credit card fraudulent transactions is illustrated in Figure 1:

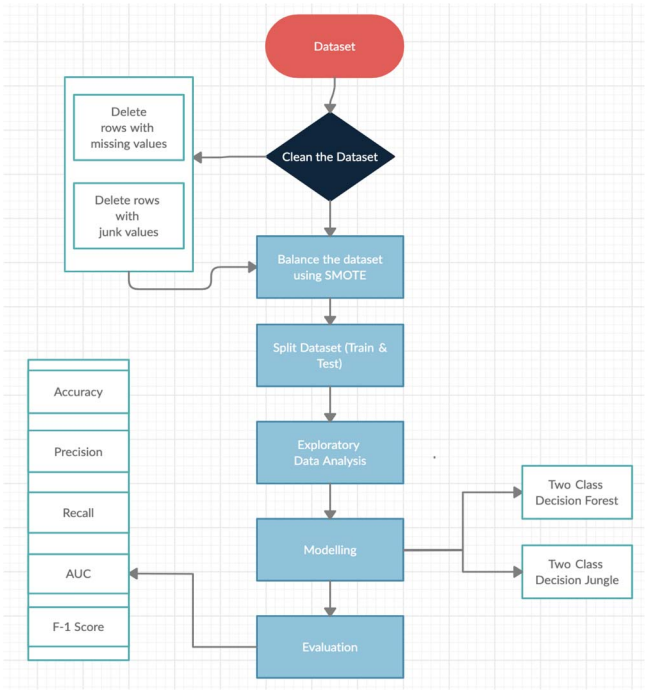

Fig. 1. Proposed Methodology for online Fraud Detection

A. Dataset Description

We have used credit card transactions dataset. This dataset has been widely used by many researchers. It has 284,807 transactions recorded for a period of two days in September 2013 [2]. This dataset is highly imbalanced; 0.172% of all the transactions were fraudulent in nature. This dataset has 30 features. In the dataset, we are not provided with original features and background. We are provided with a PCA transformed version of the features due to confidentiality and privacy issues.

B. Cleaning the Dataset

In this step we clean the dataset, so that it is appropriate to be used to train our model. Generally this step involves deleting the rows that have missing or inappropriate values.

C. Balancing the Dataset

Imbalance in dataset is usually reflected by disproportionate distribution of classes within the dataset. For convenience, we can call the class that makes up a huge proportion of the data set as majority classes and the class that makes up a smaller proportion as minority classes [9]. We need to apply a sampling technique before performing classification task on imbalanced dataset. It is difficult to train a model with an imbalanced dataset. One approach to solve the problem of dealing with imbalanced dataset is to oversample the minority class. One such technique is called Synthetic Minority Oversampling Technique (SMOTE) [10]. In the feature space, SMOTE selects samples that are close, by drawing a boundary between the samples in the feature space and picking a new sample at any point along that line. [11]. We have used SMOTE to balance our dataset.

D. Data Analysis

Our methodology focusses on trying to understand the credit card transactions dataset and come up with an effective model to catch fraudulent transactions. To get a sense of the dataset we have performed exploratory data analysis (EDA) using widely used open source libraries such as NumPy, Pandas, matplotlib, Seaborn, etc. Matplotlib and Seaborn are excellent libraries for visualization. We have obtained histograms, bar graphs, density plots, box plots, etc. to get a better sense of the dataset.

E. Azure Machine Learning (Azure ML) Platform

Azure ML is an end-to-end cloud-based environment where one can train, test, deploy, automate, manage, and track Machine Learning models [12]. Users can perform a variety of tasks like importing the dataset, training, splitting, clustering and multitude of other tasks from a simple web browser. Azure ML has built-in support for most of the widely used open-source python frameworks. It essentially has everything one needs to create an end to end ML pipeline. It has other features like Bot Frameworks, which are skeleton codes in order to build chatbots [13]. After we develop an end-to-end ML model, we can easily deploy and publish the model in the form of web service hosted on Azure. Once we deploy the model, we can have access to it from almost anywhere including custom apps, web sites, Azure Data Factory, BI tools [14]. We have trained and tested both of our classification models on Azure ML cloud platform.

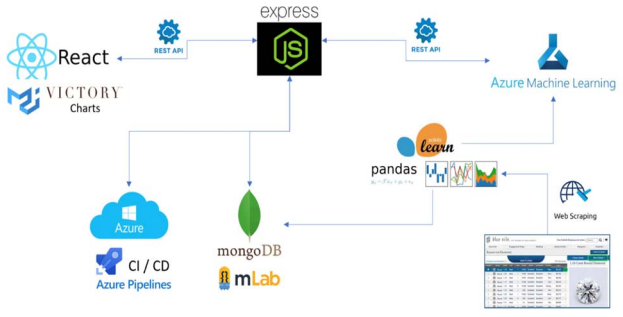

Fig. 2. Azure Machine Learning [23]

F. *Machine Learning(ML) Algorithms*

- Decision Jungle:

DJ comprises of multiple DAGs [15] and are very powerful models for classification [16]. Decision Jungles have the following advantages: 1). DAGs typically have lower memory footprints and better generalizations as they allow tree branches to merge compared to Decision Trees. 2). Decision Jungles represent decision boundaries that are non-linear in nature as they are non-parametric in nature. 3)They have very good resilience even in the presence of noisy features [15].

- Decision Forest:

Decision Forest being an ensemble learning method works on the principle that one can get better results and generalized models by creating and combining multiple related models in some way rather than a single model [17].

IV. Exploratory Data Analysis

Here, we have conducted some preliminary data analysis and have explained the rationale behind our analysis. We have analyzed the dataset using widely used open source libraries such as: Pandas, matplotlib, Seaborn, NumPy, etc.

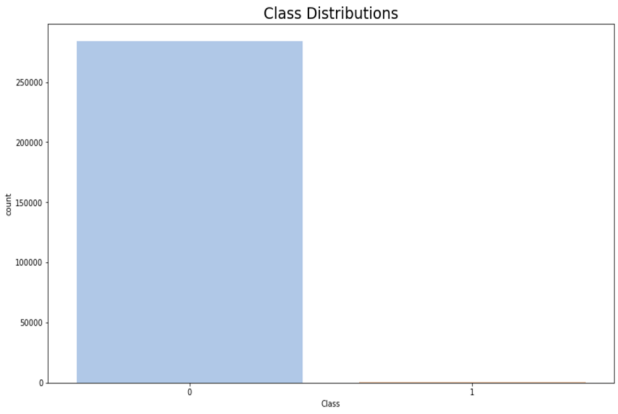

Fig. 3. Class Distributions

In the Figure 3, “0” represents Normal Transactions and “1” represents Fraudulent Transactions. Class distribution graph indicates that there are 284,315 normal transactions and 492 fraudulent transactions.

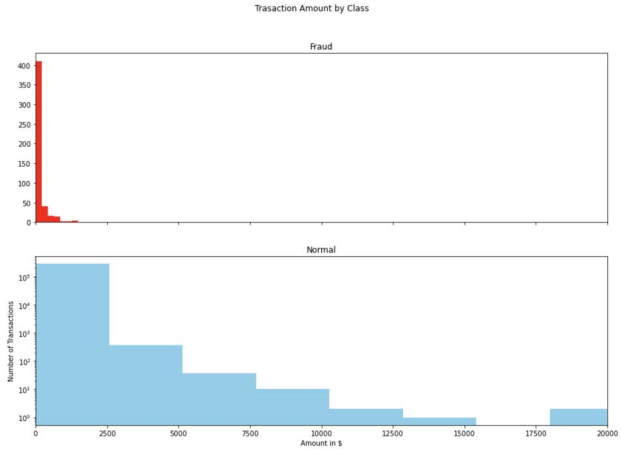

Fig. 4. Comparison of Transactional Values

Next, we are trying to understand the amount disparity between normal and fraudulent transactions. From Figure 4, It is evident that fraudulent transactions have very small transactional amount when compared to normal transactions which have higher transactional value.

We need to find accurate correlations between input and target variables. Correlational analysis is crucial to get more insight into the data. By looking at correlational matrix we get a sense of how one or more features are related to each other. Positive correlation implies when there is an increase in one of the features/variables, the value of another feature/variable also increases. Negative correlation implies, the increase in one of the feature’s value causes a decreases in another feature’s value. Two features might not have any correlation if they are independent of each other [18].

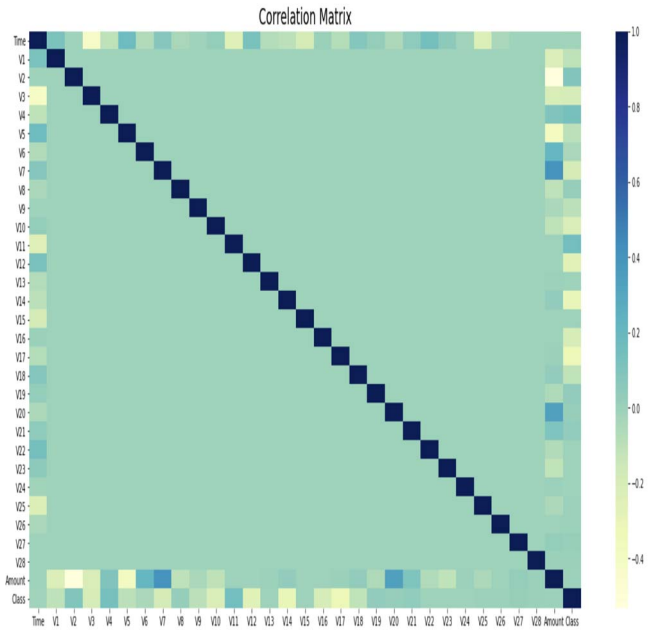

Fig. 5. Correlation Matrix

From the above correlation matrix, we can see that V14 and V12 are negatively correlated and V4 and V11 are positively correlated. We are going to use box plots to get a sense of how these features are distributed in fraudulent and non-fraudulent transactions.

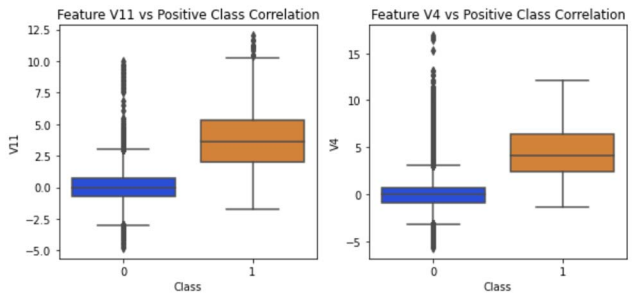

Fig. 6. Features V11 & V14 vs Positive Class Correlation

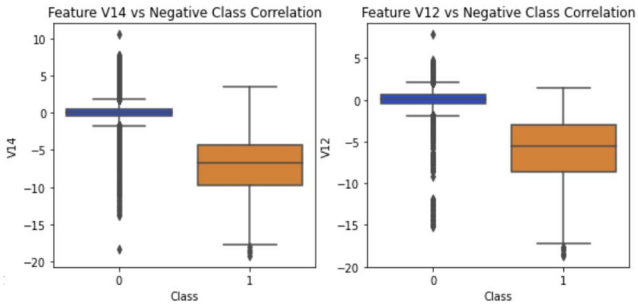

Fig. 7. V14 & V12 vs Negative Class Correlation

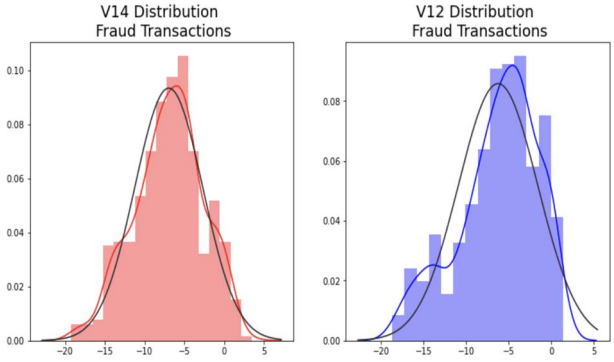

Fig. 8. V14 and V12 Distributions

Figure 8. gives us a sense of how features V14 and V12 are distributed for fraudulent transactions. It can be implied from the above figure that they are normally distributed.

V. MODEL IMPLEMENTATION

In this paper, we have used DF and DJ algorithms to implement credit card fraudulence detection system. We have conducted the experiments on Azure Machine Learning Platform under default parameter settings. We have compared the overall performance of the implementations using the following metrices: AUC, Accuracy, Recall, Precision and F-1 Score. Also, we have computed the confusion matrix for both the implementations to compare the predictions of both the models.

Confusion Matrix is computed using the following parameters:

- True Positive (TP): A fraudulent transaction is predicted as fraudulent transaction.
- True Negative (TN): A normal transaction is predicted as a normal transaction.
- False Positive (FP): A normal transaction is predicted as fraudulent transaction.
- False Negative (FN): A fraudulent transaction is predicted as a normal transaction.

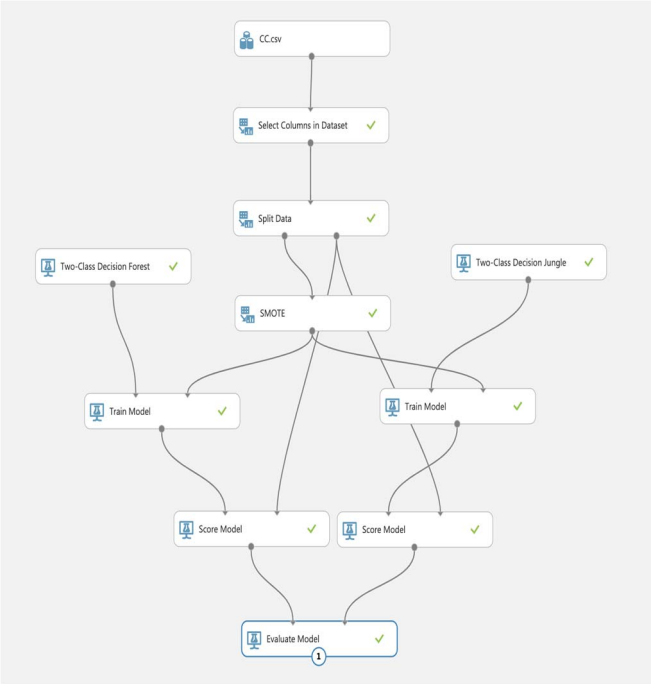

Fig. 9. Azure ML model

Following are the metrics of evaluation:

- Precision: It is the ratio of accurately predicted positive observations to total positive occurrences predicted.
  Precision = TP / TP+FP

- Recall: It is the ratio of true positives and true positives plus false negatives
  Recall = TP / TP+FN

- Accuracy: It is the ratio of observations correctly predicted to total observations.
  Accuracy = (TP+TN) / (TP+FP+FN+TN)

- F1 score: F-1 Score is defined as the harmonic mean of Precision and Recall (it takes into account both false positives and false negatives)
  F1 Score = 2*(Recall * Precision) / (Recall + Precision)

- Area under ROC Curve (AUC):

  • AUC value of 1 is a perfect score.
  • Higher the value of AUC better is the performance of the classifier [19].
  • ROC curve indicates the TP vs FP rate.

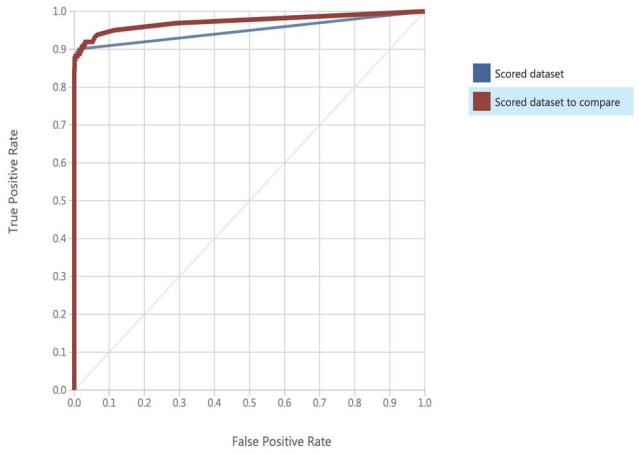

Fig. 10. ROC Curve

VI. Results

In this paper we have implemented a model to detect online credit card fraudulent transactions by training DF and DJ algorithms on Azure ML platform. Both DF and DJ algorithms are giving close to 99.9% accuracy. But the performance of DJ is better than DF as it has a better AUC value and Recall value.

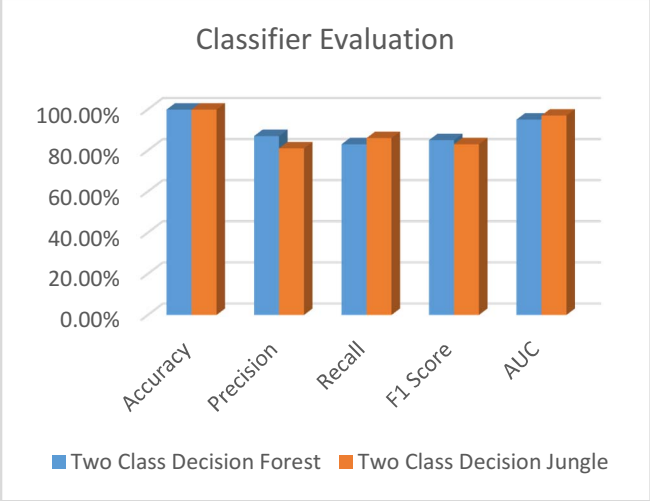

Fig. 11. Classifier Evaluation

TABLE I. CONFUSION MATRIX

	TP	TN	FP	FN
Two Class Decision Forest	134	85262	19	27
Two Class Decision Jungle	136	85249	32	25

TABLE II. CLASSIFICATION RESULTS

	Accuracy	Precision	Recall	F1 Score	AUC
Two Class Decision Forest	99.90%	0.87	0.83	0.85	0.95
Two Class Decision Jungle	99.90%	0.81	0.86	0.83	0.97

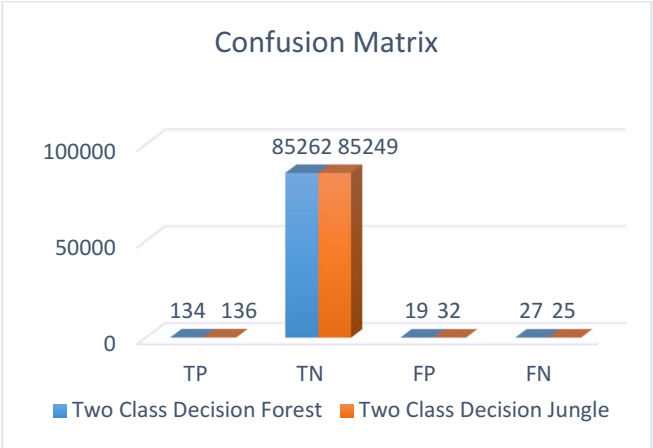

Fig. 12. Confusion Matrix

VII. Conclusion and Future Work

The usage of credit cards for online transactions is increasing very rapidly. In this day and age, it is very critical to correctly identify online credit card fraudulent transactions. In conclusion, we have proposed a credit card fraudulence detection method which identifies online credit card fraudulent transactions. Decision Jungle algorithm has shown promising results to be adapted in any fraudulence detection system. We will extend our work in the future by using neural networks to build more advanced fraudulence detection system.

References

[1] "Annual Fraud Statistics Released by The Nilson Report", Nov 2019, [online] Available: https://www.prnewswire.com/news-releases/payment-card-fraud-losses-reach-27-85-billion-300963232.html

[2] "Credit Card Fraud Detection Dataset", Mar 2018, [online] Available: https://www.kaggle.com/mlg-ulb/creditcardfraud

[3] 3- Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines," Int. Multiconference Eng. Comput. Sci., vol. I, pp. 442–447, 2011.

[4] V. Van Vlasselaer et al., "APATE: A novel approach for automated credit card transaction fraud detection using networkbased extensions," Decis. Support Syst., vol. 75, pp. 38–48, 2015.

[5] A. Shen, R. Tong, Y. Deng, "Application of classification models on credit card fraud detection", Service Systems and Service Management 2007 International Conference, pp. 1-4, 2007.

[6] M. J. Islam, Q. M. J. Wu, M. Ahmadi, M. A. Sid-Ahmed, "Investigating the Performance of Naive-Bayes Classifiers and KNearestNeighbor Classifiers", IEEE International Conference on Convergence Information Technology, pp. 1541-1546, 2007.

[7] You Dai, Jin Yan, et al. (2016) "Online Cedit Card Fraud Detection: A Hybrid Framework with Big Data Technologies". IEEE Trust Com-BigSE-ISP

[8] Philip K Chaan (2016), "Distributed Data Mining in Credit Card Fraud Detection", Florida Institute of Technology.

[9] "Imbalanced Data", Google Developers, Jan 2019, [Online]https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data

[10] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," J. Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[11] SMOTE for Imbalanced Classification with Python, Jan 2020[Online]https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/

[12] "Azure Machine Learning", Jan 2020, [Online] https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-ml data.

[13] "6 reasons Microsoft has become the go-to for machine learning", May 2019, [online] Available: https://medium.com/datadriveninvestor/6-reasons-microsoft-has-become-the-go-to-for-machine-learning-e642864ef5f5

[14] "AZURE MACHINE LEARNING: AN OVERVIEW", Feb 2019, [online] Available: https://www.bluegranite.com/blog/bid/404378/azure-machine-learning-an-overview.

[15] "Two-Class Decision Jungle", Aug 2019, [online] Available: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-decision-jungle

[16] Decision Jungles: Compact and Rich Models for Classification Jamie Shotton Toby Sharp Pushmeet Kohli Sebastian Nowozin John Winn Antonio Criminisi.

[17] "Two-Class Decision Forest", Oct 2019, [online] Available: https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-decision-forest

[18] "Understanding Value Of Correlations In Data Science Projects", Jan 2019, [online] Available: https://medium.com/fintechexplained/did-you-know-the-importance-of-finding-correlations-in-data-science-1fa3943debc2

[19] "Classification: ROC Curve and AUC", Jan 2020, [online] Available: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc#:~:text=AUC%20represents%20the%20probability%20that,has%20an%20AUC%20of%201.0.

[20] Stolfo, S.J., Fan, D.W., Lee, W., Prodromidis, A.L.: Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results (1999)

[21] Maes, S., Tuyls, K., Vanschoenwinkel, B., Manderick, B.: Credit Card Fraud Detection using Bayesian and Neural Networks. In: Proc. of NAISO Congress on Neuro Fuzzy Technologies (2002)

[22] Kou, Y., Lu, C., Sinvongwattana, S., Huang, Y.P.: Survey of Fraud Detection Techniques. In: Proc. of IEEE Networking, Taiwan, March 21-23 (2004)

[23] "Full Stack Machine Learning on Azure", Oct 2019, [online] Available: https://towardsdatascience.com/full-stack-machine-learning-on-azure-f0f6b77be07e