A Case Study: Fraud Detection Models of Credit Card Using ML

Shalini Lodhi[1], Ravi Kumar Singh Pippal[2]
[1]MTech Scholar, [2]Professor
Department of CSE, Vedica Institute of Technology,
RKDF University, Bhopal, M.P, India
ravesingh@gmail.com

Abstract: Unique: The COVID-19 pandemic has restricted individuals' portability partially, making it hard to buy labor and products disconnected, which has driven the formation of a culture of expanded reliance on web-based administrations. One of the critical issues with utilizing Visas is misrepresentation, which is a serious test in the domain of online exchanges. Subsequently, there is an immense need to foster the most ideal way to deal with utilizing AI to forestall practically all false Mastercard exchanges. This contextual investigation an AI models in view of two phases of assessment. A true Visa extortion discovery dataset of European cardholders is utilized in each model alongside delineated K-fold cross-approval. In the primary stage, nine AI calculations are tried to identify deceitful exchanges. The best three calculations are named to be utilized again in the subsequent stage, with 19 resampling methods utilized with every single one of the best three calculations. Out of 330 assessment metric qualities that required almost one month to get, the All K-Nearest Neighbors (AllKNN) undersampling method alongside CatBoost (AllKNN-CatBoost) will be view as the best proposed model in future.

Keywords: fraud detection; credit card fraud; machine learning; XGBoost; CatBoost; random forest.

I. Introduction

As the world is going to a credit only economy, there will be increasingly more reliance on making on the web exchanges. Current misrepresentation doesn't expect fraudsters to be actually in the wrongdoing areas. They can play out their fiendish exercises at the solace of their homes with numerous approaches to concealing their personalities. Such character concealing procedures incorporate utilizing a VPN, directing the casualty's traffic through the Tor organization, and so forth, and it isn't not difficult to follow them back. The effect of online monetary misfortunes can't be undervalued. When fraudsters take card subtleties, they can utilize the actual cards or sell the card subtleties to others, similar to the case in India, where the card subtleties of around 70 million individuals are being sold on the dim web [1].

One of the most serious Mastercard extortion episodes in ongoing memory that occurred in the UK brought about GBP 17 million aggregates in monetary misfortunes. The episode happened after a gathering of worldwide fraudsters schemed to take the detail data of in excess of 32,000 charge cards during the 2000s [2]. This occurrence is viewed as the greatest card misrepresentation ever. In this way, the absence of successful security frameworks brings about billion-dollar misfortunes because of Mastercard extortion [3]. The two cardholders, while utilizing their cards, and card backers, while handling the exchanges, are being consoled that all exchanges are harmless. On the other hand to this conviction, fraudsters expect to trick monetary organizations and cardholders into accepting that the deceitful exchanges are authentic. Moreover, there are a few false exchanges that happen ceaselessly to get monetary benefit without the information on both card guarantors and cardholders. Both approved organizations and cardholders now and again don't realize that they have deceitful exchanges, and this is the haziest side of Visa exchanges. In light of that, it is an extremely provoking cycle to identify fake

action among great many certified exchanges, particularly when deceitful exchanges are fundamentally not exactly the real ones [4].

There are numerous extortion recognition methods that assistance to forestall misrepresentation in the monetary business, for example, prescient examination and information mining, particularly displaying calculations that consolidate bunching procedures and oddity identification [5]. Notwithstanding, this large number of methods can't be performed without AI calculations, whether they are managed or solo, which can be viable in Visa misrepresentation grouping [6]. Notwithstanding, those AI calculations experience incalculable quantities of difficulties while attempting to distinguish all false action [7]. In the ideal AI model, the generally utilized assessment measurements should be at the most elevated values. With expectations of drawing nearer to this ideal model, there are numerous enhancements required in this field. The difficulties confronting Mastercard misrepresentation discovery rely upon many variables, for example, AI calculations, cross-approval procedures, and looking like methods. Taking into account these elements can upgrade the presentation of the model that can be approved by the assessment measurements. In a true issue, it is very uncommon to have a reasonable dataset to work with, and that implies that the order calculation subverts the significance of the minority class in the dataset as a rule.

Truly, the minority class is the main part of the arrangement cycle, particularly in Mastercard misrepresentation discovery. Because of the lopsided conveyance of the classes in the dataset, the proposed approach features the awkwardness class issue utilizing different resampling strategies subsequent to picking the best AI calculations. Not exclusively are the resampling strategies thought about in this paper, however so too are the better CV (cross-validation) methods also. This paper proposes a high level methodology as far as picking the best AI calculation in mix with the best resampling strategy. This approach depends on an investigation at two phases utilizing execution assessment measurements.

The main stage intends to examine nine AI calculations with their default boundaries. The nine calculations are K-Nearest Neighbors (KNN) [9], Logistic Regression (LR) [8], Decision Tree (DT) [10], Random Forest (RF) [9], Naïve Bayes (NB) [11], Gradient Boosting Machines (GBM) [12], Extreme Gradient Boosting (XGBoost) [14], Light Gradient Boosting Machine (LightGBM) [13], and Category Boosting (CatBoost) [15]. Out of these nine calculations, hands down the best three calculations are assigned to be use in the subsequent stage. The subsequent stage intends to dissect 19 resampling strategies with every single one of the chose three calculations from the primary stage. These 19 resampling methods are sorted as follows: 11 undersampling, 6 oversampling, and 2 with mixes of both undersampling and oversampling strategies immediately. Moreover, this stage expects to choose the best blend of both calculation and resampling procedure to get the best proposed model in light of the best by and large execution. This imaginative methodology stands apart by investigating various ways of tending to the class lopsidedness issue in the dataset. This is displayed as far as contrasting the best AI calculations and utilizing defined K-fold CV and resampling procedures. Utilizing this number of different calculations and procedures gives a promising outcome, particularly given that it required almost one month just to get all the assessment metric qualities.

2. Related Work

Given the significance of charge card misrepresentation, there are prestigious procedures to obstruct this fiendish action. Not in the least do monetary foundations and banks give the accommodation of monetary

administrations, yet they likewise make sure to the cutting edge defenders of cardholders. Likewise, they contribute and foster different strategies, including the cutting edge AI procedures that numerous frameworks intensely rely upon. One of these procedures utilized is DT. It is not difficult to execute, however it necessities to check every exchange individually [16].

Khatri et al. [17] dissected different models with an imbalanced European Visa extortion discovery (ECCFD) dataset. They didn't think about utilizing any resampling strategies. The outcomes demonstrated that DT was by and large awesome, with great Recall (79.21%), Precision (85.11%), and time (5 s), while KNN was better as to Recall (81.19%), Precision (91.11%), however not with time (463 s). Another procedure includes utilizing LightGBM. Taha and Malebary [18] directed their trial on two datasets utilizing LightGBM. The first is the ECCFD dataset, and the subsequent one is the UCSD-FICO Data Mining Contest 2009 dataset. Utilizing a 5-overlay rendition of the K-Fold CV, they determined the normal qualities. They contrasted it with their Optimized Light Gradient Boosting Machine (OLightGBM), which included hyper-boundary tuning with other cutting edge calculations. They found that OLightGBM accomplished the most noteworthy scores in both datasets. In the first dataset, OLightGBM accomplished 90.94% in Area under the Receiver Operating Characteristic Curve (AUC) measures, 98.40% in Accuracy, 40.59% in Recall, 97.34% in Precision, and 56.95% in F1-Score. Likewise, in the second dataset, OLightGBM accomplished 92.88% in AUC, 98.35% in Accuracy, 28.33% in Recall, 91.72% in Precision, and 43.27% in F1-Score. One more exploration bearing zeroed in on LR and KNN, by which Vengatesan et al. [19] analyzed the exhibition of LR and KNN on the imbalanced ECCFD dataset. The discoveries were that KNN accomplished the best Precision of 95%, Recall of 72%, and F1-Score of 82%.

Likewise, Puh and Brki'c [20] concentrated on the presentation of various calculations, in particular, RF, the Support Vector Machine (SVM), and LR, on the dataset of European cardholders. They handled the irregularity class issue in the dataset utilizing the Synthetic Minority Oversampling Technique (SMOTE). Destroyed and LR were utilized to make their models for certain progressions in the boundaries of the calculation. The LR boundary C was set to 100 and L2-Regulation was utilized. They made two models involving LR as far as the growing experience. The first includes static learning and the other one includes steady learning. The outcomes showed that the AUC score was 91.14% with static learning and the AUC score was 91.07% with steady learning. The Average Precision score was 73.37% with static learning, and the Average Precision score was 84.13% with steady learning. Different specialists zeroed in on RF.

Hema [21] assessed the ECCFD dataset without tending to the imbalanced class issue utilizing three calculations, which were RF, LR, and Category Boosting (CatBoost). Hema found that RF gave better generally brings about terms of Accuracy (99.95%), Precision (91.95%), Recall (79.2%), F1-Score (85.1%), MCC (85.31%), and AUC (89%). Kumar et al. [22] led a fundamental report utilizing RF on the ECCFD dataset. They observed that the exactness of RF was 90%. A few different specialists thought about utilizing an Artificial Neural Network (ANN), which reenacts how a human cerebrum functions [23].

Asha and KR [24] utilized SVM, KNN, and ANN models on the ECCFD dataset. The outcomes showed that the ANN was awesome among different models, with an Accuracy of 99.92%, a Precision of 81.15%, and a Recall of 76.19%. Dubey et al. [1] directed a trial on the Credit Card Customer dataset with the

utilization of ANN. That handled the information in the principal layer, which was the information layer, and afterward the secret layer, which had 15 neurons and the utilization of the RELU enactment capacity, and afterward the result layer, which utilized the Sigmoid actuation work. Ultimately, their methodology brought about 99.92% Accuracy, 99.96% Recall, 99.96% Precision, and 99.96% F1-Score values.

Varmedja et al. [25] led their exploration by parting the ECCFD in a 80:20 proportion. The calculations they utilized were LR, RF, NB, Multilayer Perceptron, and ANN. The imbalanced issue of the dataset was tended to utilizing SMOTE. Each model was refreshed through numerous ages relying upon the capacity to bear the streamlining. Their discoveries showed that RF accomplished the best outcomes with regards to Accuracy (99.96%), Recall (81.63%), and Precision (96.38%).

Taking a gander at related work, there are a few viewpoints should be considered to identify false movement in Mastercard exchanges. Each approach has its own strategy to improve the general presentation of their proposed models. In any case, an AI calculation can have a specific outcome in one methodology, and various outcomes in different ones. To get a superior thought on which calculation plays out the best, expanding the quantity of calculations utilized in the trial ought to be thought of. The lopsidedness class issue is extremely normal in datasets. Subsequently, not resolving this issue can prompt lackluster showing. This issue can be handled by utilizing defined CV and resampling procedures, with a critical number of resampling methods that can be utilized in tests. Besides, the quantity of assessment measurements is significant for assessing the model's exhibition from various points. A few past works need at least one of these viewpoints. Thus, a remarkable methodology is proposed.

3. Algorithms

There are so many calculations, we can used to recognize the misrepresentation in charge cards utilizing AI and profound learning techniques, some are examine beneath:

3.1. Logistic Regression

This is one of the conventional AI calculations that is as yet utilized today because of its fast investigation and basic technique for handling the highlights of a class. The capacity of LR to relate different elements, particularly with solid ones, and its capacity to conform to various elements rely upon indicator factors and the result. LR utilizes values that are more prominent than 1 and under 0 to treat the peculiarities in the dataset, and it isn't restricted to arranging and foreseeing binominal results, yet additionally multinomial results also, and it utilizes the sigmoid capacity to assess the upsides of boundaries' coefficients [7]. At the point when LR inspects the upsides of the properties during a continuous exchange, it lets us know regardless of whether the exchange ought to proceed, as it is utilized for bunching [26].

3.2. K-Nearest Neighbors

This is a classifier that is utilized for grouping and relapse issues. One of its benefits is that it builds the Mastercard discovery rate and diminishes phony problem rates. It works in view of closeness measures. Hence, it stores all occasions that are open and orchestrates new ones [9]. It utilizes factual learning strategies, and it functions admirably in managed AI frameworks, as there is a learning stage that is utilized to get the vital information, which empowers it to characterize significant contrasts. Be that as it may, in solo learning methods, the preparation stage isn't needed. There are three variables in KNN: the distance measurements, the distance rule, and the K worth. Distance measurements help to find the closest

neighbor of any approaching data of interest, the distance rule characterizes the new data of interest into a class when it is handling its elements corresponding to the data of interest in its area, and the K worth is the quantity of neighbors with which to look at it [7].

3.3. Decision Tree

This is a computational technique that expects to group and foresee. It has a tree-like interaction, including a root hub, leaf hub, and branch. Each inward hub shows a test in light of its credits, the result of the test demonstrates each branch, and the class name holds each leaf hub [10]. The interaction happens recursively when a similitude tree is characterized, where the hubs are marked with property names and edges. Those marked qualities are credits that fulfill a few circumstances and leaves that contain a power factor, which is characterized as the proportion of the quantity of exchanges that fulfill these circumstances over the all out number of real exchange in the way of behaving [27].

3.4. CatBoost

There are a few classifiers that have a place with the DT family, for example, CatBoost, which is a bleeding edge, open-source type of gradient boosting for the DT library created by Yandex scientists and specialists [15]. CatBoost is exceptionally flexible and can be utilized with a large number of utilizations and issues.

3.5. XGBoost

Another classifier is XGBoost, which has a place with the DT family, and it is a choice treebased group AI classifier. XGBoost utilizes a slope supporting structure that comprises of a bunch of CART (classification and regression trees) [14].

3.6. GBM

GBM is another classifier that has a place with the DT family that utilizes a troupe method. It means to further develop Accuracy utilizing a troupe of trees rather than a solitary tree, and this calculation is utilized for relapse and grouping, as illustrated by Friedman [12,28].

3.3.7. LightGBM

One more calculation is LightGBM, which is ordinarily utilized as a quick, conveyed, highperformance, and open-source inclination supporting structure. LightGBM is created by Microsoft and depends on DT calculations [13].

3.8. Naïve Bayes

This is a directed AI classifier that can be prepared to foresee future occasions of the objective class. NB is known to be a strong probabilistic strategy that exploits component and class data from the dataset, which empowers it to foresee cases in the future [29, 30]. "Gullible" is a depiction of how this strategy functions since it treats each trait freely founded on the class variable, whether it is available or missing, and "Bayes" is a portrayal of how it computes the likelihood of the class rightness [11]. Despite the fact that NB has a straightforward system, this calculation creates great outcomes in many muddled true issues.

3.9. Random Forest

While experiencing a characterization or relapse issue, the gathering strategy, otherwise called the RF technique, can manage both by developing numerous choice trees such that each tree goes about as a frail

student, and those trees are added together, and they become a hearty student [9]. One of the upsides of a RF is that it is viable and quick while dealing with imbalanced datasets, in any event, when the imbalanced datasets have large number of highlights [31]. The way that a RF works is that each tree gives a grouping vote to a class. The new article is made, and it is given most extreme votes into the class.

4. Resampling Techniques

Resampling methods are normally used to handle the irregularity class issue in a dataset [32]. While taking a gander at the dataset utilized here, the all out number of substantial cases is 284,315 and the complete number of misrepresentation cases is 492. That implies that the legitimate cases are 99.827% of the all out number of cases; in the interim, the extortion cases are just 0.173% of the complete number of cases. Unquestionably, the dataset is exceptionally lopsided. Subsequently, resampling strategies prove to be useful, as the lopsidedness class issue is connected with the presentation of the calculations [33]. There are 3 fundamental classifications of resampling methods: undersampling, oversampling, and the blend of both undersampling and oversampling.

4.1. Undersampling

Undersampling procedures are known to give a minimal adjusted preparing set, and one of the upsides of this sort of strategy is that it lessens the expense of the learning stage [34]. One of the issues of undersampling strategies is the expulsion of an enormous lump of the preparation set, particularly when the greater part class occasions are colossally immense, which prompts the deficiency of critical cases that would, thus, lead to challenges in characterization and expectation.

4.2. Oversampling

Dissimilar to undersampling, the advancement of oversampling strategies means to save the greater part class cases and reproduce the minority class occurrences to handle the issue of imbalanced preparation set. The issue with this sort of strategy is that it might prompt terrible showing of the model now and again on the grounds that it could be difficult to create the minority information in the preparation set [35, 36].

4.3. Oversampling and Undersampling Combination

This blend expects to utilize both undersampling and oversampling methods simultaneously. By consolidating these strategies, the irregularity class issue is tended to in an unexpected way.

5. Hardware, Software and Dataset Requirement

There are so many requirements to develop a model to detect credit card fraud detection with ALLKNN and resembling methods

5.1. Simulation Environment

The simulation environment can be categorized as follows:

5.1.1. Software

The experiment is performed using a 64-bit Windows 10 virtual machine on a server that is equipped with Anaconda Navigator 1.10.0, Jupyter Notebook 6.1.4, and Python 3.8. The libraries used in the Anaconda Navigator environment are Scikit-Learn, Pandas, Numpy, Seaborn, Matplotlib, and Imbalanced-Learn, along with machine learning classifiers.

5.2. Hardware

The accompanying focuses mirror the equipment climate: • Processor: Intel(R) Xeon(R) CPU D-1527 @ 2.20 GHz 2.19 GHz. • RAM: 7.00 GB. The actual server has four CPU centers, and the virtual machine just purposes three CPU centers with six dangers.

6.2. Dataset

The dataset that is utilized with this proposed approach is a genuine world dataset got from Kaggle [29]. It contains exchanges made by Visas in September 2013 by European cardholders that happened north of two days distributed by Universite Libre de Bruxelles. There are 492 occurrences of extortion out of 284,807 exchanges with 31 highlights, to be specific 'Time', 'V1' to 'V28', 'Sum', and 'Class'. This dataset is broadly utilized by numerous analysts and specialists found in the connected work segment; consequently, this dataset is decided to analyze the assessment metric upsides of our proposed model with theirs.

7. Conclusions and Future Work

With expanded reliance on internet based exchanges and Mastercards, fraudsters and hoodlums are fostering their means to hold onto others' cash. In any case, a proactive methodology should be considered by saddling man-made brainpower and AI devices to forcefully handle this issue, paying little mind to how modern the countermeasures are. The methodology which will build in view of two phases. The primary stage plans to choose the best three AI calculations out of nine calculations. The subsequent stage intends to incorporate the best three calculations with nineteen resampling methods. Each model in the two phases is assessed in light of the Area under the Accuracy, Receiver Operating Characteristic Curve (AUC), Recall, F1-Score and Precision. In the principal stage, the nine calculations are K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB), Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and Category Boosting (CatBoost). In the subsequent stage, the 19 resampling strategies are partitioned as follows: 11 undersampling, 6 oversampling, and 2 mixes of both undersampling and oversampling procedures. The complete numbers of models in the two phases are 66, with their 330 assessment metric qualities that required almost one month to get. The best model out of every one of these will be according to investigation AllKNN alongside CatBoost (AllKNN-CatBoost). At long last, AllKNN-CatBoost is contrasted and past works with the equivalent dataset and comparable methodologies.

REFERENCES

[1] Dubey, S.C.; Mundhe, K.S.; Kadam, A.A. Credit Card Fraud Detection using Artificial Neural Network and BackPropagation. In Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Rasayani, India, 13–15 May 2020; pp. 268–273.

[2] Martin, T. Credit Card Fraud: The Biggest Card Frauds in History. Available online: https://www.uswitch.com/credit-cards/guides/credit-card-fraud-the-biggest-card-frauds-in-history/ (accessed on 22 January 2022).

[3] Zhang, X.; Han, Y.; Xu, W.; Wang, Q. HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. Inf. Sci. 2019, 557, 302–316. [CrossRef]

[4] Makki, S.; Assaghir, Z.; Taher, Y.; Haque, R.; Hacid, M.-S.; Zeineddine, H. An experimental study with imbalanced classification approaches for credit card fraud detection. IEEE Access 2019, 7, 93010–93022.

[5] McCue, C. Advanced Topics. Data Mining and Predictive Analysis; Butterworth-Heinemann: Oxford, UK, 2015; pp. 349–365.

[6] Berad, P.; Parihar, S.; Lakhani, Z.; Kshirsagar, A.; Chaudhari, A. A Comparative Study: Credit Card Fraud Detection Using Machine Learning. J. Crit. Rev. 2020, 7, 1005.
[7] Jain, Y.; Namrata, T.; Shripriya, D.; Jain, S. A comparative analysis of various credit card fraud detection techniques. Int. J. Recent Technol. Eng. 2019, 7, 402–403.
[8] Tolles, J.; Meurer, W.J. Logistic regression: Relating patient characteristics to outcomes. JAMA 2016, 316, 533–534. [CrossRef][PubMed]
[9] Shirodkar, N.; Mandrekar, P.; Mandrekar, R.S.; Sakhalkar, R.; Kumar, K.C.; Aswale, S. Credit card fraud detection techniques–A survey. In Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Shiroda, India, 13–15 May 2020; pp. 1–7. [CrossRef]
[10] Gaikwad, J.R.; Deshmane, A.B.; Somavanshi, H.V.; Patil, S.V.; Badgujar, R.A. Credit Card Fraud Detection using Decision Tree Induction Algorithm. Int. J. Innov. Technol. Explor. Eng. IJITEE 2014, 4, 66–67.
[11] Zareapoor, M.; Seeja, K.; Alam, M.A. Analysis on credit card fraud detection techniques: Based on certain design criteria. Int. J. Comput. Appl. 2012, 52, 35–42. [CrossRef]
[12] Friedman, J.H. Greedy function approximation: A gradient boosting machine. Ann. Stat. 2001, 29, 1189–1232. [CrossRef]
[13] Microsoft. LightGBM. Available online: https://github.com/microsoft/LightGBM (accessed on 22 January 2021).
[14] XGBoost Developers. Introduction to Boosted Trees. Available online: https://xgboost.readthedocs.io/en/latest/tutorials/model.html (accessed on 22 January 2022).
[15] Yandex Technologies. CatBoost. Available online: https://yandex.com/dev/catboost/ (accessed on 22 January 2022).
[16] Delamaire, L.; Abdou, H.; Pointon, J. Credit card fraud and detection techniques: A review. Banks Bank Syst. 2009, 4, 61.
[17] Khatri, S.; Arora, A.; Agrawal, A.P. Supervised machine learning algorithms for credit card fraud detection: A comparison. In Proceedings of the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 29–31 January 2020; pp. 680–683. [CrossRef]
[18] Taha, A.A.; Malebary, S.J. An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine. IEEE Access 2020, 8, 25579–25587. [CrossRef]
[19] Vengatesan, K.; Kumar, A.; Yuvraj, S.; Kumar, V.; Sabnis, S. Credit card fraud detection using data analytic techniques. Adv. Math. Sci. J. 2020, 9, 1185–1196. [CrossRef]
[20] Puh, M.; Brki´c, L. Detecting credit card fraud using selected machine learning algorithms. In Proceedings of the 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Zagreb, Croatia, 20–24 May 2019; pp. 1250–1255. [CrossRef]
[21] Hema, A. Machine Learning methods for Discovering Credit Card Fraud. Int. Res. J. Comput. Sci. 2020, 8, 1–6.
[22] Kumar, M.S.; Soundarya, V.; Kavitha, S.; Keerthika, E.; Aswini, E. Credit card fraud detection using random forest algorithm. In Proceedings of the 2019 3rd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 21–22 February 2019; pp. 149–153. [CrossRef]
[23] Patidar, R.; Sharma, L. Credit card fraud detection using neural network. Int. J. Soft Comput. Eng. IJSCE 2011, 1, 32–38.
[24] Asha, R.; KR, S.K. Credit card fraud detection using artificial neural network. Glob. Trans. Proc. 2021, 2, 35–41. [CrossRef]
[25] Varmedja, D.; Karanovic, M.; Sladojevic, S.; Arsenovic, M.; Anderla, A. Credit card fraud detection-machine learning methods. In Proceedings of the 2019 18th International Symposium Infoteh-Jahorina (Infoteh), Novi Sad, Serbia, 20–22 March 2019; pp. 1–5. [CrossRef]
[26] Breunig, M.M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying density-based local outliers. In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, TX, USA, 15–18 May 2000; pp. 93–104. [CrossRef]

[27] Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Ballarat, VIC, Australia, 15–19 December 2008; pp. 413–422. [CrossRef]
[28] John, H.; Naaz, S. Credit card fraud detection using local outlier factor and isolation forest. Int. J. Comput. Sci. Eng 2019, 7, 1060–1064. [CrossRef]
[29] Dal Pozzolo, A.; Caelen, O.; Johnson, R.A.; Bontempi, G. Calibrating probability with undersampling for unbalanced classification. In Proceedings of the 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, 7–10 December 2015; pp. 159–166. [CrossRef]
[30] Sahin, Y.; Duman, E. Detecting credit card fraud by ANN and logistic regression. In Proceedings of the 2011 International Symposium on Innovations in Intelligent Systems and Applications, Istanbul, Turkey, 15–18 June 2011; pp. 315–319.
[31] Kokkinaki, A.I. On atypical database transactions: Identification of probable frauds using machine learning for user profiling. In Proceedings of the 1997 IEEE Knowledge and Data Engineering Exchange Workshop, Nicosia, Cyprus, 4 November 1997; p. 109.
[32] Piryonesi, S.M.; El-Diraby, T.E. Data analytics in asset management: Cost-effective prediction of the pavement condition index. J. Infrastruct. Syst. 2020, 26, 04019036. [CrossRef]
[33] Maes, S.; Tuyls, K.; Vanschoenwinkel, B.; Manderick, B. Credit card fraud detection using Bayesian and neural networks. In Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies, Brussel, Belgium, 16–19 January 2002; pp. 261–270.
[34] Syeda, M.; Zhang, Y.Q.; Pan, Y. Parallel granular neural networks for fast credit card fraud detection. In Proceedings of the 2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No. 02CH37291), Atlanta, GA, USA, 12–17 May 2002; pp. 572–577. [CrossRef]
[35] Seeja, K.; Zareapoor, M. Fraudminer: A novel credit card fraud detection model based on frequent itemset mining. Sci. World J. 2014, 2014, 1–10. [CrossRef]
[36] Scikit-Learn-Contrib. Imbalanced-Learn. Available online: https://github.com/scikit-learn-contrib/imbalanced-learn (accessed on 22 January 2022).