|  |  |
| --- | --- |
|  | |
| Scientific Research + Literature –Assessment 2  TU060 : Individual Annotated Bibliography on Credit Card Fraud Topic | |
|  | |
| Ciaran Finnegan – Part Time – First Year 2021/2022  MSc in Computer Science (Data Science)  Student No : D21124026  27/3/2022 |  |
|  |  |

Table of Contents

1 Technical Topic – Data Analytics Specialism 3

1.1 Assignment Submission Topic 3

1.2 High Level Description of Journal / Articles in this Report 3

2 Bibliography of Sources 4

2.1 Anomaly Detection Algorithms for Credit Card Fraud Detection 4

2.2 Feature Reduction for Credit Card Fraud Datasets 6

2.3 Assessment of SVM and Random Forest Algorithms in Unbalanced Credit Card Datasets 8

2.4 Deep Learning ML for Credit Card Fraud Detection 10

2.5 Optimising XGBoost for Credit Card Fraud Detection 12

# Technical Topic – Data Analytics Specialism

## Assignment Submission Topic

The chosen topic for this research bibliography assignment is;

*Techniques to improve the Machine Learning workflow processes when attempting to detect credit card fraud with highly imbalanced datasets.*

## High Level Description of Journal / Articles in this Report

Section 2 of this assignment report provides a detailed overview of the articles/journals selected and reviewed for this particular topic of Machine Learning techniques in Credit Card Fraud detection.

Below is a high-level list of the subjects covered in the selected articles/journals;

* Specialist anomaly detection algorithms for unbalanced credit card fraud datasets. (Section 2.1)
* Feature Engineering within unbalanced credit card fraud datasets. (Section 2.2)
* Assessing SVM and Random Forest in credit card fraud detection. (Section 2.3)
* Deep Learning techniques in credit card fraud detection. (Section 2.4).
* Optimising XGBoost for credit card fraud detection. (Section 2.5)

# Bibliography of Sources

## Anomaly Detection Algorithms for Credit Card Fraud Detection

***Reference***

Ceronmani Sharmila, V., R., K., R., S., D., S., & R., H. (2019). Credit Card Fraud Detection Using Anomaly Techniques. *2019 1St International Conference On Innovations In Information And Communication Technology (ICIICT)*, *1*(1), 1-4. doi: 10.1109/iciict1.2019.8741421

***Aim***

To demonstrate the efficiencies of Local Outlier Factor (LoF) and Isolation Forest (IF) algorithms in detecting credit card fraud. The paper describes that credit card fraud can be a relatively rare occurrence, and that specialist anomaly detection algorithms can be shown to be less computationally expensive than more ‘generic’ algorithms in terms of predictive ML model building.

***Methods***

A historical (Kaggle) dataset of 27K credit card transaction is loaded into an ML experiment run in Jupyter Notebooks, with all required processing libraries preloaded.

The data is split between test and training sets, in line with standard ML workflow practice, and processed in sequence by the unsupervised LoF and IF algorithms to build models and assess accuracies. These algorithms provide an aggregate framework to identify the ‘anomalies’, which are the transactions marked as fraudulent.

LoF/IF detection performance is measured against output from models trained with more established algorithms.

Data visualisation techniques are used to display the key features in the dataset, and their relevance to the anomaly scoring.

***Conclusions***

This conference article declares that the LoF/IF algorithm used by the authors is quicker and more efficient in fraud detection than general Decision Tree approaches, and is also more understandable than Neural Network techniques.

This assertion is not backed up by any tables of actual metric data, just by the narrative text from the authors stating that their LoF/IF approach is better.

***Gaps in the Research:***

The paper does not contain any tabular data with which to strengthen the argument that LoF and IF algorithms are more accurate and faster at detecting fraud.

The range of other algorithms against which LoF and IF are compared is limited to a brief discussion on Decision Trees and Neural Networks.

***Critique***

There is important reference in the paper than speed can be important in credit card fraud detection, particularly with real-time requirements. The implementation of the aggregation LoF and IF algorithms is well thought out and presented with sound logic.

However, although the visualisations on feature importance are interesting, what is missing is tabular data on actual model accuracy, including the False Negative rate, and benchmark on speeds against other algorithmic approaches.

## Feature Reduction for Credit Card Fraud Datasets

***Reference***

Lima, R. and Pereira, A., 2017. Feature Selection Approaches to Fraud Detection in e-Payment Systems. *Lecture Notes in Business Information Processing*, [online] pp.111-126. Available at: <https://www.researchgate.net/publication/313731885\_Feature\_Selection\_Approaches\_to\_Fraud\_Detection\_in\_e-Payment\_Systems> [Accessed 11 September 2020].

***Aim***

E-Payment systems tend to generate not just a high volume of records but also a significant number of features for each transaction. Effective feature engineering of such datasets is adversely impacted by the high data imbalance of fraud and non-fraud credit card transaction records. This research paper demonstrates a scenario that shows that certain resampling techniques can counter this imbalance and deliver effective feature reduction and improved fraud detection accuracy.

***Methods***

A real-world historical dataset is obtained from a Latin American e-payment system (PagSeguro). All the transactions in this dataset are labelled ‘fraud’ or ‘non-fraud’, although the volume of ‘fraud’ records is only just over 1% of the entire dataset.

A series of Supervised Machine Learning experiments are conducted using a combination of seven chosen re-sampling techniques and three feature selection approaches. Three Classification models (Bayes, Decision Tree, and Logistic Regression) are applied, using 8-Fold Cross Validation, and accuracy is measured primarily using the AUC (Area Under Curve) metrics, with particular attention to the True Positive rate.

***Conclusions***

A tabular set of results using AUC metrics showed the relative performance in feature selection and subsequent credit fraud detection rates after various re-sampling techniques were applied.

Within the scope of these experiments, the best results were obtained by a re-sampling technique called ‘Random Undersampling’ followed by the ‘Relief’ approach to feature reduction.

The researchers conclude the paper by declaring that after applying a financial cost measure to the detected credit card fraud records their techniques delivered a nearly 58% improvement in reducing fraud losses, when compared to analysis conducted in the past by the PagSeguro Corporation itself.

***Gaps in the Research:***

There are no details provided on the fraud detection methods previously employed by the PagSeguro Corporation. The assertion from the researchers that their methods would have reduced fraud losses by 58% lacks a degree of context.

The number of records in the dataset is not given in this paper, and this is an unusual omission. (This dataset appears in other research papers, so it is known to contain over 1 million entries).

***Critique***

The source dataset contains 380 columns and the experiment described in the research paper applies a logical and effective approach to reduce this number of features prior to classification. The difficulties for feature reduction with highly imbalanced datasets are well described, and the results of the experiments are well tabulated.

A proficient Data Analyst should be able to replicate these experiments with other datasets and generate new benchmarks. Only three Classification models are applied in these experiments so a more comprehensive suite of algorithms would be desirable in future to further ground the accuracy assertions of the researchers.

## Assessment of SVM and Random Forest Algorithms in Unbalanced Credit Card Datasets

***Reference***

Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. Decision Support Systems, 50(3), 602–613. <https://doi.org/10.1016/J.DSS.2010.08.008>

***Aim***

To evaluate the effectiveness of Support Vector Machine (SVM) and Random Forest (RF) algorithms in credit card fraud detection in a highly imbalanced dataset. The effectiveness of the ML experiments will be compared against what are considered to be ‘established’ results from prior Logistic Regression (LR) analysis.

***Methods***

Research is conducted through a series of Supervised Machine Learning experiments, with a labelled dataset generated from real-world information by an international credit card operator.

The original dataset contained nearly 50 million records, but this was reduced to a sample of 350K fraud/non-fraud records, within which fraud records constitute less than 1% of the overall volume of credit card transactions.

The ML experiments conduct an iterative set of resampling, feature engineering, and modelling steps and accuracy is primarily measured by AUC metrics.

***Conclusions***

The ML experiments using RF algorithms generated better fraud detection performance metrics than both SVM and Logistic Regression.

The primary benchmark table in the paper was produced using an oversampled fraud dataset (10% fraud), which produced an AUC value of **0.953** for RF, **0.942** for LR, and **0.908** for SVM.

The tabular and graphical results from many of the research experiments showed that Logistic Regression outperformed many of the SVM runs, despite being considered a more ‘traditional’ model.

***Gaps in the Research:***

The dataset upon which the ML experiments were conducted did not contain any timestamp features. It was therefore not possible to look at time series or sequence analysis and derive any observations. The paper acknowledges this deficiency and sees the inclusion of timestamp data as an important future consideration.

There is also very little effort invested by the researchers in hyper-parameter tuning for any of the algorithms. The authors concede that this could have contributed to the relatively poorer performance by the SVM models.

The source dataset is from 2007, so would not contain more contemporary credit card attributes such as ‘chip and pin’ data. However, it is still a feature rich dataset that does not contain any obsolete characteristics.

***Critique***

Although this paper is nearly twelve years old, much of the underlying analysis remains valid, particularly as there are still relatively few credit card fraud datasets in general circulation.

This research paper compliments the other articles in this bibliography report in that there is significant upfront emphasis and guidance on re-sampling techniques due to the general imbalance against fraudulent records. The experiments are described in detail in the paper and could be reapplied to more contemporary datasets, and still provide valuable insight.

## Deep Learning ML for Credit Card Fraud Detection

***Reference***

Nguyen, T., Tahir, H., Abdelrazek, M., & Babar, A. (2020). Deep Learning Methods for Credit Card Fraud Detection. Retrieved 25 March 2022, from https://doi.org/10.48550/arXiv.2012.03754

***Aim***

To demonstrate that recent (2020) advances in Deep Learning can improve on credit card fraud detection rates when compared against previously established Machine Learning techniques.

***Methods***

Three separate (highly imbalanced) datasets were used as the input for Machine Learning experiments to produce a performance matrix on credit card fraud detection. Six models were created for each dataset, four with Deep Learning methods, and two others using Support Vector Machine (SVM) and Random Forest (RF) algorithms respectively.

Three of the four Deep Learning approaches were a variation on Convolutional Neural Networks, and one was a Long Short Term Memory (LSTM Recurrent Neural Network) classifier.

Standard ML workflow approaches were applied, starting with data pre-processing steps, mainly for resampling, and the splitting of all datasets into Training Data (80%) and Validation Data (20%) in order to train the models. A tabular output was generated to compare model performance based on the following metrics; *Accuracy*, *Precision*, *Recall*, and *F1 Score*. Accuracy is not strong metric to distinguish model performance with highly imbalanced data, so particular focus was given to the F1 score and the ability to decrease False Negatives (high Recall).

***Conclusions***

When all methods were compared, the LSTM approach produced the highest F1 score, which was **84.85%**. Recall rates were relatively consistent across RF and the Deep Learning approaches, although SVM scored noticeably lower.

Another key observation from the experiments was that with the application of GPU computing and TensorFlow Deep Learning libraries the training time for LSTM, and the CNN approaches, was significantly less than with SVM and RF. One set of experiments, which used a dataset with 10 million rows, had to be abandoned completely for SVM and RF because of the training time required by those classifier algorithms.

***Gaps in the Research:***

The paper admits that the Deep Learning approaches could possibly deliver even better performance with hyper parameter tuning but this was not considered in these research experiments.

The comparison diagrams are effectively incomplete because the last dataset was too large to be classified by SVM and RF in the time allocated for the experiments. It is unlikely that this would have impacted on the final conclusions around LSTM Deep Learning performance, but it remains a gap in the research.

***Critique***

It terms of presentation, the paper would have benefitted from the inclusion of a comprehensive tabular display of performance metrics. The graphical bar charts are somewhat cluttered in the conclusion section of the paper.

The datasets used are a good cross section of available real-world information, with a good range of row size and features. However, the paper is somewhat inconclusive about the optimal resampling approach that could be applied with Deep Learning techniques.

## Optimising XGBoost for Credit Card Fraud Detection

***Reference***

Priscilla, C., & Prabha, D. (2020). Influence of Optimizing XGBoost to handle Class Imbalance in Credit Card Fraud Detection. *2020 Third International Conference On Smart Systems And Inventive Technology (ICSSIT)*, pp. 1309-1315. doi: 10.1109/icssit48917.2020.9214206

***Aim***

Resampling techniques, when applied to highly imbalanced datasets such as those common in credit card fraud detection, can introduce problems with the generation of unrepresentative ‘artificial’ minority class data (oversampling), or the loss of important classification information by the reduction of of the majority class (undersampling). This paper demonstrates an approach to optimise the XGBoost classifier algorithm to improve detection of credit card fraud without the need for resampling.

***Methods***

A Machine Learning experiment was set up to directly compare the performance of the models generated separately by the optimised XGBoost (OXGBoost) and standard XGBoost algorithms.

Two real-world datasets were used as inputs; one with 569K rows of which 3% represented fraudulent credit card transactions, and another where fraud constituted only 0.1% of the records in a dataset of 284K rows.

The Training Data for the XGBoost model was subjected to a number of data resampling techniques before modelling against the Test Data. OXGBoost did not resample the data before modelling.

A series of hyper parameter setting changes were applied to XGBoost, using the ***RandomizedSearchCV*** procedure in the *Scikit-learn* Python library, to create OXGBoost. These steps are described in detail in the paper, along with the mathematical underpinning of XGBoost itself.

A graphical matrix of results was produced comparing the model performances using ROC-AUC analysis, and Precision-Recall Curve values (FScore).

***Conclusions***

The OXGBoost model performed substantially better on the dataset with 3% imbalance, with an AUC score of **0.966** and FScore of **0.816** (versus **0.94** and **0.627** respectively for XGBoost).

The difference in results for the highly imbalanced (0.01%) dataset were slightly less striking but OXGBoost still scored better with an AUC metric of **0.981** and FScore of **0.873** (versus **0.977** and **0.857** respectively for XGBoost).

The significance of the performance results are that XGBoost employed a range of resampling techniques, but none outperformed the optimisations that were applied to create OXGBoost.

***Gaps in the Research:***

Only one optimisation technique is described by the authors in the creation of the OXGBoost algorithm. Although the experiment results clearly show that the improvements are effective at avoiding the need for resampling, there is clearly an opportunity to further research in additional XGBoost optimisation approaches.

***Critique***

This paper is an interesting comparison to other research in this report, all of which rely, to some degree, on the resampling of highly imbalanced data. Resampling has the potential to corrupt the base data used for classification, in credit card fraud or other domains. This paper is a concise and well-constructed examination of how to avoid the need for resampling with highly imbalanced datasets through algorithm tuning.