# Artificial Intelligence: Evaluating machine learning algorithms on credit card fraud detection

Thesis submitted in fulfillment of the requirements for the degree **Post Graduate Diploma in Data Analytics** at The Independent Institute of Education, Varsity College**.**

***Kashik Ramnath***

ST10117137

*Supervisor: Mr. R, Ponnan*

*06 / 11 / 2023*

## Declaration

I hereby declare that the Research Report submitted for the Post-Graduate Diploma in Data Analytics to The Independent Institute of Education is my own work and has not previously been submitted to another University or Higher Education Institute for degree purposes.

Signature

07/11/2023

Date

## Acknowledgments

I thank Mr. R. Ponnan (Varsity College Durban North) for guidance on research proposal and research paper.

## Title

Artificial Intelligence: Evaluating machine learning algorithms on credit card fraud detection

## Abstract

The modern progresses of ecommerce and digital internet payments have assisted in the development and popularization of credit card fraud, making it one of the most common types of fraud. It is therefore important to assess which of various machine learning algorithms are most effective at detecting fraudulent transactions.

The research will be conducted on true world data from European cardholder transactions from September 2013 gathered secondarily from Kaggle. The algorithms being tested in this paper includes Logistic Regression (LR), Decision Trees (DT), Random Forest (RF), Gaussian Naive Bayes (NB), K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). The results will be evaluated using the metrics of accuracy, precision, recall, f1 score, and confusion matrixes for classification models.

This study concludes that Random Forest was evaluated to be the highest performing and most consistent model for detecting fraudulent and non-fraudulent credit card transactions for both balanced and skewed datasets.

The recommended continuation of study may include the evaluation of neural networks/deep learning algorithms on credit card fraud detection, and the evaluation of performance on hyperparameter-tuned machine learning models on credit card fraud detection.

# Key terms

Algorithm – A set of instructions or procedure computed by software or hardware to solve an issue; Tech Target (2023).

Artificial Intelligence – Artificial intelligence or AI is the uniting of computer science, deep learning, machine learning, sets of algorithms and datasets to empower problem solving and simulation of human intelligence; IBM (2023).

Card fraud – Criminal act of identity thievery through unauthorized taking of a person's credit card information to remove funds from or charging purchases; Cornell Law School (2023).

Dataset – The building-blocks of a database, datasets are collections of information and data points in the form of numerical, variable, categorical or correlations structured to be related to specific subject; Technopedia (2023).

Machine learning – Machine learning or ML is a sub-division/category of AI which is the process of using pattern recognition computers use to make predictions based on fed datasets; Hewlett Packard Enterprise (2023).

Machine-learning-based fraud detection – Another term for 'algorithmic fraud detection'. Follows the method of a rules-based fraud detection system, however, doesn't require human intervention for updating of rules as it is automatic, Fraud (2023).

Model – Refers to a machine learning model or file that has been trained with datasets to identify specific patterns using an algorithm to develop reasoning over and the learning from the data; Microsoft (2023).

Multicollinearity – Statistical concept where several independent variables in a model are correlated; (Investopedia, 2023).

# Contents

# List  of tables and figures

## Tables

## Figures

# Introduction

Artificial intelligence was built upon the work of Alan Turing in 1950, then developed in 1952 by Arthur Samuel in 1952 then soon gained its name in 1955 by John McCarthy; Tableau (2023). Ever since this time-period leaps in artificial intelligence technology have been taken and continue to do so to this very day. Artificial intelligence (AI) is a computers systems ability to perform tasks, humans are usually known for solely operating, IBM (2023). Machine learning (ML) is AI's set of learning algorithms and statistical models to identify and learn from data patterns, Hewlett Packard Enterprise (2023).

These concepts can be applied to studying a user's financial history to identify irregularities or abnormal behavior to predict, detect and decrease instances of card fraud (fraud committed by criminals using a payment card which often belongs to another person).

The current traditional methods of card fraud detection and prevention includes the rule-based fraud detection (the identification of fraud built upon recognizing abnormal attributes, which then flags the attribute as potentially fraudulent, according to Fraud (2023)) and security methods the holders of the credit cards follow, which in a lot of cases ceases the opportunity for card fraud to take place. However, criminals are finding ever-evolving methods to commit card fraud which results in the rule-based system to be constantly updated; this is cumbersome work. Instances of card fraud always slip through and therefore an updated method or another line of defense in the form of artificial intelligence/machine learning card fraud detection systems can benefit users.

Although this is not a foreign concept, recent major leaps in AI technology makes this a lucrative and enticing tool for organizations/institutions to adopt. It is important that it is researched how machine learning card fraud detection model algorithms compare to another in effectiveness, uses, performance, benefits, and shortcomings.

## Research Questions

The hypothesis that will be investigated in the research includes the following:

Which algorithm is most effective at detecting credit card fraud?
**$H_0$:** There is no specific algorithm that consistently performs above the others.
**$H_1$:** Random Forest algorithm outperforms other algorithms.
**$H_A$:** There exists one classification algorithm that outperforms other models consistently.

Which machine learning model is most successful at classifying card fraud in a skewed dataset?
**$H_0$:** There exists no model that outperforms others consistently.
**$H_1$:** Logistic Regression outperforms other classification models.
**$H_A$:** There exists one algorithm that outperforms other models consistently.

## Literature review

Theoretical frameworks used in the study is the Fraud Triangle Theory and Fraud Diamond Theory.

According to the National Whistleblower Center (2023), the Fraud Triangle Theory is a framework of anti-fraud criminological research model by Edwin Sutherland and Donald R that conceives the factors – motivation/incentive/pressure, opportunity, and rationalization - that leads to card fraud and higher risk of fraud. According to Sujana, Yasa, and Wahyuni (2018); Fraud Diamond Theory is the further developed idea of 1953's the Fraud Triangle Theory, with the addition of another factor, by Wolfe and Hermanson in 2004, known as 'capacity' or 'capability'.

The motivation/incentive/pressure according Abdullahi and Mansor (2015), is the perpetrator being pressured or motivated into performing card fraud, such as financial needs or issues (debt for example), or in non-financial cases such as greed or stress.

Opportunity is the window in which someone has the chance to use their abilities/skills to commit card fraud; Sujana, Yasa, and Wahyuni (2018). This window or opportunity is provided by an unsuccessful/ineffective control or governance system, or it can be perceived as is by the perpetrator rather than being real existing issues with fraud control.

Rationalization is simply the perpetrator developing a mental idea of acceptance of why his/her unlawful actions are morally acceptable, such as needing to provide income to support oneself or their family; Abdullahi and Mansor (2015).

Capability, as introduced in the Fraud Diamond Theory, is the perpetrator possessing the resources, traits, abilities/skills, intelligence/creativity, or are in organizational position of power, to participate in card fraud, as even with the previous mentioned elements, if is not capable of performing fraud, they will not be able to do so; Abdullahi and Mansor (2015).

These collective elements will be addressed in the research, as by training a model that is efficient at detecting, minimizing and by nature decreasing card fraud. The ability for machine learning to quickly learn and adapt to fraudulent transaction detection will higher the bar for 'capability', meaning the skill level to commit fraud will need to be higher, the 'motivation' for committing card fraud will be discouraged, and the higher security will leave less 'opportunity' for perpetrators to abuse weaknesses in systems.

A study has performed similar research that found due to increasing popularity of cashless transactions, one of the most common frauds are credit card frauds. It is stated that, 'Although there is a tremendous volume increase in credit card transactions, the amount of frauds is proportionally the same or have decreased due to sophisticated fraud detection systems. However, fraudsters are constantly coming up with new ways to steal information', Karanovic, Sladojevic, Arsenovic, Varmedja, and Anderia (2019:1). Various machine learning algorithms such as RF, NB, MLP, LR, needs analyzing for the determination of the most appropriate algorithm for detection of credit card fraud detection. A confusion matrix was used to identify that random forest (RF) was best the performing algorithm, Karanovic, Sladojevic, Arsenovic, Varmedja, and Anderia (2019).

According to Khatri, Arora, Agrawal (2020), credit card fraud is strongly correlated with the rise of online transactions, and these transactions occur extremely quickly, and therefore a quicker response to detecting fraud in the form of machine learning rather than current conventional fraud detection and prevention methods. Supervised machine learning models performance measured in sensitivity, precision and time are therefore necessary. In the study naïve bayes classifier was the best performing and hence the better approach for detecting card fraud.

According to Geetha, Vaishnavi (2019), banks in recent years adopted to EMV smart cards which improved security on certain on-card payments however card-not-present fraud rates increased. This study used measures of performance measuring like the previously mentioned study, in the form of accuracy, precision and MCC. The results were measured before and after SMOTE, an oversampling technique for balancing the dataset. Both before and after applying SMOTE random forest algorithm, logistic regression and decision trees yielded the highest performances, the performance scores where so similar that a clear victor was not necessary to commend.

Banks and financial institutions are offering credit card fraud detection applications as they are in high value and demand. Fraudulent transactions have increased to an extreme level, significantly affecting the economy through card-not-present and card-present frauds. The most efficient accuracy score in real-time credit-card fraud detection recorded was SVM with 91% accuracy. Performance in the methods of accuracy, precision, recall, true positive, false positive, f1, ROC and MCC was measured; Thennakoon, Bhagyani, Premadasa, Mihiranga, and Kuruwitaarachchi (2019).

## Research methodology

The data analysis is quantitative and follows the positivism paradigm and therefore the methods are related to analyzing and observing numerical, devoid of bias, factual data with the goal of identifying the highest performing model at predicting credit card fraud.

The study will follow an objective approach (quantitative) with a positivism paradigm. This is because as a researcher and objective analyst no forms of bias will occur as the study will be based upon and interpreted from quantifiable, observable data that will be statistically analyzed (judged as science via logic), rather than personal or related values, which is near definition of what a positivism paradigm is.

Empirical-analytical science will be performed. This will be ensued in the research as it is an experiment of observing and measuring data (real financial variables) in a tabular dataset to test different models' performances at predicting and classifying if the transaction is fraudulent. The performance calculation of the models will follow statistics-based performance measuring methods of measuring accuracy, precisions, f1 score, recall and R-squared. These findings will be compared and observed. Thus, the study follows mathematical principles, interpretation, and observation without any form of personal values or bias, adhering to a positivism paradigm, quantitative analysis, approach, and an empirical-analytical science.

The study intends to follow epistemology position of positivism. This is depicted in the study amidst reality being measurable rather than there being only one truth/reality, PEDIAA (2023); meaning that the phenomena are observable and measurable and provides credible facts (involved with the methods of attaining knowledge and its nature).

This study will follow an experimental and predictive research design, as it looks at the problem from a strictly scientific background understanding the impacts of the features/independent variables (v1 through v28, time and amount) and the dependent variable (class; whether a variable is fraudulent or not). The goal is accurately predicting the outcomes of fraudulent and non-fraudulent transaction, hence predictive research design.

The data analysis process consists of 5 main parts in specified order; performed in Visual Studio Code utilizing Python Jupyter notebooks:
Data collection, Data preparation, Data Exploration, Data Modeling, Result interpretation.

The step of data collection is the gathering of the datasets and other relevant information such as resources and research papers.

The data that will be used in this research is secondary and not primary data as it was collected and or sorted by person's/institutions outside the author and supervisors of this specific study; namely Machine Learning Group – ULB (Université Libre de Bruxelles), and Wordline.

The study makes use of datasets from Kaggle (Kaggle.com), a large renowned data science community and company. The datasets used will not be placeholder nor simulated transaction data, rather real-world sourced information, for the purpose of testing algorithms in real-case scenarios, yielding accurate-to-life results.

'The dataset contains transactions made by credit cards in September 2013 by European cardholders. This dataset presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions', Machine Learning Group – ULB (2021).

The dataset consists of 31 variables. 28 of the variables are PCA transformed variables, which means that they are a summary of a collection of data in numerical/quantifiable form. These 28 features/columns are labelled V1, through V28, without further context being given – this was done as the information being portrayed are real, sensitive, confidential information and therefore the protection of the anonymous people whose data it is needs to be followed (Ethical considerations). This can prove difficulty in understanding the data but should not affect the process of developing models.

The remaining 3 columns that are known is Time, Amount, and Class. Time is the duration in-between the first and last transactions in the dataset. Amount is the cost or money involved in the transaction. Class is the binary 1 and 0, representing fraudulent transaction and nonfraudulent transaction respectively.

The dataset is unbalanced, meaning that methods of balancing the data such as SMOTE, undersampling or oversampling will be needed to be evaluated.

The dataset aligns perfectly with the research being done as it is of the topic of credit card fraud, in which the purpose is to build models to classify and predict if a transaction is fraudulent, which is the purpose of the research and related the research questions and hypotheses. Finding credit card fraud datasets that are not simulated data is difficult as it involves the use of confidential information and therefore options are severely limited and resources are lacking, however, this dataset does fill all requirements necessary for the research to commence. The dataset is also vast in size, which results in a stronger models compared to other smaller simulated datasets.

Data preparation involves the cleaning of the dataset. Cleaning will involve removing nulls, redundant values, outliers, and values that will hinder the performance of the model in future steps. Data wrangling will not entirely be necessary as all the data is in usable forms and datatypes (numerical in nature).
However, the dataset is unbalanced and will require a method such as undersampling, oversampling, or SMOTE to balance the dataset. After this step's completion the data would be suitable for understanding and model building methods.

Data exploration involves evaluating each variables relationship on effecting the outcome and its existing or lack of relationships on other variables.
This will be visualized using graphs, plots, statistics, and correlations. It is important to improve understanding of the dataset during this stage. Redundant data will be excluded from the model-making process.

Data modelling is the process of creating and training the models. The data is to be split into train and test, which is the percentage of the dataset that will be used to train the models, and the percentage of the dataset which will be tested on the model to observe its ability to predict fraudulent transactions.

This ratio will be 0.2, which means 20% of the dataset will be used for testing or unseen data, while 80% of the dataset will be used for training the models.

The modeling process will be performed for various types of models and each model will be employ a relevant tuning technique to best portray its performance. These models will include Linear Regression, Logistic Regression, Decision Trees, KNN, SVM, Gaussian Naïve Bayes

Result interpretations are the final step. For each tuned model the performance in terms of accuracy, precision, R-square, f1 score and recall score will be compared to identify the highest performance across all evaluation variables. The scores will be ranked from highest to lowest performance and visualized using graphs of the statistical outputs.

Therefore, the data analysis methods are interpreted and observed statistical values following the quantitative, positivism paradigm in which there is no room for biases to occur or influence the outcome of the research and methods involved in the data analysis processes.

Ethical considerations that have been included are that the user’s sensitive information such as transaction specific information, identities, and context of the variables in the dataset are not available; they have PCA transformed into a quantifiable summary of a collection of statistics to protect privacy and information of the people the transactions belong to. The secondary data was originally sourced through collaboration of Worldline and the Machine Learning Group of ULB (Université Libre de Bruxelles); trusted computer science institutions that adhere to policies and ethically source resources.

Therefore, sensitive information has been operated appropriately with confidentiality and anonymity. The limitation to this is that due to lack of context, analyzing and understanding of the data has an added layer of difficulty, however this should not negatively affect results or model building, rather just require more work.

As a researcher, thine duty is to not falsify information, distort results, exert bias, misuse information or methods. All models in the methodology will incorporate their own training and tuning methods appropriately, regardless of time required to do so. All results of the methods will be strictly following statistical procedures to calculate accuracy, precision, score, etc, and cannot be biased. The data can’t be misused as the information is already publicly available with anonymous users/participants.

# Empirical analysis and results

## Empirical Analysis

The experiment was conducted on both, the dataset when undersampled, and the dataset without undersampling. Undersampling the dataset involved incorporating a balancing technique in which the data of the minority class is used as a cap to decrease the size of the majority class. This was performed via Imbalanced-learns's RandomUnderSampler library.

To address reliability, the training and testing dataset percentages remained the same ratio for all model methods used in the research. During this process, a set of random-states (a hyperparameter that shuffles the dataset values into random orders to prevent biases in the data training and data testing phase) variables will be used to test if similar results are returned each time the experiment methods are performed; this will be done for every detection model in testing. The tests will also be performed numerous times to further ensure reliability. All models tested are thus utilizing a dataset test-train split of 20% testing data and 80% training data, which all models are using a random state of 1.

Validity of the study is direct, as the purpose is to test, and examine models to identify the performance of different classification and prediction models to detect credit card fraud which is adhered to as the performance being measured and compared are the scores of accuracies, precision, recall , f1 score, including a confusion matrix – which are the data analytics recognized standards of statistical methods in measuring performance of machine learning models. The dataset being used is large, containing real-world data which further cements that it is likely the results of the model will return similar results in other use-cases with same dataset categories.

A Jupyter Notebook in Visual Studio Code, utilizing Python was used to conduct the full analysis and model creation process. This included imported libraries such as Seaborn (for statistical graphs), Matplotlib (visualisations), Pandas (data manipulation), numpy (mathematical operations), imblearn's under-sampling (resampling library), sklearn metrics (model result and performance evaluation) and sklearn model selection (Machine Learning model library). The dataset was examined and processed to ensure the feature data types were correct, finding and removing of null values and duplicated values that could hinder model performance. Except for a small percentage of duplicates, the dataset was primarily cleaned, however, very unbalanced.

Exploratory data analysis of the dataset has shown the imbalances and skewness of the target variable, being 99.83% non-fraudulent transactions, while the remaining 0.17% are fraudulent transactions. This implies that models trained without balancing the dataset beforehand will result in bias towards the majority class.

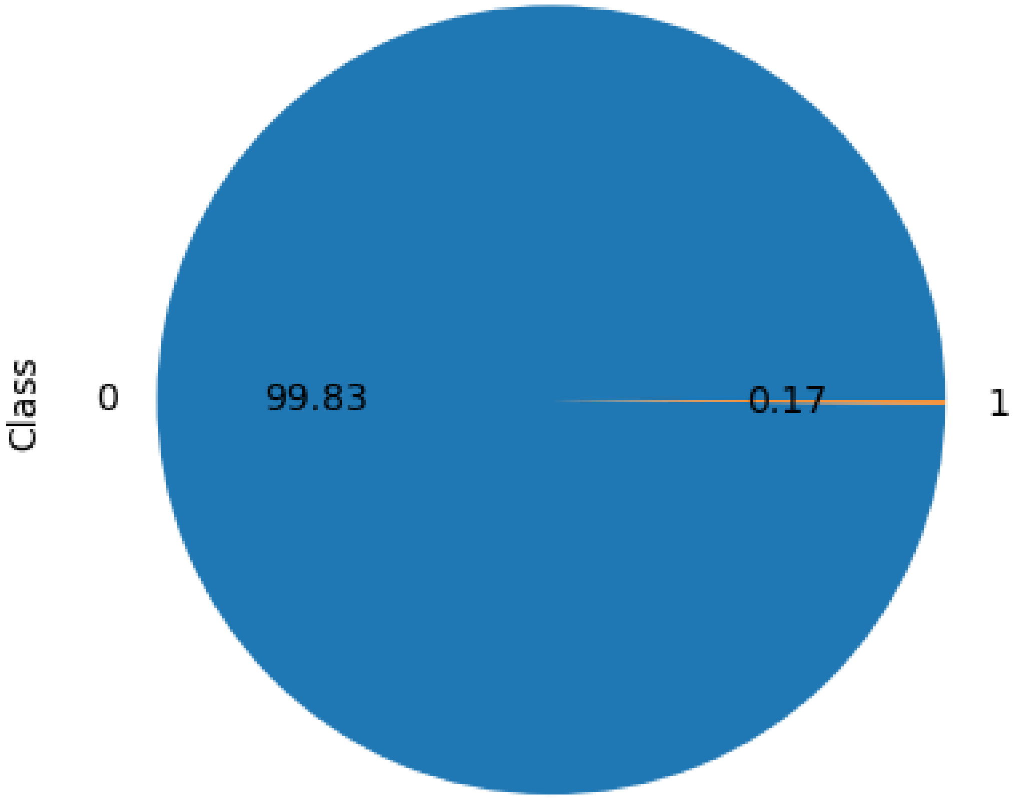


*Figure 0: Fraudulent & Non-Fraudulent data split*

The distribution of fraudulent transactions was found to be spread across the dataset increasing and decreasing along non-fraudulent transactions. The fraudulent transactions are skewed slightly to the left. Due to the very low amount of fraud transactions throughout the dataset it would be appropriate to use as much fraud samples as possible, which can be made possible through undersampling while avoiding hindering model performance.

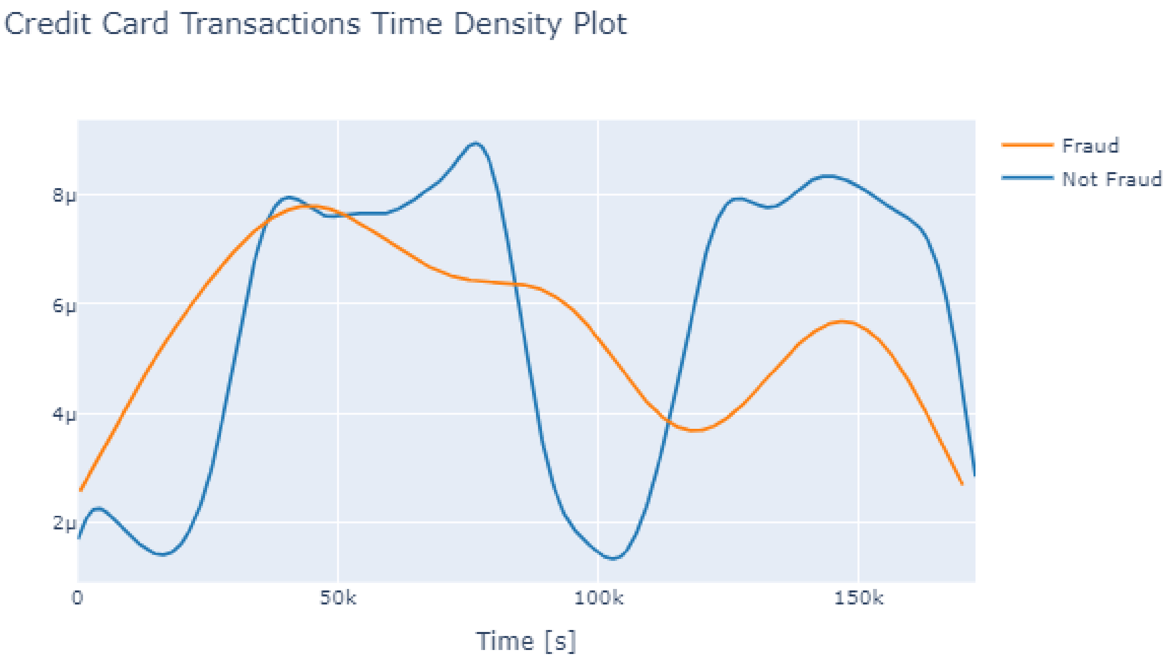


Figure 1: Credit Card Transactions Time Series Density Plot

This is correlation heatmap. It shows which variables have relationships or correlations. This concept of independent variables being correlated is known as multicollinearity; it is appropriate to avoid this phenomenon as it can become sensitive to small changes in the model. The heatmap portrayed no significant visual occurrences of multicollinearity that could lead to high variance of estimated coefficients and therefore no feature removal was required.

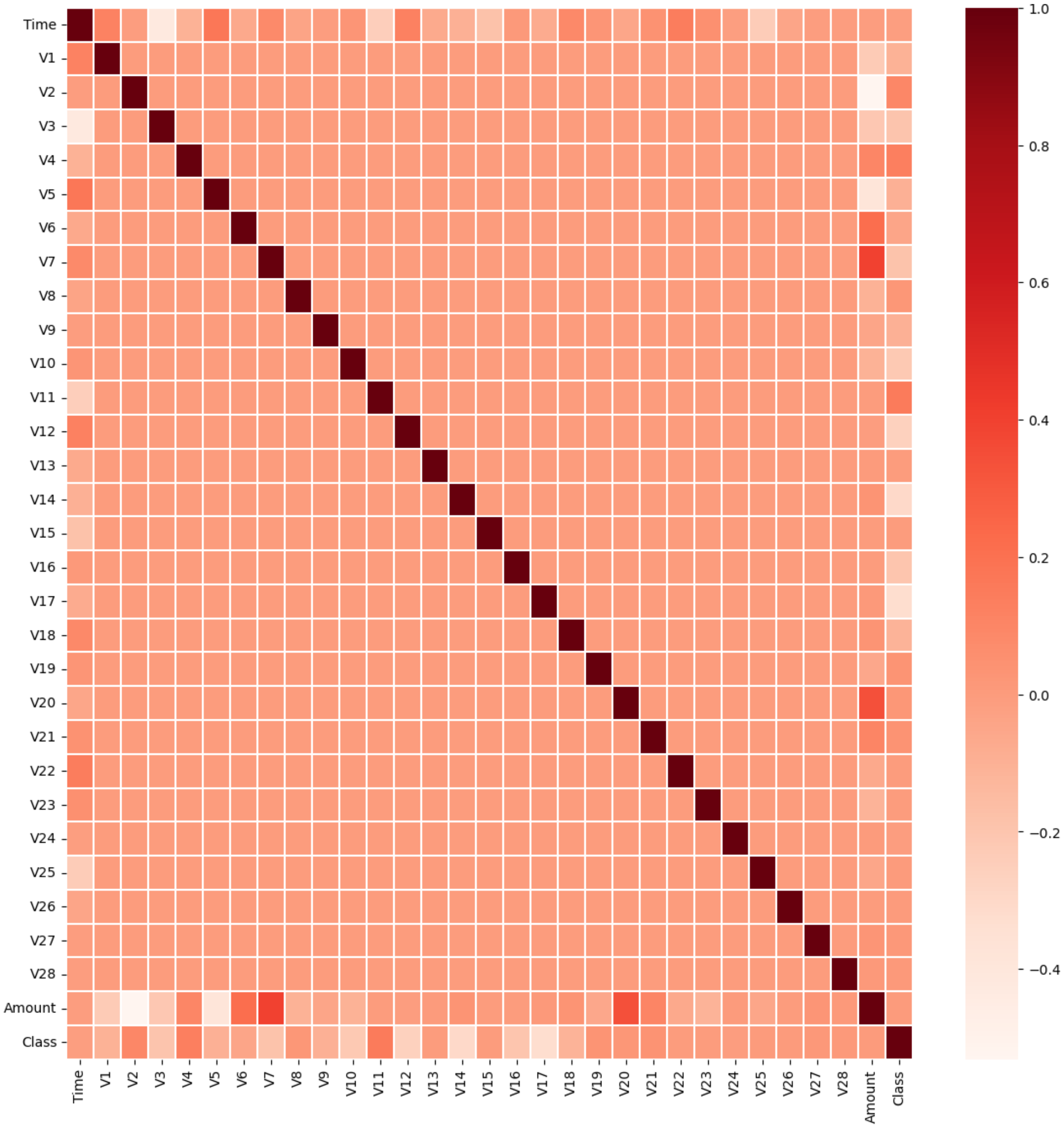
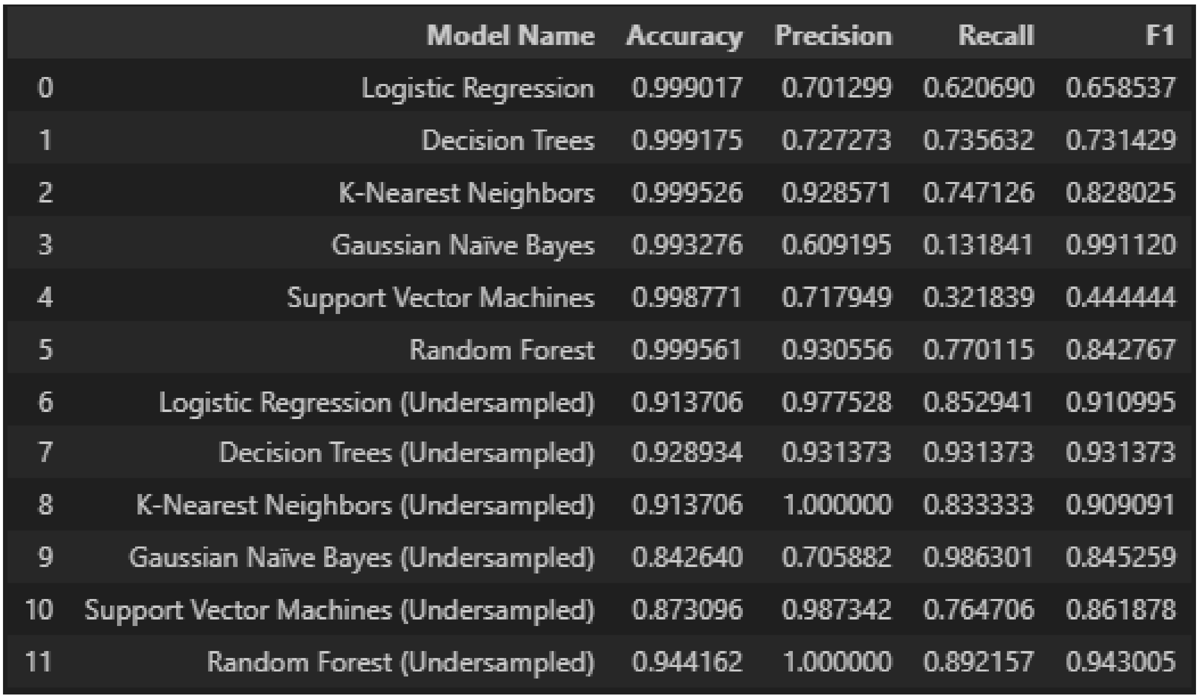


*Figure 1: Dataset Variable Correlation*

## Results

Refer to table 1 below; Accuracy of the models are the total number of predictions that each model has performed correctly, however, a high accuracy may not necessarily indicate a well-developed model, as it may be able predict a non-fraudulent transaction well, but not a fraudulent transaction. Precision refers to how many of the results that were predicted positive, are relevant/correct. It is how accurate the positive predictions are. Recall refers to the number of true positives predictions in the pool of actual positive instances, however, there exists a tradeoff between precision and recall, as increasing one decreases the other. F1 score is the harmonic mean of precision recall that balances both recall and precision (and considers the balance between false positives and false negatives) thus making it appropriate for unbalanced datasets. Therefore, the appropriate metric for assessing the classification model of an unbalanced dataset is the F1 score, while accuracy is more appropriate for balanced datasets.

*Table 01: Machine Learning Model Results - Evaluation Metrics*



| | Model Name | Accuracy | Precision | Recall | F1 |
|---|---|---|---|---|---|
| 0 | Logistic Regression | 0.999017 | 0.701299 | 0.620690 | 0.658537 |
| 1 | Decision Trees | 0.999175 | 0.727273 | 0.735632 | 0.731429 |
| 2 | K-Nearest Neighbors | 0.999526 | 0.928571 | 0.747126 | 0.828025 |
| 3 | Gaussian Naïve Bayes | 0.993276 | 0.609195 | 0.131841 | 0.991120 |
| 4 | Support Vector Machines | 0.998771 | 0.717949 | 0.321839 | 0.444444 |
| 5 | Random Forest | 0.999561 | 0.930556 | 0.770115 | 0.842767 |
| 6 | Logistic Regression (Undersampled) | 0.913706 | 0.977528 | 0.852941 | 0.910995 |
| 7 | Decision Trees (Undersampled) | 0.928934 | 0.931373 | 0.931373 | 0.931373 |
| 8 | K-Nearest Neighbors (Undersampled) | 0.913706 | 1.000000 | 0.833333 | 0.909091 |
| 9 | Gaussian Naïve Bayes (Undersampled) | 0.842640 | 0.705882 | 0.986301 | 0.845259 |
| 10 | Support Vector Machines (Undersampled) | 0.873096 | 0.987342 | 0.764706 | 0.861878 |
| 11 | Random Forest (Undersampled) | 0.944162 | 1.000000 | 0.892157 | 0.943005 |

As portrayed above, the models tested on the unbalanced and balanced datasets made use of scikit-learn’s imported libraries of Logistic Regression (LR), Decision Tree Classifier (DT), K-Neighbors Classifier (K-Nearest Neighbors / KNN), GaussianNB (Gaussian Naïve Bayes / NB), and SVM (Support Vector Machines).

During the splitting of the data for test and train, a test size of 0.2 was employed, indicating 80 percent training and 20 percent test size. The training split is the data that will be used to train the model, while the test split is the unseen data that the model will try to predict. The ‘random_state’ parameter of all models were set to 1 for reliability – this is the seed for a random number generator which we set stagnant for the purpose of reproducibility of results. The models that were tested on the undersampled dataset involved the same process, except prior, the dataset was balanced using the imbalanced-learn’s ‘RandomUnderSampler’, incorporating the ‘”not minority’’ ‘sampling_strategy’ parameter. The results were then calculated using scikit-learn’s (python library) confusion matrixes and metric reports as will be shown henceforth:

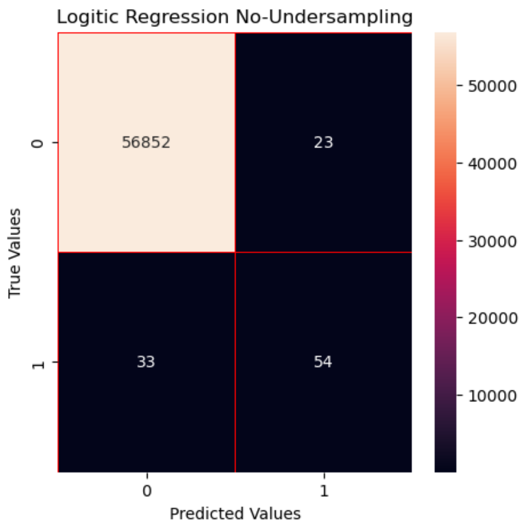


Figure 3: LR Confusion Matrix

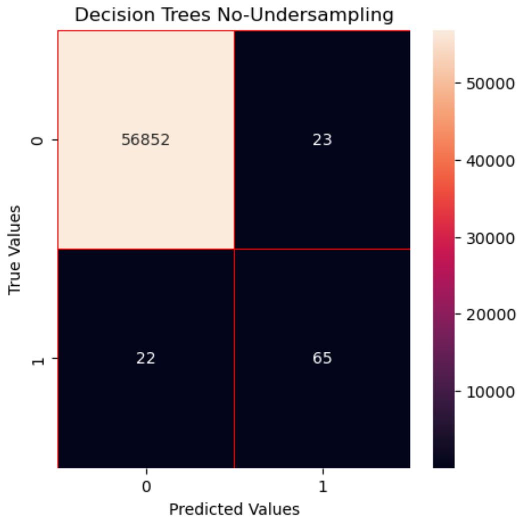


Figure 4: DT Confusion Matrix

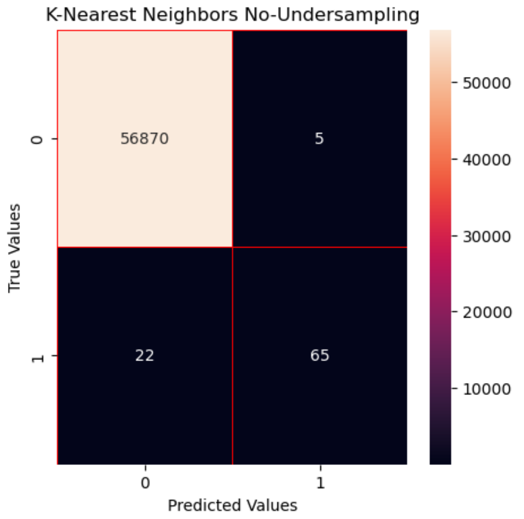


Figure 5: KNN Confusion Matrix

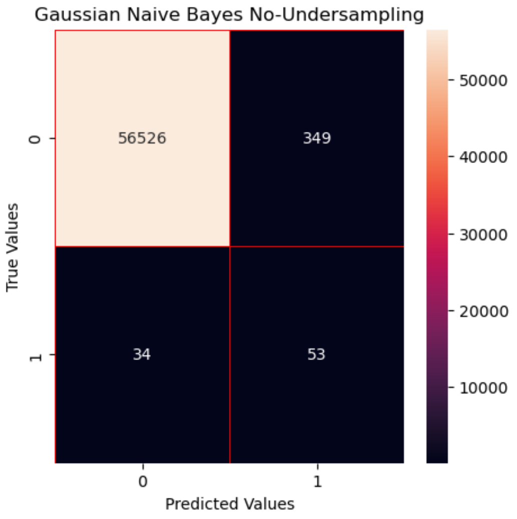


Figure 6: NB Confusion Matrix

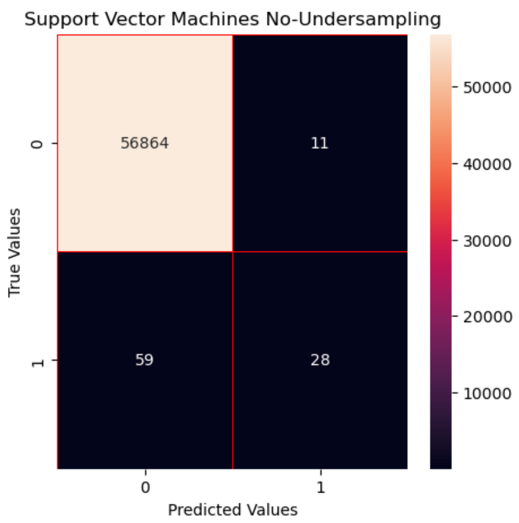


*Figure 7: SVM Confusion Matrix*

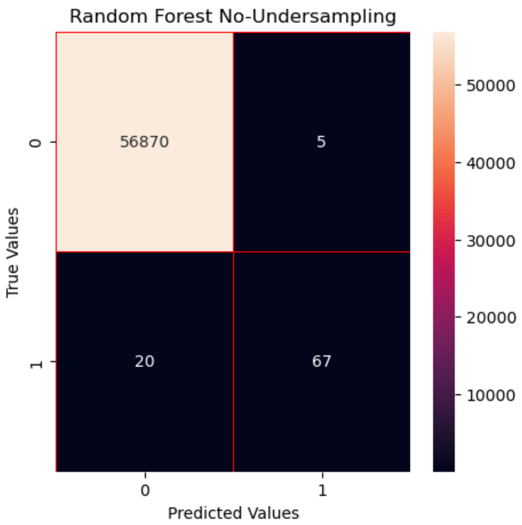


*Figure 8: RF Confusion Matrix*

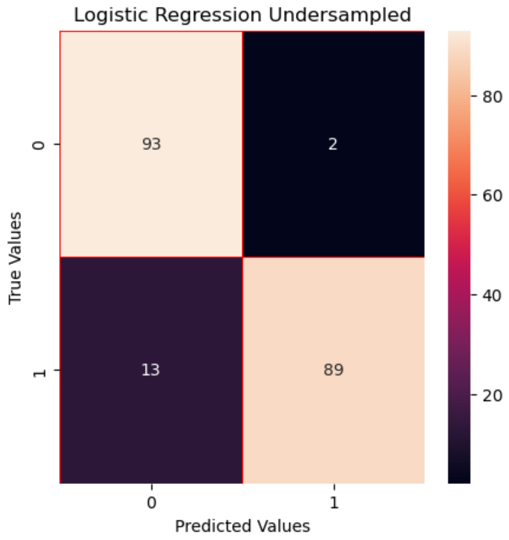


*Figure 9: LR Undersampled Confusion Matrix*

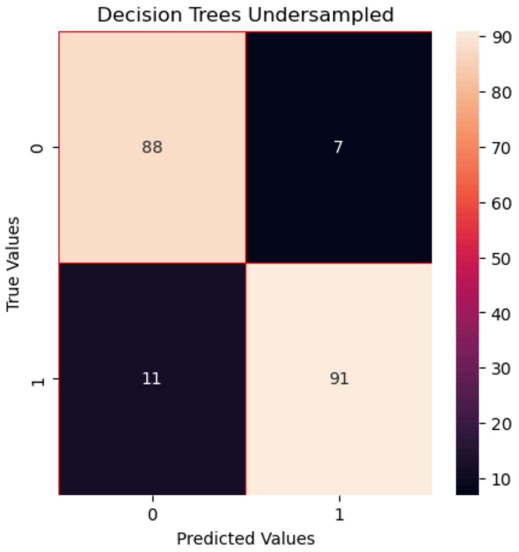


*Figure 10: DT Undersampled Confusion Matrix*

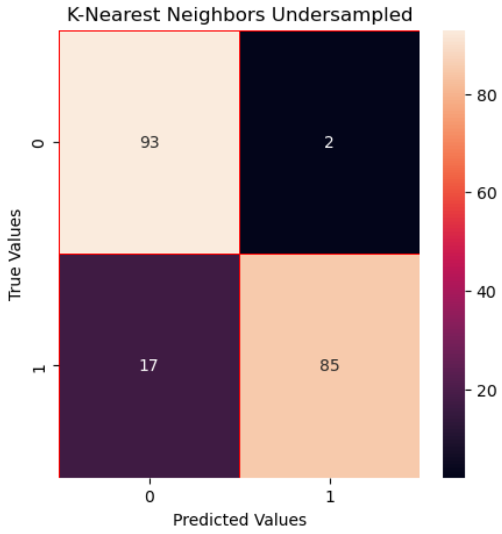


*Figure 11: KNN Undersampled Confusion Matrix*

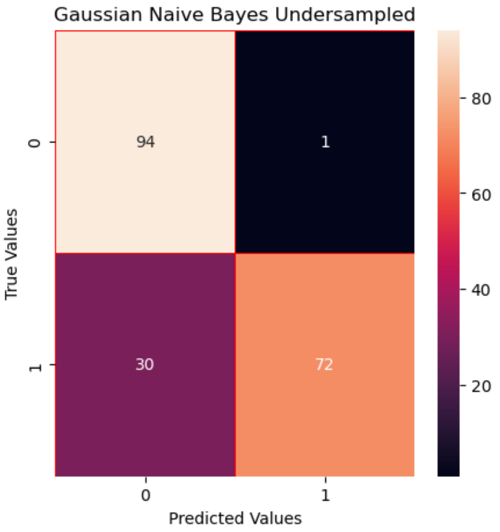


*Figure 12: NB Undersampled Confusion Matrix*

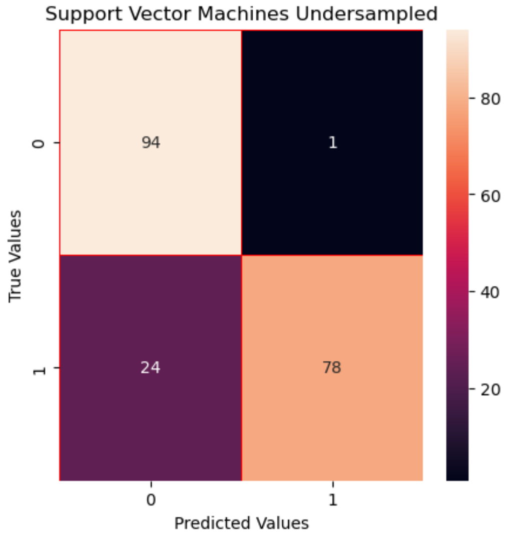


*Figure 13: SVM Undersampled Confusion Matrix*

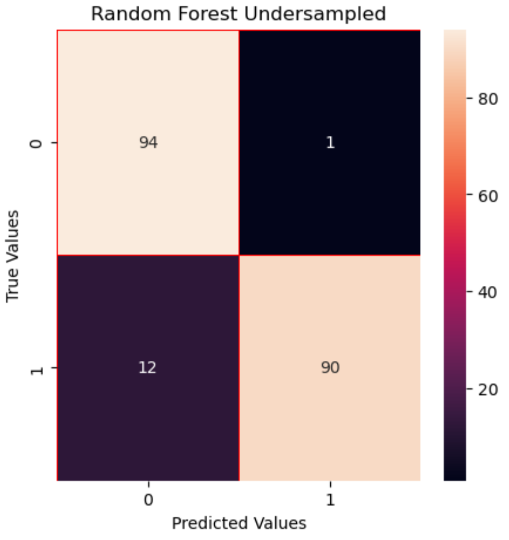


*Figure 14: RF Undersampled Confusion Matrix*

The confusion matrixes showed above use 1 and 0 to represent true and false respectively for the presence or absence of a fraudulent transaction. These quadrants can be labelled True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). These quadrants are the building blocks of the formulas that calculate accuracy, precision, recall and  f1 scores.

*Table 1: Confusion Matrix Quadrants Explained*

| | |
|---|---|
| True Positive (TP) | The model predicted positive (for fraudulent transaction), and it is true (bottom right quadrant). |
| True Negative (TN) | The model predicted negative, and it is true (top left quadrant). |
| False Positive (FP) | The model predicted positive, and it is false (top right quadrant). |
| False Negative (FN) | The model predicted negative, and it is false (bottom left quadrant). |

# Findings, Recommendations, and Conclusion

## Findings

All algorithms tested on the unbalanced dataset yielded high accuracy due to having a large training sample for identifying non-fraudulent transactions, but the ability of said models to identify fraudulent transactions were poor. Judging a model’s effectiveness if the dataset is unbalanced on the accuracy metric alone is not a correct evaluation, as the balancing of the dataset has not been considered, and will require the evaluation of recall, precision and f1 scores, however, the model with the highest accuracy metric on the unbalanced dataset is Random Forest with 99.9% accuracy.

The algorithms with the highest accuracy metric on the balanced dataset are the Random Forest algorithm with 94.4% accuracy, followed by Decision Trees (92.9%), and K-Nearest Neighbors (91,4%).

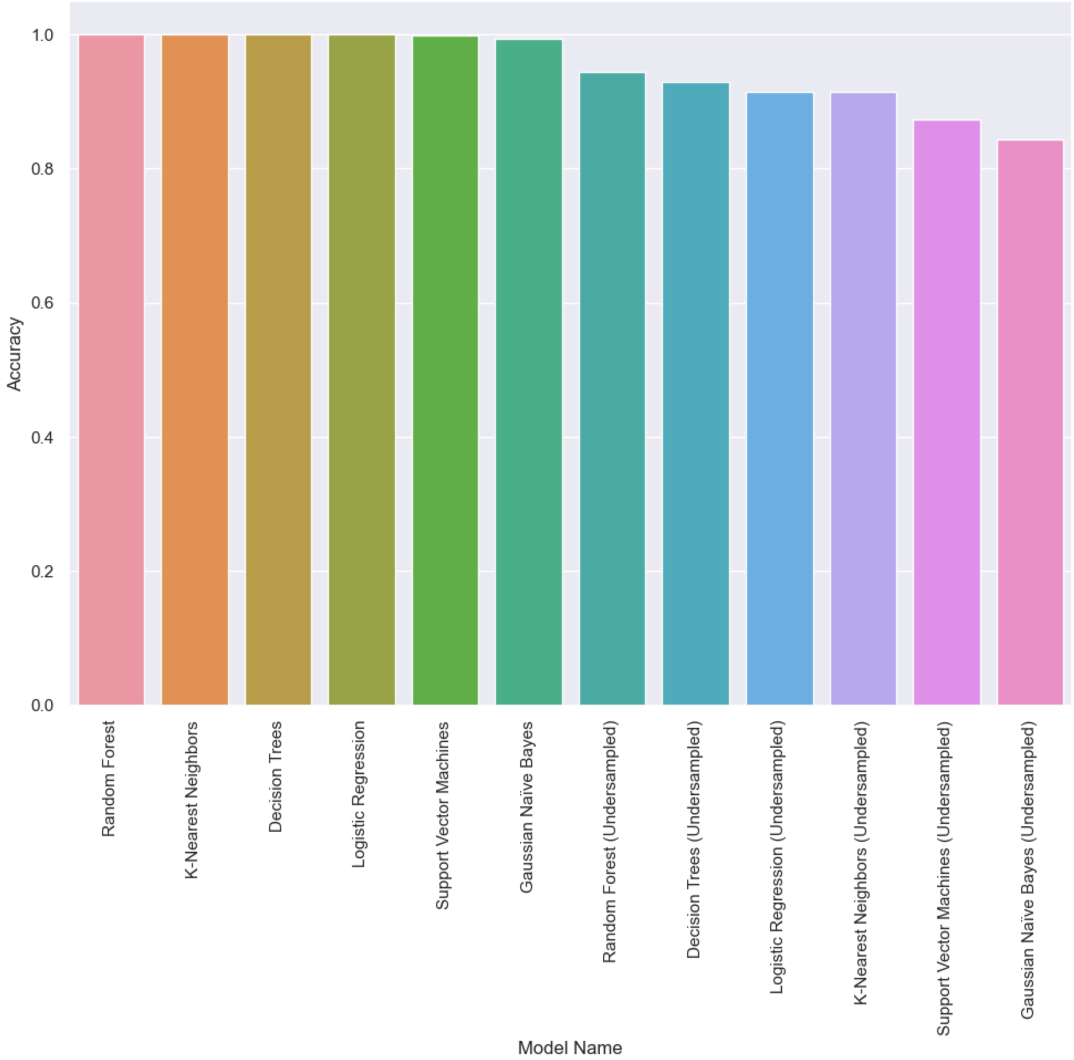


*Figure 15: Algorithm accuracy bar chart*

Following the metric of precision, all models tested on the unbalanced dataset had varying results, with Random Forest being the highest, 93.1%.

Most algorithms that were tested of the balanced dataset had higher precision, with K-Nearest Neighbors and Random Forest having perfect precision (100%), followed by Support Vector Machine (98.7%).

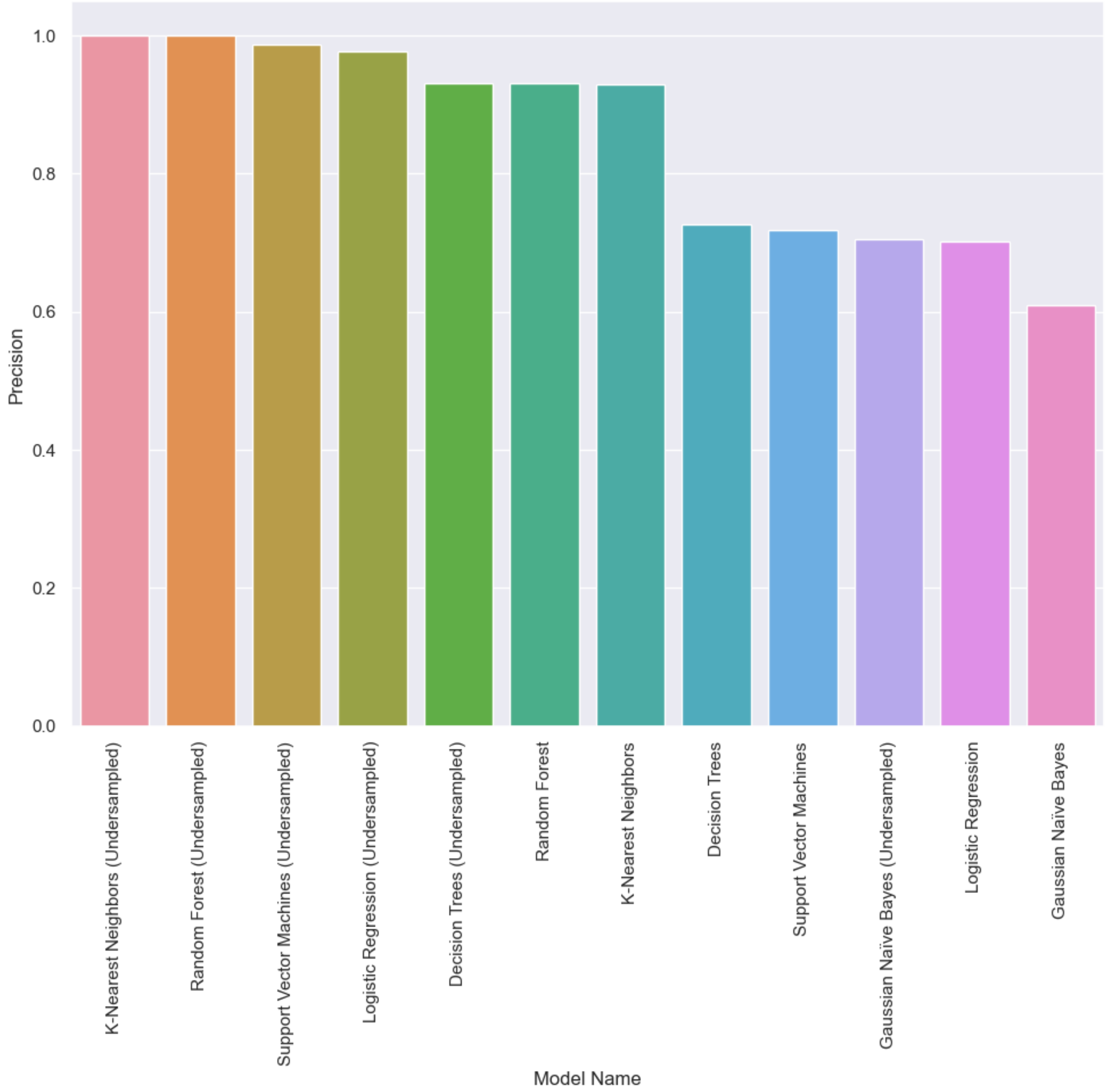


*Figure 16: Algorithm precision bar chart*

The algorithm with the highest recall metric for the unbalanced dataset is Random Forest with 77%.

The algorithm with the highest recall for the balanced dataset is Gaussian Naïve Bayes with 98.6% followed by Decision Trees(93.1%) and Random Forest (89.2%).

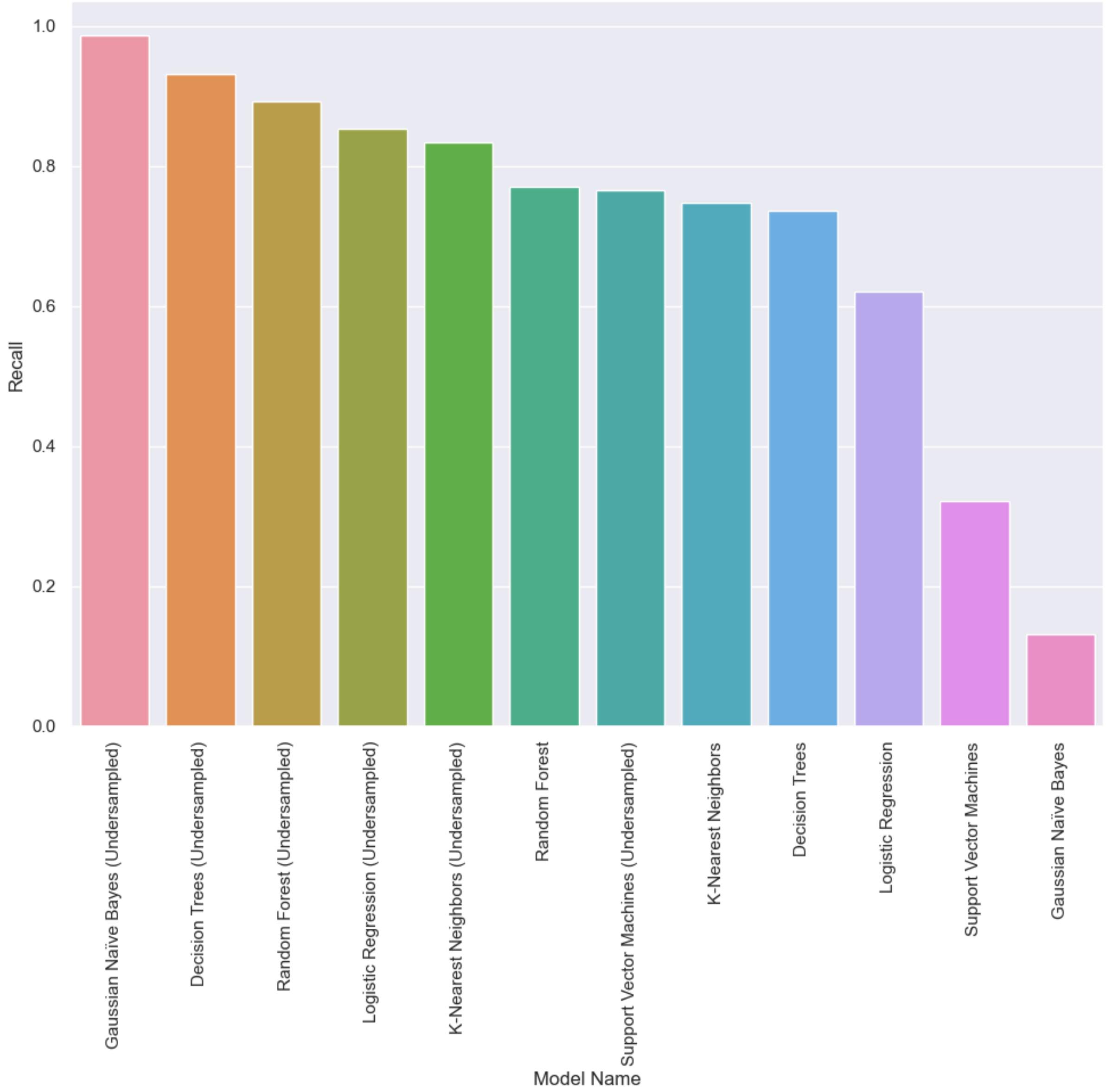


*Figure 17: Algorithm recall bar chart*

For the unbalanced dataset, Gaussian Naïve Bayes attained the highest F1 score (99.1%), however, considering the low recall and precision scores and the model is very sensitive to imbalances, the model overall, performed poorly.

The greatest performing algorithm for F1 score for the balanced dataset is Random Forest with 94.3%, followed by Decision Trees (93.1%) and Logistic Regression (91.1%).

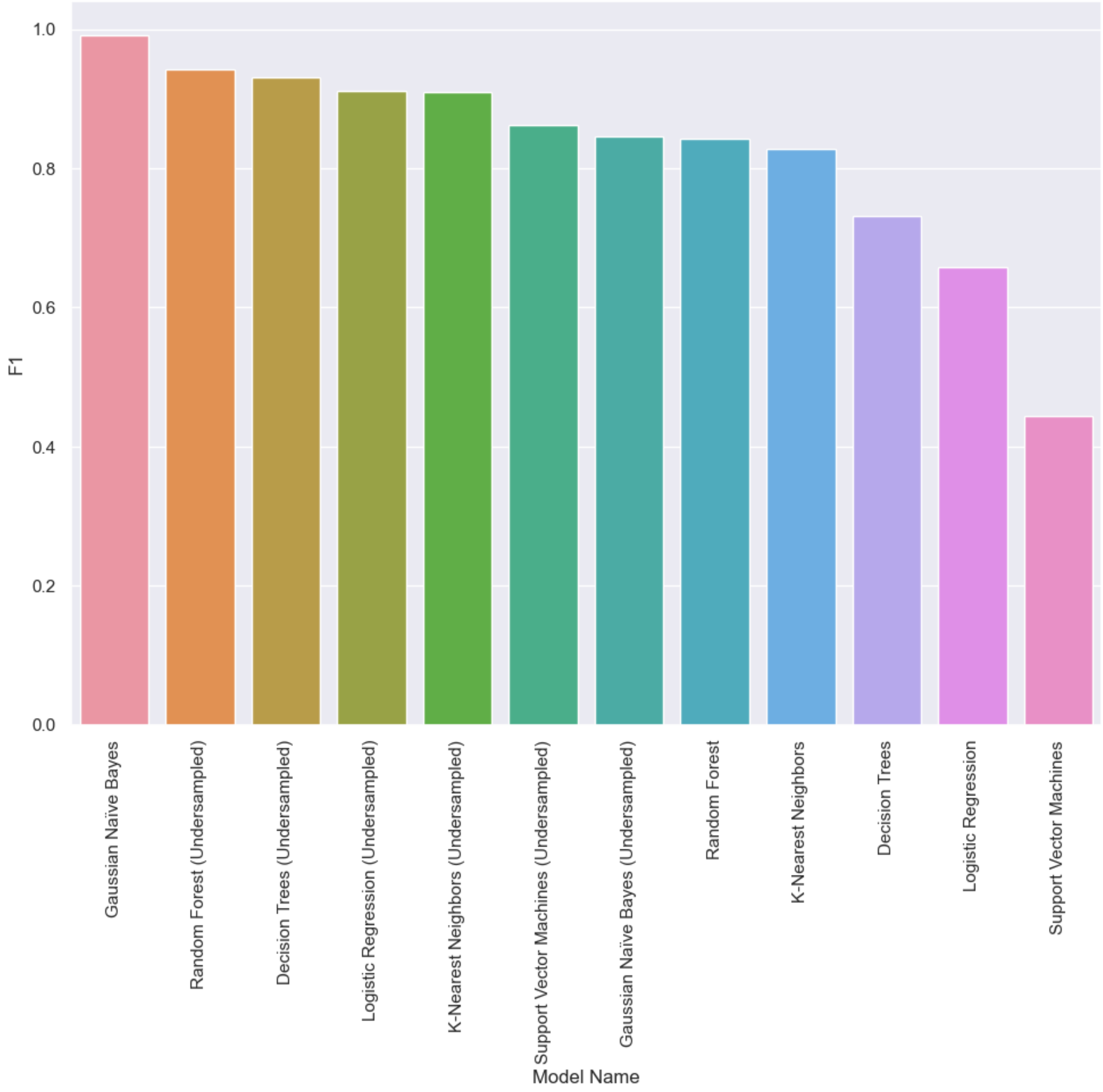


*Figure 18: Algorithm F1 bar chart*

Results show that the classification algorithm which outperforms other models is the Random Forest algorithm, thus proving; '**$H_1$:** Random Forest algorithm outperforms other algorithms', answering the research question of which model is most effective at detecting credit card fraud. Therefore, the null and alternative hypothesis is rejected.

The algorithm that was most successful at classifying credit card fraud in a skewed dataset is Random Forest, which proves the alternative hypothesis of '**$H_A$:** There exists one algorithm that outperforms other models consistently.', answering the research question of which machine learning model is most successful at classifying card fraud in a skewed dataset. Therefore, the hypothesis 1 and null hypothesis is rejected.

## Recommendations

Recommendations for future research may involve the evaluation of various deep learning models and neural network performance on identifying fraudulent credit card transactions, as it is a less performed approach than evaluating machine learning models performance. While this study evaluates a plethora of machine learning models, an extension of research would benefit from evaluating these model performances before and after the tuning and fine-tuning of their hyperparameters, to improve performance on detecting credit card fraud. The studies could benefit from evaluating algorithmic performance after different forms of dataset balancing techniques – while this study made use of undersampling the dataset using a random under-sampler, different forms of resampling and resampling methods such as oversampling, SMOTE, and related techniques could be investigated.

Limitations of the study include that due to the nature of the datasets having a basis of real-world confidential information, finding authentic non-simulated datasets are not easily available nor on the public domain (understandable as participants are required to grant consent on sharing their private financial information anonymously), making it difficult to perform research on variety of datasets.

Most related studies did not appear to explore a plethora of machine learning models to test their performance rather a singular selected or few model algorithms, however, few studies performed analysis on a short list of models each with varying results. No mentions of a followed framework were spoken of nor adapted in the work. This research will add to the existing body of knowledge by testing more machine learning algorithms, including lesser tested algorithms to extend the reach towards conclusions of related studies.

# Conclusion

This evaluation of machine learning algorithms on detecting card fraud is vital as the criminal phenomenon fraud in the digital realm is increasing yearly at high rates. This study has evaluated that the Random Forest machine learning model is most effective at detecting card fraud, which aligns with the results of studies performed by Karanovic, Sladojevic, Arsenovic, Varmedja, and Anderia (2019) and a study by Geetha, Vaishnavi (2019). Future research possibilities can include deep learning and neural network performance evaluating on card fraud, including hyper-parameter tuning on models to evaluate results.

# References

Abdullahi, R. and Mansor, M. 2015. Fraud Triangle Theory and Fraud Diamond Theory. Understanding the Convergent and Divergent For Future Research. International Journal of Academic Research in Accounting, Finance and Management Sciences, 5(4):38-45. [Online]. DOI: 10.6007/IJARAFMS/v5-3/1823 [Accessed 11 June 2023].

Abdullahi, R. and Mansor, M. 2015. Fraud Triangle Theory and Fraud Diamond Theory. Understanding the Convergent and Divergent For Future Research. International Journal of Academic Research in Accounting, Finance and Management Sciences, 5(4):38-45. [Online]. Available at: https://www.researchgate.net/profile/NoorhayatiMansor/publication/310755230_Fraud_Triangle_Theory_and_Fraud_Diamond_Theory_Understanding_the_Convergent_and_Divergent_For_Future_Research/links/59855b5e458515605844f69b/Fraud-Triangle-Theory-and-Fraud-Diamond-Theory-Understanding-theConvergent-and-Divergent-For-Future-Research.pdf [Accessed 11 June 2023].

Cornell Law School. 2023. Credit card fraud, n.d. [Online]. Available at: https://www.law.cornell.edu/wex/credit_card_fraud [Accessed 1 November 2023].

Fraud. 2023. Rules-Based Fraud Detection, n.d. [Online]. Available at: https://fraud.net/d/rules-based-fraud-detection/ [Accessed 1 November 2023].

Geetha, S., Vaishnavi V., N. 2019. Credit Card Fraud Detection using Machine Learning Algorithms. INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019, 165(-):631-641. [Online]. Available at: https://www.sciencedirect.com/science/article/pii/S187705092030065X [Accessed 25 May 2023].

Hewlett Packard Enterprise. 2023. Machine learning, n.d. [Online]. Available at: https://www.hpe.com/za/en/what-is/machine-learning.html [Accessed 1 November 2023].

IBM. 2023. What is artificial intelligence (AI)? n.d. [Online]. Available at: https://www.ibm.com/topics/artificial-intelligence [Accessed 1 November 2023].

Investopedia. 2023. Multicollinearity: Meaning, Examples, and FAQs, 25 February 2023. [Online]. Available at: https://www.investopedia.com/terms/m/multicollinearity.asp#:~:text=Multicollinearity%20is%20a%20statistical%20concept,in%20less%20reliable%20statistical%20inferences [Accessed 5 November 2023].

Karanovic, M., Sladojevic, S., Arsenovic M., Varmedja, D., and Anderia, A. 2019. Credit card fraud detection - machine learning methods. Faculty of Technical Sciences University of Novi Sad Novi Sad, 1(1):1-6. [Online]. DOI: 10.1109/INFOTEH.2019.8717766 [Accessed 25 May 2023].

Karanovic, M., Sladojevic, S., Arsenovic M., Varmedja, D., and Anderia, A. 2019. Credit card fraud detection - machine learning methods. Faculty of Technical Sciences University of Novi Sad Novi Sad, 1(1):1-6. [Online]. Available at: https://ieeexplore.ieee.org/abstract/document/8717766 [Accessed 25 May 2023].

Khatri, S., Arora, A., Agrawal A., P. 2020. Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison. 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp: 680-683. [Online]. Available at: https://ieeexplore.ieee.org/abstract/document/9057851 [Accessed 7 November 2023].

Khatri, S., Arora, A., Agrawal A., P. 2020. Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison. 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp: 680-683. [Online]. DOI: 10.1109/Confluence47617.2020.905785 [Accessed 25 May 2023].

Microsoft. 2023. What is machine learning model? 30 December 2021. [Online]. Available at: https://learn.microsoft.com/en-us/windows/ai/windows-ml/what-is-a-machine-learningmodel#:~:text=A%20machine%20learning%20model%20is,and%20learn%20from%20those %20data [Accessed 3 June 2023].

National Whistleblower Center. 2023. The Fraud Triangle, n.d. [Online]. Available at: https://www.whistleblowers.org/fraud-triangle/ [Accessed 11 June 2023].

PEDIAA. 2023. Difference Between Ontology and Epistemology, 4 December 2016. [Online]. Available at: https://pediaa.com/difference-between-ontology-and-epistemology/ [Accessed 3 November 2023].

Sujana, E., Yasa, P. E., and Wahyuni, M. A. 2018. Testing of Fraud Diamond Theory Based on Local Wisdom on Fraud Behavior. Advances in Economics, Business and Management Research, 69(1):12-15. [Online]. DOI: 10.2991/teams-18.2019.3 [Accessed 11 June 2023].

Sujana, E., Yasa, P. E., and Wahyuni, M. A. 2018. Testing of Fraud Diamond Theory Based on Local Wisdom on Fraud Behavior. Advances in Economics, Business and Management Research, 69(1):12-15. [Online]. Available at: https://www.atlantispress.com/proceedings/teams-18/55911610 [Accessed 11 June 2023].

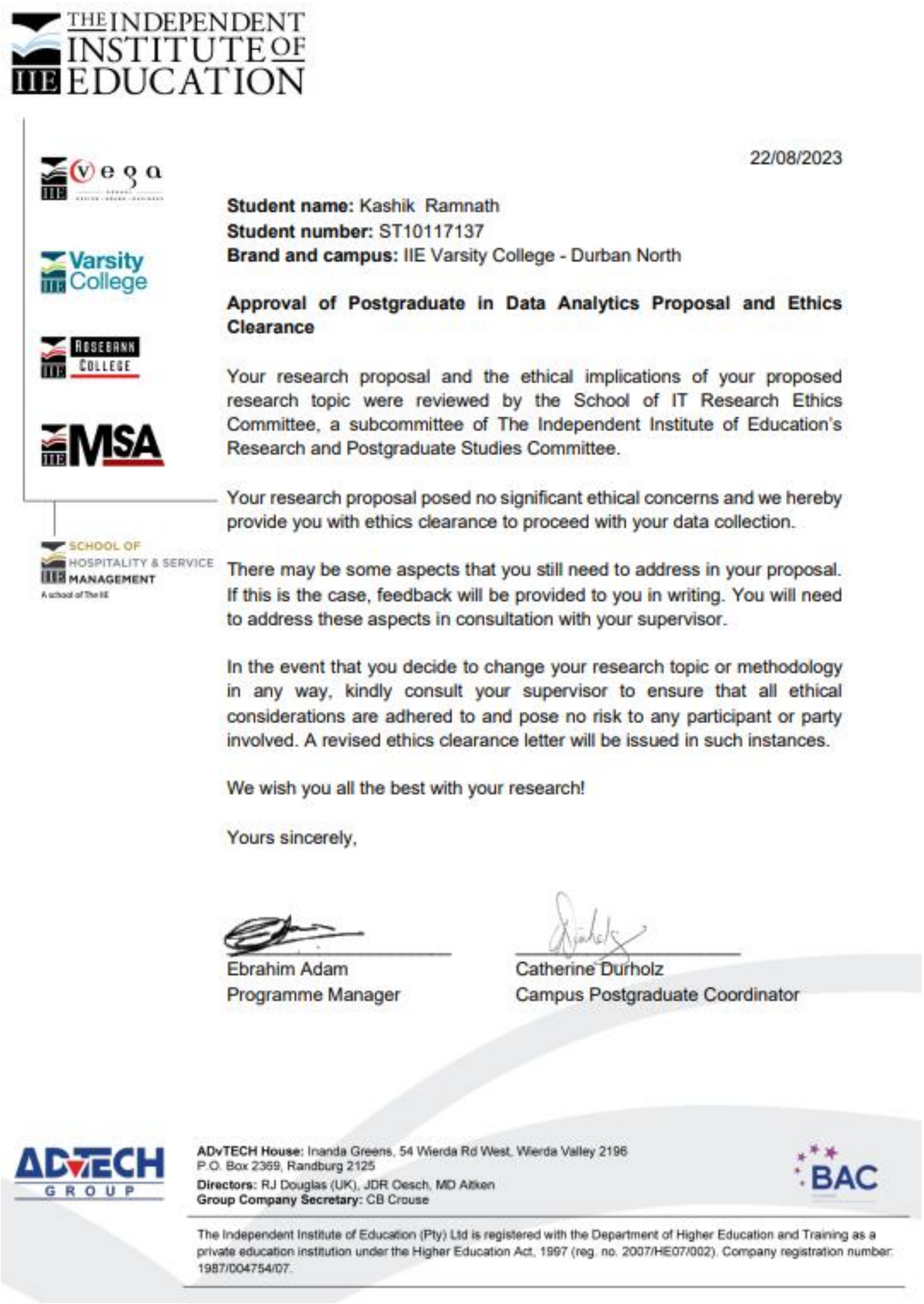
Tableau. 2023. What is the history of artificial intelligence (AI)? n.d. [Online] Available at: https://www.tableau.com/data-insights/ai/history#:~:text=Birth%20of%20AI%3A%201950-%201956&text=into%20popular%20use.-%20,Dates%20of%20note%3A,ever%20learn%20the%20game%20independently [Accessed 2 November 2023].

Tech Target. 2023. What is an algorithm? n.d. [Online]. Available at: https://www.techtarget.com/whatis/definition/algorithm [Accessed 3 June 2023].
Technopedia. 2023. Data Set, 11 May 2022. [Online]. Available at: https://www.techopedia.com/definition/3348/data-set-ibm-mainframe [Accessed 3 November 2023].

Thennakoon, A., Bhagyani, C., Premadasa, A., Mihiranga, S., and Kuruwitaarachchi, N. 2019. Real-time Credit Card Fraud Detection Using Machine Learning. 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 488-493. [Online]. DOI: 10.1109/CONFLUENCE.2019.8776942 [Accessed 25 May 2023].

Thennakoon, A., Bhagyani, C., Premadasa, A., Mihiranga, S., and Kuruwitaarachchi, N. 2019. Real-time Credit Card Fraud Detection Using Machine Learning. 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 488-493. [Online]. Available at: https://ieeexplore.ieee.org/abstract/document/8776942 [Accessed 25 May 2023].

# Appendix



THE INDEPENDENT INSTITUTE OF EDUCATION

IIE Vega

IIE Varsity College

IIE Rosebank College

IIE MSA

IIE School of Hospitality & Service Management
A school of The IIE

22/08/2023

**Student name:** Kashik Ramnath
**Student number:** ST10117137
**Brand and campus:** IIE Varsity College - Durban North

**Approval of Postgraduate in Data Analytics Proposal and Ethics Clearance**

Your research proposal and the ethical implications of your proposed research topic were reviewed by the School of IT Research Ethics Committee, a subcommittee of The Independent Institute of Education's Research and Postgraduate Studies Committee.

Your research proposal posed no significant ethical concerns and we hereby provide you with ethics clearance to proceed with your data collection.

There may be some aspects that you still need to address in your proposal. If this is the case, feedback will be provided to you in writing. You will need to address these aspects in consultation with your supervisor.

In the event that you decide to change your research topic or methodology in any way, kindly consult your supervisor to ensure that all ethical considerations are adhered to and pose no risk to any participant or party involved. A revised ethics clearance letter will be issued in such instances.

We wish you all the best with your research!

Yours sincerely,

Ebrahim Adam
Programme Manager

Catherine Durholz
Campus Postgraduate Coordinator

ADvTECH GROUP

**ADvTECH House:** Inanda Greens, 54 Wierda Rd West, Wierda Valley 2196
P.O. Box 2369, Randburg 2125
**Directors:** RJ Douglas (UK), JDR Oesch, MD Aitken
**Group Company Secretary:** CB Crouse

BAC

The Independent Institute of Education (Pty) Ltd is registered with the Department of Higher Education and Training as a private education institution under the Higher Education Act, 1997 (reg. no. 2007/HE07/002). Company registration number: 1987/004754/07.

**Word Count: 5721**