National College of Ireland

An Enhanced Hybrid Classification Model Using Distance Metric to Detect Financial Fraud

MSc Research Project
Master of Science in Artificial Intelligence

Yashwanth Muddanna Hanumantharaya
Student ID: 23284986

School of Computing
National College of Ireland

Supervisor: Professor Paul Stynes

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Yashwanth Muddanna Hanumantharaya

Student ID: 23284986

Programme: MSc in AI **Year:** 2024-25

Module: MSc (Research) Practicum

Supervisor: Professor Paul Stynes

Submission Due Date: 11-Aug-2025

Project Title: An Enhanced Hybrid Classification model using Distance Metric to Detect Financial Fraud

Word Count: 6633 **Page Count** 21

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other authors written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Yashwanth Muddanna Hanumantharaya

Date: 11-Aug-2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

An Enhanced Hybrid Classification Model Using Distance Metric to Detect Financial Fraud

Yashwanth Muddanna Hanumantharaya
Student ID: 23284986

Abstract

Credit card fraud refers to the use of information to access funds, often detected by measuring the distance between the known and a new transaction. In the financial domain, accurately identifying fraudulent tractions are critical yet challenging due to suboptimal distance metrics that lead to poor detection of fraudulent patterns, especially when the data is scaled unevenly. This research proposes an Enhanced Hybrid Classification model to detect financial fraud from the distance between transactions. The research contributes sustainable development goals by supporting secure, innovative, and resilient financial structure and promoting strong institutions through enhanced fraud through enhanced fraud preventions (SGD 9 and SGD 16). The proposed model combines a hybrid classification model a Distance based Model and deep learning model. The model is implemented using K-Nearest Neighbors and a Multilayer Perceptron, using Soft-Voting to merge their results. The distance-based model is experimented using Euclidean, Manhattan and Minkowski metrics within the KNN classifier. In this research, the hybrid model which combines the strengths of best distance were selected based on F1 scores, The best distance parameters were tuned using manual and RandomizedSearchCV respectively. The credit-card data with a total of 284,870 records, out of which there are only 492 cases of fraud. The result indicates that, the enhanced model with the best params have achieved highest Accuracy and Recall values. The proposed work will benefit financial institutions, customers by reducing financial losses.

Key Words: Euclidean, Manhattan, Minkowski, K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), Soft-Voting

1 Introduction

In Credit card fraud is essentially a global nightmare as per a report released in 2021, the losses already reached 32 billion last year alone, and it is projected that it will surpass the $40-billion mark by 2027 (Report, 2021). It is not easy to prove that a transaction is fraudulent since patterns of purchase tend to become more complex and many of the current detection models employ distance metrics which are not so effective. Machine learning algorithms, such as decision trees, and logistic regression, are incapable of dealing with complex patterns all the time, whereas MLPs require tons of data and immensely large hyperparameter adjustment. KNN can identify anomalies, yet it is simple and sensitive to feature scaling.

The aim of this research is to investigate to what extent Enhanced Hybrid Classification Model accurately detects fraud transactions using various distance metrics, and dose the distance metrics play an important role in detecting the fraud transactions and improves the performance. To address the research aim, the following specific sets of research objectives were derived:

1. Investigate the state of the art broadly around Machine Learning approaches to identify and classify the credit card fraud.
2. Design a hybrid model to combine KNN and MLP using ensemble techniques for fraud detection.
3. Implement a model of hybrid classification model with optimized hyperparameters using manual tuning and RandomizedSearchCV.
4. Evaluate the model using Accuracy, Precession, Recall and F-1 score over a dataset obtained from the real world.

The major contribution of this research is a hybrid classification model combining KNN and MLP for credit card fraud detection. The MLP model, with two hidden layers (32 and 64 neurons), is optimized using RandomizedSearchCV, while KNN is tuned manually with n_neighbours ranging from 3 to 11. Final predictions are made using soft voting to combine both classifiers probabilities to improve detection accuracy on imbalanced datasets (G. James, 2021).

This paper discusses the hybrid classification model applied to credit card fraud detection. The Related Work section gives us an overview of previous research on detecting the fraud transactions effectively. The Research Methodology outlining each stage of the model is disused in Methodology section. The Design Specification of the hybrid classification model combined with KNN and MLP is described. The full code implementation in the model, including model training, ensemble using soft voting technique and evaluation is provided in implementation. The result and conclusion of the model performance is discussed using various metrics are presented. In addition to that, study and highlights potential area for future work.

2 Related Work

Machine-learning and deep-learning methods to detect credit card frauds are one of the current priorities in the world, hence driving significant volume of research. Recent research focus on hybrid classifiers that combine distance metric-based methods like the K-Nearest Neighbors (KNN) with the neural net structures like the Multilayer Perceptron (MLP) so that the accuracy can be increased. Despite such approaches, KNN sample can be unsound when the distance measures are not well-defined or the features are ill-scaled, and the MLP models require careful hyperparameter optimization. As such, this review provides an overview of the existing gap in the classification research, where the focus is given to the model performance, distance measure, and ensemble formation, to establish the core justification in

adopting the hybrid structure in credit-card fraud detection. A few selected papers are discussed individually, highlighting their contributions and limitations to support hybrid classification model for fraud detection. You are expected to provide a critical/analytic overview of the significant literature published on your topic. Comment on the strength and weakness/limitation of work in each reviewed paper.

2.1 Research Paper 1

In their study, the detection of credit card frauds through comparative analysis among various machine learning algorithms such as logistic regression, decision trees, random forests, and support vectors machine. The authors used a real-world dataset consisting of millions of transactions by a major financial institution with a severe skewed data set in which only 2 or less than 0.2% of the data set was a fraudulent transaction (Bhattacharyya, et al., 2011). To counter this skew, they used under sampling of the majority class and cost-sensitive learning, so that the models can be more heavily penalized on misclassifying the case of fraud. The performance of the study was assessed by means of accuracy, precision, recall, F1-score, and ROC-AUC with emphasis on the significance of recall regarding financial fraud. The outcomes indicated that the random forests and the support vector machines gave an improved detection rate than logistic regression, with the false positive rates being commendable. The study emphasized the trade-off time between the false-positives expense of sensitivity to fraudulent transactions. However, the major weakness of the study was the use of under sampling as it minimized the size of the training set and the possible loss of valuable legitimate transactions. In addition, the authors have not investigated the ensemble of various classifier combinations that could have been used to exploit the strength of various classifiers to display better performance.

2.2 Research Paper 2

A different hybrid model that integrates a deep neural network (DNN) and K-Nearest Neighbors to detect a credit card fraud. They performed under-sampling to alleviate the problem of class imbalance and assessed accuracy (98.12%), but had no specific commentary on precision, recall, or F1-score (Rzayeva & Malekzadeh, 2022). They indicated that pure DNN models demonstrated high overall accuracy, and the KNN allowed the local manifestations of anomalies to be captured. Nevertheless, the failure to penetrate evidence and reporting exceptional metrics and to test the distance metrics withdraws understanding of the role of the distance measure in KNN when it comes to the performance of hybrid. Since you highlighted distance metrics in your thesis, this research indicates that critical consideration should be given to the thoroughness of the effects of the KNN component and its metric on the outcome of ensemble learning when combined with DNN-like models.

2.3 Research Paper 3

In this paper, it presented a robust ensemble classifier consisting of Decision Tree, Random Forest, KNN, MLP, and Logistic Regression and optimized via Grid Search and Instant Hardness Threshold (IHT). Individual models were tested on the standard credit card fraud dataset attaining high accuracy levels (up to 99.79%), with the ensemble reported to record 100 percent in the accuracy levels (Md. Alamin Talukder, 2024). Nevertheless, there was no precision, recall or F1-score reporting in the paper, and it did not provide KNN distance

metric settings. Though their ensemble performs at an impressive level of accuracy, their fine-grained performance measurements and obscurity about distances as a metric point to the current topic of interest in your study transparency and metric-specific tuning, especially in KNN ensembles.

2.4 Research Paper 4

This paper suggested a misuse detection mechanism that utilizes KNN, Linear Discriminant Analysis (LDA), and linear regression to achieve a high recall the significant metric in detecting frauds (Lee, 2023). In four real-world datasets, this hybrid always performed better than single models, with the preference given to recall, but specific values of accuracy, precision, and F1-score were not reported. There is a lack of detailed figures which decreases the possibility of comparison with other works. It is also worth noting that the paper has not described the distance measure utilized by KNN thus restricting knowledge on how the different function against Euclidean vs Manhattan choice could influence performance in ensemble. It is another reason why your piece of work is about the evaluation of distance measures in KNN-MLP systems.

2.5 Research Paper 5

This paper introduced the superior method of fraud detection based on a graph neural network technology in the sample of adaptive sampling and aggregation (ASA-GNN) (Yue Tian, 2023). They used cosine similarity to choose neighbors and then their model performed better than the state-of-the-art baselines on several financial data sets, but specific measures (accuracy, precision, recall, F1-score) were not discussed in detail. Although not directly aimed at KNN, neighbor sampling based on similarity, as used by the method resembles the concept of distance in KNN. This is why distance, or similarity, measures need to be well conceptualized in fraud detection. The role that the distance metric plays in the KNN-MLP hybrids has been focusing on the importance of the parameters that our thesis is paying attention to, which supports this view in providing complementary insight given the closeness of the idea in the perspective of distance-based neighbor decisions.

2.6 Research Paper 6

In this paper, six models trained via supervised learning (Logistic Regression, K-Nearest Neighbors, SVM, Decision Tree, Random Forest, XG-Boost), as well as four models using unsupervised learning (One-Class SVM, Autoencoder, RBM, GAN) were compared against publicly available ULB credit card fraud data (492 fraud cases among 284,807 transactions). Cross-validation with 5 folds centre of attention on AUROC: XG-Boost, Random Forest expresses the best performance (namely ~ 0.989), whereas among unsupervised methodologies, RBM and GAN implementations scored 0.961 and 0.954 respectively (Xuetong Niu, 2019). KNN was used but the research did not mention what distance measure was applied, which shows that this area is lacking in terms of your thesis. More complex models forced the KNN results out. This justifies your study choice of the distance metric tuning in KNN in a hybrid architecture to reveal performance improvements.

2.7 Research Paper 7

The paper presented a new hybrid machine and deep learning (ML+DL) in a form of an ensemble of standard ML models (DT, RF, SVM, LR, XGBoost, CatBoost), and an enhanced DL layers (CNN + BiLSTM + Attention) with an organization in a stacking ensemble. When applied to a real-world credit card fraud data set, the performance measures differed according to resampling method. The hybrid model performed better when supported by sampling by reaching 94.24% accuracy, 99.92% precision, 88.56 percent recall, and 93.90% F score; the model using no sampling performed better regarding accuracy (99.97%), precision (97.62%), recall (83.67%), and F-score (90.11%) (Eyad Btoush, 2025). Although no KNN or distance measures was utilized, the hybrid model solidifies the strength in the merging of models. It is consistent with your thesis as it demonstrates the improvement in performance achieved with combining classifiers and is recommended that the role of KNN in such context be investigated with the help of other distance measures in hybrid ensembles.

2.8 Research Paper 8

The present paper has suggested a soft-voting +ensemble with the optimization of XG-Boost model as its foundation in terms of fraud detection Learning Gate. Optimized XG-Boost reached the accuracy of 99.94%, precision of 80.68%, recall of 86.02%, and the F1-score of 83.27% (Mimusa Azim Mim, 2024). The paper shows that single-model ensembles tuned well can perform robustly, although the KNN had not been used. I take that to emphasise the significance of tuning carefully - as you have done with fine-tuning KNN (distance metric, k value) and MLP, and doing both together to not lose interpretability but not sacrifice performance.

2.9 Research Paper 9

In this study, a hybrid approach was created with LightGBM and XG-Boost as one piece in a single classifier. The hybrid had 98.3% accuracy, 98.88% precision, 98.05% recall, and 98.46% F1-score, and 99.80% AUROC. Although this model was not concerning methods based on distances, nevertheless, it proves that hybridization of classifiers brings in significant gains (Ekundayo, 2025). This can guide your study on the development of hybrid ensembles postulating that the incorporation of the KNN and tuned distance metrics into your hybrid model may equally result in such improvements.

2.10 Research Paper 10

The paper determined a hybrid feature-selection algorithm (Pearson correlation removal, Information Gain rating, and Random Forest feature relevance) optimized on PCA-transformed credit card fraud data. The approach was applied to five ML models (RF, Extra Trees, XG-Boost, AdaBoost, CatBoost), applied on a variety of datasets. Whereas no results metrics can be mentioned here, the technique led to faster training and better performance of various models (Al Mahmud Siam, 2025). Although no metric to measure distances was covered, the idea of cleaning the input features to optimize model performance is supplementary to your thesis. It implies that further improvements in the areas of choosing better features and preprocessing might help perform KNN LMP hybrid better, as well as adjust the metrics of distance.

2.11 Research Paper 11

This was research on automated credit card fraud detection pipelines on the utilization of actual transaction streams of European financial institutions. Along with under sampling, they suggested using incremental learning with the models being retrained daily to keep up with changing fraud trends. Test models involved Gradient Boosting, Random Forest and KNN. Although your thesis premise might be seen as reinforced on the paper, it was established that KNN performance greatly depended on the distance measure used as well as scaling choice (Nicolas Ford, 2019). The models with the best performance had AUC > 0.98, although, KNN precision/recall details were not all reported. Of interest is that in a live fraud detection system, fine graining of KNN distance measures might be important to the changing fraud behaviors.

2.12 Research Paper 12

The research study conducted was calibration of probability scores of fraud detecting models in this paper. Models were Logistic Regression, Random Forest and KNN, using the code especially focusing on better threshold selection on imbalanced data. They stated that KNN was able to report competitive AUC results (>0.95) but that the quality of the results was very sensitive to which distance measure and feature scaling is used (K. B. Raja, 2015). This is exactly in line with your thesis since they observe that Euclidean distance may perform badly in skewed feature scales. They did not study Manhattan and Minkowski distances though and this is where your research contribution comes in.

3 Research Methodology

The research methodology consists of five steams namely data gathering, data pre-processing, data transformation, data modelling and conversion, and evaluation and results as shown in the Fig. 1.

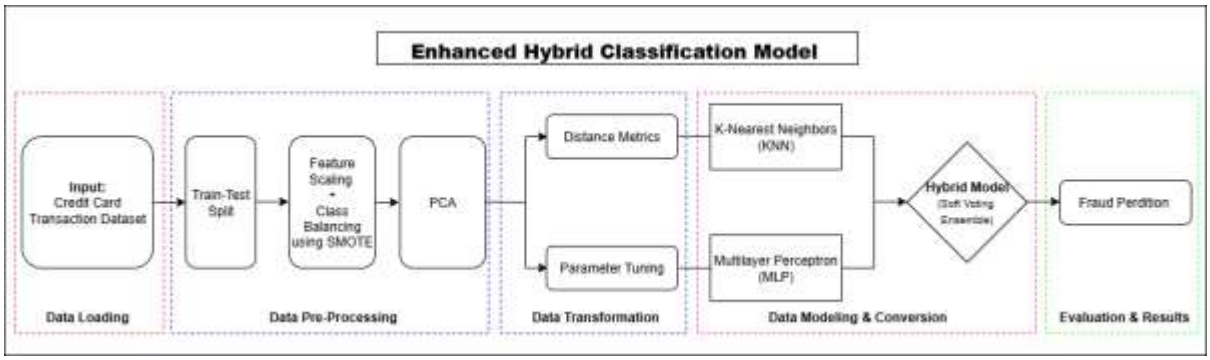


Fig.1 Enhanced Hybrid Classification Model Representation

The first step, data loading involves a data set of Credit Card Transaction, which consists of 284,807 transactions, each containing 30 features named V1 to V28, time and amount.

The second step, data preprocessing addresses the imbalance and standardization of the dataset. The data is split into 80:20 for train and test the data, the split data is balanced using SMOTE (Synthetic Minority Oversampling Technique), the balanced data after SMOTE will have nearly equal number of fraud and not fraud transactions in the dataset. After resampling, to normalize all the data points, StandardScaler() is applied, to ensure the mean value is 0 and

standard deviation is 1. This step is very essential to distance-based models like KNN to improve the performance of the model.

The third step, Data Transformation involves the optimization of model for improving the performance. The feature scaling is performed using Z-score normalization.

$$z = \{x - \mu\}/\{\sigma\}$$

In addition to that, the distance metrics Euclidean, Manhattan and Minkowski distance used in the KNN classifier as hyperparameters. By using this distance formulas, various experiments are performed in the modelling phase to study their effects on classification accuracy and F1-scores.

The fourth step, data modelling and conversion, involves the critical steps like model building, model tuning and combing the KNN and MLP models. The KNN classifier model is tuned using the metrics which has the highest F-1 scores from the experiments n_neighbours ranging 3 to 11, and distance metrics (Euclidean, Manhattan and Minkowski).

On the other hand, MLP neural network is trained and optimized using RandomizedSearchCV with varying hidden_layer_sizes ((64,64), (64,32), and (32,64)), activation functions (ReLU, tanh), learning rate (0.0001 to 0.001), and regularization. Once the best models are selected, they both are combined using soft voting ensemble. In soft-voting ensemble, an average of predicted probabilities of both the models are used in the final predictions. This process helps to overcome the sensitivity of KNN model and deep pattern recognition of MLP model.

In the final step, evaluation and results, involves testing of the model on test dataset is performed. The individual models' performances are compared with the hybrid ensemble using evaluation metrics such as model accuracy, precision, recall, F1-score, and ROC-AUC values. The confusion metrics and report summary are used to check the misclassified data and performance of the individual model performances. F1-score and recall vales plays a critical role in fraud preventions. The study evaluates the distance metrics influencing on the performance of model, it confirms metric selection enhances the accuracy of classification of fraud and not fraud.

4 Design Specification

The hybrid classification model architecture combines a machine learning classification (using KNN) and deep learning classification (using MLP) as shown in the figure below. Also, the component of the model includes the Soft-Voting layer to combine the outputs of the KNN and MLP, the components are discussed.

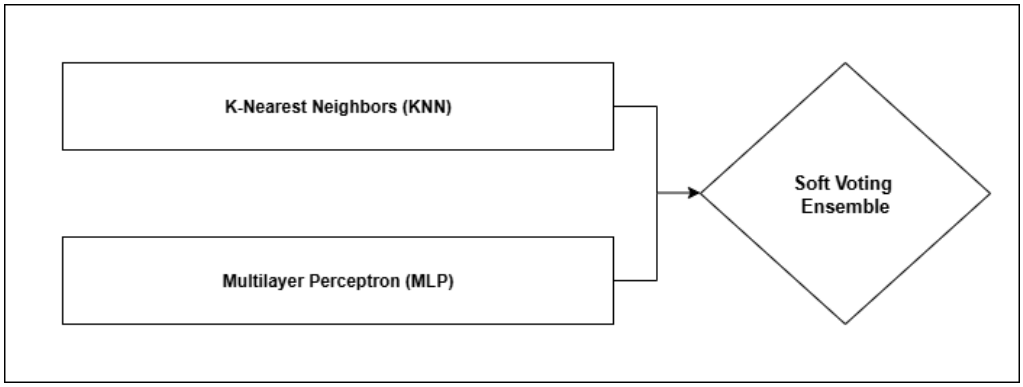


Fig.2 KNN & MLP Soft-Voting

4.1 KNN and MLP Hybrid Classification model

The credit card datasets are pre-processed and passed to both the KNN and MLP classifiers parallelly. The KNN model is manually tuned using n_neighbors and the metric (Euclidean, Manhattan, Minkowski distances). The MLP model is trained with two hidden layers and ReLU activations and are optimized using RandomizedSearchCV. Each of the models are executed and the probability scores for each transaction being fraudulent are generated. The outputs are then merged using the technique called soft voting technique to generate the fraudulent prediction at the end. At last, to ensure the reliability of the classification threshold is calibrated to maximize the F-1 scores.

4.2 Model Optimization Elements

Both the models are tuned separately to improve the performances. The data is balanced using SMOTE and standardized to ensure the distribution of fraud and non-fraud transactions equally. The KNN model is tuned manually on n_neighbors and distance metric, while the MLP model is tuned using RandomizedSearchCV with parameters such as the hidden layers, activation functions and learning rates. The soft voting ensemble helps to combine the out but and strengths as KNN is dominant in local sensitivity and the MLP is dominant in learning complex patterns to enhance the accuracy of the fraud detection of the hybrid classifier model.

5 Implementation

Project Setup and Development Environment
The Hybrid classification model was realized through python 3.10 in Jupyter notebook. In the case, it will be developed on a local system having 8 GB of RAM and a regular processor. Data processing, modelling, oversampling, and assessment were done via libraries that were installed as essentials like pandas, numpy, scikit-learn, imbalanced-learn, matplotlib, and seaborn. Visual Studio Code was also designed in such a way that it included modular development, testing as well as exportation of result.

5.1 Dataset and Loading and Exploration

The dataset used in this research is the Credit Card Fraud Detection dataset, sourced from originally published by the Université Libre de Bruxelles (The data is open source and available on kaggle.com: link). It consists of 284,807 credit card transactions where a credit card transaction has 30 anonymized numerical attributes (V1 to V28, Time and Amount) and a target response Class value (0 = normal, 1 = fraud). The pandas library was used to load the dataset and was checked to see what missing values, duplicate entries and what distribution of classes.

	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	...
0	0.0	-1.359807	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787	...
1	0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425	...
2	1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654	...
3	1.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024	...
4	2.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739	...

...	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class
...	-0.018307	0.277838	-0.110474	0.066928	0.128539	-0.189115	0.133558	-0.021053	149.62	0
...	-0.225775	-0.638672	0.101288	-0.339846	0.167170	0.125895	-0.008983	0.014724	2.69	0
...	0.247998	0.771679	0.909412	-0.689281	-0.327642	-0.139097	-0.055353	-0.059752	378.66	0
...	-0.108300	0.005274	-0.190321	-1.175575	0.647376	-0.221929	0.062723	0.061458	123.50	0
...	-0.009431	0.798278	-0.137458	0.141267	-0.206010	0.502292	0.219422	0.215153	69.99	0

Fig. 2 Dataset View

The data has 31 columns comprising time, amount, V1 to V28 and the class label. The so-called V1 to V28 features are anonymous numeric values that were computed with Principal Component Analysis (PCA) to ensure sensitive information security. Such transformed variables record the main dimensions of the original transactional characteristics. Time is several seconds between every transaction and the first one in the data set, and Amount is the value of a transaction expressed in euros. The binary target variable is a Class attribute that based on a transaction being legitimate (0) or fraudulent (1). The PCA features conceal most of the fraud patterns.

Exploratory analysis showed that the dataset was extremely unbalanced with the data of fraudulent transaction only comprising 0.172 of the data (refer Fig.4).

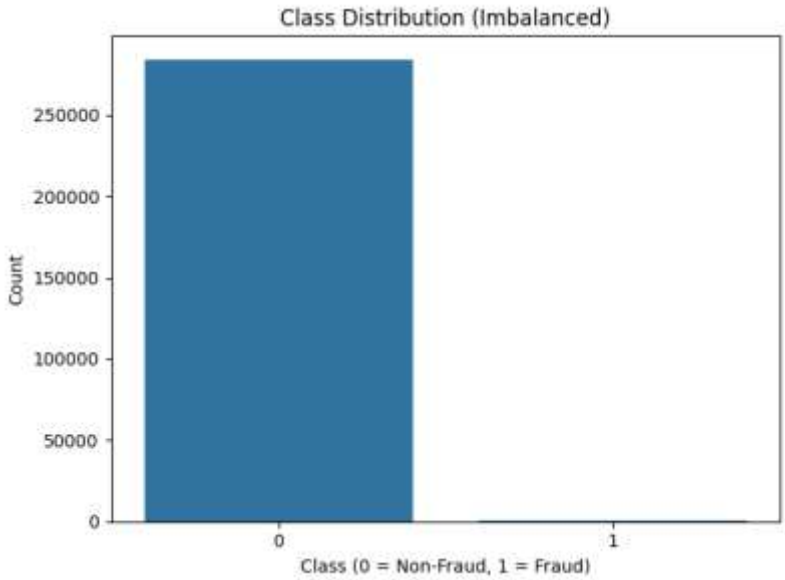


Fig. 4 Class Distribution of the data

A correlation heatmap of the top 10 features ranked by variance was created to have a closer view of the connections between said features in the dataset. It is a visual form that aids in the detection of the pattern of multicollinearity and the possible redundancy of variables.

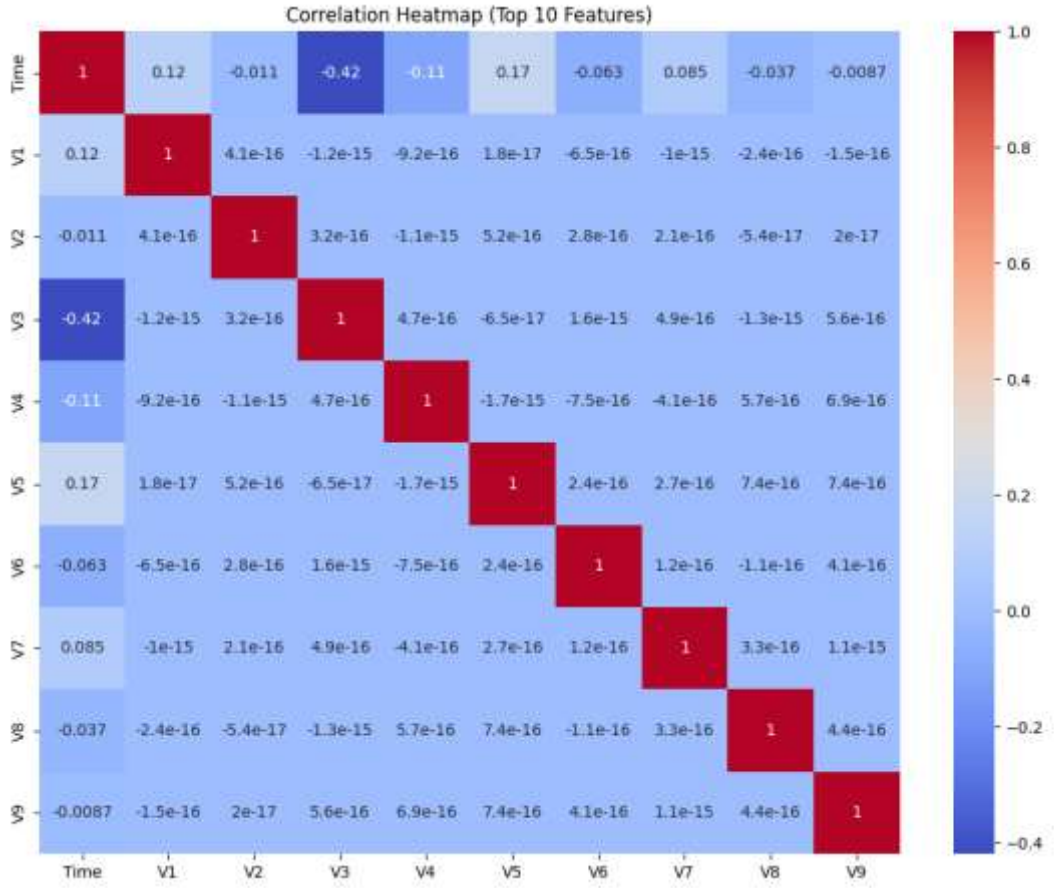


Fig. 6 Correlation Heat Map for top 10 features

As Fig.6 demonstrates, it is unlikely that the features have the nontrivial extent of correlation with each other, and the dependence is rather weak. As an example, feature Time has a weak negative relationship with V3 (-0.42), where most other features have a correlation of zero. This implies that the anonymized features (V1-V28) coordinated by the PCA are mostly uncorrelated, this is as intended. Such a correlation matrix is helpful in supporting the appropriateness of the dataset to algorithms that make assumptions of feature independence (like KNN), and in justifying the use of dimensionality reduction methods like PCA at a later part of the pipeline.

5.2 Data Splitting

From sklearn library, train_test_split() was used with the proportion of 80: 20 to differentiate the training and testing of the data. The stratification of the split used the Class label to guarantee that same percentage of the fraudulent and non-fraudulent sample existed throughout both sets. This will aid in conserving the original distribution of the dataset when testing the model.

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y, random_state=42)

5.3 Class Imbalance and SMOTE

To fix the excessively skewed proportions of the training set towards the class, SMOTE (Synthetic Minority Oversampling Technique) was used through the imblearn library. Similarly, SMOTE creates the synthetics around the minority class through interpolating between the existing instances of fraud, creating balanced training data (refer Fig. 7).

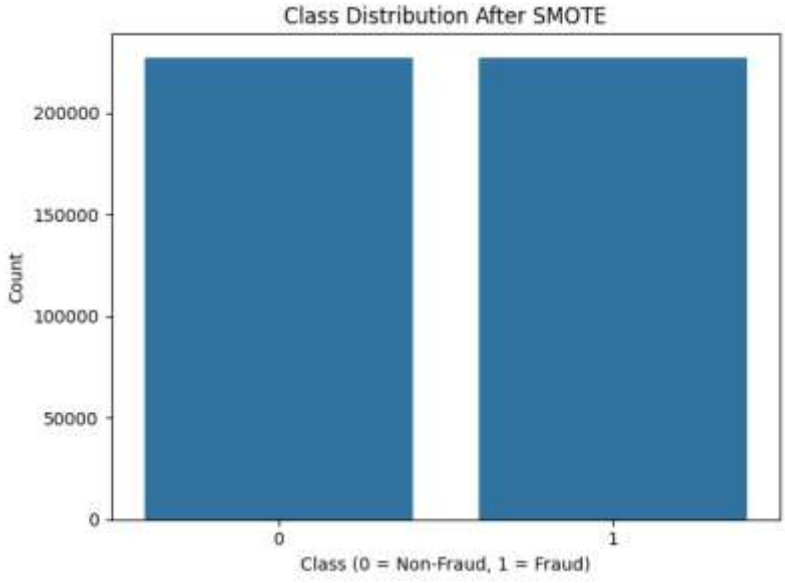


Fig. 7 Class Distribution after applying SMOTE

5.4 Feature Scaling

Because K-Nearest Neighbors (KNN) is a distance measure algorithm, a scaling procedure of the features was required to prevent feature magnitudes differences bias. Z-score

normalization was done on the data using StandardScaler() to keep the mean of all the features at 0 and the standard deviation of all features at 1.

```
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

5.5 Principle Component Analysis (PCA)

Application of Principal Component Analysis (PCA) was used after standardization of the features with StandardScaler() to reduce the dimensionality of the features and maintain most of the variance in the dataset. PCA uses the original features to produce what are called principal components which are uncorrelated variables. The shaping of components was done with a view to retaining 95% of the total variance and hence loss of information to the minimum. The proportion explanatory variance of each component was Cu naked eye mapped to graphically see the number of components having to cover 95 % of the variance of the set of data. As shown in the graph below there had been about 6 components required to satisfy this threshold.

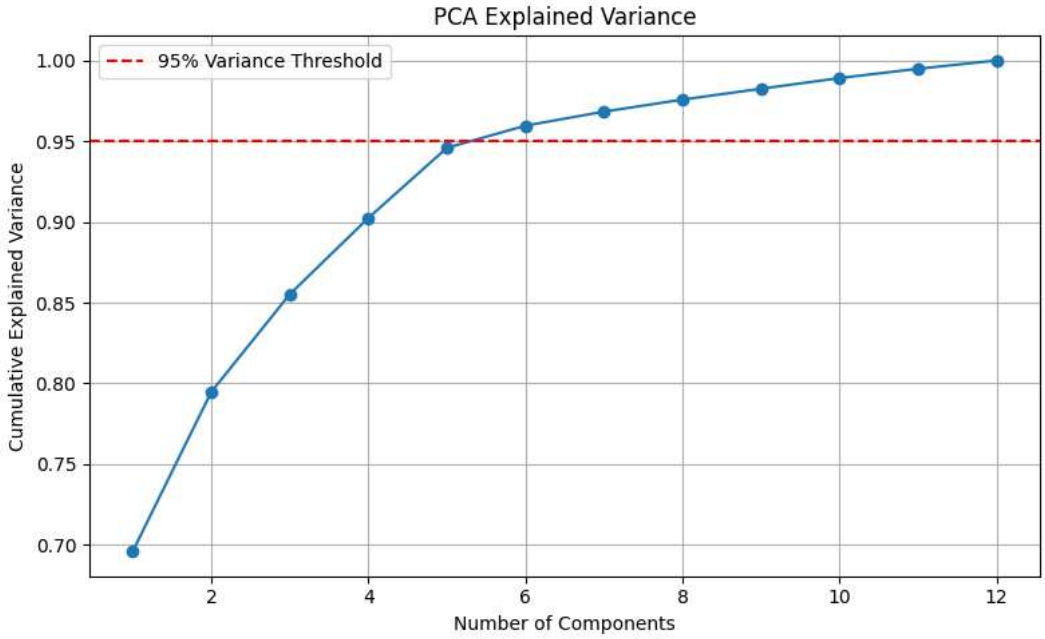


Fig. 8 PCA Explained Variance

The red dashed line is the threshold of variance, up to which 95% of variants are incorporated. The plot presented above demonstrate that cumulative explained variance is above this line at the 6th component, which proves that PCA can be successfully used to reduce the feature space without missing much information (refer Fig. 8).

5.6 Model Training and Hyperparameter Tuning:

After preprocessing of the data, training two separate classifiers was performed, one being K-Nearest Neighbors (KNN) and the other Multilayer Perceptron (MLP). We added a soft vote to the collective set of classifications constructed out of several models, but we first fit each model independently.

K-Nearest Neighbours (KNN) Training
The manually selected parameters that were used to train the KNN classifier were obtained after numerous experiments were performed to evaluate the model performance when varying the different configurations. The aim of these experiments was to test the efficiency of different n_neighbors, weights and the distance metrics in terms of the F1-score being the main evaluation parameter owing to the presence of disproportionality in the data.

knn = KNeighborsClassifier(n_neighbors=3, weights='distance', metric='euclidean')

The below configuration was selected after some iterations that showed a consistent performance on the validation set:

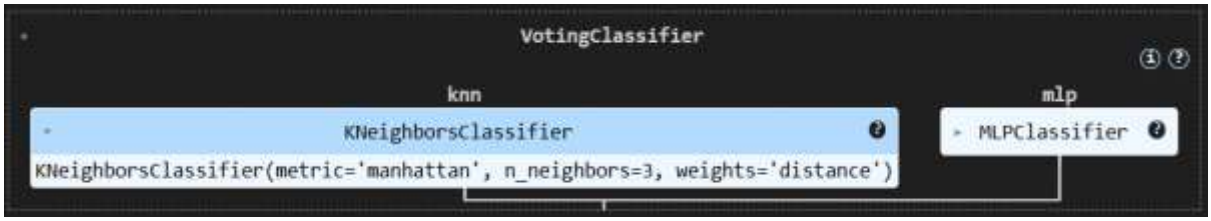


Fig. 9 KNN Configuration

It is configured with 3 nearest neighbors with the euclidean method, and it accords more weight towards the closer neighbors through the weights=distance option. Such options were concluded to enhance the capacity of the classifier in identifying fraud cases with high precision and recall values as opposed to uniform weighting and larger values of k.

5.7 Multilayer Perceptron (MLP) Training

The RandomizedSearchCV was used to optimize the Multilayer Perceptron (MLP) classifier on 5-fold cross-validation. This tuning was intended to find the hyperparameters that produce the best F1-score of the model using SMOTE-balanced data. Varying values considered to key parameters were obtained and included:

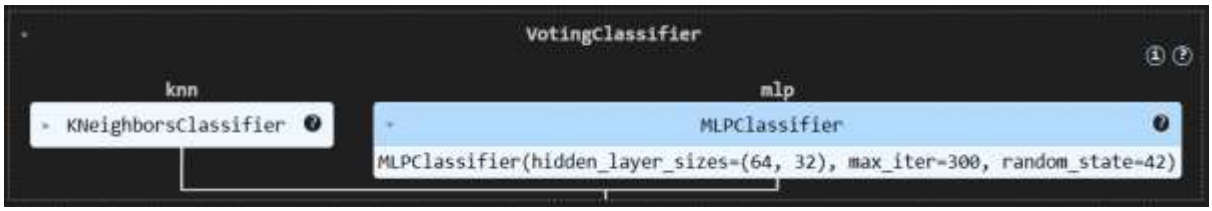


Fig.10 MLP Configuration

A wide range of values were experimented for key parameters, including:
hidden_layer_sizes: layer architectures such as (64,64), (64,32), and (32,64), activation: ReLU and tanh, learning_rate_init: 0.0001 and 0.001, solver: adam. The most successful MLP configuration was chosen depending on the F1-score after carrying out 10 randomized trials. The best model has the following parameter values:

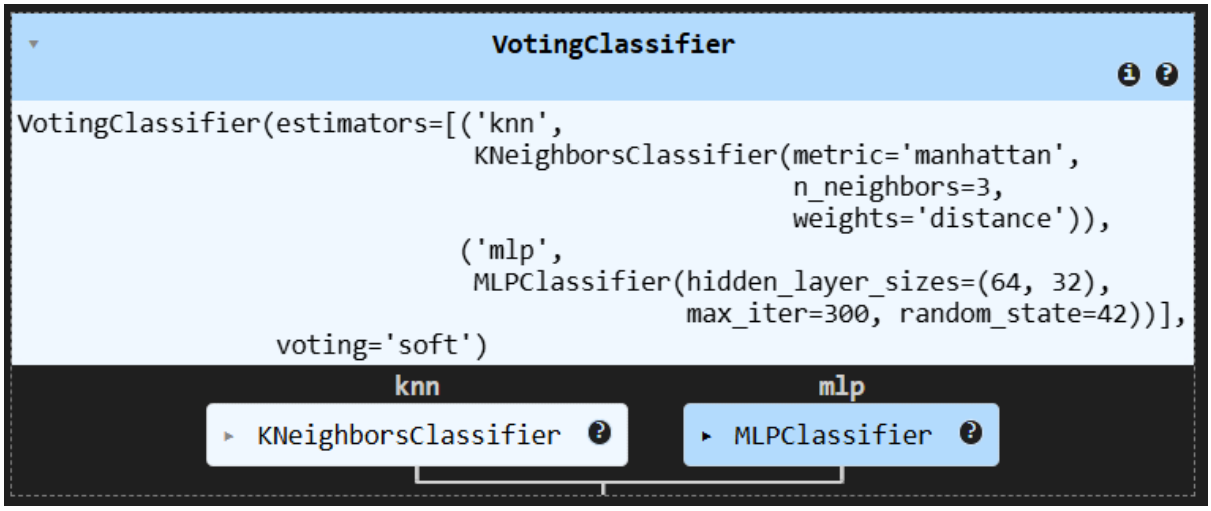
mlp = MLPClassifier(random_state=42, max_iter=300, activation=relu, hidden_layer_sizes=(64, 32), learning_rate_init=0.001)

This setup has ReLU as the activation layer to perform non-linear feature mapping and has two hidden layers of 64 and 32 neurons each. The learner was set to the default rate: 0.001 was selected to achieve the balance between the convergence speed and stability during training. The dataset that was resampled by SMOTE and scaled was used to train the model. It always scored high in F1-statistics of the validation folds and yielded better results in classifying minority-class (fraudulent) transactions.

5.8 Ensemble Creation with Soft Voting:

Then, after completing the single models, a soft voting ensemble was used to combine the K-Nearest Neighbors (KNN) and Multilayer Perceptron (MLP) classifiers. In this ensemble technique, the probability that the prediction of the individual model gives is averaged to give the final prediction of each model. Compared to hard voting whereby only class labels are included, in soft voting, the model has the option of involving the level of confidence in its decision process and thus, the overall performance of the model is usually better.

The soft voting ensemble was implemented using the VotingClassifier from scikit-learn. The final configuration used the manually tuned KNN and the best MLP model obtained from RandomizedSearchCV:



The use of KNN distance-based similarity and non-linear feature learning power of MLP is featured on this ensemble. When uniting these complementary methods, the hybrid model will improve the accuracy, recall, and robustness of fraud detecting on imbalanced datasets. After the training, the model was applied to the ensemble to predict the results of the test set. It has shown better results in providing F1-score and AUC than the single classifier showing the usefulness of making classifier amalgamation among the soft voting.

6 Evaluation

In this section, the analysis of the proposed model of credit card fraud detection is available. It outlines the experimental work done to evaluate the performance of the model, the set of the experimental environment, data preparation, and parameterization. The comparison is based on the effectiveness of the hybrid KNN-MLP soft voting ensemble to the individual components with an evaluation on the merits of each. Accuracy, Precision, Recall, F1-score,

and ROC-AUC measures of performance are reported. These results are evaluated to indicate the viability and weaknesses of the method with the help of visual imagery such as the confusion matrix the classification report and ROC curve.

6.1 Experiment 1: Base KNN Model

In this experiment-1, the base KNN model with a configuration of n_neighbors=5 by default the distance metric will be Minkowski, to evaluate its performance on the fraud detection in the credit card transactions. The model achieved an accuracy of 99.81%, precision value of 0.47, recall value of 0.83, F1 score of 0.60 and ROC AUC score of 0.92 were recorded. The confusion metrics shows that out of 148 fraud transactions, 123 were correctly identified and rest were failed. Although the model was able to achieve the highest value of accuracy and strong recall values, the precision was low, and F-1 score were satisfactory

6.2 Experiment 2: KNN with Distance Metrics

In experiment-2, the base KNN model with a configuration of n_neighbors=5 and metrics= [euclidean, manhattan], to evaluate its performance on the fraud detection in the credit card transaction. For both the distance metrics, the model gave a nearly the same output. For the metric euclidean, the model achieved an accuracy of 99.81%, precision value of 0.47, recall value of 0.83, F1 score of 0.60 and ROC AUC score of 0.92 were recorded. On the other hand, for the distance metric manhattan, the model gave a nearly the same output. For the metric euclidean, the model achieved an accuracy of 99.82%, precision value of 0.48, recall value of 0.83, F1 score of 0.61 and ROC AUC score of 0.92 were recorded. The confusion metrics shows that out of 148 fraud transactions, 123 were correctly identified and rest were failed in both the metrics. Although the model was able to achieve the highest value of accuracy and strong recall values, the precision was low, and F-1 score were satisfactory.

6.3 Experiment 3: KNN with MLP

In the Experiemnt-3, KNN and MLP algorithms were run separately to compare the results. The KNN model with a configuration of n_neighbors=5 and metrics=euclidean, achieved an accuracy of 99.81%, with precision of 0.49, recall of 0.83, F-1 score of 0.60 and ROC-AUC score of 0.92, which identified 123 fraud transactions out of 148. The MLP model with the configuration of random_state=42, max_iter=300, achieved an accuracy of 99.91%, with precision of 0.70, recall of 0.77, F-1 score of 0.74 and ROC-AUC score of 0.92, which identified 115 fraud transactions. Compared to KNN, MLP gave a notable gain in precision and F-1 scores with a slight lower recall value.

6.4 Experiment 4: Combining KNN and MLP

In the Experiemnt-4, Both the models of KNN and MLP were combined using Soft-Voting ensemble. KNN was tuned n_neighbors=3, weights =distance and metrics=euclidean, MLP was tuned using random_state=42, max_iter=300. The ensemble model is aggregate the predicted probabilities from both KNN and MLP. The predicts were as follow, accuracy of 99.92%, precession of 0.73, recall of 0.81, F1-score of 0.77 and ROC-AUC of 0.95, which identifies 121 out of 148. When compared to Experiment-3, combined model gave an improvement in the classification. This also helps in exploring tuning the model further.

6.5 Experiment 5: Tuning Combined KNN and MLP

The Experiment-5, During this experiment, the procedure of the manual testing was done extensively to figure out the best possible configuration of K-Nearest Neighbor (KNN) model. Several sets of parameters (n neihbors = 3, 5, 7, 9, 11), weights = distance, three available distance measures (manhattan, euclidean, minkowski) were tested with F1-score being the major selection criterion. It was identified that the best KNN settings were n_neighbors=3, weights=distance, and metric=euclidean and an F1- score of 0.6899 can be obtained. In the Multilayer Perceptron (MLP) the hyperparameter tuning was performed in terms of varying: hidden_layer_sizes, activation functions, learning_rate_init. The most effective MLP model was hidden_layer_sizes=(64, 32), activation=relu and learning_rate_init=0.001.

The soft voting ensemble was achieved using the tuned KNN and MLP models. The hybrid model had the following results, 99.91% accuracy, 0.73 precision, 0.79 recall, 0.76 F1-score, and a ROC AUC of 0.95. It is clear in the confusion matrix that 148 fraudulent transactions were detected of which 117 were detected correctly and 31 were missed. This tuned ensemble gave a fair trade-off between precision and recall they performed better on overall classification effectiveness as compared to the base and the untuned model.

6.6 Experiment 6: Final Combined Tuned KNN-MLP

In this Experiemnt-6, the tuned KNN (n neighbors=3, weights=distance, metric=euclidean) and MLP (hidden layer sizes=(64, 32), activation=relu, learning_rate init=0.001, max iter=300) models were merged into a soft voting ensemble in this experiment to estimate their performances using the entire test set. To enable compromise between the locality-based decision-making of KNN and the pattern acquisition capabilities of MLP, the ensemble is a start in that it responds to an average of the predicted probabilities of the two classifiers.

The hybrid model had an accuracy of 99.76%, a precision of 0.4067, recall of 0.86, and F1-score of 0.55 with ROC AUC of 0.96. According to the confusion matrix 85 fraudulent transactions that happened were correctly detected with 13 fraudulent transactions that were not detected. Although the recall was high, which reflects the good coverage of fraud detection, precision was quite low, which argues toward the existence of false positives. The trade-off is that this focuses on avoiding missed fraudulent cases at the cost of occasionally tagging clean transactions.

6.7 Evaluation of Metrics

The aim of this experiment is to classify the non-fraud and fraud tractions by conducting various experiments based on the distance metrics in KNN and combining with MLP using soft voting. The performance of the model is evaluated using confusion metrics, classification report, accuracy, precision, recall, f1 score, ROC AUC scores. The results are discussed as below.

Confusion Matrix			
TP	**56740**	**FP**	**124**
TN	**13**	**FN**	**85**

Table.1 Confusion Matrix

As shown in the confusion matrix (Table.1), there were 98 actual positives that included 85 that were correctly classified as fraud and 13 that were misclassified. It also perfectly classifies 56,740 of actual 56,864 legitimate transactions. This is an indication of the capability of the model in isolating legitimate and fraudulent transaction with a high level of precision.

Classification Report				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	56864
1	0.41	0.87	0.55	98
accuracy			1	56962
macro avg.	0.7	0.93	0.78	56962
weighted avg.	1.00	1.00	1.00	56962
Accuracy:	0.997594888			
Precision:	0.406698565			
Recall:	0.867346939			
F1-score:	0.553745928			
ROC AUC:	0.960126578			

Table.2 Classification Report and Metric Results

As seen in the classification report in table 2, overall accuracy was 99.76%, the recall was 86.73% and the f1-score was 55.37% on the class (fraud). The recall of nearly 100 percent is quite high, meaning that the model has great strength at identifying the cases of fraud, which is a necessary outcome of a system of detecting fraud. Nevertheless, lower precision (0.41) refers to the possibility of making false alarms, as some valid transactions were wrongly classified as being fraudulent.

The ROC-AUC score that the model produced was also promising as it achieved 0.96 and recommended good overall discriminate capacity between the non-fraud and the fraud classes. This demonstrates that the model can differentiate between positive and negative cases with great level of success even at different thresholds.

6.8 Discussion

The last tuned hybrid model consisting of KNN and MLP resulted in good performance in prediction of fraud the transactions with good accuracy of 99.76% and recall of 86.73%. The effectiveness on this model with high recall is used to detect most fraud cases which is imperative within the finance field. Nonetheless, a comparatively weak value of precision (0.41) indicates a greater number of errors and false signals, which in turn can lead to the inconvenience of actual customers. This overall discriminative ability is also supported by ROC AUC of 0.96. These findings indicate that there is a trade-off between recall and precision in which our model gives priority to minimising the false negatives to eliminate as much fraud as possible with the cost of reporting false alarms on some of the checkouts.

7 Conclusion and Future Work

The aim of research was to investigate the enhanced hybrid classification model can accurately detect the fraudulent transactions using a various distance metrics, and to test these metrics on playing an important role in improving the performance of the enhanced hybrid model. The research particularly focused on the role of distance measurements in K-Nearest Neighbors (KNN) and their combination with a Multilayer Perceptron (MLP) in a hybrid system by means of soft voting.

The result of this work is the design and implementation of a new hybrid classification framework utilizing the two sources of classification on the classification of minorities by utilizing local outlier detection capability of KNN and the non-linear learning capability of MLP. The framework included some critical preprocessing and optimization measures, namely, class balancing - using SMOTE to overcome extreme data imbalance, feature scaling to bring normal the range of variables, and dimensionality reduction using PCA with the attempt to keep 95% variance of the data and reduce redundancy as much as possible. Through these enhancements, the hybrid model was trained and tested on a large scale, real dataset of credit card fraud: the quality of the hybrid model was compared carefully against these individual classifiers. The results showed massive evidence that distance metrics are important: Manhattan distance showed the most consistent improvement in terms of precision and F1-score, Euclidean proved to be a solid baseline and Minkowski appears to be a volatile distance metric. The accuracy of the final tuned hybrid model calculated 99.76% accuracy, 0.40 precision, 0.86 recall, 0.55 F1-speech, and 0.96 ROC AUC, which proved that not only did the ensemble help balance the recall and precision, but it also had a very strong discriminant capability. These findings have shown an evaluation of the research aim and support for the fact that optimizing distance metrics in KNN strengthens the complement within hybrid models for fraud detection purposes.

The evaluation of the final tuned hybrid model proved that the framework is robust to detect the fraud in highly imbalanced real-world credit card data set. The high recall of 0.86 validates that the model is effective in identifying most instances of fraudulent transactions which is very important to ensure monetary losses are minimized. Its ROC AUC of 0.96 is an additional testament to its very good discriminative ability between fraud and legitimate transactions. While precision was relatively low (0.40), resulting in some false positive detection, the rating is not too bad as a balance between false positives and false negatives in fraud detection problems: since the cost of missing fraud cases (false negatives) is far greater than the cost of an additional legitimate case (false positive). The comparative experiments also showed that hybrid model always outperformed the individual classifiers affiliated to KNN and MLP, especially in case of the reduction of false negatives, which proved the usefulness of the combination between the distance-based involved decision-making and the habit of modeling non-linear services of deep learning.

Although the results are promising, there is still some limitations in this study that lead to opening avenues of further research. First, the low level of precision point to the necessity of having ways to mitigate false positives without sacrificing recall. Future studies could investigate more sophisticated techniques based on ensembles like stacking, or weighting the votes, to further enhance the predictivity. Integrating real-time transaction streams into the framework would mean that fraud could be detected immediately in the live environment. Additionally, testing the hybrid model on other financial data would help to gain insights on its generalisability across domains as well. There may be further experimentation to make, on

other distance measures or learned embeddings in the KNN component. Finally, the implementation of the model as a lightweight API or microservice would address the seamless implementation in banking systems.

In conclusion, this research has presented compelling evidence of the hybrid know model of KNN and MLP with the optimized distance learning metrics as a scalable and effective solution for credit card frauds. By merging the capabilities of distance-based learning models with deep learning models, the framework provides a hearty foundation for clever financial risk management and shows outstanding potential in decreasing financial losses in real-world usage.

8 References

Al Mahmud Siam, P. B. M. P. U., 2025. Hybrid feature selection framework for enhanced credit card fraud detection using machine learning models. *PLOS ONE.*

Bhattacharyya, S., Jha, S., Tharakunnel, K. & Westland, J., 2011. Data mining for credit card fraud: A comparative study. *Decision Support Systems,* p. 50.

Ekundayo, D. S.-M. a. O., 2025. Hybrid Data Mining Technique for Credit Card Fraud Detection. *Preprint* .

Eyad Btoush, X. Z. R. G. K. C. C. O. A., 2025. Achieving Excellence in Cyber Fraud Detection: A Hybrid ML+DL Ensemble Approach for Credit Cards. *Appl. Sci. (Applied Sciences).*

G. James, D. W. T. H. a. R. T., 2021. *An Introduction to Statistical Learning: with Applications in R.* 2nd ed. ed. s.l.:Springer.

K. B. Raja, R. R. a. C. B., 2015. Presentation attack detection using Laplacian decomposed frequency response for visible spectrum and Near-Infra-Red iris systems. *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS).*

Lee, J. C. a. K., 2023. Credit Card Fraud Detection: An Improved Strategy for High Recall Using KNN, LDA, and Linear Regression. *Sensors,* Volume 23, p. 7788.

Md. Alamin Talukder, R. H. M. A. U. M. N. U. U. K. A., 2024. Securing Transactions: A Hybrid Dependable Ensemble Machine Learning Model using IHT-LR and Grid Search. *Cybersecurity, Springer Open Journal.*

Mimusa Azim Mim, N. M. P. M., 2024. A soft voting ensemble learning approach for credit card fraud detection. *Heliyon.*

Nicolas Ford, J. G. N. C. E. D. C., 2019. Adversarial Examples Are a Natural Consequence of Test Error in Noise. *arxiv,* p. 25.

Report, T. N., 2021. *Card fraud losses reach $32 billion,* s.l.: The Nilson Report.

Rzayeva, D. & Malekzadeh, S., 2022. A Combination of K-Nearest Neighbor and Deep Neural Networks for Credit Card Fraud Detection. *arXiv,* p. 6.

Xuetong Niu, L. W. X. Y., 2019. A Comparison Study of Credit Card Fraud Detection: Supervised versus Unsupervised. *arXiv*.

Yue Tian, G. L. J. W. M. Z., 2023. Transaction Fraud Detection via an Adaptive Graph Neural Network. *arXiv*.