USING FIVE MACHINE LEARNING MODELS WITH RESAMPLING TECHNIQUES

CREDIT CARD FRAUD DETECTION

Data Source: Machine Learning Group - ULB, Andrea. (2017). Credit Card Fraud Detection, Version 3.
Available at: https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud.

About Dataset

The objective of this classification task is to identify fraudulent credit card transactions to prevent customers from being charged for unauthorised purchases.

Time of Dataset

2013

The dataset consists of credit card transactions conducted by European cardholders in September 2013 over a period of two days.

Transactions

284k+

The dataset exhibits significant class imbalance, with 492 fraudulent transactions out of 284,807, accounting for 0.173% of all transactions.

Features

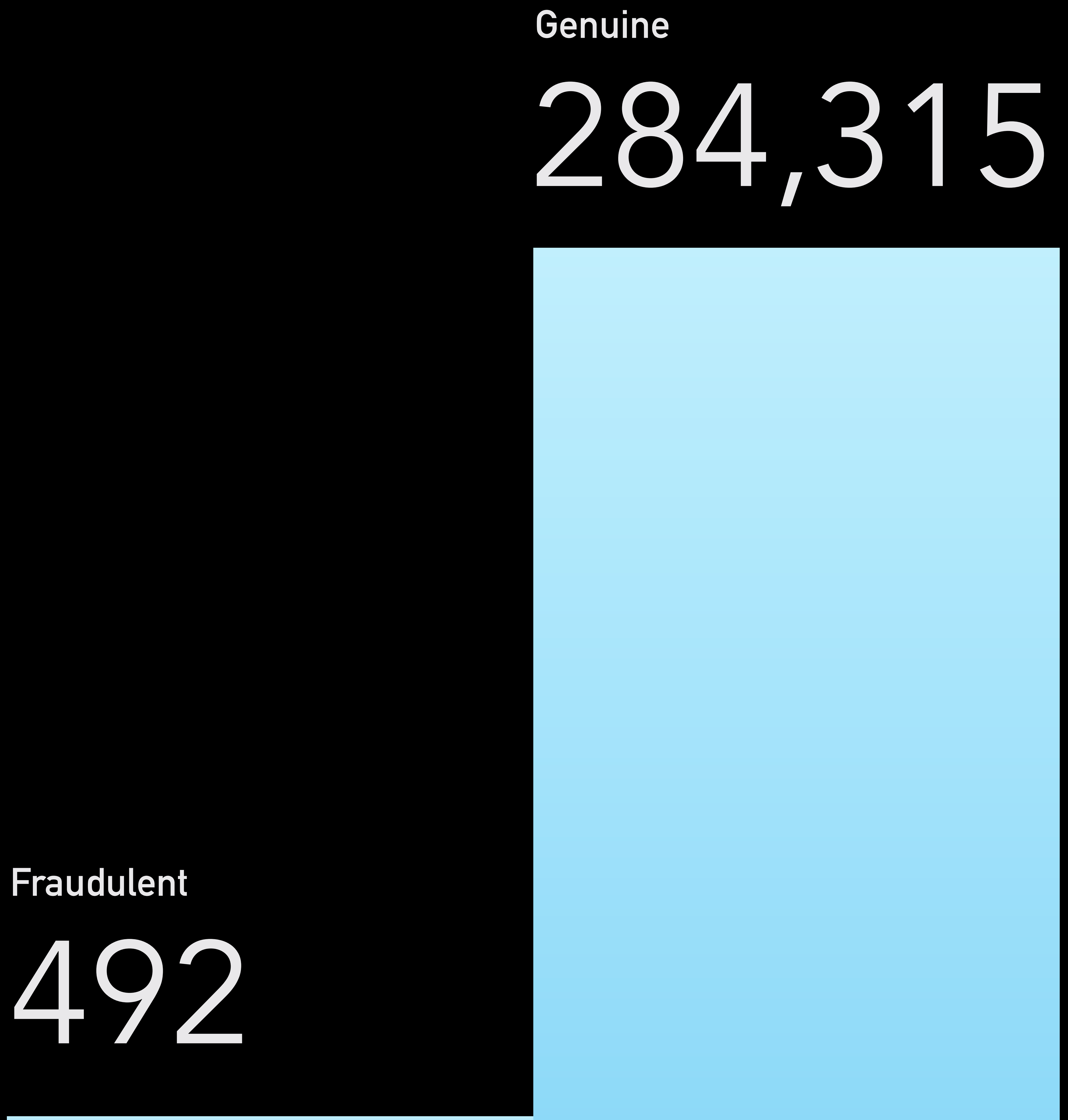
31

It contains numerical input variables resulting from a PCA transformation. Due to confidentiality concerns, the additional background information is not provided.

DATASET CHARACTERISTIC

Imbalanced Dataset

The dataset comprises 284,807 credit card transactions. Within this dataset, the fraudulent transactions constitute 0.173% of all transactions, indicating a significant class imbalance.

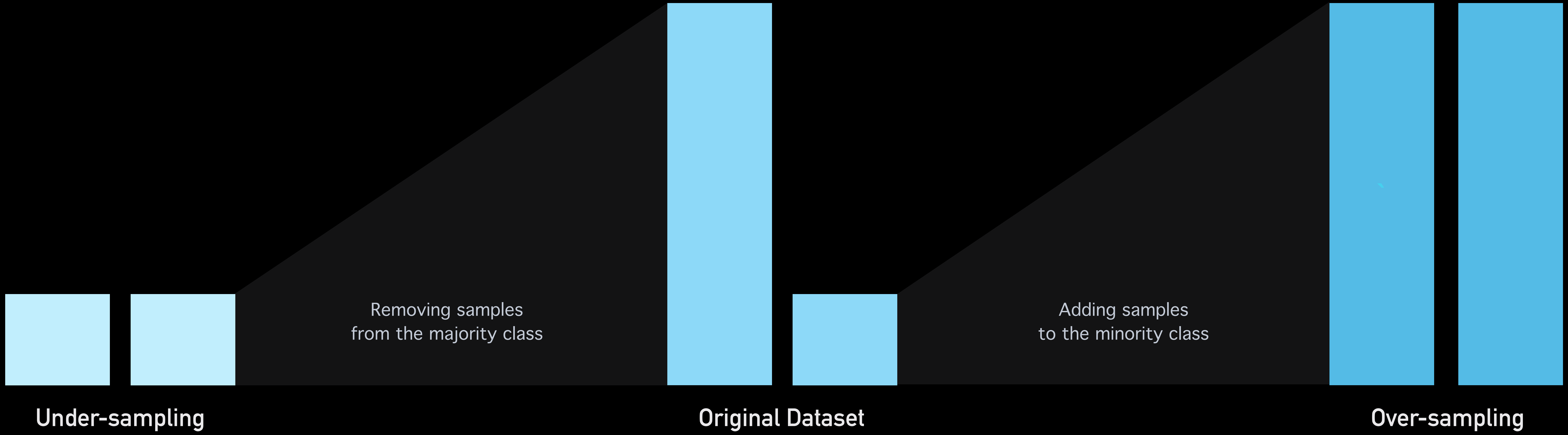


APPROACHES TO CLASS IMBALANCE PROBLEM

Under-sampling and Over-sampling

Addressing imbalanced data analysis poses a significant challenge in machine learning. Classifiers trained on imbalanced datasets often exhibit a bias towards the majority class, leading to over-prediction.

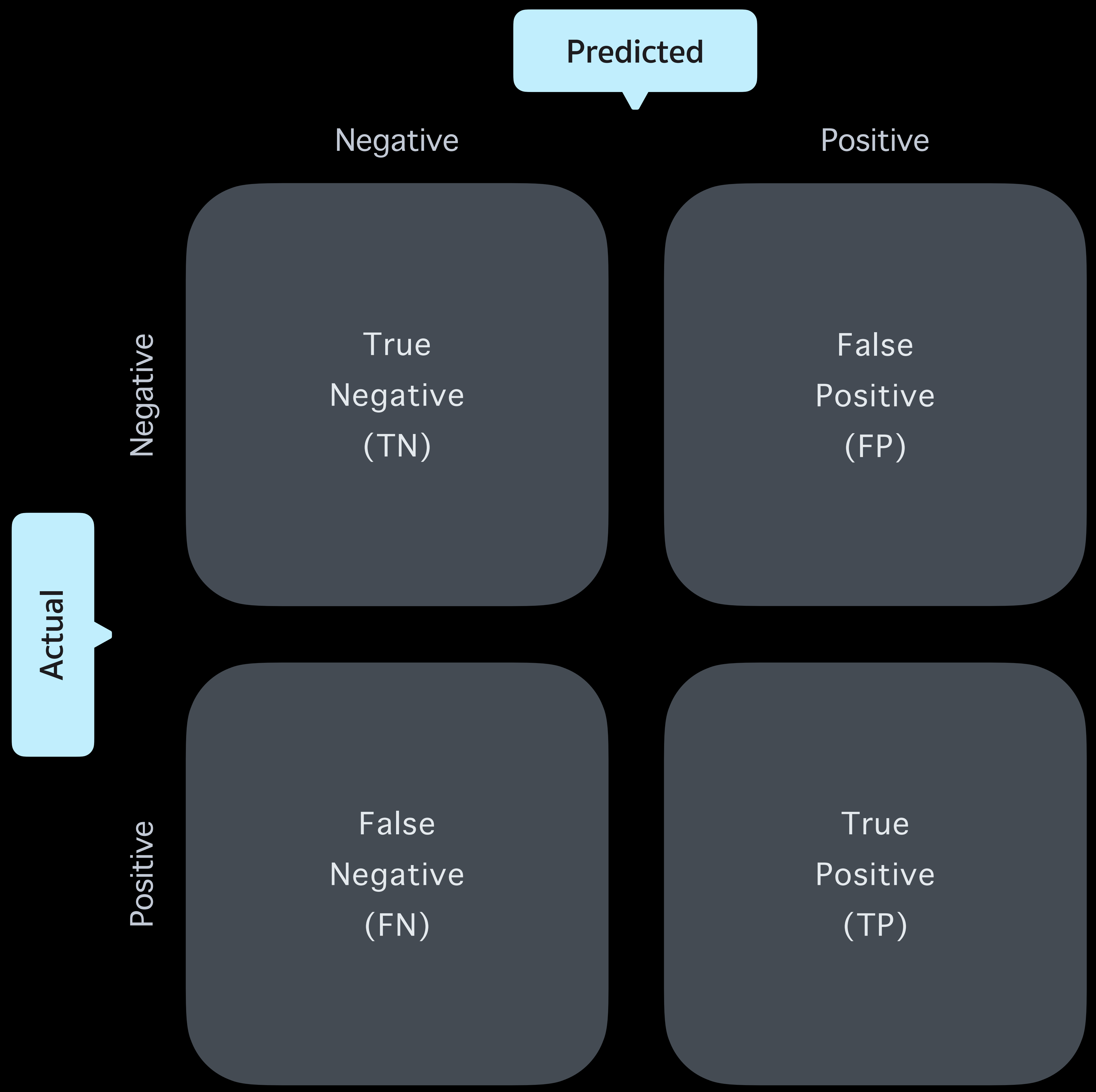
Various approaches can mitigate this issue, including under-sampling and over-sampling techniques. In this classification task, we will explore the effectiveness of Random Under-sampling and SMOTE (Synthetic Minority Oversampling Technique) to balance the dataset.



EVALUATION METRICS FOR IMBALANCED DATA

Confusion Matrix

A confusion matrix represents the prediction summary made by a model compared to the actual labels in the dataset, as illustrated. The confusion matrix is essential for computing a diverse range of evaluation metrics, such as accuracy, precision, recall, and the F1 score.



		Predicted	
		Negative	Positive
Actual	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

Metrics

The most commonly utilised metrics for imbalanced datasets include F1 score, precision, recall, ROC-AUC score, average precision score (AP), etc. In this classification task, we will focus on recall, precision, and F1 score.

Accuracy is a metric that indicates the proportion of correct overall predictions. However, accuracy alone may not be the most informative metric in highly imbalanced datasets. This is due to the fact that employing a simplistic strategy, such as blind guessing by favouring the majority class (genuine transactions), would still yield a high accuracy score of around 99.8% in this classification task.

Metrics	Formula	Interpretation
Recall	TP / (TP + FN)	The proportion of actual positive instances that were correctly identified.
Precision	TP / (TP + FP)	The proportion of positive identifications that were correct.
F1 Score	(2 × Precision × Recall) / (Precision + Recall)	The harmonic mean of a model’s precision and recall. A high F1 score indicates both high precision and recall.

WITHOUT UNDER-/OVER-SAMPLING TECHNIQUES

MACHINE LEARNING MODEL PERFORMANCES

Confusion Matrices

Before implementing under-sampling or over-sampling techniques, it is evident that KNN excelled in capturing the actual fraudulent transactions compared to other models.

True Label

Random Forest

	Genuine	Fraudulent
Genuine	100.00%	0.00%
Fraudulent	23.65%	76.35%

Logistic Regression

	Genuine	Fraudulent
Genuine	99.98%	0.02%
Fraudulent	37.16%	62.84%

KNN

	Genuine	Fraudulent
Genuine	99.98%	0.02%
Fraudulent	21.62%	78.38%

SVM

	Genuine	Fraudulent
Genuine	99.96%	0.04%
Fraudulent	22.30%	77.70%

Neural Networks Using TensorFlow

	Genuine	Fraudulent
Genuine	99.98%	0.02%
Fraudulent	27.03%	72.97%

Predicted Label

Metrics

Before implementing under-sampling or over-sampling techniques , it is evident that:

- KNN achieved the highest recall, indicating its effectiveness in capturing positive instances (correctly classifying fraudulent transactions).
- Random Forest achieved the highest precision, indicating a very low rate of false positives (incorrectly classifying fraudulent transactions).
- Random Forest has the highest F1 score, reflecting a balanced performance between precision and recall.

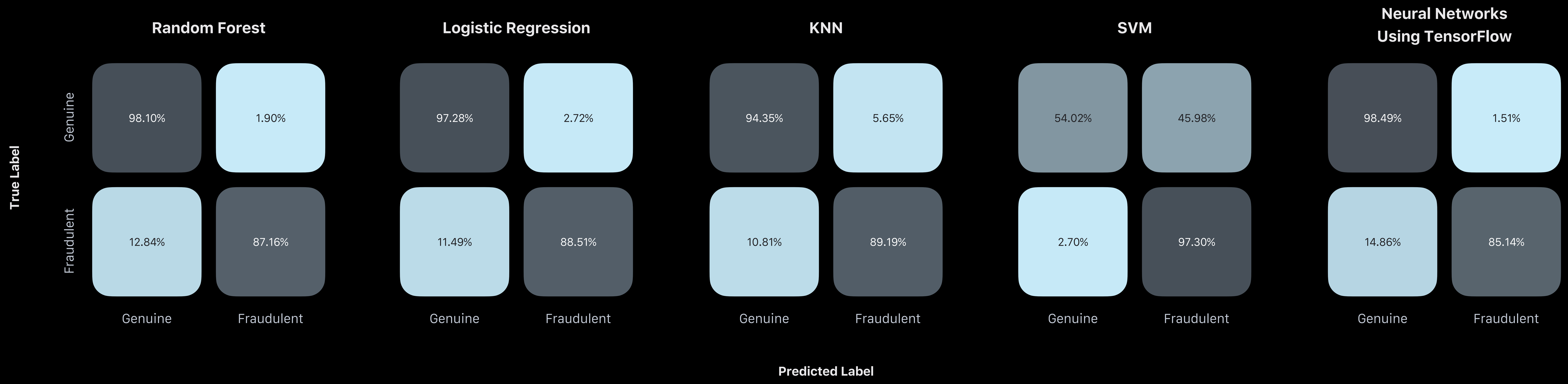
Models	Recall	Precision	F1 Score	Accuracy
Random Forest	0.763514	0.965812	0.852830	0.999544
Logistic Regression	0.628378	0.861111	0.726562	0.999181
KNN	0.783784	0.899225	0.837545	0.999473
SVM	0.777027	0.787671	0.782313	0.999251
Neural Networks Using TensorFlow	0.729730	0.837209	0.779783	0.999286

WITH RANDOM UNDER-SAMPLING TECHNIQUE

MACHINE LEARNING MODEL PERFORMANCES

Confusion Matrices

After implementing under-sampling, all models demonstrate an improvement in capturing actual fraudulent transactions.
However, this improvement comes at the cost of an increased rate of false positives, as the models become more sensitive to the minority class (fraudulent transactions).



True Label

Random Forest

	Genuine	Fraudulent
Genuine	98.10%	1.90%
Fraudulent	12.84%	87.16%

Logistic Regression

	Genuine	Fraudulent
Genuine	97.28%	2.72%
Fraudulent	11.49%	88.51%

KNN

	Genuine	Fraudulent
Genuine	94.35%	5.65%
Fraudulent	10.81%	89.19%

SVM

	Genuine	Fraudulent
Genuine	54.02%	45.98%
Fraudulent	2.70%	97.30%

Neural Networks Using TensorFlow

	Genuine	Fraudulent
Genuine	98.49%	1.51%
Fraudulent	14.86%	85.14%

Predicted Label

Metrics

After implementing under-sampling, it is evident that:

- All models show an improvement in recall after under-sampling, emphasising their enhanced ability to identify positive instances.
- Precision drops for all models, a common consequence of under-sampling, leading to an increased rate of false positives.
- The trade-off between precision and recall is evident in the F1 scores for all models.
- SVM exhibits exceptionally high recall but struggles with precision, resulting in the lowest F1 score and accuracy.
- Random Forest, Logistic Regression, and TensorFlow Neural Network maintain relatively balanced performance across metrics.

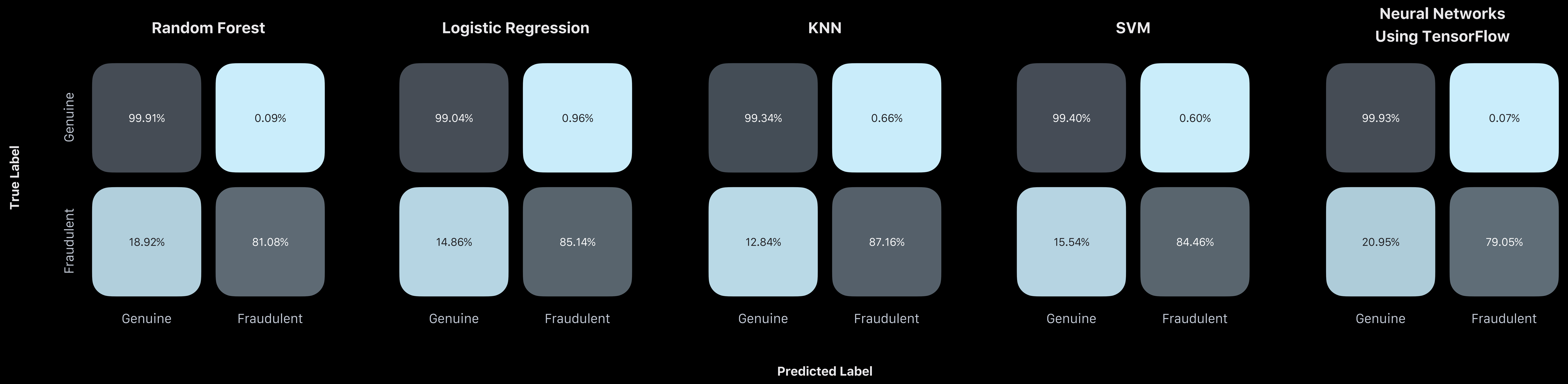
Models	Recall	Precision	F1 Score	Accuracy
Random Forest	0.871622	0.073756	0.136004	0.980818
Logistic Regression	0.885135	0.053491	0.100886	0.972672
KNN	0.891892	0.026667	0.051785	0.943424
SVM	0.972973	0.003658	0.007289	0.540910
Neural Networks Using TensorFlow	0.851351	0.088983	0.161125	0.984645

WITH SMOTE

MACHINE LEARNING MODEL PERFORMANCES

Confusion Matrices

After implementing SMOTE, there is a consistent improvement in identifying actual fraudulent transactions across all models. However, similar to under-sampling, this improvement with SMOTE comes at the cost of an increased rate of false positives.



Random Forest

	Genuine	Fraudulent
Genuine	99.91%	0.09%
Fraudulent	18.92%	81.08%

Logistic Regression

	Genuine	Fraudulent
Genuine	99.04%	0.96%
Fraudulent	14.86%	85.14%

KNN

	Genuine	Fraudulent
Genuine	99.34%	0.66%
Fraudulent	12.84%	87.16%

SVM

	Genuine	Fraudulent
Genuine	99.40%	0.60%
Fraudulent	15.54%	84.46%

Neural Networks Using TensorFlow

	Genuine	Fraudulent
Genuine	99.93%	0.07%
Fraudulent	20.95%	79.05%

Metrics

After implementing SMOTE, it is evident that:

- There is a consistent improvement in recall across all models after applying SMOTE, indicating better identification of positive instances.
- However, this improvement in recall comes at the cost of precision, leading to an increased rate of false positives in all models.

Comparing to random under-sampling, it is evident that:

- Random under-sampling tended to achieve higher recall, but at the cost of significantly lower precision for all models.
- SMOTE generally outperformed random under-sampling for all models by providing a more balanced combination of recall and precision, resulting in higher F1 scores.

Models	Recall	Precision	F1 Score	Accuracy
Random Forest	0.810811	0.603015	0.691643	0.998748
Logistic Regression	0.851351	0.133192	0.230347	0.990145
KNN	0.871622	0.187228	0.308244	0.993224
SVM	0.844595	0.195618	0.317662	0.993715
Neural Networks Using TensorFlow	0.790541	0.676301	0.728972	0.998982

Summary

Comparing the results of the models before applying resampling techniques, after applying random under-sampling, and after applying SMOTE, it is evident that:

- The models had varying levels of precision and recall before resampling
- Random under-sampling increased recall but at the cost of precision for most models
- SMOTE provided a more balanced improvement, with a trade-off between recall and precision

In conclusion, SMOTE generally outperformed random under-sampling by providing a more balanced improvement in both recall and precision for these models. Ultimately, the choice of the best model and resampling techniques depends on the business goal, such as the importance of precision versus recall, as well as considerations like computational efficiency. In this classification scenario, our objective is to identify fraudulent credit card transactions to ensure customers are not charged for items they did not purchase. While recognising fraudulent transactions is crucial, avoiding misclassification of transactions as fraud is also important to prevent unnecessary investigations or actions. For instance, if customers regularly make purchases with their credit cards and the model incorrectly classifies these transactions as fraud, it could lead to increased customer complaints and dissatisfaction.

Therefore, balancing recall and precision is crucial in this classification scenario. The F1 score, which balances precision and recall, can be a useful metric. Given the importance of both recall and precision, models that provide a higher F1 score after SMOTE can be preferable, as models after SMOTE generally offer a better balance between recall and precision compared to random under-sampling. Notable examples include Random Forest and neural networks using TensorFlow. Determining the best model ultimately depends on specific business requirements, such as the overall objectives of the fraud detection system.

THANK YOU