Machine Learning Models:

Original Data:

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.9352	0.9358	0.9352	0.9247
Decision Tree	0.9133	0.9128	0.9133	0.9130
Random Forest	0.9330	0.9325	0.9330	0.9223
Gradient Boosting	0.9317	0.9295	0.9317	0.9213
SVM	0.9361	0.9364	0.9361	0.9263
K-Nearest Neighbors	0.8981	0.9021	0.8981	0.8600

Even though there is a class imbalance the model performance is pretty consistent. The model is able to detect the minority class pretty well. The reason is quite simple, because it is text data, both classes are quite distinct and with proper vectorization technique, the model can learn the intricacies of the minority class properly as well. This is further proven by the following table.

After Applying Oversampling Method (SMOTE-ENN)

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.9192	0.9287	0.9192	0.9228
Decision Tree	0.8822	0.9020	0.8822	0.8901
Random Forest	0.9289	0.9230	0.9289	0.9237
Gradient Boosting	0.8998	0.9093	0.8998	0.9038
SVM	0.9426	0.9393	0.9426	0.9379
K-Nearest Neighbors	0.8889	0.8741	0.8889	0.8794

The traditional machine learning model performed well enough and they took very little time to train, on the other hand deep learning models took quite a while for similar performance. This is because the dataset is in English language. The existing vectorization techniques are extremely

effective and that's why even traditional machine learning models can achieve such high performance.

	Model	Accuracy	Precision	Recall	F1 Score
Deep Learning Models	Feed Forward NN	0.9233	0.9175	0.9233	0.9112
	CNN	0.9225	0.9183	0.9225	0.9085
	RNN	0.8871	0.7870	0.8871	0.8341
Transformers	BERT	0.9328	0.6598	0.8356	0.7374