

Model Mavericks

**Problem Statement 3- Bank Customer
Complaint Classification**

Objectives



Data Collection and Pre-processing

Classification model development

Sentiment Analysis

Model Training and Evaluation

Techstack



NLTK



BeautifulSoup

Proposed Workflow

Data preprocessing

CFPB dataset

Pre-processing
(Tokenization, Stop word removal, Label Encoding, etc.)

Exploratory Data Analysis

Model Development

Model selection
(RFC, Multinomial NB, etc.)

Train - test split

Fine tuning hyperparameters

Sentiment Analysis

Sentiment analysis
(list of keywords)

SVC model training

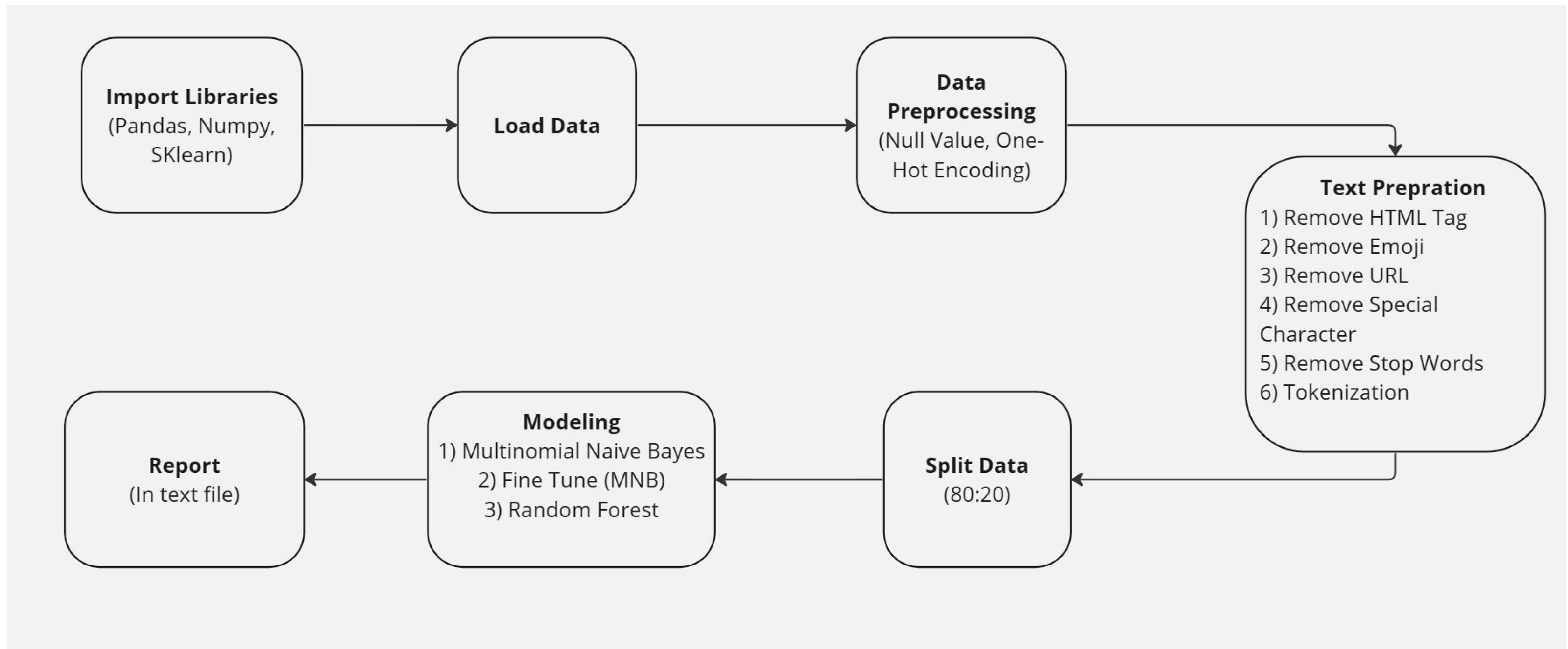
Gain insights

Model training and Evaluation

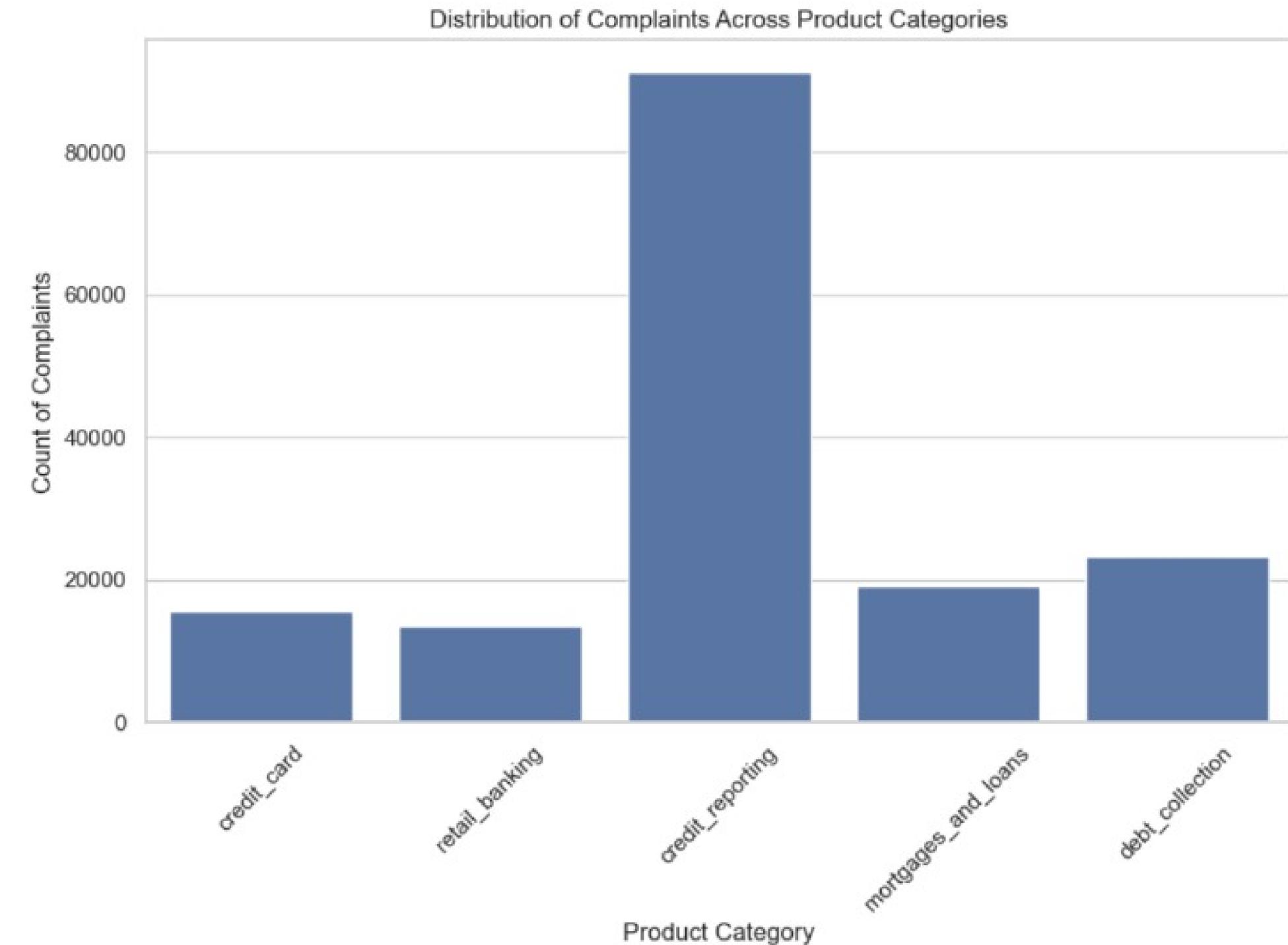
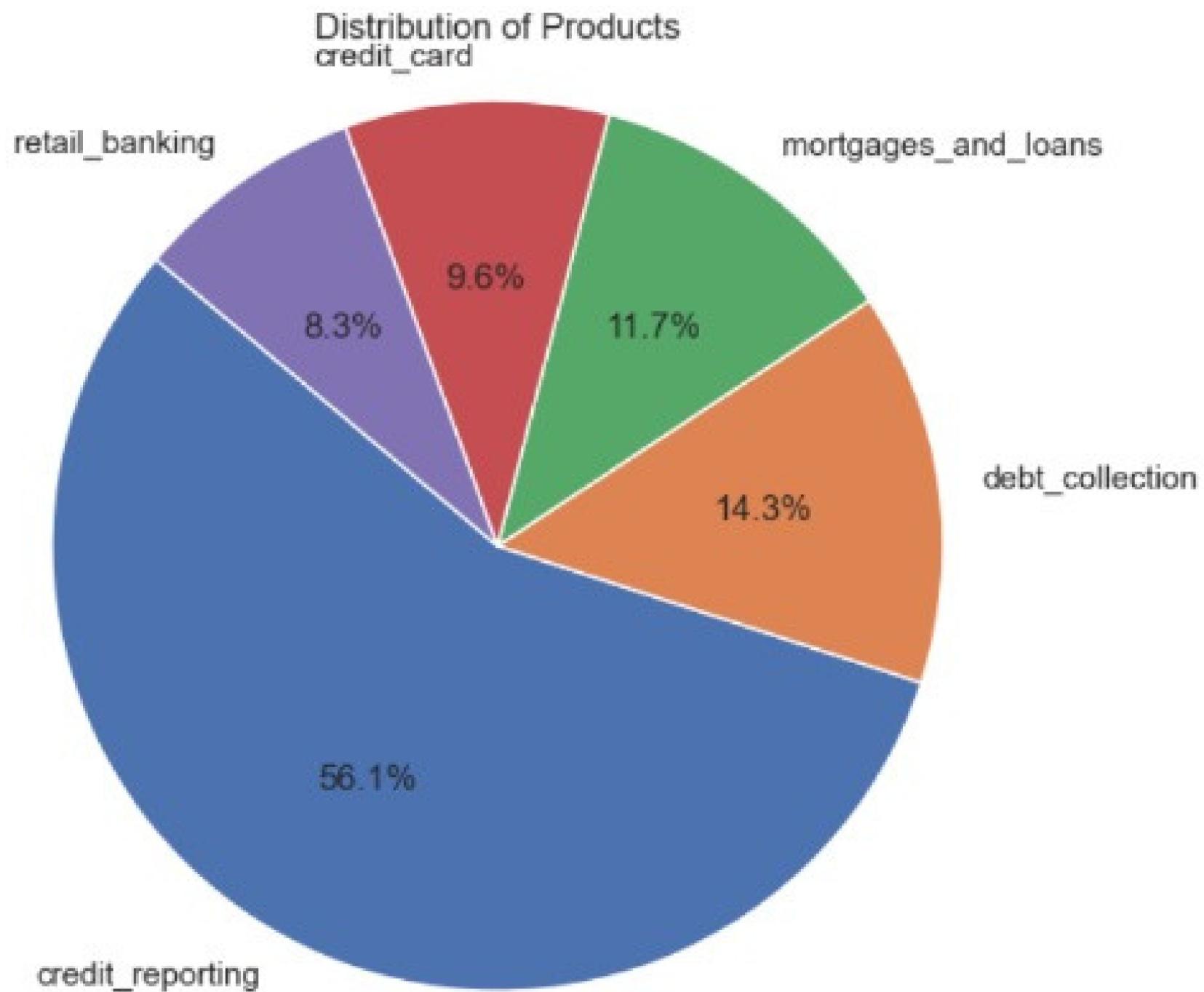
Evaluation metrics

Cross-validation accuracy

Performance analysis

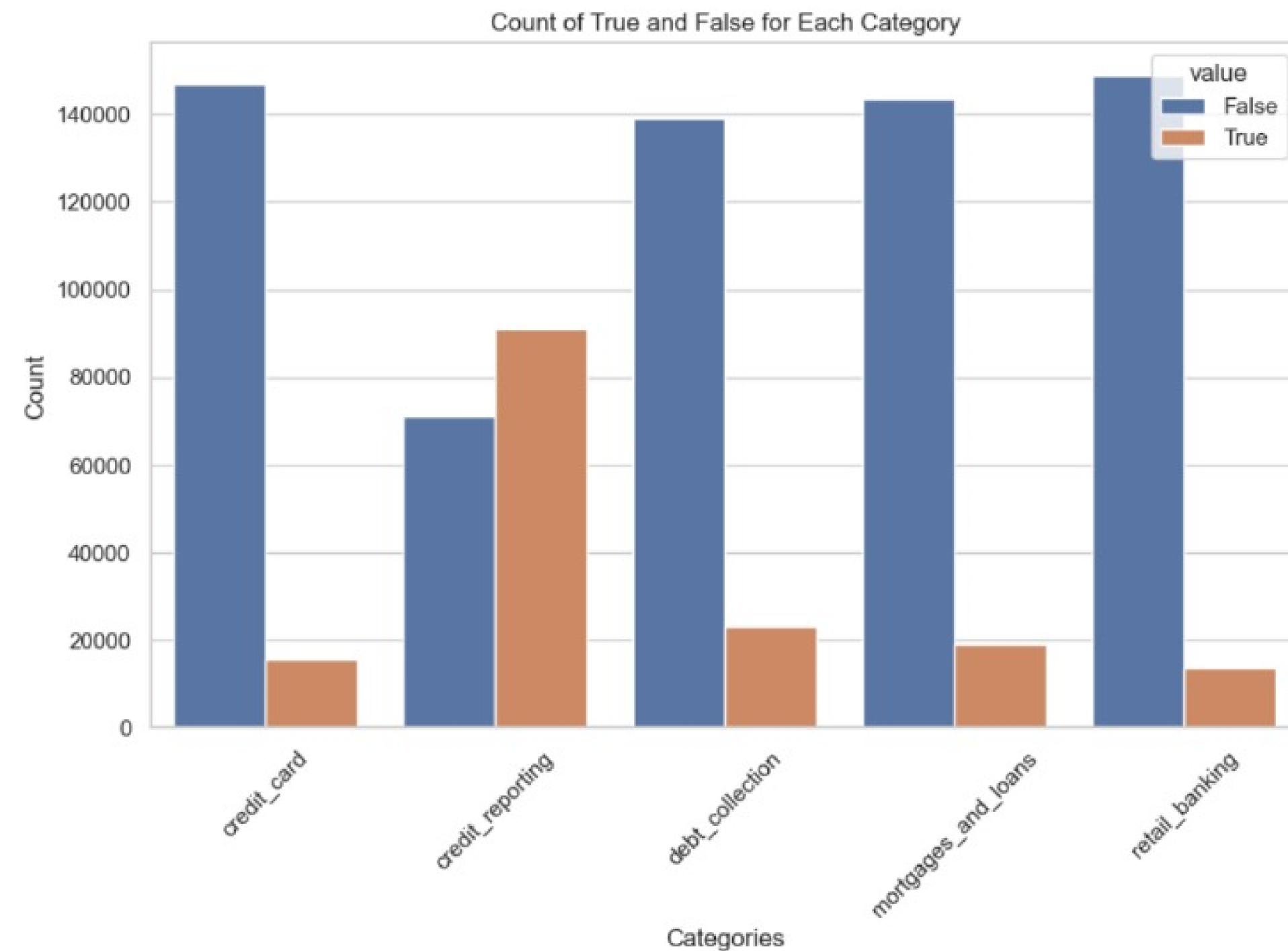


EDA

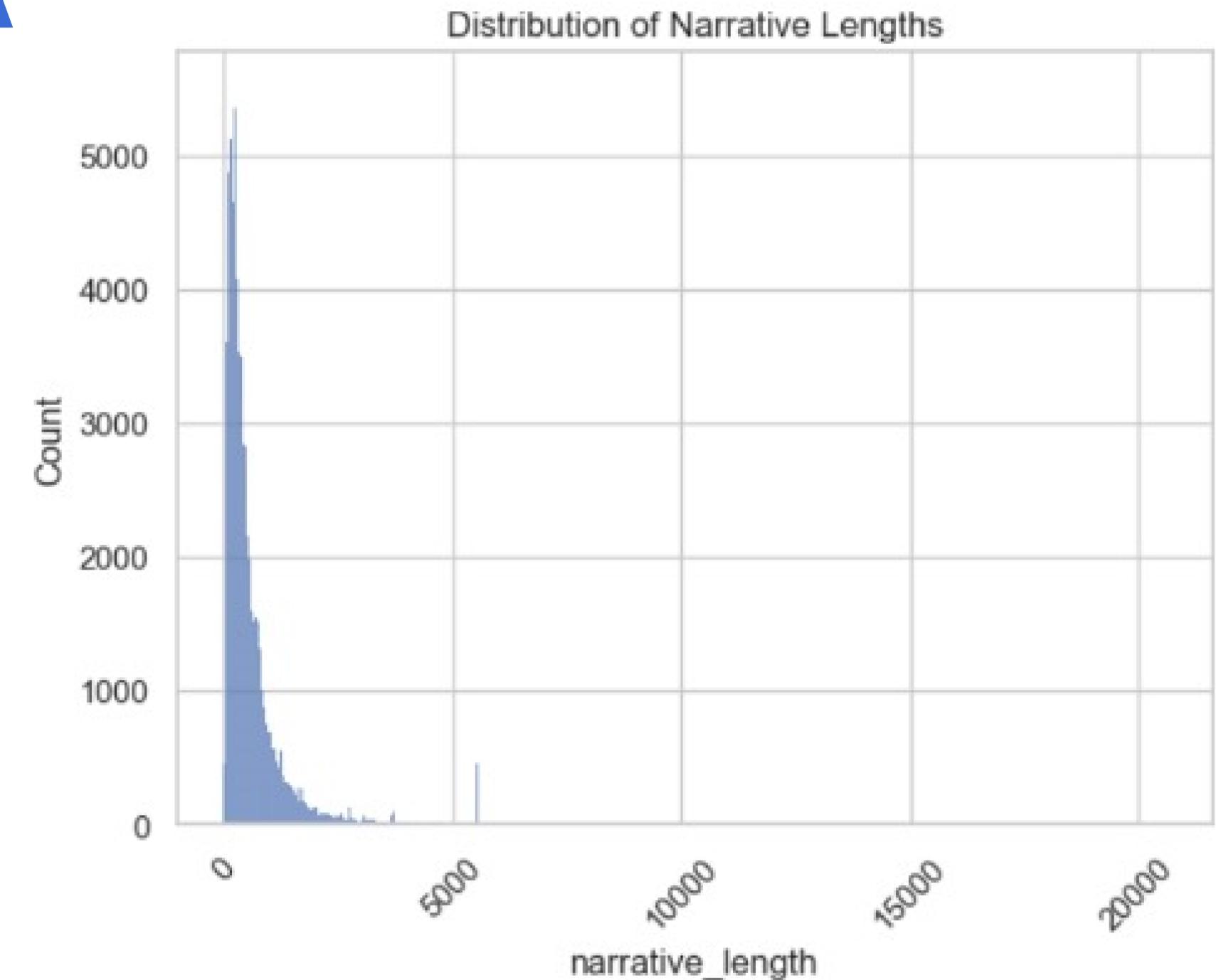


Credit reporting has the highest percentage of complaint count as compared to other classes.

EDA



Total count of that particular class with respect to the entire dataset



Histogram of the lengths of narratives
(Left - skewed data)

Results

Model	Accuracy	Precision	Cross validation accuracy
Multinomial NB	81	83	81
Multinomial NB (BOW)	82	76	81
Multinomial NB (Fine tuning)	88	83	
Random Forset Classifier	88	87	
Random Forset Classifier (Fine Tuning)	82	85	

Results

Compliance Classification

Prediction model

Enter your complaint

I'm extremely displeased with the performance of my credit card account. Despite meeting all payment deadlines and maintaining good credit, I've encountered numerous issues. Hidden fees and unclear terms have resulted in unexpected charges on my statement. Furthermore, the customer service response has been slow and unhelpful when attempting to resolve these matters. The lack of transparency and communication is unacceptable. As a loyal cardholder, I demand clarity in fees, improved customer service, and prompt resolution of disputes. This experience tarnishes my trust in your credit card services and prompts me to consider alternatives if these issues persist.

Classify

Predicted Category: Credit card

UI displaying correctly predicted output on a randomly generated complaint

Results

The screenshot shows a Jupyter Notebook interface with the following content:

```
# Train a classifier (Support Vector Machine in this example)
clf = SVC(kernel='linear')
clf.fit(X_train_vec, y_train)

# Evaluate the classifier
y_pred = clf.predict(X_test_vec)
print(classification_report(y_test, y_pred))

[98] ✓ 26m 13.3s
... precision recall f1-score support
...
negative 0.99 0.95 0.97 5263
neutral 0.98 1.00 0.99 24424
positive 0.99 0.88 0.93 2796
accuracy 0.98 0.98 0.98 32483
macro avg 0.98 0.94 0.96 32483
weighted avg 0.98 0.98 0.98 32483
```

```
import joblib
joblib.dump(clf, 'sentiment_analysis_model.pkl')

[100] ✓ 0.0s
... ['sentiment_analysis_model.pkl']

new_complaints = ["This is illegal"]
new_complaints_vec = vectorizer.transform(new_complaints)
predicted_sentiments = clf.predict(new_complaints_vec)
print(predicted_sentiments)

[99] ✓ 0.0s
... ['negative']
```

The notebook also displays a classification report table:

	precision	recall	f1-score	support
negative	0.99	0.95	0.97	5263
neutral	0.98	1.00	0.99	24424
positive	0.99	0.88	0.93	2796
accuracy			0.98	32483
macro avg	0.98	0.94	0.96	32483
weighted avg	0.98	0.98	0.98	32483

Sentiment Analysis using SVM gives an accuracy of 98% and correctly predicts when given new input.

However, this model was developed using smaller dataset.

Future Scope

1. Sentiment Analysis
2. Multimodal Analysis
3. Multilingual Integration
4. Personalized Resolution
5. Real-time Monitoring
6. Feedback loop (human intervention)
7. LLM integration

Thank You