Title Page

Fusemachines

Youtube Sentimental Analysis

Title Page Proprietors Introduction

Sentimental Analysis

Introduction

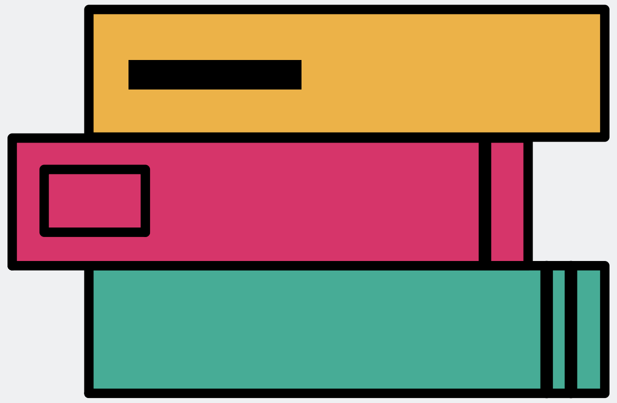
Sentimental Analysis involves the classification of human language sentences into predefined categoriees.

Our youtube comment sentimental analysis involves three categories namely positive , neutral and negative.

Title Page | Proprietors | Introduction

Steps

Introduction



Steps done

1. Scraped comment data via youtube api
2. Preprocessed the scraped data
3. Labelled the dataset with textBlob package

4. Naiive Bayes Model was trained
5. Model performance measured
6. Retraining and reevaluating model with some changes

Scraping

The First Attempt

Initial attempt was done with selenium which was very slow and unfeasible

The Second Attempt

Involved using the youtube api to get json of requested number of comments (16333)

Preprocessing

Html tags and links removed

Emoji replaced with words

Removal of non english alphabets

Labelling TextBlob

Positive Comments

4609 comments were labelled as positive

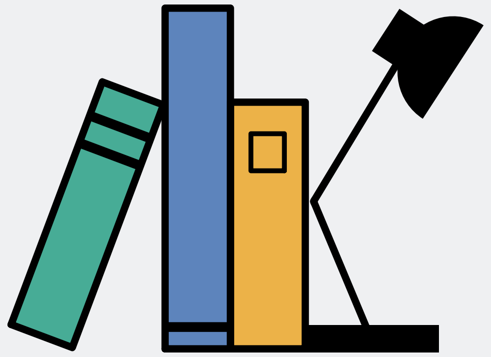
Neutral Comments

10865 comments were labelled as neutral

Negative Comments

859 comments were labelled as negative

The dataset is imbalanced



Model training

Remove special characters

Remove stop words

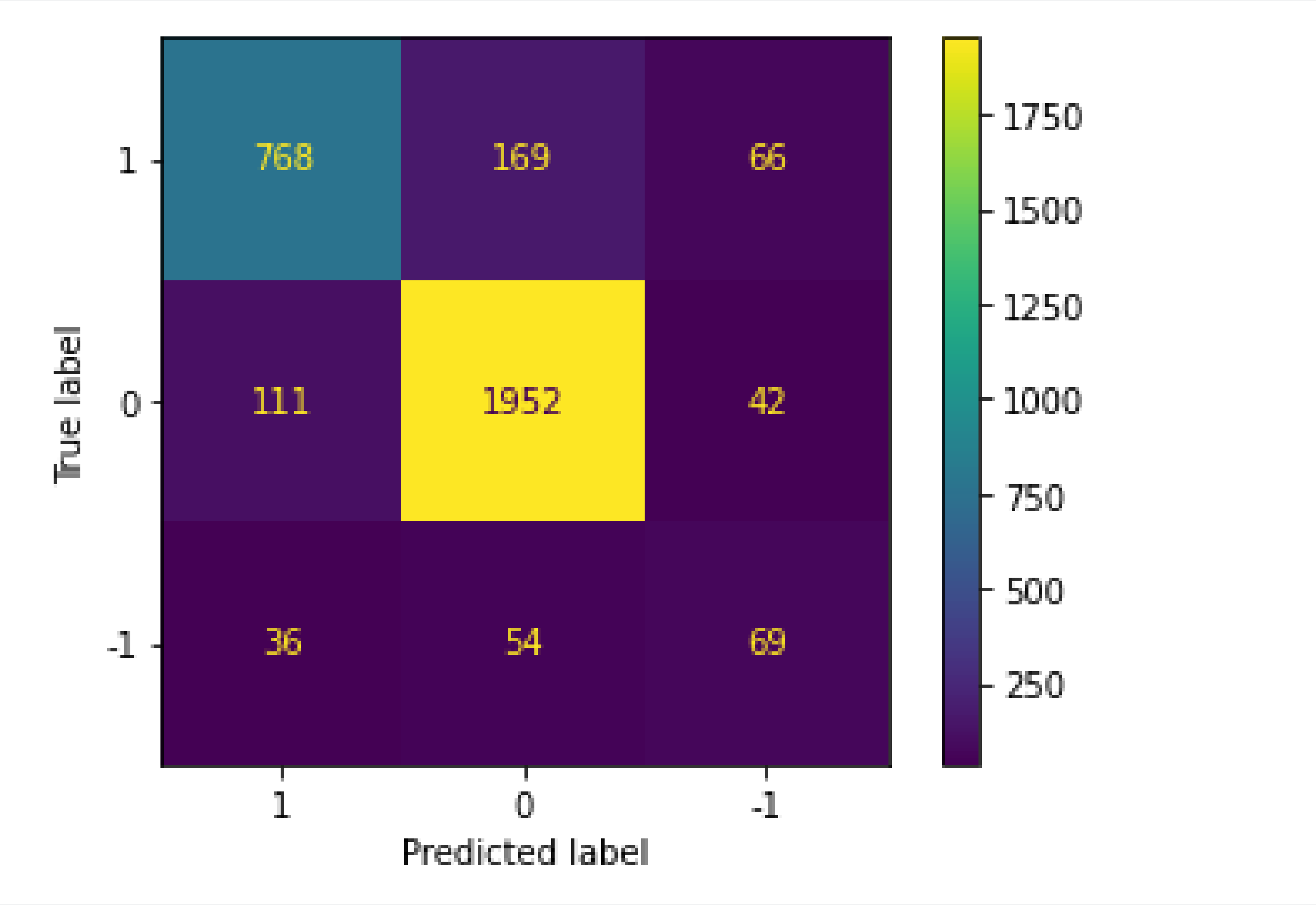
Lemmatize instead of stemming

Vectorize by bag of word model

Multinomial Naiive bayes classifier model used

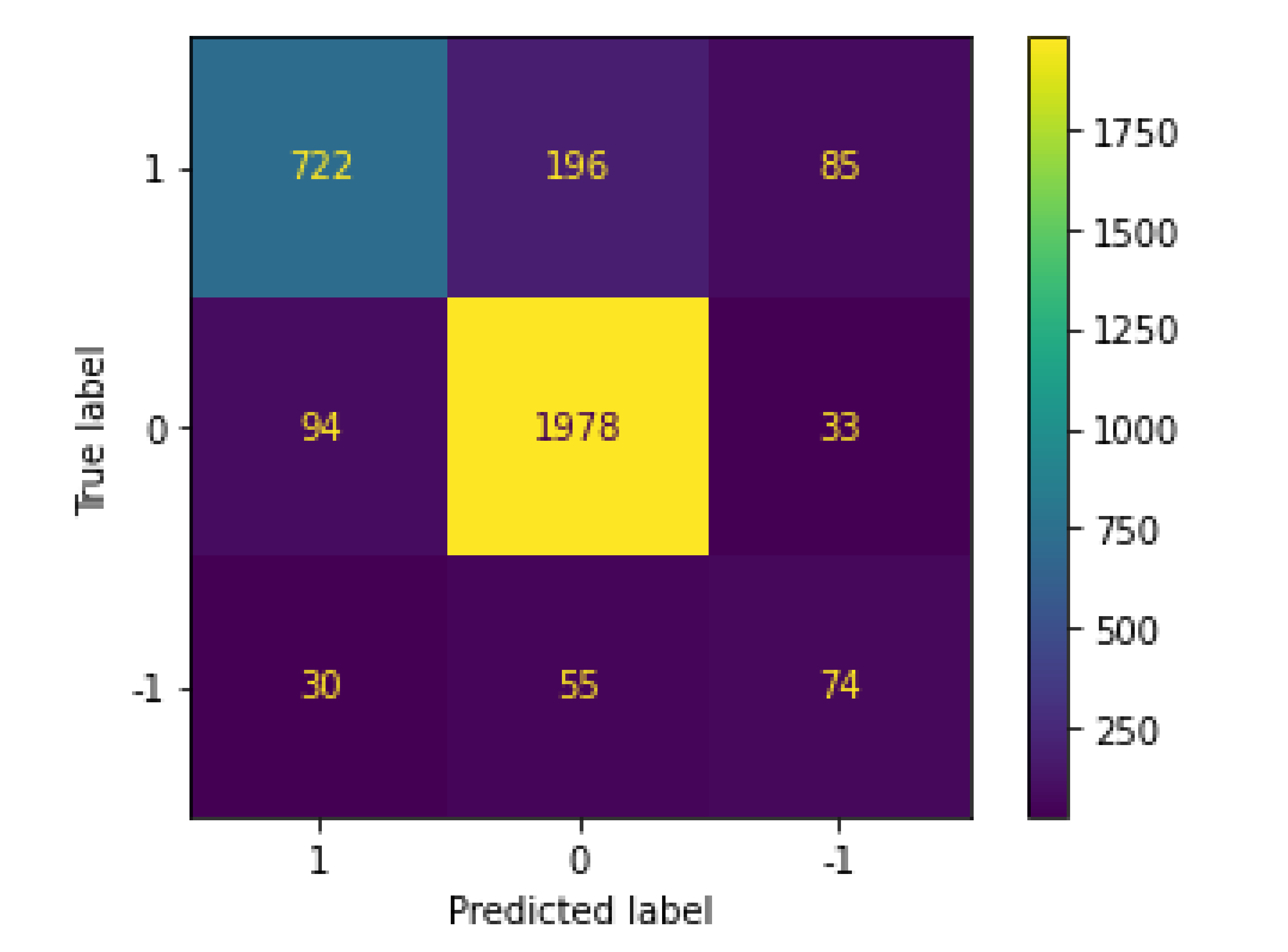
Results 1

No consideration for the skewed dataset taken



Results 2

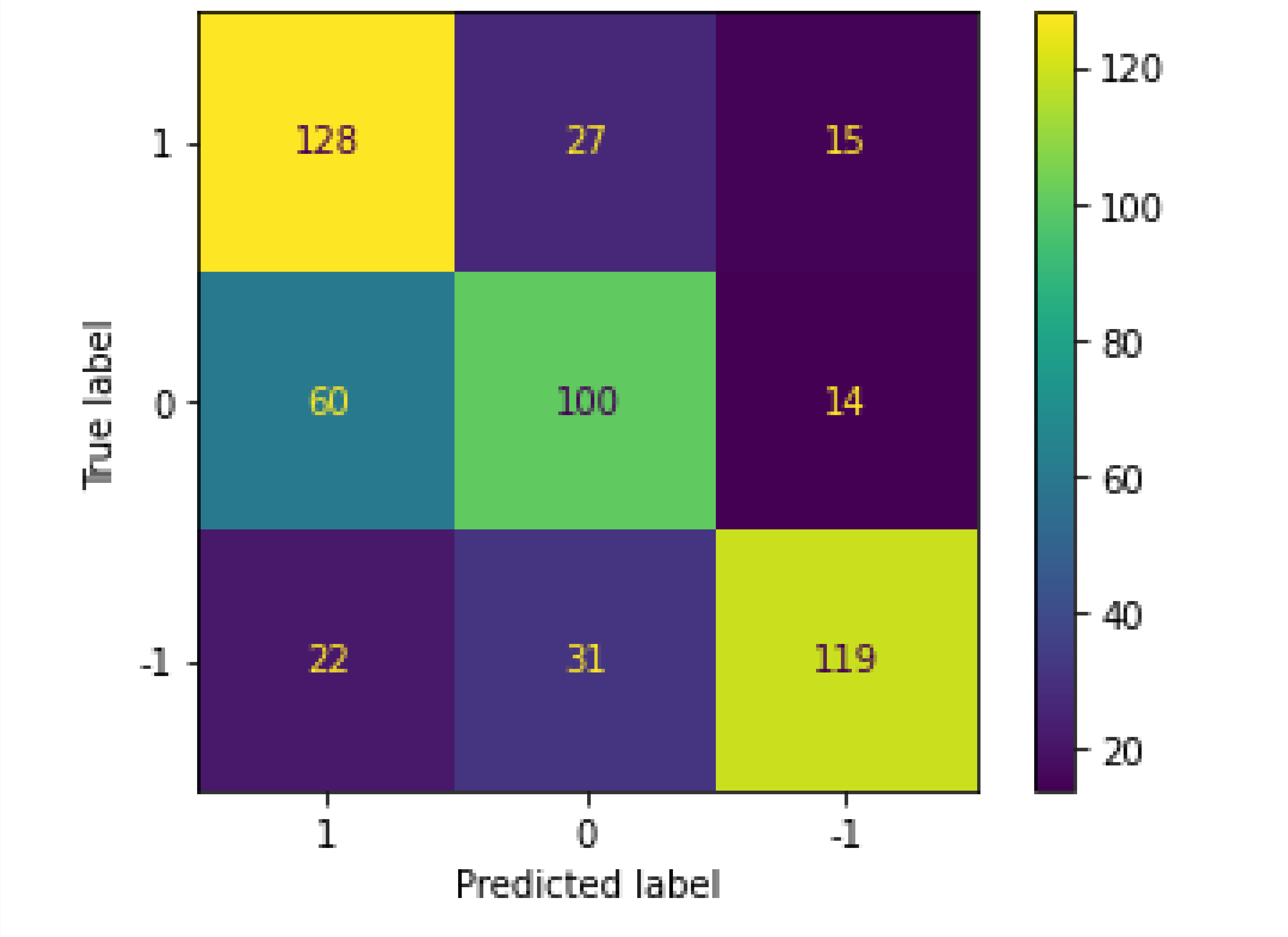
Number of features limited to only 25k



	1	0	-1
1	722	196	85
0	94	1978	33
-1	30	55	74

Results 3

Sampalled similar number of each category



True label \ Predicted label	1	0	-1
1	128	27	15
0	60	100	14
-1	22	31	119

Conclusion

The model with equally sampled training data performed the best for the classificatiion task. This emphasises the analysis of dataset skewness before the model training steps