# Comparison report

## Vader Sentiment vs Logistic Regression (with RandomOverSampling)

The dataset contains 1288 rows and it is not balanced. There are 561 negative tweets, 455 positive tweets, 271 neutral tweets.

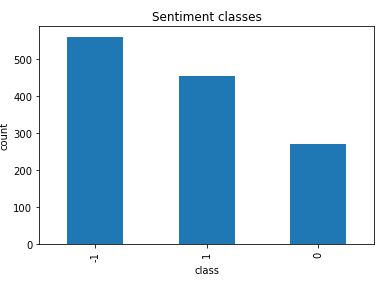


Figure 1 Dataset class imbalance

I built and compared several classification models – SVM and Logistic Regression. First, I trained the models with imbalanced data. Then, I built same models with SMOTE and RandomOverSampling techniques to balance the training data. Results showed, that Logistic Regression with RandomOverSampling performed the best and achieved the overall accuracy score of 0.60. It is still much lower than the result obtained with Vader Sentiment analysis tool (overall accuracy 0.77).

In the table below we can observe the accuracy metrics. Vader sentiment analysis tool produced much better accuracy metrics. Only f1-score for neutral class is close in both results, .50 for Logistic Regression and .55 for Vader. Logistic Regression precision metric is considerably low (.44), while the recall for the is quite high (.58). High recall but low precision returns many results, but most of its predicted labels are incorrect when compared to the training labels. And this can be seen in confusion matrix. Logistic Regression predicted neutral sentiment correctly 14 times out of 32 total negative predictions, while Vader predicted neutral sentiment correctly 11 times out of 16 negative predictions.

Both classifiers were most accurate when predicting negative sentiments, less accurate when predicting positive sentiments, and worst accuracy was achieved when predicting neutral sentiments. This might be due to the fact that the test data is not balanced. Most tweets are negative (61 cases in the test set) and only few neutral (24 cases in the test set).

|  |  |
| --- | --- |
| Logistic Regression + RandomOverSampling | Vader Sentiment |
|  |  |
|  |  |

Table 1 Accuracy metrics and confusion matrices

Table below shows broken down predictions. For example, both models correctly predicted negative for the same tweets 40 times1. When the classifiers disagreed, logistic regression was correct only once­2. 8 times3 both classifiers predicted negative sentiment when the actual class was positive (same tweets).

|  |  |  |  |
| --- | --- | --- | --- |
| Actual class | Logistic Regression prediction | Vader Prediction | Count |
| -1 | -1 | -1 | 401 |
| -1 | 0 | 12 |
| 0 | -1 | 2 |
| 0 | 0 | 3 |
| 0 | 1 | 1 |
| 1 | -1 | 6 |
| 1 | 1 | 83 |
|  |  |  |  |
| 0 | -1 | -1 | 3 |
| -1 | 0 | 1 |
| 0 | -1 | 1 |
| 0 | 0 | 9 |
| 0 | 1 | 4 |
| 1 | -1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 4 |
|  |  |  |  |
| 1 | -1 | -1 | 1 |
| -1 | 1 | 8 |
| 0 | -1 | 2 |
| 0 | 0 | 1 |
| 0 | 1 | 9 |
| 1 | 1 | 23 |