Sentimental analysis for yelp dataset

]Bhaskar Jupudi jupudi@ucsc.edu]Trivikram Bollempal-
litbollemp@ucsc.edu

Abstract

In this project, we aim to perform sentiment anal-
ysis i.e., classifying whether the review is pos-
tive or negative using the yelp dataset based on
reviews and ratings. The classification problem
can be solved by a set of algorithms. Every algo-
rithm has its own advantages and disadvantages
in terms of accuracy and model complexity.

For example, Naive Bayes classifier is faster to
compute than Logistic Regression classifier for
huge datasets. But the disadvantage with the for-
mer classifier is that it assumes that features are
independent where as the latter has no such as-
sumptions which can lead to better precdiction.
Our work mainly concentrates on implement-
ing these two classifiers and techniques to make
them perform much better. We have adopted
multi-processing for feature extraction to make
it way faster and also implemented two differ-
ent approaches of Logistic regression for both
binary and multi-class classfication. We have
also implemented Naive Bayes classifier. Finally,
we contrast these two algorithms based on time
taken for execution and performance metrics like
accuracy, precision and recall.

1. Problem statement

asasd

2. Feature Extraction

asdasdas

3. Model Formulation

4. Evaluations

5. Results

References

Langley, P. Crafting papers on machine learning. In Lang-
ley, Pat (ed.), *Proceedings of the 17th International Con-
ference on Machine Learning (ICML 2000)*, pp. 1207–
1216, Stanford, CA, 2000. Morgan Kaufmann.