

## Comparison of Machine Learning Classifiers for Sentiment Analysis in Hotel Reviews

PLU Kaushalya# and WU Wickramaarachchi

*Department of Computing, Rajarata University of Sri Lanka, Mihintale, Sri Lanka*

#udanikaushalya95@gmail.com

Sentiment analysis or opinion mining refers to the process of identifying people's sentiments, opinions, attitudes and emotions behind a written text. In recent years, sentiment analysis studies have become an active research area under natural language processing. Understanding the opinion behind the user-generated text can be applied to various applications. When it comes to the hotel sector and travel planning, user reviews and comments are quite useful. Therefore, guest reviews are becoming a prominent factor, which influence people's booking decisions. In addition, knowing about these comments is important for quality control of the hotel management too, because it may be worth checking out some stats over time. The fundamental objective of this research is to compare several machine learning classifiers and find out the best classifiers to develop a sentiment analysis model for the hotel reviews, to tackle customers' sentiment. Under this research, a comparative analysis was established among Multinomial Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB), Logistic Regression (LR), Stochastic Gradient Descent Classifier (SGD), Linear Support Vector Classifier (SVC), Random Forest Classifier and Multi-layer Perceptron Classifier (MLP) classifiers. Moreover, two feature extraction techniques called Count Vectorizer and Term Frequency Inverse Document (TF-IDF) are also compared to find out the best approach to perform the feature extraction. The result from this research shows that the highest results were obtained in Logistic Regression with TF-IDF method (Accuracy 87.39%) and SGD algorithms with TF-IDF (Accuracy 87.71%), while the lowest accuracy was obtained for Bernoulli NB classifier with Count Vectorizer (Accuracy 64.67%). Every time when using Count Vectorizer as the feature extraction method, the accuracies decreased, than when the TF-IDF method was used.

**Keywords:** *sentiment analysis, machine learning classifiers, feature extraction techniques*