Abstract

Sentiment Analysis is the study of people’s opinions and emotional feedbacks towards an entity which can be products, services, individuals or events. The opinions are most presumably be expressed as reviews or comments. With the advent of social networks, forums and blogs, these reviews emerged as an important factor for the customers’ decision for the purchase or choice of any item. Nowadays, a vast scalable computing environment provides us with very sophisticated way of carrying out various data-intensive natural language processing (NLP) and machine-learning tasks to analyze these reviews. One such task is text classification, a very effective way of predicting customers’ sentiment. This paper investigates the different ways of sentiment analysis from customers’ review using machine learning algorithms. For classifying text from overall sentiment, we considered two class, i.e. predicting whether a comment or review is positive or negative. In our study, we used two popular public datasets and six different machine learning algorithms – Naïve Bayes (Multinomial and Bernoulli), Logistic Regression, SGD (Stochastic Gradient Descent), Linear SVM (Support Vector Machine) and RF (Random Forest). Moreover, we applied parameter optimization on SVM and SGD classifiers on
different threshold values to identify and analyze the differences in the accuracy of the classifiers and to obtain the optimal outcome from the model.

References


Index Terms

Computer Science Algorithms

Keywords

Natural Language Processing, Sentiment Analysis, Opinion mining, Machine Learning.