**Amazon Product Review Analysis**

Sanjeev Ramasamy Seenivasagmani Siddharth Kumar S

School of Information Studies School of Information Studies

Syracuse University Syracuse University

Syracuse, NY 13244 Syracuse, NY 13244

[saramasa@syr.edu](mailto:saramasa@syr.edu) -email-

**Abstract**

As people started providing product reviews in abundance, it is difficult to find a middle ground to evaluate the reviews and determine the reason behind their reviews. In this paper, we will be analyzing the reviews provided on the Amazon website using Multinomial Naïve Bayes and Support Vector Machine algorithms to predict the sentiment of the users towards the product and perform topic modelling using Latent Dirichlet Allocation algorithm to determine the topics discussed or in simple terms, the reason behind such ratings/reviews.

**1** **Introduction**

Classifying data is what we will be doing in this project and Support Vector Machines (SVM) are the frontier in classifying the data with the n-dimensional hyperplane, n being the number of features. Alongside SVM, we will also be using another classifier called Multinomial Naïve Bayes (MNB) classifier which is a specific instance of a Naïve Bayes Classifier which uses a multinomial distribution for each of the features. Latent Dirichlet Allocation (LDA) is a generative probabilistic model of a collection of documents made up of words and/or n-grams (set of words) which we will be using for performing topic modelling.

We will be applying the above-mentioned algorithms to the data and build a model that will predict the sentiment of the provided comment/review and cluster the word according to the topics that were discussed.

**2** **Method**

With the help of SVM & MNB algorithms we will be building models to accurately classify the comments based on their sentiment. The algorithm with the highest accuracy will provide the sentiment based on which we will be performing topic modelling.

**2.1 Data**

**2.1.1 Data Collection**