**Title:**

**Sentiment analysis of reviews using different machine learning algorithms on textual data.**

**Project Member:**

• Soma Dasgupta

**Purpose of the project:**

The main aim of the project is to perform text mining and to analyze textual data by classifying based on polarity (Negative or Positive) of reviews. The dataset used for this project is reviews given by various users to different products bought from Amazon. Our intention is to understand how each machine learning algorithm is handling the textual data. By carrying out performance analysis we want to determine how well each classifier is working by computing their accuracy.

**Approach:**

• As the dataset is in XML format, it should be preprocessed accordingly to make the data organized, to eliminate unnecessary tags (i.e. text) and find ways to handle the inconsistent or incomplete data.

• Now the organized data needs to be cleaned such as emoticons, hashtags, punctuations etc. using regular expressions.

• Using Natural language processing module, we remove Stop words, Stemming and perform Tokenization.

• Different machine learning algorithms such as Recurrent Neural Networks, Naïve Bayes, Support Vector Machines are used to build the predictive models and the reviews are classified into positive and negative.

• The data is divided into train and test sets in the ratio 70 and 30 to serve this purpose.

• Performance analysis will be carried out by determining sensitivity and specificity which will help in determining ROC and AUC. Also, accuracy and time complexity will be calculated.

• The project will be implemented in Python environment using Scikit-learn, nltk, pandas, numpy, matplotlib, etc.,

**Deliverables:**

• Source Code

• Project Report

• Project Demo