**Movie review prediction using NLP and**

**Naïve Bayes**

Submitted by:

**Mohit Singh ( 9916103239 )**

**Paras Kaushik ( 9916103021 )**

**Divik Shrivastava ( 9916103048 )**

Under the supervision of

**Dr. Shikha Mehta**



**Department of CSE/IT**

**Jaypee Institute of Information Technology University, Noida**

**NOVEMBER 2018**

**ACKNOWLEDGEMENT**

In performing our assignment, we had to take the help and guideline of some respected persons, who deserve our greatest gratitude. The completion of this assignment gives us much Pleasure. We would like to show our gratitude to our lecturer

**Mrs. Shikha Mehta** for giving us a good guideline for assignment throughout numerous consultations. We would also like to expand our deepest gratitude to all those who have directly and indirectly guided us in writing this assignment.

Many people, especially our classmates and team members itself, have made valuable comment suggestions on this proposal which gave us an inspiration to improve our assignment. We thank all the people for their help directly and indirectly to complete our assignment.

**Problem Statement :**

To classify numerous reviews of any movie as a positive one or a negative one based on the content present in the review using Natural Language Processing along with various Machine learning techniques used to classify it.

**Reason/Motivation to choose the topic :**

In past years the number of movies releasing or filming every year increased exponentially with the course of time. Such movies include films based on various titles, stories, contents, topics, etc. With now a lot of movies as an option to the film enthusiasts and common public only selective ones really make an impact and are screened. Even after getting screened only movies with specific contents are accepted by the audience and so by the viewers. With this people face a lot of problem selecting the movies they can watch by spending three hours out of their precious work and personal times for entertainment.

Thus movie reviews became very important and reliable source of consultation for anyone who wants to watch any movie. But even after this stage of sorting out the specific movies still there are hundreds and thousands of reviews for a single movie given by common people as well as critics. So we created a tool to find out overall review of a movie thus saving people who seek entertainment their precious time.

**Software Requirement :**

Domain- Natural Language Processing , Machine Learning , Data Analytics

Back end- Python NLTS, Python Sklearn

Operating system- Windows 10

**Methodology :**

First we intend use dataset of a review and analyse the words used in it, after that using techniques of Natural Language Processing we remove all the stopwords which are common through NLTS, these include those words which do not affect the state of a sentence actively or directly and thus can be ignored.

Then the resulting data is fed in a Naïve Bayes Classifier which uses Bayes Theorem to predict the nature of a review either good or bad hence determining whether movie is worth watching or not.

We also compared the results of Naïve Bayes with some other machine learning techniques based on accuracy achieved.

**Conclusion :**

Thus we applied the knowledge of Natural language processing with machine learning techniques to predict whether a movie is worth watching or not based on it’s user and critic review