**Uber Text Reviews Analysis**

Author: HARSH NATVARLAL SINDHWA, ATHARVA KOUSADIKAR, APOORVA ANUMALLA, VAMSHI REDDY KOMATIREDDY, RAJANI KUMARI JANGITI

NORTHEASTERN UNIVERSITY

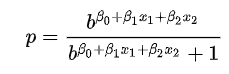


**Uber Text Reviews Analysis**

Customer satisfaction and their feedback is the most important thing that matters in any kind of business. Every time you take a ride on Uber, you are asked to rate the ride and the driver’s performance at the end of the ride. It is well known that why these ratings are important to the Company. Reviews and Ratings are a great way to enhance the quality of their drivers and to develop the company with a good reputation. Also, they get an opportunity to understand what they are lacking in. Through this project, we began to analyze the text reviews given by Uber customers as a feedback and have categorized them with the help of Bag-of-words. Later, the Logistic Regression model is added to the modified data and has been evaluated.

It is easy to analyze the ratings, as they are discrete data and can be manipulated using mathematical operations. The actual task is to decode the text reviews where customers write their feedback in words with complaints, complements and many more with respect to their uber ride and this feedback serve for the changes expected from customers.

**Technologies**: PySpark and using logistic regression.



**Benefits**: To Save time and effort and to get a solution with high performance, sentimental analysis has been used here as the algorithm is fully automated. It extracts important information regarding the emotions and attitudes, understanding how the customers liked their ride. The analysis gets more accurate and smarter when trained with more data.

**Drawbacks:** In the code, unigrams have been used for the analysis. Using bigrams or trigrams would have added for more accurate results.

**Challenges:** After getting used to the SMS language, many people use shortcuts instead of the full words. For example, people use “U” for representing “You”. Cleaning and sorting such kind of text to their original format is a difficult task.

**Installation**: We have used databricks spark cluster to do the project.