IBM Advanced Data Science Capstone Project

Harsh Vardhan Singh

Stakeholders' Presentation

Contents

Sentiment Analysis of Amazon Customer Reviews

Introduction to the dataset

- This project involves performing **Sentiment Analysis** of **Amazon Customer Reviews** using -
  - Natural Language Processing
  - Supervised Deep Learning
- The data is sourced from a publicly shared dataset on **Kaggle**
- It consists of 3 million customer reviews that will be used for training and another 650 thousand reviews for testing
- The datasets consist of 3 columns -
  - Customer rating *(Scale of 1 - 5)*
  - Review heading *(Unstructured text)*
  - Review text *(Unstructured text)*
- For the purpose of this project, we will consider reviews with **Ratings < 3** to be **Negative** while **Ratings > 3** will be marked as **Positive**
- Thus, we will try to predict the sentiment of any customer review as Negative or Positive

Business use case

- In today's day and age, e-commerce is an integral part of the consumer industry
- The global ecommerce market is estimated to have reached **US$ 4.3 trillion** as of 2021 and online shopping is one of the most popular online activities worldwide.
- Any e-commerce business would benefit from **analyzing the textual reviews** posted by customers on their platform.
    - *Sentiment analysis of the reviews can uncover **valuable information about the quality** of products*
    - *It can also help in exploring the **shortcomings and/ or defects** in particular products*
    - *It can also enable the business to better understand their **customers' overall experience** and the feedback received*

Analysis and results

- We successfully conducted a thorough analysis of the Amazon Customer Reviews data
    - Using various techniques in Natural Language Processing, we were able to extract and quantify textual data from the customer reviews
    - We then developed advanced deep machine learning algorithms to predict the sentiment of any given customer review

Our models achieved an accuracy of ***92%*** *in correctly predicting the review sentiment of the test data*

Word cloud representing the most common words in the review texts >

Proposed data product

- Given the success of the first stage of our data analysis of Amazon Customer Reviews, we propose the following **next steps** -
  - **Deploy the model on a live stream** of customer review data to analyze customer feedback in real time
    - With real time sentiment analysis, we can make **better business decisions** regarding product offerings
    - We can also **proactively address service-related issues** faced by customers
  - Deploy **resources to improve the prediction models** by reviewing some of the modeling stages which were skipped due to resource constraints (e.g. hyperparameter tuning)