

Predicting Business Owners Response on Social Media Using Data Science Approach

Faculty Supervisor: Dr. Naveen Kumar

Report By

Purushotham Vadde
purushotham.v@ou.edu

113459984

Spring 2021

Credit hours:4

Table of Contents

1. Introduction-----	3
2. Objective-----	3
2.1 Technical Project Objectives	
2.2 Individual Learning Objectives	
3. Plan-----	4
3.1 Data Collection	
3.2 Data Cleansing	
3.3 Exploratory Data Analysis	
3.4 Feature Engineering	
3.5 Modeling and Hyper Parameters Tuning	
3.6 Model Evaluation	
4. Deliverables-----	5
5. Schedule-----	6

1. Introduction

93% of customers will read reviews of local businesses to determine their quality(**BrightLocal**). Social media is filled with a lot of reviews there are thousands of reviews available for each business on social media and we can also see the responses from the business owners for the reviews given by the public. In this project, we will develop a Data science approach to determine what causes business owners to respond to a comment on social media and how likely is a business owner to respond to a comment about their firm.

The outcome of this project will help in providing the key insights about the motivations and actions of a business owner who respond to online reviews for their firms. Developing a machine learning-based model and researching the root causes of engagement by owners will also help to expose other interesting insights into the use of social media by the firms.

2. Objective:

2.1 Technical Project Objectives:

This project aims to build and implement a supervised machine learning model to classify the business reviews on the Yelp website, we will classify the public reviews for which the business owner going to respond.

2.2 Individual Learning Objectives

By this project, we will be developing skills in the field of Natural Language Processing, Machine learning, and Deep learning using packages such as NLTK, sci-kit learn, TensorFlow.

The project will also provide me with skills such as:

- Learning Machine learning and Deep Learning Algorithms for classification of data.
- Hyperparameter Tuning to find the optimal parameter for Machine learning Algorithms.
- Learning the Data Visualization skills.
- Storytelling skills from the insights got from the project.

3. Plan

3.1 Data Collection:

In this step, we are going to collect the business reviews by the public and Business owner and details related to restaurants from the Yelp websites for a city (Seattle, WA).

3.2 Data Cleansing:

The data which we collected in the previous stage will be processed and cleansed here, During this phase, the data will go through multiple data cleaning techniques in NLP such as Punctuation, stop words, lemmatization, stemming, Parts of Speech tagging, and Tokenization. Using these techniques, we will be able to remove unwanted text.

3.3 Exploratory Data Analysis

In this step we will analyze the dataset to summarize the main characteristics with visualization methods, apart from that this step(EDA) also helps in finding important features, detect outliers and anomalies;

3.4 Feature Engineering

In this step the Text data is converted into a numerical format called Vectors using different word embedding techniques, we also use different feature engineering techniques such as Feature scaling, encoding, and others to improve the Model's accuracy.

3.5 Modeling and Hyper Parameters Tuning

In this step we will build the Machine learning Model to classify the types of reviews for which the Business owners are likely to respond by using the Classification Algorithms, also perform the hyperparameter tuning to obtain the optimal parameters for the learning algorithm.

3.6 Model Evaluation

In this step, we will evaluate the different model's performance using evaluation metrics and find the best Model with better performance results. We use the evaluation metrics such as Accuracy, AUC curve, Confusion Matrix, etc...

4. Deliverables

Design a Model that will classify the reviews that are likely to get review comment by the Business owner using Machine learning and Deep Learning Algorithms

5.Schedule

No	Task	Expected Completion Data
1	Data Collection From Yelp	2/12/2021
2	Project Repository setup	2/19/2021
3	Data Cleaning	2/26/2021
4	Present initial collected data	2/26/2021
5	Exploratory Data Analysis	3/5/2021
6	Review	3/5/2021
7	Feature Engineering	3/12/2021
8	Understanding the Deep Learning Algorithms	3/19/2021
9	Modeling and Performance Tuning	4/9/2021
10	Code Review	4/9/2021
11	Model Performance Evaluation	4/16/2021
12	Building Dashboard	4/23/2021
13	Presentation and Final Report	4/30/2021