

Data Foundation IS 6713

Final Project Proposal

Maryam Khadivian\_ues970

Amazon Alexa Reviews

Dr. Gianluca Zanella

Spring 2021

## **Amazon Alexa Reviews**

1) What you want to do?

My main goal of this project is a review analysis on a product known as Amazon's "Alexa". The goal of my analysis is to compare the rating reviews of the product amongst its vast customer base along with the pro's and con's associated with the product. The analysis will also provide accurate feedback to the most popular models that the "Alexa" feature is used and the type of consumer base it attracts in a marketing aspect.

2) Why should we care?

By using this data, we can analyze the success rate of Amazon's Alexa product. The feedback from the information gathered utilizing customer reviews is essential in any company in order to monitor the success of the product or also the downfall of a product amongst its consumer base.

3) Where will your data come from?

I searched through the <https://www.kaggle.com/datasets> website. Then specifically under <https://www.kaggle.com/sid321axn/amazon-alexa-reviews> and, I also extracted the most current information from Amazon's own website based on this product. This dataset consists of a nearly 3000 Amazon customer reviews (input text), star ratings, date of review, variant and feedback of various Amazon Alexa products like Alexa Echo, Echo dots, Alexa Firesticks etc.

4) How do you want to process and analyze the data?

The first step will require applying pre-processing steps in order to transform the textual words into numeric features that can then be used utilizing machine learning algorithm. I will be exploring the text pre-processing techniques in Python using the NLTK (Natural Language Toolkit) library. Once this process is complete, I will be using the algorithm to perform a Sentiment Analysis on the dataset.

5) List of max 6 Keywords – To mark topic and domain of the idea.

1. Amazon-Alexa
2. Customer review
3. Analyze the data
4. Predict the model
5. Machine learning
6. Alexa NLP
7. Business

6) A subset of the 9 Heilmair questions:

1. Why is it hard?

The review feedback from customers or consumers is challenging to interpret due to the fact they are layered in human emotions and feelings. Due to this the feedback is almost neutral as the comments are both negative and positive in the same sentences.

For instance:

- I do not dislike my Alexa.
- I would really truly love using my Alexa while I'm not home.

So, I believe sorting all these sentiments in to a correct classification is the hardest part of doing this project. A misunderstanding of the customer reviews would lead to many areas of the provided data to be sorted into wrong classifications thus giving us a false and inaccurate result.

2. Who cares?

The data analysis conducted on Amazon's "Alexa" program will provide essential and valuable information to the Corporate brand and marketing departments.

3. If you're successful, what difference will it make?

In aspirations of a successful project, the information gained will be up to date in accurate format. This information will provide real insight to what's making the product a success and also pointing out its failures. This analysis would provide the information in ways of redirecting marketing for bigger sales and also ways to improve the product with its customers in a first-person response.