**DATA MINING AND PREDICTIVE ANALYSIS LAB**

**(PROJECT SYNOPSIS)**

**FISH THE BEST PHONE FROM A SEA OF VARIETY**

**Abstract:**

* Our aim is to select the best phone from a vast array of different phones available in the market (of comparable budget ranges) specific to each brand based on customer reviews and value for money.
* Clean the unstructured data and pre-processing for further analysis
* Analyse customer reviews using NLP
* Feed the data to a machine learning model which provides a recommendation and helps the user select the best suited phone for themselves which falls in their comfortable budget range.

**Dataset description:**

This data has information about the mobile phones which are present on the amazon. This data-set is scraped from the amazon.com. Below gives a description of each attribute in the dataset:

1. Product\_name: Title of product
2. by\_info: Company name
3. Product\_url: URL of product from amazon
4. Product\_img: URL of image
5. Product\_price: Price of product in INR
6. rating: 5 star rating
7. total\_review: Customer review number
8. ans\_ask: Customer asked for the answer
9. prod\_des: Short description of product
10. feature: complete description of product and its features
11. cust\_review: review of the customer

**Methodology:**

Rapid evolution in technology and internet brought us to the era of ecommerce. Many customers share their good or bad opinions about a product online. These opinions become a part of the decision making process to make impact on a business model.

Our approach to the project will be as follows:

1. Data extraction

This stage comprises of downloading the most suitable dataset with maximum attributes for our analysis.

1. Data cleaning

Data cleaning is the process of preparing datafor analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted.

In our project, we will truncate non numeric characters from an alphanumeric attribute to convert it to a numeric attribute. Next, we will perform removal and/or replacement of null values as per requirement.

1. Data pre-processing
2. Analysing customer reviews using NLP
3. Encoding categorical features:

This will be performed using label encoding and get dummies

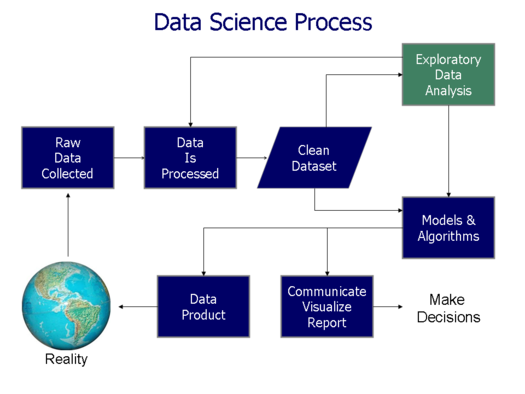
1. Using a machine learning algorithm to help the user choose the best phone

**Language used:**

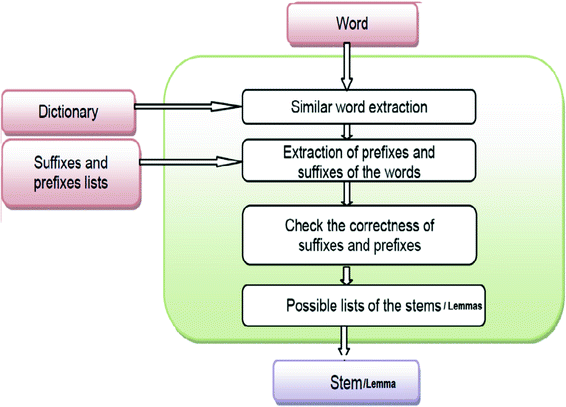
Python language for the entire coding of the project.

**Flow Diagrams:**

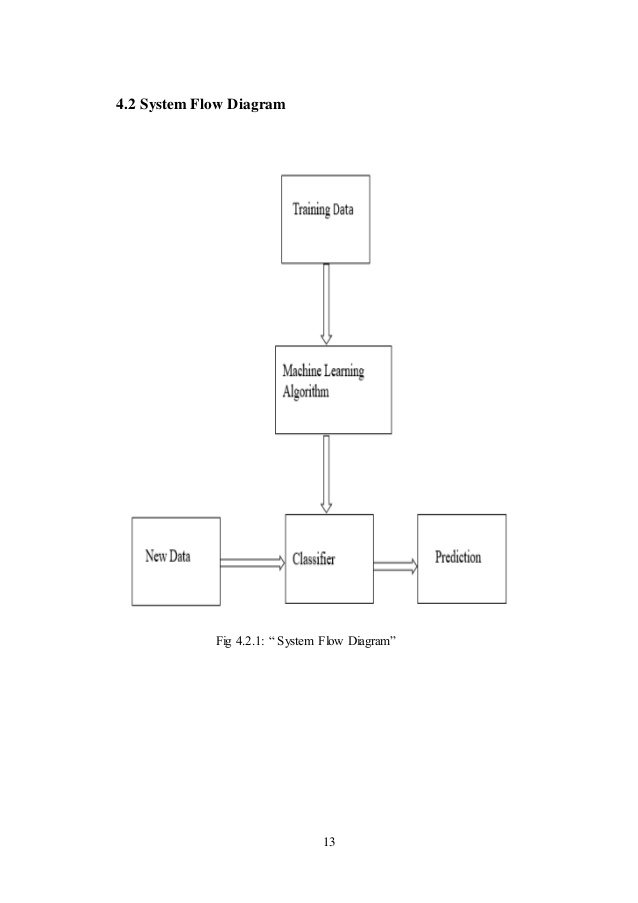
1. General flow of the project:



1. For NLP



1. For ML



Submitted by:

|  |  |  |
| --- | --- | --- |
| Name | Registration Number | Roll Number |
| Anvita Sharma | 170953264 | 64 |
| Meena Sirisha U. | 170953268 | 65 |
| Gunveen Batra | 170953270 | 66 |