PROJECT SYNOPSIS

ON

Review Analysis

INDEX & TABLES

1. Project Description
2. Modules
3. Data flow diagram
4. ER-Diagram
5. Language/Libraries/Tools

DESCRIPTION OF PROJECT

Introduction:

Sentiment Analysis refers to the practice of applying Natural Language Processing and Text Analysis techniques to identify and extract subjective information from a piece of text.

Sentiment Analysis is the detection of attitudes: “enduring, affectively colored beliefs, dispositions towards objects or persons”.

Problem Statement:

The process of summing up all the reviews received and then putting an abstract of all the data present is very hectic for a human. Thus, to overcome this problem, a software is made in order to get this job done efficiently.

Problem Identification:

#Customer reviews from customers who have actually purchased and used the product in question can give you more context to the product itself.

#Each reviewer rates the product from 1 to 5 stars, an provides a text summary of their experiences and opinions about the product.

#Sometime customers given ratings differ with their comments.

Objective:

->NRC emotion lexicon is used that categorizes customers’ reviews into eight emotions:

Anger Anticipation Disgust

Fear Sadness Joy

Trust Surprise

Methods:

The methods used are defined below:

NB classifier is used that computes a set of potentialities by counting the frequency and amalgam of valuation in a given dataset. It is very functional to codify emails appropriately. The fidelity and recall of this method is cognized to be very potent.

Text feature extraction method that is based on VSM(Vector Space Model),and it uses a set of predefined keywords. It is a process of taking out a list of words from the text data and then transforming them into a feature set which is usable by a classifier.

2. MODULES:

1. **Login**:

In this module, user will enter the User Id and Password. It is then checked, and only valid user id and password will get entry into search area. This is a security feature in order to avoid entry of unauthorized users.

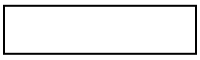
2. Input and Prediction:

Through this window, user enters a review message and gets the predicted result as output.

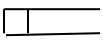
3. DATA FLOW DIAGRAM

DFD:-

The Data flow Diagram shows the flow of data. It is generally made of symbols given below :

(1) A **square** shows the Entity : - 

(2) A **Circle** shows the Process: - 

(3) An **open Ended Rectangle** shows the data store :- 

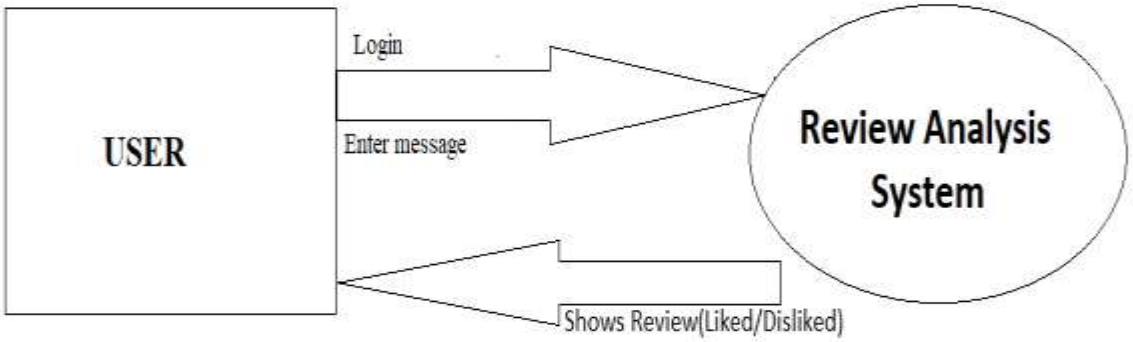
(4) An **arrow** shows the data flow :-

The DFD can be up to several levels. The 0 level DFD states the flow of data in the system as seen from the outward in each module.

The first level DFD show more detail, about the single process of the 0 level DFD

The second level DFD can show even more details and so on.

Context Level DFD



4. E-R Diagram

Definition:

An entity-relationship (ER) diagram is a specialized graphic that illustrates the interrelationships between entities in a database. ER diagrams often use symbols to represent three different

types of information. Boxes are commonly used to represent entities. Diamonds are normally used to represent relationships and ovals are used to represent attributes.

Entity Relationship (ER) diagram:

This diagramming technique is used to visually present a database schema or data model and was original proposed by Chen in the 1970s. There are many different data modeling notations; some are very similar to UML class diagrams (with the exception of operations). However, the notation the used here is slightly different, as proposed by Elmasri, et al. The database schema for this system is shown in figure. The table object has been left out of the diagram because the table management feature set had been dropped from the requirements before this stage of the design process.

Some important database design decisions are as follows:
_ To store the total price of an order with the order rather than calculating it on the fly when looking at past orders. This is because the price of menu items could change at any time, so the total price at the time of ordering must be stored so that the total price is not incorrectly calculated in future.

_ Similar to the previous point, the order receipt is stored as a hard-copy and not regenerated when reviewing past orders because things such as the restaurant name or VAT percentage are subject to change. Receipts stored need to be exactly the same as the customer copy in case of dispute.

Note: In this project, we have not used any database table since it is related to analysis and we read data from docx files which are unstructured so there is no need of ERD .

5. Language/Libraries/Tools

Front End :

Python tkinter

Back End :

Python and Data Science

Libraries :

- ✓ tkinter
- ✓ pandas
- ✓ String
- ✓ PIL
- ✓ re
- ✓ sklearn
- ✓ pyinstaller

Other S/W :

- ✓ Python3.x
- ✓ IDLE
- ✓ Anaconda (jupyter lab)