Project Proposal

On Sentiment Analysis

Darshan MR (1NT19CS063), JVSS Pavan Kumar (1NT19CS087),

Saiel K Gaonkar(111NT19CS166), Vishesh KR (1NT19CS217)

Introduction:

Every day we come across various products in our lives, on the digital medium we swipe across hundreds of product choices under one category. It will be tedious for the customer to make selection. Here comes 'reviews' where customers who have already got that product leave a rating after using them and brief their experience by giving reviews. As we know ratings can be easily sorted and judged whether a product is good or bad. But when it comes to sentence reviews, we need to read through every line to make sure the review conveys a positive or negative sense.

Sentiment Analysis is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative our neutral. Understanding people's emotions is essential for businesses since customers can express their thoughts and feelings more openly than ever before. It is quite hard for a human to go through each single line and identify the emotion being the user experience. Now with technology, we can automatically analyze customer feedback, from survey responses to social media conversations, brands are able to listen attentively to their customers, and tailor products and services to meet their needs.

Python sentiment analysis is a methodology for analyzing a piece of text to discover the sentiment hidden within it. It accomplishes this by combining machine learning and natural language processing (NLP). Sentiment analysis allows you to examine the feelings expressed in a piece of text.

Data Mining Task:

- **Data preprocessing:** Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.
- **Data Selection:** It is defined as the process of determining the appropriate data type and source, as well as suitable instruments to collect data. Data selection precedes the actual practice of data collection.

Data set:

We have taken the dataset of people who have taken journey for a particular airline. This file has reviewer ID, User ID, Reviewer Name, Reviewer text, helpful, Summary (obtained from Reviewer text), Overall Rating on a scale 5, Review time

Methods And Models:

Normalization: Database normalization is the process of structuring a database, usually a relational database, in accordance with a series of so-called normal forms to reduce data redundancy and improve data integrity. In machine learning and data mining, data normalization is used to make model training less sensitive to feature scale. As a result, our model can converge to better weights, resulting in a more accurate model. It is generally useful for classification algorithms.

Associative Rule Mining: Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently an itemset occurs in a transaction.

Model Building: The resultant output of the normalization process will be taken as the input for the model building. Since the primary goal of the project is to predict an attribute., we will have to apply one of the regression algorithms to obtain the desired result. First select the best performing model by using cross validation. Let's consider all the classification algorithm and perform the model selection process.

Assessments:

As a classification problem, Sentiment Analysis uses the evaluation metrics of Precision, Recall, F-score, and Accuracy. Also, average measures like macro, micro, and weighted F1-scores are useful for multi-class problems

Presentation and Visualization:

We have visualized the dataset in the form of pie-chart and bar-graph for better understanding of dataset.

Roles:

Darshan MR: Drafting the project proposal, gathering dataset.

JVSS Pavan Kumar: Focusing on dividing the dataset for training & testing using Regular expression(RE), visualizing the dataset into graphs.

Saiel Gaonkar: Train the dataset, apply Machine learning(ML) algorithm Random Forest classifier to train dataset.

Vishesh KR*:* Making Predictions and Evaluating the Model, drafting report.

Schedule:

06/01/2022	Data Preprocessing
09/01/2022	Loading dataset and performing data mining tasks
12/01/2022	Data Visualization
15/01/2022	Model Evaluation
17/01/2022	Project Report

Bibliography:

https://techvidvan.com/tutorials/python-sentiment-analysis/

https://www.analyticsvidhya.com/blog/2021/06/rule-based-sentiment-analysis-in-python/

https://www.digitalocean.com/community/tutorials/how-to-perform-sentiment-analysis-in-python-3-using-the-natural-language-toolkit-nltk

https://www.kaggle.com/benroshan/sentiment-analysis-amazon-reviews#Table-of-Contents

https://www.kaggle.com/sid321axn/natural-language-processing-sentiment-analysis