ATSS's

Institute of Industrial & Computer Management &

Research, Nigdi Academic Year 2020-2021

MCA II - SEM-1

Mini Project Synopsis

Title of the Project: Movies reviews Analysis using Machine Learning

Details of Team Members

Name of Team Member 1	**Rwittick Sadhu**
Roll Number	**43**
Contact Number	**8906118059**
E-mail ID	**sadhurwittick@gmail.com**
Name of Team Member 2	**Chetan Kailas Chaudhari**
Roll Number	**75**
Contact Number	**7262083845**
E-mail ID	**cc3862558@gmail.com**

Drawbacks of Existing System:

- Survivor usually provide a feedback form & enter manually all the details. And analyze one by one.
- That is time consuming, hectic, error could prone
- data could be lost.
- Not owner-friendly environment.
- Accident easily occur either burning of papers or water damage.

Need of Proposed System:

- The aim of this project is to build a Movies reviews analysis model which will allow us to categorize words based on their sentiments, that is whether they are positive, negative and also the magnitude of it.

- This project could show a path to reduce customer churn.

Scope of the proposed System:

- Movies reviews Analysis is a process of extracting opinions that have different polarities. By polarities, we mean positive, negative or neutral. It is also known as opinion mining and polarity detection.

- With the help of sentiment analysis, you can find out the nature of opinion that is reflected in documents, websites, social media feed, etc.

- Movies reviews Analysis is a type of classification where the data is classified into different classes. These classes can be binary in nature (positive or negative) or, they can have multiple classes (happy, sad, angry, etc.).

- It will be a great project to understand how to perform Movies reviews Analysis and it is widely being used nowadays.

- This will be one of the most popular machine learning projects. The reason behind this is every company is trying to understand the sentiment of their customers if customers are happy, they will stay.

- Easy to operate and maintain .

.

.

Modules Description:

In order to build our project on Movies reviews Analysis, we will make use of the tidytext package that comprises of sentiment lexicons that are present in the dataset of 'sentiments'.

Syntax:

```
library(tidytext)
Sentiments
```

We will make use of three general purpose lexicons like –

- **AFINN**
- **bing**
- **loughran**

These three lexicons make use of the unigrams. Unigrams are a type of n-gram model that consists of a sequence of 1 item, that is, a word collected from a given textual data.

In the AFINN lexicon model scores the words in a range from -5 to 5. The increase in negativity corresponds the negative sentiment whereas an increase in positivity corresponds the positive one.

The bing lexicon model on the other hand, classifies the sentiment into a binary category of negative or positive.

And finally, the loughran model that performs analysis of the shareholder's reports. In this project, we will make use of the bing lexicons to extract the sentiments out of our data. We can retrieve these lexicons using the get_sentiments() function. We can implement this as follows –

Syntax:

```
get_sentiments("bing")
```

- **Performing Movies reviews Analysis with the Inner Join**

In this step, we will import our libraries ‘janeaustenr’, ‘stringr’ as well as ‘tidytext’. The janeaustenr package will provide us with the textual data in the form of books authored by the novelist Jane Austen. Tidy text will allow us to perform efficient text analysis on our data. We will convert the text of our books into a tidy format using unnest_tokens() function.

We have performed the tidy operation on our text such that each row contains a single word. We will now make use of the “bing” lexicon to and implement filter() over the words that correspond to joy. We will use the book Sense and Sensibility and derive its words to implement out sentiment analysis model.

In the last step, we will visualize the words based on their corresponding positive and negative scores.

Beneficiary:

The project is identified by the merits of the system offered to the Users. The merits of this project are as follows: -

- This project offers user to enter the data through simple and interactive forms. This is very helpful for the client to enter the desired information through so much simplicity
- Decision making process would be greatly enhanced because of faster processing of information since data collection from information available on computer takes much less time then manual system
- Easier and faster data transfer through latest technology associated with the computer and communication
- Through these features it will increase the efficiency, accuracy and transparency

Technology Description:

Front End Language	HTML, AJAX, JQUERY, JAVASCRIPT, CSS
Back End Language	Python Machine Learning
Library	
Tools	Jupiter, Visual Studio Code.
Framework	

Hardware-Software Requirements :

Hardware Requirement		
Processor	**RAM**	**Free Space**
Intel(R) Pentium(R) CPU N3700 @ 1.60 GHz	4 GB or Above	500MB or Above
Software Requirements		
Operating System	**Application/Software**	
64-bit Operating System, x64 based processor	Web Browser: Mozilla, Google Chrome,IE8, etc.	
Server-Side Requirements		
Operating System	**Processor**	**Storage Space**
Win-7, Win-10, Linux or any other higher version	Intel core i3 or i5	Ram 4GB or more and 5GB or More Free Space

Expected GUI Screens:

- Statement Entering
- Analyzed Text Statement
- Percentage.

Future Enhancement:

Signature of the Student:

Rwittick Sadhu

Chetan Chaudhari

Signature of the Project Coordinator :