INTERNSHIP: PROJECT REPORT

Internship Project Title	TCS iON RIO-45: Automate detection of different sentiments from textual comments and feedback.
Project Title	TCS iON RIO-45
Name of the Company	TCS ION
Name of the Industry Mentor	Debashis Roy
Name of the Institute	PrepInsta Technologies Pvt. Ltd.

Start Date	End Date	Total Effort (hrs.)	Project Environment	Tools used
1/6/2022	13/6/2022	45	GoogleCollab,Chrome,Windows10	Python3

Project Synopsis

Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. The applications of sentiment analysis are broad and powerful. The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world. It can also be an essential part of your market research and customer service approach. Not only can you see what people think of your own products or services, you can see what they think about your competitors too. The overall customer experience of your users can be revealed quickly with sentiment analysis, but it can get far more granular too. It can also be used to categorize feedback in movies, products etc.. Hence, developing a tool for automating Sentiment Analysis can be very beneficial for various industries.. The applications of sentiment analysis are broad and powerful. Sentiment analysis can do the following-

1. Choose the model.
2. Choose the classifier
3. Import the data. You can import data from an app or upload a CSV or Excel file
4. Set the conditions for positive and negative analysis to train your sentiment analysis classifier.

5. Test your classifier. Once the model has been trained with some examples, you can paste your own text to see how they're classified
6. Once your model is trained, you can upload any amount of data.

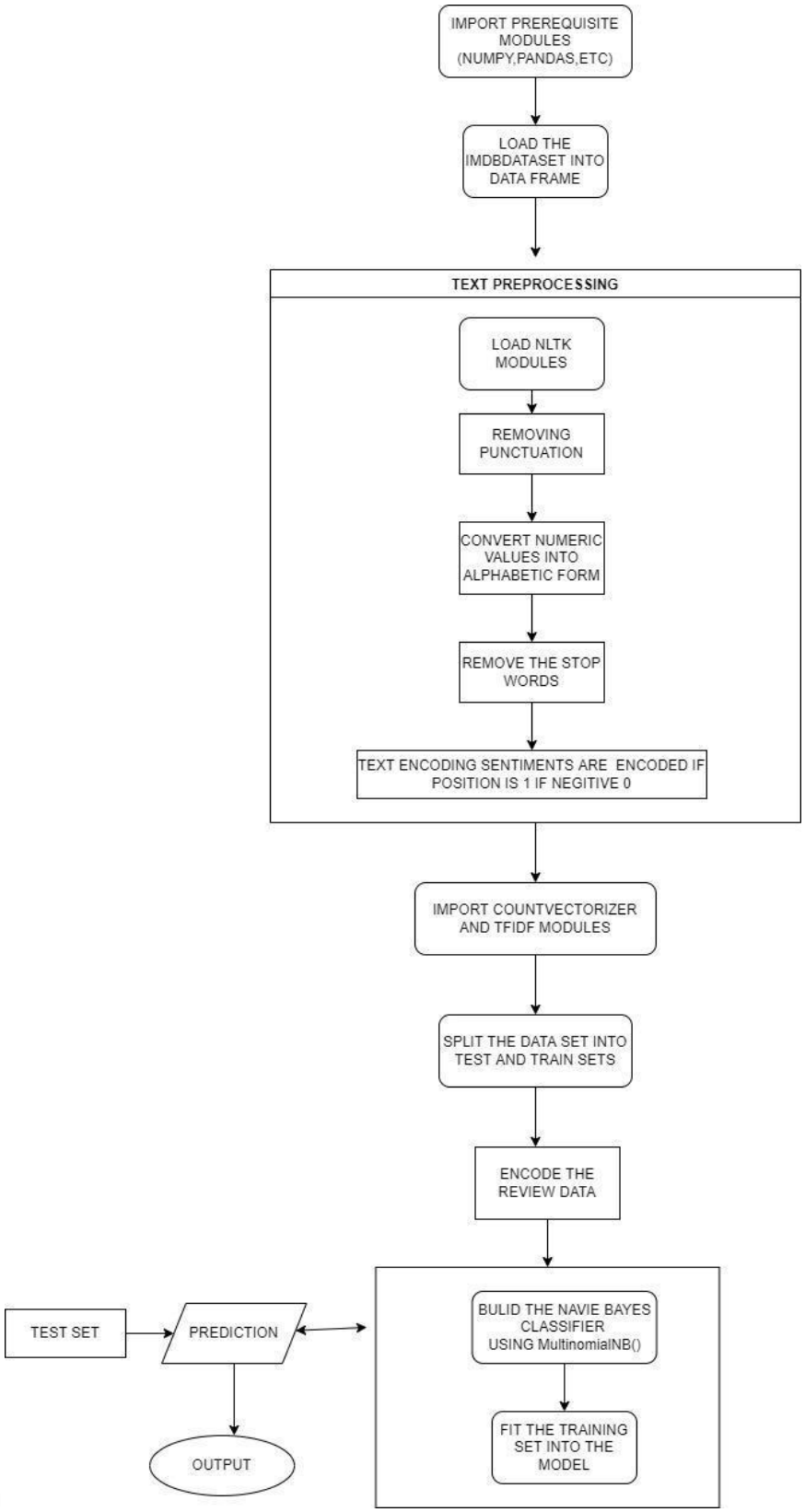
Solution Approach

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 of a movie review can rate how positive or negative a movie review is and hence the overall rating for a movie. Therefore, the process of understanding if a review is positive or negative can be automated as the machine learns through training and testing the data.
Classification Algorithms such as Naïve Bayesian classifier Multinomial NB can be used for the task. This model is used to perform sentiment analysis on movie reviews from the IMDB dataset.
The CountVectorizer can be used to break down the sentiments into numerical values.
On supplying a test set to the model , on the outcome it will try to predict whether it is a positive or negative feedback.

Assumptions:

IMDB dataset that contains the text of 50,000 movie reviews from the Internet Movie Database.
These are split into 25,000 reviews for training and 25,000 reviews for testing. The training and
testing sets are balanced, meaning they contain an equal number of positive and negative reviews So,
its a binary or two-class-classification without a 'neutral' label.We Use the MultinomialNB as the classifier because the The multinomial Naive Bayes classifier is suitable for classification with discrete features
Here we aim to predict whether a record is a 1 or a 0 such as Positive and Negative .

Project Diagrams



u

THE ABOVE IS THE PROJECT WORKFLOW DIAGRAM

Algorithms:

Naïve Bayesian classifier:

TheNaïve Bayesian classifier works as follows: Suppose that there exist a set of training data,D, in which each tuple is represented by ann-dimensional feature vector,X=x1,x2,..,xn, indicating measurements made on the tuple from attributes or features. Assume that there are classes,C1,C2,...,Cm. Given a tupleX, the classifier will predict that belongs to if and only if:P(Ci|X)>P(Cj|X), where i,j∈[1,m]andi≠j.P(Ci|X) is computed as:

$$P(C_i|X) = \prod_{k=1}^{n} P(x_k|C_i)$$

Outcome:

"I really enjoyed this film. Think of a cross between Go, Fear and Loathing in Las Vegas and Two Hands. The action didn't let up and the characters were all given life by a strong cast of young and old hands. I enjoyed the chaos and the messiness, and I think the low-budget feel of the film adds to rather than detracts from its appeal. For those of you who get the film out on the basis of Kylie being on the cover, you'll be disappointed. Although she plays a strong character, screen-time is shared pretty evenly throughout the cast. She's back to her grass roots here. Standout performers for me were Simon Lyndon, Paula Arundell, Nathan Page and Matt Wilkinson but the whole cast's effort was the highlight of this film."

POSITIVE

"What a nasty cynical film. Apparently this sad excuse for a dramatic urban look at what 20 year olds do whilst crawling through the gutter of Sydney nightlife is supposed to be somehow connected with its target market. Made by some Industry nobody and pals who seemingly thought they could cobble together any sleazy behavior with a young cast and pour it into multiplexes, SAMPLE PEOPLE deservedly failed miserably at the Australian box office. It is so offensive in its clichéd depictions of obvious and easy targets it was fully rejected by the very audience it was intended. Shoddy and cruel and with no attempt to offer quality or resonance to the young audience who might have been attracted by the marketing or casting SAMPLE PEOPLE might have been interesting or even informative if not botched by its exploitive view of 'what teens want to see in a movie'. The character played by Ben Mendelsohn is particularly offensive and Kylie Minogue is again wasted by poor material and untalented film makers. It is as if the producers thought teens would watch any ugly trash and just slung-together scenes and characters who were shallow and soulless. Well they were very wrong. A mini budget film made in 1983 called GOING DOWN got this topic right and is an excellent antidote to this poison."

NEGATIVE

''Superhero comics, and much of their adaptations, have long taken an outsized, soap opera-like approach to storytelling. At their best, they can take these fantastical ideas and make them emotionally resonant, even if there's obviously no real-world phenomenon to connect them to. In some respects, Endgame pulls this off beautifully, like how the character Nebula confronts her past self through time travel, giving physical form to her personal growth. But as fun as the movie is, there's an undeniable hollowness at its core induced by its unwillingness to follow through on certain ideas and symbols''

POSITIVE/NEUTRAL

FINAL ACCURACY VALUES

AFTER TESTING THE ACCURACY SCORE OF THE MODEL ,FOUND THE MODEL ACCURACY IS UP TO 83 PERCENT AND THE MODEL MOSTLY FOUND OUT GIVING THE ACCURATE PREDICTIONS

```
Test accuracy
83.46000000000001
```

CLASSIFICATION REPORT

	precision	recall	f1-score	support
0	0.78	0.93	0.85	7411
1	0.92	0.74	0.82	7589
accuracy			0.83	15000
macro avg	0.85	0.84	0.83	15000
weighted avg	0.85	0.83	0.83	15000

Exceptions considered:

Accuracy Function values vary slightly on every compilation of the model. This may give a slight difference in the prediction value which may impact the neutral sentiments.

Enhancement Scope:

The Naive Bayes classifier model performance can be calculated by the hold-out method or cross-validation depending on the dataset
Increasing the accuracy is the approachable way of enhancement scope. This can be achieved by

using the clean dataset,increasing the training set batch that helps the model to train more and can give accurate results or Using a different model like LSTM or random forest can increase the accuracy.

Link to Code and executable file:

htt0