TWITTER SENTIMENT ANALYSIS

A Project Work

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

CSE-IBM BIG DATA ANALYTICS

Submitted by:

Vansh Sharma

17BCS3905

Under the Supervision of:

MRS. SUMAN SARKAR

APEX INSTITUTE OF TECHNOLOGY

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB

NOVEMBER& 2019

DECLARATION

I, '**MANISH GUPTA'**, student of **'Bachelor of Engineering in CSE-IBM BIG DATA ANALYTICS'**, **session: 2019 - 2020**, Apex Institute of Technology, Chandigarh University,Punjab, hereby declare that the work presented in this Project Work entitled '**TWITTER SENTIMENT ANALYSIS'** is the outcome of our own bona fide work and is correct to the best of ourknowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

<div align="right">

MANISH GUPTA
UID: 17BCS3921

</div>

Date:

Place:

CERTIFICATE

This is to certify that the work embodies in this dissertation entitled **'TWITTER SENTIMENT ANALYSIS'**being submitted by **MANISH GUPTA 17bcs3921**for partial fulfillment of the requirement for the award of **Bachelor of Engineering** in *CSE-IBM BIG DATA ANALYTICS*discipline to Apex Institute of Technology, Chandigarh University, Punjabduring the academic year 2019 - 2020 is a record of bonafide piece of work, undertaken by him/her the supervision of the undersigned.

Approved and Supervised by

Signature of Supervisor
(Mr. SUMAN SARKAR)

Professor , Ait-cse

Forwarded by

Signature of Project Coordinator
(MR.GURJOT SINGH SODHI)
Professor, Ait-cse

EXTERNAL EXAMINER

Signature of External Examiner
(External Examiner's Name)

ABSTRACT

Sentiment analysis is the automated process of analyzing text data and classifying opinions as *negative*, *positive* or *neutral*. Usually, besides identifying the opinion, these systems extract attributes of the expression e.g.:

- *Polarity*: if the speaker express a *positive* or *negative* opinion,
- *Subject*: the thing that is being talked about,
- *Opinion holder*: the person, or entity that expresses the opinion.

Currently, sentiment analysis is a topic of great interest and development since it has many practical applications. Companies use sentiment analysis to automatically analyze survey responses, product reviews, social media comments, and the like to get valuable insights about their brands, product, and services.

ACKNOWLEDGEMENT

I would like to take this opportunity to express my profound sense of gratitude and respect to all those who helped me throughout the duration of this project. The **'APEX INSTITUTE OF TECHNOLOGY, CHANDIGARH UNIVERSITY'** in particular has been the source of inspiration for me. I acknowledge the effort of those who have contributed significantly to my project.

I feel privileged to offer my sincere thanks and deep sense of gratitude to Mr.Gurjot Singh Sodhi , Project Coordinator (AIT-CU) and Mr.Suman Sarkar(Project Supervisor) for expressing their confidence in me by letting me work on a project of this magnitude and using the latest technologies and providing his support, help & encouragement in implementing this Project.

Table of Contents

1. INTRODUCTION :-

Sentiment analysis is the automated process of analyzing text data and classifying opinions as *negative*, *positive* or *neutral*. Usually, besides identifying the opinion, these systems extract attributes of the expression e.g.:

- *Polarity*: if the speaker express a *positive* or *negative* opinion,
- *Subject*: the thing that is being talked about,
- *Opinion holder*: the person, or entity that expresses the opinion.

Currently, sentiment analysis is a topic of great interest and development since it has many practical applications. Companies use sentiment analysis to automatically analyze survey responses, product reviews, social media comments, and the like to get valuable insights about their brands, product, and services.

1.1 PROBLEM DEFINITION:-

It's estimated that 80% of the world's data is unstructured and not organized in a pre-defined manner. Most of this comes from text data, like emails, support tickets, chats, social media, surveys, articles, and documents. These texts are usually difficult, time-consuming and expensive to analyze, understand, and sort through.

Sentiment analysis systems allows companies to make sense of this sea of unstructured text by automating business processes, getting actionable insights, and saving hours of manual data processing, in other words, by making teams more efficient.

1.2 :-<u>**PROJECT OVERVIEW/ PROJECT SPECIFICATION:-**</u>

This project is designed to give you a hands-on experience with natural language processing and work towards to develop a classification machine learning model on a real-world dataset. Sentiment analysis (also known as opinion mining) is one of the many applications of Natural Language Processing.

 It is a set of methods and techniques used for extracting subjective information from text or speech, such as opinions or attitudes. In simple terms, it involves classifying a piece of text as positive, negative or neutral.

1.3:- <u>HARDWARE SPECIFICATIONS:-</u>

Recommended System Requirements:

Processors:

Intel® Core™ i5 processor 4300M at 2.60 GHz or 2.59 GHz (1 socket, 2 cores, 2 threads per core), 8 GB of DRAM

   Intel® Xeon® processor E5-2698 v3 at 2.30 GHz (2 sockets, 16 cores each, 1 thread per core), 64 GB of DRAM

   Intel® Xeon Phi™ processor 7210 at 1.30 GHz (1 socket, 64 cores, 4 threads per core), 32 GB of DRAM, 16 GB of MCDRAM (flat mode enabled)

Disk space: 2 to 3 GB

Operating systems:

 Windows® 10, macOS*, and Linux*

Minimum System Requirements:

Processors:

Intel Atom® processor or Intel® Core™ i3 processor

   Disk space: 1 GB

Operating systems:

Windows* 7 or later, macOS, and Linux

   Python* versions: 2.7.X, 3.6.X

*Included development tools:*conda*, conda-env, Jupyter Notebook* (IPython)

1.4 SOFTWARE SPECIFICATION:-

Operating systems:

Windows* 7 or later, macOS, and Linux

   Python* versions: 2.7.X, 3.6.X

*Included development tools:*conda*, conda-env, Jupyter Notebook* (IPython)

1.4.1 Python:-

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. Van Rossum led the language community until stepping down as leader in July 2018.

Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative,

functional and procedural. It also has a comprehensive standard library.

Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of Python's other implementations. Python and CPython are managed by the non-profit Python Software Foundation

1.5 Environment:-

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

2. LITERATURE SURVEY :-

Sentiment analysis can be applied at different levels of scope:

- **Document level** sentiment analysis obtains the sentiment of a complete document or paragraph.
- **Sentence level** sentiment analysis obtains the sentiment of a single sentence.
- **Sub-sentence level** sentiment analysis obtains the sentiment of sub-expressions within a sentence.

It's estimated that 80% of the world's data is unstructured and not organized in a pre-defined manner. Most of this comes from text data, like emails, support tickets, chats, social media, surveys, articles, and documents. These texts are usually difficult, time-consuming and expensive to analyze, understand, and sort through.
Sentiment analysis systems allows companies to make sense of this sea of unstructured text by automating business processes, getting actionable insights, and saving hours of manual data processing, in other words, by making teams more efficient.

Some of the advantages of sentiment analysis include the following:

- **Scalability**:
Can you imagine manually sorting through thousands of tweets, customer support conversations, or customer reviews? There's just too much data to process manually. Sentiment analysis allows to process data at scale in a efficient and cost-effective way.

- **Real-time analysis**:
We can use sentiment analysis to identify critical information that allows situational awareness during specific scenarios in real-time. Is there a PR crisis in social media about to burst? An angry customer that is about to churn? A sentiment analysis system can help you immediately identify these kinds of situations and take action.

- **Consistent criteria**:
Humans don't observe clear criteria for evaluating the sentiment of a piece of text. It's estimated that different people only agree around 60-65% of the times when judging the sentiment for a particular piece of text. It's a subjective task which is heavily influenced by personal experiences, thoughts, and beliefs. By using a centralized sentiment analysis system, companies can apply the same criteria to all of their data. This helps to reduce errors and improve data consistency.

3.Methodology:-

The following methodology has been followed to achieve the objectives defined for proposed research work:

Supervised Learning: Supervised learning is the machine learning task of learning function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal).

Data Preparation: obtain dataset from internal and external sources.
Data consistency check in term of definition of fields , unit of measurements , time period etc.

Data Exploration and Conditioning: missing data handling ,outliers , graphical or visual analysis, Transformation ,Normalization, Partitioning into training , validation and test datasets
.
Model Planning : Determining data mining task such as prediction , classification etc.
Select appropriate data modeling method such as KNN and Logistics Regression.
Model Building: Building different candidate model using selected techniques and their variant using training data.

Evaluate the final model on test data.

Model Interpretation: Model evaluation using key performance matrices.

Model deployment: Pilot project to integrate and run the model on the operational system.

4.CONCLUSION :

This project was successfully completed within the time span allotted. The project TWITTER SENTIMENT ANALYSIS has been developed . All the modules are tested separately and put together to form the main system. Finally the system is tested with real data and everything workedsuccessfully. Thus the system has fulfilled the entire objective identified.The system had been developed in an attractive dialogs fashion. So user with minimumknowledge about computers can also operate the system easily. It will make easy interactions between users and store. The speed and accuracy are maintained in proper way.

5. REFERENCE & APPENDICES:

WEBSITES REFERRED:

The following links were searched and exploited extensively for the project development and implementation:

1. https://www.kaggle.com/c/twitter-sentiment-analysis2
2. https://www.udemy.com/machinelearning/
3. http://index-of.es/Varios-2/Hands%20on%20Machine%20Learning%20with%20Scikit%20Learn%20and%20Tensorflow.pdf
4. https://monkeylearn.com/sentiment-analysis/.
5. https://www.w3schools.com/python/
6. https://github.com/zfz/twitter_corpus