“SENTIMENT ANALYSIS IN SOCIAL MEDIA ON FACEBOOK”

THESIS SUBMITTED IN PARTIAL FULFILMENT OF REQUIREMENT

FOR THE AWARD OF THE DEGREE OF

Master of Technology

in

Software Engineering

Under the guidance of

Vinod Kumar

(Associate Professor

Computer Science and Engineering)

Delhi Technological University

Submitted By:

Zenawi Haileslassie Berhe

(Roll No. - 2K17/SWE/20)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahabad Daulatpur, Main Bawana Road, Delhi-110042

June 2019

M.TECH (SOFTWARE ENGINEERING)

ZENAWI HAILESLASSIE BERHE

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Rod, Delhi-110042

CANDIDATE'S DECLARATION

I Zenawi Haileslassie Berhe (2k17/SWE/20), student of M.Tech in Software Engineering, hereby declare that the thesis work entitled "**Sentiment Analysis in Social media on Facebook**" which is being submitted by me to the Department of Computer sciences and engineering, Delhi Technological University, Delhi in partial fulfilment of requirements for the award of degree of Master of Technology is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Association, fellowship or other similar title or recognition

Place: Delhi (Zenawi Haileslassie Berhe)

Date ....

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Rod, Delhi-110042

CERTIFICATE

I hereby certify that project Dissertation titled "**Sentiment Analysis in Social media on Facebook**" submitted by **Zenawi Haileslassie** (Roll No. 2K17/SWE/20) Department of computer sciences and engineering, Delhi Technological Engineering, Delhi in partial fulfilment of the requirement for the award of degree Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Signature ………………………..

Place: Delhi (associate prof." supervisor"Vinod Kumar)

Date …………………………

ACKNOWLEDGEMENT

First of all, I would like to express my deep sense of respect and gratitude to my project supervisor Associate Professor Vinod Kumar (Associate Professor, Computer Engineering Department) for providing the opportunity of carrying out this project and being the guiding force behind this work. I am deeply indebted to his for the support, advice and encouragement she provided without which the project could not have been a success. Secondly, I am grateful to Dr. Rajni Jindal, HOD, Department of Computer Science & Engineering, DTU for her immense support. I would also like to acknowledge Vinod Kumar (Associate Professor, Department of Computer Science & Engineering), Delhi Technological University library and staff for providing the right academic resources and environment for this work to be carried out. Last but not the least I would like to express sincere gratitude to my Family and friends for constantly encouraging me during the completion of work.

Signature ……………………

Place: Delhi Zenawi Haileslassie Berhe (M.Tech)

ABSTRACT

In this advanced stage technology where giving & devouring managements through online trade has turned into day to day activities (task), everybody wants to take an interest and give suppositions about the administrations he/she spends. these days are occurring substantially more on social areas as opposed to in a feedback box. Facebook is a standout amongst the most regularly utilized online networking areas where individuals voice their sentiments on attractive much everything. such a significant number of specialist organizations take the stage to advance their administrations among the clients. Media transmission section is one of them. It is very evident that numerous Telecommunication organization keeps up a Facebook page or gathering to raise their managements to the clients and get disapproval from them. Limitless is being delivered along these lines every day.

In this thesis, we will separate every one of the sentiments (remarks as content information) from each individual profile pages utilizing Facebook graph API is given by Facebook application & experience commotion scrubbing, applying sentiment & classifier for computing opinion extremity go to choice whether offers are given by organization is receiving great input or not. And in this thesis for our sentiment analysis I used naïve Bayes classifier techniques.

Key words: -Social network, Facebook graph API, Naïve Bayesian classifier, Sentiment Analysis, Facebook.

Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NLP:	Natural language processing
SVM:	Support vector machine
KNN:	K-nearest Neighbor
AI:	Artificial intelligence
API:	Application program interface
SA:	Sentiment analysis
URL:	Unified resource locator

CHAPTER ONE

INTRODUCTION

1.1 Introduction to sentiment Analysis

Sentiment analysis is a Computational investigation of sentiments, opinions, assessments, frames of mind, evaluation, influences, feelings, subjectivity, and so on., communicated in content. And sentiment Analysis additionally called Opinion Mining is a field inside Natural Language Processing (NLP) that fabricates systems that attempt to recognize just as to isolate assessments inside substance. Ordinarily, other than recognizing the assessment, these frameworks separate traits of the articulation for example.: -

- ✓ **Polarity:** on the off chance that the speaker communicates a positive or negative sentiment.
- ✓ **Subject:** what is being discussed,
- ✓ **Opinion holder:** the individual, or substance that communicates the opinions (supposition).

At present time sentiment analysis is a subject for incredible premium & advancement because it has numerous pragmatic applications. since openly & secretly accessible data over Internet is always growing, countless communicating feelings are accessible in survey destinations, discussions, online journals, and web-based life. With the help of opinion examination systems, this unstructured information could be consequently changed into sorted out data of open evaluations about things, administrations, brands, authoritative issues, or any subject that people can express emotions about. This data can be amazingly profitable for business applications like promoting examination, publicizing, thing overviews, net sponsor scoring, thing info, and customer organization.

What is Opinion mining?

Text data can be comprehensively arranged into two fundamental sorts: certainties & feelings. Realities are target articulations about something. opinions are generally subjective expressions that depict individuals' conclusions, examinations, and sentiments toward a subject or point.

Sentiment analysis, similarly the same number of other natural language processing issues, can be shown as a grouping issue where two sub-issues must be settled:

- ✓ Categorizing a sentence as objective or subjective, known as **subjectivity characterization**.

- ✓ Categorizing a sentence as communicating neutral, positive or negative assessment also known as **polarity characterization**.

In a sentiment, the substance for content discussions about the article, the parts, the angles, the qualities, or the highlights. It might likewise be an item, an administration, an individual, an association, an event, or a point. For instance, investigate the supposition text below:

"Battery life time for this camera is excessively short."

Negative conclusion is communicated about a component or life time of a substance (devices).

a) Comparative versus direct opinion

Two main categories of assessments or feelings in sentiment analysis are direct feelings and comparative feelings. Direct sentiments give an assessment about an element straightforwardly, for instance:

"The image nature of camera X is too poor."

This immediate or direct assessment expresses is shows negative feeling for devise X.

The second type of assessment is comparative or relative feelings, supposition is communicated by contrasting the elements and another, for instance see below:

"The image nature of camera X is superior to that of camera Y."

Typically, the relative opinion nonstop similarities or contrasts between at least two substances or devices utilizing a similar or superlative type of a descriptive word or verb modifier. In the past model, there is positive assessment for devices X & on the other hand, a negative conclusion about camera Y.

b) Implicit versus explicitly opinions

For explicit opinion regarding the matter is a sentiment explicitly communicated in a subjective text or content. the accompanying text or contents communicates an explicit positive supposition or sentiment as shown below.

"The voice nature of this telephone is stunning."

For implicit opinion regarding the matter for the sentiment inferred in an objective text or content. And the accompanying text or content communicates is Implicitly negative feeling:

"The headphone is breaking after two days."

Inside the implicit opinion we can incorporate analogies might be for the most troublesome sort of conclusions for investigating or incorporate great deal of semantic data.

1.2 Sentiment analysis scopes

Sentiment analysis can be connected in various dimension degrees or scopes but the most common scopes of sentiment analysis are three as listed below:

- ✓ **The first is document level:** document level sentiment analysis gets different opinions as a total report or passage(paragraph).
- ✓ **The second is sentence level:** sentence level sentiment analysis gets different opinions of solitary or single.
- ✓ **The third is sub-sentence level** sentiment analysis or opinion examination gets the slant of sub-expressions inside a sentence.

1.3. Problem of Statement

Currently, Sentiment analysis is widely used in politics, Business, advertisement, public actions and advancement since it has numerous practical applications. That is privately and publicly accessible data over Internet is continuously growing, huge number of writings text feelings or opinions are accessible in survey locales, gatherings, social media and online journals.

With the support (assistance) of assumption examination systems, this unstructured information could be naturally changed into organized or sorted out data of prominent sentiments about things, organizations, brands, governmental issues, or any subject that people can express evaluations about. This information is exceptionally helpful in business applications such as showcasing examination, advertising, item surveys, review products, item input, and client administration.

1.4. Objectives of the Project

In many organizations and associations, a client's impression of item or administration is very important data. from the information picked up from an examination, for example, this an organization can distinguish issues with their items, spot drifts before their challengers, make improved interchanges with their intended interest group, and again significant knowledge into how successful their advertising efforts were. through this information organizations increase important criticism which enables them to additionally build up the up and coming age of their item.

The main objective of this project is to classify a number of text files or sentiment analysis at different level of scopes as follows:

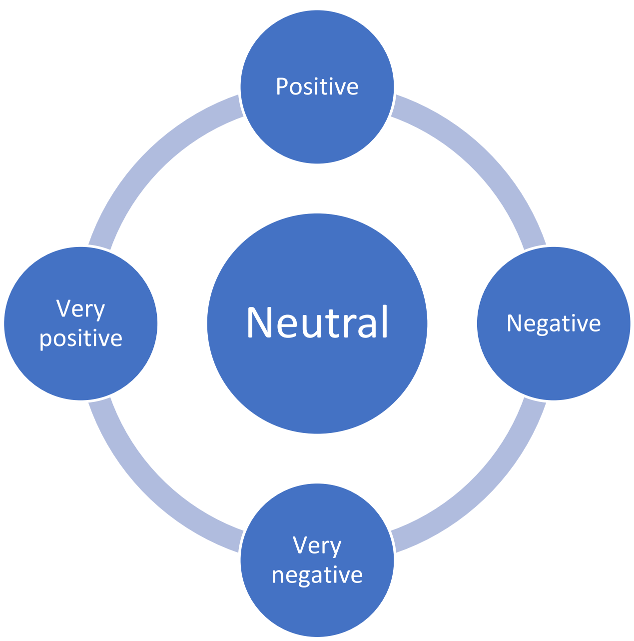
- ✓ To classify the sentiment analysis as sentence level
- ✓ To classify the sentiment analysis as document level
- ✓ To classify the sentiment analysis as sub-sentence level
- ✓ To classify the text as **positive accuracy** and **negative accuracy**

1.5 Sentiment analysis types

There are numerous types or kinds of sentiment analysis and SA apparatuses extend from frameworks for emphasis the polarity (neutral, positive, negative) in different frameworks for distinguish sentiments and opinions (irate, happy, sad & so on) or recognize aims (for example intrigued vs not intrigued). The following areas are used to spread out for the furthermost vital things.

a) Fine grained sentiment analysis

In some cases, it might be likewise intense on actuality progressively exact for the dimension of polarity of feeling, because rather than simply discussing negative, neutral or positive opinions we can think about the following classes basic classes of sentiment analysis:



This is generally alluded to as fine-grained sentiment Analysis. This can be, for instance, plotted from first star to five-star rating in the survey, for example.: for Very negative is 1 star and Very Positive = 5 stars. A few frameworks likewise give distinctive kinds of polarity by distinguishing to negative or positive fleeing is related with a specific opinion, for example, outrage, trouble, or stresses (for example negative sentiments) or joy, love, or energy (for example positive sentiments).

All that you have to know to begin with sentiment analysis is as shown in figure 1.1 below the rating scale is from five star to one star only that is peoples can rate based on their fleeing starting from one star to five star, that is if somebody rates **five star** the sentiment analysis is **very positive**, for **four star** indicates positive sentiment analysis, for **three star** the sentiment analysis is **neutral**, for **two star** the sentiment analysis is **negative**, for the **one star** the sentiment analysis is **very negative.**

Figure 1.1 rating of any product in sentiment analysis

b) Emotion detection

Feeling or emotion detection goes for recognizing feelings like happiness, dissatisfaction, horror, trouble. Numerous emotion (feeling) recognition frameworks resort to vocabularies or lexicons (for example list of words and the feelings or emotion they pass). One of the drawbacks of depending on lexicon or dictionaries is that the manner in which individuals express their feelings fluctuates a ton thus do the lexical things that we used. a couple of words that would usually express hatred like homicide (for instance in your thing is a touch of or your customer backing is executing me) may in like manner express bliss (for instance in compositions like this is the crap or you are murdering it).[1].

c) Aspect based sentiment analysis

Generally, while dissecting opinions in subjects, for example instance items it may be powerful on not just whether the individuals is conversing to negative, neural or positive polarity for the item, but additionally which specific perspectives of the item individuals talks clearly about it. That is the thing that aspect-based sentiment analysis is about. In the above example:

"Battery life time for this camera is excessively too short."

The text or content is communicating negative sentiment for the devices life time, but more unequivocally, about the battery life, which is a specific component parts of the devices.

d) Intent Analysis

Intent is one type of sentiment analysis that essentially recognizes what individuals need to do with the content instead of what individuals state with the content. Let see three different examples below:

"Your client support is a debacle. I've been on hold for 10 minutes".

"I might want to realize how to supplant the cartridge".

"Would you be able to enable me to round out this structure?"

A person has no issues identifying the objection for the first content, since the inquiry for the second content & the demand is more focused on the third content. Nevertheless, different machines may have a few issues for recognize them. At times the expected activity must be deduced from the content, yet now and again, deriving it requires some relevant learning.

e) Multilingual Sentiment Analysis

Multilingual is also one type of sentiment analysis it can be a troublesome assignment. Generally, for great deal of preprocessing is required & that preprocessing types utilization for various assets. For the large portion such kinds of assets are accessible on the web (for example sentiment lexicon), however numerous others must be made (for example interpreted corpora or commotion recognition calculations). The utilization of the assets accessible requires a great deal of coding knowledge and can take long to execute.

An option in contrast to that would identify language in writings naturally, at that point practice model for your preferred language (if writings are not written in English), lastly, play out the investigation.

1.6 Reasons the importance's of sentiment analysis

Around 80% of the world's information's are unstructured & not sorted out in the form of pre-characterized way. A large portion of this originates from content information, similar to messages, bolster tickets, talks, web-based life, reviews, articles, and reports. These writings are typically troublesome, tedious and costly to examine, comprehend, and sort through.

Sentiment analysis frameworks enables organizations to understand this ocean of unstructured content via computerizing business forms, getting significant experiences, and sparing long periods of manual information handling, at the end of the day, by making groups increasingly proficient.

Some portion of the advantages of sentiment Analysis incorporate the accompanying:

- ✓ **Scalability**: Would you be able to envision physically dealing with a huge number of tweets, client bolster discussions, or client surveys? There's simply an excess of information for processing physically. sentiment analysis permits for processing information in scaling productive & financially perception mode.
- ✓ **Real-time analysis**: real time analysis is utilizing for different sentiment analysis for distinguish basic data to permit situational mindfulness amid explicit situations continuously. Is there a PR emergency in online networking going to blast? An irate client that is going to agitate? sentiment analysis framework can help you promptly distinguish these sorts of circumstances and make a move.
- ✓ **Consistent criteria:** most of peoples do not watch pure criteria to assess opinions of different of content. the evaluated distinctive individuals possibly concur around 60-65% of the occasions when making a decision about the opinion for a specific bit of content. It's an emotional errand which is intensely impacted by close to home encounters, contemplations, and convictions. By utilizing a brought together feeling examination framework, organizations can apply similar criteria to the majority of their information. This diminishes mistakes and improve information consistency.

  Look at the Use Cases and Applications area to see instances of organizations & associations for utilizing opinion examination to an assorted arrangement of different contents.

CHAPTER TWO

Background and Motivation

2.1 Background

Sentiment analysis or Estimation examination tries to comprehend a subject's demeanor or passionate response toward a particular point (or brand). Conclusion examination or sentiment analysis (SA) utilizes specific devices, strategies, and techniques to comprehend what individuals state about an issue.

SA does not need to be confused and specialized. It could be something as straightforward as getting an individual in your group to discover what is being said about your image via web-based networking media and recognize its amount is great and the amount of isn't. There is no requirement for a major spending plan and venture into confused programming (it's incredible in the event that you can bear the cost of it, however on the off chance that you can't, that is fine as well).

You can apply the standards of feeling examination or SA to:

- ✓ Item audits, studies, and reactions
- ✓ People and associations
- ✓ Issues, subjects, occasions, preparing
- ✓ On the web and online networking content including Facebook posts, webcasts, Tweets, blog remarks, gathering posts, recordings, illustrations, and pictures.

Instruments, procedures or techniques utilized in notion examination differ broadly and may incorporate web-based life observing, catchphrase/content handling devices, biometric devices, computational semantics, Natural Language Processing (NLP) devices, or a basic evaluation by someone else.

2.2. Motivation

It's crucial for entrepreneurs to focus on clients' criticism about their administrations. It's additionally fundamental for organizations decide the amount of informal exchange can be considered resource or risk to their products' repute. Over estimation investigation (SA) it would be considerably additional proficient to discover how clients feel about administrations, items, offers, occasions & the general population who are the essences of the commercial. Recognizing what a client is now’s suspecting after each web-based social networking post, entrepreneurs can frame and change their promoting procedures as needs be. It can help them get profitable setting on the best way to react to the input and how to approach the clients for the following administration. Ways how assessment examination on clients' remarks can elevate business circumstances are talked about it below:

I. Think of Business Solutions Beforehand:

At the point when an update on offer or administration is exhibited to clients via web-based networking media, business pioneers can acquire complete data utilizing the negative, positive and neutral remarks that will help them to additionally assess, make reports & think of flexible arrangements. It's gainful with regards to benchmarking contenders and markets. In addition, Sentiment investigation can likewise assist organizations with analyzing how the most recent administration is considered among their clients & get a universal thought of which statistic section creates the most enthusiasm for the business.

II. Valuable measure ROI of marketing campaign:

Straightforward count of quantity of the preferences, remarks (devotees) don't give the genuine image as far as it’s achievement of promoting effort. Through sentiment analysis (SA) entrepreneurs can consolidate subjective & measurable estimations & amounts of the sincere ROI rate are advertising effort utilizing the negative or positive talks of clients.

III. Boost (Lift) customer service:

Supposition investigation or SA is an exceedingly powerful strategy that encourages organizations to connect with their clients before any negative inclination nearby the specific administration (organization's) repute spreads wide. So, organizations can transform terrible client understanding into a positive side by giving fulfilling administration.

CHAPTER THREE

LITERATURE REVIEW

3.1 Sentiment Analysis

Sentiment Analysis (SA) shows the utilization of NLP (natural language processing), content examination, computational linguistics & biometrics to efficiently recognize, quotation, evaluate & ponder full of feeling states also, emotional data. It is immensely utilized in client surveys, reactions in on the web or online and social media sites or destinations. Feeling investigation or sentiment analysis is turning into a mainstream consider nowadays, principally in light of the way that social organizing destinations incorporate online clients who are allowed for expressing their considerations, emotions & imitations about a particular point. Online surveys are given to clients in the public organizing destination is critical for the two clients and specialist organizations. With fast web individuals commonly look for things in the web. Indeed, these days, any sort of promoting business is as of now drenching to the new patterns of organizations. Aside from their composed studies, the organizations likewise broaden their consumer loyalty examination through the web, so as to assemble a vast measure of information. [2].

Authors of [3] focused on designing understudies' Twitter presents on comprehend issues and glitches in their instructive encounters. They originally directed a subjective investigation on tests taken from around 25,000 tweets related to designing understudies' school life. They discovered building understudies experience issues, for example, inadequacy of social commitment, overwhelming investigation burden, and rest insufficiency. In view of these outcomes, we actualized a multi-name grouping calculation to classify tweets mirroring understudies' issues.

They at that point utilized the calculation to prepare an indicator of understudy issues from around 35,000 tweets issued at the geo-area of Purdue University. This work, displays a methodology and results that show how easygoing web-based social networking information can give bits of knowledge into understudies' encounters. [3].

Masses of clients share their emotions via web-based networking media like Facebook, making it a profitable stage for following and investigating open notion. Internet based life is one of the greatest stages where monstrous texts are distributed each day which makes it a perfect hotspot for catching the sentiments towards different inquisitive points, for example, items, products or big names, and so on. The principle objective of this paper is to give a diagram of most recent updates in feeling investigation and characterization techniques and it incorporates the concise talk on the difficulties of Sentiment analysis for which the work needs to be finished. We likewise discovered that the vast majority of the works done are in view of machine learning strategy as opposed to the lexicon-based technique [4].

Sentiment analysis has likewise ended with conceivable for investigate the inclinations of an individual. Sentiment analysis push to choose the negative, neutral or positive perspectives on an individual dependent on his frame of mind on a given point. Beforehand, it was utilized for lexical or sentence structure highlight extraction, appointing an extremity mark to each record or content unit. Nowadays, long range informal communication destinations such as Facebook, Twitter demonstrate the impact of environment have on online clients [5].

Sentiment Analysis is put together not just with respect to the negative or then again positive polarity of arguments & ideas, it also additionally on the linguistic diagram of the sentence being investigated. different frameworks attempt toward peruse among appearances, recognizing informal or casual articulations, offering understanding to invalidations, altering extremity of arguments constructing on the related verb modifiers, descriptors, action words, considering explicit utilitarian rationale supplements.[6].

The writers plan to investigate the remarks or comments on the long-range informal communication site Facebook and to build up the module that can decide the quantity of remarks, likes, reposts. From that point, we can distinguish inclines in client intrigue (interest), and build up their Facebook to expand the quantity of observing visits, remarks [7].

CHAPTER FOUR

PERPOSED METHOODOLOGY

4.1 Working principle of Sentiment Analysis?

There are numerous techniques and algorithms for executing sentiment analysis frameworks, and can be named as follows:

- ✓ **Rule-based system:** rule-based system it performing different sentiment analysis dependent in different physically made standards.
- ✓ **Automatic system:** this system is depending in AI (artificial intelligent) or machine learning strategies to gain from information or data.
- ✓ **Hybrid system:** hybrid system that join both Rule based system as well as automatic approaches.

1) Rule-based approaches

As a rule, rule-based methodologies characterize lot of rules in some kinds of scripting language for recognize polarity and subjectivity of a feeling [1].

Rules are utilizing an assortment of sources of info as shown below:

- ✓ Classic Natural language processing strategies like parts of speech tagging, stemming, tokenization.
- ✓ Different assets, for example, lexicon or dictionaries (for example expression and list of words).

A fundamental case of a Rule based execution would be the accompanying:

1. Define two arrangements of separated words (for example negative words, for example, terrible, evil and poorest and positive words, for example, great, best, beautiful, and so on).

2. Given a content:

I. Count quantity of positive words shown up on content.

II. Count quantity of negative words shown up on content.

3. If the quantity of positive word presences is more noteworthy than the quantity of negative word presences returns a positive sentiment, alternately, return a negative sentiment. Something else, return neutral. This framework is "naive" because it doesn't consider how words are joined for the arrangement. for further developed preparing it may be made, but these frameworks become intricate rapidly. They can be difficult for keeping up the new

standards might be expected for include help the vocabulary & expressions. Plus, including new standards may have undesired results because of the connection with past guidelines. Thus, these frameworks require vital interests in physically tuning and keeping up the rules.

2) Automatic Approaches

Automatic strategies or method, as opposed to Rule based frameworks, don't depend on physically created guidelines, however on AI procedures. The sentiment analysis assignment is normally demonstrated as a classification problem where a classifier is fed with a content or text and returns the comparing classification, for example positive, negative, or neutral (on the off chance that polarity analysis is being performed).

Said AI or machine learning classifier can more often than not be executed with the accompanying advances and parts:

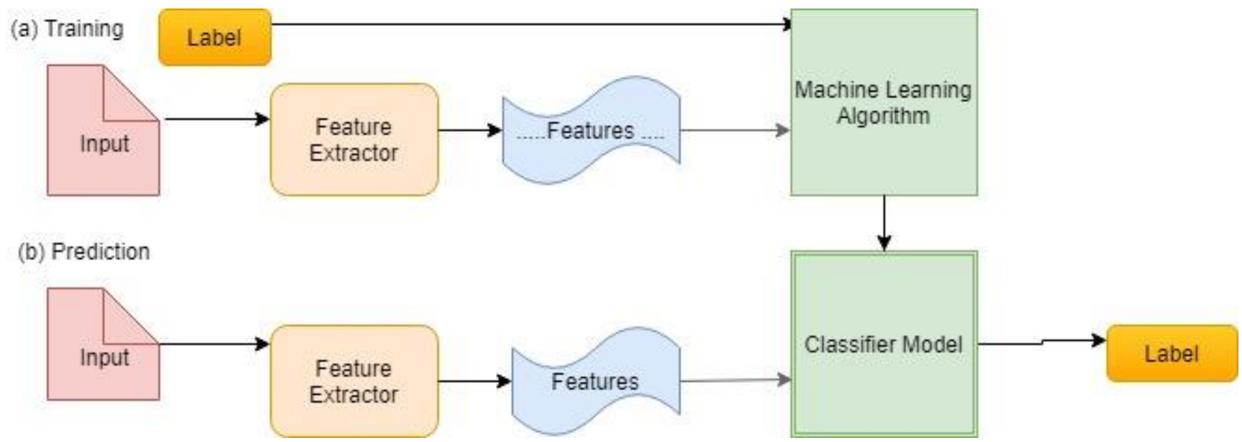


Figure 4.1 flow chart how to train the model

The Training and Prediction Processes

In the training (preparation) procedure (a) our model figures out how to relate a specific information (for example a content (text)) to the relating yield (tag) in light of the test tests utilized for preparing. The component extractor exchanges the content contribution to an element vector. Sets of highlight vectors and tags (for example positive, negative, or neutral) are fed into the AI calculation to produce a model.

In the expectation procedure (b), the element or feature extractor is utilized to change inconspicuous content contributions to include vectors. These element vectors are then fed into the model, which produces anticipated labels (once more, positive, negative, or neutral). [8]

Feature extraction from text

The first step in a machine learning text classifier is to transform the text into a numerical representation, usually a vector. Usually, each component of the vector represents the frequency of a word or expression in a predefined dictionary (e.g. a lexicon of polarized words). This process is known as feature extraction or text vectorization and the classical approach has been bag-of-words or bag-of-n grams with their frequency.

More recently, new feature extraction techniques have been applied based on word embeddings (also known as word vectors). This kind of representations makes it possible for words with similar meaning to have a similar representation, which can improve the performance of classifiers.

Classification Algorithm

The classification step more often than not includes an accurate model like Naïve Logistic Regression, Bayes, Support Vector Machines, or Neural Networks:

- ✓ **Naïve Bayes**: a group of probabilistic calculations or algorithms that utilizes Bayes' Theorem to foresee the classification of a content or text and specifically for this project we are applying naïve Bayes classification algorithm.
- ✓ **Linear Regression:** a very outstanding calculation in insights used to anticipate some value (Y) given a lot of highlights (X).
- ✓ **Support Vector Machines (SVM):** a non-probabilistic model which utilizes a representation of text or content precedents as focuses in a multidimensional space. These models are mapped with the goal that the instances of the diverse classes (Sentiments) have a place with particular districts of that space... At that point, new messages are mapped onto that equivalent space and anticipated to have a place with a classification dependent on which locale they fall into.
- ✓ **Deep Learning:** a differing set of calculations or algorithms that endeavors(attempts) to impersonate how the human mind functions by utilizing artificial neural systems to process information.

Sentiment analysis evaluation and metrics

There are numerous manners by which you can get execution measurements or metrics for assessing a classifier and to see how exact an opinion investigation demonstrate is. A standout amongst the most every now and again utilized is called cross-validation.

What cross-validation is part the preparation information for a specific number of preparing folds (around 70% of the information or text) & also similar number of test data (around 30% of the training information), utilize the preparation folds to prepare the classifier, & the test is against the testing folds for acquiring execution measurements. The procedure is repetitive on various occasions & a normal to every one of the measurements is determined.

On the off chance that the testing set is dependably to the equivalent, it may be overfitting to the testing set, that implies it may change the investigation for the given arrangement of information so it may neglect for breaking down an alternate set. Cross-validation keeps as the maximum information is available and as the crassness is increasing almost certainly utilizes.

Recall, accuracy & precision

Recall, accuracy & precision are normal measurements utilizes for assess the execution of a classifier. Precision estimates how many contents were anticipated effectively having a place with a specified classification out of the majority of the text that were anticipated (accurately & none accurately) as having a place with the classification. Recall estimates what number of texts were anticipated accurately as having a place with a given classification out of the considerable number of texts that ought to have been anticipated as having a place with the classification. We likewise realize as number the of information increase the classifiers indicates that better recall.

Accuracy estimates what number of writings were anticipated effectively (both as having a place with a class and not having a place with the class) out of the majority of the texts in the corpus. Maximum recall and precision are utilized to measure execution because accuracy only cannot that much regarding how to fortunate (unfortunate) a classifier is.

For a troublesome undertaking like sentiment analysis, recall & precision levels are probably going to decrease at the beginning. If we feed the classifier with maximum information, execution will improve. Nonetheless, in feature if we see underneath, because of clarified information isn't probably going to be exact, the odds are that accuracy levels won't get excessively high. Notwithstanding, in the event that you feed the classifier reliably labeled information, results will be in the same class that means the result is the creates some other order issue.

Inter annotator agreement

With regards of the inter annotator agreement (for example understanding of people on a given comment duty), a standout amongst the greatest oftentimes utilized measurements for Krippendorff's Alpha. As per Saif et al., best inter annotator agreement for Facebook and Twitter sentiment analysis achieves 0.655 estimation of Krippendorff's Alpha. [9]. This implies is a decent arrangement for understanding (because alpha is always greater than 0), but we trust still a long way to extraordinary (fore example: about 0.8 is base reliability quality edge societal researchers utilizes so as for state information is dependable).

All things considered, this 0.655 is a pointer of the trouble of feeling examination discovery for people also. Mulling over that machines gain from the information they are encouraged with, programmed forecasts are probably going to reflect the human contradiction implanted in the information.

3) Hybrid Approaches

The idea of hybrid techniques is extremely instinctive: simply join the best of the two universes, the rule based and the automatic approaches. As a rule, by consolidating the two methodologies or techniques can improve precision and accuracy.

Assessment Analysis Challenges

The greater part of the effort in sentiment analysis as of late has been about growing increasingly sentiment classifier is managing a portion to primary difficulties & restrictions in the field.

Subjectivity & Tone

The discovery in objective & subjective contents is similarly to critical breaking down their quality. Actually, assume objective texts don't contain explicit feelings. for instance, you mean to investigate the opinion of the two contents below:

- ✓ This bundle is decent.
- ✓ This bundle is red.

A great many people can state the positive certain for to first text & also neutral to the second text. All establishes (descriptive words, action words, and some nouns) ought not be dealt with the equivalent concerning how they make assessment. In the models above, subjective is more abstract than red.

Polarity and Context

All speeches are expressed eventually in time, in some spot by and to certain individuals you get the point. All expressions are articulated in setting. Breaking down estimation without setting gets entirely troublesome. Be that as it may machines can't find out about settings on the off chance that is not referenced clearly. some of the issues emerge to changes the setting polarity. The following reactions to an overview:

- ✓ All are ok.
- ✓ Literally nothing!

To visualize the reactions listed before originate from responses to the inquiry What did you like about the occasion? The essential response would make certain and the second one would be negative, isn't that so? By and by, envision the responses begin from answers to the request What did you Dislike about the event? The negative in the request will make incline examination change totally. A better than average course of action of preprocessing or post handling will be required if we are to consider at any rate some bit of the setting where compositions were made. Regardless, how to preprocess or post process data in order to get the bits of setting that will help break down slant isn't clear.

Comparisons

Step by step instructions to treat examinations in sentiment analysis is another test worth handling. Take a gander at the writings beneath:

- ✓ This item is best in class.
- ✓ This is superior to old instruments.
- ✓ This is superior to nothing.

There are a few comparisons like the first over that needn't bother with any logical pieces of information so as to be characterized accurately.

The second and third messages are somewhat harder to group, however. OK group them as positive or neutral? Most likely, you are bound to pick positive feeling to the second & neutral feeling to the third, isn't that so? Indeed, setting can have any kind of effect. For instance, if the old instruments the second content discussions about were viewed as futile in setting, at that point the second content ends up being truly like the third content. In any case, if there is no setting given those contents sense extraordinary [8].

Emojis

There are two sorts of emoticons as per Guibon et al.. Western emoticons (for example :D) are encoded in just a single character or in a mix of several them though Eastern emoticons (for example ¯ \ _ (ツ) _/¯) are more drawn out blend of fonts of a perpendicular sort. Especially in tweets, emoticons assume like job for the supposition of writings.

Estimation investigation (SA) performed over tweets requires unique thoughtfulness regarding character level just as word level. Be that as it may, regardless of how much consideration to pay for every great deal of preprocessing may require. For instance, you may need to preprocess web-based life text & change eastern & western emoticons for whitelist & tokens (for example continuously accept the components to group purposely) so, to support the improve assumption investigation execution. This is very thorough rundown for emoticons & their Unicode fonts to prove to be useful while preprocessing.

Defining of neutrality

To characterizing what we mean by neutral is another test in handling so as for performing precise assessment examination. for all arrangement issues, characterizing for different classes & for such situation the unbiased tag is a standout amongst the imperative pieces of different issue. What do you mean the unbiased, negative or positive it makes a difference when you train supposition investigation representations? because labeling information needs to labeling principles be predictable the great meaning of different issues is an absolute necessity.

The following certain thoughts in which an unbiased tag may include:

1. The first one is **Objective content**. This is alleged target writings don't contain express assessments, so you ought to incorporate those texts into the neutral class.
2. The second one is **Irrelevant data**. On the off chance that you haven't preprocessed your information to sift through superfluous data, you can label it nonpartisan. Be that as it may, be cautious! Possibly do this on the off chance that you know how this could influence by and large execution. At times, you will add commotion to your classifier and execution could deteriorate.
3. The third is **content containing wishes**. A few wishes like I wish the item difficult to incorporation commonly unbiased. In any case, those including correlations, similar to wish the item in which is better are truly hard to sort.

Accuracy of sentiment analysis

This sentiment analysis is a massively troublesome undertaking notwithstanding for people. All things considered; opinion investigation classifiers probably won't be as exact as different kinds of classifiers. Keep in mind that between annotator understanding is entirely low and that machines gain from the information they are nourished with.

All things considered, you may state, is it worth the exertion? The appropriate response is straightforward: it beyond any doubt is justified, despite all the trouble! Odds are that feeling examination expectations will not be right occasionally, yet by utilizing supposition investigation you will get the chance to take care of business around 70-80% of the occasions you present your writings for grouping.

In the event that you or your organization have not utilized supposition examination previously, at that point you'll see some improvement actually rapidly. For run of the mill use cases, for example, ticket directing, brand observing, and VOC examination or Analysis, this implies you will spare a great deal of time and cash - which you are probably going to put resources into internal manual effort these days spare your groups about disappointment, & increment your productivity [10].

4.2 Procedures of Sentiment Analysis

Now a day most of peoples of the world are using social network sites like Facebook and twitter, but specifically this project is focused on Facebook and this Facebook contains 5000 characters for all status of comments and updates. From Facebook we have to collect different data in the corpus & save all of them as text file using Facebook graph API with the support of python script [1].

Figure 4.2 flow chart of the process

I. **Data collection:** -includes sources of significant information from the chose social media sites (Facebook, for instance). It ought to be focused on, that the exploration material gathered at this stage is spoken to by disordered information, with everyday language, slang, neologisms, truncations or augmentations.

II. **Text-Preprocessing: -**at this stage, unnecessary data is removing out and non-content information eliminated. Information quality should be careful with the following things:

- ✓ Eliminating site locations and replace them by word URL (unified resource locator),
- ✓ In expressions, for example, #name eliminate #, and leave the name as it were,
- ✓ Eliminate spaces and punctuation marks from the content.

III. **Training data: -** In this stage we have already the clean data from above stage, we train our data using different techniques of calculating polarity of the text.

IV. **Implementation or result**: - here is the final stage of our work, that is classifying our text using naïve Bayes' classification algorithm using final result of Accuracy, Precision and Recall.

4.3 Facebook graph API

To use the Facebook Graph API effectively, we need a record on Facebook and make a planners account in the Facebook for designers' area. at this point Facebook will empower you to expel all the open data from any open gatherings, pages or even from the open or open profile. For this circumstance we should be secure application token which is given by the Facebook application. By this token Facebook get attestation or affirmation about the validity of the customers. this token gets terminated inside a brief span roughly inside 60 minutes. On the other hand, another methodology is to utilize the link type of Facebook application ID & Facebook Secret ID to utilize it as the access token. It can't never terminate [11].

In the figure 4.3 below Facebook graph API is access only by using unique user **access token** for each and every profile or group page. Let's say for my profile is the unique access token is "**EAADX5UbkyhUBAORZBs8jW4pl4wmeoAX9md3LeDd5H5DIcDZB3P4k9JBqNBC9ydVjLZCcmsXHcgDq7gqYjFHs0OA8nZChhQZAPuz39MKaKpNLSM3P85pUc2lD8RLm7goBZC44V2BADgOmfTJPliLVZAazWmvTKZCZAfawYv4JyniHt8AV20zMJwdAC5cZBmzXcCcdQZD"** and this access toke is accessible only for limited time that is after a few time minutes it will expired.

4.3.1 Python script

In the wake of accomplishment of the verification Facebook is utilize from Facebook Graph API developer is prepared data by extracting the information or comments. For such kinds of situation Python libraries and Python content or python scrips are used to to extract the information from their website of Facebook graph API [12]. Our content is working good in Python 2.9 or more. it needs a few exceptional included libraries in our content, for example, Jason, Naïve Bayes Classifier, time and csv.

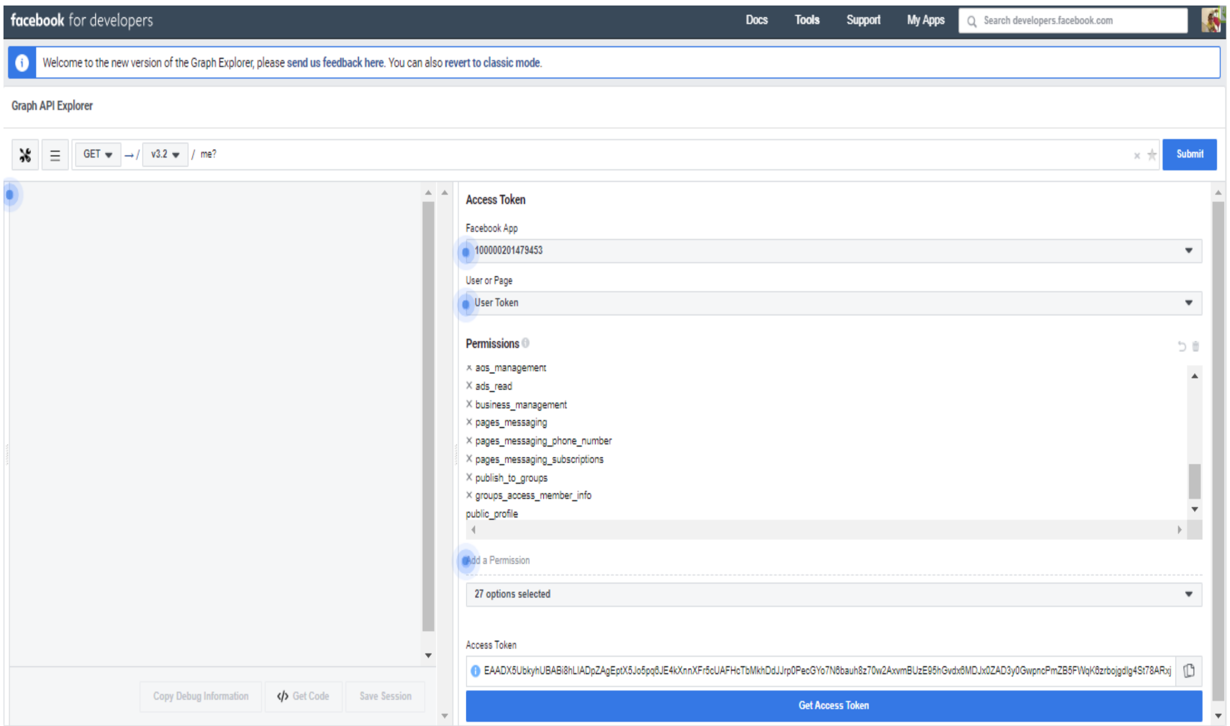


Figure 4.3 graph Facebook for developers

The above figure 4.3 graph API Facebook developer is an application to extract Facebook information as developer from Facebook profile, but to access this application first of all we have to create Facebook app account as developer in Facebook graph API application with a unique access token to access the whole Facebook profile like groups, pages, number of likes, number of comments, number of shares from specific Facebook profile and also we have to hit the access permission perimetries before star the navigation or searching.

CHAPTER FIVE

EXPERIMENTAL DESIGN

5.1 Experimental Result

Now we will talk about the experimental structure setup required to get the outcomes of sentiment analysis. The sentiment analysis classifier includes a progression of steps like data collection, data extraction, naive Bayes classifier, and creating confusion matrix and lastly making an expectation model. In this way, we have executed an apparatus in python that plays out every one of these procedures on the informational index that contains naive Bayes classifier just as interpretation for each example.

Text Blob: - Text Blob is found at the top of NLTK tool and it includes python 2 & python 3 libraries to handle fictional content or information. It gives a steady API to plunging into public NLP (natural language processing) assignments; for example, POS, thing phrase extraction, conclusion examination, and the sky are the limit from there.

Text Blob means to give access to basic text-processing tasks through a natural interface. You can treat Text Blob protests as though they were Python strings that figured out how to do Natural Language Processing (NLP).

Text Blob is a python library and offers a straightforward API to get to its techniques and perform fundamental NLP assignments. something worth being thankful for about Text Blob is that they are much the same as python strings. Along these lines, you can change and play with it same as we did in python.

The supposition (assessment) property reestablishes a named tuple of the structure Sentiment (subjectivity, polarity). The polarity score is a float inside the range [-1.0, 1.0]. The subjectivity is a float inside the range [0.0, 1.0] where 0.0 is amazingly target and 1.0 is emotional. On the off chance that no source language is determined, Text Blob will endeavor (attempt) to identify the language. You can indicate the source language explicitly, as so. Raises Translator Error if the Text Blob can't be converted into the mentioned language or Not Translated if the deciphered outcome is equivalent to the info string.

```
 RESTART: C:\Users\zed\AppData\Local\Programs\Python\Python37\Lib\site-packages\
textblob.py
['__add__', '__class__', '__contains__', '__delattr__', '__dict__', '__dir__', '
__doc__', '__eq__', '__format__', '__ge__', '__getattribute__', '__getitem__', '
__gt__', '__hash__', '__init__', '__init_subclass__', '__iter__', '__le__', '__l
en__', '__lt__', '__module__', '__ne__', '__new__', '__reduce__', '__reduce_ex__
', '__repr__', '__setattr__', '__sizeof__', '__str__', '__subclasshook__', '__we
akref__', '_cmpkey', '_compare', '_create_sentence_objects', '_strkey', 'analyze
r', 'classifier', 'classify', 'correct', 'detect_language', 'ends_with', 'endswi
th', 'find', 'format', 'index', 'join', 'json', 'lower', 'ngrams', 'noun_phrases
', 'np_counts', 'np_extractor', 'parse', 'parser', 'polarity', 'pos_tagger', 'po
s_tags', 'raw', 'raw_sentences', 'replace', 'rfind', 'rindex', 'sentences', 'sen
timent', 'sentiment_assessments', 'serialized', 'split', 'starts_with', 'startsw
ith', 'string', 'strip', 'stripped', 'subjectivity', 'tags', 'title', 'to_json',
 'tokenize', 'tokenizer', 'tokens', 'translate', 'translator', 'upper', 'word_co
unts', 'words']
[('This', 'DT'), ('TextBlob', 'NNP'), ('is', 'VBZ'), ('an', 'DT'), ('interesting
', 'JJ'), ('because', 'IN'), ('it', 'PRP'), ('is', 'VBZ'), ('used', 'VBN'), ('fo
r', 'IN'), ('all', 'DT'), ('text', 'NN')]
Sentiment(polarity=0.625, subjectivity=0.5)
Positive accuracy = 83.41399032116146% via 4546 samples
Negative accuracy = 100.0% via 2979 samples
>>>
```

5.1 Text Blob classifying of texts in to positive and negative accuracy

Since, from the above figure 5.1 we can conclude that from around 7525 given sample texts around 83.41% are positive accuracy via 4546 samples and 100% are negative accuracy via 2979 samples this indicates that almost all the negative texts or sentence are all pure negative sentence or texts because 100% pure negative accuracy is the result showing and contrasting from around 4546 sample sentence or texts we get 83.41% positive accuracy that's good most of around 83% of the sentence or texts are positive.

```
>>> from textblob.classifiers import NaiveBayesClassifier
>>> train = [
    ('I love this sandwich.', 'pos'),
    ('this is an amazing place!', 'pos'),
    ('I feel very good about these beers.', 'pos'),
    ('this is my best work.', 'pos'),
    ("what an awesome view", 'pos'),
    ('I do not like this restaurant', 'neg'),
    ('I am tired of this stuff.', 'neg'),
    ("I can't deal with this", 'neg'),
    ('he is my sworn enemy!', 'neg'),
    ('my boss is horrible.', 'neg')
    ]
>>> test = [
    ('the beer was good.', 'pos'),
    ('I do not enjoy my job', 'neg'),
    ("I ain't feeling dandy today.", 'neg'),
    ("I feel amazing!", 'pos'),
    ('Gary is a friend of mine.', 'pos'),
    ("I can't believe I'm doing this.", 'neg')
    ]
>>> cl = NaiveBayesClassifier(train)
>>> print ("\t ##### Naive Bayes Classifier for training text #####")
         ##### Naive Bayes Classifier for training text #####
>>> print(cl.classify("This is an amazing library!"))
pos
>>> print(cl.classify('I do not like this restaurant'))
neg
>>> c2 = NaiveBayesClassifier(test)
>>> print ("\t ##### Naive Bayes Classifier for testing text #####")
         ##### Naive Bayes Classifier for testing text #####
>>> print(c2.classify('the beer was good.'))
pos
>>> print(c2.classify('I do not enjoy my job'))
neg
>>> prob_dist = cl.prob_classify("This one's a doozy.")
>>> print(prob_dist.max())
pos
>>> print(round(prob_dist.prob("pos"), 2))
0.63
>>> print(round(prob_dist.prob("neg"), 3))
```

Figure 5.2 Text Blob classifying of iisample text as pos and neg for training and test data

From the above figure 5.2 text blob tool we have two sample text training text and test tex and from this text using text blob and Naïve Bayes classifier we classify the text in two as pos and neg and also, we calculated the probability district matrix value.so that the first sentence or text “the beer was good” is a positive sentence the message is beer was nice and it’s probability district value is 0.63 that is 63% of the sentence is positive accuracy and from the second sentence or text “This one’s a doozy” is a negative sentence means the things are doozy and it’s probability district value is 0.37 that is 37% of the sentence is negative accuracy.

VADER Sentiment Analysis: - VADER Sentiment Analysis is described as VADER (Valence Aware Dictionary and sentiment Reasoner) is a standard based and a vocabulary notion examination apparatus that is expressly delicate to estimations imparted in web-based life, and capacities outstandingly on compositions from various regions. VADER (Valence Aware Dictionary for sentiment Reasoning) is a model utilized for content conclusion investigation that is more sensitive to both polarity (positive/negative) and intensity (quality) of feeling. Presented in 2014, VADER text opinion investigation utilizes a human-driven methodology, consolidating subjective examination and experimental approval by utilizing human raters and the astuteness of the group.

In this post, we'll talk about how VADER feeling investigation ascertains the sentiment score of an information content. It consolidates a word reference, which maps lexical highlights to feeling power, and five straightforward heuristics, which encode how relevant components addition, decrement, or invalidate the assessment of content. think about the accompanying sentences: "The food is interesting." And "I hate that man."

Do you get a feeling of the emotions that these sentences infer? The first plainly passes on positive feeling, while the second passes on negative feeling. People partner words, expressions, and sentences with feeling. The field of Text Sentiment Analysis endeavors to utilize computational calculations so as to unravel and measure the feeling contained in media, for example, content, sound, and video.

Text Sentiment Analysis is a huge field with a great deal of academic literature behind it. Nevertheless, its tools actually simply come down to two methodologies: the lexical methodology and the machine learning approach.

Lexical methodologies expect to outline to sentiment by structure a vocabulary or a 'word reference of feeling.' We can utilize this lexicon to evaluate the opinion of expressions and sentences, without the need of taking a gander at whatever else. Conclusion can be straight out –, for example, {negative, neutral, positive} – or it tends to be numerical – like a scope of forces or scores. Lexical methodologies take a gander at the sentiment class of each word in the sentence & choose what notion classification of the entire sentence is. The intensity of lexical methodologies lies in the way that we don't have to prepare a model utilizing named information, since we have all that we have to survey the supposition of sentences in the lexicon of feelings. VADER is a case of a lexical technique. machine learning approaches, then again, take a gander at recently marked information so as to decide the feeling of at no other time seen sentences. The machine learning approach includes preparing a model utilizing recently observed content to anticipate/group the conclusion of some new information content. The good thing about machine learning approaches is that, with a more remarkable volume of information, we for the most part show signs of improvement expectation or order results. In any case, in contrast to lexical methodologies, we need recently named information so as to really utilize machine learning models.

VADER (Valence Aware Dictionary and sentiment Reasoner) is a vocabulary (lexicon) and rule-based SA tool that is explicitly sensitive feelings communicated in different media life. It is completely publicly released under the [MIT License]

Part of speech tagging

These are grammatical features. Since Text Blob is found at the top of NLTK, the grammatical form tags are equivalent. Here are the definitions blow table 5.1:

One of the more dominant parts of NLTK for Python is the part of speech tagger that is inherent. Remarkably, this part of speech tagger isn't perfect, yet it is pretty darn great. On the off chance that you are searching for something better, you can buy a few, or even adjust the current code for NLTK. Part of speech tagging is with the goal that you can comprehend the sentence structure and start to utilize your program to some degree pursue the importance of a sentence dependent on the word utilized, it’s part of speech, and the string it makes.

To go with the video, here is the example code for NLTK part of speech tagging with heaps of remarks and data also:

Table 5.1 POS Tags lists

POS Tags lists	Descriptions
CC	Coordinating conjunction
CD	Cardinal digits
DT	Determiner
EX	Existential there (like this is, there exists)
FW	Foreign word

IN	Preposition/subordinating conjunction
JJ	Adjective ‘big’
JJR	Adjective, comparative ‘bigger’
JJS	Adjective, superlative ‘bigger’
LS	List marker
MD	Model could will
NN	Noun, singular ‘desk’
NNS	Noun plural ‘desk’
NNP	Proper noun, singular ‘Harrison’
NNPS	Proper noun, plural ‘Americas’
PDT	Predeterminer ‘all the kids’
POS	Possessing ending parent\’s
PRP	Personal pronoun I, he, she
PRP$	Possessive pronoun my, his, hers
RB	Adverb very, silently
RBR	Adverb, comparative better
RBS	Adverb, superlative best
RP	Particle give up
TO	To go ‘to’ the store
UH	Interjection erm
VB	Verb, base from take
VBD	Verb, past tens took
VBG	Verb, gerund/present participle taking
VBN	Verb, past particle taken
VBP	Verb, sing. present, non-3d take
VBZ	Verb, 3rd person sing present takes
WDT	WH-determiner which
WP	WH-pronoun who, what
WP	Possessive wh-pronoun whose
WRB	Wh-adverb where, when

```
 RESTART: C:\Users\zed\AppData\Local\Programs\Python\Python37\Lib\site-packages\
vender.py
Sentiment(polarity=0.25, subjectivity=0.55)
{'neg': 0.0, 'neu': 0.798, 'pos': 0.202, 'compound': 0.5081}
[('He', 'PRP'), ('must', 'MD'), ('be', 'VB'), ('arrested', 'VBN'), ('if', 'IN'),
 ('he', 'PRP'), ('is', 'VBZ'), ('not', 'RB'), ('a', 'DT'), ('criminal', 'NN')]
Positive accuracy = 69.4298574643661% via 5332 samples
Negative accuracy = 57.764441110277566% via 5332 samples
>>> 
```

Figure 5.2 Classifier accuracy for training data

The above figure 5.3 indicates that for more than 10.000 sample sentence or text files we get sentiment analysis so that around 5332 samples are negative samples and around 5332 also positive samples. Then from those samples 69.43% is positive accuracy via 5332 sample sentence and 57.76% is negative accuracy via 5332 sample sentences or texts. So, we can conclude that It would seem that our positive accuracy is tolerable, however the negative sentiment accuracy isn't all that good. It resembles our positive accuracy is not too bad, yet the negative sentiment accuracy isn't too great.

5.3 Naïve Bayesian classifier

Naive Bayes is a standout amongst the most utilized procedures with regards to text classification problem which include high dimensional preparing informational collections per training data. It depends on Bayes' likelihood hypothesis. It introduces less trouble than different calculations. It isn't just quick yet additionally profitable as far as making forecasts dependent on generally little measure of as of now provided data. Despite the fact that it depends on the Bayes' hypothesis, that means the labeled "naive" since the presumption that the event of a specific element in a class is free events in different highlights to paying little respect to any connection of events among them. For instance, a natural product might be viewed as an apple on the off chance that it is red, round and around 3 creeps in width. Regardless of whether the highlights – color, shape and diameter across rely upon one another, these properties add to the decision that this regular item is an apple expected to features free of each other.

$$\boldsymbol{P(A/B) = \frac{P(B/A)P(A)}{P(B)}} \text{.......................................} \textbf{(*)}$$

The conditional probability of Bayes' hypothesis chooses the probability of an event subject to the past data of the events. In the condition referenced above, there are two events A and B.

P(A) demonstrates the likelihood of the events of occasion some time P(B) demonstrates the likelihood of the events of occasion B. hence as per Bayes' speculation, P(A/B) or likelihood of A give B is equivalent to P(B/An) or likelihood of B give A duplicated likelihood of An up on likelihood of B. that is An is suggestion and B is proof. P(B|A) indicates probability or how well the model predicts the data, P(A) appears earlier likelihood or the amount we believe the model exactly delineates reality set up together data as for all our earlier information, P(B) exhibits the Normalizing relentless or on the other hand the steady that impacts the back thickness to facilitate to one and P(A|B) shows back likelihood or the amount we believe a given model definitely portrays the condition given the additional open data including all our earlier information [13].

For the credulous Bayes' classifier for instance, let us acknowledge that there is a corpus of data reliant on comments left by Facebook customers on a post. The datasets are occurrences of positive and negative words set away in two one of a kind classes s and using the classifier figuring we can appraise the likelihood of explicit words' going on or count estimation and their positive and negative thoughts. This is required as pre-arranged advisers for get ready datasets. Thusly, the components in this dataset can be seen as inclinations and the probability that a variable will happen given the verification in the sentence can be conveyed as appeared underneath condition (**).

$$\boldsymbol{Probability(conclusion/Sentence)} \\ = \frac{\boldsymbol{Probability(Sentence/conclusion)Probability(Sentence)}}{\boldsymbol{Probability(conclusion)}} \dots\dots\dots\dots\dots. \ (**)$$

In this case I can expect the words in a sentence as tokens and P(sentence/conclusion) on account of P(token/conclusion) over all of the words in a sentence. This will consider the amount of times a word has occurred in a sentence.

$$\textit{Probability}(\textbf{token/conclusion}) = \textbf{count (the token of the class} + \frac{\mathbf{1}}{\textbf{count (the token in the class)}} + \textbf{count(all token)} \ldots\ldots\ldots\ldots\ldots \quad (***)$$

The option of adding 1 in the check above is called Add-One Smoothing which is used to kill any likelihood of duplication with zero, implies that is keeping away or removing from a word happening multiple times or to demonstrate that a word has happened in any event yet again than the esteem exhibited in the preparation information(training data). Along these lines, the classifier immediately registers the earlier likelihood that is the likelihood of a word being sure or negative earlier any preparation information reliant on the amount of positive and negative word points of reference. at that point then for each class, the tokens are copied or increased with the likelihood of each word being in that class. After that the last result is evaluated and the most elevated scoring class is returned as an extremity which chooses whether it is decidedly or contrarily of the content or remark.

5.4 Training data

The training investigation is concentrating in naïve Bayes classifier to decide the text collected from Facebook is either a negative or positive message input based on the remarks. That is, we have to focused on the status and concentrate every one of the remarks. Later on, we have to arranged these remarks for our preparation in test data and training data.

Around 1500 remarks from the status and arranged up to 60 percent of the remarks. physically denoted the data into negative and positive extremity. We made two clusters (shows) for prescient positive words and prescient negative words. By then we absolutely the recurrence of all these positive and negative words in the sentences. As demonstrated by the occasion of a prescient positive word for the genuine(true) positive sentences and in the real negative sentences we use the mean measurable equation to find a weight for each word. For this the readiness informational collection will work all the more capably or proficiently. For instance, "As" is a prescient positive word. In any case, it may not for the most part be in the certified positive sentence. Let say event(occurrence) of "LIKE" in the authentic positive sentences are 60% occasions and again in the real negative sentences in 25% occasions. Thus, the probability (likelihood) of this word for being a veritable positive word is 85%. So, in all likelihood, if "LIKE" exists in a sentence conceivably the sentence is multiple times positive inside 100% occasions.

Table 5.2 Positive and Negative polarity sample comments

Positive polarity sample text or comments	**Negative polarity sample text or comments**
Very effective, I like the offer.	I can’t offer, it is costly.
This is good, when you compared with others.	This is too exaggerated.
For students the offer is good.	Not satisfactory totally the offer.
The food is delicious.	I am not happy with restaurant.
I get free services.	Quality is not good.
The company is great.	Environment condition is not that much good.

CHAPTER SIX

RESULT AND ANALYSIS

6.1 Analysis

In this thesis or project to perform the sentiment analysis of different text reviews from social media Facebook we used three techniques or methods Precision, Recall and Accuracy and those three techniques are very useful and effective in sentiment analysis. Accuracy it performs the overall results of our analysis, but it may or may not be like we expected or precise exactly. Precision is always measuring the exactness of the classifiers and in precision it indicates a high precision shows less false positive, and lower precision shows more false positives. Therefore, if we want to improve precision value, we have to decrease recall values. Recall is also measuring the sensitivity of the classifier or completeness of classifiers. A high recall shows fewer false negative, whereas lower recall shows higher false negatives. The same thing if we want to increase or improve recall value, we have to decrease the precision values. In order to calculate the precision, recall and accuracy percentages of both negative and positive set of words we have to initialize dataset as actual zero (0), actual one (1), predicated zero (0) and predicated one (1) as shown below in table 6.1.

Table 6.1 Confusion matrix table to find out precision, accuracy and Recall

	0 (predicted)	1 (predicated)
0 (Actual)	A (TP)	B (FN)
1 (Actual)	C (FP)	D (TN)

Where A: True positive (TP)

B: false negative (FP)

C: false positive (FN)

D: true negative (TN)

Accuracy = $\frac{A+D}{A+B+C+D}$

Recall (positive) = $\frac{A}{A+C}$

Recall (negative) = $\frac{D}{B+D}$

Precision (positive) = $\frac{A}{A+B}$

Precision (negative) = $\frac{D}{C+D}$

In which FP, TP, TN and FN indicates that to the number of genuine or true positive occasions, the quantity of false negative cases, the quantity of false positive occurrences and the quantity of genuine negative cases as defined above table 6.1.

Now with the help of above formula we have to find out the percentage of each method (techniques) and also compared and contras with all percentage results with different training data sets like 100, 150, 200 and 250 as shown below table 6.6, table 6.5, table 6.4, table 6.3 and table 6.2 in each different methods that is in negative recall, positive recall, negative precision, positive precision and accuracy [8].

Table 6.2 Result for accuracy

Training datasets	Accuracy (%)
100	80%
150	80%
200	85%
250	82%

Table 6.3 Result positive corpus for precision

Training datasets	Positive corpus (%) for precision
100	57.14%
150	66.667%
200	75%
250	71.43%

Table 6.4 Result negative corpus for precision

Training datasets	Negative corpus (%) for precision
100	92.31%
150	88.88%
200	91.667%
250	89.65%

Table 6.5 Result positive corpus for recall

Training datasets	Positive corpus (%) for recall
100	80%
150	80%
200	85.7%
250	83.33%

Table 6.6 Result Negative corpus (%) for Recall

Training datasets	Negative corpus (%) for Recall
100	80%
150	80%
200	84.6%
250	81.25%

The lexicon-based techniques more often than not give scores to the highlights or words present in the records and afterward we pick top N features out of them which are most helpful for the classification reason and overlook or evacuate the remainder of the highlights or features. We have utilized the term number of training datasets to represent to the quantity of features that have been picked as best N features by the utilized lexicon-based strategy. The execution of the classifier changes a ton with the change in the estimation of N, i.e., number of training datasets, in this way we have determined the execution measurements for both the techniques over a scope of N. In our investigation, we changed the estimation of number of training datasets (N) from 100 to 250, with step-size of 1 and recorded the different execution measurements. The outcomes we acquired are as appeared below:

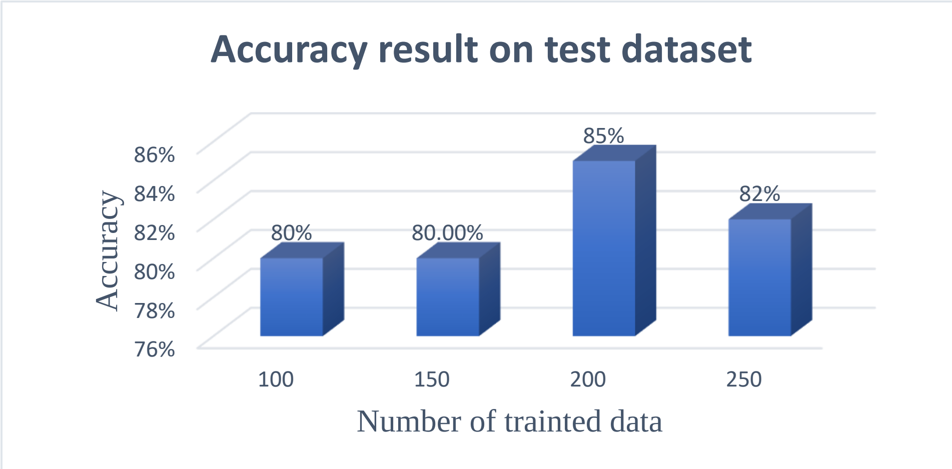


Figure 6.1 Result of accuracy

From the above figure 6.1 when we utilized naive Bayes strategy as the component determination technique, at that point the best accuracy accomplished is 85% when the number of datasets is 200. Additionally, the average accuracy accomplished when number of datasets was differed from 100 to 250 was 81.75%. additionally, the accuracy graph is significantly more stable when our strategy is utilized. As Fig. 6.1 shows, the accuracy is fluctuating a great deal with the adjustment in the number of training datasets when naive Bayes strategy is utilized, so our technique gives substantially more stable in the accuracy of the classifier in naive Bayes technique.

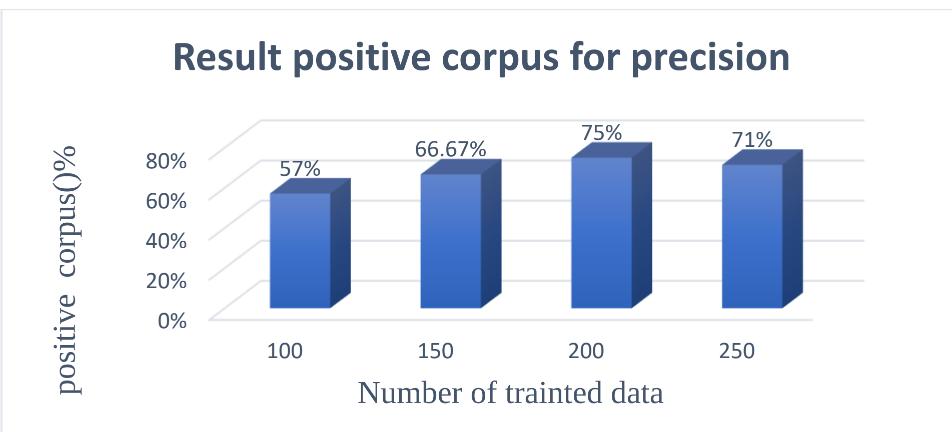


Figure 6.2 Result positive corpus for precision

When we utilized naive Bayes strategy as the component determination technique as shown figure 6.2 above, when the number of datasets is 200 the best positive precision corpus is accomplished 75%. And the average positive precision corpus accomplished when number of datasets is differed from 100 to 250 is 67.4%.

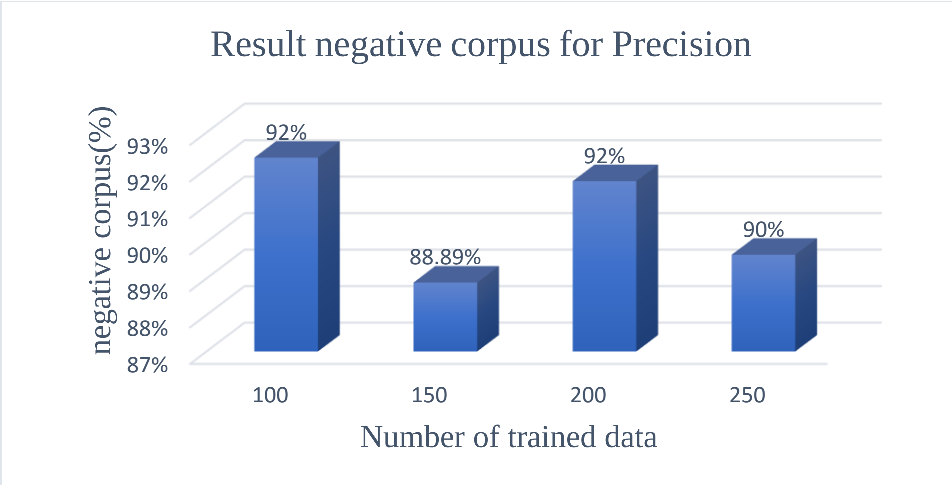


Figure 6.3 Result negative corpus for precision

The same thing when we utilized naive Bayes strategy as the component determination technique as shown figure 6.3 above, when the number of datasets is100 and 200 the best negative precision corpus is accomplished 92%. And the average negative precision corpus accomplished when number of datasets is differed from 100 to 250 is 90.7%.

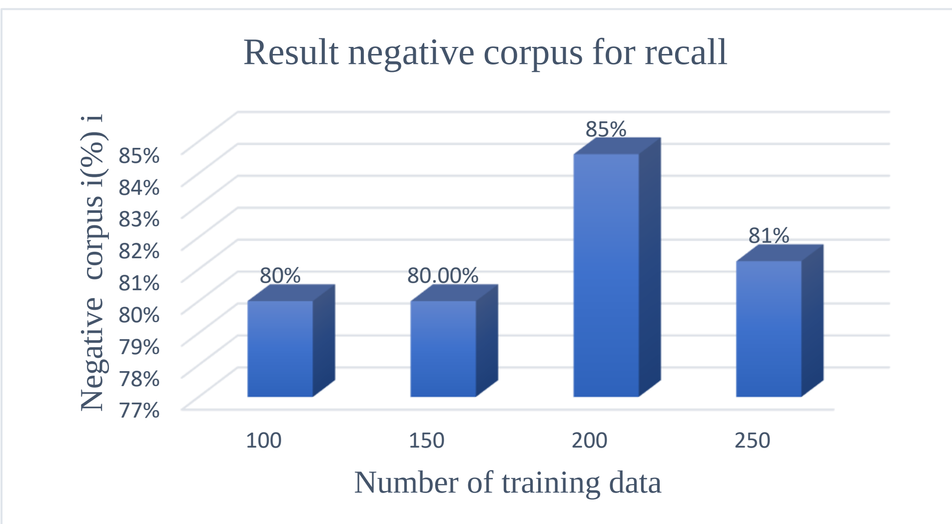


Figure 6.4 Result negative corpus for recall

When we utilized naive Bayes strategy as the component determination technique as shown figure 6.4 above, when the number of datasets is 200 the best positive Recall corpus is accomplished 85%. And the average positive Recall corpus accomplished when number of datasets is differed from 100 to 250 is 81.5%.

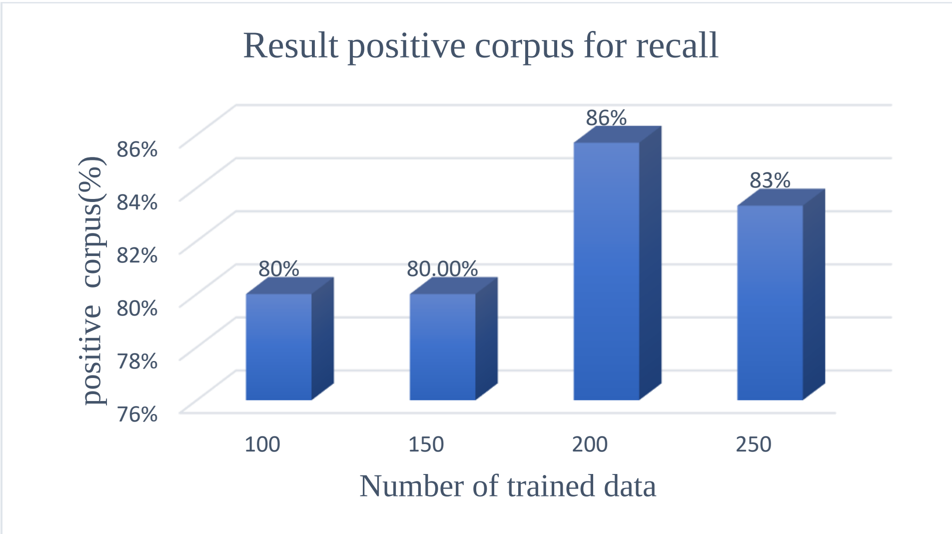


Figure 6.5 Result positive corpus for recall

Finally, when we utilized naive Bayes strategy as the component determination technique as shown figure 6.5 above, when the number of datasets is 200 the best negative Recall corpus is accomplished 86%. And the average negative Recall corpus accomplished when number of datasets is differed from 100 to 250 is 82.25%.

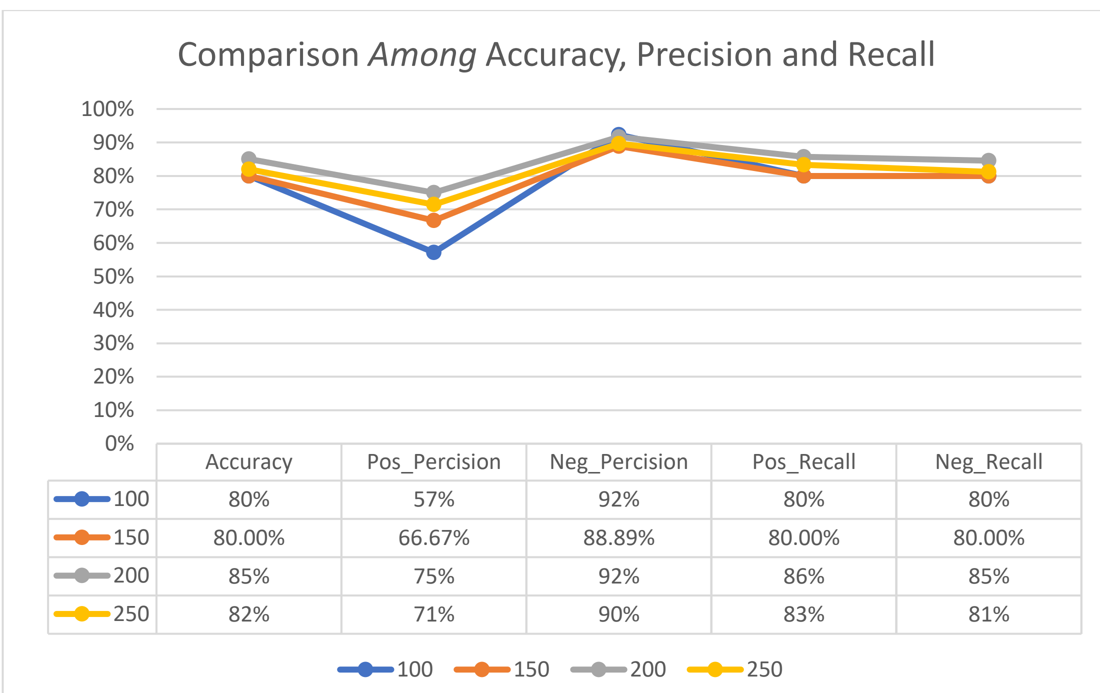


Figure 6.6 Comparison Accuracy, Precision and Recall

From the above figure 6.6 the comparison in Accuracy, pos precision, neg precision, pos recall and neg recall the minimum value is in pos precision 66.6% not only pos precision but also accuracy, neg precision, neg recall and pos recall is minimum value when you compared with Accuracy, neg precision, pos recall and neg recall. The pick or maximum value of the graph from the above figure 6.6 is neg precision around 92% even though at this point is maximum and at pos precision is also minimum around 66.6% we can conclude that the almost the manual calculated value and expected value is the same.

CHAPTER SEVEN

CONCLUSION AND FEATURE WORK

7.1 Conclusion

In this research or thesis, I used machine learning automatic approaches for sentiment analysis using social media Facebook posts of different companies. from the posts of Facebook company, I collect data with the help of Facebook graph API after I getting the required data, I have applied naïve Bayes' machine learning classifier techniques. The comparison in Accuracy, pos precision, neg precision, pos recall and neg recall the minimum value is in pos precision 66.6% not only pos precision but also accuracy, neg precision, neg recall and pos recall is minimum value when you compared with Accuracy, neg precision, pos recall and neg recall. The pick or maximum value of the graph from the above figure 6.6 is neg precision around 92% even though at this point is maximum and at pos precision is also minimum around 66.6% we can conclude that the almost the manual calculated value and expected value is the same.

7.2 Feature work

This sentiment analysis in social media on Facebook is now a day is very important and critical specially in politicos, online shopping market, advertising of products, Election … etc. So, that as a feature work to improve effectiveness of the outcome we have to apply with different techniques like SVM or Support vector machine and KNN or k- nearest Neighbors algorithm techniques in order to compare and contrast the result with naïve Bayes classifier.

Reference

[1] . schwaiger, josef, lang, markus, ritter, christian, "assessing the accuracy of sentiment analysis of social media posts at small and medium sized enterprises in southern Germany", twenty-fourth European conference on information systems (ecis), Istanbul, turkey, 2016.

[2] . Bushra Shourin and Kazi Sayef Shawgat, Analyzing Data of Social Media to Evaluate Customer Behavior of Companies Using Sentiment Analysis (2018).

[3] . Xin Chen,Mihaela Vorvoreanu, and Krishna Madhavan," Mining Social Media Data for Understanding Students' Learning Experiences" *IEEE* trans. on learning technologies, vol. 7, no. 3, JulySeptember 2014.

[4] . Harshali P. Patil, Dr. Mohammad Atique, "Sentiment Analysis for Social Media: A Survey", 978-1-4673-8611-1/15/$31.00 ©2015 IEEE.

[5] . M. Kumar, Dr. A. Bala, "Analyzing Twitter Sentiments Through Big Data", 2016 International Conference on Computing for Sustainable Global Development (INDIACom).

[6] . F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, T. By, "Sentiment Analysis on Social Media", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 919-926, 2012

[7] . Thao Thanh Nguyen, Alla G. Kravets "Analysis of the Social Network Facebook Comments", Hint,> the publication date is not mentioned in the paper.

[8] . Sindhu C, Dyawanapally Veda Vyas, Kommareddy Pradyoth, "Sentiment Analysis Based Product Rating Using Textual Reviews", International Conference on Electronics, Communication and Aerospace Technology ICECA 2017

[9] . Krippendorff, K. (2013) pp. 221–250 describes the mathematics of *alpha* and its use in content analysis since 1969.

[10] .https://monkeylearn.com/sentiment-analysis/#why-sentiment-analysis-is-important.

[11] . Bushra Shourin and Kazi Sayef Shawgat , Analyzing Data of Social Media to Evaluate Customer Behavior of Companies Using Sentiment Analysis (2018)

[12] T. Wilson, J. Wiebe, P. Hoffman, “Recognizing contextual polarity in phrase level sentiment analysis”, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347-354, 2005.

[13] a. Goel, J. Gautam, S. Kumar, “Real Time Sentiment Analysis of Tweets Using Naïve Bayes, 2nd International Conference on Next Generation Computing Technologies (NGCT-2016) Dehradun, India 14-16 October 2016.

LIST OF PUBLICATIONS

[1] . Paper work under the guidance of 'Associate prof. Vinod kumar' with title "Sentiment Analysis on Social Media Facebook in different OS Versions of iPhones", I participated National confference on communication and data sicence (NCCDS-19) 26th April 2019 and published in BJIT – International Journal of Information Technology by SPRINGER NATURE.

[2] . Review paper work under the guidance of 'Associate prof. Vinod kumar' with title "Knowledge Discovery and Data Mining Review Papers", International Journal of Engineering Trends and Technology (IJETT) - Volume 67 Issue 4 - April 2019.