TO STUDY OF SENTIMENTAL MARKETING AND ANALYSIS IN REFERENCES OF ARTIFICIAL INTELLIGENCE.

A Project submitted to

University of Mumbai for partial completion of the degree of

Bachelor in Commerce (Management Studies)

Under the Faculty of Commerce

By

YASH JANARDAN GAVANKAR

Under the Guidance of

Prof. Shweta Soman

M.L. Dahanukar College of Commerce

Dixit Road, Vile- Parle East

Mumbai- 57.

APRIL 2021

INDEX

M.L. Dahanukar College of Commerce

Dixit Road, Vile- Parle East

Mumbai- 57

CERTIFICATE

This is to certify that Mr. Yash Janardan Gavankar, has worked and duly completed his projec work for the degree of Bachelor in Commerce (management studies) under the Faculty o Commerce in the subject of Marketing and his project is entitled, 'Study of sentimenta marketing and analysis in references of artificial intelligence' under my supervision.

I further certify that the entire work has been done by the learner under my guidance and that no part of it has been submitted previously for any Degree or Diploma of any University.

It is his own work and facts reported by his personal findings and investigations.

Date of Submission:

Prof. Shweta Soman

Guiding Teacher

Signature of Principal:

Signature of Coordinator:

College Seal:

DECLARATION BY LEARNER

I the undersigned Mr. Yash Janardan Gavankar hereby, declare that the work embodied in thi project work titled ‘Study of sentimental marketing and analysis in references of artificia intelligence’

forms my own contribution to the research work carried out under the guidance of **Prof. Shweta Soman** is a result of my own research work and has not been previously submitted to any other university for any other Degree/ Diploma to this or any other University.

Wherever Reference has been made to previous work of others, it has been clearly indicated as such and included in the bibliography.

I, here by further declare that all the information of this document has been obtained and presented in accordance with academic rules and ethical conduct.

Yash Janardan Gavankar

Certified by

Prof. Shweta Soman

ACKNOWLEDGMENT

To list who all have helped me is difficult because they are so numerous and the depth is so enormous.

I would like to acknowledge the following as being idealistic channels and fresh dimensions in the completion of the project.

I take this opportunity to thank the **University of Mumbai** for giving me chance to do this project.

I would to thank my **Principal, Dr. Dyaneshwar Doke** for providing the necessary facilities required for completion of this project.

I take this opportunity to thank our **Coordinator Dr. Mitali Shelankar**, for her moral support and guidance.

I would also like to express my sincere gratitude towards my project guide **Professor. Shweta Soman**

whose guidance and care made the project successful.

I would like to thank my **College Library**, for having provided various reference books and magazines related to my project.

Lastly, I would like to thank each and every person who directly or indirectly helped me in the completion of the project especially **my Parents and Peers** who supported me throughout the project

CHAPTER 1 INTRODUCTION

1.1 Introduction

Reaching your pockets hopefully, your phone's still there our phones do a lot for us they check the weather they remind us to turn on our alarm just in case we don't wake up the next morning but there is one thing our phones can't do yet tell us how we are ‘hey Siri how am I doing today?’ ‘Okay Google how are my emotions today?’ these seem like ridiculous questions but with advancements in sentiment analysis and machine learning our machines are becoming closer to answering these very questions let me give you a sentence. For e.g., I love that movie and I asked you to rate it out of the 10 with 0 being negative and 10 being positive now we'd all agree that this is a positive sentence and give it around to 10.

Now let's change the verb a bit I liked that movie here still pretty positive but lower on the scale now let's go to the other end of the spectrum I hated that movie now whoever said this feels negative about the subject and so we'd probably give this around zero now sentiment analysis is simply using machine learning to teach computers to do just this extract the sentiments out of our sentences now how does this work what is machine learning is simply like a function in math you give it one or many numbers and it spits out another in machine learning these functions are called models now these models are often neural networks.

Neural networks simulate the structures of our brains to set to get inputs and their associations to build models predicting future inputs now here's joe who's joe you might ask joe's our friendly neighbourhood machine. The learning model of course now we want joe to tell us whether.

E.g. the image is a tiger see joe's sad right now because he has no clue what to do here is where we train our model for joe to tell us what a tiger is and what a tiger isn't we as humans must need to first tell him what a tiger looks like and what it doesn't there's a slight problem here.

However, Joe doesn't see this image as we do the one thing joe can do however is interpret numbers so what we can do is give these images to joe as a list of numbers of RGB vectors for

each pixel now let me break that down RGB vectors are red green blue vectors for each pixel denoting the colour of each pixel.
The image now what effectively allows us to do is convert these images into numbers that are great joe can now understand what we're trying to give him he knows what he's trying to do now what we can do when given an unfamiliar image is turn this into RGB vectors give it to joe and hey what do you know he thinks it's a tiger he's 97 sure so this was the case with pictures with. Images but we're talking about sentences the thing is it's the same how do we turn words into numbers now some of you might be thinking let's just slap a number on each word and call it a day, but the thing is if we train our models using those vector inputs we'd run into a problem.

This method struggles to recognize the semantic similarity between words for example this method fails to recognize the similarity between a word such as love and liked as opposed to a low similarity between loved and hated here is where we run into one of the most fundamental concepts in sentiment analysis word vectors now what are verb vectors well they're exactly as they would seem they're vectors corresponding to each word much like the RGB vectors for each pixel now unlike the RGB vectors however these word vectors can span from 25 up to a thousand components now conveniently as these vectors are still simply a list of numbers.
They can be plotted on an n-dimensional space but for the sake of visualization and your brains let's reduce that down to two on this coordinate plane what word vectors allow us to do is to demonstrate and evaluate the relationships between words as distances between points now somewhere on this coordinate plane lion and cat would be near each other related by their Fellini while somewhere else on the plane.

Honda and Ford would be clustered together related by their car manufacturing status now this seems to be working great what's the problem well the problem comes in when we add-in more words what would we do about the word jaguar now it is a feline so does it go somewhere near cat and lion no but it is a car manufacturer somewhere in the middle, we have no clue we've run into the dilemma which makes it possible for word vectors to be multi-dimensional by adding more vectors by adding more dimensions to these vectors.

We're able to express the relationships between words in the English language with more nuance great now we have these word vectors and we can associate them to the words in our sentence converting them to numbers that joe loves theoretically now we can feed these numbers to joe and joe will now be able to predict the sentiment of any sentence we give it so naturally.

The company decided to put that to the test where would I get my data to train this model well after some searching, The company decided to go with Kaggle’s twitter sentiment data set consisting of 1.5 million tweets manually categorized by either zero for negative or one for positive now you might be thinking hold up 1.5 A million tweets it's a lot of tweets and that is true it is a lot of tweets but just as you and the company will be better at identifying something.

The more examples you got of it joe can benefit from as much data as we can give it as for our word vector, however, went with Stanford university's gloves and for global vectors now this word vector set was pre-created which means that these researchers had to go through thousands and thousands of sentences look at instances for each word and evaluate their context to create word vectors for each word there was one more step the company had to take before they could train joe however let's look at this tweet stopped at McDonald’s for lunch.

I’m excited nuggets now if we fed this right to our model, we'd see a problem see us as humans can see through the Twitter clutter can see through the various distractions in this tweet but for joe needs a bit of help and that's why we need to clean this data set show you what they mean the first things to go were punctuation.

Along with punctuation when Twitter artefacts such as mentions hashtags and links. The second went is what recalled in natural language processing as stop words such as if I and that doesn’t necessarily add to the meaning of the sentence.

Now finally and arguably the trickiest part of cleaning this data set was how to deal with the internet slang now it's impossible to go anywhere on the internet without encountering some sort of abbreviation some sort of slang the tough part about dealing with this is that there is no set way of evaluating these words now to be fair common words such as law or lmao all have their entries in word vector sets such as glove now misspellings such as the one we see in this tweet can be caught with a spell check but some words and phrases do end up slipping through our fingers and that does make or break some sentiment analysis models.

Now regardless we've caught that and now we were able to condense that original tweet into the four words that you see on the bottom there now that we've cleaned our tweets. We can associate the word vectors in gloves to each word and now again we have our numbers to word association and can now train our model that's exactly what the company did.

Now how is a model you might be asking how good was it well luckily for the safety of the internet world as we know it wasn't that a successful model reached around a 60°/°accuracy which meant that it was able to correctly identify the sentiments of around 60 per cent of the sentences that it got, however, considering that this is a the problem that has yet to be solved this number is a sign of hope for things to come now throughout this talk, you might have been asking yourself why do we care who asked Andy what's next and I'm here to tell you this it's true this technology is bringing us ever so closer to our inevitable robot overlord world but I still believe that this technology is imperative and essential to our technological development for the benefits that it can provide currently the applications of sentiment analysis are purely commercial we see the movie producers using sentiment analysis to evaluate audience feedback on their recent project.

We see corporations including this technology to assess how consumers are reacting to their products but in the future, as this technology gets better we can see that this technology can be applied to a myriad of problems for example sentiment.

The analysis could be used to provide help for people with mental health issues many people with these issues find refuge on the internet and so with this technology we'll be able to provide help for people that might have been reluctant to seek it furthermore, this technology could be used to gauge radicalism on the internet as the internet has become a hub for radicalization we can see that this technology can be used by governments to make the internet safer for us all and hey if that didn't reach all of you then maybe our phones can become a therapist one day.

1.2 What Is Marketing ?

Marketing is a major business function that's made up of a bunch of other parts, one of which of course is advertising, but there's also a ton other like PR, market research, social media, content marketing, search engine marketing or search engine optimization, pricing and pricing psychology, copywriting and one of the personal favourites, direct response or direct response marketing and quite a few more. So, saying that marketing is just advertising and kind of like saying that finance is just taxes or HR is just hiring people or legal is just not getting sued. Sure, these are all important but there just one piece of the entire puzzle.

Therefore, if you're just getting started in marketing can seem pretty overwhelming which is why my suggestion is to find an area that you find interesting, fascinating, like say social media or content marketing, start there really get some roots and branch out later. Alright, so now that we've got that covered, what exactly is marketing? Well, as I've just covered, you know that marketing is made up of a bunch of different sub-segments things like content marketing, email marketing and social media, all the things we talked about.

But what exactly is the nature of marketing, which sounds like some kind of documentary "The Nature of Marketing" - on this week's episode of The Nature of Marketing - Well, one of the first things you learn in any first-year marketing course is the four P´s of marketing: product, price, place and promotion. The product being the details around whatever product or service is being sold. Price is the price, kind of obvious, but there's a lot more that goes into it than just that.

The place is where the products being sold and promotion, we're just kind of fun stuff. This is essentially how you're going to sell more of the product and what you're going to do to get the word out about your service and while all of that is accurate and true and the four P´s do make up a part of marketing, well, I prefer a simpler easier to understand the definition. Marketing in its most basic sense is communicating value to your customers.

It's essentially answering the question of your customers "Why should I care?" Marketing helps people solve their problems by clearly defining and delivering solutions and explaining the benefits of the solution, so they can get better results.

Marketing is about connecting with people, understanding their pains and their problems and their frustrations, making them feel understood so that you can position your business as the solution to their problems and essentially just make them feel better and marketing is a powerful force. As anyone that's been in business for any length of time can tell you, it's not always the best product or service that wins, in fact, it rarely is, rather it's the product or service with the best marketing.

Like it or not, that's just kind of how it works which is why having good marketing is just so important. So, my favourite definition of marketing is that it's communicating value to your customers but there's another side entirely that's rarely talked about but still equally powerful and that's creating value for your customers. You see, marketing can create value and your marketing in and of itself can be valuable.

An example could be a blog post that helps someone out, even before any money has changed hands or an advertisement someone sees that brightens their day and makes them laugh whether they choose to buy or not or the way that a product or service is delivered, the packaging let's say which is so luxurious and over-the-top that it makes the customer feel special just for having purchased it. You see, economics assumes that people make buying decisions rationally, logically and with perfect information, but this is rarely, pretty much never the case.

We as humans are emotional, often highly illogical and rarely have the full set of facts when making any decision, which explains why marketing is so important and so powerful. After all, if we made all of our purchasing decisions based solely on logic and utility and the value we would get from these products, well, the entire luxury goods market wouldn't even exist but not all marketing is created equal. You see, when it comes to marketing like when it comes to pretty much anything, there's good marketing and there's bad marketing. Bad marketing is all the reasons that marketing gets such a bad reputation. It's the stuff that looks cheap, makes people feel dirty, even just looking at it and promotes bad products or bad services that helps nobody.

It's the fake countdown timers you sometimes see on websites the going-out-of-business sales that never end and the pushy promotion of useless products. That's bad marketing. Good marketing, on the other hand, helps customers achieve their goals, makes them feel better about themselves and has the power to truly change the world. So, my question to you is: What kind of marketer do you want to be? That that was a rhetorical question.

If you say the bad guy, I got nothing for you but if you say the good kind and I know you did, then make sure to check out this video right here which I've got linked up on the page which is going to give you even more practical and effective marketing strategies to help take your business and your marketing to the next level and way beyond that.

1.3 Ai In Marketing

Many companies - and thus the marketing teams that support them - are rapidly adopting intelligent technology solutions to encourage operational efficiency while improving the customer experience. Through these platforms, marketers are ready to gain a more nuanced, comprehensive understanding of their target audiences. The insights gathered through this process can then be used to drive conversions while simultaneously easing the workload for marketing teams. What is AI (AI) Marketing? AI marketing uses AI technologies to make automated decisions supported by data collection, data analysis, and additional observations of audience or economic trends which can impact marketing efforts. AI is usually utilized in marketing efforts where speed is important.

AI tools use data and customer profiles to seek out out the thanks to best communicate with customers, then serve them tailored messages at the right time without intervention from marketing team members, ensuring maximum efficiency. for several of today's marketers, AI is used to strengthen marketing teams or to perform more tactical tasks that require less human nuance.

AI marketing use cases include: data analysis natural language processing media buying automated deciding content generation real-time personalization Components of AI in Marketing It's clear that AI holds an important role in helping marketers connect with consumers. the subsequent components of AI marketing structure today's leading solutions that are helping to bridge the gap between the expansive amounts of customer data being collected and thus the actionable next steps which can be applied to future campaigns: Machine Learning is driven by AI, and it involves computer algorithms which can analyse information and improve automatically through experience. Devices that leverage machine learning analyses new information within the context of relevant historical data which will inform decisions supported what has or hasn't worked within the past.

Big Data and Analytics

The emergence of digital media has brought on an influx of big data, which has provided opportunities for marketers to understand their efforts and accurately attribute value across channels. This has also led to an over-saturation of data, as many marketers struggle to determine which data sets are worth collecting.

AI Platform Solutions

Effective AI-powered solutions provide marketers with a central platform for managing the expansive amounts of data being collected. These platforms have the ability to derive insightful marketing intelligence into your target audience so you can make data-driven decisions about how to best reach them. For example, frameworks such as Bayesian Learning and Forgetting can help marketers gain a clearer understanding of how receptive a customer is to a specific marketing effort.

Challenges for AI Marketing

Modern marketing relies on an in-depth understanding of customer needs and preferences, and then the ability to act on that knowledge quickly and effectively. The ability to make real-time, data-driven decisions has brought AI to the forefront for marketing stakeholders. However, marketing teams must be discerning when deciding how to best integrate AI into their campaigns and operations. The development and use of AI tools are still in the early stages. Therefore, there are a few challenges to be aware of when implementing AI in marketing.

Training Time and Data Quality

AI tools do not automatically know which actions to take to achieve marketing goals. They require time and training to learn organizational goals, customer preferences, historical trends, understand overall context, and establish expertise. Not only does this require time, but it also requires data quality assurances. If the AI tools are not trained with high-quality data that is accurate, timely, and representative, the tool will make less than optimal decisions that do not reflect consumer desires, thereby reducing the value of the tool.

Privacy

Consumers and regulating bodies alike are cracking down on how organizations use their data. Marketing teams need to ensure they are using consumer data ethically and in compliance with standards such as GDPR, or risk heavy penalties and reputation damage. This is a challenge where AI is concerned. Unless the tools are specifically programmed to observe specific legal guidelines, they may overstep in what is considered acceptable in terms of using consumer data for personalization.

Getting Buy-In

It can be difficult for marketing teams to demonstrate the value of AI investments to business stakeholders. While KPIs such as ROI and efficiency are easily quantifiable, showing how AI has improved customer experience or brand reputation is less obvious. Marketing teams need to ensure they have the measurement abilities to attribute these qualitative gains to AI investments.

Deployment Best Practices

Because AI is a newer tool in marketing, definitive best practices have not been established to guide marketing teams' initial deployments.

Adapting to a Changing Marketing Landscape

With the emergence of AI comes a disruption in the day-to-day marketing operations. Marketers must evaluate which jobs will be replaced and which jobs will be created. One study suggested that nearly 6 out of every 10 current marketing specialist and analyst jobs will be replaced with marketing technology.

How to Use AI in Marketing

It's important, to begin with a thorough plan when leveraging AI in marketing campaigns and operations. This will ensure marketing teams minimize costly challenges and achieve the most value from their AI investment in the least amount of time.

Before implementing an AI tool for marketing campaigns, there are a few key factors to consider:

Establish Goals

As with any marketing program, it is important that clear goals and marketing analytics are established from the outset. Start by identifying areas within campaigns or operations that AI could stand to improve, such as segmentation. Then establish clear KPIs that will help illuminate how successful the AI-augmented campaign has been – this is especially important for qualitative goals such as “improve customer experience.”

Data Privacy Standards

At the outset of your AI program, be sure that your AI platform will not cross the line of acceptable data use in the name of personalization. Be sure privacy standards are established and programmed into platforms as needed to maintain compliance and consumer trust.

Data Quantity and Sources

To get started with AI marketing, marketers need to have a vast amount of data at their disposal. This is what will train the AI tool in customer preferences, external trends, and other factors that will impact the success of AI-enabled campaigns. This data can be taken from the organization’s own CRM, marketing campaigns, and website data. Additionally, marketers may supplement this with second and third-party data. This can include location data, weather data, and other external factors that may contribute to a purchasing decision.

Acquire Data Science Talent

Many marketing teams lack employees with the necessary data science and AI expertise, making it difficult to work with vast amounts of data and deliver insights. To get programs off the ground, organizations should work with third-party organizations that can assist in the collection and analysis of data to train AI programs and facilitate ongoing maintenance.

Maintain Data Quality

As machine learning programs consume more data, the program will learn how to make accurate, effective decisions. However, if the data is not standardized and free of errors, the insights will not be useful and can actually cause AI programs to make decisions that hinder marketing programs. Prior to implementing AI marketing, marketing teams must coordinate with data management teams and other lines of business to establish processes for data cleansing and data maintenance. When doing so, consider the seven essential data dimensions:

- Timeliness
- Completeness
- Consistency
- Relevance
- Transparency
- Accuracy
- Representativeness

Selecting an AI Platform

Selecting the right platform or platforms is a crucial step in getting an AI marketing program off the ground. Marketers should be discerning in identifying the gaps that the platform is trying to fill and select solutions based on capabilities. This will revolve around the goal marketers are trying to achieve – for example, speed and productivity goals will require different functionality than tools used to improve overall customer satisfaction with AI. One thing to keep in mind when selecting a tool is the level of visibility you will need regarding why an AI platform made a certain decision. Depending on the algorithm in use, marketing teams may get a clear report on why a certain decision was made and which data influenced the decision, while algorithms working on a more advanced level with deep learning may not be able to give as definitive reasoning.

Benefits of Leveraging Artificial Intelligence in Marketing

There is a myriad of use cases for AI in marketing efforts, and each of these use cases yields different benefits such as risk reduction, increased speed, greater customer satisfaction, increased revenue, and more. Benefits may be quantifiable (number of sales) or not quantifiable (user satisfaction). There are a few overarching benefits that can be applied across AI use cases:

Increased Campaign ROI

If leveraged correctly, marketers can use AI to transform their entire marketing program by extracting the most valuable insights from their datasets and acting on them in real-time. AI platforms can make fast decisions on how to best allocate funds across media channels or analyse the most effective ad placements to more consistently engage customers, getting the most value out of campaigns.

Better Customer Relationships & Real-Time Personalization

AI can help you deliver personalized messages to customers at appropriate points in the consumer lifecycle. AI can also help marketers identify at risk customers and target them with information that will get them to re-engage with the brand.

Enhanced Marketing Measurement

Many organizations have trouble keeping pace with all of the data digital campaigns produce, making it difficult to tie success back to specific campaigns. Dashboards that leverage AI allow for a more comprehensive view into what is working so that it can be replicated across channels and budgets allocated accordingly.

Make Decisions Faster

AI can conduct tactical data analysis faster than its human counterparts and use machine learning to come to fast conclusions based on campaign and customer context. This gives team members time to focus on strategic initiatives that can then inform AI-enabled campaigns. With AI, marketers no longer must wait until the end of a campaign to make decisions but can use real-time analytics to make better media choices.

1.4 SEVEN EXAMPLES OF ARTIFICIAL INTELLIGENCE IN MARKETING

AI is being used in marketing initiatives in a multitude of use cases, across a broad array of industries including financial services, government, entertainment, healthcare, retail, and more. Each use case offers different results, from improvements to campaign performance, to enhanced customer experience, or greater efficiency in marketing operations.

There are numerous ways businesses can take advantage of machine learning to create a more comprehensive marketing plan. Consider the following:

a) **Bidding on Programmatic Media Buys**

A problem that marketing teams often encounter is deciding where to place advertisements and messaging. Marketing teams can create informed plans based on user preferences, but these teams are not flexible or agile enough to alter the plan in real-time based on the latest consumer information. AI is being used by marketers to mitigate this challenge through programmatic advertising. Programmatic platforms leverage machine learning to bid on ad space relevant to target audiences in real-time. The bid is informed by data such as interests, location, purchase history, buyer intent, and more. This enables marketing teams to target the right channels at the correct time, for a competitive price. Programmatic buying exemplifies how machine learning can increase marketing flexibility to meet customers as their needs and interests evolve.

b) **Select the Right Message**

Across channels, different consumers respond to different messages – some may resonate with an emotional appeal, some humour, others logic. Machine learning and AI can track which messaging consumers have responded to and create a more complete user profile. From there, marketing teams can serve more customized messages to users based on their preferences. For example, Netflix uses machine learning to understand the genres a certain user is interested in. It then customizes the artwork that the user sees to match up with these interests. On the Netflix Tech Blog, they explain how they use algorithms to determine which artwork will most entice a viewer to watch a certain title, saying:

"Let us consider trying to personalize the image we use to depict the movie Good Will Hunting. Here we might personalize this decision based on how much a member prefers different genres and themes. Someone who has watched many romantic movies may be interested in Good Will Hunting if we show the artwork containing Matt Damon and Minnie Driver, whereas, a member who has watched many comedies might be drawn to the movie if we use the artwork containing Robin Williams, a well-known comedian."

Credit: Netflix Tech Blog

When AI and machine learning are used, these platforms can gather valuable data on customers that allow marketing teams to increase conversion rates and improve the customer's experience. Marketing teams can then analyse all of this data to create a more nuanced view of the customer, even considering additional factors such as if a user would have watched a title regardless of the image, and how that plays into future messaging.

c) **Granular Personalization**

A highly granular level of personalization is expected by today's consumers. Marketing messages should be informed by a user's interests, purchase history, location, past brand interactions, and a host of other data points. AI helps marketing teams go beyond standard demographic data to learn about consumer preferences on a granular, individual level. This helps brands create curated experiences based on a customer's unique tastes. For example, Spotify uses AI to create customized playlists based on what a customer has listened to in the past, current hits across genres, and which music is being talked about. It uses these datasets to create customized playlists for users and to create genre playlists based on artists that appear in conversation, in articles, etc. This has helped Spotify to become a top streaming service and emphasize customer experience through personalization.

Another trend based on AI-enabled personalization is atomic content. Here, AI learns customer preferences and pulls pieces from a library of content to create a customized email or offer for a client featuring relevant images, videos, or articles.

d) **Chatbots and Conversational Experiences**

With the development of natural language processing through AI, chatbots are now being used to augment customer service agents. Customers with more basic queries can refer to chatbots which will give immediate, accurate answers. They will be able to leverage past questions and historical data to deliver personalized results. This gives time back to customer service agents to work on complicated requests that need more human nuance.

e) **Predictive Marketing Analytics**

With so much data coming, marketing teams are having a hard time actually deriving insights from it. AI allows marketing teams to make the most of this data using predictive analytics, which leverages an assortment of machine learning, algorithms, models, and datasets to predict future behaviour. This can help marketing teams understand the types of products a consumer will be looking for and when – allowing them to position campaigns more accurately.

For example. Amazon uses predictive analytics to suggest products to consumers based on past purchases and behaviours, increasing conversions and customer satisfaction. AI can also be used to help marketing teams more accurately track attribution, allowing teams to see which campaigns contributed most to ROI.

Credit: Woo Commerce

f) **Marketing Operations**

Another key use case for AI in marketing is to increase efficiency across various processes. AI can help to automate tactical processes such as the sorting of marketing data, answering common customer questions, and conducting security authorizations. This allows marketing teams more time to work on strategic and analytical work.

g) **Dynamic Pricing**

AI can help make brands more competitive by enabling dynamic pricing. AI platforms can suggest optimal prices for products in real-time by evaluating huge quantities of historical and competitive data. This strategy has been especially effective in retail. It allows brands to adjust prices to reflect demand for certain products, boost sales, and edge out the competition.

1.5 PREDICTIONS AND TRENDS FOR AI MARKETING

While AI is still largely new to the marketing space, it promises to only grow in popularity. There are a few AI trends marketers will see over the next few years and should begin to adapt to:

AI is Growing:

- Gartner has predicted that by 2022, AI will replace about 33% of data analysts in marketing.
- Tech giants realize the benefits and potential for AI. In 2016, they were already spending on average between $20-$30 billion. 90 per cent of this budget was focused on deployment and research.
- Additionally, in 2020 Gartner predicted that more than 40% of data science tasks will be automated

Teams Will Scale Through AI

Marketing teams will be put under increased pressure to demonstrate marketing value and ROI to executive stakeholders. Teams will leverage AI solutions to drive these targets and better allocate funds towards successful campaigns and provide the marketing metrics that demonstrate the value of campaigns.

Marketing Leaders Who Don't Leverage AI Will Be Replaced by Those Who Do

According to Gartner, those responsible for marketing insights will no longer be as competitive in this changing marketing landscape. The majority of those surveyed by Gartner employs AI solutions in their marketing strategy or are planning to. Only 13 per cent do not see a use for it in the next three years.

1.6 Context-Based Sentiment Analysis on Amazon Product Customer Feedback Data

The sentiment is a mindset, opinion, or decision that a sensation trigger. Sentiment Analysis is the procedure of using text analytics to mine different sources of data for opinions. It is also referred to as opinion mining and explores the emotions of a person about certain products. The internet is indeed rich with resources relating to sentiment analysis such as feedback of a product on an e-commerce website and tweets sent by an individual on Twitter social networking site. Context-based sentiment analysis is an exciting tool that continues to evolve as Artificial Intelligence and Machine Learning become more sophisticated. It is a precise science that can be used to aid new product development as well as for marketing and public relations. The analysis will be able to process product data at scale in an efficient way. Customer behaviour can be analysed, and it can be found out what products are in demand and which products are not doing well. This model plans to perceive the properties of products and stores which play a pivotal role in growing sales. Word Sense Disambiguation is used to distinguish two similar sounding words from each other. Customer behaviour can be analyzed and it can be found out what products are in demand and which products are not doing well. Through this, it is proposed to find out sentiments involved in each product whether they be positive or negative, based on which sentiment analysis will be carried out. Naïve Bayes and Support Vector Machine Algorithms have been used for classification. The model will be robust with high precision and straightforward addition of attributes from various sources to the set.

1.6.1 Introduction

Recently, sentiment analysis of product customer feedback, an application topic in text mining, and computer linguistics research has become very popular. We would like here to analyse the association of product feedback from Amazon with customers' ratings. We use conventional algorithms, including analysis by Naïve Bayes Algorithm (NBA) as well as Support Vector Machine (SVM) algorithm. The core of sentiment analysis is the function of the classification of texts, and different words contribute differently. For current sentiment analysis research, distributed word representation is most often used. Moreover, distributed word representation takes only semantic word information into consideration, but does not take the meaning of the language into account. One is able to improve one's interpretation of these algorithms by analysing the results. These may also be an alternative to other forms of rating fraud methods. The fixed aspects of each product and store will be described.

Analysis will be able to process product data at scale in an efficient way. It builds on strategies for the recovery of knowledge and defines key terms that represent feelings.

The contexts will be used to generate supervised learning features. Project proposes to build a predictive model for acquired sales data and predict sales of each product at a specific store. Understand attributes of stores and products which will play a crucial role in intensifying sales. Most social networking sites publish their Application Programming Interface (API).

Based on an analyst's viewpoints, encouraging researchers and developers to collect and analyse data. Users can publish their own content through various blogs, social networking channels, and websites, based on their point of view. Nonetheless, there are certain limitations in such online data that may impede the sentiment analysis process. Some flaws include that people may post feedbacks whose quality may not be guaranteed. Like online Spammers will post spam on the forum. Some spam is obsolete while some spam can have fake opinions.

Word Sense Disambiguation (WSD) is a challenge in Natural Language Processing (NLP) to decide the "sense" of a word by using the said word in a specific context, an often-unconscious mechanism in people. WSD is a problem of conventional classification: If a word and its possible senses are specified in a dictionary, classify the word into one or more sensory classes in context. Description characteristics (adjacent words in this case) justify the classification.

WSD is a Resolution process whose word meaning in a given context is triggered by the use of the term, a mechanism that does not seem to be usually present in individuals.

Some challenges we try to address include a sentiment analytical system that relies upon sentiment-based tokens for its analysis. For comments that absolutely have underlying emotions, the program may not work well. These underlying emotions are known as inherent sentiments. An inherent sentiment is generally expressed by neutral terms, making it quite difficult to judge its sentiment polarity.

Like an instance, a phrase like “This Item is described as” has a challenge, this positive feedback consists of only neutral terms, which occur intermittently. With these limitations in mind, this model aims to solve these problems. In order to increase the categorization of study levels, more Context-Based Sentiment Analysis … 517 characteristics will be extracted and grouped into vectors. To perceive the existence of such an opinion within the reach of a certain product the subsequent action is for the emergence of such a sensation in the sense of a substance to be recognized. This paper contains data from a variety of customer feedbacks obtained by Amazon.com from February to April 2014. Every product analysis must first be tested and put under scrutiny before it can be made available. Subsequently, each analysis should eventually have a ranking, which can be used as the basic truth. To explain the ground truth, in case of positive, negative, or neutral thought, it is like a sticker of certain opinions.

The rating is in fact based on the system is positive, negative, or neutral. Customer behavior can be analyzed and it can be found out which products tend to do well and which do not. Algorithms such as NBA and SVM will be implemented to detect pivotal terms in text and analyze its context. Use detected terms to generate supervised learning features. It will improve sales based on sentiments portrayed for each product. This approach aims to distinguish customers’ positive and negative opinions about various products and create and design supervised research to polarize large quantities of feedback. Our repository contains customer feedbacks and assessments, which we obtained from Amazon Consumer Feedback. On the basis of this, we extracted the repository features and developed several supervised models. Not only do conventional algorithms like NBA, SVM, and K-nearest neighbor (KNN) come in these models.

But a few other algorithms are used which will be talked about in Related Work. The precise nature of these models is contrasted and the biased behaviors towards attitudes are better understood. 2 Related Work Further research has been done in previous years to understand the significance of text resources.

The studies published in previous years on sentiment analysis have increased, as can be seen from the Knowledge Web statistics. Sentimental analyzing or mining of opinions is one of the themes of this study, based on a bunch of texts, we can research people's opinions, perceptions, behaviours, feelings, issues, incidents, topics, and their characteristics. This approach is used in different ways. Like an instance, corporations always want to gain feedback and perceptions about their products and services for the public or customers. Prospective clients would also like to learn current users' thoughts and feelings before using a service or purchasing a product. Eventually, analysts use this knowledge to examine market trends and customer perceptions in-depth that may lead to better bond market prediction.

Nevertheless, to look up, track webpages, and spread information about what they contain continues to be a daunting challenge thanks to various sites augmentation. Generally speaking, each site contains a lot of thought, not always easy to decode in lengthy posts. In general, it is arduous for the common human reader to locate the related webpages and to summarize the information and opinions.

In fact, it actually is complicated and difficult to teach a sarcasm recognition system,518 C. Sindhu et al. since machines are currently not yet able to think like people. In sentimental analytics, profound neural networks are also common. Some researchers have utilized a convolutional network to mark the semantic position so that task-specific design engineering cannot be overtaken.

Instead, the authors suggested that Recurrent Neural Networks (RNN) be used to improve composition in tasks such as sensing feelings. We want to use all conventional algorithms in this model including NBA, KNN, SVM, and tricks for profound learning. As we compare the precision of these models, we want to better know how these calculations function in activities such as the study of emotions.

A considerable amount of papers were surveyed for this project. The scope of this survey includes all there is in an online supermarket which is an abundance of items, it is a matter of what to choose in the end. Though not all items belong to the same crux some items may not be selling so much. Due to which it might be a waste in buying them.

The applications of sentiment analysis include product feedback, customer support, and reputation management. Through our specific model, we can identify which items in an online supermarket are the ones that sell the most and which are the ones that are not doing well in sales. A rather important usage of our project includes word sense disambiguation. Work hereafter includes testing the categorization scheme using other datasets. Most of the datasets used in the survey papers are either from Amazon or Twitter. The commonly used algorithms are Naïve Bayes, Support Vector Machine, Latent Dirichlet Allocation (LDA), Rule-based Sentiment Analysis (RBSA) , bidirectional long short term memory (BiLSTM) and Random Forest Algorithm (RFA).

The discussed the algorithms, methodology, measures, and demerits of various papers used for the survey. Word Sense Disambiguation According to a word and its potential meanings, as defined in a dictionary, a word instance must be categorized into a class or a number of its meaning classes is the definition of WSD. The context attributes (like the adjacent terms) produce the evidence for data. Such as a mouse can be both the rodent and the electronic mouse for computers.

Dataset The data used for this study is a compilation of Amazon product feedbacks. The four main categories include fashion, books, electronics, and home decor. Each feedback is divided into (a) ID of Consumer, (b) ID of product (c) Ratings (d) time schedule of feedback (e) whether the feedback was helpful or not. There were over 5 million feedbacks for 18,000 products. Dataset chosen is quite large and contains many

reviews as given above. Having such a large dataset has its advantages as there is a large range between the worst review and the best review. In these cases having more reviews is a boon as we can look at the larger picture. we have talked about datasets that were frequently used in previous papers. Sentiment Sentence Extraction In this data, all idiosyncratic content was derived for future study. All feelings are contained in the subjective material. At least one positive or negative word will be in a sentiment phrase. Sentences are initially first tokenized into English words which were isolated. The full sentence is used as sometimes taking lone words doesn’t tell us the full picture. Taking the full sentence makes us understand the context of the sentence or paragraph.

Sentiment Phrase Identification Adjectives and verbs are words that through negative prefixes, can express opposing feelings. Like instance, the following statement contained in the analysis of an electronic device “The built-in camera on this phone has its uses but so far nothing great.” The word, “great” is a positive word.

Nonetheless, a sentence like “nothing great” depicts unpleasant experiences mostly. Those sentences, therefore, must be classified. Another example we can use is when one says that the doll looks pretty ugly. The word is pretty is a positive word but ugly is a negative word. Though when one uses both at once the overall result of the word is negative.

Word Sense Disambiguation This decides what meaning of a word in a context is triggered by using the word. In one or more of its sense types, an incident will be classified. Features like adjacent terms help in distinguishing. Such as a mouse can mean a computer mouse as well as a mammal. Though when someone types mouse cage we know it means the animal. While if someone types a keyboard with a mouse we know they mean the computer mouse. Neighbouring words help in classifying in this case.

The workflow for the model and steps involved in the WSD Methodology. Dataset Amazon’s customer product feedback data used. Sentiment Sentence Extraction Sentence tokenized into English words Sentiment Phrase Identification Identifies positive and negative phrases, Sentiment Score Computation Scores are graded on range 1-5 stars. Feature Vector Generation Each entry of data to be transformed into a vector. Sentiment Polarity Categorizati on Ratings Given from 1-5 stars Result Elucidation Which products are doing well and which are not? Word Sense Disambiguati on Word classified due to adjacent words Word sense disambiguation methodologyContext-Based Sentiment Analysis … 523 3.5 Sentiment Score Computation Scores are graded on range 1–5 stars. A rating of 1–2 stars is taken as negative; 3 stars is neutral and 4–5 stars is taken as positive.

a. star is generally known as poor reviews which means that the product itself was abysmal. (2) stars mean average nothing special.
b. stars are good.
c. is for a very good product while.
d. stars are for exceptional products.

Feature Vector Generation Features are usually taken from sentiment tokens whose information is drawn out from the training dataset. A vector shall not have a copious load of features or else it will suffer from curse of dimensionality which makes available data become sparse while the volume of the space increases at high speed. The vector function consists of four components: 2 hashes centred on a binary array, an average sentiment rating, and a basic truth label.

1.6.2 Sentiment Polarity Categorization

When one goes through the procedure of polarity categorization of sentiments it can be found to be bipartite: given that categorization happens at both sentence and review level. When a sentence is given, the target of categorization at sentence-level is to segregate into negative or positive about the sentiments conveyed.

Analysis of the product reviews. The number of product reviews has been compared with product reviews with total words used. For this method, training data require fundamental real-world tags that specify the positive or negative value of the given sentence.

Nonetheless, ground truth tagging can become a genuinely difficult issue, owing to the huge number of data that is present. Considering manual tagging of each of the sentences is quixotic, a machine tagging method is then chosen for an elucidation. If a greater number of negative tokens are present than positive ones, sentences will get tagged as negative, and conversely having a greater number of positive tokens will render the sentence as positive.

Result Elucidation From the results obtained we can identify which product is selling more by the amount of positive reviews for them and the ones which are not doing well albeit the ones with the negative reviews can be looked into and we can find out what are the problems with these products and try to improve upon them.

So that sales of the product gaining negative reviews increases. In our model we have used two algorithms; NBA and SVM. The NBA works like suppose a set of training details occurs, such as T, where the n-dimensional function variable is each different.

Thus, a variety of training data are available. Y = ya, yb … yz, implies z measurements contrived z characteristics tuple.

Then we infer m classes where ca, cb, …, cn. According to tuple m, the classifier forecasts m resides in Ci providing: $P(cj|m) < P(ci|m)$, where $j, i \in [1, m]$ and $j \neq i$. $P(ci|y)$ will get calculated as: $P(C\,j|Y) = \Pi\, P(y|C\,j)$

For the linear and nonlinear data set, SVM conducts data classification. SVM is a conventional algorithm for the classification of linear and nonlinear data in the Machine Learning (ML).

Given training data with binary outputs, SVM tries to find a hyperplane as the decision-making field, to maximize the differentiation of positive and negative samples. SVM converts the data into a bigger dimension and solves the conundrum by locating a linear hyperplane if data is not linear.

The kernel used for such nonlinear data is called Gaussian Radial Basis function (RBF). SVM was used to classify each feedback to a label of “positive”, “neutral”, or “negative”. Once data can be differentiated linearly, the optional boundary separates data from one class to another from a vector kernel via SVM. A linear kernel can be entered as a logical one: $V * X + d = 0$ (2) X works as a training tuple, V is a function of weight and V = va, vb … vn, while d works as scalar.

The Relation between the two classifiers mainly SVM and NBA has been compared with the three evaluation criteria mainly precision,Context-Based Sentiment Analysis … 525

Relation between three evaluation criteria and two classifiers SVM NBA Precision 0.81 0.84 Recall 0.75 0.823 F1 0.78 0.831 0.7 0.75 0.8 0.85 Precision Recall F1 Recall, and F1 measure.

The conundrum essentially converts to a reduction in order to improve the kernel, calculated decidedly as: Σaiyixi (3) where ai is numeric and zi is labelled with vector help, Xi. It concurs: if zi = 1 then vixi ≥ 1; if zi = −1 then vixi ≥ −1.

If the data is inseparable linearly, to convert the data into a greater dimension, SVM allows the use of nonlinear simulation. Subsequently, it solves the conundrum by locating a linear kernel. Functions to achieve such conversions are known as kernel.

The kernel attribute that our work promotes: the Gaussian radial base function (GRBF): J (Y k, Y j) = e|−b|Y k−Y j||2 (4) yk is a tuple test and β is a different parameter with a predefined value while yj is a support vector.

1.6.3 Discussion

This model proposes to build a predictive model for acquired sales data and predict sales of each product at a particular store. Customer Behavior can be analyzed and it can be found which products are in demand and which products are not doing well. Through this, it will try to improve sales based on sentiments portrayed for each product. System should be robust with high precision and applications from various sources easy to add.

Nonetheless, this analysis still has a few constraints. One downside is that the scheme for evaluating our emotions in this analysis concentrates on the nature of stimuli, the program may be inadequate for only unconscious emotional analysis. A definition is typically implicitly conveyed in neutral terms, which allows deciding its polarity complicated.

A sentence like, for instance, “Item described is”, positive evaluations often have neutral terms only. The future works on526 C. Sindhu et al. solving these issues in the context of these restrictions. Specific functionalities should be clearly derived and organized into functional vectors to increase the grouping of feedback. The next phase in the area of intrinsic sensation research is to determine that it exists in the continuum of a specific product. Future analysis will include the validation of our categorization framework with other data sets.

1.7 Subjectivity Detection for Sentiment Analysis on Twitter Data

1.7.1 Abstract

With the quick increment in the quantity of web clients, the Internet has an enormous measure of data produced by the clients. Many people share their views regarding a topic on social media platforms such as Facebook and Twitter and give their feedback or review about a product on e-commerce web sites such as Amazon and Flipkart which leads to a huge amount of data. The identification of subjective statements from the data is known as subjectivity detection. To automate the analysis of such data, sentiment analysis is used. The aim is to find the opinionative data and classify it according to its polarity, i.e. positive, negative or neutral feedback, known as sentiment classification and then analysing it which is known as sentiment analysis. However, before performing sentiment examination, the information is exposed to different pre-processing procedures which finally give the desired optimized output. This allows us to get to know about the public's mood or opinion about a particular topic. This summarization helps the concerned organization or public to improve their product or service based on the feedback received.

1.7.2 Introduction

With the enormous development of number of web users, the quantity of tweets every day on Twitter has likewise expanded definitely. Mining of sentiment from these tweets is helpful for the organizations and associations. For instance, it very well may be utilized as a sub-module in suggestion motors and so forth. Sentiment analysis is relevant mining of content that plans to group content into positive, negative and impartial. Sentiment analysis is an issue which incorporates different NLP sub-problems which are to be settled which incorporate mockery identification, element acknowledgement and subjectivity recognition and so forth.

Subjectivity detection has picked up significance with the quick development of data created via web-based networking media which requires the identification of subjective data (opinion) and objective data (fact). Subjectivity identification can be substantially more testing than polarity recognition; however, it has been underexplored because of the supposition that most of the information via web-based networking media is objective.

For example, “My favourite pair of shoes is sold out” is an objective statement because it is a fact, and “This pair of shoes is very good” is a subjective statement because it tells about the opinion of the person.

Subjectivity detection helps to get to know about the opinion of the users about a particular product and topic which indeed helps the concerned organization or public to improve their service or product dependent on the feedback received.

Related Work Systematic literature audit process is used in this overview. First, researchers scanned for some related papers, research reports that are comprehensively worried about subjectivity identification or opinion mining from the content. Detection of user’s opinion and classifying its polarity, i.e. positive, negative and neutral, is known as polarity detection (PD). Previously done work in sentiment analysis was either knowledge-based or sentiment-based. But recently there have been various studies that utilize various machine learning techniques to classify the text. Supervised machine learning techniques are comparatively better than unsupervised machine learning techniques in performance, but it is expensive to acquire the huge amount of labelled data required for supervised learning, whereas it is comparatively easy and less expensive to acquire unlabelled data for unsupervised learning.

Numerous specialists are putting their endeavours to identify the best technique for subjectivity identification. Albeit, a portion of the algorithms give great outcomes such as support vector machine (SVM), maximum entropy, Naïve Bayes and so forth; however, no technique can resolve every one of the difficulties. The vast majority of the researchers detailed that SVM has high precision than different algorithms.

The different algorithms and the data sets used in different papers have been mentioned in Subjectivity Detection Approach A framework was implemented in which the first step is to classify messages as subjective and objective tweets (subjectivity detection). The second step is to classify the subjective tweets into positive and negative (polarity detection). Usually, a purely objective sentence does not carry any sentiment, and a purely subjective sentence usually tends to lean towards a positive or a negative sentiment.

Though there are a few exceptions, for example, “The food made me sick” is an objective sentence with sentiment, and “I believe he came to the college yesterday” is a subjective sentence with no sentiment. Classifying a sentence as objective or subjective is done by using libraries such as TextBlob created by Steven Loria, and tools such as Opinion Finder (http://mpqa.cs.pitt.edu/opinionfinder/). A filtering mechanism is also implemented to have a control on the level of subjectivity in the training set by using a subjectivity threshold. Another approach that was previously implemented, which was later scraped due to inconsistencies in the results and lack of accuracy, was from a given tweet, we map its POS using a POS dictionary (http://wordlist.sourceforge.net/pos-readme). POS tags are used to indicate sentiment tagging in a tweet. Objective messages usually consist of adjectives or interjections. We get the prior subjectivity and polarity

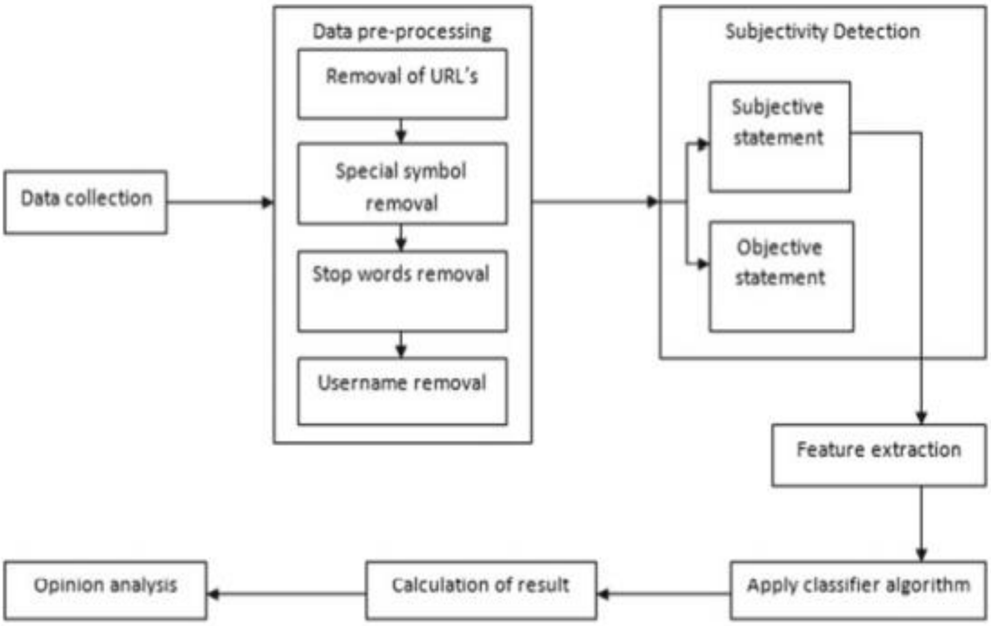


Fig. 1 Subjectivity detection for sentiment analysis on Twitter data

Table 2 Data set annotation scheme

Sentiment	Annotation
Positive	4
Neutral	2
Negative	0

Data Collection The data set that is used consists of around 1.6 million tweets for training and 5000 tweets for testing.

The tweets in the data set are categorized into positive, negative and neutral. The data set is very versatile and consists of various categories such as company, movie, location, person, product, event and misc. The emoticons were removed for the training and the test data (Table 2).

Data Pre-processing The data extracted from Twitter contains various contents which do not contribute to the sentiment of the user; therefore, it has to be first pre-processed. Pre-processing Subjectivity Detection for Sentiment Analysis on Twitter Data 471 [20] includes four basic steps—removal of URL, removal of special symbols, removal of stop words and removal of username.

In removal of URL, any kind of link which is tweeted by the user and does not contribute to the sentiment analysis is removed. Removal of special symbol step deals with removing various symbols which do not have any actual sentiment, e.g. full stop (.), punctuation mark (!) and so forth.

Stop words removal step removes the stop words, words such as a, the which do have no effect on sentiment analysis should be removed and the conversion of emoticons to its equivalent word. Finally, in the username removal step every user’s username starts with @ which has no effect on the sentiment analysis is removed, e.g. @username. 3.3 Subjectivity Detection As previously mentioned, the first step is to classify the tweet into subjective and objective and remove the tweets based on their subjectivity scorekeeping only the tweets having score higher than the specified threshold.

This step is introduced to achieve higher accuracy. The pre-processed data is taken and is classified into subjective or objective statement using a subjectivity classifier. All the tweets having a subjectivity score lesser than the specified threshold are filtered out, and the classifier is trained with only the remaining tweets. It is observed that as the subjectivity threshold is increased, significant amount of tweets gets filtered out.

Feature Extraction A data set contains numerous ascribes that add clamour to the data and influence exactness. The commotion likewise bit by bit expands the time required to assemble the model. Feature extraction basically combines ascribes into a reduced feature set. The selected features and their blend assume a significant job for identifying the sentiment of the text. Selection of features [22, 23] from the extracted features can possibly improve the arrangement exactness, restricted in on a key feature subset of opinion discriminators and give more prominent understanding into habitually happening ascribes and qualities. The extracted features focus on a document vector whereupon machine learning strategies are applied to group the extremity of the content utilizing the got document vector.

3.2 Data Pre-processing The data extracted from Twitter contains various contents which do not contribute to the sentiment of the user; therefore, it has to be first pre-processed. Pre-processingSubjectivity Detection for Sentiment Analysis on Twitter Data 471 [20] includes four basic steps—removal of URL, removal of special symbols, removal of stop words and removal of username. In removal of URL, any kind of link which is tweeted by the user and does not contribute to the sentiment analysis is removed. Removal of special symbol step deals with removing various symbols which do not have any actual sentiment, e.g. full stop (.), punctuation mark (!) and so forth. Stop words [21] removal step removes the stop words, words such as a, the which do have no effect on sentiment analysis should be removed and the conversion of emoticons to its equivalent word. Finally, in the username removal step every user's username starts with @ which has no effect on the sentiment analysis is removed, e.g. @username.

3.3 Subjectivity Detection As previously mentioned, the first step is to classify the tweet into subjective and objective and remove the tweets based on their subjectivity scorekeeping only the tweets having score higher than the specified threshold. This step is introduced to achieve higher accuracy. The pre-processed data is taken and is classified into subjective or objective statement using a subjectivity classifier. All the tweets having a subjectivity score lesser than the specified threshold are filtered out, and the classifier is trained with only the remaining tweets. It is observed that as the subjectivity threshold is increased, significant amount of tweets gets filtered out.

3.4 Feature Extraction A data set contains numerous ascribes that add clamour to the data and influence exactness. The commotion likewise bit by bit expands the time required to assemble the model. Feature extraction basically combines ascribes into a reduced feature set. The selected features and their blend assume a significant job for identifying the sentiment of the text. Selection of features [22, 23] from the extracted features can possibly improve the arrangement exactness, restricted in on a key feature subset of opinion discriminators and give more prominent understanding into habitually happening ascribes and qualities. The extracted features focus on a document vector whereupon machine learning strategies are applied to group the extremity of the content utilizing the got document vector.

3.5 Apply Classifier Algorithm There are generally three approaches which include: Supervised learning [24, 25] is a sort of learning in which we train the machine with the information which is well labelled. The machine learning approach pertinent to sentiment examination, for the most part, belongs to supervised classification. In machine learning-based methods, two sets of records are required: training set and a test set. Machine learning techniques such as naïve Bayes, SVM, maximum entropy and so forth are used. Unsupervised learning [26] is a sort of learning wherein we train the machine with the information which is not labelled. Classification is performed by comparing the features of a given text with sentiment lexicons whose sentiment values are determined prior to their use. Clustering methods such as k-means, mean shift clustering and so forth are used. Reinforcement learning (RL) [27] is the field that reviews the problems and procedures that attempt to retro-feed its model to improve. To achieve this, RL needs to be able to "sense" signals, consequently choose an activity and afterwards look at the result against a "reward" definition. RL attempts to make sense of what to do to boost these prizes, yet it does this without any support. Supervised learning methods for classification by using machine learning [28] algorithms such as Naive Bayes, SVM and maximum entropy have been found to give good accuracy. SVM was used as vast majority of researchers claimed it to be more accurate than the other algorithms, so we decided to use SVM to build the classifier.

Calculation of Result Calculating the polarity of the user's statement using the approach described. The most generally used assessment measurements are accuracy, recall, precision and F-score. The confusion matrix is shown in Table 3.

Calculating the polarity of the user's statement using the approach described. The most generally used assessment measurements are accuracy, recall, precision and F-score. The confusion matrix is shown in Table 3.

Table 3 Confusion matrix showing the performance of a sentiment analysis method

	Is positive	Is negative
Positive prediction	TP	FP
Negative prediction	FN	TN

Evaluation First we see the effects of the subjectivity threshold parameter. From the results obtained, it can be clearly observed that the tweets get filtered out to an everincreasing extent with an increase in subjectivity threshold parameter as shown in Fig. 2. Here we see the tweets remaining after the filtering process from Text Blob and Opinion Finder tool. Text Blob utilizes a function that finds a tweet's subjectivity level, whereas Opinion Finder tool denotes which segments of the message are subjective. This helps us to find the level of subjectivity of a tweet. Sent Outlook is used, which we created, to find the best-suited filtering for our experiment. Subjectivity Level = Length of subjective parts Total length of the tweet (1) For the experiment, we pick an optimal threshold value of 0.5, factoring that the model should be trained on a progressively conventional data set and the subjectivity level can be calculated using (1). The relation of the accuracy with subjectivity threshold can be seen in Fig. 3. Using the SVM classifier, we obtained satisfactory accuracy, precision, recall and F-score as shown in Table 4.

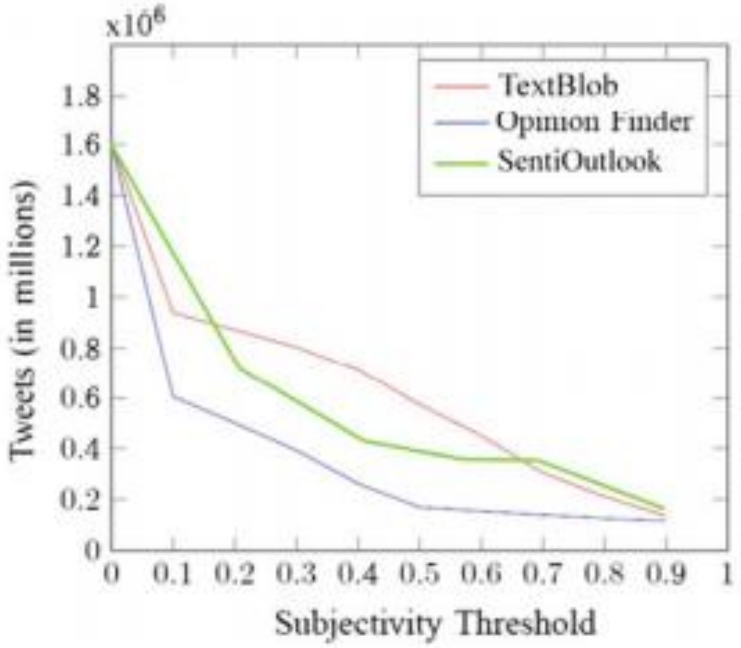


Fig. 2 Remaining tweets with subjectivity threshold

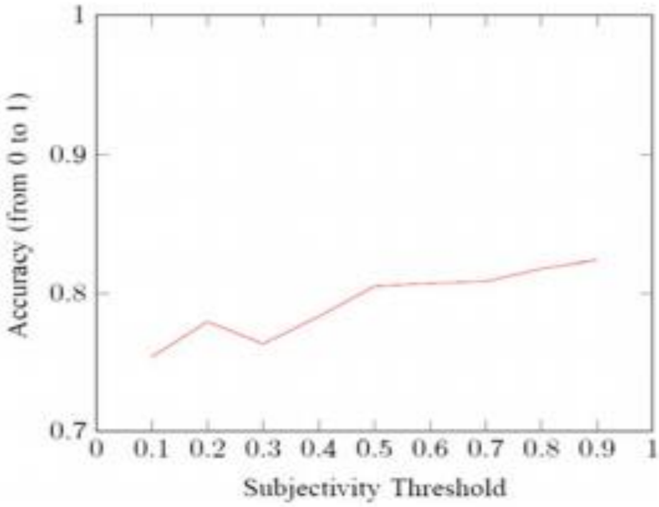


Fig. 3 Accuracy against subjectivity threshold

Table 4 Performance matrices using SVM

Matrices	Values	Formula used
Accuracy	82.66	$\frac{TN+TP}{TN+FN+TP+FP}$
Precision	78	$\frac{TP}{TP+FP}$
Recall	86	$\frac{TP}{TP+FN}$
F-score	84	$\frac{2\times\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$

1.7.3 Discussion

Twitter has a large amount of data in the form of tweets which includes the comments, opinions and reviews of the public regarding a particular product or service. Therefore, sentiment analysis comes into play to mine the opinion of the users. Many researches have been done on this but there is still a lot of scope in increasing the accuracy of the system. We came across various techniques which can be used to improve the accuracy but hardly any work is accomplished [29] on them such as oxymoron words, misspelled words, etc., and these problems should be considered in any future work done on this topic. Also, the example we took in the introduction "This pair of shoes is very good" is actually a subjective statement; however, our system detects it as an objective statement. It will also be quite interesting to go beyond just the positive and negative and extract more information and patterns from these data

CHAPTER 2 REVIEW OF LITERATURE

In the past, many researchers have emphasized the importance of emotionalism in different areas of knowledge. **In the year 1913; Charles Darwin** was the first to present strong base for emotions, representing their importance, usefulness and adaptive value. One of the appealing utility of emotions is its application during decision making stages. In the past, typically friends and families were asked for opinion, whenever decision is to be made. Organizations conducted surveys to identify the opinions of their customers about its services/products. Due to the arrival of social networks and tremendous growth in its content, locating the opinion sites and continuously examining them is a complicated task as there are large number of different sites having huge volume of text in unstructured and disorganized form .Researchers are seeking to automate document processing techniques that includes models of emotions. One of its outcomes is "Sentiment analysis or Opinion Mining".

Sentiments, emotions, feelings and other effective states having been researched in many disciplines .Emotions have been studied since the time of Darwin and different disciplines within the psychology have formed different theories showing various ways of construing emotional phenomenon (**Davidson R.F., 2003**)15. Some of the remarkable work is done by (**Osgood C.E., 1975**)16. They tried to find connection on how different people from different cultures communicate, and how people express their feelings / emotions through text. They also tried to find out how text activates different emotions.

One more approach to the textual emotion detections is to build models and then apply them to recognize the emotional tone in the text. This approach is termed as Sentiment Analysis, Subjectivity Analysis, Opinion Extraction or Opinion Mining. 15 Davidson R.F., K. S. (2003). Handbook of Affective Sciences. New York: Oxford University Press, USA. 16 Osgood C.E., W. M. (1975).

Cross Cultural Universals of Affective Meaning. Urb, University of Illinois Press.

The term “Sentiment” with reference to analyzing text automatically emerged in (Das & Chen, 2001)17 and (**Tong, 2001**)18. Subsequently it appeared in (**Pang, Lee, & Vaithyanathan, 2002)19 and (Turney P., 2002**)20.

The term ‘Sentiment Analysis’ gained popularity in the research community, and researchers in the domain of natural language processing and computational linguistics explored many dimensions of it. (**Liu B., 2010**)21 is of the opinion that many researchers have researched almost all major facets of the domain, but the solutions reported so far are not that perfect. One of the most important issues is that the existing studies are still coarse and at abstract level, as not much is being done on finer details.

Das, S., & Chen, M. (2001). Yahoo! for Amazon: Extracting market sentiment from stock message boards. Asia Pacific Finance Association Annual Conference (APFA). 18 Tong, M. R. (2001). An operational system for detecting and tracking opinions in online discussion. Proceedings of the Workshop on Operational Text Classification (OTC) .

19 Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs Up? Sentiment Classification using machine learning techniques. Conference on Empirical Methods in Natural Language Processing **(EMNLP), (pp. 79-86). 20 Turney, P. (2002).**

Thumbs Up or Thumbs Down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the Association for Computational Linguistics (ACL), (pp. 417-424). **21 Liu, B. (2010).** Sentiment Analysis: A Multi-Faceted Problem. IEEE Intelligent Systems.

Sentiment analysis can be exercised at various levels. According to (Lee, 2002)22 sentiments reside in small linguistic elements. (Wiebe J.M., 2004) 23 defined sentiment analysis as “a field that studies usually single words, phrases or sentences”. (**Kamps, Marx, Mokken, & de Rijke, 2004)24 , (Kim & Hovey, 2004)25 , (Takamura, Inui, & Okumura, 2007)** also analysed sentiments at word level.

To carry out word level sentiment analysis, various manual and semi automated dictionaries were constructed (Hatzivassiloglou & Wiebe, 2000). 26 22 Lee, P. a. (2002). Thumbs Up? : Sentiment Classification using Macine Learning Techniques. ACL-02 Conference on Empirical Methods in NLP, (pp. 79-86). 23 Wiebe J.M., W. T. (2004). Learning Subjective

Language. Computational Linguistics , 277- 308. **24 Kamps, J., Marx, M., Mokken, R. J., & de Rijke, M. (2004).**

Using WordNet to measure semantic orientation of adjectives. Proceedings of 4th International Conference on Language Resources and Evaluation, VI, pp. 1115-1118. 25 Kim, S. M., & **Hovey, E. H. (2004).**

Determining the sentiment of Opinions. Proceedings of 20th International Conference on Computational Linguistics, (pp. 1367-1373). 26 Hatzivassiloglou, V., & Wiebe, J. (2000). Effects of adjective orientation and gradability on sentence subjectivity. International conference on Computational Linguistics.

(Agarwal, 2003)27 , (Turney P.D., 2003)28 and (Hu & Liu, 2005)29 studied entire document as a sentiment unit. The core of such analysis is that a document or a statement in it contains a combination of positive and negative emotions. (Hu & Liu, 2005)30 , (Turney P. , 2002), 31 (Dave, Lawrence, & Pennock, 2003)32 advocated that semantic orientation of phrases or words can be aggregated to find the semantic orientation of the overall sentence or document. 27 Agarwal, R. S. (2003). Mining newsgroups using network arising from ssocial behavior.

Twelfth International World Wide Web Conference. 28 Turney, P., & Litman, M. L. (2003). Measuring praise and criticism: Inference of Semantic Orientation from association . ACM Transactions on Information Systems (IOIS) , 315-346. 29 Hu, M., & Liu, B. (2005). Mining and Summarizing customer Review. Proceedings of the Conference of Human Language Technology and Emperical Methods in Natural Language Processing. 30 Hu, M., & Liu, B. (2005). Mining and Summarizing customer Review. Proceedings of the Conference of Human Language Technology and Emperical Methods in Natural Language Processing. 31 Turney, P. (2002). Thumbs Up or Thumbs Down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the Association for Computational Linguistics(ACL), (pp. 417-424). 32Dave, K., Lawrence, S., & Pennock, D. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews.

sentiment analysis determines the polarity class of the text. Once the subjective sentences (sentences bearing sentiments/opinions) are identified, polarity classification, aims to classify the sentiments as negative or positive (bipolar Classification) or place its position 42 Hatzivassiloglou, V., & Wiebe, J. (2000). Effects of adjective orientation and gradability on sentence subjectivity. International conference on Computational Linguistics. 43 Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. Proceedings of the Conference on Empirical Methods in Natural Language Processing. 44 Hatzivassiloglou, V., & Wiebe, J. (2000). Effects of adjective orientation and gradability on sentence subjectivity. International conference on Computational Linguistics. 45 Beamara, F., Cesarano, C., Picariello, A., Reforgiato, D., & Subrahmanian, V. (2007). Sentiment Analysis: Adjectives and Adverbs are better than adjectives alone. International Conference in Weblogs and Social Media.on the range between positive and negative polarities (multiclass classification) (Pang & Lee, 2008) 46 .

Another important task in sentiment analysis is to recognize an object or feature on which sentiments are expressed. (Hu & Liu, 2005)47 (Liu B. , 2006) 48 and (Popescu & Etzioni, 2005) 49are of an opinion that not only an object but its attributes or components are significant for a more accurate and detailed sentiment analysis as a reviewer can have different sentiments on different components or features. (Yi, Nasukawa, Niblack, & Bunescu, 2003) 50 , (Hu & Liu, 2004) 51 , (Popescu & Etzioni, 2005)52 investigated into an area of feature extraction in sentiment analysis. 47 Hu, M., & Liu, B. (2005). Mining and Summarizing customer Review. Proceedings of the Conference of Human Language Technology and Emperical Methods in Natural Language Processing 48 Liu, B. (2006). Web Data Mining Chapter Opinion Mining. Springer. 49 Popescu, A., & Etzioni, O. (2005). Extracting the product feature opinions from reviews. Human Language Technology and Emperical Methods in Natural Language Processing. 50 Yi, J., Nasukawa, T., Niblack, W., & Bunescu, R. (2003). Sentiment Analyzer: Extracting sentiments about a given topic using natural language processing techniques. Proceedings of the 3rd IEEE international conference on Data Mining, (pp. 427-434). 51 **Hu, M., & Liu, B. (2004).** Mining opinion features in customer reviews. Proceedings of AAAI, (pp. 755-760). 52 Popescu, A., & Etzioni, O. (2005). Extracting the product feature opinions from reviews. Human Language Technology and Emperical Methods in Natural Language Processing.

Opinion holder extraction is also one of the significant job in sentiment analysis. (Choi, Cardie, Riloff, & Patwardhan) 53 applied extraction patterns and conditional random fields to identify the opinion holder. The problem was also addressed by (Bethard, Yu, Thornton, Hatzivassiloglou, & Jurafsky, 2004)54 . Many times a sentence can be a comparative sentence, where there are more than one objects and sentiments are expressed on them. (Liu B. , 2006) 55 stated that “by using comparative adjectives and adverbs (more, less, better, longer) and superlative adjectives (most, least, best), these sentences can be identified.” 53 Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. Identifying sources of opinions with conditional random fields and extraction patterns. Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language PROCESSING, (p. 2005). 54 Bethard, S., Yu, H., Thornton, A., Hatzivassiloglou, V., & Jurafsky, D. (2004). Automatic extraction of opinion propositions and their holders. Proceedings of the AAI Spring Symposium on Exploring Attitudes and Affect in Text. 55 Liu, B. (2006). Web Data Mining Chapter Opinion Mining. Springer.

Most of the work (Lee, 2002) 56 , (Turney P. , 2002)57 , (Hu & Liu, 2005) 58; (Popescu & Etzioni, 2005)59 , (Liu B. , 2006) 60 has been carried out on movie and product reviews, as it is easy to recognize the subject / topic of the text. In the literature, the problem of sentiment analysis is addressed by two techniques viz. supervised (machine learning) and unsupervised (semantic orientation). Machine learning is a branch of Artificial Intelligence. It is all about constructing a system that learns from data. In this technique there are two sets of documents; one is called training set and the other is called test set. An automatic classifier learns the distinguishing characteristics from the training set. Using test set, the performance of the automatic classifier is validated. Several machine learning methods have been employed in analyzing sentiments.

Most 56 Lee, P. a. (2002). Thumbs Up? : Sentiment Classification using Macine Learning Techniques. ACL-02 Conference on Empirical Methods in NLP, (pp. 79-86). 57 Turney, P. (2002). Thumbs Up or Thumbs Down? Semantic orientation applied to unsupervised classification of reviews. Proceedings of the Association for Computational Linguistics(ACL), **(pp. 417-424). 58 Hu, M., & Liu, B. (2005).**

Mining and Summarizing customer Review. Proceedings of the Conference of Human Language Technology and Empirical Methods in Natural Language Processing. 59 Popescu, A., & Etzioni, O. (2005). Extracting the product feature opinions from reviews. Human

Language Technology and Empirical Methods in Natural Language Processing**. 60 Liu, B. (2006).** Web Data Mining Chapter Opinion Mining.

Springer prominent of them are Support Vector Machine, Maximum Entropy Classification and Naïve Bayes Classification. Naïve Bayesian classification is one of the extensively used mechanisms for analyzing sentiments. It uses Bayes rule as its main equation for text categorization. The naïve part of this model is the assumption of the word independence.

It has shown good results in the research conducted by (Lewis & Ringuette, 1994) 61and (McCallum & Nigam, 1998) 62 . It is used by (Melville & Wojciech, 2009) 63 , (Qiang, Ziqiong, & Rob, 2009) 64 , (Rui, Chengqing, & Shoushan, 2011) 65 , (Songbo & Jin, 2008) 66 , (Ziqiong, Qiang, Zili, & Yijun, 2011) 67for the 61 Lewis, D., & Ringuette, M. (1994).

A comparison of two learning algorithms for text categorization. Third Annual Synposium on Document Analysis and IR, (pp. 81-9). 62 McCallum, A., & Nigam, K. (1998). A comparison of event models for Naive Bayes text classification. Proceedings of Third Annual Symposium on Document Analysis and IR . 63 Melville, & Wojciech, G. (2009).

Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification. KDD 09. Paris, France. 64 Qiang, Y., Ziqiong, Z., & Rob, L. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Systems with Applications, (pp. 6527-6535). 65 Rui, X., Chengqing, Z., & Shoushan, L. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. Information Sciences 181 , 1138-1152. 66 Songbo, T., & Jin, Z. (2008).

An empirical study of sentiment analysis for chinese documents. Expert Systems with Applications , 2622-2629. 67 Ziqiong, Z., Qiang, Y., Zili, Z., & Yijun, L. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. Expert Systems with applications .sentiment analysis task. (Pang, Lee, & Vaithyanathan, 2002) 68experimented with these approaches and classified reviews pertaining to movies into two classes, namely positive and negative.

They compared Maximum Entropy [ME], Support Vector Machine [SVM] and Naïve Bayes [NB], and by considering different characteristics unigrams(1-gram), bigrams(2-grams) or both in combination, only adjectives, POS tags etc. They reported that Support Vector Machine outperforms other two by achieving 82.9 %, highest classification accuracy.
A support vector machine is one of the best statistical classification method that was proposed by Vapnik. SVM technique was used by (Rui, Chengqing, & Shoushan, 2011) 69, (Ziqiong, Qiang, Zili, & Yijun, 2011) 70, (Songbo & Jin, 2008) 71for sentiment analysis. Multiple variants 68 Pang, B., Lee, L., & Vaithyanathan, S. (2002).

Thumbs Up? Sentiment Classification using machine learning techniques. Conference on Empirical Methods in Natural Language Processing (EMNLP), (pp. 79-86). 69 Rui, X., Chengqing, Z., & Shoushan, L. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. Information Sciences 181 , 1138-1152. 70 Ziqiong, Z., Qiang, Y., Zili, Z., & Yijun, L. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. Expert Systems with applications . 71 Songbo, T., & Jin, Z. (2008).

An empirical study of sentiment analysis for chinese documents. Expert Systems with Applications , 2622-2629.of SVM were developed. (Kaiquan, Liao, Jiexun, & Yuxia, 2011) 72 used multiclass SVM for sentiment classification.

Maximum entropy attempts to preserve maximum uncertainty as much as possible. By maximizing entropy, it is ensured that no biases are initiated into the system. (Nigam, Lafferty, & McCallum, 1999) 73and (Berger) 74 investigated that maximum entropy outperforms Naïve Bayes technique.

Many algorithms are offered by machine learning approach, but classifying data according to the presence of sentiments in it, presents many unique challenges. (Pang & Lee, 2002) 75used term presence instead of frequency to improve the performance of their system, because it is more advantageous to look out for most unique ones instead of paying attention to most frequent items. (Kim & Hovy , 2006) 76 added to the above thought that, not only term presence but term positions are also essential for sentiment analysis.

Term position determines, and 72 Kaiquan, X., Liao, S., Jiexun, L., & Yuxia, S. (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. Decision Support

Systems, 743-754. 73 Nigam, K., Lafferty, J., & McCallum, A. (1999). Using maximum entropy for text classification. Proceedings of IJCAI-99 Workshop on machine learning for information filtering. 74 Berger, A. (n.d.).

A brief Maximum Entropy Tutorial. 75 Pang, B., & Lee. (2002). Thumbs Up? Sentiment Classification Using Machine Learning Techniques. Proceedings of ACL-02 Conference on Empirical Methods in Natural Language Processing 10:79-86. 76 Kim, S. M., & Hovey, E. H. (2004). Determining the sentiment of Opinions. Proceedings of 20th International Conference on Computational Linguistics, (pp. 1367-1373). Sometimes reverses, the polarity of the phrase.

(Turney & Litman, 2003) 77, used part-of-speech tagging and used adjectives & or adverb for detecting sentiments at the document level. Syntax information such as negation, intensifiers and diminisher are used by (Kudo & Mastumoto, 2004) and shown that it performs better than the bag of words approach. (Das & Chen, 2001a) 78 suggested to take negations into consideration by appending them to the terms that are close to negation.

They considered negation as an integral task in sentiments analysis as negation reverses the polarity of a word. Machine learning approaches tend to be more accurate than lexicon based approaches, but it requires huge training data to train the classifier. Lexicon based methods does not require prior training and it has better generality than machine learning approach. In the literature there are many references of lexicon based approach for sentiment classification.

It usually works well on small sentences, which is the prominent feature of social networking sites (Bermingham & Smeaton, 2010) 79 . According to (Chaovalit & Zhou, 2005)80 , “it is also appropriate for realtime sentiment classification given its relatively lower computation requirement”. 77 **Turney, P., & Litman, M. L. (2003).**

Measuring praise and criticism: Inference of Semantic Orientation from association . ACM Transactions on Information Systems (IOIS) , 315-346. 78 Das, S., & Chen, M. (2001). Yahoo! for Amazon: Extracting market sentiment from stock message boards. Asia Pacific Finance Association Annual Conference (APFA). 79 Bermingham, A., & Smeaton, A. F. (2010).

Classifying Sentiment in Microblogs: Is Brevity an Advantage? Proceedings of the 19th ACM international conference Proceedings of the 19th ACM international conference on Information and knowledge management, (pp. 1833-1836). 80Chaovalit, P., & Zhou, L. (2005). Movie Review Mining: A Comparison between Supervised and Unsupervised Classification Approaches. Proceedings of the 38th Annual Hawaii International Conference on System Sciences, (pp. 112c-112c).(Liu B. , 2012) 81 termed sentiment lexicon as an opinion lexicon. According to (Pang B., 2008) 82, "it is a collection of words or phrases that are commonly used to express feelings." Words that belong to sentiment lexicons are called sentiment words. Each word in it is associated with a score representing its sentiments.

To perform sentiment analysis, every word in the text is matched against the words in the lexicon. If a match is found, sentiment scores are recorded. Total sentiment score of a document is computed using positive and negative scores. (Fu, Abbasi, Zeng, & Chen, 2012) 83and (Hu & Liu, 2005) 84used the equation "difference between the score obtained from the positive words (+ve words) and that from (minus) the score obtained from negative words (-ve score). (Das and Chen 2001) 85 imposed more weights on certain words, for example, higher weights are assigned to adjectives and adverbs as they are the strong indicators of sentiment words. To eliminate the problem of ambiguities, some studies e.g. (Wilson, Wiebe, & Hoffmann, 2005b)86 attached POS tags to every word in the lexicon. 81 Liu, B. (2012). Sentiment Analysis and Opinion Mining.

Synthesis Lectures on Human Language Technologies , 1-167. 82 Pang B., L. L. (2008). Opinion Mining and Sentiment Analysis. Foundation and trends in Information Retrieval , 2(1-2) pp 1-135. 83 Fu, T., Abbasi, A., Zeng, D., & Chen, H. (2012). Sentimental Spidering: Leveraging Opinion Information in Focused Crawlers . ACM Transactions on Information Systems , 24. 84 Hu, M., & Liu, B. (2005). Mining and Summarizing customer Review.

Proceedings of the Conference of Human Language Technology and Emperical Methods in Natural Language Processing 85 Das, S., & Chen, M. (2001). Yahoo! for Amazon: Extracting market sentiment from stock message boards. Asia Pacific Finance Association Annual Conference (APFA). 86 Wilson, T., Wiebe, J., & Hoffmann, P. (2005b).

"Recognizing Contextual Polarity in PhraseLevel Sentiment Analysis . Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing: Association for Computational Linguistics, (pp. 347-354).Lexicon based methods are based on the language skills and are reliable in all the texts and domains (Taboda, Brooke, Tofiloski, Voll, & Stede, 2011) 87 .

The effectiveness of lexicon based approach depends upon the lexical resources on which it relies. Widely used general purpose lexicons to detect sentiments are SentiWordNet (SWN) (Esuli & Sebastiani, 2006) 88 , General Inquirer (GI) (Stone, Dunphy, & Smith, 1966) 89 , Opinion Lexicon (OL) (Hu & Liu, 2005) 90 , Multi Perspective Question Answering (MPQA) (Wilson, et al., 2005a)91 etc. According to (Brody and Diakopoulos 2011) "general lexicons are robust across different domains (Liu B. , 2012) 92, however, their performance usually suffers from two aspects, insufficiency and inaccuracy**." 87 Taboda, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011).**

Lexicon-based methods for Sentiment Analysis. COmputational Linguistics , 37 (2), 267-307. 88 Esuli, A., & Sebastiani, F. (2006). SentiwordNet: A Publicly available lexical resource for Opinion Mining. Proceedings of 5th Conference on Language Resources and Evaluation. 89 Stone, P., Dunphy, D. C., & Smith, M. S. (1966). The General Inquirer: A Computer Approach to Content Analysis. **90 Hu, M., Liu, B., & Cheng, J. (2005).**

Opinion Observer: Analysing and Comparing Opinions on the Web. Proceedings of International World Wide Web Conference. 91 **Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., et al. (2005a).** Opinion finde A System for Subjectivity Analysis Proceedings of HLT/EMNLP on Interactive Demonstrations. Association for Computational Linguistics , 34-35. 92 Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies , 1-167.

In a sentiment lexicon, for each word, a value representing its sentiment score is assigned either manually or by using some automatic technique. Largely used sentiment lexicons namely, General Inquirer (Stone, Dunphy, & Smith, 1966)93 and Multi-Perspective Question Answering (MPQA) Subjectivity Lexicon (Wilson, Wiebe, & Hoffmann, 2005b) 94, are compiled by experts manually. But it requires a considerable efforts and time and of experts to construct it. Automatic methods usually make use of language resources i.e. dictionary or corpus (Liu B., 2012)95. In the "dictionary-based approach", a small set of sentiment words are handpicked. These words are called seed words. New words representing sentiments are recognized by finding out the association between the new word and seed words.

This relation is defined by the dictionary that is being used. For instance, the Opinion Lexicon (Hu & Liu, 2005) 96by Liu, consist of sentiment adjectives those are determined by antonyms and synonyms of a seed words in a lexicon. The efficiency of this technique greatly depends upon the antonym and synonym entries of the dictionary employed. Even though these relations between the entries are accurate, they do not include any domain-specific information.

According to **(Hatzivassiloglou & Wiebe, 2000)** 97 " in the corpus-based approach, new 93 Stone, P., Dunphy, D. C., & Smith, M. S. (1966). The General Inquirer: A Computer Approach to Content Analysis. 94 Wilson, T., Wiebe, J., & Hoffmann, P. (2005b). "Recognizing Contextual Polarity in Phrase Level Sentiment Analysis . Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing: Association for Computational Linguistics, (pp. 347-354). 95 Liu, B. (2012).

Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 1-167. 96 Hu, M., & Liu, B. (2005). Mining and Summarizing customer Review. Proceedings of the Conference of Human Language Technology and Emperical Methods in Natural Language Processing. 97 Hatzivassiloglou, V., & Wiebe, J. (2000).

Effects of adjective orientation and gradability on sentence subjectivity. International conference on Computational Linguistics. Sentiment words are recognized based on their relationships with seed words in the corpus. They identified sentiment words using conjunctions in a 21-million-word WSJ (Wall Street Journal) corpus". In their algorithm (Hatzivassiloglou & Wiebe, 2000)97, they did not use a seed lexicon. Instead, they were dependent upon the linguistic knowledge to identify the polarity of a word. Other corpus-based techniques also used seed words. (Turney P.D., 2003) 98 used a "Pointwise Mutual Information and Information Retrieval (PMI-IR)" technique to determine the text polarity.

This approach directly determines the polarity of the sentiment words in the text. PMI is applied to gauge the co-occurrence between two phrases/ words. The two seed words that were selected by (Turney P.D., 2003) 98are “excellent and poor”. (Wiebe J.M., 2004) 99 iteratively annotated a corpus automatically and then applied it to dig out the subjective information.

However, this approach did not determine the polarity of subjective words. A broad range of lexicons have been developed to be applied in Sentiments Analysis. They are constructed by expanding the existing general-purpose lexicons. One classic example is that of (Subasic & Huettner, 2001) 100.

They manually built a lexicon and associated with it the words with affect categories, specifying intensity and centrality. The most extensively used lexicon for sentiment analysis is SentiWordNet (**Esuli & Sebastiani, 2006**) 101 . SentiWordNet offers 98 Turney, P., & Litman, M. L. (2003). Measuring praise and criticism: Inference of Semantic Orientation from association . ACM Transactions on Information Systems (IOIS) , 315-346. 99 Wiebe J.M., W. T. (2004). Learning Subjective Language. Computational Linguistics , 277- 308. 100 Subasic, & Huettner. (2001).

Affect analysis of text using fuzzy semantic typing. IEEE - FS , 9, 483-496. 101 Esuli, A., & Sebastiani, F. (2006). SentiwordNet: A Publicly available lexical resource for Opinion Mining. Proceedings of 5th Conference on Language Resources and Evaluation.annotation based on three numerical sentiment scores (positivity, negativity and objectivity) for each synsets of the WordNet . WordNet is an electronic lexical database of English. It is one of the most valuable tools for computational linguistic and natural language processing and. In WordNet, Adverbs, Verbs, Adjectives and Nouns are grouped into cognitive synonyms called “synsets”. These synsets are connected to each other by means of lexical and semantic relations. SentiWordNet assigns 3 sentiment scores positive, negative and objective to each synset such that “Pos(S) + Neg(S) + Obj(S) = 1”. WordNetAffect is another extension to WordNet that is developed by (Strapporava & Viltutti, 2004) 102.

It is an expansion of WordNet that supports for multilingual sentiment detection. It was developed at ITC-irst. WordNetAffect assigns at least one domain label (e.g. Sports, Medicines, and Politics etc) to each synsets. Multi-Perspective Question Answering (MPQA), is a subjective lexicon of 8,222 terms gathered from several sources. It comprises of a list of words that is associated with its polarity and intensity of polarity. SenticNet is also one of the popular lexicons used for sentiment analysis. It provides sentiment scores in a range between -1 and 1 for 14000 common sense concepts.

The sentiments conveyed 102 Strapporava, C., & Viltutti, A. (2004). Wordnet - afect: An feefective extension of WordNet. 4th International Conference on Language Resources and Evaluationare defined on the basis of the intensity of 16 basic emotions defined in a model called Hourglass of Emoticons.

Sentiment Analysis is an essential means to measure the polarity of the customer reviews on the products. Customer reviews are generated by various web resources like social networks, blogs, newspapers, e-commerce websites and forums.

It helps to express the opinions and generates enormous amount of data, which leads to big data. In past years, countless attention was received through web resources as a new resource of separate opinions [Din, 08] [Moh, 12]. The opinions are utilized to understand the general public and consumers on social events, political movements, company strategies, marketing campaigns, product preferences and monitoring reputations [Sal, 11]. Many of the organizations are using real-time applications for analyzingthese data with the combination of web mining, data mining, text mining and natural language processing techniques. The vast amounts of data related to customer opinions/reviews are quite difficult to analyze and need a novel approach to get a generalized opinionated summary. This leads the research community and industrialists to depend on big data analytics.

In July 2000 by Francis Diebold of University of Pennsylvania in his work of Econometrics and Statistics (2000) [Die, 00]: “Big data refers to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology. In this new and exciting world, sample sizes are no longer fruitfully measured in “Number of observations”, but rather in, say, megabytes. Even data accruing at the rate of several gigabytes per day are not uncommon”. There are some definitions given by some organizations. Gartner defined big data as [Gar, 1up, “Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making”.

The Big Data Commission at the Tech America Foundation offers the following definition: "Big data is a term that describes large volumes of high-velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information" (Tech America Foundation, 2012) [Ami, 15]. Researchers at McKinsey propose an intentionally subjective definition: "Big data refer to datasets whose size is beyond the ability of the typical database software tools to capture, store, manage, and analyse". (McKinsey Global Institute May 2011) [Jam, 11].

Meanwhile, Jerry Smith, Data Scientist Insights contributor, developed a mathematically defined big data [Smi,12]: "Big data represents the historical debris (observed data) resulting from the interaction between 70 and 77 independent variable / subjects, from which non-random samples of unknown populations, shifting in composition with a targeted time frame, can be taken".

Big data analytics denotes the process of gathering, analyzing and organizing the massive amounts of data to determine patterns and other useful information. Big data analytics are involved in making understandings out of huge volumes of varied data. These varied data are in a raw form that lacks a data model to define what each element means in the context [Ora, 13]. This technology typically includes large-scale and complex programs under specific analytical methods.

CHAPTER 3 RESEARCH METHODOLOGY

3.1 INTRODUCTION:

Today's world is completely flooded with reviews and ratings in online markets and websites.

Television, films, videos, billboards, movies, magazines, newspapers, internet and several others are the channels through which producers advertise their products, services or their brand to the public by ratings and reviews.

Among all the marketing weapons, advertising is more popular for its long-lasting impact on viewer's mind, as its exposure is much broader. through all media influence audiences, but television and the internet, especially the social media networks and products ratings have become one the strongest medium of advertising as it has massive reach and can influence not only the individual's attitude, but also his/her behaviour, lifestyle, exposure and in the long run, even the culture of the entire society.

The primary objective of marketing is to impact on buying behaviour and brand preference of the audience and also result into change or strengthening in the memories of the brand in the consumer's mind. Memories about the brand consist of those associations that are related to brand name in consumer mind. These brand cognition influence consideration, evaluation and finally purchases. However, the most important thing that the marketer or the advertiser want to identify is the influence of advertising that consumers have in their brand preference. Organizations spend heavily on the advertising of its products and services so as to attract and influence the consumers towards their brand. The organizations try to entice the consumers by showcasing the benefits or advantages the consumers will have when they get associated with their brand. This ensures that the sentimental analysis and marketing will create a favourable place in the minds of the consumers, thus resulting into their loyalty towards the brand. Therefore, it is very essential to study the role and impact of ratings on brand preference of the consumers.

3.2 RESEARCH OBJECTIVES

1: To find out if the buying behaviour changes according to the reviews.

2: Do people believe in polls?

3: To find out if people watch the recommended content.

4: How many people are aware of sentimental analysis and marketing.

5: To find out whether Artificial intelligence is feasible for opinion computing.

3.3 HYPOTHESIS

Hypothesis An explanation is a set of statements usually constructed to describe a set of facts which clarifies the causes, context, and consequences of those facts. This description may establish rules or laws and may clarify the existing rules or laws in relation to any objects, or phenomena examined.

H0: There is no impact on people by the usage of sentimental analysis in consumer behaviour of people who shop online as well as people who use social media and other OTT platforms.

H1: There is a positive impact of by the usage of sentimental analysis in consumer behaviour of people who shop online as well as people who use social media and other OTT platforms.

3.4 RESEARCH METHODOLOGY:

3.3.1 Definition:

Research methodology is a description of what the activity of research is, how to carry it on, how to measure its progress, and what continues its success. Research methodology is the study of the principles of methods, rules, and postulates employed by a certain discipline (Kothari, 2003). This chapter signifies the "research methodology", which explains the research objectives, significance of the study, scope, sampling, sources of data collection, statistical techniques used for data analysis, etc. Basically, the purpose of this chapter is to communicate the nature of the problem and the tools used to find solutions of the problem.

3.3.2 Research design, sample and data:

Universe	Mumbai
Sampling Method	Simple Random Sampling Method
Sample Size	99
Methods of Data Collection	Primary Data & Secondary Data
a) Primary Data	Questionnaire
b) Secondary Data	Blogs, Websites, Journals, Reference books
Representation of Data Analysis	Pie Charts, Graphs

• Area of the study: Mumbai.

• Sampling Size: 99 , consumers who shop through various online websites at different locations in Mumbai.

• Data Source: Primary data and Secondary data were both used for information generation. The findings were drawn mainly from primary source by making them fill a questionnaire.

• Sampling Procedure: The sampling procedure used is Simple random sampling method. Simple random sampling method means each member of the population is equally likely to be chosen as the part of the sample.

3.3.3 Research Universe: The Universe consists of all survey elements that have the quality for inclusion in the research study. The accurate definition of the universe for a particular study is the set by which it specifies who or what is of the interest of the research study. The universe may be individuals, groups of individuals or organizations.

Research Design: Research design is a framework or blueprint for conducting the marketing research project. It details the procedures necessary for obtaining the information needed to solve the marketing research problems.

3.3.4 Sampling Method: The sampling method is the process of selection of certain percentage of a whole group of items as per pre-determined plan. Sampling is a better choice from the point of view of time and money.

Simple Random Sampling Method: Simple random sampling is a sampling technique where every item in the population has an even chance and likelihood of being selected in the sample. Here, the selection of items completely depends on chance or by probability and therefore, this sampling technique is also sometimes known as a "method of chances". It is the purest and most straightforward probability sampling strategy. It is also the most popular method for choosing a sample among population for a wide range of purposes. For the present study, this method has been selected so as to obtain data from a sample of two hundred and sixty respondents. After going through and screening the filled-up questionnaires, all respondents were found to be complete in all aspects and hence fit for the analysis. Utmost care has been taken to ensure that the respondents of all age groups, genders and occupations were included in the study.

3.3.5 Sample:

Sample means a small portion of the population taken up for intensive study purpose. It is a small part of the entire population having similar characteristics within the population. It is selected at random form large number of individuals who are suitable for the study.

Sample Size:

The sample size of the survey most typically refers to the numbers of units that were chosen from which data was gathered. However, sample size can be number of sample units selected for contact or data collection. There is also the final sample size, which is the number of completed interviews or units for which the data is actually collected. The final sample size may be much smaller than the designated sample area (universe), if there are considerable non-responses, ineligibility, or both. Not all the units in the designated sample need to be processed if productivity in collecting the data is much higher than anticipated to achieve the final sample size.

3.3.6 Data Collection:

Data is the information which will be collected from various sources. It concerns with gathering of accurate and precise information about the problem of the study.

Methods of Data Collection:

There are two methods which can be used to collect relevant data, which can be essential for this study. They are:

a) Primary Data: Primary data constitutes of first-hand information which is collected for the first time in order to solve research problem. It is the data collected from primary sources which are the original sources of data collection. The researcher himself/herself collects primary data or through trained assistants. Some of the primary sources of data collection include interview method, questionnaire method and observation method.

➢ Questionnaire method: Questionnaire is the most evident method of data collection, which is comprised of a set of questions related to the research problem. This method is very convenient, when the data is to be collected from the diverse population. In mainly includes set of questions, either open-ended or close-ended, which the respondents are required to answer on the basis of their knowledge and experience with the issue concerned. This method is used in the current study, as it provides convenience and can be accomplished using minimal time and money.

b) Secondary Data: Secondary data refers to data which is collected by someone, other than the user. Secondary data is the information that already exists, collected by someone else. It can also be referred as the second-hand data. Secondary data analysis can save time that would otherwise be spent on collecting and gathering data, particularly in the case of quantitative data and can provide larger and higher-quality databases that would be unfeasible for any individual researcher to collect on their own.

The main sources of secondary data collection in this study were:

➢ Blogs: A blog (a truncation of the expression "weblog") is a discussion or informational website published on the web consisting of discreet, often informal diary-style text entries (posts). Posts are typically displayed in reverse chronological order, so that the most recent post appears first, at the top of the web page.

➢ Websites: A website is a collection of related network web resources, such as web pages, multimedia content, which are typically identified with a common domain name, and published on at least one web server. Notable examples are Wikipedia.org, Google.com and Amazon.com. Websites can be accessed via a public Internet Protocol (IP) network, such as the internet or a private Local Area Network (LAN), by a Uniform Resource Locator (URL) that identifies the site.

➢ Reference books: Reference books generally refers to books which are written by research scholars or other knowledgeable people, which can be referred by others to gain more knowledge and guidance. 54 These books are very useful for the study as they provide discrete and essential information regarding various topics which can be favorable by the researcher. Some other essential sources of secondary data collection are as follows:

➢ Articles ➢ Journals ➢ Newspapers ➢ Magazines, etc.

3.5 SIGNIFICANCE OF THE STUDY:

Seeing the current scenario, we seem to be in a world, in which the fittest survives at the expense of the weakest. Anything and everything can sell in this world, if it is advertised effectively. And as the trend is going now-a-days, customers have become more conscious towards preferences of the brands they use or are associated with. Due to the super-fast media and rapidly changing technologies, the preference of brand has been largely impacted and influenced by the role of marketing.

The study is targeted towards the impact of different types of marketing and impacts of sentiments on the consumer brand preference.

This study demystifies the influence of ratings on the choices that consumers makes while making purchase decisions. This study will help organizations and brands to understand and emphasize on the importance of ratings and reviews on consumer brand preference.

It will also enable them to structure their adverts and brands to make them more appealing in order to improve sales and lead to better performance. The significance of this study is to understand the overall role and impact that ratings and reviews have on the consumers while making brand preferences.

This study will not only be helpful for the organizations and brands, but also can be very significant to the consumers and the society in large to understand the details and peculiarities of advertisements and the impact it has on them.

The findings and recommendations of this study will go a long way in helping organizations and brands in adopting unique and creative advertising strategies, and appealing brand designs to help in getting more consumers for their products and services

3.6 LIMITATIONS OF THE STUDY:

• Bias: The data is based on individual opinion which may bring to some bias opinions.

• The study is restricted to Mumbai. Hence it may not be possible to generalize the finding to the entire population of the country.

• Primary source of data collection: Primary source of data is the main source of gathering information, hence manipulation at the respondent's end cannot be avoided.

• Many of the people approached responded to the questionnaires based on their mood and feelings at that moment of time.

• Time factor can be considered as a major limitation.

• This study had mainly focused on advertisements as its center point for evaluating consumer brand preference, hence, other factors such as price, quality, etc. are not taken into consideration.

• The information which has been gathered through the internet, newspapers, research articles, etc. carry their own limitations, as most of them are based on the author's thoughts and perception, true picture behind it cannot be portrayed.

CHAPTER 4 ANALYSIS AND INTERPRETATION OF DATA

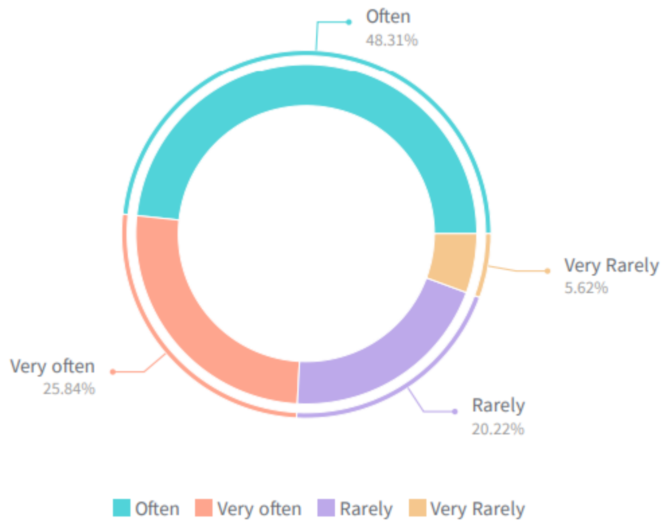
Q. 1 Demographics

- Name
- Email Id
- gender
- employment status
- age

This question was compulsory

QUESTION 02 | MULTIPLE CHOICE

How often do you shop online?



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Often	43	48.31%
Very often	23	25.84%
Rarely	18	20.22%
Very Rarely	5	5.62%

Interpretation: According to the survey,

(48.31%) Often do online shopping.

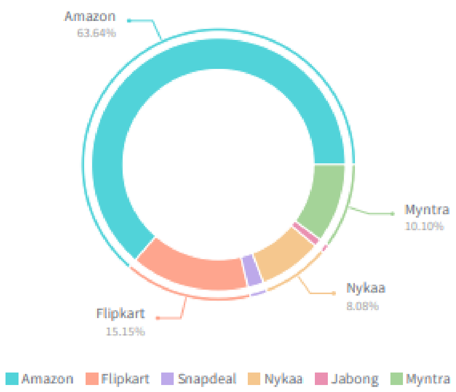
(25.84%) Are the people who shop online very often.

(20.22%) Who shop Rarely

(5.62%) Are the people who shop very rarely.

QUESTION 03 | PICTURE CHOICE

Which apps or websites you usually shop



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Amazon	63	63.64%
Flipkart	15	15.15%
Snapdeal	2	2.02%
Nykaa	8	8.08%
Jabong	1	1.01%
Myntra	10	10.10%

Interpretation: According to the survey,

(63.64%) Use Amazon.

(15.15%) Are the people who use Flipkart.

(2.02%) Use Snapdeal.

(8.08%) Use Nykaa.

(1.01%) Use Jabong.

(10.10%) Use Myntra.

QUESTION 04 | MULTIPLE CHOICE

When did you last buy something online ?

ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Less than 1 month ago	53	53.54%
Between 1 and 6 months ago	31	31.31%
Between 6 months and 1 year ago	9	9.09%
I don't remember	6	6.06%

Interpretation: According to the survey,

(53.54%)Less than 1 month ago.

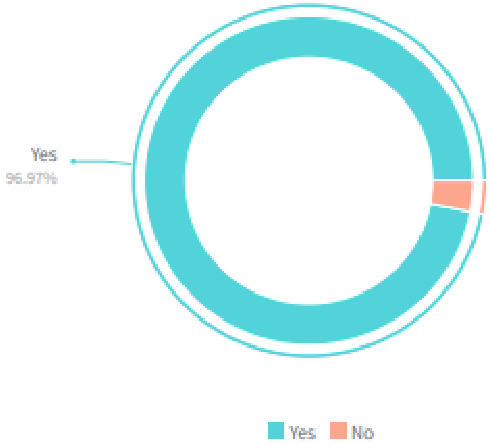
(31.31%) Between 1 and 6 months ago.

(9.09%) Between 6 months and 1 year ago.

(6.06%) I don't remember.

QUESTION 05 | YES OR NO

Do you consider reviews and ratings before buying the product?



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Yes	96	96.97%
No	3	3.03%

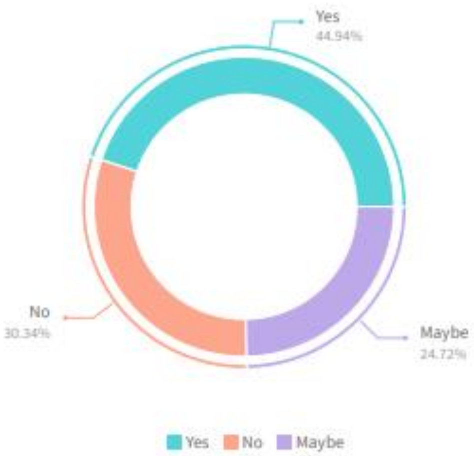
Interpretation: According to the survey,

(96.97%) Say Yes.

(3.03%) Say No.

QUESTION 06 | MULTIPLE CHOICE

Do you feel comfortable trying a new product that has no reviews or ratings?



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Yes	40	44.94%
No	27	30.34%
Maybe	22	24.72%

Interpretation: According to the survey,

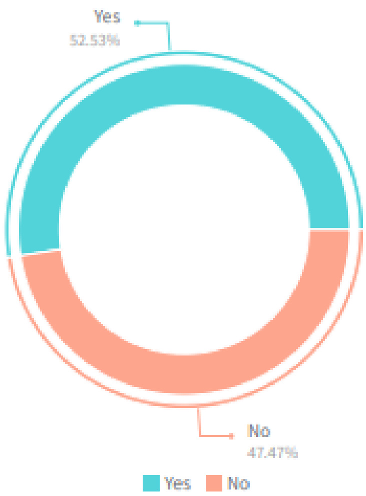
(44.94%) Are okay trying out a product without a review.

(30.34%) Are not okay without reviews.

(24.72%) Are not sure.

QUESTION 07 | YES OR NO

Do you watch twitter polls?



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Yes	52	52.53%
No	47	47.47%

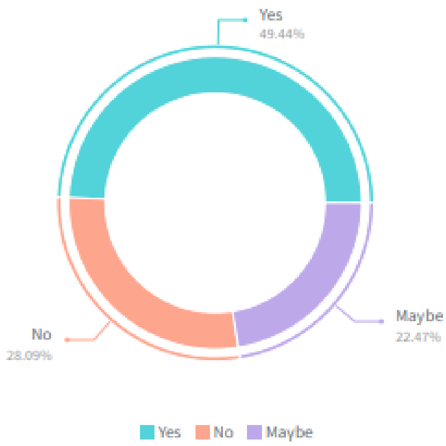
Interpretation: According to the survey,

(52.53%) Say Yes, they do watch twitter polls.

(47.47%) Say No they do not watch twitter polls.

QUESTION 08 | MULTIPLE CHOICE

Do you believe or trust twitter polls and product ratings?



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Yes	44	49.44%
No	25	28.09%
Maybe	20	22.47%

Interpretation: According to the survey,

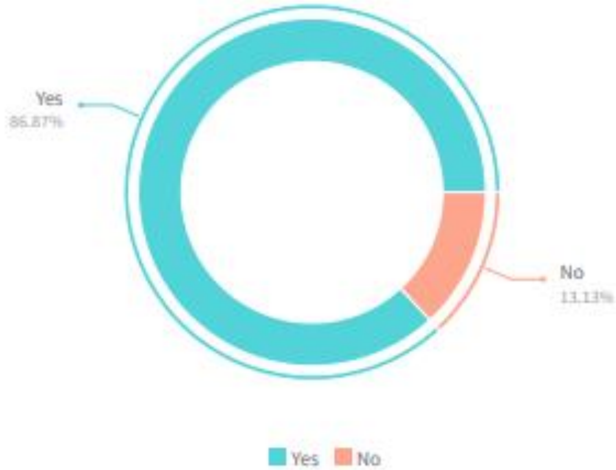
(49.44%) Say Yes, they trust twitter polls.

(28.09%) Say No they don’t trust twitter polls.

(22.47%) Say they are not sure about twitter polls.

QUESTION 09 | YES OR NO

Do you watch recommended videos on youtube or any other OTT platforms?



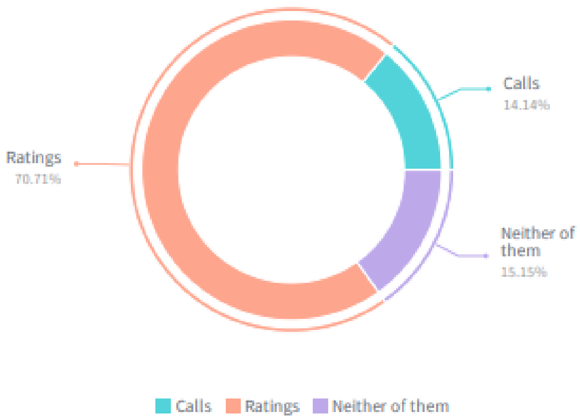
ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Yes	86	86.87%
No	13	13.13%

(86.87%) Say yes, they do watch recommended options.

(13.13%) Say No they do not watch recommended options.

QUESTION 10 | MULTIPLE CHOICE

What do you find more comfortable receiving a call for reviews or giving a star rating on an app or website.



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Calls	14	14.14%
Ratings	70	70.71%
Neither of them	15	15.15%

Interpretation: According to the survey,

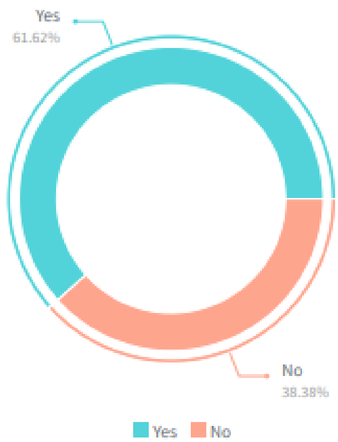
(14.14%) Say they are okay with calls.

(70.71%) Say they are fine with rating a product.

(15.15%) Say they don’t want any.

QUESTION 11 | YES OR NO

Are you aware about the term sentimental analysis/marketing?



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Yes	61	61.62%
No	38	38.38%

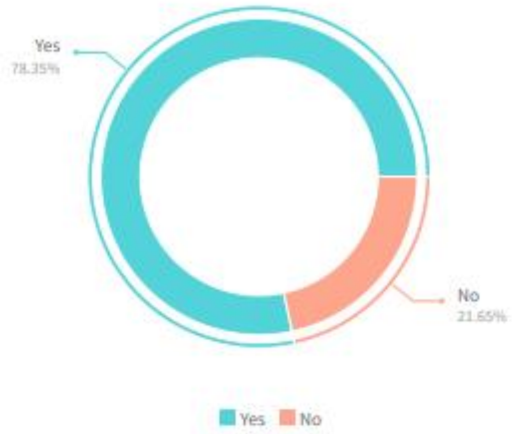
Interpretation: According to the survey,

(61.62%) Say Yes, they know what sentimental analysis/marketing is.

(38.38%) Say No they don't know about it.

QUESTION 12 | YES OR NO

Is Artificial intelligence a good option for opinion computing?



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Yes	76	78.35%
No	21	21.65%

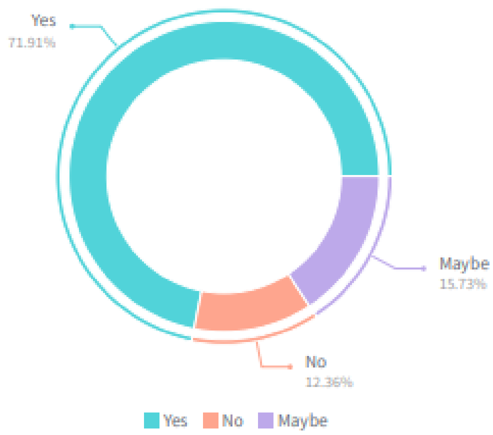
Interpretation: According to the survey,

(49.44%) Say Yes, they think A.I is a good option for marketing.

(28.09%) Say No they don’t find A.I a good option.

QUESTION 13 | MULTIPLE CHOICE

Will like to know more about sentimental analysis/marketing?



ANSWER CHOICES	RESPONSES	RESPONSE PERCENTAGE
Yes	64	71.91%
No	11	12.36%
Maybe	14	15.73%

Interpretation: According to the survey,

(71.91%) Say Yes, they will like to know about sentimental analysis/marketing.

(12.36%) Say No they don’t.

(15.73%) Say they are not sure.

CHAPTER 5 FINDINGS AND CONCLUSION

From the above interpretation of data and analysis on the basis of the objective of studying the following findings can be established that a company must consider artificial intelligence and sentimental analysis for their marketing purposes as referring to the successful model of sentimental analysis and marketing of big companies such as Amazon Twitter.

By looking at the feasibility of artificial intelligence in reference to marketing it seems that use of artificial intelligence in to the companies and in daily lives and in human psychology is very imminent.

people are more tended towards the reviews the ratings and recommended videos or content via any other OTT platforms in order to stand out in the market a company must consider ratings of their products serious seriously.

By looking at the findings It might be safe to say that the ratings are to be considered as new kind PR

As the data suggest the people are more likely to buy and consider the products by looking at their reviews and ratings it is most likely to be not changed and just improved in these aspects.

CHAPTER 6 CONCLUSION

- The growth of social data is exponential, which has given rise to new aspects, such as the subjectivity detection. Subjectivity detection is a natural language processing task that consists of differentiating subjective data (opinions) from objective data (facts).
- By using subjectivity detection, we can filter out the tweets that are objective and find the tweets that are subjective and carry out sentiment analysis only on the subjective data.
- The accuracy of the sentiment analysis can further be increased by implementing methods to fix the misspelled words and correcting the use of any abbreviated or shortened words.
- The use of oxymoron words is another factor that can affect the accuracy of sentiment analysis which can be corrected by using a new model to detect and replace the oxymoron words with equivalent words that are effectively analysed by the sentiment analyser.
- Sentiment Analysis is a precise science that facilitates the development of new ideas as well as promotions and the public. Data sets taken from Amazon.com have been used which contain product feedbacks. Precision, Recall, and F1 has been used as evaluation measures.
- The algorithms used are NBA and SVM. By choosing a judgment limit that maximizes distance from the nearest data points in all categories, SVM varies from other conventional algorithms. SVM not only considers a limit for the decision but finds the best limit for options. The best choice is to be as far away from the closest point of all groups as possible.
- The nearest points to the boundary of judgment, which maximizes the distance from the boundary for judgments, are considered to help vectors as shown. The decision boundary when using vector-assisted systems is regarded as a fixed margin or median margin when using vector-assisted devices.
- Through context-based sentiment analysis, the model will analyze customer behaviour for each product and then try to improve the sales based on the sentiments portrayed for each product whether they be positive or negative.

6.1 SUGGESTIONS AND RECOMMENDATIONS

- Company should give utmost attention towards the research and innovativeness as to survive in the market campaign should be such that the people are made to give reviews.
- Social media websites should consider accuracy more in terms of oxymoron or euphemism or a metaphor.
- Company should be bold and shouldn't hesitate to embrace the Artificial intelligence In there company.
- Proper machine and human intelligence should be managing in the company environment.
- To maximizes the distance from the boundary for judgments, are considered to help vectors as shown.
- Try to improve the sales based on the sentiments portrayed for each product whether they be positive or negative.
- The algorithms used are NBA and SVM. By choosing a judgment limit that maximizes distance from the nearest data points in all categories, SVM varies from other conventional algorithms.
- To use subjectivity detection, we can filter out the tweets that are objective and find the tweets that are subjective and carry out sentiment analysis only on the subjective data.

BIBLIOGRAPHY:

www.marketingevolution.com

www.ama.org

www.cmswire.com

shodhganga.inflibnet.ac.in

https://www.researchgate.net/

https://scholar.google.com/

https://www.sciencedirect.com/science/article/pii/S2090447914000550

https://www.sciencedirect.com/science/article/abs/pii/S1751157709000108

https://www.sciencedirect.com/science/article/pii/S1018363916300071

https://shodhganga.inflibnet.ac.in/handle/10603/279998

https://shodhganga.inflibnet.ac.in/simple-search?query=SENTIMENTAL+MARKETING&go=

https://shodhganga.inflibnet.ac.in/handle/10603/310405

https://shodhganga.inflibnet.ac.in/handle/10603/261292

ANNEXURE

A study on sentimental analysis/marketing in reference to Artificial intelligence

1.

Name

Email

Gender: Male Female Transgender

what's your employment status?

Age

2. How often do you shop online?

☐ Often

☐ Very often

☐ Rarely

☐ Very Rarely

3. Which apps or websites you usually shop?

☐ 

☐

☐ snapdeal

☐ NYKAA

☐ JABONG .COM

☐

4. When did you last buy something online ?

- ☐ Less than 1 month ago
- ☐ Between 1 and 6 months ago
- ☐ Between 6 months and 1 year ago
- ☐ More than 1 year ago
- ☐ I don't remember

---

5. Do you consider reviews and ratings before buying the product?

- ☐ Yes
- ☐ No

---

6. Do you feel comfortable trying a new product that has no reviews or ratings?

- [ ] Yes
- [ ] No
- [ ] Maybe

---

7. Do you watch twitter polls?

- [ ] Yes
- [ ] No

---

8. Do you believe or trust twitter polls and product ratings?

- [ ] Yes
- [ ] No
- [ ] Maybe

9. Do you watch recommended videos on youtube or any other OTT platforms?

☐ Yes

☐ No

---

10. What do you find more comfortable receiving a call for reviews or giving a star rating on an app or website.

☐ Calls

☐ Ratings

☐ Neither of them

---

11. Are you aware about the term sentimental analysis/marketing?

☐ Yes

☐ No

12. Is Artificial intelligence a good option for opinion computing?

- [ ] Yes
- [ ] No

---

13. Will like to know more about sentimental analysis/marketing?

- [ ] Yes
- [ ] No
- [ ] Maybe