major project

by Saumya Satija

Submission date: 27-May-2019 09:16AM (UTC-0400)
Submission ID: 1136523649
File name: doc1.pdf (1.61M)
Word count: 10653
Character count: 70475

Report on

TO EVALUATE THE PERFORMANCE OF DIFFERENT SUPERVISED LEARNING ALGORITHMS IN PREDICTING THE FAKE ONLINE HOTEL REVIEWS

Submitted By:

Komal Sharma: 2K17/MBA/726
Saumya Satija: 2K17/MBA/748

Under the guidance of:

Ms. Kusum Lata

Assistant Professor

UNIVERSITY SCHOOL OF MANAGEMENT & ENTREPRENEURSHIP

Delhi Technological University

Vivek Vihar, Phase 2, Delhi-110095

May 2019

CERTIFICATE

This is to certify that the project report titled "To evaluate performance of the different supervised learning algorithms in detecting fake online hotel reviews" is a record of the project work carried out by Saumya Satija and Komal Sharma under the guidance of Ms Kusum Lata.

This project is for the fulfillment of Masters of Business Administration from University School of Management And Entrepreneurship, Delhi Technological University

_____

Signature of Mentor

DECLARATION

We hereby declare that the project work entitled "*TO EVALUATE THE PERFORMANCE OF DIFFERENT SUPERVISED LEARNING ALGORITHMS IN PREDICTING THE FAKE ONLINE HOTEL REVIEWS*" submitted to the USME, DTU, is a record of an original work done by us under the guidance of Ms.Kusum Lata, Assistant Professor, USME, DTU and this project work is submitted in the partial fulfillment of the requirements for the award of the degree of Master of Business Administration. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Saumya Satija

2K17/MBA/748

Komal Sharma

2K17/MBA/726

ACKNOWLEDGEMENT

It is our pleasure to be indebted to various people, who directly or indirectly contributed in the development of this work and who influenced our thinking, behavior, and acts during the course of study.

We express our sincere gratitude to the authorities who gave us an opportunity to take this project.

We are highly intended and extremely thankful to Ms. Kusum Lata for her support, cooperation, guidance and suggestions that helped us in completing this project.

Saumya Satija

Komal Sharma

ABSTRACT

In this era of digitization electronic-commerce has taken a vital place in everyone's life and online reviews help people get more information from the others opinion using different platforms. For product designing the online media become an important platform and it is also get target for the opinion spamming. The potential customers tend to make decision according to these reviews before purchasing the products. However, driven by profit, spammers post fake reviews to mislead the customers by promoting or demoting target product or services offered online. As there is no mechanism to control the reviews posted, anybody can write anything unanimously which conclusively leads to fake reviews.

Recently fake review detection has attracted significant attention of both businesses and the academicians. Researchers for years working on the fake review detection using supervised learning , ground truth of the larger scale datasets were still mat been used and mostly approaches of supervised learning are based on pseudo fake reviews rather than real fake reviews. For reviews to have genuine user experiences and opinions differentiating between fake and genuine reviews is an important thing. Supervised learning has been one of the approaches used for solving the problem. Precisely, we provide comparison of four supervised machine learning algorithms: Naïve Bayes (NB), Decision Tree (DT-J48), K-Nearest Neighbor (K-NN), K-star (K*) and Support Vector Machine (SVM) for sentiment classification using two datasets of reviews. In this paper, we have tried to collect reviews from online hotel booking platforms and used the sentiment analysis technique along with supervised classification algorithms and conducted a comparative analysis.

Keywords: Electronic-commerce, Online reviews, spam reviews, sentiment analysis, supervised learning.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

Abreviation	Full Form
FTA	Foreign tourist arrivals
FEE	Foreign Exchange Earnings
WEF	World Economic Forum
GST	Goods and services tax
OTA	Online Travel Arrangements
POS	Part-of-speech
WEKA	Waikato Environment for Knowledge Analysis tool
TN	True Negative review
FN	False Negative review
TP	True Positive review
FP	False Positive review
SVM	Support Vector Machine
K-NN	K-Nearest Neighbor
STWV	StringToWordVector

INTRODUCTION

Availability of millions of products and services online makes it difficult to find the best suitable product according to the needs because of availability of numerous alternatives. To remove the confusion the most popular approach used by people all over is to look for online reviews of others who have already experienced. Online reviews become an important assets for buyers for their decision making while making online purchases. These reviews are used by individuals, manufacturers, buyers and retailers for purchasing and business decisions'. With the growing popularity of websites such as Goibibo and TripAdvisor where people can state their opinion on different hotels and rate them, hotel industry is replete with reviews and ratings. Thus, it is easy to find reviews on specific type of staying.

Customers make use of these online reviews for reference about products and also give their own reviews regarding their own experiences.' With increase in the number of online platforms it has led to an increase information available. In order to get short term profit, increased publicity influencers try create fake reviews and recently there has been considerable growth. These reviews are produced by people who have not personally experienced on the subjects but have reviewed these types of reviews are called spam, fake, deceptive or reviews. During this process, users often want to obtain more about detailed information about the hotels, especially from the experiences of past customers. As a result, online hotel reviews are becoming vital for customers to obtain additional user-centered knowledge about the specific property. However, because both owners and customers are aware of the importance of the online reviews, some owners are paying customers to write good reviews so they can boost their revenue through online sales. On the other hand, malicious users also use fake reviews to attack their competitors, resulting in unfair market competition and financial loss of the customers The reviews are of two types the positive reviews which are written in order to build reputation and generate a positive online word of mouth and the negative reviews which are written in order to defame, demote and generate a negative online word of mouth. At times, to create better ratings for the venue, hotel owners pay employees to fabricate false reviews.

Traditional methods of data analysis have long been used to detect fake/fraudulent reviews. Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. Some of these techniques facilitate useful data

interpretations and can help to get better insights into the process behind data. To go beyond a traditional system, a data analysis system has to be equipped with considerable amount of background data, and be able to perform reasoning tasks involving that data. In effort to meet this goal researchers have turned to the fields of machine learning and artificial intelligence. A review can be classified as either fake or genuine either by using supervised and/or unsupervised learning techniques. These methods seek reviewer's profile, review data and activity of the reviewer on the Internet mostly using cookies by generating user profiles. Using either supervised or unsupervised method gives us only an indication of fraud probability.

With the use of supervised as well as unsupervised techniques one can identify the fakeness of the review. Supervised machine learning is a machine learning algorithm which has labelled training data. Unsupervised machine learning algorithm is a technique in which there is unlabeled training data. Sentiment analysis is another approach used in the detection process. No standalone statistical analysis can assure that a particular review is fraudulent one. It can only indicate that this review is more likely to be suspicious. Detection and filtering of genuine reviews is an interesting problem for the researchers and e-commerce sites. One such review site that filters fake reviews is yelp.com. The filter used in yelp.com to hide fake reviews from public is a trade secret.

What is learning from data?

For a machine to make such choices, the intuitive way is to model the problem into a mathematical expression.

- The knowledge acquired from human understating is called domain knowledge.
- The knowledge learned from given training data is called data driven knowledge.

Machine learning is a mix of different techniques, the methods for learning can typically be categorized as three general types:

Supervised learning: The learning algorithm is given labeled data and the desired output. For example, pictures of dogs labeled “dog” will help the algorithm identify the rules to classify pictures of dogs.

Supervised learning is classified into two types:

- Regression: Problems where output is continuous or a number are known as regression problems.
- Classification: Problems where output is discrete are known as classification problems.

Examples of supervised learning are spam prediction, fault prediction. In supervised learning, classification techniques such as decision tree, neural networks, and support vector machines are used.

Unsupervised learning: The data given to the learning algorithm is unlabeled, and the algorithm is asked to identify patterns in the input data. For example, the recommendation system of an e-commerce website where the learning algorithm discovers similar items often bought together. In unsupervised learning, clustering methods are used to identify patterns from unlabeled samples.

Reinforcement learning: The algorithm interacts with a dynamic environment that provides feedback in terms of rewards and punishments. For example, self-driving cars being rewarded to stay on the road.

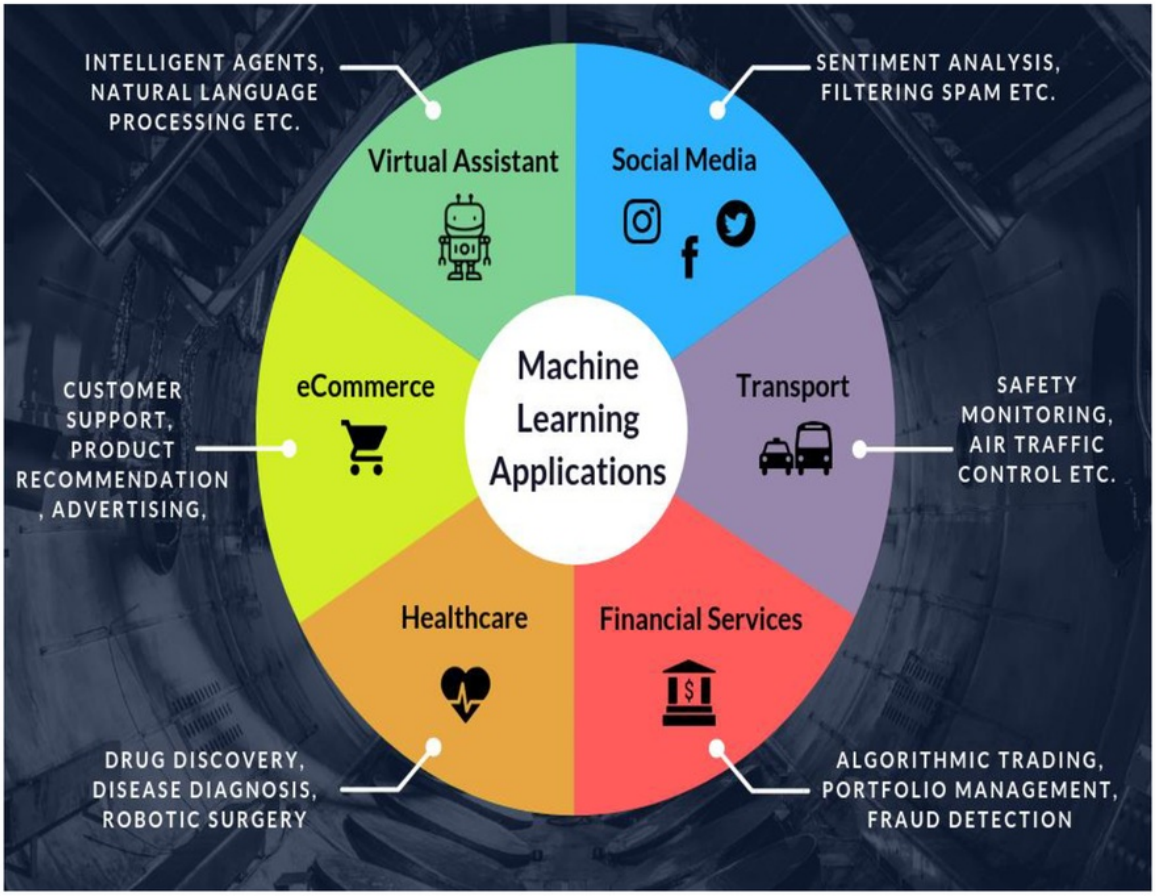


Figure 1: Applications of Machine Learning

INDUSTRY PROFILE

The Indian tourism and hospitality industry has emerged as one of the key drivers of growth among the services sector in India. Tourism in India has significant potential considering the rich cultural and historical heritage, variety in ecology, terrains and places of natural beauty spread across the country. Tourism is also a potentially large employment generator besides being a significant source of foreign exchange for the country. During 2018, FEEs from tourism increased 4.70 per cent* year-on-year to US$ 28.59 billion. FEEs during January 2019 was US$ 2.55 billion. [11] Tourism is a major engine of economic growth and an important source of employment & foreign exchange earnings in many countries including India. It has great capacity to create large scale employment of diverse kind – from the most specialized to the unskilled and hence can play a major role in creation of additional employment opportunities. It can also play an important role in achieving growth with equity and sustainability.

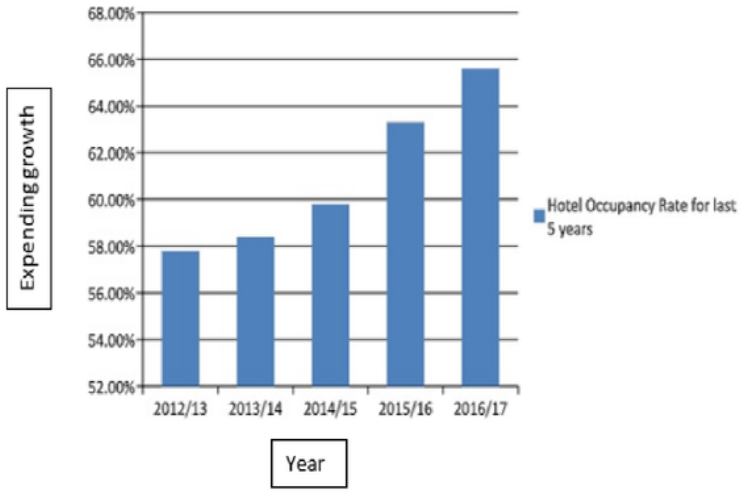
India is the most digitally-advanced traveller nation in terms of digital tools being used for planning, booking and experiencing a journey, India's rising middle class and increasing disposable incomes has continued to support the growth of domestic and outbound tourism.

During 2018, foreign tourist arrivals (FTAs) in India stood at 10.56 million, achieving a growth rate of 5.20 per cent year-on-year. FTAs in January 2019 stood at 1.10 million, up 5.30 per cent compared to 1.05 million year-on-year. [11] The travel & tourism sector in India accounted for 8 per cent of the total employment opportunities generated in the country in 2017, providing employment to around 41.6 million people during the same year. The number is expected to rise by 2 per cent annum to 52.3 million jobs by 2028.

International hotel chains are increasing their presence in the country, as it will account for around 47 per cent share in the Tourism & Hospitality sector of India by 2020 & 50 per cent by 2022. The Government has also been making serious efforts to boost investments in tourism sector. In the hotel and tourism sector, 100 per cent FDI is allowed through the automatic route. A five-year tax holiday has been offered for 2, 3 and 4 star category hotels located around UNESCO World Heritage sites (except Delhi and Mumbai). Total FDI received by Indian hotel & tourism sector was US$ 12 billion between April 2000 and December 2018. India is a large market for travel and tourism. It offers a diverse portfolio of niche tourism products - cruises, adventure, medical, wellness, sports, MICE, eco-tourism, film, rural and religious tourism. India

has been recognized as a destination for spiritual tourism for domestic and international tourists.

The ever expanding growth of the Indian travel and tourism industry has a cascading effect on the hotels industry, boosting the occupancy rate and the average room rate.



Source: Ministry of tourism (annual report 2017-18)

Key trends

The key trends that have emerged in the Indian hotel industry are mentioned below:

Asset-light model: In the recent times, the asset-light model has gained significant importance with both international as well domestic biggies in the hotel industry opting for it. The trend was started in India by the major global hotel groups like Marriott and Accor and soon the domestic players like Taj, Oberoi, ITC and Leela are following suit.

Earlier this year (February 2018) Taj (from the Indian Hotels Company Ltd.) announced its five- year strategy revealing that by the year 2022, 60% of its assets will not be owned by the company. "Our strategy is three pronged: Restructure, Reengineer and Reimagine our portfolio to achieve 8% point EBIDTA margin improvement.[11]

This will be driven by a deep commitment to service excellence as well as implementation of revenue and profit-driving initiatives. Integral to our strategy of reinforcing the multi-product, multi-segment brand scape is our customer" said, Puneet

Chhatwal, Managing Director and Chief Executive Officer, Indian Hotels Company Limited.

Even budget hotels like LemonTree which started with solely owned properties are adapting to this new model. In fact, 22 of its upcoming hotels which 1,775 rooms are under management contracts.

Mid-market boom: The mid-market refers to hotels falling under two/three/four-star properties categorized as business hotels, resorts, boutiques, havelis, and full-service or limited- service hotels with average pricing being INR 4,000 or less for a night.

An upsurge in travel of the middle class, increase in business and leisure travel, growing urbanization, strengthened economic growth and the doubling of air travel over the past seven years have boosted the mid-market brands in the hotel industry in India.

The mid-market of the Indian hotel industry mainly caters the domestic business and leisure travelers of the country. Driven by the travels of upper-middle class and middle-class the budget hotels hotels have a key role to play in the growth of this sector in the times to come.

Local flavor: Modern travelers including the millennials are more and more seeking to connect with the local people, culture and food while travelling. Many hotels adopting the strategy to weave in the destination and culture through various encounters for the guests by curating experiences around local cuisine, art, architecture, and rituals. For example, ITC Hotels' Food Sherpa programme - where the hotel chef takes guests on a culinary tour of the city. Experts say, it is no more about a room, creating unique experiences will be key to win the guests.

Travellers want to mingle with locals, and hotels will need to facilitate that. Jean-Michel Cassé, chief operating officer, India and South Asia, AccorHotels, says, "At Novotel hotels, we are introducing opportunities for our guests to mix with locals through activities such as yoga sessions, guitar lessons and art classes to make our hotels more sociable places for locals and visitors alike. We will also create shared workspaces and invite local entrepreneurs into our hotels to work and socialize."

- The international travel and tourism industry continues to be one of the largest global industries and a major engine of economic growth. At present, 1 in every 11 people worldwide are employed by the tourism sector, with the direct contribution of travel and tourism to GDP US$ 234 billion (9.4% of total GDP) in 2017[11].
- The Indian hospitality industry has been instrumental in contributing to the nation's economic growth. This trend is expected to continue especially with the

introduction of e-visa for foreign tourists and with the domestic economy improving, there are clear signs of increased domestic travel. The growth rate in room demand (about 6.8%) has been consistently outpacing the supply (about 3%) growth in India for the past few years.

- As per a report by World Economic Forum (WEF), India was ranked 12th in the Asia Pacific region and 40th overall in the list of the world's attractive destinations. Further, India ranked 7th among 184 countries in terms of travel & tourism's total contribution to GDP in 2016.
- India's Tourism sector has been performing well with Foreign Tourist Arrivals (FTAs) growing by 9.7% to 8.8 million and Foreign Exchange Earnings (FEEs) at 8.8% to US$ 22.9 billion in 2016. FTAs during 2017 were 10.2 million, with a growth of 15.6%, while FEEs from tourism were US$ 27.7 billion, with a growth of 20.8% over 2016[11].
- As per the Ministry of Tourism, FTAs on e-Tourist Visa grew by 143% to 10.8 lakh in 2016, and further grew by 57.2% to 17.0 lakh during 2017. The growth here was attributable to the introduction of e-TV for 161 countries from 113 countries earlier [11].
- Domestic travel spending also witnessed impetus, attributing 87.2% to the direct Travel & Tourism GDP. The appreciation of the US dollar has made international travel unattractive to many who are now seeking to travel within the country for holidays.
- Amongst the Emerging Market and Developing Economies, China's economic growth in 2017 stood at 6.9% while Indian economy grew at 7.1%. India continues to be among the world's fastest growing major economies, despite temporary hiccups caused by demonetisation and goods and services tax (GST) implementation.
- The Union Cabinet has approved a MoU between India and South Africa, aimed at expanding bilateral cooperation in the tourism sector through exchange of information and data, establishing exchange programmes and increasing investments in the tourism and hospitality sector.
- The Ministry of Tourism has the main objective of increasing and facilitating tourism in India. Augmenting tourism infrastructure, easing of visa regime, assurance of quality standards in services of tourism service providers, projecting the country as a 365 days' tourist destination, promoting tourism in a sustainable manner, etc. are some of the policy areas which need to be constantly worked upon to increase and facilitate tourism in India.

- As a step in this direction, Ministry of Tourism has recently launched the 'Adopt A Heritage' project. Heritage sites are being offered for adoption by the public sector, private sector and individuals to become 'Monument Mitra's for developing amenities and facilities at these sites under this programme.
- In the Union Budget 2017-18, the government has proposed to establish five special tourism zones and increase the focus towards rural infrastructure development and introduction of bio-toilets. Under Budget 2017-18, the government allotted US$ 142.8 million for integrated development of tourist circuits under Swadesh Darshan scheme.
- Further, US$ 14.8 million was allocated for promotion & publicity of various programmes & schemes of the Tourism ministry.

In the long term, the demand-supply gap in India is very real and that there is need for more hotels. The shortage is especially true within the budget hotels and the mid-market hotels segment. There is an urgent need for budget and mid-market hotels in the country as travellers look for safe and affordable accommodation. Various domestic and international brands have made significant inroads into this space and more are expected to follow as the potential for this segment of hotels becomes more obvious.

The past year saw the start of an uptrend in the hospitality sector but was also marked by multiple challenges like curbs on liquor sale and the GST rollout. With these issues behind them, hotel operators expect business growth to pick up momentum in 2018. The growth is expected to come from the rise in online bookings. Hotel bookings is one of the least penetrated segments in the travel categories in India.

With a rise in online competition, popular models have come up with online travel agents (OTAs) offering a single marketplace for all travel-related needs. There are also seen Meta search engines like TripAdvisor and MakeMyTrip, that operate like travel discovery platforms. Further, online accommodation reservation services like Oyo Rooms have gained popularity. Apart from this, branded hotels are seen operating direct bookings through their websites [12].

Apart from the above, the Indian government has realised the country's potential in the tourism industry and has taken several steps to make India a global tourism hub. The "Clean India" campaign and development of inland waterways for transport and

tourism are projects that have gained momentum over the previous year. Additionally, programmes such as "Make in India" and the "Smart Cities" initiative have highlighted the Government's support to skill development and investments in Hospitality and Tourism.

Apart from the above initiatives, the government has proactively sought foreign investment from countries such as China, the United States and Japan, leading to an increase of business related travel to the country.

It should be noted that that the base for tourism in India is still very low. The spurt in demand for hotel accommodation over the last few years has inflated hotel rooms in the country. However, a number of international brands across all hotel segments are planning to or have recently entered the Indian market. Furthermore, domestic hotel chains, too, are embarking on strong expansion and development plans across all hotel segments.

India, after China, is considered as one of the most lucrative hotel markets in the world and has the second largest construction pipeline in Asia. Growing affluence, potential for economic growth, increases in disposable incomes and the burgeoning middle class are expected to drive both leisure and business travel in the coming years.

The long-term outlook for the Indian hospitality business continues to be positive, both for the business and leisure segments. The sector has potential for growth on the back of increases in disposable incomes, increase in foreign tourist arrivals, momentum from government-led initiatives, and the burgeoning middle-class population.

Online consumers tend to utilize numerous websites like Yatra, Trivago, Goibibo, Cleartrip, Expedia etc. as a primary tool for booking travel products due to the variety of product offerings, quick price comparisons, time savings and ease of use when requesting services to fulfill their needs. Online booking availability not only benefits customers by making travel arrangements easier, it also increases the profits of businesses such as airlines, hotels and other package tour companies.

ORGANIZATION PROFILE

Hotel booking has become a real easy job with the development of internet and rise of many online hotel booking websites. These websites provide a detail description of the hotels with which they have a tie up and a person can, just by login into the booking portals, can get the required room of their choice fitting into their budget. The site provides a detail view of the hotel with the help of a picture gallery. Planning the whole trip has become really easy with the help of online booking portals.

With regard to gathering information for a hotel stay, business travellers most often follow their company's recommendation for a hotel, although many of them use search engines or online travel agents to learn more about available hotels. In contrast, recommendations of friends and colleagues are most important to leisure travellers, followed by travel-related websites, search engines, and OTAs, Online Travel Arrangements. Once the information is gathered, however, travellers of all kinds turn more to such sources as the brand website, OTAs, and TripAdvisor. Late in the decision process, the respondents tended to land on the brand websites or go to an OTA, where they can book their room.

Price transparency of online channels adds more pressure to hotel room rates thereby forces hotels to keep rate parity in all channels, keep online rates as low as possible, or provide "low price guarantees" on hotel websites. The travel intermediaries consist of third-party travel agencies (e.g., Bookmyhotel.com), social media sites (e.g., Yatra.com) and search engines (e.g., Google, yahoo). Most consumers are concerned with acquiring good value for their money instead of solely seeking the lowest possible price.

Apart from the above, the Indian government has realized the country's potential in the tourism industry and has taken several steps to make India a global tourism hub. The "Clean India" campaign and development of inland waterways for transport and tourism are projects that have gained momentum over the previous year. Additionally, program such as "Make in India" and the "Smart Cities" initiative have highlighted the Government's support to skill development and investments in Hospitality and Tourism. It should be noted that that the base for tourism in India is still very low.

The spurt in demand for hotel accommodation over the last few years has inflated hotel rooms in the country. However, a number of international brands across all hotel segments are planning to or have recently entered the Indian market. Furthermore, domestic hotel chains, too, are embarking on strong expansion and development plans across all hotel segments.

The deepening penetration of internet usage and smartphones in India has led to increased booking of hotels through online portals and applications in recent times, the report added. While online travel agents like Makemytrip, Cleartrip, Yatra and Goibibo continue to dominate the travel bookings industry on the internet, online accommodation reservation services like Oyo Rooms, Stayzilla are gaining popularity. In addition, meta search engines like TripAdvisor and Kayak, that operate like travel discovery platforms have been able to establish presence in the Indian market and this segment is also expected to attract competition in coming times. Internet accessibility and usage are happening more in our everyday lives than ever before and as such have also become an important factor in modern travel behavior.

With OTAs and online accommodation booking services starting loyalty schemes, its market share compared to the direct bookings channels from branded hotels is expected to improve drastically. Overall, the Indian hotel industry is expected is register a huge surge in online bookings and the low penetration levels coupled with the increasing smartphone and internet usage is expected to turn into a big opportunity for home-grown start-ups and international brands alike.

TRIPADVISOR

TripAdvisor, enables travelers to unleash the full potential of every trip. With 702 million reviews and opinions covering the world's largest selection of travel listings worldwide covering 8 million accommodations, airlines, experiences, and restaurants. TripAdvisor provides travelers with the wisdom of the crowds to help them decide where to stay, how to fly, what to do and where to eat. TripAdvisor also compares prices from more than 200 hotel booking sites so travelers can find the lowest price on the hotel that's right for them. TripAdvisor-branded sites are available in 49 markets, and are home to the world's largest travel community of 490 million average monthly unique visitors, all looking to get the most out of every trip.

TripAdvisor Media Group operates 25 travel brands. TripAdvisor is where we go to praise, criticise and purchase our way through the inhabited world. It is, at its core, a guestbook, a place where people record the highs and lows of their holiday experiences for the benefit of hotel proprietors and future guests. But this guestbook lives on the internet, where its contributors continue swapping advice, memories and complaints about their journeys long after their vacations have come to an end.

Every month, 456 million people – about one in every 16 people on earth – visit some tentacle of TripAdvisor.com to plan or assess a trip. For virtually every place, there exists a corresponding page. The Rajneeshee Osho International Meditation Resort in Pune, India, has 140 reviews and a 4 out of 5 rating, Cobham Service Station on the M25 has 451 reviews and a rating of 3.5, while Wes Anderson's fictional Grand Budapest Hotel currently has 358 reviews and a rating of 4.5.[12]

Over its two decades in business, TripAdvisor has turned an initial investment of $3m into a $7bn business by figuring out how to provide a service that no other tech company has quite mastered: constantly updated information about every imaginable element of travel, courtesy of an ever-growing army of contributors who provide their services for free. Browsing through TripAdvisor's 660m reviews is a study in extremes. As a kind of mirror of the world and all its wonders, the site can transport you to the most spectacular landmarks.

Despite its recent difficulties, the number of reviews on TripAdvisor keeps growing. At present, more than 200 new posts are uploaded to TripAdvisor every minute.

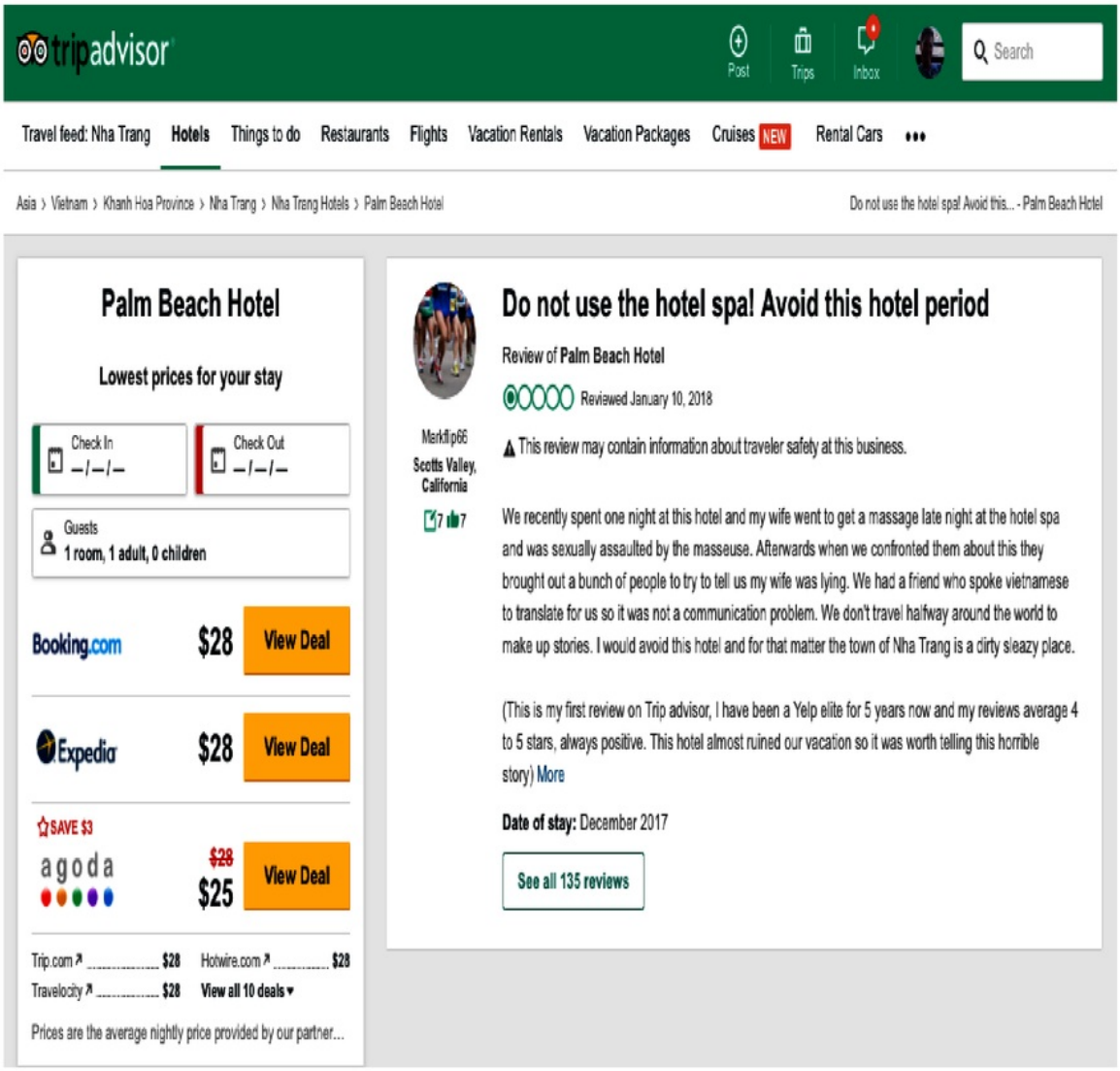


Figure 1: Hotel Reviews shared on Tripadvisor

GOIBIBO

Goibibo.com is among top three most searched sites for making a hotel reservation online. The site was started in the year 2009 and over is a flagship company of the larger group, ibibo, which provides completes travel solutions and has many companies operating under it. One can find a brief descriptions about the place for which one is finding hotels. The site also provides an option of sub search for example within a city, one can make a search on the basis of area he wants to stay in like near railway station, city center and many more.

Ibibo Group is an online travel organization founded in January 2007 by Ashish Kashyap. The company is a subsidiary of Naspers, which owns an 80% stake in Ibibo Group. In February, Naspers announced plans to increase its stake in Ibibo Group to 90% by investing additional $250 million in the company.

With PayU, a business can process payment by using credit cards and debit cards (for more than 50 banks), and online banking for banks including ICICI, HDFC, SBI, and Axis Bank. PayU provides an API for integration and transaction analytics to improve the speed and security of payment processing.PayU India launched of a first-of-its kind premium deferral payment facility for consumers 'Lazy Pay' aimed at those who transact digitally for any amount between Rs 500 and Rs 2,500 and is an option to pay later. [12]

Businesses have also developed more subtle tactics designed to stop critical reviews from appearing in the first place. In July, many were fined for suppressing negative reviews of its rental apartments by withholding the email addresses of disgruntled guests from this site, ensuring that the company could not prompt them to write a review.

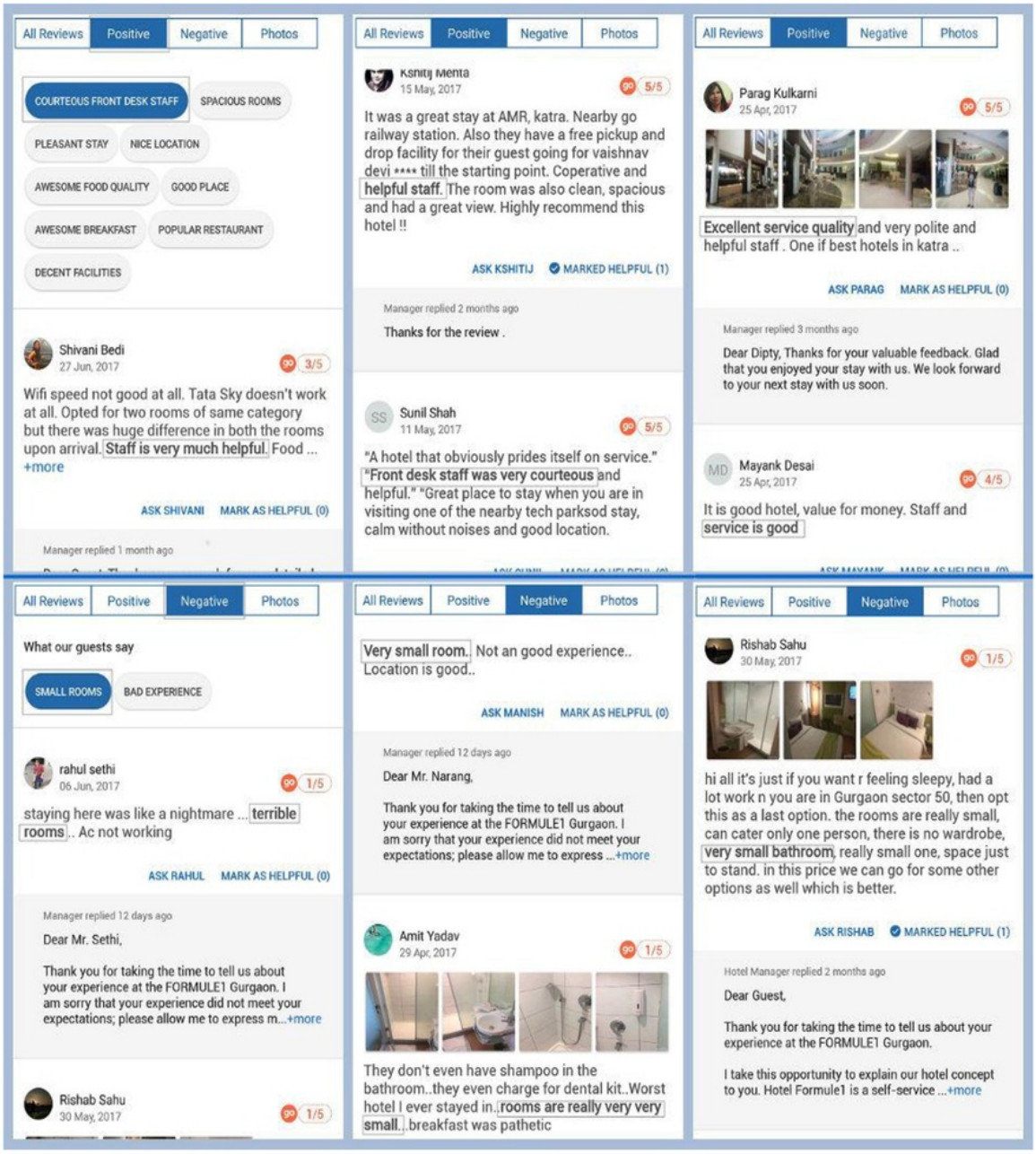


Figure 2: Hotel reviews on goibibo

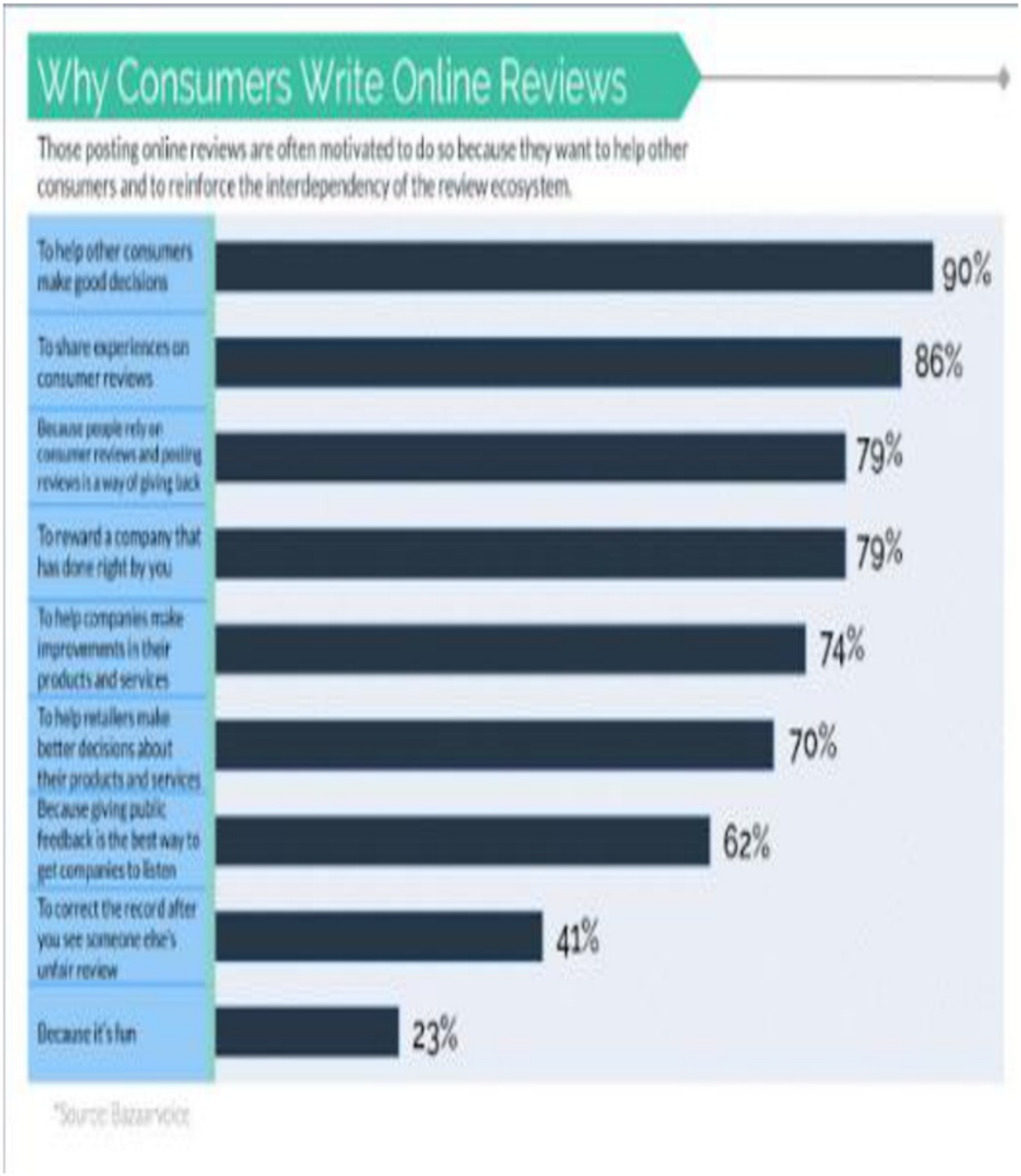


Figure 2: Why Consumers post online

Figure 3: Importance of Online Reviews for Hotels

OBJECTIVE OF STUDY

In this project, our objective is to study the performance of different approach using supervised learning classification algorithms in context of predicting fake and real online hotel portals reviews. This approach is based on sentimental features of a reviews present on online hotel booking portals.

The main contribution in this project are stated as follows:

- To find out which classification algorithm will more accurately predict the nature of reviews.
- Two datasets from different sources is obtained, and the features are identified to carry out the study.
- Supervised learning methods are applied to detect fakeness of review separately for both the datasets.

LITERATURE REVIEW

Fake review detection has attracted significant attention from both businesses and the research community. For reviews to reflect genuine user experiences and opinions, detecting fake reviews is an important problem. Online reviews are increasingly used by individuals and organizations to make purchase and business decisions. Positive reviews can render significant financial gains and fame for businesses and individuals. Unfortunately, this gives strong incentives for imposters to game the system by posting fake reviews to promote or to discredit some target products or businesses. It is true that any hotel or restaurant can be the victim of fake reviews, particularly if they are negative. While rare, it can happen, whether it's an unethical competitor or an individual who has decided to cause problems for a business.

Mukerjee et al. [1] used filtered and unfiltered data from yelp.com to analyze real and pseudo reviews. While analyzing reviews, authors noticed that the quantity of fake reviews on yelp.com is much smaller than real reviews. POS (part-of-speech) based features to build classifiers. All experiments were based on 5-fold cross validation and yielded 67.8% accuracy.

McCallum and Nigam [6] performed an evaluation of two event models for the Naïve Bayes Text Classification. The first model was a multi-variate Bernoulli event model. This model calculates the probability of the document, by multiplying the probability of all attribute values, including the probability of non-occurrence for words that were not present in the document. The second model was the multinomial event model. This model calculates the probability of the document, by multiplying the probability of words that occurred in that document. Authors completed the empirical study with different data sets, such as the Newsgroups and WebKB. Results showed that the second model outperformed the first one in terms of large vocabulary sizes. The multinomial model reduced accuracy error by 27%.

There are thousands of reviews online, which makes it convenient for people to make decisions, but the amount of data makes it difficult to sort through [4]. The real value of online reviews is in its content and the certainty that reviewer indeed received products or services prior to writing the review. Promotion or demotion of the products and services is one of the main reasons for deceptive reviews. At times, to create better ratings for the venue, hotel owners pay employees to fabricate false reviews [5]. Alternatively, some reviewers write negative reviews for malicious reasons, like to distort the reputation of the business reviewed [6].

Ott et al. [7] define deceptive opinion spam as "fictitious opinions that have been deliberately written to sound authentic in order to deceive the reader." To analyze deceptive opinion spam, authors performed a hotel review analysis, using a data set of 800 reviews. Turker was used and tasked to write a deceptive review on one of the 20 Chicago hotels. Previous works mainly use rating score and the feedback score as indicator detect the spam reviews. Rating is regarded as representation of reviewer's sentiment orientation, generally the five star represent the high satisfaction while one star means poor satisfaction. Nevertheless these methods have shortcoming by using rating score as indicator. First, the rating will not necessarily completely represent the sentiment of the reviewer. There exist some positive reviews with low rating and some negative reviews with high rating. All these case belong to the inconsistent of the reviews. At the same time, even though two reviews have same ratings, the different content will produce different influence to the reader. The case mentioned above should not be regarded as noisy data. Potential customer will make decision after reading the content carefully. Compared with rating score, the content of the reviews will represent more accurate sentiment of the reviewer. Therefore it will indeed influence the potential customer. [1]

Yoo and Gretzel (2009) [15] gather 40 truthful and 42 deceptive hotel reviews and, using a standard statistical test, they have manually compared the psychologically relevant linguistic differences between them.

According to a recent study vendors with the best reputation have an increased number of sales. However, promoting trustworthy participation also bears an incentive for malicious actors to push their reputation unfairly to gain more benefit. Dishonest reviews or ratings have already become a serious problem in practice. Thus, in this research, our primary goal is detecting unfair reviews on reviews through Sentiment Analysis using supervised learning techniques in an E-Commerce environment. Our research is fundamentally focused at the document level of Sentiment Analysis, precisely on datasets of reviews. Sentiment Analysis methods will have a fundamental positive effect on reputation systems, especially in unfair reviews detection processes in an e-commerce environment and other domains. Feedback reviews in e-commerce is an important source of information for customers to reduce product uncertainty when making purchasing decisions. However, with increasing volume of feedback reviews, customers sometimes make product buying decisions based on unfair or fake feedback reviews.

Because majority spam reviews are written by humans (there is also a small portion of synthetically generated reviews), [2] one of the most effective ways to distinguish spam

and non-spam reviews is by using machine learning techniques, which has proved its ability in such problems especially when dealing with text and natural language. When training classifiers for spam review detection, one essential task is to obtain a sufficient number of training data for the learning and evaluation phases. In many situations, it is rather difficult to distinguish spam and non-spam reviews manually from a novice or expert user.

In the domain of spam and non-spam review detection, non-spam reviews are often the majority population, and the spam or fake reviews are relatively rare and difficult to obtain. This naturally leads to imbalanced data distributions where positive samples (i.e. spams) are only a small portion of the training samples, and majority samples belong to the negative class (i.e., non-spams). The main challenge of learning from imbalanced data distributions is that the classifiers learned from imbalanced are biased to the prior class distributions, and will tend to classify all examples into the majority class group.

In machine learning, imbalanced data distributions often happens because of the lack of examples from the minority class (which is often of user's special interest). Imbalanced data easily compromises the performance of any learning algorithm [9]. Technically speaking, the imbalanced dataset is the dataset that has large number of examples or instances for one class compared to the other class, so when the number of instances for class A is less than or greater than class B, the dataset is considered imbalanced.

RESEARCH METHODOLOGY

Research methodology process includes a number of activities to be performed. These are arranged in proper sequence of timing for conducting research. One activity after another is performed to complete the research work. The research methodology was divided into following steps which involved supervised algorithm and WEKA (Waikato Environment for Knowledge Analysis) tool. Weka is open source software for data mining under the GNU General public license. This system is developed at the University of Waikato in New Zealand. "Weka" stands for the Waikato Environment for knowledge analysis. Weka is freely available at http://www.cs.waikato.ac.nz/ml/weka. The system is written using object oriented language java. Weka provides implementation of state-of-the-art data mining and machine learning algorithm. User can perform association, filtering, classification, clustering, visualization, regression etc. by using weka tool [14] .Our study's main goal is to study two reviews datasets which contain fair reviews or unfair reviews to find out which supervised learning techniques perform better.

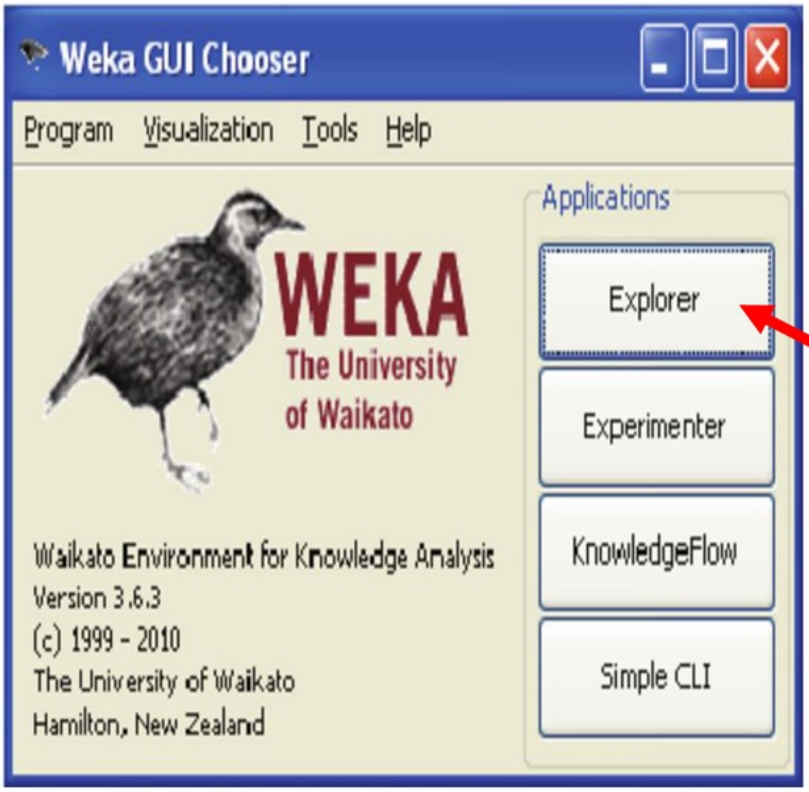


Figure 4: WEKA Application Interface

WEKA tool provides following applications:

Explorer provides functions and support environment for performing data mining processes it helps in accessing to databases, exploration and selection of data and data processing which is aimed at importation, transformation (application of filters) and data extraction. The visualize functionality aimed at the visualization of data using graphic techniques.

Predictive and descriptive modelling compiles a wide range of data mining procedures for the knowledge models.

The Simple CLI allows the user to modify Weka by integrating new functionality developed in Java code, using its structure and object oriented functional design.

Step1: Data collection

We collected the reviews from online hotel booking portals mainly trip advisor and goibibo for Indian hotels. We have based our experiment on analyzing the standard dataset's sentiment value using machine learning algorithms. We have used the

reviews' original datasets to test our reviews classification methods. Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes. The data collection component of research is common to all fields of study.

In order to validate the performance of our method for detecting review spams, we implemented the proposed method using WEKA data mining tool. We classified the entire data into two datasets one from TripAdvisor and other from goibibo.we used a standard parameter for extracting data manually, the table 1 describe the summery of the data we have collected.

Step 2: Data Cleaning

Data cleaning, or *data cleansing*, is an important part of the process involved in preparing data for analysis. Data cleaning is a subset of *data preparation*, which also includes scoring tests, matching data files, selecting cases, and other tasks that are required to prepare data for analysis.

Missing and erroneous data can pose a significant problem to the reliability and validity of study outcomes. Many problems can be avoided through careful survey and study design. During the study, watchful monitoring and data cleaning can catch problems while they can still be fixed. At the end of the study, multiple imputation procedures may be used for data that are truly irretrievable.

The dataset is obtained from tripadvisor and goibibo online portal and data is divided into five scale rating: 1 star, 2 star, 3 star, 4 star, 5 star for tripadvisor and for goibibo 0 star, 1 star, 3 star, 4 star. The original data cannot be used for forming the model and usually it was not in the clean form. We have delete some blank space and special character like hashtag, exclamation mark, parentheses, ampersand, apostrophe etc. table 1 show the after cleaning dataset attribute

Figure 5: Steps of analysis

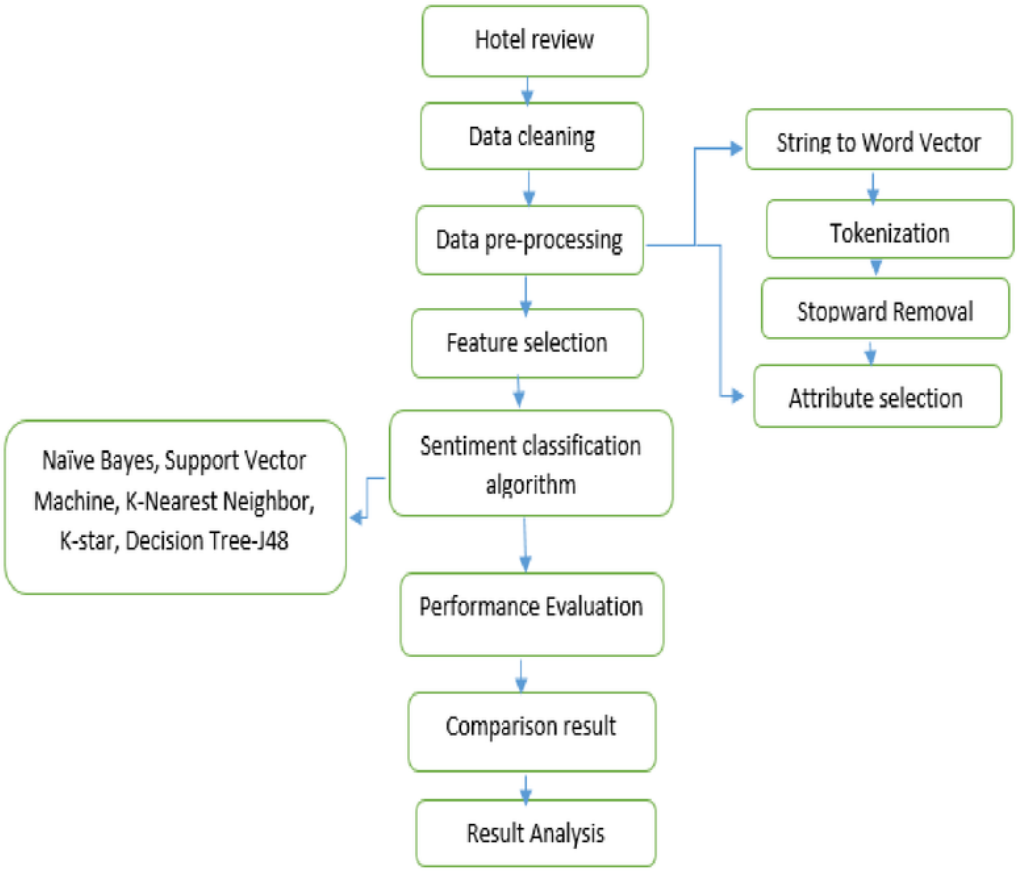


Figure 6: Steps of analysis

Table 1: Datasets before and after processing

Before Pre processing				After Pre processing	
Dataset	Attributes	Class rating	Number of reviews	Attributes	Class rating
Tripadvisor	Review, Rating, Image_count, length, Location, date, verified user, Image shared	1 star, 2 star, 3 star, 4 star, 5 star	483	Rating, Length, Location_mentioned, date of posting review, verified user and profile picture	Real Fake
Goibibo	Review , Rating, property_type, Image_count, length, Location, date, profile_picture, Rating_text	0 star, 1 star, 2 star, 3 star, 4 star	498	Rating, Length, Location_mentioned, date of posting review, verified user and profile picture	Real Fake

Step 3: Data Preprocessing

Data preprocessing is the important step in text mining process, it is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Data preprocessing prepares raw data for further processing. We have broken down this step as below:

StringToWordVector (STWV)

A string vector is defined as an ordered finite set of strings. StringToWordVector filter is the main text analysis tool in Weka and it makes the transformed datasets attribute

value either Positive, Negative or Neutral for all single-words, depending on the word appearing in the document or not. String vectors are structured data representing raw data and should be distinguished from other structured data such as numerical vectors and bags of words. There are three advantages in representing documents into string vectors. The first advantage is to avoid the aforementioned two main problems completely: the huge dimensionality and the sparse distribution. The second advantage is that string vectors are characterized as more transparent representations of documents than numerical vectors; it is easier to guess the content of documents only from their representations. The third advantage is that there is the potential possibility of tracing more easily, as to why documents are classified so. However, in order to use string vectors more freely, it is necessary to set foundations that are more mathematical.

In the weka tool the string to word vector is the main text analysis filter which transform the dataset attribute value according to the word appear in it either positive or negative. It is the filtration process which have two sub-process: stopward removal and tokenization.

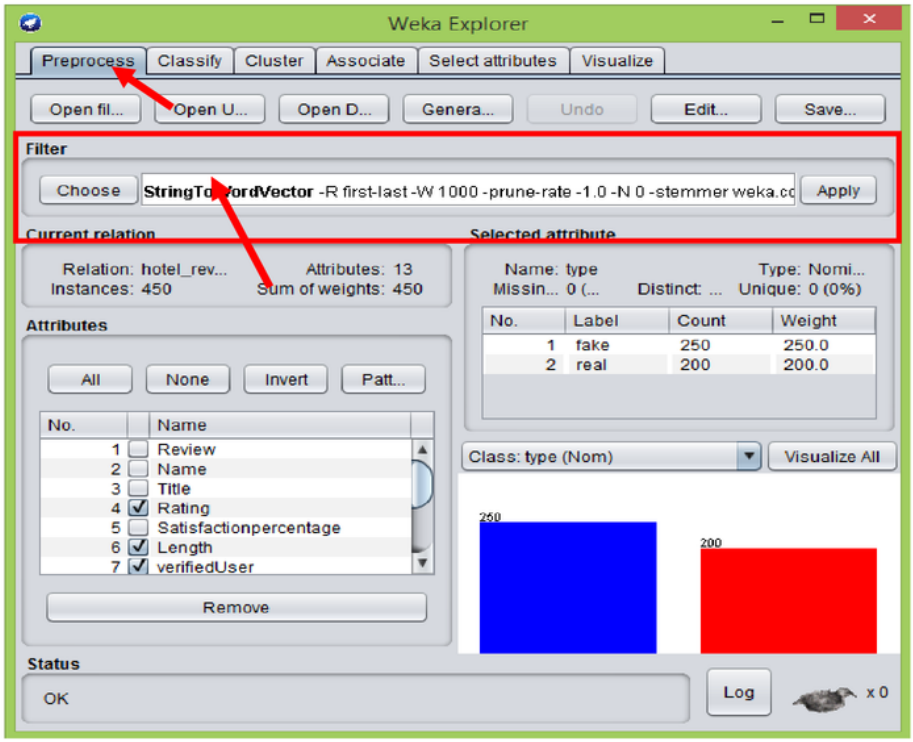


Figure 7: Preprocess in WEKA

Stopword removal and Tokenization

Stopword basically filter out the common words (eg. “a”, “an”, “the”, “is” etc.) which do not add any significant information for datasets and didn’t add any value to the sentence meaning. Any group of words can be chosen as the stop words for a given purpose. For some search engines, these are some of the most common, short function words, such as the, is, at, which, and on. In this case, stop words can cause problems when searching for phrases that include them, particularly in names such as "The Who", "The", or "Take that". Other search engines remove some of the most common words including lexical words, such as "want" from a query in order to improve performance.

These words were ignore so we have specific words for classifier. To make the provided document classifiable using Machine Learning we need to do Feature extraction that is converting the normal text to a set of features that can then be used by the ML Algorithm to discriminate between corn and non-corn.

Attribute selection

This is similar to feature selection, where we select the relevant feature subset which can be used in model construction. Attribute selection increase the chances of accuracy for better clarity.

Step 4: Feature Selection

Whether you select and gather sample data yourself or whether it is provided to you by domain experts, the selection of attributes is critically important. It is important because it can mean the difference between successfully and meaningfully modeling the problem and not. Feature Selection or attribute selection is a process by which you automatically search for the best subset of attributes in your dataset. The notion of “best” is relative to the problem you are trying to solve, but typically means highest accuracy.

A useful way to think about the problem of selecting attributes is a state-space search. The search space is discrete and consists of all possible combinations of attributes you could choose from the dataset. The objective is to navigate through the search space and locate the best or a good enough combination that improves performance over selecting all attributes.

Three key benefits of performing feature selection on your data are:

- **Reduces Overfitting**: Less redundant data means less opportunity to make decisions based on noise.
- **Improves Accuracy**: Less misleading data means modeling accuracy improves.
- **Reduces Training Time**: Less data means that algorithms train faster.

The Attribute Evaluator is the method by which a subset of attributes are assessed. For example, they may be assessed by building a model and evaluating the accuracy of the model.

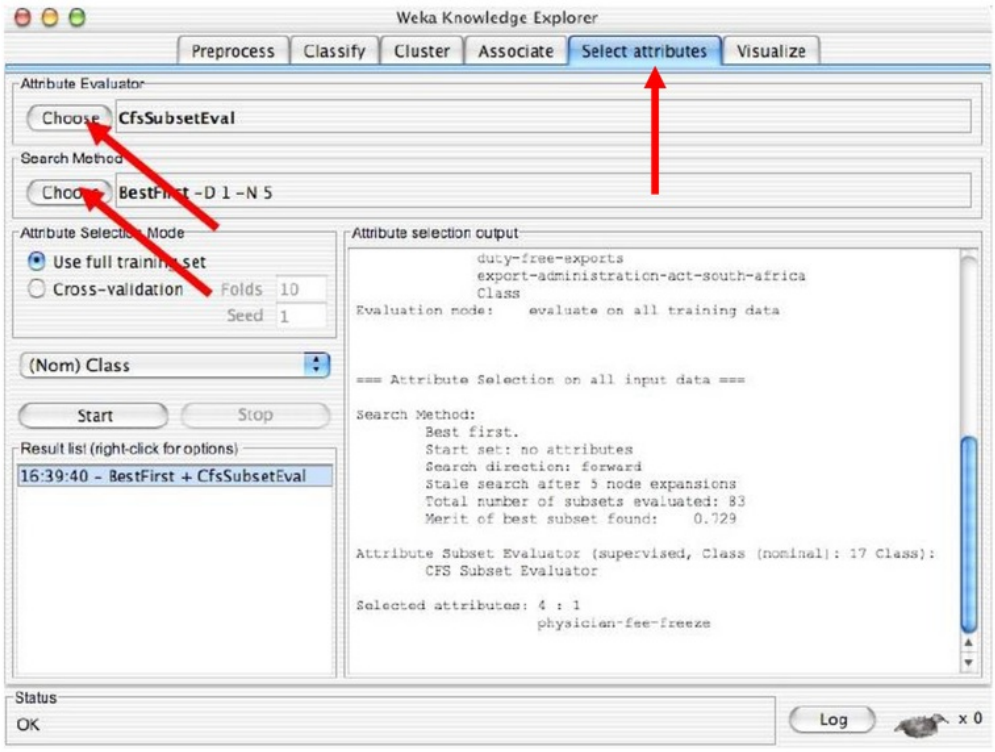


Figure 8: Attribute Selection in WEKA

Some examples of attribute evaluation methods are:

- CfsSubsetEval: Values subsets that correlate highly with the class value and low correlation with each other.
- ClassifierSubsetEval: Assesses subsets using a predictive algorithm and another dataset that you specify.
- WrapperSubsetEval: Assess subsets using a classifier that you specify and n-fold cross validation.

The Search Method is the structured way in which the search space of possible attribute subsets is navigated based on the subset evaluation. Baseline methods include Random Search and Exhaustive Search, although graph search algorithms are popular such as Best First Search.

Some examples of attribute evaluation methods are:

- Exhaustive: Tests all combinations of attributes.
- BestFirst: Uses a best-first search strategy to navigate attribute subsets.
- GreedyStepWise: Uses a forward (additive) or backward (subtractive) step-wise strategy to navigate attribute subsets

Feature selection is a significant role in the classification of better accuracy and to find the relevant attributes. Our research used one feature selection method **(CfsSubset +BestFirst)** which is largely used in sentiment analysis classification along with stop word removal.

Step 5: Sentiment Classification Algorithm

We used SA algorithm for classify the document either fake or real. In our research study we used five popular supervised learning algorithm: Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, K-star, Decision Tree-J48 classifiers.

Classifier Type	Classifiers
Bayes	BayesNet, Complement Naïve Bayes, DMNBtext, Naïve Bayes, Naïve Bayes Multinomial, Naïve Bayes Multinomial Updatable, Naïve Bayes Simple, Naïve Bayes Updateable
Functions	LibLINEAR, LibSVM, Logistic, Multilayer Perceptron, RBF Network, Simple Logistic, SMO
Lazy	IB1, Ibk, Kstar, LWL
Meta	AdaBoostM1, Attribute Selected Classifier, Bagging, Classification via clustering, Classification via Regression, Cost Sensitive Classifier, CVParameter Selection, Dagging, Decorate, END, Filtered Classifier, Grading, Grid Search, LogitBoost, MetaCost, MultiBoost AB, MultiClass Classifier, Multi Scheme, Ordinal Class Classifier, Raced Incremental Logit Boost, Random Committee, Random Subspace
Mi	Citation KNN, MISMO, MIWrapper, SimpleMI
Rules	Conjuntive Rule, Decision Table, DTNB, Jrip, Nnge, OneR, PART, Ridor, ZeroR
Trees	BFTree, Decision Stump, FT, J48, J48graft, LAD Tree, LMT, NB Tree, Random Forest, Random Tree, REP Tree, Simple Cart, User Classifier

Figure 9: Summary of Classifiers supported in WEKA

Naïve Bayes:

One of the easiest ways of selecting the most probable hypothesis given the data that we have that we can use as our prior knowledge about the problem. Bayes' Theorem

provides a way that we can calculate the probability of a hypothesis given our prior knowledge.

Bayes' Theorem is stated as:

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

Where

- **P(A|B)** is the probability of hypothesis A given the data B. This is called the posterior probability.
- **P(B|A)** is the probability of data B given that the hypothesis A was true.
- **P(A)** is the probability of hypothesis A being true (regardless of the data). This is called the prior probability of A.
- **P(B)** is the probability of the data (regardless of the hypothesis).
  You can see that we are interested in calculating the posterior probability of P(A|B) from the prior probability p(A) with P(B) and P(B|A).

The naïve Bayes algorithm is based on the bayes rule of conditional probability with independent assumptions between the features, this is an important machine learning algorithm.

A list of probabilities are stored to file for a learned naive Bayes model. This includes:

- **Class Probabilities**: The probabilities of each class in the training dataset.
- **Conditional Probabilities**: The conditional probabilities of each input value given each class value.

This classifier assumes the features (in this case we had words as input) are independent. Hence the word naive. Even with this it is powerful algorithm used for

- Real time Prediction
- Text classification/ Spam Filtering
- Recommendation System

Decision Tree-J48:

It is a predictive machine learning technique which help you to decide the target value of the sample with the help of several attributes. In the testing option, we have choose percentage split as the preferred method. A decision tree is a decision support system that uses a tree-like graph decisions and their possible after-effect, including chance event results, resource costs, and utility. A Decision Tree, or a classification tree, is

used to learn a classification function which concludes the value of a dependent attribute (variable) given the values of the independent (input).attributes (variables). This verifies a problem known as supervised classification because the dependent attribute and the counting of classes (values) are given [14].

Decision trees are the most powerful approaches in knowledge discovery and data mining. It includes the technology of research large and complex bulk of data in order to discover useful patterns. This idea is very important because it enables modelling and knowledge extraction from the bulk of data available. All theoreticians and specialist are continually searching for techniques to make the process more efficient, cost-effective and accurate. Decision trees are highly effective tools in many areas such as data and text mining, information extraction, machine learning, and pattern recognition.

Decision tree offers many benefits to data mining, some are as follows:-

- It is easy to understand by the end user.
- It can handle a variety of input data: Nominal, Numeric and Textual
- Able to process erroneous datasets or missing values
- High performance with small number of efforts
- This can be implemented data mining packages over a variety of platforms

A tree includes: - A root node, leaf nodes that represent any classes, internal nodes that represent test conditions (applied on attributes) figure 1

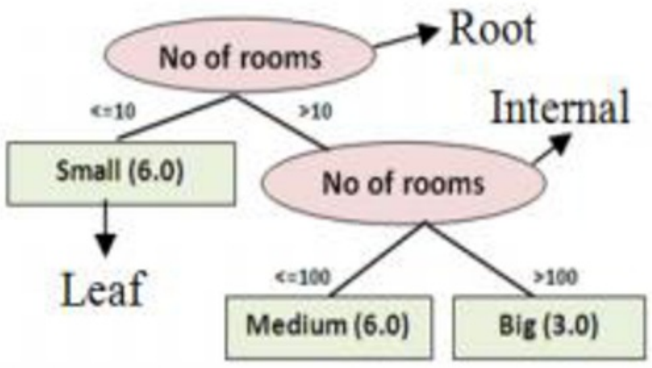


Figure 10: Decision Tree

Support Vector Machine (SVM): SVM is one of the supervised learning model which examine the data and then identify the patterns which again can be used for regression and classification analysis [8].

K star: K* is an instance-based classifier. The class of the test instance is established in the class of those training instances similar to it, as decided by some similarity function [10].

K-Nearest Neighbor (K-NN): K-NN is a lazy learning algorithm and one of the non-parametric approach for classifying the object based on the training data. The algorithm is costly in terms of memory and computation cost. It chooses the nearest neighbor on the basis of majority voting. The output is based on the different key factors like similarity measure, distance measure, k parameter [9]. We choose 'k=3' for our study. This number decide how many neighbor should influence our classification.

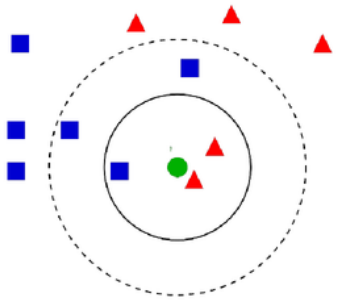


Figure 11: NN selection in KNN

Step 6: Performance Evaluation

The output is predicted via generated confusion matrix which classifies the reviews with the following attributes which are true positive which the actual positive reviews are in the testing data, which are correctly classified positive by the model as well, false positive: fake positive reviews which are incorrectly classified by the model, true negative the actual negative reviews which are correctly classified by the model and false negative which are incorrectly classified by the model. The confusion matrix is very important tool used in our study. The performance was measured on the basis accuracy, precision, F-measure and recall.

Accuracy is a ratio of correctly predicted instances to the total number of instances. Accuracy is useful measure only when we have balanced datasets where values of false positive and false negative instances are almost same.

Precision is the ratio of correctly predicted positive instances to the total predicted positive instances.

Recall is the ratio of correctly predicted positive instances to total number of instances in actual positive class.

F Measure is weighted average of Precision and Recall and take false positives and false negatives into account.

Table 2: Confusion Matrix

	Real	**Fake**
Real	True Negative Reviews (TN)	False Positive Reviews (FP)
Fake	False Negative Reviews (FN)	True Positive Reviews (TP)

Formulas:
True Positive Rate = TP/ TP + FN
True Negative Rate = TN/ TN + FP
False Positive Rate = FP/ FP + TN
False Negative Rate = FN/ FN + TP
Accuracy = TP+TN / TP+TN+FN+FP
Precision = TP/ TP + FP
Recall = TP/ TP + FN
F measure = 2* ((Precision * Recall) / (Precision + Recall))

Step 7: Results

The following table and graph present the performance evaluation in terms of accuracy, precision, recall, F-measure and time taken to build the model by different classification algorithms.

Table 3: Results of Accuracy, Precision, Recall, F-Measure and Time Taken to build model for tripadvisor

Classification Algorithm	Accuracy	Precision	Recall	F-Measure	Time Taken (seconds)
NB	63	64.5	64.8	64.3	0.02

J48	64.1	50.9	55.6	49.7	0.04
SVM	**68.99**	68.2	68	68.1	2.86
KNN	57	55.5	56	55.9	0
K*	56.05	55.1	55.6	55.3	0

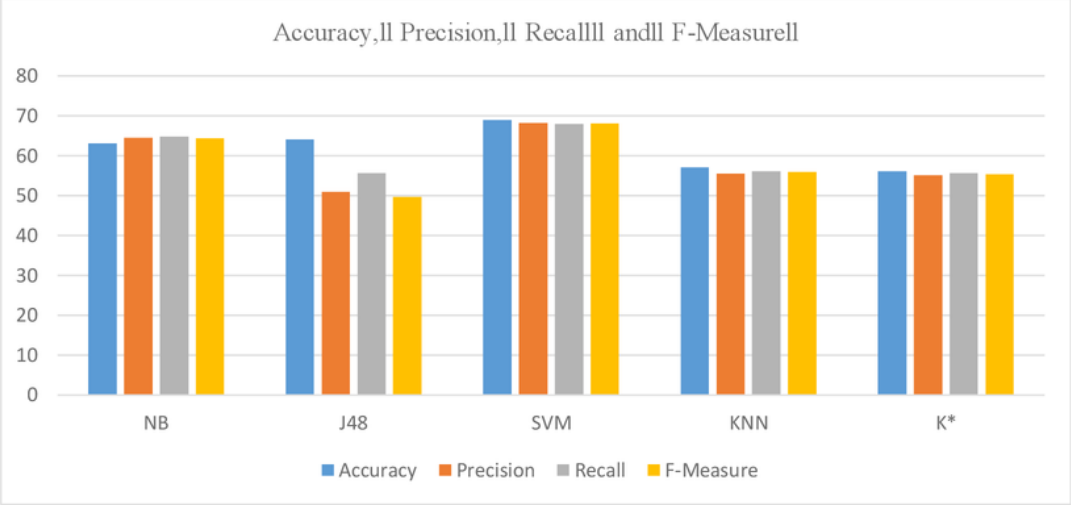


Figure 12: Comparative analysis for tripadvisor

From the above experiment we observed that accuracy of SVM classification algorithm is highest. We observed that K-star and KNN took the least amount of time in building the while Naïve Bayes took 0.02 second, DTJ48 took 0.04 seconds and SVM takes the maximum amount of time 2.86 seconds to build a model.

The following table and graph present the performance evaluation of different classification algorithm used in terms of accuracy, precision, recall, and F-measure calculated for goibibo.

Table 4: Results of Accuracy, Precision, Recall, F-Measure and Time Taken to build model for goibibo

Classification Algorithm	Accuracy	Precision	Recall	F-Measure	Time Taken (seconds)
NB	63	58.4	63.1	59.8	.02
J48	63.9	56.9	63.4	59	.04

SVM	67	63.4	69	61.1	2.40
KNN	61.63	58.1	65.2	59.3	.01
K*	64.3	57.2	64.4	58.7	0

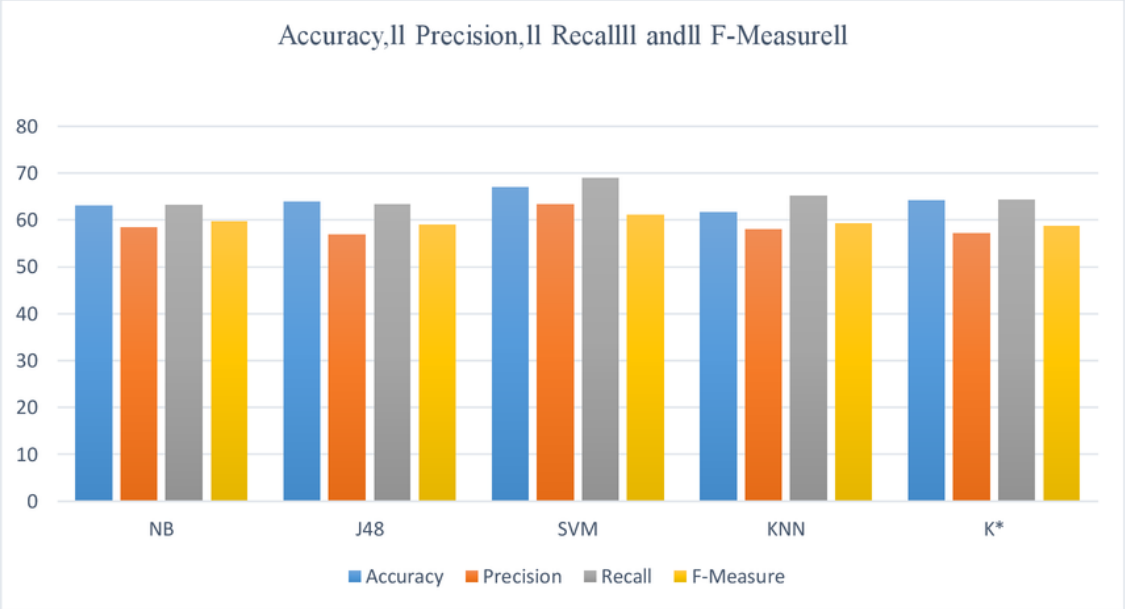


Figure 13: Comparative analysis for goibibo

From the above observations we can see that accuracy of SVM classification algorithm is highest for goibibo as well. We observed that K-star and KNN took the least amount of time in building the while Naive Bayes took 0.02 second, DTJ48 took 0.04 seconds and SVM takes the maximum amount of time 2.40 seconds to build a model.

CONCLUSIONS

In this paper we tried to analyze online hotel reviews from two different datasets (tripadvisor and Goibibo based on supervised learning techniques. The experiment was carried out to identify the best classification algorithm using five performance measures. We found that machine learning algorithms can be used in classifications of the reviews. As five supervised learning algorithm is used to classifying sentiment of our two datasets.

Validation is a techniques in machine learning used to find the error rate of the build model, which can be considered as close to the true error rate. Following are three types of validation techniques are:

Holdout cross validation techniques: in this techniques the data is split into two different datasets labeled as a training and a testing dataset. The ratio of split can be either 60/40 or 70/30 or 80/20. This technique is called the hold-out validation technique. In this case, there is a likelihood that uneven distribution of different classes of data is found in training and test dataset. To fix this, the training and test dataset is created with equal distribution of different classes of data. This process is called stratification.

K-Fold Cross-Validation: In this technique, k-1 folds are used for training and the remaining ones are used for testing as shown in the picture given below.

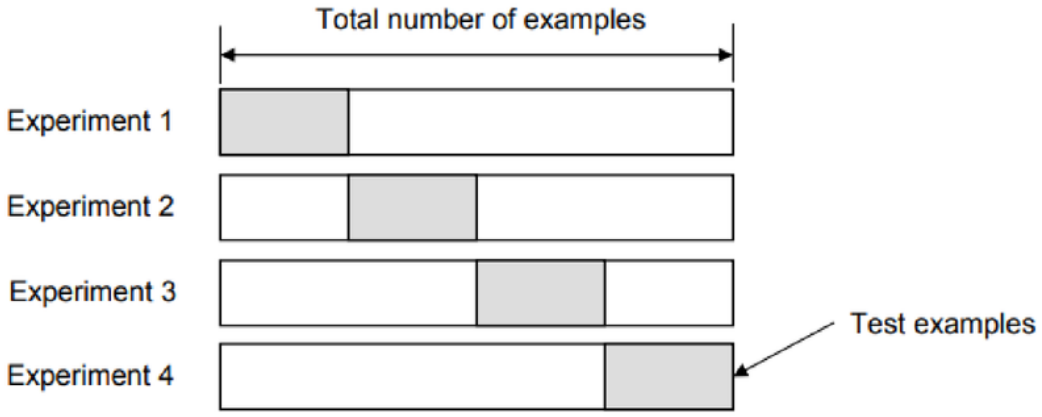


Figure 14: K-Fold cross validation

The advantage is that entire data is used for training and testing. The error rate of the model is average of the error rate of each iteration. This technique can also be called

a form the repeated hold-out method. The error rate could be improved by using stratification technique.

Leave-One-Out Cross-Validation: In this technique, all of the data except one record is used for training and one record is used for testing. This process is repeated for N times if there are N records. The advantage is that entire data is used for training and testing. The error rate of the model is average of the error rate of each iteration.

The following diagram represents the LOOCV validation technique.

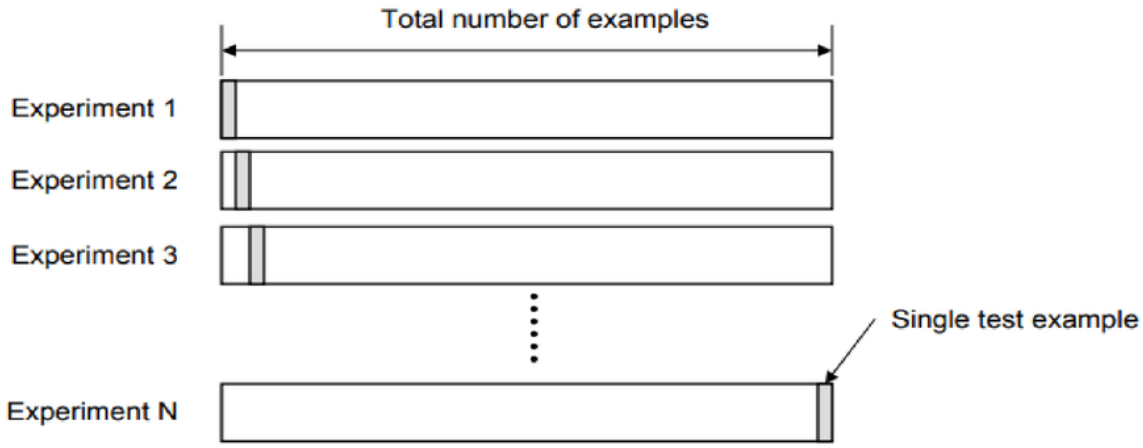


Figure 15: LOOCV validation technique

Random Subsampling: In this technique, multiple sets of data are randomly chosen from the dataset and combined to form a test dataset. The remaining data forms the training dataset. The following diagram represents the random subsampling validation technique. The error rate of the model is the average of the error rate of each iteration.

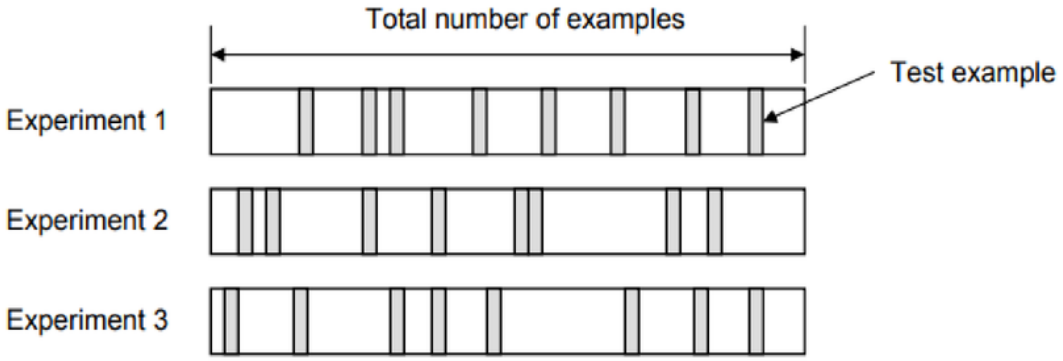


Figure 16: Random Subsampling Validation

In our study we have used tenfold cross validation method. In terms of accuracy, SVM is the best algorithm for all tests since it correctly classified 69% of the reviews in tripadvisor and 67% of the reviews in goibibo.

FUTURE WORKS

For future work, we would like to extend our study to other online portals like zomato, swiggy dataset and used different feature selection methods. As the same sentiment were used in these online services portals. Furthermost, we may apply these sentiment classification algorithm on various tools like python just to evaluate the performance of our work on that tool to check the variation on the results.

major project

ORIGINALITY REPORT

1%	1%	1%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.huang-verlag.de Internet Source	1%
2	rua.ua.es Internet Source	<1%
3	Oz Shy, Rune Stenbacka. "Investment in customer recognition and information exchange", Information Economics and Policy, 2013 Publication	<1%
4	www.dspace.espol.edu.ec Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		