CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

# Identification of Extremist Reviewer Groups using Product Reviews

by

Haleema Sadia

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Computing
Department of Computer Science

2021

*My dissertation work is devoted to My Family and My Teachers. I have a special feeling of gratitude for My beloved parents, whose prayers and support enabled me to succeed in all spheres of life. Special thanks to my supervisor, whose uncountable confidence helped me to reach this milestone.*

# CERTIFICATE OF APPROVAL

## Identification of Extremist Reviewer Groups using Product Reviews

by

Haleema Sadia

(MCS191022)

## THESIS EXAMINING COMMITTEE

| S. No. | Examiner | Name | Organization |
|---|---|---|---|
| (a) | External Examiner | Dr. Waseem Shahzad | Fast, Islamabad |
| (b) | Internal Examiner | Dr. Nayyer Masood | CUST, Islamabad |
| (c) | Supervisor | Dr. M. Shahid Iqbal Malik | CUST, Islamabad |

Dr. M. Shahid Iqbal Malik
Thesis Supervisor
December, 2021

Dr. Nayyer Masood
Head
Dept. of Computer Science
December, 2021

Dr. Muhammad Abdul Qadir
Dean
Faculty of Computing
December, 2021

# *Author's Declaration*

I, **Haleema Sadia** hereby state that my MS thesis titled "**Identification of Extremist Reviewer Groups using Product Reviews**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

**(Haleema Sadia)**

MCS191022

# *Plagiarism Undertaking*

I solemnly declare that research work presented in this thesis titled "**Identification of Extremist Reviewer Groups using Product Reviews**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

**(Haleema Sadia)**
MCS191022

# *Acknowledgement*

Thanks to Allah Almighty for blessing me with wisdom and strength to complete the dissertation. Being an MS graduate at Capital University of Science and Technology has been a magnificent and challenging experience. During the degree, I have found clear guidelines in shaping my academic career. Here is a humble tribute to all those people. I would like to express my sincerest appreciation to my enthusiastic supervisor, **Dr. M. Shahid Iqbal Malik**, for his supervision, assistance, and immense knowledge. I am sincerely thankful to him for his constant support, motivation, and patience. His invaluable help of constructive comments and suggestions throughout the thesis work has contributed to the success of this research. It has been a fantastic experience, and I thank him wholeheartedly, not only for his tremendous support. I want to thank my family, who had motivated me continuously to achieve this milestone. A word of applause for my friends and classmates who had assisted me in sharing knowledge and other resources required to conduct research. Thank you all.

**(Haleema Sadia)**

# *Abstract*

In today's world populated by online platforms, review platforms and online sites play a key role in buyer's decisions for their next buy/purchase. There are various online review platforms in which reviewers write their opinion about different products. eBay, Yelp, TripAdvisor, and Amazon are some popular review platforms. Reviewers write reviews on various products in order to exploit consumers' widespread opinions in their favor. Individual reviewers write reviews to assist other consumers while several individual reviewers create an intricate network, and they end up being a significant influence on the general impression of the product since they share a higher number of people reviewing. By exploiting these reviews facility, some reviewers write fake reviews in the form of groups. They share common characteristics in their nature of reviews and try to influence a product. They target different products on the brand level and products categories. Studies on product reviews, fake reviews identification, and spam reviewer groups identification have been explored in prior studies. A little has been studied to identify groups targeting a brand as a whole rather than just a product.

This research uses a publicly available amazon extremist reviewer's dataset to identify extremist reviewer groups. For features selection, both filter and wrapper method techniques were used. Potential features are selected by the wrapper forward method using the 3-layer perceptron as the learning model. It is a binary classification problem. ML models such as Support Vector Machine, Logistic Regression, Random Forest, Decision Tree, Gaussian Naïve Bayes, Stochastic Gradient Descent, K-Nearest Neighbor, 3-Layer Perceptron, 4-Layer Perceptron, Extreme Gradient Boosting used to classify the extremist and moderate reviewer groups. The 3-layer perceptron well performed with the precision of 99.16%, recall 99.16%, F1-measure 99.16%, and AUC 99.12%. This study aims to design an effective framework for extremist reviewer group's identification using online product reviews. To provide buyer awareness in online marketplaces, they can differentiate between moderate reviewers and extremist reviewer groups without any extra struggle.

# Contents

**Author's Declaration** **iv**

**Plagiarism Undertaking** **v**

**Acknowledgement** **vi**

**Abstract** **vii**

**List of Figures** **xi**

**List of Tables** **xii**

**Abbreviations** **xiv**

**1 Introduction** **1**
1.1 Social Media Platforms . . . . . . . . . . . . . . . . . . . . . . . . . 2
1.1.1 e-Bay . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
1.1.2 Yelp . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
1.1.3 Tripadvisor . . . . . . . . . . . . . . . . . . . . . . . . . . 9
1.1.4 Amazon . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
1.2 Problem Statement . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
1.3 Scope . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
1.4 Research Questions . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
1.5 Research Objectives . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

**2 Literature Review** **17**
2.1 Studies on Product Reviews . . . . . . . . . . . . . . . . . . . . . . 17
2.2 Fake Reviews Identification . . . . . . . . . . . . . . . . . . . . . . 21
2.3 Reviewer Groups and Extremist Groups Identification . . . . . . . . 25
2.4 Research Gap . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

**3 Proposed Methodology** **32**
3.1 Dataset Description . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
3.2 Features . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
3.2.1 Average Rating . . . . . . . . . . . . . . . . . . . . . . . . 34

3.2.2 Average Upvotes . . . . . 34
3.2.3 Average Sentiment . . . . . 35
3.2.4 Group Time Window . . . . . 35
3.2.5 Review Count . . . . . 36
3.2.6 Rating Deviation . . . . . 36
3.2.7 Early Time Window . . . . . 37
3.2.8 Verified Purchase . . . . . 37
3.3 Features Selection . . . . . 37
3.3.1 Filter Methods . . . . . 39
3.3.1.1 Pearson Correlation Method . . . . . 39
3.3.1.2 Information Gain Method . . . . . 39
3.3.1.3 Gain Ratio Method . . . . . 40
3.3.2 Wrapper Method . . . . . 40
3.3.2.1 Forward Selection . . . . . 41
3.4 Machine Learning Models . . . . . 42
3.4.1 Support Vector Machine . . . . . 43
3.4.2 Logistic Regression . . . . . 43
3.4.3 Random Forest . . . . . 44
3.4.4 Decision Tree . . . . . 44
3.4.5 Gaussian Naive Bayes . . . . . 45
3.4.6 Stochastic Gradient Descent . . . . . 45
3.4.7 K-Nearest Neighbor . . . . . 46
3.4.8 3-Layer Perceptron . . . . . 46
3.4.9 4-Layer Perceptron . . . . . 47
3.4.10 Extreme Gradient Boosting . . . . . 47
3.5 Evaluation Metrics . . . . . 48
3.5.1 Precision . . . . . 48
3.5.2 Recall . . . . . 48
3.5.3 F1-Measure . . . . . 48
3.5.4 Area Under the Curve . . . . . 49
3.6 Tools and Languages . . . . . 49

**4 Results and Analysis** **50**
4.1 Experimental Setup . . . . . 50
4.2 Experiment 1 Classifiers Performances for Extremist Reviewer Groups Detection . . . . . 51
4.2.1 Classifiers Performances using Micro-Average . . . . . 51
4.2.2 Classifiers Performances using Macro-Average . . . . . 53
4.3 Experiment 2 Impact of Feature Selection on Classification . . . . . 54
4.3.1 Feature Selection using Pearson Correlation Method . . . . . 54
4.3.1.1 Top-Five Features . . . . . 54
4.3.1.2 Top-Four Features . . . . . 56
4.3.1.3 Top-Six Features . . . . . 58
4.3.2 Feature Selection using Information Gain Method . . . . . 59

4.3.2.1 Top-Five Features . . . . . . . . . . . . . . . . . . 59
4.3.2.2 Top-Four Features . . . . . . . . . . . . . . . . . . 61
4.3.2.3 Top-Six Features . . . . . . . . . . . . . . . . . . . 63
4.3.3 Feature Selection using Gain Ratio Method . . . . . . . . . 64
4.3.3.1 Top-Five Feature . . . . . . . . . . . . . . . . . . . 65
4.3.3.2 Top-Four Feature . . . . . . . . . . . . . . . . . . . 66
4.3.3.3 Top-Six Feature . . . . . . . . . . . . . . . . . . . . 68
4.3.4 Feature Selection using Wrapper method . . . . . . . . . . . 70
4.4 Experiment 3 Performance of Individual Features on Classification . 72

**5 Conclusion and Future Work** **76**
5.1 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . 76
5.2 Future Work . . . . . . . . . . . . . . . . . . . . . . . . . . 77

**Bibliography** **78**

# List of Figures


1.1 Social media buying behavior using online value proposition [6] . . . 4
1.2 Net revenue in U.S dollars from eBay [15]. . . . . . . . . . . . . . . . 6
1.3 Number of Yelp reviews made each day [27]. . . . . . . . . . . . . . 10
1.4 Tripadvisor ecosystem [37]. . . . . . . . . . . . . . . . . . . . . . . . 12
1.5 An example of an Amazon product review [46]. . . . . . . . . . . . . 14

3.1 Block diagram of proposed methodology. . . . . . . . . . . . . . . . . 33
3.2 Methodology adopted by the wrapper method for features selection [114] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 41

4.1 Individual feature performance using precision. . . . . . . . . . . . . 72
4.2 Individual feature performance using recall. . . . . . . . . . . . . . . 73
4.3 Individual feature performance using F1-score. . . . . . . . . . . . . 74
4.4 Individual feature performance using AUC. . . . . . . . . . . . . . . 75

# List of Tables


2.1 Review on latest approaches. . . . . . . . . . . . . . . . . . . . . . . . 29

3.1 Extremist reviewers dataset . . . . . . . . . . . . . . . . . . . . . . 33

4.1 Hardware requirements. . . . . . . . . . . . . . . . . . . . . . . . . 51
4.2 Hardware requirements. . . . . . . . . . . . . . . . . . . . . . . . . 51
4.3 Classifiers performances using micro-average . . . . . . . . . . . . . 52
4.4 Classifiers performances using macro-average . . . . . . . . . . . . . 53
4.5 Top five feature selection using pearson correlation method (micro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
4.6 Top five feature selection using pearson correlation method (macro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56
4.7 Top four feature selection using pearson correlation method (micro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
4.8 Top four feature selection using pearson correlation method (macro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 57
4.9 Top six feature selection using pearson correlation method (micro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 58
4.10 Top six feature selection using pearson correlation method (macro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 59
4.11 Top five feature selection using information gain method (micro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60
4.12 Top five feature selection using information gain method (macro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
4.13 Top four feature selection using information gain method (micro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62
4.14 Top four feature selection using information gain method (macro-average) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62
4.15 Top six feature selection using information gain method (micro-average) 64
4.16 Top six feature selection using information gain method (macro-average)64
4.17 Top-five feature selection using gain ratio method (micro-average). . .65

4.18 Top-five feature selection using gain ratio method (macro-average). . .66
4.19 Top four feature selection using gain ratio method (micro-average). . .67

4.20 Top four feature selection using gain ratio method (macro-average). . .68

4.21 Top six feature selection using gain ratio method (micro-average). . . 69
4.22 Top six feature selection using gain ratio method (macro-average). . 69
4.23 Feature selection using wrapper method (micro-average) . . . . . . 71
4.24 Feature selection using wrapper method (macro-average) . . . . . . 71

# Abbreviations

| | |
|---|---|
| **AUC** | Aluminium gallium nitride |
| **DNN** | Deep Neural Network |
| **DT** | Decision Tree |
| **LCM** | Latent Collusion Model |
| **LR** | Logistic Regression |
| **ML** | Machine Learning |
| **MLP** | Multi-Layer Perceptron |
| **RF** | Random Forest |
| **RFG** | Review Factor Graph |
| **SVM** | Support Vector Machine |

# Chapter 1

# Introduction

In today's online marketplace-dominated world, review platforms and internet sites play a critical part in a buyer's choice to make their next order. Without a doubt, many people will post evaluations that are less than accurate to influence most consumers' decisions to their benefit. Individually or collectively, those people are acting. Individual users may write such evaluations out of irritation or delight, but they do not even have a significant impact on the general judgment of an item, but they do assist other purchasers by sharing their experiences. A more convincing instance is when several individuals build a complicated network, and the sheer volume of reviews significantly impacts the overall opinion of the products. Opinion spam reviews are not the only thing that has this kind of power.
Prior studies demonstrate that 10-15 percent of reviews repeat the first few evaluations, making a false initial review even more powerful [1]. This one is broad viewpoint fraud, and every review site should be informed of it and take necessary steps to detect and prevent it. It is a typical case of group fraud, in which multiple users from a company's network collaborate to focus and affect a specific product. It is a relatively well-known occurrence, but most organizations use specific tactics to hide their cooperation. Furthermore, because such organizations are financially or indeed motivated and numerous of them are typically controlled by the same company, organizations have several hits for viewpoint spam, which frequently have some similar behaviors in the structure of their evaluations.

The main issue is that there is no easy method to determine the distinction flanked by a genuine comment and a fraudulent review, but people are frequently unwilling to distinguish between the two. Users submit comments to give an opinion by discussing any positive or negative experiences with everyone. Consequently, this provides benefits and possibilities for buyers to be manipulated through the use of totally bogus comments. Thoughts can be written to help or hurt a company's or manufacturer's image. To use a comprehensive and extensive analytic approach these features may be used to categorize them better. To avoid comment trolling, an E-commerce giant has implemented a new rule that restricts the number of comments goods may receive in a single day, as mentioned in the article [2].

## 1.1 Social Media Platforms

Individuals utilize social media to connect with friends, relatives, and other members of their organizations. Companies utilize social media to sell and marketplace their brand and keep track of consumer complaints. Marketing items, promoting companies, connecting with clients, and fostering new companies are all online. Social networking encourages buyer reviews and allows people to express their feelings about a business as a means of communicating. Companies can immediately react to excellent and nasty comments, resolve consumer issues, and retain or restore consumer trust. Content aggregation is indeed done over the internet. It is the technique of using digital platforms to acquire information, commodities, or ideas. Organizations utilize outsourcing to solicit suggestions from workers, buyers, and the broader population to improve current goods or build new ones.
In social networks, users use such systems to communicate with others and exchange material, opinions, and views. The consumer is generally at the hub of such communities. Customer biographies aid in the identification of many other users who share similar tastes or problems. Such examples are LinkedIn and Facebook.
In media sharing platforms, platforms are primarily concerned with information. Engagement on YouTube, for instance, revolves upon user-created content. In community-based platforms, that feature of the international community, similar

to a blogging community, focuses on conversation. People provide conversation suggestions that turn into lengthy comments sections. In review panel platforms, the emphasis of such platforms seems to be on an evaluation, which is generally of a good or amenity.

The use of social networking sites and the generation of different social networks has exploded in the last ten years. People may socialize and share data with other people in an online world using this latest tech. Over 3.8 billion people take advantage of social media on a worldwide scale. Social networking applications, including TripAdvisor and yelp, appear each year, following in the footsteps of major platforms like Amazon and eBay. By 2023, the amount of people using social media inside the country Is expected to reach 257 million [3]. Technology's advancement has both beneficial and harmful consequences in a variety of areas, including partnerships.

In 2016, the number of people using social media worldwide reached 2 billion. Social networking services provide up new avenues for connecting people from various areas. 2.14 billion individuals will most likely utilize internet services to purchase products by the conclusion of 2021. In 2023, the worldwide e-commerce market is anticipated the exceed 6.5 billion worth this year., accounting for 22% of merchandise trade. Social media is expected to account for 95 percent of all buys by 2040 [4].

Some consumers use the reviewability to create false assessments in the shape of a group, sharing standard features in their nature of comments and attempting to influence a product. They target different products on the brand level and products categories. Social media and social networking sites can interact with others with various cultural and ethical beliefs [5]. Figure 1.1 shows that how social networks use the web brand value to encourage customer purchase behavior. Furthermore, due to the user's continual exposure to such actions and tactile components, their knowledge and attention grow significantly, and they can examine their wishes and how they may be met through their products.

Today's time of life of internet portals, review platforms, and web websites play an essential role in a customer's decision to buy or purchase something. Reviewers create product evaluations in order to take advantage of the public's broad

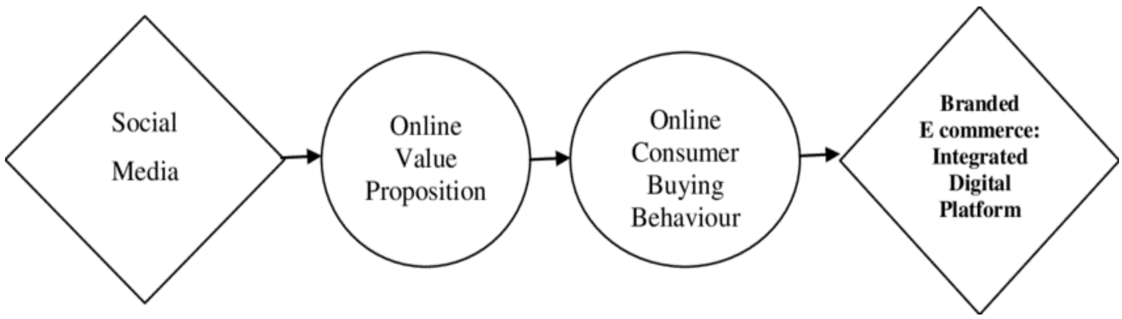


FIGURE 1.1: Social media buying behavior using online value proposition [6]

views to their benefit. Individual reviewers submit evaluations to help other customers, but when multiple users work together, they form a complex network that significantly impacts the overall perception of the product because they have a more significant number of individuals evaluating. Individual reviewers submit evaluations to help other customers, but when multiple users work together, they form a complex network that significantly impacts the overall perception of the product because they have a more significant number of individuals evaluating. Some consumers use the reviewability to create false assessments in the shape of a group, sharing standard features in their nature of comments and attempting to influence a product. They target different products on the brand level and products categories. There are a variety of online review sites where users may express their thoughts on numerous things. eBay, Yelp, TripAdvisor, and amazon are some popular review platforms.

### 1.1.1 e-Bay

With the rapid change of eBay platforms on the Web, sales have made their way inside millions of consumers. It offers the platform for the purchase and implements control through its bidding and advertising operations, but it will not buy the things posted for purchase here on the website,

nor will it acquire absolute ownership of goods [7]. The social media website powerhouse eBay is perhaps the most famous. Online sales have already lately been one of the most widespread and effective kinds of online trade.

Investigate the drivers of buyer and vendor behavior using longitudinal data of eBay currency bids. Initially, a variety of realities are documented. Then create and test a systemic estimator of eBay bids. The champion's plague is measured using estimations after this system, and bidder income is simulated using alternative reservation pricing [8].

eBay, the world's most significant internet online marketplace, reported a net income of $1.39 billion during its last quarterly income statement for 2006, representing a 35 percent year-over-year increase. eBay listing information can be collected in many formats.

They utilize eBay's APIs (Application Programming Interaction), a website spider, or purchasing data collection options. The uploaded sales sites are eBay's primary HTML reviews; it does not provide the purchaser's rating documents or the buying record. People may get the transaction numbers of completed bids in a given segment by eBay's search function or a java applet named Harvex. Surprisingly, this is a very well finding in the bidding concept also that vendor must provide any knowledge about the offered item that might be useful to the buyers [9].

That used a statistical model of historical money transactions on eBay, attempted to uncover such characteristics, and discovered that participation, the usage of reservation pricing, and vendor repute are all predictors of the ultimate cost. Three primary conclusions.

Firstly, competing eBay customers' review evaluations get a demonstrable impact on a purchaser's bidding pricing. Bad publicity has a far more significant impact than good input [10]. Others target specific behaviors, such as last-minute bids. Referring patients through both vendors or eBay buyers suggesting buyers may as well declare their utmost capacity to invest only, soon within the sale, steadily for the past buying, often known as snipe, occurs. eBay, for instance, gives an analogy of a successful initial offer to teach buyers about supply and demand underlying initial bids [11]. eBay already has 27,000 categories, eight of which have total sales of much more than $1 billion annually [12]. Every several months, eBay invites a group of up to a few sellers and customers to give them information about their

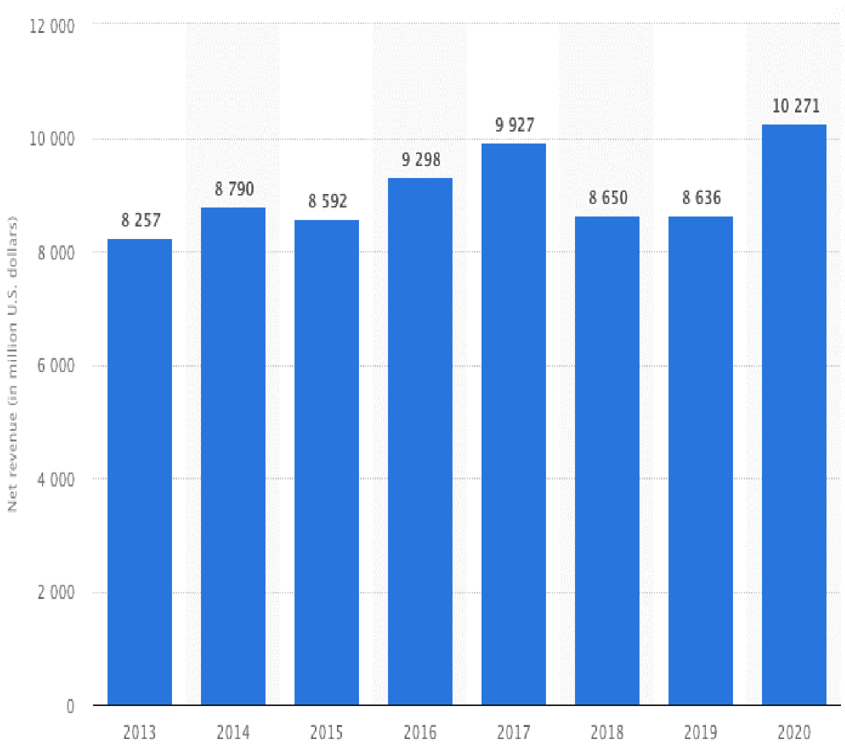


FIGURE 1.2: Net revenue in U.S dollars from eBay [15].

processes, so what else eBay should do. Present a method that combines NLP, information extraction, and topic modeling approaches to mine reviews for dimensional scores and values [13].

Knowledge discovery and empirical methods have been used to analyze previous sales, which has resulted in a significant quantity of studies.

The primary aspect is whether ebay.com platforms have begun to take the place of further conventional souk middlemen like relics, booster packs, and other collections sellers.

The top purchasers began paying their bills on eBay around two years later. Barbie Dolls seem to be favorite plush toys between enthusiasts. Internet sales feature the number of ads plus strong search engines, resulting in fluid marketplaces of different types of products.

Due to reduced management fees, certain middlemen, such as the previous estate sale, were compelled to quit the industry [14]. As Figure 1.2 shows, after falling to a bottom of US $2,129 million during the first period of 2020, the firm soared to its most significant net earnings of US $3,023 in 2021.

### 1.1.2 Yelp

Customer review websites, such as yelp, have become great attractions for expressing people's opinions in current history. To evaluate favorable Yelp.com reviews' influence on food ordering accessibility, use an interrupted time series approach. The market sells up a nineteen percent share faster if given an additional quarter score, having more significant effects if alternative data is limited. Such results imply how chefs are enticed to write false evaluations; however, a complete package of tests confirms that eateries do not alter rankings inside a perplexing, disjointed fashion. While Yelp rankings impact appointment capacity, it is conceivable that such variations are related to consumers who might have strolled in booking online instead. If customers interpret Yelp rankings should mean that more excellent businesses seem to be more consistent with the long delays, they could go out of their way to reserve a spot [16].

Reviews groups try to enhance the authority of conventional brand management with open dialogues regarding as well as through consumers, based on the democratizing ideals of Web 2.0. The majority of commodities activity in Yelp emerges as a favorable leaning towards regionalism, per a text study of customer reviews (n = 1972) or surveys (n = 18). Customers discourse to build a visual for validity surrounding regionalism which follows equal rationale underlying employer logos; critics, if return, take image locally for parts of their whole genuine identity based upon moral virtue to someone's neighborhood. Its consequences of that kind of reasoning are being questioned under the light that commodities; appointment's dedication to particular, individualized levels of self, which has been overshadowed with broader communal as well as communal conflicts [17].

Reviewer groups may create a context for commodities action, politicizing purchase as a community service worker or culture opposition, even though they support traditional materialism ideals. In engaging with these views, Market Engagement not only enters the debate regarding products in politicians but also gives content to investigate broader themes of regional freedom versus participation and inner against outer movement tactics [18]. It can have severe ramifications for emerging nations under the influence of money to mold and sell their history to global customers. Dish has been highlighted to strengthen nationhood, maintain local organic culture inside the current competition between countries, and enhance public gastronomy as a strategy to resist globalization [19].

Users wrote a description only if they think it will assist customers, which might be determined by various factors such as helpfulness, the number of likes from other customers, the size of the comment, rating system, clarity. With a wide variety of merchandise, buyer feedback is becoming more accessible available. They complement additional data in automated marketplaces, like marketing material, professional evaluations, even tailored ways made by the automatic recommender. Testimonials having high rankings are much less helpful for experiencing products than evaluations having medium rankings. Review depth positively affects the review's benefit for both product types, but the product type moderates the effect of review complexity on the review's benefit [20].

Customers are more willing to be subjected to an abundance of data due to the prominence of user evaluations. Past surveys have enthused emphasis towards the determinants of consumer advertisement, explicitly exploring what qualities contribute to either a comment that is seen as principally beneficial by internet users, in opposition to work on the results of customer reviews. The research has empirically evaluated utilizing nonexistent negatively logistic modeling on 16343 online reviews from Yelp.com. In contrast to the substantive qualities of comment, these findings showed that ratings variance and appreciation of both the author had a substantial influence on customer impressions of intervention effectiveness. Such consequences underwrite information in the internet comment usefulness study, and they have ramifications for website quality suppliers who want to anticipate potential worthwhile comments [21].

Attempts were made to identify characteristics utilized by Yelp filtration to detect aberrant behavior. Creating an effective classification model necessitates knowledge of either the characteristics and behaviors of consumer desires. Predicting these biases in some sectors, including style, may be highly challenging allocated to the requirement to represent the current aesthetic of items and their change through time. Style development presents particular issues because of its complex meanings and pro behavior, particularly given the sparseness and enormous size of the source data [22]. Moreover, it was discovered that users who wrote false comments had behavioral and cognitive tendencies of misuse of top frequent phrases. Yelp employs a screening system to screen fraudsters and place these inside a sorted listing to verify the legitimacy of comment threads expressed on the site [23].

Such fast-developing and changing state indicators are provided through this process, indicating the need to expand existing concepts to incorporate increasingly dynamic networks. Scholars frequently seek the interpersonal confinement that organizations could also provide to understand why leaders keep from having traditional rank signals swamped by followers, inspired by previous projects on civilization [24].

Figure 1.3 shows that the number of reviews made each day. Clubs in the 19th century provided Haute food to topmost Citizens with royal aspirations [25]. Style development presents particular issues because of its complex meanings and pro behavior, particularly given the sparseness and enormous size of the source data A novel category arose in the U.s, which offered a large selection of alternatives and device reaches to pick from all [26].

### 1.1.3 Tripadvisor

Several sites for internet evaluations in the tourist industry, like Tripadvisor, Rough World, and Google Maps [28], have grown increasingly popular in the study and practice in the latest generations. Tripadvisor seems to be an Amer-ican tourism online firm that provides evaluations for tourists regarding actual encounters inside resorts, eateries, and landmarks, conferring to Wikipedia.Cited as a significant was opened in January 2000 by Stephen Kaufer and Langley Stein-ert, along with many others, as a website that listed material from handbooks, periodicals, and journals. InterActiveCorp bought the website in 2004 then split out its tour operator division, Booking, annually after [29]. Their findings reveal that the essential mental topics are space appearances, community, proximity, buy-ing behavior, total score, and space described, as determined by Python's textual data and topic extractor.

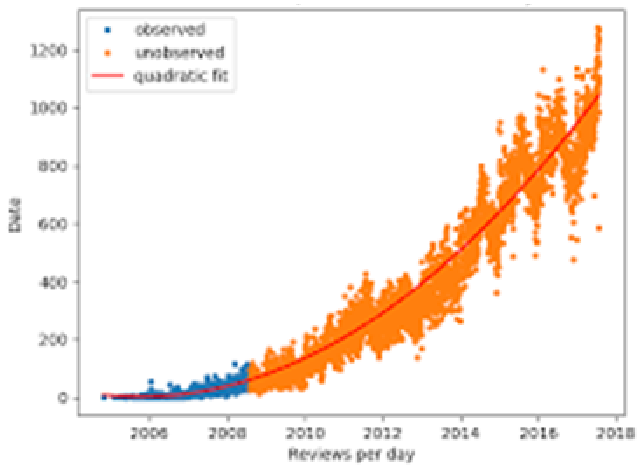


FIGURE 1.3: Number of Yelp reviews made each day [27].

Multiple scorers assessed trustworthiness impressions, and a convenient forecasted them. They were causing an increase in the handful of studies on consumer advertisement and the application of new analysis tools now it has grown, becoming the leading social networking group, with five billion monthly registered users and 56 billion evaluations and comments on over 7 million hotels, eateries, including activities in 59 industries across the globe [30]. Furthermore, the material is confidential into various categories once the consistent architecture of an accordingly leading is discovered. According to the findings, tourist study through webpages, forums, and online networks is on the rise, indicating a cyclical swing wherein the data and bargaining leverage imbalance between the upstream and downstream sides is shifting towards supplier to customer [31].
Indicates that more than a quarter of the text reads concentrate on guesthouses and use quantitative approaches past experimental, the connection among consumer advertisement but instead purchase of stocks and also gratification and proposed control strategy has got increasing courtesy, and opinion mining of consumer advertisement, inspiration to leave feedback, and indeed the involvement of reviews are pretty spread.

Cloud computing, convenience and ease of collecting data, and non-intrusiveness with participants are all great benefits of reviewing studies [32]. On the other hand, sophisticated language analyzers extract significance from appreciated suggestions left by users. Such analyses are usually content, and they frequently include enormous data warehouses, so about Data Management, beyond the advanced analytics of standard approaches [33]. TripAdvisor is the world's biggest traveling group. Various fundamentals will impact consumers' trust in e, but trustworthiness is the most significant [34]. The study also looks into online TripAdvisor is the most widely researched of all the channels created and rated by many other members. It shows that word-of-mouth has a beneficial effect on perceived risk based on source trustworthiness to the degree of knowledge uptake and that users utilize eWOM to mitigate risks involved while deciding. Businesses may tailor existing advertising methods to understand customers' requirements [35]. TripAdvisor has more than 5 million customer accounts using 30 million things each month. Depending on the rating system, they were able to determine how people felt about the reviews. Both quantitative tests of inferential investigation and correlation and deterioration were used to investigate the study topics. The usefulness of assessments varied depending on the emotion of the assessments, regardless of the merchandise category. Furthermore, the connection between source credibility and comment usefulness differed depending on comment mood and making process [36]. As Figure 1.4 shows, with over 260 million users and 22 additional tourism boards throughout a platform, creating a profile on TripAdvisor would put in touch with the whole TripAdvisor ecosystem, which might help grow income and company.

### 1.1.4 Amazon

Consumers are altering their lives due to the fast expansion of Online businesses and industries, including expressing their ideas, insights, remarks, and evaluations on a dedicated website or communities, such as marketplaces Amazon web services. Users choose Walmart when internet purchases during the night since it is one of several e-commerce behemoths, enabling them to peruse thousands of

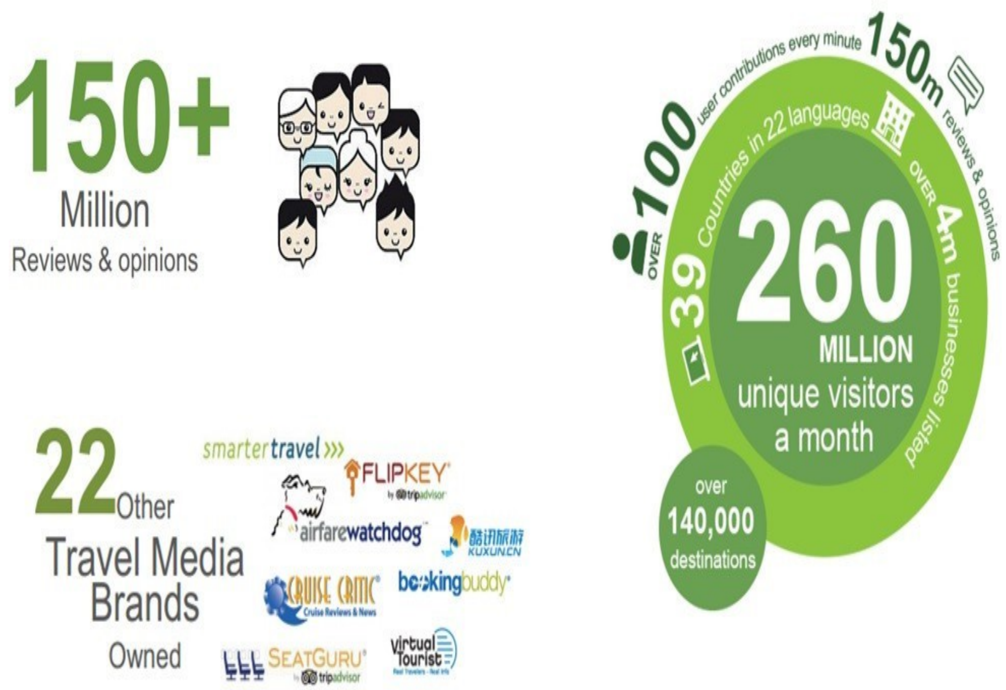


FIGURE 1.4: Tripadvisor ecosystem [37].

customer reviews for such goods customers were attractive throughout. The data used in this study is a set of beauty product reviews from Amazon that is collected from the Snap dataset [38]. Amazon.com developed fast since its inception as an online retailer in 1994 and has served as a model for user-generated feedback and focused on misleading wrong opinion phishing, which are evaluations intended to harm the repute of other businesses [39].

Because of the significant emphasis on consumer internet evaluations, companies are increasingly encouraged to seek and create deceiving thought false evaluations that are purposefully made to appear genuine and fool the user. They discovered that traditional n-gram word classification algorithms outperform expert assessors in detecting nasty misleading comment spamming. Further explore the possible connections among emotion and deceit, using the abovementioned good reviews data, and give preliminary findings suggesting this link's future investigation [40]. Identifying false reviews posted by another individual under various titles and publishing every review below a new identity was an issue. Two techniques for detecting good reports are presented, with the reports showing that they exceed rasterized statistical features employed in previous studies.

The first technique takes the statistical matching score of terms all the way down to the bottom of evaluations. The second technology is founded on recommender systems, so it includes two methods to leverage the commonality of reviewer-founded segmentation: sack and bag-of-opinions. The tests have been carried out on evaluations using three different sets of data: Yelp 57K comments, Trustpilot 9K comments, and the Otto collection of 57K comments [41]. Internet cafe evaluations to identify four eating aspects (taste/food, pleasure, significance, and placement). The Latent Dirichlet Allocation (LDA) was used to identify critical elements, including nutrition, interaction, locality, and price using 294,034 Yelp.com comments or favorable and unfavorable emotion was given to another recovered feature. Nutrition was connected to good emotions, whereas price was correlated with adverse emotions. Evaluating a considerable volume of simple textual information provided by consumers contributes to the existing fiction on management principles of development.

The findings from this education might be employed as a long-term business model for comparison site creators, allowing users to rank and select functional assessments depending on individual interests.

The abstracted characteristics and their associated sentiments also aid restaurants in better understanding how to fulfill the demands of a wide range of consumers while maintaining long-term competitiveness. Knowledge internet comments may help entrepreneurs acquire a greater comprehension of company consumers' thinking and practices that can be utilized to enhance quality and build a better position in the market [42].

Following that, an investigation of a corpus of Amazon reviews found that this two reasoning in comments enhances their usefulness substantially. Researchers discovered that favorable evaluations had a more massive influence than nasty comments, but inflammatory rhetoric slightly influences.

Their discoveries directly impact e-commerce companies, and businesses may use their insights to develop their consumer reaction mechanism and offer more meaningful merchandise recommendations [43].

The Linear Text Pattern has been used to study every phase of a summary's encoding separately, and these have been used to determine not whether reviewing

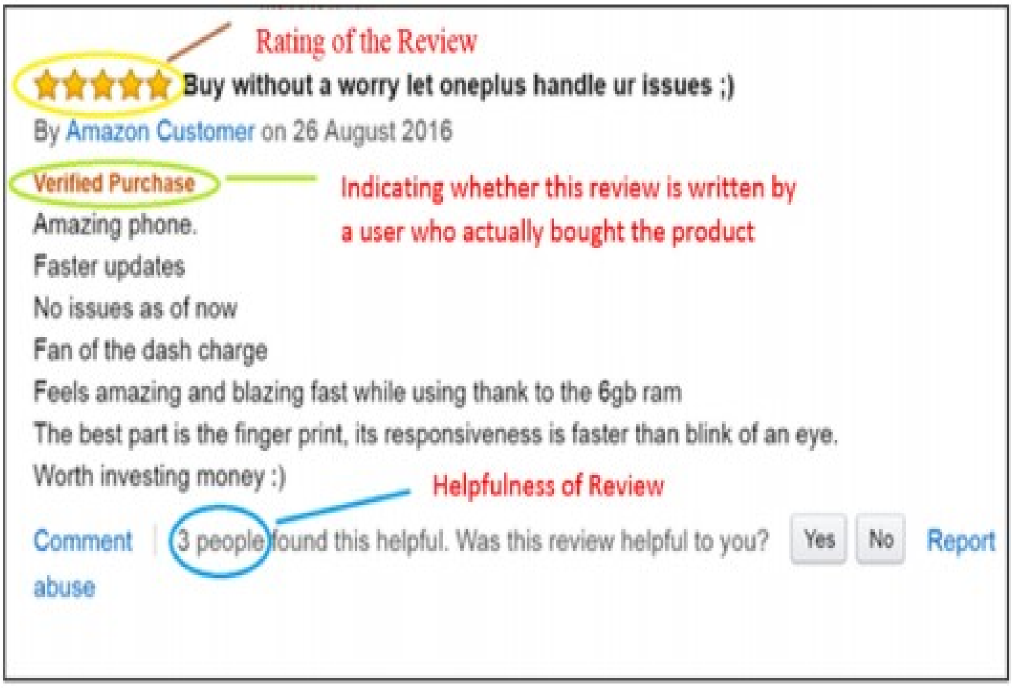


FIGURE 1.5: An example of an Amazon product review [46].

statements were multiple. Internet customer reviews play an essential part in the evaluation of customer satisfaction, and so these evaluations must be reviewed attentively [44]. The usefulness of gaming journalism on the internet Valve shop is evaluated, and the results. When version characteristics are used, modeling becomes less transferrable. One such study approaches the topic after a different perspective, positing that usefulness is a writer's intrinsic characteristic. Each review is usually followed by a fundamental question, such as "Is this comment useful to you?" that generates $2.7 billion in income for Amazon each year [45]. As shown in Figure 1.5, There are three critical components to an Amazon User Review.

I Verified Purchase.

II Rating: Product rating in between 1 to 5 stars.

III Helpfulness: The whole group of individuals who considered the review to be beneficial. These features will aid us in comprehending and analyzing the evaluations in order to gain insights.

## 1.2 Problem Statement

According to our knowledge in the literature, there is only one study that characterizes and detects extremist reviewer groups using product reviews. They used only eight features for the identification of extremist reviewers' groups. In addition, the performance metrics achieved by their proposed method are not promising. There is a need to design a more effective identification system that improves the classification performance of the prior framework.

## 1.3 Scope

The dominant aspect of this study is identifying extremist reviewer group's which make more straightforward for user and brand owners to distinguish between extremist and moderate reviewer groups. The coverage of this study covers the identification of extremist and moderate reviews on a dataset called extremist reviewers dataset. Our technique will be trained only on a selected dataset and produce results according to the extremist reviewer's dataset.

## 1.4 Research Questions

In this research, we address the following research questions.
**RQ1:** Among applied ML models, which model demonstrates the best performance for extremist reviewer groups identification?
**RQ2:** What are the most contributing features after utilization of filter and wrapper-based feature selection methods in identifying extremist reviewer groups?

## 1.5 Research Objectives

Reviews are a famous methodology for getting feedback from customers. Reviewers play an essential role in online marketplaces.

Some people exploit this facility and try to influence the product. In few Prior studies worked on spam, fraudulent, and fake reviewer identification using product reviews. The determination of this proposed study is described as follows.

1. To deliver buyer awareness in online marketplaces, they can differentiate between individual reviewers and extremist reviewers' groups without any extra struggle.
2. Customers can identify extremist reviewer groups using amazon product reviews.

# Chapter 2

# Literature Review

This chapter describes the existing literature into three parts and highlights some significant crucial problems which lead towards the suggested solution. We divided the existing literature into three parts: section 2.1 describes the studies on product reviews, section 2.2 describe fake reviews identifications, section 2.3, describes the reviewer groups and extremist groups identification. We also discussed in 2.4, and after that, in section 2.5, the critical analysis was discussed.

## 2.1 Studies on Product Reviews

Extensive research has been done on mining online reviews and categorizing them based on user sentiment. Regarding digital marketing, emotion analysis is employed. It may determine whether a business or organization is regarded favorably or poorly on the internet based on a user's general approach or attitude as expressed in social networking sites [47]. Demonstrate an algorithm for predicting kindness and generosity relying on deep learning models and examine the emotional influences of review helpfulness. Using the national research council Canada (NRC) emotion Lexicon, they demonstrate an approach for extracting significant distinct optimistic and harmful feeling elements from the written content of merchandise evaluations. Furthermore, the product type, assessor, distinguishability, readability, semantics, and sentimentality-associated criteria are considered [48]. They have provided various strategies that allow e-commerce recommendation algorithms to exploit reviews fully. Textual data mining methods enable emotional understanding of the various attributes, statistical measures, including networked depictions, which bridge the linguistic barrier amongst reviewers with marketing materials. By employing review information to deliver recommendations with explanations, recommendation algorithms overcome the cold-start problem. They also address online marketplace examples and experiences (i.e., Flipkart) [49]. Utilizing efficient provisions of the section and several reviewer's factors, construct the impactful average performance forecasting models that use the gradient descent augmenting instructional strategies. This research aimed to develop a system for extracting new evaluation gratified variables after writing assessment [50].

A reviewer usually contributes two pieces of information: one is an overall rating of the products that she or he has used, and the other is a documentary evaluation that includes her or his comprehensive thoughts proceeding with the goods. The writer promoting elevated items, including image sensors, laptops, even automobiles, generally comments solely on multiple merchandise lines to her/his rare consumption experiences [51]. Using NLP, expand the notion of mixed recommendations by mining significant beneficial aspects using public networks. Used natural language processing to enhance the idea of combined recommendations through repeatedly removing expressive characteristics after community broadcasting. It permits individuals explicitly to impact suggestions by picking aspect numbers and ranking this information from other customers in individual evaluations [52]. It has revealed that the product-of-experts model's greater flexibility allowed it to outperform the Latent Dirichlet Allocation (LDA)-based strategy in the Amazon review dataset, realizing state-of-the-art presentation. Surprisingly, the fully convolutional channel is more robust modeling associated treatments to hinder the model's capacity to perform as a preprocessing step of item descriptions [53]. Numerous issues manipulating the utility of operational evaluations and the moderating effect of product types, such as experience or search items, on the usefulness of online reviews were investigated. This research looks at the many factors that affect the usefulness of consumer advertisement and the indirect role of market segments, such as experiential or searching products, on the effectiveness of digital feedback [54].

Investigate two online information components to determine the factors influencing the apparent expediency of available consumer evaluations. (i) the qualities of reviewers, such as personal identity disclosure, reviewer expertise, and reputation, (ii) measurable (i.e., star scores and duration of evaluations) and qualitative measures are included in the reviews themselves (i.e., review readability and perceived enjoyment) [55]. Scientists tested the key aspects that influenced the usefulness of online reviews and created a helpful prediction model. Build an active prediction system regarding informativeness, five unique language features are presented, and general neural network techniques are used. Two significant Amazon review datasets were subjected to experimental analysis [56].

Investigated the effectiveness of script analysis in forecasting the usefulness of operational consumer evaluations. Toward implement the concept of mental scripts, researchers enlist human experts, who are asked to underline key terms for assessing assessment usefulness. They employed human annotation to operationalize the concept of the cognitive script [57]. Based on user-generated vacation reviews, neural networks were utilized to construct summaries that considered changing opinions over time. Only the most important of such papers are challenging for every individual user to absorb. The multimodal textual summary seems to be the most fantastic powerful method for completing this job. Researchers describe their study strategy using machine learning to produce summaries from consumer trip assessments that accommodate changing perspectives across the period [58].

Demonstrates that product reviews are valuable for judging a product's quality. According to various marketing research, reviews also play a significant role in sustaining a brand's online reputation. It generates a virtuous cycle involving purchases and good Correctional facilities in products of low companies, which benefits individual revenues and its total product attributes [59]. The assessment system is usually included in the evaluation, which significantly affects the overall score of just an item, but it does have a much more significant effect once readers share them. People read reviews only if they believe it will help them, which can be determined by various factors such as valuable upvotes from other customers, the assessment's content, score system, and accessibility [60].Several attempts have been made to determine what makes a review helpful to in-dividuals. They evaluated the influence of significant features of evaluations using 53 scientific research that yielded 191 sample dimensions, recent systematic eval-uation, and meta [61]. It Created a recommender system capable of predicting each user's perceptions of the usefulness of a particular review.

This work cre-ates algorithms that capture factual reading material of consumer advertisements, such as economic signals, aesthetic signals, customer satisfaction, service stan-dards, pricing, and competitive prices, in helping to resolve the ranking issue [62]. Investigated the potential effect of emotions and feelings upon that percentage of ballots in online merchandise assessments indicated that those evaluations were helpful [63]. This research aids in the understanding of emotionality in recommen-dations from friends and has significant ramifications for related businesses and customers. The effect of review text emotions on review usefulness was explored. Researchers discover that further severe evaluations contain higher human emo-tions than lower controversial ones, demonstrating a highly skewed dispersion of emotionality and experimentally supporting a central premise that underlies most existing web advertising work. The findings show that emotions have varied effects on helpfulness when it comes to experience and search products. These results in-dicate that leveraging Brand-related material in business growth, advertising, and system building has significant management consequences [64]. Demonstrate that evaluations with many comments, high sentiment values, and high polarity scores obtain helpful votes. The length and frequency of reviewer activity are statisti-cally significant determinants of helpfulness prediction. This research builds on previous research on intervention effectiveness by considering the essential aspects of comments and the critical indications of the user and the type of goods [65]. Re-searchers looked at how assessment, evaluator, or facilitating conditions influenced evaluation usefulness predictions.

Utilize the constant learning predictive, logistic regression, random forest, neural network, and fully convolutional approaches, and several relevance predictions systems are designed utilizing two different Amazon affiliate evaluation records [66].

## 2.2 Fake Reviews Identification

The majority of spam review detection research falls into two groups. One set of researchers mainly focuses on the reviews' text. On the other hand, other groups of investigators focus on the conduct of reviewers rather than the content of reviews. However, combining both approaches yields the best results. Various efforts have been made to detect and analyze these behaviors, collectively referred to as opinion spam.

They have worked on false-negative opinion spam, which consists of evaluations intended to harm the reputations of other businesses. They discovered that typical n-gram text categorization systems outperform human judges detecting harmful, deceptive opinion spam [40]. Addressed recognizing false reviews written by individuals under many names and posting them under others. Two approaches for detecting similar reviews have been suggested, and the findings show that they exceed the vectoral similarity measures employed in previous research. The suggested approaches leverage the similarity of the reviews topic distributions utilizing two representations: container-of-arguments and container-of-opinion phrases, based on topic modeling and semantic similarity between words to the review level [41]. They have worked on restaurant reviews that were flagged as dubious or fraudulent by yelp's screening system. Yelp's algorithm flags roughly one out of every five reviews as fraudulent, according to their study. These reviews are usually worse than other reviews and are written by individuals with a less-reputable record. Furthermore, their findings show that financial incentives play a significant role in the choice to commit fraud. When organizations face increasing competition and have a terrible or less established reputation, they are more prone to cheat the system [67]. Targeted detecting fraudulent product reviews utilizing a review's text and rating properties. The suggested method assesses an evaluation's trustworthiness, the commentator's reliability, and the merchandise's reliability [68]. Look into the inner workings of yelp's secret filtering algorithm. Put a few current research methodologies to the test using real-life Yelp data to see how they performed. The behavioral aspects were found to be highly successful, while the language features were not. According to their research and findings, yelp's filtering is acceptable, and its filtering algorithm appears to be linked to unusual spamming activities [69]. Claimed to be the first to publish a large-scale study of restaurant reviews. Dianping, a Chinese group buying website for locally sourced food delivery services, consumer items, and retail services, was used to acquire a significant quantity of data. Temporal and geographical characteristics were employed at multiple levels (reviews, users, IPs) [70].

Detected spamming networks utilize reviewer posting frequency over short periods and other people posting frequency for the same items simultaneously. Attempts were made to identify individual spammers and spammer organizations. Several fraudsters also tend to publish comments to almost a similar set of items in a quick dated, a practice known, for example, co-spammers. Researchers present a co-bursting system that refers to the co connections that would be more successful than previous techniques at detecting spamming organizations [71]. Attempts were made to identify individual spammers and spammer organizations. Researchers examined a vast collection of Yelp restaurant ratings and filtered reviews to define how opinion spamming occurs in a business environment. Employing word document, researchers found three primary advertising initiatives: earlier, middle, then latter throughout the numerous eateries [72].

Linguistic characteristics were used to distinguish fraudulent reviews from honest reviews. To classify the reviews, researchers used unsupervised learning using self-organizing maps (SOM) combined with convolutional neural networks. They turn the reviews into pictures by clustering semantically related phrases surrounding a single picture frame and perhaps a SOM square unit. The assessment pictures are given towards CNN to pattern recognition and education [13]. Two distinct feature extraction approaches and six machine learning classification algorithms are investigated and compared.

They compared state-of-the-art techniques, experimental evaluations utilizing current community datasets, and a counterfeit broadcast dataset show extremely positive and improved results [2]. Speagle is a sensor that monitors questionable individuals, evaluations, and other phishing goods, by combining information in a database (consumer tree) using content (behavioral then textual information).

Their primary addition is using a review-network-based categorization job that reflects prior information of the node's class distribution derived through metadata [73]. Their suggested architecture is unsupervised, but it can readily incorporate labels. Without any retraining or modifications to its core algorithms, their technique may rapidly and smoothly incorporate moderately, i.e., a (limited) collection of tags if provided. Using three existing Yelp.com evaluation data sets filtering phishing and suggested non-spam comments, they show that SPEAGLE beats numerous benchmarks and legislature techniques, demonstrating its efficacy and versatility [74].

Identify spam reviews; a Bayesian model was employed. For fraud review detection, they looked at two criteria. One is to review in brief bursts/periods, and another looks for people who evaluate products differently from others. Researchers offer a method for identifying fake comments that integrates these two techniques consistently and systematically, enabling identification, notably whenever one of the indicators is missing. To integrate these two techniques, they developed a Probabilistic Interpretation for Ratings Information framework, a customizable Probabilistic framework based on users' behavior. Depending on their theory, develop a percentage chance doubt indicator, Standardized Estimated Shock Overall [75]. While the classification model has been used to identify fake comments for many generations, data points of massive datasets remain inaccessible, and perhaps most monitoring component learning methods are focused on bogus fake comments instead of actual false evaluations. For false review identification, researchers employed collective positive-unlabeled learning. They offer the first documented effort on false comment identification in China, filtering comments via Dianping's false comment finding, in collaboration with Dianping, the country's main comment serving website. Researchers demonstrated Custom Hierarchical Community Classifier, a trained technique classifying a layered architecture containing evaluations, individuals, and IP addresses. Then they expanded it to include Unlabeled and Collective Positive learning (CPU) [76]. Their suggested approach is entirely unsupervised and linearly scalable. It comprises corresponding binary phases: rating operators and evaluations to identify scams and assemblage to visualize and make sense of the data. They discovered that reviewers, reviews, and goods incorporate structural signals more deeply.

They present FRAUDEAGLE, a quick and efficient system for detecting scammers and false comments in the consumer evaluations dataset. They Utilize fake facts and show that their system is efficient, with FRAUDEAGLE effectively detecting malfeasance in a significant internet assessment process dataset. They built a bipartite network for spam review identification involving goods, users, and reviews [77].

Many incorporations approaches are used for sentiment analysis to assessments of review detection. They utilized sentiment analysis with their lexicon and matched the sentiment analysis result to a customer review. Using text analytics, try to discover spamming and false evaluations and screen out evaluations that contain profanities, obscene, or foul language.

Links the submitted assessment with the computed scores of each comment, creating an opinion word with the assistance of an internal staff vocabulary, using the e-commerce dataset. Buyers may provide goods feedbacks and opinions in the type of rankings on several e-commerce platforms. If the difference in rating between the two results is more significant than a specific threshold, it is classified as spam [78].

Researchers suggested a reproductive LDA-founded issue modeling technique aimed at counterfeit evaluation identification. There seem to be 20 honest evaluations and 20 false comments (800 evaluations entire). Amazon Mechanical Turk is used to generate false ratings.

When it comes to making buying choices, clients are gradually turning to customer information resulting. The constructive broadcast can lead to substantial economic advantages.

They used the same five-fold cross-validation method then the identical data divisions in trials used to stop sentences. Their method is based on an irregular of Latent Dirichlet Allocation (LDA), and it tries to detect minor variations in the topic-word distributions of fraudulent and genuine evaluations [79].

## 2.3 Reviewer Groups and Extremist Groups Identification

Single fraud reviewers have a much more damaging and modest influence than the group of fraudulent reviewers. Rather than individual reviews, the problem of guide classification remained solved through addressing a collective of commentators. Consumers and companies are increasingly using biased digital platforms, including such brand attitude, to make decisions. Others attempt to manipulate the scheme for money or reputation by comment spammers (e.g., creating bogus evaluations) to support or relegate certain items. These bogus comments must be identified in order for evaluations to understand actual customer thoughts and arrogances. A counterfeit user team (a team of users that collaborate to produce fake comments) is much more destructive since they may monopolize the opinion about the particular brand [80].

Equity and dependability measure a participant's and a rating's dependability, accordingly, whereas usefulness measures a service performance. Naturally, a customer is equitable if it gives consistent rankings that are near to customer satisfaction. Demonstrate the interconnectedness amid ratings by proposing six assumptions, and afterward create a continuously iterative description that fulfills those assumptions [81]. They expand the approach to include behavior characteristics and handle the slow response issue. To compute these inherent rankings, including all customers, evaluations, and goods, they created the REV2 methodology. Demonstrate that this method will always be complete and has a linearization cost. REV2 surpasses nine current systems in recognizing honest and dishonest consumers, according to comprehensive trials on five rating samples [82]. On several customer purchasing marketplaces, such as Amazon, comment manipulation is exceptionally general. Fraud comment and spamming discovery systems make use of customer behavior, ranking, and substance. Offer a method for achieving both objectives at the same time. A Comment Genetic Algorithm-based framework is planned to integrate all characteristics and utilize opinion transmission among articles and users by specifying characteristics to characterize each assessment and user [83].Experiments indicate that their approach surpasses all supported techniques na-tively in productivity and correctness by a substantial margin. Further intriguing research that uses metadata to define various elements in e-commerce sites may be found [84].

Where goods, reviews, and users are all categorized simultaneously. Adopt a new strategy to detect comment fraudsters, one that uses the way that indicates comments. Discharges of evaluations can occur as a result of a company developing awareness or spamming assaults. In the same way, fraudsters prefer to cooperate with other marketers, and legitimate consumers begin to arise along with other legitimate authors, approvers, and comments coming in a blast are fre-quently connected. Synchronization is a crucial group activity. Allows researchers to create a system of evaluators who emerge in various spurts. Inside the LBP architecture of internet backbone prediction, they present many characteristics and use pattern-driven communication forwarding. Present a unique approach for dynamically evaluating identified offenders background modeling categorization of negative comments[85].

People may locate and exchange intelligence from blog posts to movies to companies using online streaming sustainability indicators. Throughout principle, such services let customers to login profiles, establish relationships, post and review material, and discover different material by combining user feedback. These websites are growing in popularity; Yelp, for example, has over 35 billion comments. However, the site's attractiveness increases fraudulent behavior, such as numer-ous impersonation assaults and the "purchasing" of recommendation systems [86]. Researchers propose Iolaus, a method that defends in contradiction of this kind of threat by using the core online community of internet pleased evaluation meth-ods. Iolaus employs two distinctive approaches to protect from various individual assaults: (a) weighting rankings to guard against specific threats and (b) comparative rankings to minimize the impact of "purchased" evaluations [87].

This signal was also utilized as a behavioral indication, and an unsupervised model for identi-fying group cooperation was presented, integrating many additional metrics.

the challenges of corrupt business thought manipulation in customer reviews platforms when sections of society collaborate to post false evaluations to manipulate the repute of consumer-packaged goods. The members (or accomplices) may avoid discovery by modifying individual behavior to avoid looking weird, making corrupt business deception considerably more challenging to fight. When blackmailers had modified individual actions together, such indicators might be useless. [88]. Notable studies in the field of user-level fraud detection include Individuals that write fake comments or comment scammers will be detected.

They find many common behaviors among comment fraudsters and analyze them in applied to measure fraudsters. They are particularly interested in modeling the right traits. Maximize overall influence, and fraudsters may indeed attack certain items or mar-ket segments. Secondly, individuals tend to differ from several other consumers in unique consumer reviews. They are applying grading techniques to an Online review collection to estimate the level of fraud for every commenter. Using in-ternet spammers assessment methods explicitly designed to customer assessment studies the selected group of highly doubtful users for more investigation via ex-pert users testers [89]. Provided scoring methods for detecting the spamminess of a reviewer based on their rating behaviors. Customers may learn a lot regard-ing facilities and goods by reading customer appraisals. On the contrary, Internet trolls stand entering the could mislead many users by submitting phony comments. Links among assessors, assessments, and firms already assessed by the criticism provide a fresh idea of a varied assessment tree. It is the first time these complex linkages have been discovered to detect fake appraisals [90].

Design an efficient technique for calculating the trustworthiness of users, the integrity of comments, or the company's dependability.

After reviewing their findings, they complement the present techniques, and detecting increasingly complex and nuanced spammer actions are approved by review petition [91].

It detected fraudulent reviewers by using burstiness in reviews.

Modeling spamicity as a latent component was acces-sible, and numerous reviewer behavioral imprints were utilized [92]. Combining reviewers, reviews, and goods suggests that behavioral cues that are constantly changing may be collected using a network of diverse reviewers [93]. Few studies have attempted to identify fraud reviewer groups at the group level. First, identifying FIM is used by spammer reviewers' clusters and ranking them according to the link between groups, goods, and individual reviewers. Other research has absorbed arranging to enhance the primary scoring method while neglecting the given technique's effectiveness in identifying groupings (FIM was used to identify groupings) [94]. An Expectation-Maximization technique was suggested to calculate every group's collusive rating discovered through frequent itemset mining [88]. FIM, it is said, has a proclivity towards detecting tiny, close-knit groups [95]. They proposed GSBP, a divide-and-conquer method that highlights the reviewers' graphs hierarchical architecture. Spammer groups with participants, target goods, and fake reviews were created immediately [96]. Many graph-based methods have also demonstrated the ability to identify fraud reviewers and scam reviews [39] simultaneously [96]. Moreover, [97] The reviewer graphs were expanded to identify collusive customers or a transitory group of employed people to spam. Again, no research was done on extremists at a group level, particularly about a brand, because extremism impacts "brand attitudes" in the end. Propose GGSPAM, an upper computational method for detecting reviewing spamming organizations by leveraging the topographical nature and extent reviewing graphs, shows co-reviews' collusiveness. It has already been much buzz around detecting dishonest, spammer, or fraud consumers in online analysis places. Suggest techniques for determining the level of spamming for each user [89]. Create a technique for calculating the trustworthiness of users, the integrity of comments, and the company's dependability [91]. Inside the chart-based classical, represent authors and their founder in bursting as a Markov random variable, then use the looping fuzzy clustering approach to determine whether or not a commenter is a bot [98]. Introduce Iolaus, a mechanism that defends numerous identification and overall score assaults using internet gratified evaluation methods [87]. Demonstrate whether Yelp's screening is acceptable and its methodology is linked to unusual marketing activity [99]. Present a publisher spamicity platform that helps us describe publisher spamicity as implicit and leverage different observable reviewing behavioral traces [92]. It identifies fraudulent buyer groupings and employs multiple behavioral predictions made from the cooperation phenomena amongst false authors and connection systems that rely on the ties between communities, single writers, and the evaluated items [80]. They expand the approach to include behavior characteristics and handle the slow response issue [82]. Create a system that can identify both false comments and comment spammers simultaneously [84]. A probabilistic viewpoint highlights the challenge of identifying corrupt business cheating in consumer advertisement and presents a new predictive method dubbed the Latent Collision Model (LCM) to mechanism corrupt business comment theft [88]. To compute these inherent rankings for all customers, reviews, and goods, create a REV2 method. Use GGSPAM to identify fraudster organizations using the reviewing sample's hierarchical architecture, which shows the founder scheming [96]. Used Defrauder, an approach for detecting cybercrime reviewing teams that is uncontrolled [97]. Individuals are grouped when they have directly evaluated (goods more) several brands. Therefore, the groupings are derived utilizing standard itemset searching on brand characteristics. To categorize prospective groupings as extremist groups, create a content supervised model. The structure of the user subgroups is based on eight attributes unique to each collection, variety combination, according to the hypothesis. Reliability in rankings, comment mood, confirmed purchases, post times, and valuable comments earned on evaluations exemplify such characteristics [100] shown in Table 2.1.

Table 2.1: Review on latest approaches.

| **Ref** | **Targeted Problem** | **Features** | **Datasets** | **Machine learning Models** | **Performance** |
|---|---|---|---|---|---|
| Mukherjee, Liu [80] | PSpotting fake reviewer groups | Group time window, group deviation, group content similarity. | Amazon product reviews | Support vector machine, Logistic regression | Accuracy = 95% |

| | | | | | |
|---|---|---|---|---|---|
| Kumar, Hooi [82] | Identify propaganda and non-propaganda articles from news outlets based on their content | Fairness of a user, reliability of a rating, goodness of a product | Flipkart and amazon user review dataset | Naïve Bayes Model | Accuracy = 84% |
| Lu, Zhang [84] | Simultaneously detecting fake reviews and review Spammers | Reviewer rating difference, review number, group rating feature, average product rating. | Amazon electronic reviews | Support vector machine, Logistic regression | Accuracy = 90% |
| Xu and Zhang [88] | Collusive fraud detection in online reviews | Homogeneity based collusive behavior measures | Amazon review dataset. | Support vector machine | Accuracy = 80% F1 = 75% |
| Wang, Gu [96] | Review spammer group detection | Average time window, rating variance. | Amazon and yelp review dataset | Decision Tree, XGBoost | F1 = 95.50% |

| | | | | | |
|---|---|---|---|---|---|
| Dhawan, Gangireddy [97] | Spotting collective behavior of online frauds in customer reviews | Review tightness, neighbor tightness, product tightness, time window | Amazon, yelp, and play store reviews dataset | Random Forest, Support vector machine | AUC = 84.10% |
| Gupta, Aggarwal [100] | Detecting and characterizing extremist reviewer groups in online product reviews | Metadata features | Amazon reviews dataset | SVM, RF, LR, DT,3-layer perceptron | Precision = 98% Recall = 98% F1 = 98% AUC = 98% |

## 2.4 Research Gap

The majority of the prior studies had focused identification of fake/fraudulent reviewers and their groups. However, only one study addresses the problem of extremist reviewer group's detection using product reviews. There is a need to further explore this topic deeply by highlighting significant features and robust machine learning models.

# Chapter 3

# Proposed Methodology

In this chapter, the framework for the proposed solution is described, as shown in figure 3.1. The division of this chapter is into different sections 3.1 defined dataset description, 3.2 defines features description, 3.3 defined feature selection details, 3.4 ML models, 3.5 evaluation metrics, and 3.6 defined the tools and languages used in our proposed methodology.

A publicly available extremist reviewers' dataset was used. First, we applied normalization. In the filter methods, we utilized the Pearson correlation method, information gain method, and gain ratio method, and in the wrapper method, the forward features selection technique was used for feature selection. Machine learning models, including support vector machine, logistic regression, random forest, decision tree, Gaussian naïve Bayes, stochastic gradient descent, k-nearest neighbor, 3-layer perceptron, 4-layer perceptron, and extreme gradient boosting were used. For training and testing, 10-fold cross-validation was used for all ML models. Precision, recall, F1 measure, AUC were used as evaluation metrics for all ML models. The output of the proposed solution is extreme or moderate reviewers.

## 3.1 Dataset Description

We use a publicly available dataset for our experimental setup of identifying

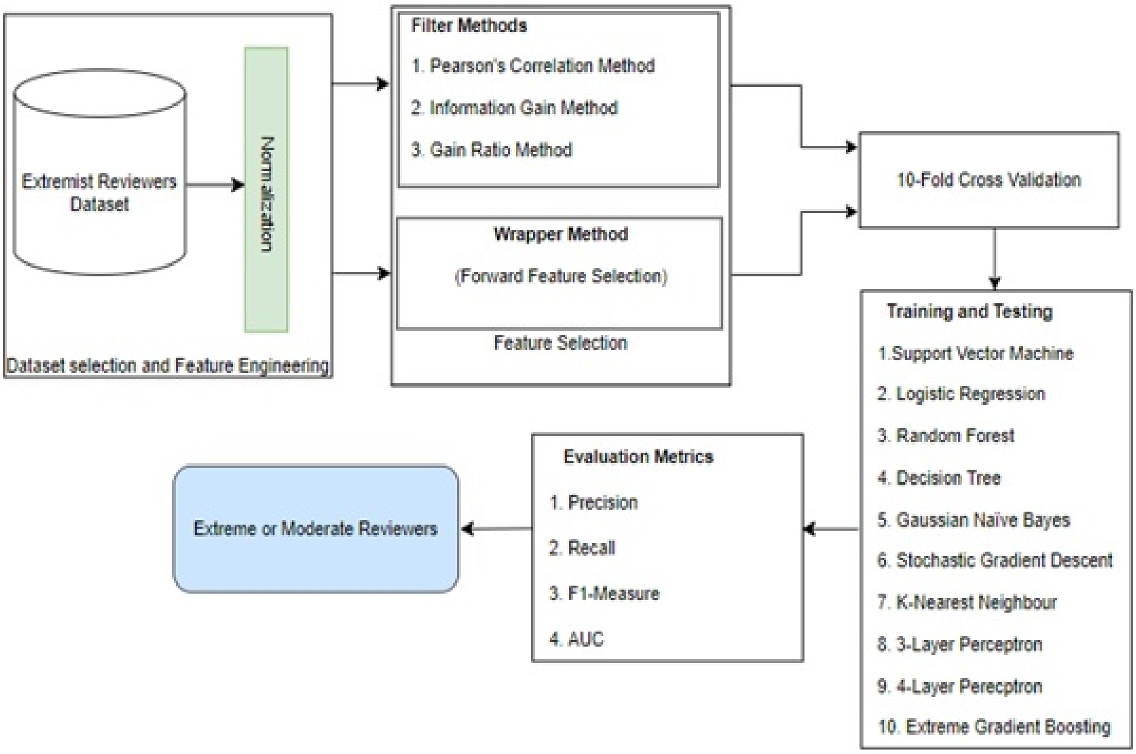


FIGURE 3.1: Block diagram of proposed methodology.

extremist reviewers groups [100]. This dataset consists of 923 potential instances assigned 1 or 0 class labels depending on extreme or moderate. Moderate and extreme groups were allocated to 454 and 469 groups, correspondingly. Amazon extremist reviewer's dataset, as shown in table 3.1 further considered for our experimental setup.

TABLE 3.1: Extremist reviewers dataset

| **No. Total Instances** | **No. Extremist class labels** | **No. Moderate class labels** |
|---|---|---|
| 923 | 469 | 454 |

## 3.2 Features

The baseline dataset [100] contains eight exceedingly beneficial features for spotting extremist groups at the brand level. These features are helpful to identify groups of extremism at the brand level.

Following are the features used in this dataset.

1. Average Rating
2. Average Upvotes
3. Average Sentiment
4. Group Time Window
5. Review Count
6. Rating Deviation
7. Early Time Window
8. Verified Purchase

The details of the features are described above.

### 3.2.1 Average Rating

Rating is the rating given to a particular brand product by the group member. The average rating detentions the average rating assumed to the specific brand B by group G. Average rating has calculated the means of the review's evaluations provided by teammates to items of a particular brand. Rating is the actual rating given to the product via a group member. The baseline paper [100] identifies how an extremist position could provide an average satisfaction score of nearer to five-stars rating or one-star rating at the extremities.

### 3.2.2 Average Upvotes

The average upvotes show the average number of upvotes concerning the given brand the given group receives. The baseline study [100] used this feature to captures the average number of upvotes of a particular group concern with a specific brand. The average of upvotes is calculated by taking the mean of upvotes received crosswise the reviews dispatched by the group members to a particular item of a specific brand. Upvote given to specific brand items by specific group members is the number of upvotes. The reviewer is a group member who posts the reviews for the product of a specific brand.

### 3.2.3 Average Sentiment

The average sentiment is used for identifying the overall sentiment of a group given to the specific brand. The nature of text used to write a review on a product indicates the sentiment of a reviewer. The baseline study discovered the average sentiment of reviews for investigating the review description given by a specific group to a specific brand. For sentiment analysis baseline study and [101] used SentiWordNet 3.0. The SentiWordNet returns between -1 to 1 of an overall sentiment for the review text. Towards a specific brand, an extremist reviewer group may have positive or negative sentiments. So, in the direction of a specific brand, an extremist reviewer group can write highly damaging (-1) or optimistic (+1) review text. The review posted by the group member on the brand product specifies the sentiment of a group member on the product of a brand.

### 3.2.4 Group Time Window

The difference between the latest review posted by the specific group and the earliest review posted by that similar group in favor of a particular brand is the group time window feature utilized by the baseline paper. The group time indicates the significant time difference between the last and the first review posted by any member of a specific group in favor of any product of a specific brand.
The group time window represents the difference between the earliest review and the latest review posted on the products of a given brand by the group members. The group is gratifying in spamming the reviews together and closely linked if the

group time window suggests a lower value of $\tau = 0.28$. The last date of a review posted by any group member on any brand product represents L(G, B), and F(G, B) represents the first date of a review posted by any group member on any brand product [100].

### 3.2.5 Review Count

Review count is the total count of the reviews written by all group members for all particular brand products. This research [102] used the review count feature to capture the total review count of all products given by a fraudulent group. The baseline study used review count to identify an extremist group. The researcher of the baseline study takes the sum of all reviews written by a reviewer who is a group member on all products belonging to the same brand. By using this feature, they found that an extremist group collectively is more likely to write more reviews slightly than other users. The member of the extremist group writes many reviews on all the products of a target brand. The extremist group members write positive reviews if their aim to promote a brand, and if the extremist group wants to demote a group, they write critical reviews.

### 3.2.6 Rating Deviation

Rating deviation is the deviation of the mean rating. To detect reviewers give fake opinions based on an irregular share of ratings that deviate from the public judgment, [103] used rating deviation.
The baseline study used rating deviation to determine how much the group gives deviate rating instead of ordinary people to a particular brand from the mean rating. If an extremist group wants to promote a specific brand, group members write positive reviews. In another case, if an extremist group wants to demote a particular brand, then the group members write a negative opinion. For a brand, the extremist reviewer group writes very coherent reviews. So an extremist group is expected to have a low deviation [100].

### 3.2.7 Early Time Window

The early time window calculates the difference between the time launch of a product on online review portal marketplaces and the last review posted time on this particular item. The time gap since the item launched in a marketplace and the last review posted on that item is capture using an early time window. The early time window feature is used to identify the ethical challenges of defining large core with perfusion [104]. The baseline study used the early time window feature and measured the time difference between the launched time of a product of a specific brand and the last review given by the specific group member. They calculate the mean value of all product reviews of a brand. The researcher found that $\beta = 0.28$ produced the best result to measure the time gap.

### 3.2.8 Verified Purchase

A verified purchase is that if the reviewer purchases a product, this reviewer has more credibility than the reviewer who writes a review only in favor of any product or brand publicity. Verified purchase feature is used to identify verified consumer review effect on sales [105].
The section of reviews posted by a group member on a particular brand product matching to amazon verified purchase reviews is determined using the verified purchase feature.
The baseline study used a verified purchase feature to measure the number of verified reviews posted through a specific group member on a particular brand product. The extremist group members write verified purchased reviews for a particular brand.

## 3.3 Features Selection

The feature selection method is used for dimensionality reduction of the dataset. The dimensionality reduction procedure represents the high dimension data into low dimension data and obtains the actual meaning of the dataset. Processing, analyzing, and visualizing are more manageable using low-dimensional data. Followings are the benefits of applying dimensionality reduction using features selection [106].

- The classification process is simplified, and efficiency is improved by using this method.
- Data storage space can be reduced by using this dimensionality reduction technique.
- Help algorithms to improve accuracy and perform efficiently.
- It helps to remove redundant, irrelevant, and noisy data.
- It improved and examined the design more clearly.
- It utilizes less computation time.
- It improved the quality of data.

In this research, we used the feature selection technique to reduce the dimensionality impact on the original dataset.
The subset of features efficiently defines the data. The possible small subset of features is constructed by obtaining the actual meaning of the data. The feature selection process is divided into two phases.
The first one is subset generation, and the second is subset evaluation [107]. We used two methods for feature selection. These are described below.

1. Filter Methods
2. Wrapper Method

### 3.3.1 Filter Methods

We used filter methods for measuring the characteristics of features. Filter methods are used to determine the measuring criteria of feature characteristics, including information, consistency, dependency, and distance.
For evaluating the ranking of the subset, the filter methods are considered to outperform. Data mining algorithms performed feature selection method independently by using this method [108]. For evaluating the ranking of the subset, we used different standard statistical measures. Those are described below.

1. Pearson Correlation Method
2. Information Gain Method
3. Gain Ratio Method

#### 3.3.1.1 Pearson Correlation Method

We used a statistical correlation measure for ranking the feature subset. The association between two or more quantitative variables is denoted by using the Pearson correlation method. The two or more quantitative variables relationship is assumed as a straight line based on the fundamental examination of this statistical measure [109]. We perform a Pearson correlation method by providing the original dataset as an input, and it gives a ranking of the features as output. Then we selected the top five features as subset features based on the ranking.

#### 3.3.1.2 Information Gain Method

Information gain is a filter selection approach used to select the feature on each node of the decision tree algorithm. Feature having a large number of values is preferred by information gain.
The ID3 algorithm introduced [110] to determine the goodness of a split ID3 algorithm is used information gain. When we perform a statistical measure named

the information gain method by providing the original dataset as an input, it ranks the features as output. Then we selected the top five features as subset features based on the ranking.

#### 3.3.1.3 Gain Ratio Method

The C4.5 is a successor of the ID3 algorithm known as a gain ratio. It overcomes the biased outcomes of information gain because information gain is preferred to select features with many values. For ranking the feature of high dimension dataset, used gain ratio.
The feature or attribute having a maximum gain ratio is selected [111]. For ranking the feature of high dimension dataset, used gain ratio. We perform a statistical measure gain ratio method by providing the original dataset as an input, and it ranks the features as output. Then we selected the top five features as subset features based on the ranking. The subset of features is selected based on having a maximum gain ratio.

### 3.3.2 Wrapper Method

The wrapper method is also a technique of feature selection introduced by [112]. The wrapper method used the learning algorithm to wraps the feature selection. The criteria for feature selection are the performance accuracy or the error rate of the classification process. The wrapper method overcomes the estimated error rate of a particular classifier and selects the most discernment subset of features. The feature selection is performed in the wrapper method based on the classifier performance. The best optimal features for the predicted classifier are selected by the wrapper method. By comparing filter methods, the wrapper method accomplishes the best accuracy and better performance [113]. There are three main techniques used under the wrapper method. These methods are described below. We also used the wrapper method for feature selection. The forward feature selections technique is used for selecting the subset of features under the wrapper method. Figure 3.2 shows the process of feature evaluation and selection of the best subset feature.

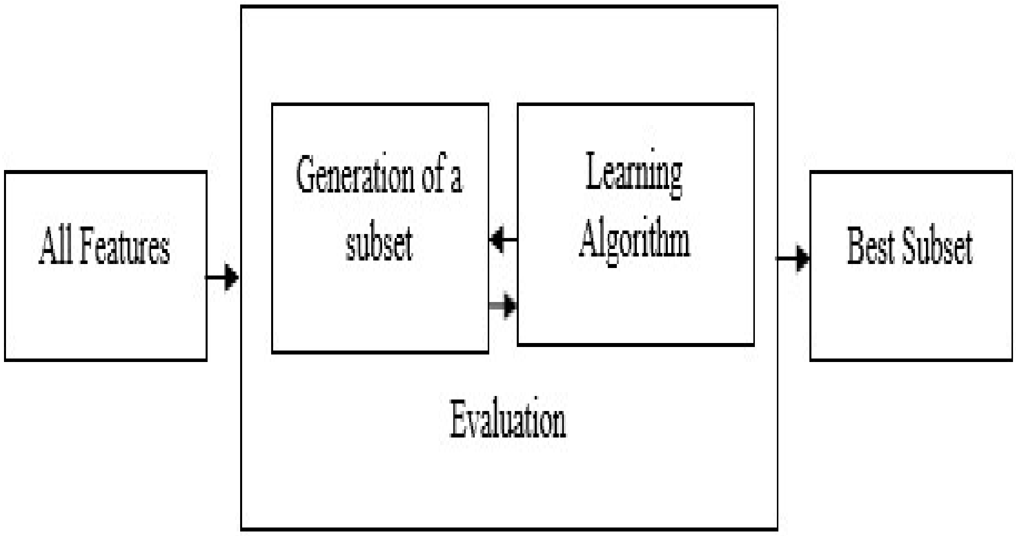


FIGURE 3.2: Methodology adopted by the wrapper method for features selection [114]

1. Backward elimination
2. Forward selection
3. By directional elimination

#### 3.3.2.1 Forward Selection

The forward selection is a technique of features selection under the wrapper method. It is also called the sequential forward feature selection because it selects the step forward feature. The sequential forward feature selection is used as an iterative method. Individually all features are evaluated at the start, and the best performance result feature is selected. After selecting the feature, in the second step, a pair of features are selected by the sequential forward selection, which produces the best classifier performance. After test all possible combinations of the remaining features with the selected feature. To reach the best performance of the model, keep adding the best performance feature in each iteration [115]. This research used the forward feature selection technique under the wrapper method and a machine model named 3-layer perceptron as a learning model. 10-fold cross-validation for train and test is used. The wrapper method provides us five best optimal features by applying a 3-layer perceptron model as a learning model.

## 3.4 Machine Learning Models

Machine learning models are build using the python language for the identification of the extremist reviewer groups. Machine learning techniques are trained and checked using state-of-the-art and proposed characteristics performing different kinds of experiments. The built-in Python packages are used for deploying the machine learning models. For testing the performance of these ML models, 10-fold cross-validation is used in all the experiments. Below is the list of all ML models we utilized in our research.

1. Support Vector Machine
2. Logistic Regression
3. Random Forest
4. Decision Tree
5. Gaussian Naïve Bayes
6. Stochastic Gradient Descent
7. K-Nearest Neighbor
8. 3-Layer Perceptron

9. 4-Layer Perceptron

10. Extreme Gradient Boosting

### 3.4.1 Support Vector Machine

Support vector machine is the machine learning model. SVM is used for classification and regression problems. Constants values in the support vector machine formula are calculated by using the training dataset. The variables and constants are used in the formula of the support vector machine. The support vector machine performs predictions based on the values of the constants, where the variables are changed as the record changes. The support vector machine has good accuracy [116]. The constants calculation of our research is as follows. Targeted feature values are anticipated for each testing record by using the support vector machine. From the Sklearn package, the SVM module is SVC used for implementation in this research. No particular parameters are used to build this model. The 10-fold cross-validation is used to perform testing and training.

### 3.4.2 Logistic Regression

There are many types of machine learning models. Logistic regression is one of them which is used for regression and classification problems. Logistic regression is outperformed in classification problems. In many situations, logistic regression may be used for regression problems. Constants values of the logistic regression formula are calculated by making a judgment based on any record of this model. The logistic regression does not provide a linear connection because an exponential function is incorporated in the logistic regression formula [117] Records of training datasets are used to calculate the values. For decision making the calculated values of constants are used. The determination performance of the models may vary because the logistic regression uses the constant values of the record. The constant values of the record may vary according to the variation in the record data.

The implementation of the logistic regression model is performed from the Sklearn package linear model class of the logistic regression model. The logistic regression model is given a penalty function of 12 and the maximum iteration of 200 as particular arguments. 10-fold cross-validation is performed for training and testing.

### 3.4.3 Random Forest

To solving the difficulties in the categorization, a machine learning model, random forest, is used. The random forest model is calculated information gain and entropy of every attribute to get the relation towards the targeted feature [118]. The entropy and information gain are calculated by using the training dataset. Different trees are utilized in this model, and every tree is used with a split objective. The change in the variable varies according to the changes in the record. Targeted feature values are anticipated for each testing record by using the random forest model. From the Sklearn package, the ensemble class module Random forest classifier is used for implementation in this research. The criteria of Gini and the maximum depth as four are given to the random forest as particular arguments. The 10-fold cross-validation is used to perform testing and training.

### 3.4.4 Decision Tree

Classification problems also solved by using a machine learning model named

the decision tree model is used. The decision tree model is calculated the information gain and entropy of every attribute to get the relation towards the targeted feature. The training dataset is used to calculate the information gain and entropy. The variable values are changed according to the changes in the recorded values. The record values are calculated: targeted feature values are anticipated using the random forest model for each testing record [119]. From the Sklearn library, the tree class module for the decision tree classifier is utilized for implementation in this research. The criteria of Gini and the maximum depth as four are given to the decision tree classifier as particular arguments. The 10-fold cross-validation is used to perform testing and training.

### 3.4.5 Gaussian Naive Bayes

The classification problem is also solved by using a machine learning model known as gaussian naïve Bayes. Gaussian naive Bayes model has a statistical notation in a mathematical formula. The model derived constants values of the formula and saved the values by using the training dataset. In the prediction process of model constants values, derived from the training dataset are employed with the usage of the testing dataset.

In the training dataset, the distribution of each class is included in the total number of classes. A class is assigned to distribution according to the prediction presence. The prediction presence of the model is predicted by examining the prediction goal and added each new record [120].

From the Sklearn library, the naïve Bayes class for the gaussian naïve Bayes module is utilized in this research. No particular parameters are used to build this model. The 10-fold cross-validation is used to perform testing and training.

### 3.4.6 Stochastic Gradient Descent

Stochastic gradient descent is a machine learning algorithm used for classification and linear regression problems. An objective function is used in stochastic gradient descent for optimization. The suitable smoothness properties of the objective function are also used in this model.

Stochastic gradient descent works like an iterative algorithm [121]. Targeted feature values are anticipated for each testing record by using the stochastic gradient descent model. From the Sklearn liner model library, the SGD classifier module is used in this research. The loss function of hinge maximum and penalty of 12 and maximum iteration of 200 are given to the Stochastic gradient boosting as particular arguments. The 10-fold cross-validation is used to perform testing and training.

### 3.4.7 K-Nearest Neighbor

Classification problems also solved by using a machine learning model named the k-nearest neighbors is used. The model provided a set of constants by defining the lacks formula and used the training dataset to compute the constant values. A rating system is based on that concept. A training dataset is required for this model to absorb how to rank records [122]. During the testing phase, every other record is compared with each record in the training dataset. For comparison purposes, the Euclidean formula is used. This model computes the distance between all existing training records and every testing record. The training records are sorted in ascending order by using the distances between each testing record. After sorting the records, the best record is obtained by using the number of neighbors. The best record is chosen based on votes. A class that appears in the most chosen record is anticipated based on the testing record. This procedure is repeated for each testing record. From the Sklearn library, the class neighbors of the k neighbors classifier module are used in this research. The k-nearest neighbors as five are given to the k-nearest neighbor classifier as particular arguments. The 10-fold cross-validation is used to perform testing and training.

### 3.4.8 3-Layer Perceptron

Multi-layer perceptron is used both for regression and classification problems. Feedforward neural network supplement is known as MLP. An input layer, hidden

layer, and output layer are the three layers used in MLP. Signals are received and processed by the input layer.

An accurate computational engine is an arbitrary number of hidden layers. The output layer performs classification and prediction [123]. Targeted feature values are anticipated for each testing record by using the three-layer perceptron model. From the Sklearn neural network library, the multi-layer perceptron classifier module is used in this research. The learning rate of le-5, random state of 0, two hidden layers with every hundred neurons, and activation function as logistic function are given to the 3-layer perceptron classifier as particular arguments. The 10-fold cross-validation is used to perform testing and training.

### 3.4.9 4-Layer Perceptron

The multi-layer perceptron is a machine learning model, and different layers are used for testing and validate the classifier or regression problems. By selecting the two hidden layers and one input layer, this research used 4 layers perceptron model [124].
Targeted feature values are anticipated for each testing record by using the four-layer perceptron model. From the Sklearn neural network library, the multi-layer perceptron classifier module is used in this research. The learning rate of le-5, random state of 0, three hidden layers with every hundred neurons, and activation function as logistic function are given to the 4-layer perceptron classifier as particular arguments. The 10-fold cross-validation is used to perform testing and training.

### 3.4.10 Extreme Gradient Boosting

Classification and regression problems are also solved using a machine learning model known as extreme gradient boost (XGBoost). The values of the constants in the formula are calculated by using the training dataset in this model. Constants and variables are including in its formula.
The prediction of this model is based on the values of the constants. The values of the variables vary according to changes in the record [125]. Targeted feature values are anticipated for each testing record by using the extreme gradient boosting model. From the Sklearn ensemble library, the XGB classifier module is used in this research. The maximum depth is four, and the learning rate of 0.1 and the number of estimators of 100 are given to the extreme gradient boosting as particular arguments. The 10-fold cross-validation is used to perform testing and training.

## 3.5 Evaluation Metrics

Evaluation metrics are used to evaluate the performance of our proposed methodology. These are the Precision, Recall, AUC, F1-Measure.

### 3.5.1 Precision

For evaluating how accurate results are produced in our applied ML models, precision is our first evaluation metric. Equation 3.1 defines the standard formula for the assessment of precision results.

$$Precision = \frac{TP}{TP + FP} \tag{3.1}$$

### 3.5.2 Recall

The recall is our second evaluation metric to evaluate how many instances are captured as actual positive for our applied ML models. Equation 3.2 defines the standard formula used for the assessment of recall results.

$$Precision = \frac{TP}{TP + FN} \tag{3.2}$$

### 3.5.3 F1-Measure

When we need to seek relationships between an uneven class distribution recall and precision, F1-measure is our third evaluation metric. Equation 3.3 defines the standard formula used for the assessment of the F1 measure.

$$F_1 = 2 * \frac{precision * recall}{precision * recall} \tag{3.3}$$

### 3.5.4 Area Under the Curve

AUC is our fourth evaluation metric. At all the classification thresholds ROC curve graph shows the performance of a classification model. Two parameters are plot in this curve as actual positive rate and false-positive rate.

## 3.6 Tools and Languages

Different tools and languages are used for the evaluation of our experimental results. These are described below.

- Python – is used for all the algorithms implementation.
- Weka – a data mining tool, is used for feature selection.
- Microsoft excel – All the calculated results are stored by using Microsoft excel.
- Google Colab

# Chapter 4

# Results and Analysis

This chapter describes the complete description of all the results and analysis composed from our set of experiments.Section 4.1 describes the Experimental setup of hardware and operating system requirements. Section 4.2 describes the Experiment 1 classifiers' performances for extremist reviewer groups identification.

Sec-tion 4.3 describes Experiment 2 Impact of feature selection on classification, which covers the experiments on all selected filter-based and wrapper-based features se-lection methods. Section 4.4 describes Experiment 3 Performance of Individual features on classification using precision, recall, F1, AUC evaluation metrics

## 4.1 Experimental Setup

For analysis following hardware and software is used.

**Hardware Requirements**

Table 4.1 shows the hardware requirements. Development software's and operating system requirements

Table 4.2 shows the development software's and operating system requirements.

Table 4.1: Hardware requirements.

| **Processor** | **RAM** | **Hard disk** |
|---|---|---|
| Intel® Core$^{TM}$ i5-5300U Processor | 16GB | 500GB |

Table 4.2: Hardware requirements.

| **Window** | **Language** | **Results** |
|---|---|---|
| Window 10 | Python 3.9 | Microsoft Excel 2017 |

## 4.2 Experiment 1 Classifiers Performances for Extremist Reviewer Groups Detection

In our first experiment, we were interested in examining the performance of the classifiers for detecting extremist reviewers by using the same dataset of our baseline study [100]. This experiment is conducted by checking the performance of ML models using the amazon extremist reviewer's dataset. We implemented Support Vector Machine, Logistic Regression, Random Forest, Decision Tree, Gaussian Naïve Bayes, Stochastic Gradient Descent, K-Nearest Neighbor, 3-Layer Perceptron, 4-Layer Perceptron, Extreme Gradient Boosting as the machine learning models. For testing and training, the 10-fold cross-validation is used for all ML models. The precision, recall, f1, and AUC are used as the evaluation metrics.

### 4.2.1 Classifiers Performances using Micro-Average

Table 4.3 illustrates the results of normalization and without normalization. The impact of normalization on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 2.9%. K-Nearest Neighbor is improved by 2.65%. Compare to the improvement Gaussian Naïve Bayes improved more than K-Nearest Neighbor after normalization. Compare to the performance K-Nearest Neighbor generates better results than

Gaussian Naïve Bayes. When we apply normalization on Support Vector Machine its improved by 3.87%. The improvement in Support Vector Machine is more than Gaussian Naïve Bayes and K-Nearest Neighbor. The Support Vector Machine also generates better results as compared to Gaussian Naïve Bayes and K-Nearest Neighbor. Decision Tree improved by 3.77% after normalization which is not more than the improvement of Support Vector Machine.
However, the Support Vector Machine and Decision Tree generate almost same results. Logistic Regression is improved by 3.89%. Compare with both improvement and performance Logistic Regression generates better results than Support Vector Machine and Decision Tree.
Stochastic Gradient Descent improved by 3.41% after normalization. Compare to the improvement Stochastic Gradient Descent improved less than Logistic Regression. Compare to the performance of Stochastic Gradient Descent and Logistic Regression generates almost same results.

TABLE 4.3: Classifiers performances using micro-average

| **Classifiers** | **Without Normalization** | | | | **With Normalization** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 87.25 | 87.25 | 87.25 | 87.31 | 90.15 | 90.15 | 90.15 | 90.21 |
| **KNN** | 88.55 | 88.55 | 88.55 | 88.48 | 91.21 | 91.21 | 91.21 | 91.21 |
| **SVM** | 90.23 | 90.22 | 90.22 | 90.22 | 94.11 | 94.11 | 94.11 | 94.11 |
| **Decision Tree** | 90.33 | 90.33 | 90.33 | 90.33 | 94.11 | 94.11 | 94.21 | 94.31 |
| **Logistic Regression** | 91.43 | 91.42 | 91.42 | 91.42 | 95.32 | 95.33 | 95.32 | 95.32 |
| **SGD** | 92.09 | 92.08 | 92.06 | 92.08 | 95.51 | 95.16 | 95.14 | 95.51 |
| **Random Forest** | 92.11 | 92.11 | 92.11 | 92.12 | 95.81 | 95.51 | 95.71 | 95.71 |
| **XGBoost** | 91.89 | 91.89 | 91.89 | 91.91 | 96.24 | 96.31 | 96.31 | 96.24 |
| **4-Layer Perceptron** | 92.89 | 92.78 | 92.78 | 92.89 | 96.52 | 96.52 | 96.52 | 96.52 |
| **3-Layer Perceptron** | **93.09** | **93.07** | **93.07** | **93.08** | **96.87** | **96.87** | **96.87** | **96.87** |

Random Forest is improved by 3.69%. Compare to the improvement Random Forest improved less than Logistic Regression. Compare to the performance the Random Forest generates better results. XGBoost is improved by 4.34%. Compare to the improvement XGBoost improved more than Logistic Regression.
Compare to the performance the XGBoost generates better than Random Forest.
4-Layer Perceptron improved by 3.63%. Compare to the improvement 4-Layer Perceptron improved less than XGBoost.
Compare to the performance the 4-Layer Perceptron generates better results than XGBoost. 3-Layer Perceptron improved by 3.78%. Compare to the improvement

3-Layer Perceptron improved less than XGBoost. Compare to the performance the 3-Layer Perceptron generates best results among all ml models.

### 4.2.2 Classifiers Performances using Macro-Average

Table 4.4 illustrates the results of normalization and without normalization. The impact of normalization on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 2.9%. K-Nearest Neighbor is improved by 2.7%.
Compare to the improvement Gaussian Naïve Bayes improved more than K-Nearest Neighbor after normalization. Compare to the performance K-Nearest Neighbor generates better results than Gaussian Naïve Bayes.
When we apply normalization on Support Vector Machine its improved by 3.95%. The improvement in Support Vector Machine is more than Gaussian Naïve Bayes and K-Nearest Neighbor.
The Support Vector Machine also generates better results as compared to Gaussian Naïve Bayes and K-Nearest Neighbor.

TABLE 4.4: Classifiers performances using macro-average

| **Classifiers** | **Without Normalization** | | | | **With Normalization** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 87.41 | 87.41 | 87.41 | 87.44 | 90.31 | 90.11 | 90.11 | 90.21 |
| **KNN** | 88.61 | 88.62 | 88.62 | 88.61 | 91.31 | 91.27 | 91.27 | 91.27 |
| **SVM** | 90.25 | 90.23 | 90.23 | 90.25 | 94.21 | 94.11 | 94.11 | 94.11 |
| **Decision Tree** | 90.34 | 90.33 | 90.33 | 90.34 | 94.21 | 94.11 | 94.11 | 94.31 |
| **Logistic Regression** | 91.44 | 91.43 | 91.43 | 91.42 | 95.41 | 95.23 | 95.23 | 95.32 |
| **SGD** | 92.11 | 92.11 | 92.11 | 92.11 | 95.61 | 95.21 | 95.21 | 95.51 |
| **Random Forest** | 92.15 | 92.12 | 92.12 | 92.15 | 95.91 | 95.81 | 95.81 | 95.71 |
| **XGBoost** | 91.88 | 91.88 | 91.88 | 91.89 | 96.31 | 96.32 | 96.32 | 96.24 |
| **4-Layer Perceptron** | 92.89 | 92.88 | 92.88 | 92.89 | 96.61 | 96.53 | 96.52 | 96.52 |
| **3-Layer Perceptron** | **93.11** | **93.11** | **93.11** | **93.11** | **96.89** | **96.89** | **96.89** | **96.89** |

Decision Tree improved by 3.87% after normalization which is not more than the improvement of Support Vector Machine. However, the Support Vector Machine and Decision Tree generate almost same results.
Logistic Regression is improved by 3.96%. Compare with both improvement and performance Logistic Regression generates better results than Support Vector Machine and Decision Tree. Stochastic Gradient Descent improved by 3.5%

after normalization. Compare to the improvement Stochastic Gradient Descent improved less than Logistic Regression.
Compare to the performance of Stochastic Gradient Descent and Logistic Regression generates almost same results.

## 4.3 Experiment 2 Impact of Feature Selection on Classification

In this type of experiment, we used the filter methods and wrapper method for feature selection. We are reporting our experiments on top-four, top-five and top-six features.
To achieve the best, we selected the top-five features based on ranking given by different filter methods. The wrapper method also selected the five best features using forward feature selection.

### 4.3.1 Feature Selection using Pearson Correlation Method

For selecting the potential features, we used a statistical measure, the Pearson correlation method. We apply a Pearson correlation method to the amazon reviewer's dataset. The Pearson correlation method gives us top-five features based on the ranking. We evaluate our results of ML models using the top-four, top-five and top-six features selected by the Pearson correlation method.

#### 4.3.1.1 Top-Five Features

The subset of top-five features selected by the Pearson correlation method is mentioned below.

1. Review count
2. Early time window

3. Rating deviation

4. Group time window

5. Verified purchase

Table 4.5 illustrates the results of without selection and with selection of top-five features. The impact of top-five features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 0.76%. K-Nearest Neighbor is improved by 4.6%. Support Vector Machine improved by 2.46%. Decision Tree improved by 2.48%. Logistic Regression is improved by 0.69%. Stochastic Gradient Descent improved by 3.5%. Random Forest is improved by 3.45%. XGBoost is improved by 0.86%. 4-Layer Perceptron improved by 0.78%. 3-Layer Perceptron improved by 0.9%. Table 4.6

TABLE 4.5: Top five feature selection using pearson correlation method (micro-average)

| **Classifiers** | **Without Selection** | | | | **With Top-5** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 90.91 | 90.91 | 90.91 | 91.09 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 95.81 | 95.81 | 95.81 | 95.81 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 96.56 | 96.55 | 96.55 | 96.45 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 96.58 | 96.11 | 96.11 | 96.81 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 96.01 | 96.01 | 96.01 | 96.01 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.41 | 96.41 | 96.41 | 95.91 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 97.15 | 97.08 | 97.08 | 97.09 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 97.11 | 97.11 | 97.11 | 97.11 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 97.31 | 97.31 | 97.31 | 97.31 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **97.77** | **97.77** | **97.77** | **97.76** |

illustrates the results of without selection and with selection of top-five features. The impact of top-five features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 1.2%. K-Nearest Neighbor is improved by 4.7%. Support Vector Machine improved by 2.41%. Decision Tree improved by 2.39%. Logistic Regression is improved by 0.69%. Stochastic Gradient Descent improved by 0.92%. Random Forest is improved by 1.3%. XGBoost is improved by 0.8%. 4-Layer Perceptron improved by 0.8%. 3-Layer Perceptron improved by 0.88%. After selection of top-five features both with micro and macro average compare to the improvement K-Nearest Neighbor is improved more than among all other applied ml models.

Table 4.6: Top five feature selection using pearson correlation method (macro-average)

| Classifiers | Without Selection | | | | With Top-5 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 91.51 | 91.01 | 89.81 | 91.09 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 96.01 | 95.81 | 95.81 | 95.81 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 96.61 | 96.55 | 96.55 | 96.45 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 96.61 | 96.11 | 96.11 | 96.81 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 96.09 | 96.01 | 96.01 | 96.01 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 96.52 | 96.52 | 96.52 | 96.52 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 97.21 | 97.08 | 97.08 | 97.09 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 97.11 | 97.11 | 97.11 | 97.11 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 97.41 | 97.31 | 97.31 | 97.31 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **97.77** | **97.77** | **97.77** | **97.76** |

Compare to the performance the 3-Layer Perceptron generates best results among all other applied ml models.

#### 4.3.1.2 Top-Four Features

The subset of top-four features selected by the Pearson correlation method is mentioned below.

1. Review count
2. Early time window
3. Rating deviation
4. Group time window

Table 4.7 illustrates the results of without selection and with selection of top-four features. The impact of top-four features on all applied ml models and metrics is clearly showing in this table along micro average.

Gaussian Naïve Bayes is decreased by 0.25%. K-Nearest Neighbor is improved by 4%. Support Vector Machine improved by 1.53%. Decision Tree improved by 2.59%. Logistic Regression is improved by 0.66%. Stochastic Gradient Descent improved by 0.59%. Random Forest is improved by 0.59%. XGBoost is improved by 0.58%. 4-Layer Perceptron decreased by 4.02%. 3-Layer Perceptron improved by 0.41%.

Table 4.7: Top four feature selection using pearson correlation method (micro-average)

| Classifiers | Without Selection | | | | With Top-4 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 89.91 | 89.91 | 89.91 | 89.97 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 95.21 | 95.21 | 95.21 | 95.18 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 95.63 | 95.63 | 95.63 | 95.61 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 96.69 | 96.69 | 96.69 | 96.71 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 95.98 | 95.98 | 95.98 | 95.93 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.09 | 96.74 | 96.41 | 96.84 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 96.39 | 96.71 | 96.51 | 96.39 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 96.82 | 96.82 | 96.82 | 96.83 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 92.51 | 92.51 | 92.51 | 92.59 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **97.28** | **97.28** | **97.28** | **97.27** |

Table 4.8 illustrates the results of without selection and with selection of top-four features. The impact of top-four features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is decreased by 0.83%. K-Nearest Neighbor is improved by 4.01%. Support Vector Machine improved by 1.55%.

Decision Tree improved by 2.47%. Logistic Regression is improved by 0.79%. Stochastic Gradient Descent improved by 0.19%. Random Forest is improved by 0.46%.

XGBoost is improved by 0.55%. 4-Layer Perceptron decreased by 6.5%. 3-Layer Perceptron improved by 0.43%.

Table 4.8: Top four feature selection using pearson correlation method (macro-average)

| Classifiers | Without Selection | | | | With Top-4 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 89.47 | 89.97 | 88.52 | 89.97 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 95.31 | 95.18 | 95.21 | 95.18 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 95.75 | 95.61 | 95.63 | 95.61 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 96.68 | 96.63 | 96.63 | 96.68 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 96.19 | 95.93 | 95.98 | 95.93 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 95.79 | 96.08 | 96.74 | 95.87 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 96.36 | 96.51 | 96.61 | 96.72 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 96.85 | 96.83 | 96.82 | 96.83 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 90.11 | 92.59 | 90.89 | 92.59 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **97.32** | **97.27** | **97.28** | **97.27** |

After selection of top-four features both with micro and macro average compare to the improvement K-Nearest Neighbor is improved more than among all other applied ml models. Compare to the performance the 3-Layer Perceptron generates best results among all other applied ml models.

#### 4.3.1.3 Top-Six Features

The subset of top-six features selected by the Pearson correlation method is mentioned below.

1. Review count
2. Early time window
3. Rating deviation
4. Group time window
5. Verified purchase
6. Average upvotes

Table 4.9 illustrates the results of without selection and with selection of top-six features. The impact of top-six features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 0.73%. K-Nearest Neighbor is improved by 4.03%. Support Vector Machine improved by 2.21 %. Decision Tree improved by 1.4%. Logistic Regression is improved by 1.31%. Stochastic Gradient Descent improved by 0.7%. Random Forest is improved by 0.59%. XGBoost is improved by 1.02%. 4-Layer Perceptron decreased by 0.13%. 3-Layer Perceptron improved by 0.28%. Table 4.10 illustrates the results of without selection and with selection of top-six features. The

TABLE 4.9: Top six feature selection using pearson correlation method (micro-average)

| Classifiers | Without Selection | | | | With Top-6 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 90.88 | 90.88 | 90.88 | 90.95 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 95.23 | 95.23 | 95.23 | 95.21 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 96.31 | 96.31 | 96.31 | 96.26 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 95.51 | 95.39 | 95.39 | 95.51 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 96.63 | 96.63 | 96.63 | 96.59 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.21 | 96.63 | 96.21 | 96.53 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 96.82 | 96.82 | 96.93 | 96.72 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 96.93 | 96.93 | 96.93 | 96.94 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 96.39 | 96.39 | 96.39 | 96.41 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **97.15** | **97.17** | **97.17** | **97.16** |

impact of top-six features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 1.18%. K-Nearest Neighbor is improved by 4.06%. Support Vector Machine improved by 2.36%. Decision Tree improved by 1.44%. Logistic Regression is improved by 1.41%. Stochastic Gradient Descent improved by 0.85%. Random Forest is improved by 0.96%. XGBoost is improved by 1.65%. 4-Layer Perceptron decreased by 0.17%. 3-Layer Perceptron improved by 0.28%. After selection of top-six features both with micro and macro average compare to the improvement K-Nearest Neighbor is improved more than among all other applied ml models. Compare to the performance the 3-Layer Perceptron generates best results among all other applied ml models.

TABLE 4.10: Top six feature selection using pearson correlation method (macro-average)

| **Classifiers** | **Without Selection** | | | | **With Top-6** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 91.48 | 90.95 | 89.71 | 90.95 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 95.36 | 95.21 | 95.22 | 95.21 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 96.56 | 96.26 | 96.31 | 96.26 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 95.65 | 95.72 | 95.39 | 95.75 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 96.81 | 96.59 | 96.63 | 96.59 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 96.45 | 96.85 | 96.72 | 96.64 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 96.86 | 96.83 | 96.72 | 96.73 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 96.95 | 96.94 | 96.93 | 96.94 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 96.43 | 96.41 | 96.39 | 96.41 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **97.17** | **97.17** | **97.17** | **97.16** |

### 4.3.2 Feature Selection using Information Gain Method

For selecting the potential features, we also used a statistical measure of information gain. We apply the information gain method to the amazon reviewer's dataset. We evaluate our results of ML models using the top-five, top-four, top-six features selected by the information gain method.

#### 4.3.2.1 Top-Five Features

The top-five features selected by the Information Gain method are mentioned below.

1. Review count

2. Average rating

3. Early time window

4. Verified purchase

5. Average upvotes

TABLE 4.11: Top five feature selection using information gain method (micro-average)

| Classifiers | Without Selection | | | | With Top-5 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 92.63 | 92.63 | 92.63 | 92.52 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 95.34 | 95.34 | 95.34 | 95.31 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 96.31 | 96.31 | 96.31 | 96.26 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 97.28 | 97.39 | 97.28 | 97.07 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 96.53 | 96.53 | 96.53 | 96.48 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.11 | 96.42 | 97.18 | 96.73 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 97.93 | 97.82 | 98.15 | 97.94 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 98.15 | 98.15 | 98.15 | 98.15 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 98.01 | 98.01 | 98.01 | 98.01 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **98.29** | **98.25** | **98.25** | **98.25** |

Table 4.11 illustrates the results of without selection and with selection of top-five features. The impact of top-five features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 2.48%. K-Nearest Neighbor is improved by 4.14%. Support Vector Machine improved by 2.21%. Decision Tree improved by 3.18%. Logistic Regression is improved by 1.21%. Stochastic Gradient Descent improved by 0.6%. Random Forest is improved by 2.13%. XGBoost is improved by 1.91%. 4-Layer Perceptron improved by 1.49%. 3-Layer Perceptron improved by 1.42%. Table 4.12 illustrates the results of without selection and with selection of top-four features. The impact of top-four features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 3.24%.

K-Nearest Neighbor is improved by 4.2%. Support Vector Machine improved by 2.35%. Decision Tree improved by 3.11%. Logistic Regression is improved by 1.3%. Stochastic Gradient Descent improved by 1.29%. Random Forest is improved by 1.96%. XGBoost is improved by 1.87%. 4-Layer Perceptron improved by 1.45%. 3-Layer Perceptron improved by 1.4%. The performance of the Gaussian Naïve

TABLE 4.12: Top five feature selection using information gain method (macro-average)

| Classifiers | Without Selection | | | | With Top-5 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 93.54 | 92.52 | 92.57 | 92.52 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 95.51 | 95.31 | 95.33 | 95.31 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 96.55 | 96.26 | 96.3 | 96.26 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 97.32 | 97.17 | 97.28 | 97.28 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 96.71 | 96.48 | 96.52 | 96.48 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 96.89 | 95.88 | 96.41 | 96.74 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 97.86 | 97.83 | 97.72 | 98.04 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 98.17 | 98.15 | 98.15 | 98.15 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 98.05 | 98.01 | 98.02 | 98.01 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **98.29** | **98.25** | **98.25** | **98.25** |

Bayes classifier is 93.54%, which is worse than other classifiers' performances because Naïve Bayes assumes an independent predictor. The performance of the 3-Layer Perceptron classifier performance is 98.29% seems to be the best classifier among all other classifiers' performances because this model has the ability of adaptive learning.

#### 4.3.2.2 Top-Four Features

The top-four features selected by the Information Gain method are mentioned below.

1. Review count
2. Early time window
3. Average rating
4. Verified purchase

Table 4.13 illustrates the results of without selection and with selection of top-four features. The impact of top-four features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 2.37%.

K-Nearest Neighbor is improved by 4.57%. Support Vector Machine improved by 2.86%. Decision Tree improved by 2.75%. Logistic Regression is improved by 0.99%. Stochastic Gradient Descent improved by 0.27%. Random Forest is improved by 2.35%.

XGBoost is improved by 1.37%. 4-Layer Perceptron decreased by 7.61%. 3-Layer Perceptron improved by 1.5%. Table 4.14 illustrates the results of without selection

Table 4.13: Top four feature selection using information gain method (micro-average)

| Classifiers | Without Selection | | | | With Top-4 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 92.52 | 92.52 | 92.52 | 92.41 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 95.77 | 95.77 | 95.77 | 95.75 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 96.96 | 96.96 | 96.96 | 96.93 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 96.96 | 96.85 | 97.17 | 96.97 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 96.31 | 96.31 | 96.31 | 96.27 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 95.77 | 96.21 | 96.31 | 95.86 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 98.15 | 98.15 | 98.15 | 98.27 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 97.61 | 97.61 | 97.72 | 97.72 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 88.91 | 88.91 | 88.91 | 89.13 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **98.37** | **98.37** | **98.37** | **98.38** |

and with selection of top-four features.

The impact of top-four features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 3.15%. K-Nearest Neighbor is improved by 4.6%. Support Vector Machine improved by 2.86%. Decision Tree improved by 2.8%. Logistic Regression is improved by 1.09%.

Stochastic Gradient Descent improved by 0.37%. Random Forest is improved by 2.17%. XGBoost is improved by 1.45%. 4-Layer Perceptron decreased by 12.56%. 3-Layer Perceptron improved by 1.5%. The performance of the Gaussian

Table 4.14: Top four feature selection using information gain method (macro-average)

| Classifiers | Without Selection | | | | With Top-4 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 93.45 | 92.41 | 92.46 | 92.41 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 95.91 | 95.75 | 95.76 | 95.75 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 97.06 | 96.93 | 96.95 | 96.93 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 97.01 | 97.19 | 96.95 | 96.97 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 96.49 | 96.27 | 96.31 | 96.27 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 95.97 | 96.73 | 95.76 | 96.74 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 98.07 | 98.17 | 97.93 | 98.39 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 97.75 | 97.72 | 97.61 | 97.72 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 84.04 | 89.13 | 85.71 | 89.13 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **98.39** | **98.38** | **98.37** | **98.38** |

Naïve Bayes classifier is 93.45%, which is worse than other classifiers' performances

because Naïve Bayes assumes an independent predictor.
The performance of the 3-Layer Perceptron classifier performance is 98.39% seems to be the best classifier among all other classifiers' performances because this model has the ability of adaptive learning.

#### 4.3.2.3 Top Six Features

The top-six features selected by the Information Gain method are mentioned below.

1. Review count
2. Early time window
3. Rating deviation
4. Group time window
5. Verified purchase
6. Average upvotes

Table 4.15 illustrates the results of without selection and with selection of top-six features. The impact of top-six features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 1.37%. K-Nearest Neighbor is improved by 3.27%. Support Vector Machine improved by 1.88%. Decision Tree improved by 3.18%. Logistic Regression is decreased by 0.12%. Stochastic Gradient Descent improved by 1.14%. Random Forest is improved by 2.02%. XGBoost is improved by 1.8%. 4-Layer Perceptron improved by 1.52%. 3-Layer Perceptron improved by 1.5%. Table 4.16 illustrates the results of without selection and with selection of top-six features. The impact of top-six features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 2.12%. K-Nearest Neighbor is improved by 3.44%. Support Vector Machine improved by 1.95%. Decision Tree improved by 3.24%. Logistic Regression is

Table 4.15: Top six feature selection using information gain method (micro-average)

| Classifiers | Without Selection | | | | With Top-6 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 91.52 | 91.52 | 91.52 | 91.46 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 94.47 | 94.47 | 94.47 | 94.42 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 95.98 | 95.98 | 95.98 | 95.95 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 97.28 | 97.28 | 97.39 | 97.41 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 95.21 | 95.21 | 95.21 | 95.61 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.64 | 96.42 | 95.88 | 96.42 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 97.82 | 97.82 | 97.82 | 97.85 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 98.04 | 98.04 | 98.04 | 98.05 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 98.04 | 98.04 | 98.04 | 98.03 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **98.37** | **98.37** | **98.37** | **98.36** |

improved by 0.04%. Stochastic Gradient Descent improved by 0.05%. Random Forest is improved by 2.07%. XGBoost is improved by 1.76%. 4-Layer Perceptron improved by 1.49%. 3-Layer Perceptron improved by 1.48%. After selection of

Table 4.16: Top six feature selection using information gain method (macro-average)

| Classifiers | Without Selection | | | | With Top-6 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 92.42 | 91.41 | 91.46 | 91.41 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 94.74 | 94.42 | 94.45 | 94.42 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 96.15 | 95.95 | 95.97 | 95.95 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 97.45 | 97.31 | 97.39 | 97.31 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 95.44 | 95.15 | 95.19 | 95.61 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 95.65 | 96.95 | 97.83 | 95.43 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 97.97 | 97.74 | 97.51 | 97.73 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 98.06 | 98.05 | 98.04 | 98.05 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 98.09 | 98.03 | 98.04 | 98.03 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **98.39** | **98.36** | **98.36** | **98.36** |

top-six features both with micro and macro average compare to the improvement K-Nearest Neighbor is improved more than among all other applied ml models. Compare to the performance the 3-Layer Perceptron generates best results among all other applied ml models.

### 4.3.3 Feature Selection using Gain Ratio Method

For selecting the potential features, we also used a statistical measure of gain ratio. We applied the gain ration method to the amazon reviewer's dataset. We evaluate our results of ML models using the top-five, top-four, top-six features selected by the gain ratio method.

#### 4.3.3.1 Top-Five Feature

The subset of top-five features provided by the gain ratio method is mentioned
.

1. Review count
2. Early time window
3. Average rating
4. Average sentiment
5. Verified purchase

Table 4.17 illustrates the results of without selection and with selection of top-five features. The impact of top-five features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 2.37%. K-Nearest Neighbor is improved by 3.59%. Support Vector Machine improved by 2.1%. Decision Tree improved by 3.18%. Logistic Regression is improved by 0.88%. Stochastic Gradient Descent improved by 1.35%. Random Forest is improved by 2.35%. XGBoost is improved by 1.48%. 4-Layer Perceptron improved by 0.65%. 3-Layer Perceptron improved by 1.6%. Table 4.18 illustrates

TABLE 4.17: Top-five feature selection using gain ratio method (micro-average)

| Classifiers | Without Selection | | | | With Top-5 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 92.52 | 92.52 | 92.52 | 92.41 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 94.79 | 94.79 | 94.79 | 94.76 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 96.21 | 96.21 | 96.21 | 96.17 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 97.28 | 97.39 | 97.28 | 97.31 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 96.21 | 96.21 | 96.21 | 96.61 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.85 | 95.44 | 96.31 | 96.74 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 98.15 | 98.04 | 97.93 | 98.06 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 97.72 | 97.72 | 97.72 | 97.72 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 97.17 | 97.17 | 97.17 | 97.18 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **98.47** | **98.47** | **98.47** | **98.48** |

the results of without selection and with selection of top-five features. The impact of top-five features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 3.14%. K-Nearest Neighbor is improved by 3.63%. Support Vector Machine improved by 2.16%. Decision Tree improved by 3.13%. Logistic Regression is improved by 1.04%. Stochastic Gradient Descent improved by 0.84%. Random Forest is improved by 2.27%. XGBoost is improved by 1.46%. 4-Layer Perceptron improved by 0.62%. 3-Layer Perceptron improved by 1.6%. The performance of the Gaussian

Table 4.18: Top-five feature selection using gain ratio method (macro-average)

| **Classifiers** | **Without Selection** | | | | **With Top-5** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 93.44 | 92.41 | 92.46 | 92.41 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 94.93 | 94.76 | 94.78 | 94.76 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 96.36 | 96.17 | 96.19 | 96.17 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 97.34 | 97.41 | 97.28 | 97.41 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 96.44 | 96.15 | 96.19 | 96.61 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 96.64 | 97.62 | 96.21 | 96.22 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 98.17 | 98.06 | 98.15 | 98.17 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 97.76 | 97.72 | 97.72 | 97.76 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 97.22 | 97.18 | 97.17 | 97.18 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **98.49** | **98.48** | **98.47** | **98.48** |

Naïve Bayes classifier is 93.44%, which is worse than other classifiers' performances because Naïve Bayes assumes an independent predictor. The performance of the 3-Layer Perceptron classifier performance is 98.49% seems to be the best classifier among all other classifiers' performances because this model has the ability of adaptive learning.

#### 4.3.3.2 Top-Four Feature

The subset of top-four features provided by the gain ratio method is mentioned below.

1. Review count
2. Early time window
3. Average rating
4. Verified purchase

Table 4.19 illustrates the results of without selection and with selection of top-four features. The impact of top-four features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 2.37%. K-Nearest Neighbor is improved by 4.57%. Support Vector Machine improved by 2.86%. Decision Tree improved by 1.51%. Logistic Regression is improved by 0.99%. Stochastic Gradient Descent improved by 0.7%. Random Forest is improved by 1.24%. XGBoost is improved by 1.4%. 4-Layer Perceptron decreased by 7.61%. 3-Layer Perceptron improved by 1.01%. Table 4.20

TABLE 4.19: Top four feature selection using gain ratio method (micro-average)

| **Classifiers** | **Without Selection** | | | | **With Top-4** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 92.52 | 92.52 | 92.52 | 92.41 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 95.77 | 95.77 | 95.77 | 95.75 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 96.96 | 96.96 | 96.96 | 96.93 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 95.61 | 95.62 | 95.51 | 95.39 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 96.31 | 96.31 | 96.31 | 96.27 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.21 | 96.61 | 96.21 | 95.97 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 97.04 | 97.04 | 97.03 | 97.04 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 97.64 | 97.64 | 97.64 | 97.63 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 88.91 | 88.91 | 88.91 | 89.13 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **97.88** | **97.87** | **97.87** | **97.89** |

illustrates the results of without selection and with selection of top-four features. The impact of top-four features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 3.15%. K-Nearest Neighbor is improved by 4.6%. Support Vector Machine improved by 2.86%. Decision Tree improved by 1.01%. Logistic Regression is improved by 1.09%. Stochastic Gradient Descent improved by 1.27%. Random Forest is improved by 1.2%. XGBoost is improved by 1.36%. 4-Layer Perceptron decreased by 12.56%. 3-Layer Perceptron improved by 0.98%. The performance of the Gaussian Naïve Bayes classifier is 93.45%, which is worse than other classifiers' performances because Naïve Bayes assumes an independent predictor. The performance of the 3-Layer Perceptron classifier performance is 97.87% seems to be the best classifier among all other classifiers' performances because this model has the ability of adaptive learning.

Table 4.20: Top four feature selection using gain ratio method (macro-average)

| Classifiers | Without Selection | | | | With Top-4 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 93.45 | 92.41 | 92.46 | 92.41 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 95.91 | 95.75 | 95.76 | 95.75 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 97.06 | 96.93 | 96.95 | 96.93 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 95.22 | 95.61 | 95.39 | 95.29 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 96.49 | 96.27 | 96.31 | 96.27 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 96.87 | 96.41 | 96.31 | 96.78 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 97.11 | 97.11 | 97.12 | 97.15 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 97.66 | 97.63 | 97.63 | 97.63 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 84.04 | 89.13 | 85.71 | 89.13 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **97.87** | **97.85** | **97.85** | **97.88** |

#### 4.3.3.3 Top-Six Feature

The subset of top-six features provided by the gain ratio method is mentioned below.

1. Review count
2. Early time window
3. Rating deviation
4. Average rating
5. Verified purchase
6. Average upvotes

Table 4.21 illustrates the results of without selection and with selection of top-six features. The impact of top-six features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 1.83%. K-Nearest Neighbor is improved by 3.81%. Support Vector Machine improved by 2.43%. Decision Tree improved by 0.97%. Logistic Regression is improved by 1.75%. Stochastic Gradient Descent improved by 0.6%. Random Forest is improved by 1.35%. XGBoost is decreased by 4.26%. 4-Layer Perceptron improved by 0.85%. 3-Layer Perceptron improved by 0.61%. Table 4.22 illustrates

Table 4.21: Top six feature selection using gain ratio method (micro-average)

| Classifiers | Without Selection | | | | With Top-6 | | | |
|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | AUC | Precision | Recall | F1 | AUC |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 91.98 | 91.98 | 91.98 | 91.85 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 95.01 | 95.01 | 95.01 | 95.01 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 96.53 | 96.53 | 96.53 | 96.48 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 95.07 | 95.18 | 95.51 | 95.17 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 97.07 | 97.07 | 97.07 | 97.03 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.11 | 95.88 | 96.96 | 96.52 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 97.15 | 97.26 | 97.82 | 97.37 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 91.98 | 91.98 | 91.98 | 91.85 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 97.37 | 97.37 | 97.37 | 97.37 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **97.48** | **97.48** | **97.48** | **97.47** |

the results of without selection and with selection of top-six features. The impact of top-six features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 2.76%. K-Nearest Neighbor is improved by 3.84%. Support Vector Machine improved by 2.53%. Decision Tree improved by 1.11%. Logistic Regression is improved by 1.83%. Stochastic Gradient Descent improved by 1.47%. Random Forest is improved by 1.28%. XGBoost is decreased by 3.24%. 4-Layer Perceptron improved by 0.81%. 3-Layer Perceptron improved by 0.63%. The performance of the

Table 4.22: Top six feature selection using gain ratio method (macro-average)

| Classifiers | Without Selection | | | | With Top-6 | | | |
|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 | AUC | Precision | Recall | F1 | AUC |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 93.06 | 91.85 | 91.91 | 91.85 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 95.14 | 95.01 | 95.01 | 95.01 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 96.73 | 96.48 | 96.52 | 96.48 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 95.32 | 95.39 | 95.28 | 95.31 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 97.23 | 97.03 | 97.06 | 97.03 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 97.07 | 96.07 | 97.06 | 97.38 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 97.18 | 97.15 | 97.15 | 97.26 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 93.06 | 91.85 | 91.91 | 91.85 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 97.41 | 97.37 | 97.37 | 97.37 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **97.52** | **97.47** | **97.48** | **97.47** |

Gaussian Naïve Bayes classifier is 93.06%, which is worse than other classifiers' performances because Naïve Bayes assumes an independent predictor. The performance of the 3-Layer Perceptron classifier performance is 97.52% seems to be the best classifier among all other classifiers' performances because this model has the ability of adaptive learning.

### 4.3.4 Feature Selection using Wrapper method

We also used the wrapper forward feature selection using the Amazon reviewer's dataset and the 3-layer perceptron as an ML model to select the potential features. The wrapper forward feature selection selected top-five features. We evaluate our results of ML models using the top-five features selected by the wrapper-forward feature selection. Below is the name of the five best optimal features selected by the wrapper method.

1. Review count
2. Average Rating
3. Average Upvotes
4. Rating Deviation
5. Verified purchase

Table 4.23 illustrates the results of without selection and with selection of top-five features selected by the wrapper method. The impact of top-five features on all applied ml models and metrics is clearly showing in this table along micro average. Gaussian Naïve Bayes is improved by 0.83%. K-Nearest Neighbor is improved by 1.96%. Support Vector Machine improved by 1.12%. Decision Tree improved by 3.18%. Logistic Regression is improved by 1.75%. Stochastic Gradient Descent improved by 0.92%. Random Forest is improved by 2.13%. XGBoost is improved by 1.69%. 4-Layer Perceptron improved by 1.85%. 3-Layer Perceptron improved by 2.29%. Table 4.24 illustrates the results of without selection and with selection of top-five features selected by the wrapper method. The impact of top-five features on all applied ml models and metrics is clearly showing in this table along macro average. Gaussian Naïve Bayes is improved by 1.22%. K-Nearest Neighbor is improved by 2.17%. Support Vector Machine improved by 1.22%. Decision Tree improved by 3.24%. Logistic Regression is improved by 1.82%. Stochastic Gradient Descent improved by 1.19%. Random Forest is improved by 1.74%. XGBoost

Table 4.23: Feature selection using wrapper method (micro-average)

| Classifiers | Without Selection | | | | With Top-5 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.15 | 90.15 | 90.15 | 90.21 | 90.98 | 90.98 | 90.98 | 91.06 |
| **KNN** | 91.21 | 91.21 | 91.21 | 91.21 | 93.16 | 93.16 | 93.16 | 93.11 |
| **SVM** | 94.11 | 94.11 | 94.11 | 94.11 | 95.22 | 95.22 | 95.22 | 95.18 |
| **Decision Tree** | 94.11 | 94.11 | 94.21 | 94.31 | 97.28 | 97.18 | 97.18 | 97.19 |
| **Logistic Regression** | 95.32 | 95.33 | 95.32 | 95.32 | 97.07 | 97.07 | 97.07 | 97.03 |
| **SGD** | 95.51 | 95.16 | 95.14 | 95.51 | 96.42 | 97.51 | 95.55 | 97.19 |
| **Random Forest** | 95.81 | 95.51 | 95.71 | 95.71 | 97.93 | 97.61 | 97.71 | 97.84 |
| **XGBoost** | 96.24 | 96.31 | 96.31 | 96.24 | 97.93 | 97.93 | 97.93 | 97.94 |
| **4-Layer Perceptron** | 96.52 | 96.52 | 96.52 | 96.52 | 98.37 | 98.37 | 98.37 | 98.37 |
| **3-Layer Perceptron** | **96.87** | **96.87** | **96.87** | **96.87** | **99.16** | **99.16** | **99.16** | **99.12** |

is improved by 1.65%. 4-Layer Perceptron improved by 1.8%. 3-Layer Perceptron improved by 2.27%.The performance of the Gaussian Naïve Bayes classifier is 91.52%, which is worse than other classifiers' performances because Naïve Bayes assumes an independent predictor. The performance of the 3-Layer Perceptron classifier performance is 99.16% seems to be the best classifier among all other classifiers' performances because this model has the ability of adaptive learning.

Table 4.24: Feature selection using wrapper method (macro-average)

| Classifiers | Without Selection | | | | With Top-5 | | | |
|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1** | **AUC** | **Precision** | **Recall** | **F1** | **AUC** |
| **Gaussian Naive Bayes** | 90.31 | 90.11 | 90.11 | 90.21 | 91.52 | 91.06 | 89.81 | 91.06 |
| **KNN** | 91.31 | 91.27 | 91.27 | 91.27 | 93.47 | 93.11 | 93.14 | 93.11 |
| **SVM** | 94.21 | 94.11 | 94.11 | 94.11 | 95.42 | 95.18 | 95.21 | 95.18 |
| **Decision Tree** | 94.21 | 94.11 | 94.11 | 94.31 | 97.45 | 97.09 | 97.28 | 97.19 |
| **Logistic Regression** | 95.41 | 95.23 | 95.23 | 95.32 | 97.22 | 97.03 | 97.06 | 97.03 |
| **SGD** | 95.61 | 95.21 | 95.21 | 95.51 | 96.79 | 96.83 | 96.31 | 96.84 |
| **Random Forest** | 95.91 | 95.81 | 95.81 | 95.71 | 97.64 | 97.73 | 97.71 | 97.84 |
| **XGBoost** | 96.31 | 96.32 | 96.32 | 96.24 | 97.95 | 97.94 | 97.93 | 97.94 |
| **4-Layer Perceptron** | 96.61 | 96.53 | 96.52 | 96.52 | 98.41 | 98.37 | 98.37 | 98.37 |
| **3-Layer Perceptron** | **96.89** | **96.89** | **96.89** | **96.89** | **99.16** | **99.16** | **99.16** | **99.12** |

**RQ 1** Among applied ML models, which model demonstrates the best performance for extremist reviewer groups identification?

**Answer:** As discussed above, the 3-layer perceptron model demonstrates the best performance, 99.16%, for extremist reviewer groups identification.

**RQ 2** What are the most contributing features after utilization of filter and wrapper-based feature selection methods in identifying extremist reviewer groups?

**Answer:** The most contributing features are Review count, Average Rating, Average Upvotes, Rating Deviation, Verified Purchased; these are selected by the wrapper-based feature selection method in identifying extremist reviewer groups.

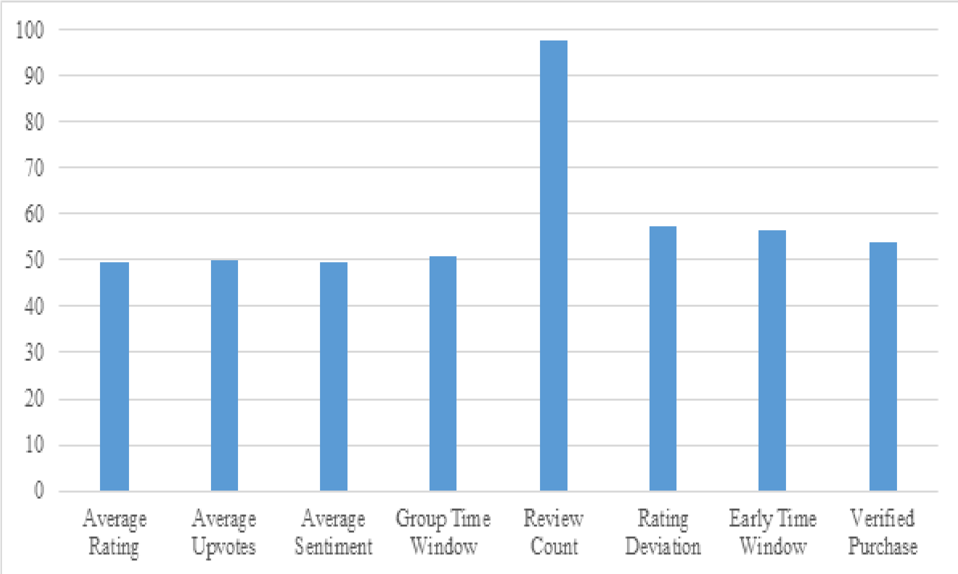


FIGURE 4.1: Individual feature performance using precision.

## 4.4 Experiment 3 Performance of Individual Features on Classification

The results, as mentioned earlier, show that the 3-layer perceptron is only a machine learning model that produces the best results in all experiments. In this section, the experiments are conducted by checking the impact of individual features using a 3-layer perceptron model. The impact of individual features is checked concerning the standalone evaluation metrics precision, recall, F1, accuracy using micro average. Figure 4.1 shows the individual feature performance using precision. As is discussed earlier, the only 3-layer perceptron machine learning model is implemented for this experiment. The following graphical representation indicates that the review count feature is outperformed among all other features with the precision of 97.50% and rating deviation as the second-best performance with the precision of 57.51%. The early time window as third-best performance with the precision of 56.66%, the verified purchase as the fourth-best performance with the precision of 53.70%. The group time window is the fifth-best performance with the precision of 50.92%, the average rating with the precision of 49.51%. The average sentiment also has 49.51% precision and the performance of average upvotes with the precision of 49.83%. This experiment indicated that the review count feature is outperformed among all other features. Figure 4.2 shows the individual feature

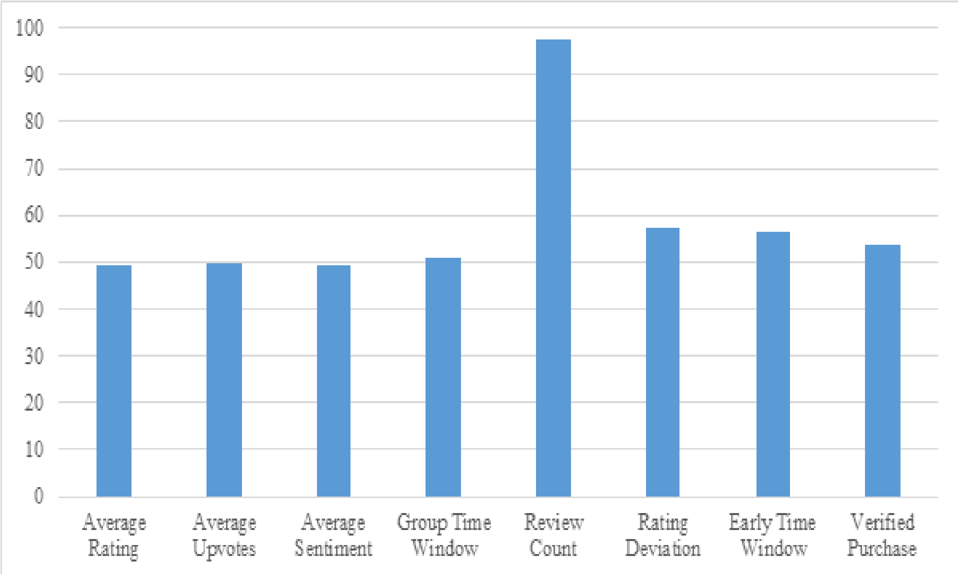


FIGURE 4.2: Individual feature performance using recall.

performance using recall. As is discussed earlier, the only 3-layer perceptron machine learning model is implemented for this experiment. The following graphical representation indicates that the review count feature is outperformed among all other features with the recall of 97.50% and rating deviation as the second-best performance with the recall of 57.51%. The early time window as third-best performance with the recall of 56.66%, the verified purchase

as the fourth-best performance with the recall of 53.70%. The group time window is the fifth-best performance with the recall of 50.92%, the average rating with the recall of 49.51%. The average sentiment also has 49.51% recall and the performance of average upvotes with the recall of 49.83%. This experiment indicated that the review count feature is outperformed among all other features. Figure 4.3 shows the individual feature performance using the F1-score. As is discussed earlier, the only 3-layer perceptron machine learning model is implemented for this experiment. The following graphical representation indicates that the review count feature is outperformed among all other features with the F1-score of 97.50% and rating deviation as the second-best performance with the F1-score of 57.51%. The early time window as third-best performance with the F1-score of 56.66%, the verified purchase as the fourth-best performance with the F1-score of 53.70%.

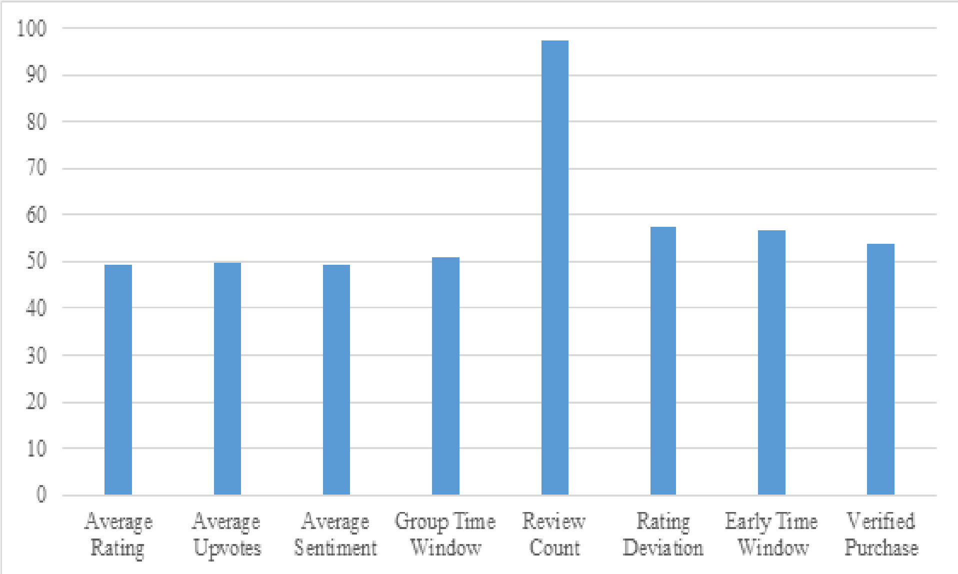


FIGURE 4.3: Individual feature performance using F1-score.

The group time window is the fifth-best performance with the F1-score of 50.92%, the average rating with the F1-score of 49.51%. The average sentiment also has 49.51% F1 score and the performance of average upvotes with the F1-score of 49.83%. This experiment indicated that the review count feature is outperformed among all other features.

Figure 4.4 shows the individual feature performance using AUC. As is discussed earlier, the only 3-layer perceptron machine learning model is implemented for this experiment.

The following graphical representation indicates that the review count feature is outperformed among all other features with the AUC of 97.49% and rating deviation as the second-best performance with the AUC of 57.91%. The early time window as third-best performance with the AUC of 56.67%, the verified purchase as the fourth-best performance with the AUC of 54.18%. The group time window is the fifth-best performance with an AUC of 51.48%, the average rating with the AUC of 50.00%. The average sentiment also has 50.00% AUC and the performance of average upvotes with the AUC of 50.21%. This experiment indicated that the review count feature is outperformed among all other features.

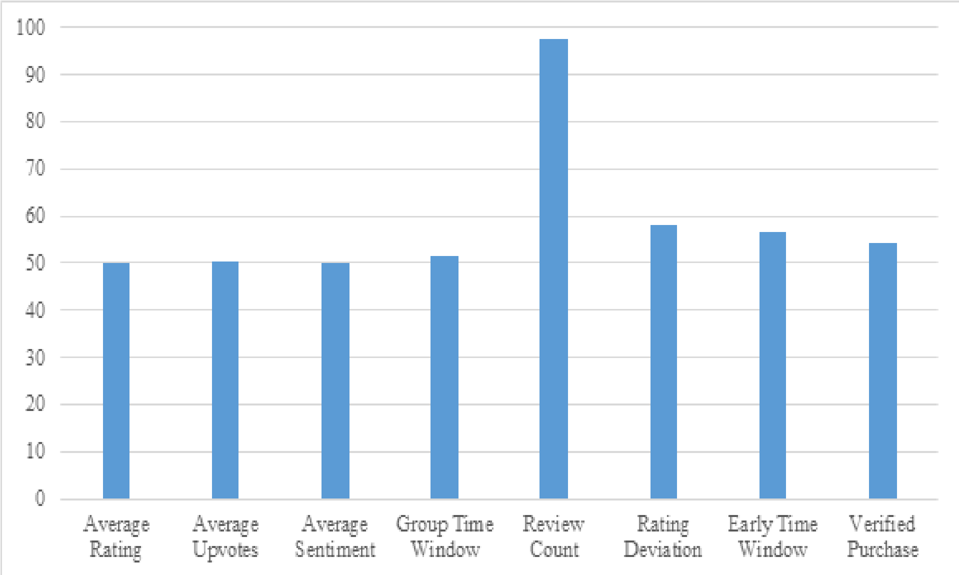


FIGURE 4.4: Individual feature performance using AUC.

# Chapter 5

# Conclusion and Future Work

The organization of this chapter is in two parts: the conclusion and the future work. In this chapter, a summary of our work done in the last four chapters is concluded. The future perspective of this research is in which factors there is a need to do more research in the form of future work is also mentioned in this chapter.

## 5.1 Conclusion

In this research, we discuss the problem of extremist reviewer groups' identification. The spammers post extreme positive or negative reviews and try to transform the reputation of the target brand as a whole. The reviewer groups write reviews that mainly target a specific brand, not only a product. The reviewer group's purpose is to promote or demote the reputation of the targeted brand. This study explores the link between the brand-level group actions and extremism in reviews that discover essential understandings about marketplace activities. The amazon extremist reviewer's dataset uses for the identification of the extremist reviewer groups. For dimensionality reduction, two different feature selection techniques, filter and wrapper methods, are adopted to select optimal features: three statistical measures, Pearson correlation, information gain, and gain ratio, used in the filter methods. The forward features selection technique is used in the wrapper method. The wrapper forward method technique outperformed and selected the best optimal features. For the classification, ten different ML models are used to classify the extreme and moderate reviewer groups. After the experimental setup, we compared all classification methods based on the classifier's performance. A 3-layer perceptron outperformed with an accuracy of 99.16%. Thus, the features selected by the forward wrapper selection are the best indicator for identifying extremist and moderate reviewer groups by reducing the dimensionality. This study helps to deliver buyer awareness in online marketplaces; they can differentiate between individual reviewers and extremist reviewers' groups without any extra struggle. Customers can identify extremist reviewer groups using amazon product reviews.

## 5.2 Future Work

This study can be additionally extended to different levels. The research on extremist reviewer groups identification is not much explored, especially on the brand level group spamming. More research is need to identifying extremism on the brand level. The researcher can enhance this research by constructing a dataset from amazon reviews and reviewers' history to highlight and introduce more robust features for identifying extremist reviewer groups in their studies. The researcher can also explore this topic by constructing the dataset from eBay or Yelp reviews and reviewers' history and identify extremist reviewer groups.

# Bibliography


[1] E. Gilbert and K. Karahalios, “Understanding deja reviewers,” in *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pp. 225–228, 2010.

[2] H. Ahmed, I. Traore, and S. Saad, “Detecting opinion spams and fake news using text classification,” *Security and Privacy*, vol. 1, no. 1, p. e9, 2018.

[3] T. Highfield and T. Leaver, “Instagrammatics and digital methods: Studying visual social media, from selfies and gifs to memes and emoji,” *Communication research and practice*, vol. 2, no. 1, pp. 47–62, 2016.

[4] S. Lee and J. Y. Choeh, “Exploring the determinants of and predicting the helpfulness of online user reviews using decision trees,” *Management Decision*, vol. 1, no. 1, 2017.

[5] R. Irfan, C. K. King, D. Grages, S. Ewen, S. U. Khan, S. A. Madani, J. Kolodziej, L. Wang, D. Chen, A. Rayes, *et al.*, “A survey on text mining in social networks,” *The Knowledge Engineering Review*, vol. 30, no. 2, pp. 157–170, 2015.

[6] M. Al Amin, N. Nowsin, I. Hossain, and T. Bala, “Impact of social media on consumer buying behaviour through online value proposition: A study on e-commerce business in bangladesh,” *BUFT Journal of Business & Economics*, vol. 1, pp. 209–228, 2020.

[7] P. Tsai, “Viacom international inc. v. youtube, inc. 718 f. supp. 2d 514 (sdny 2010),” *Intell. Prop. L. Bull.*, vol. 15, p. 103, 2010.

[8] P. Bajari and A. Hortaçsu, "The winner's curse, reserve prices, and endogenous entry: Empirical insights from ebay auctions," *RAND Journal of Economics*, vol. 15, pp. 329–355, 2003.

[9] D. Van Heijst, R. Potharst, and M. van Wezel, "A support system for predicting ebay end prices," *Decision Support Systems*, vol. 44, no. 4, pp. 970–982, 2008.

[10] D. Lucking-Reiley, D. Bryan, N. Prasad, and D. Reeves, "Pennies from ebay: The determinants of price in online auctions," *The journal of industrial economics*, vol. 55, no. 2, pp. 223–233, 2007.

[11] A. E. Roth and A. Ockenfels, "Last-minute bidding and the rules for ending second-price auctions: Evidence from ebay and amazon auctions on the internet," *American economic review*, vol. 92, no. 4, pp. 1093–1103, 2002.

[12] K. E. Reynolds, J. H. Gilkeson, and R. W. Niedrich, "The influence of seller strategy on the winning price in online auctions: A moderated mediation model," *Journal of Business Research*, vol. 62, no. 1, pp. 22–30, 2009.

[13] A. Neisari, L. Rueda, and S. Saad, "Spam review detection using self-organizing maps and convolutional neural networks," *Computers & Security*, vol. 106, p. 102274, 2021.

[14] P. Bajari and A. Hortaçsu, "Economic insights from internet auctions," *Journal of Economic Literature*, vol. 42, no. 2, pp. 457–486, 2004.

[15] Y.-E. Park and A. Allui, "ebay: Smart entry strategy into the korean market through m&a and its post-merger integration," *The Journal of Distribution Science*, vol. 17, no. 1, pp. 47–56, 2019.

[16] M. Anderson and J. Magruder, "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database," *The Economic Journal*, vol. 122, no. 563, pp. 957–989, 2012.

[17] K. M. Kuehn, "Brand local: Consumer evaluations as commodity activism on yelp. com," *Journal of Consumer Culture*, vol. 17, no. 2, pp. 205–224, 2017.

[18] R. Mukherjee and S. Banet-Weiser, *Commodity activism: Cultural resistance in neoliberal times*, vol. 15. NYU Press, 2012.

[19] C.-H. Kim, "Let them eat royal court cuisine! heritage politics of defining global hansik," *Gastronomica*, vol. 17, no. 3, pp. 4–14, 2017.

[20] S. M. Mudambi and D. Schuff, "Research note: What makes a helpful online review? a study of customer reviews on amazon. com," *MIS quarterly*, vol. 15, pp. 185–200, 2010.

[21] G. Yin, Q. Zhang, and Y. Li, "Effects of emotional valence and arousal on consumer perceptions of online review helpfulness," vol. 15, pp. 205–224, 2014.

[22] R. He and J. McAuley, "Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering," in *proceedings of the 25th international conference on world wide web*, vol. 19, pp. 507–517, 2016.

[23] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[24] A. C. Herman, "Dynamic status signaling: How foodies signal cosmopolitanism on yelp," *Poetics*, vol. 18, p. 101592, 2021.

[25] A. P. Haley, *Turning the tables: Restaurants and the rise of the American middle class, 1880-1920*, vol. 15. UNC Press Books, 2011.

[26] D. Jurafsky, V. Chahuneau, B. Routledge, and N. Smith, "Linguistic markers of status in food culture: Bourdieu's distinction in a menu corpus," *Journal of Cultural Analytics*, vol. 1, no. 1, p. 11064, 2016.

[27] A. Anemogiannis and V. Sachidananda, "Forecasting ratings and review counts for yelp businesses,"

[28] V. Taecharungroj and B. Mathayomchan, "Analysing tripadvisor reviews of tourist attractions in phuket, thailand," *Tourism Management*, vol. 75, pp. 550–568, 2019.

[29] A. Valdivia, M. V. Luzón, and F. Herrera, "Sentiment analysis in tripadvisor," *IEEE Intelligent Systems*, vol. 32, no. 4, pp. 72–77, 2017.

[30] X. Cheng, S. Fu, J. Sun, A. Bilgihan, and F. Okumus, "An investigation on online reviews in sharing economy driven hospitality platforms: A viewpoint of trust," *Tourism Management*, vol. 71, pp. 366–377, 2019.

[31] T. C. Chuang, J. S. Liu, L. Y. Lu, F.-M. Tseng, Y. Lee, and C.-T. Chang, "The main paths of etourism: trends of managing tourism through internet," *Asia Pacific Journal of Tourism Research*, vol. 22, no. 2, pp. 213–231, 2017.

[32] M. Schuckert, X. Liu, and R. Law, "Hospitality and tourism online reviews: Recent trends and future directions," *Journal of Travel & Tourism Marketing*, vol. 32, no. 5, pp. 608–621, 2015.

[33] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tourism management*, vol. 59, pp. 467–483, 2017.

[34] A. Reyes-Menendez, J. R. Saura, and J. G. Martinez-Navalon, "The impact of e-wom on hotels management reputation: exploring tripadvisor review credibility with the elm model," *IEEE Access*, vol. 7, pp. 68868–68877, 2019.

[35] S. Hussain, W. Ahmed, R. M. S. Jafar, A. Rabnawaz, and Y. Jianzhou, "ewom source credibility, perceived risk and food product customer's information adoption," *Computers in Human Behavior*, vol. 66, pp. 96–102, 2017.

[36] A. Y. Chua and S. Banerjee, "Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality," *Computers in Human Behavior*, vol. 54, pp. 547–554, 2016.

[37] K. Xie, L. Kwok, and W. Wang, "Monetizing managerial responses on tripadvisor: Performance implications across hotel classes," *Cornell Hospitality Quarterly*, vol. 58, no. 3, pp. 240–252, 2017.

[38] D. Mawhirter, S. Reinehr, W. Han, N. Fields, M. Claver, C. Holmes, J. McClurg, T. Liu, and B. Wu, "Dryadic: Flexible and fast graph pattern matching at scale," in *International Conference on Parallel Architectures and Compilation Techniques (PACT)*, vol. 58, p. 11064, 2021.

[39] J. McAuley, "Amazon product data," *Amazon Review. URL: http://jmcauley. ucsd. edu/data/amazon/[accessed 2020-03-10]*, vol. 8, no. 1, p. 11.

[40] M. Ott, C. Cardie, and J. T. Hancock, "Negative deceptive opinion spam," in *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*, vol. 14, pp. 497–501.

[41] V. Sandulescu and M. Ester, "Detecting singleton review spammers using semantic similarity," in *Proceedings of the 24th international conference on World Wide Web*, vol. 25, pp. 971–976.

[42] Y. Luo and X. Xu, "Predicting the helpfulness of online restaurant reviews using different machine learning algorithms: A case study of yelp," *Sustainability*, vol. 11, no. 19, p. 5254, 2019.

[43] B. Lutz, N. Pröllochs, and D. Neumann, "Understanding the role of two-sided argumentation in online consumer reviews: A language-based perspective," *arXiv preprint arXiv:1810.10942*, vol. 19, p. 525.

[44] M. K. Baowaly, Y.-P. Tu, and K.-T. Chen, "Predicting the helpfulness of game reviews: A case study on the steam store," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 4731–4742, 2019.

[45] Y. Yang, Y. Yan, M. Qiu, and F. Bao, "Semantic analysis and helpfulness prediction of text for online product reviews," in *Proceedings of the 53rd*

*Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, vol. 36, pp. 38–44, 2015.

[46] K. Kaushik, R. Mishra, N. P. Rana, and Y. K. Dwivedi, "Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on amazon. in," *Journal of retailing and Consumer Services*, vol. 45, pp. 21–32, 2018.

[47] K. Mouthami, K. N. Devi, and V. M. Bhaskaran, "Sentiment analysis and classification based on textual reviews," in *2013 international conference on Information communication and embedded systems (ICICES)*, vol. 32, pp. 271–276, IEEE, 2013.

[48] M. Malik and A. Hussain, "Helpfulness of product reviews as a function of discrete positive and negative emotions," *Computers in Human Behavior*, vol. 73, pp. 290–302, 2017.

[49] M. Chelliah and S. Sarkar, "Product recommendations enhanced with reviews," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 398–399, 2017.

[50] M. Malik and A. Hussain, "An analysis of review content and reviewer variables that contribute to review helpfulness," *Information Processing & Management*, vol. 54, no. 1, pp. 88–104, 2018.

[51] L. Chen and F. Wang, "Preference-based clustering reviews for augmenting e-commerce recommendation," *Knowledge-Based Systems*, vol. 50, pp. 44–59, 2013.

[52] J. Feuerbach, B. Loepp, C.-M. Barbu, and J. Ziegler, "Enhancing an interactive recommendation system with review-based information filtering.," in *IntRS@ RecSys*, vol. 15, pp. 2–9, 2017.

[53] A. Almahairi, K. Kastner, K. Cho, and A. Courville, "Learning distributed representations from reviews for collaborative filtering," in *Proceedings of the 9th ACM Conference on Recommender Systems*, vol. 6, pp. 147–154, 2015.

[54] S. Lee and J. Y. Choeh, "The determinants of helpfulness of online reviews," *Behaviour & Information Technology*, vol. 35, no. 10, pp. 853–863, 2016.

[55] Z. Liu and S. Park, "What makes a useful online review? implication for travel product websites," *Tourism management*, vol. 47, pp. 140–151, 2015.

[56] M. Malik and K. Iqbal, "Review helpfulness as a function of linguistic indicators," *International Journal of Computer Science and Network Security*, vol. 18, no. 1, pp. 234–240, 2018.

[57] T. L. Ngo-Ye, A. P. Sinha, and A. Sen, "Predicting the helpfulness of online reviews using a scripts-enriched text regression model," *Expert Systems with Applications*, vol. 71, pp. 98–110, 2017.

[58] S. Pecar, "Towards opinion summarization of customer reviews," in *Proceedings of ACL 2018, Student Research Workshop*, vol. 36, pp. 1–8, 2018.

[59] N. N. Ho-Dac, S. J. Carson, and W. L. Moore, "The effects of positive and negative online customer reviews: do brand strength and category maturity matter?," *Journal of marketing*, vol. 77, no. 6, pp. 37–53, 2013.

[60] G. Yin, L. Wei, W. Xu, and M. Chen, "Exploring heuristic cues for consumer perceptions of online reviews helpfulness: The case of yelp. com," vol. 7, no. 8, p. 37, 2014.

[61] Y. Wang, J. Wang, and T. Yao, "What makes a helpful online review? a meta-analysis of review characteristics," *Electronic Commerce Research*, vol. 19, no. 2, pp. 257–284, 2019.

[62] M. Salehan, M. Mousavizadeh, and M. Koohikamali, "A recommender system for online consumer reviews," vol. 7, no. 6, p. 53, 2015.

[63] R. Ullah, A. Zeb, and W. Kim, "The impact of emotions on the helpfulness of movie reviews," *Journal of applied research and technology*, vol. 13, no. 3, pp. 359–363, 2015.

[64] R. Ullah, N. Amblee, W. Kim, and H. Lee, "From valence to emotions: Exploring the distribution of emotions in online product reviews," *Decision Support Systems*, vol. 81, pp. 41–53, 2016.

[65] M. Malik and A. Hussain, "Exploring the influential reviewer, review and product determinants for review helpfulness," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 407–427, 2020.

[66] M. S. I. Malik, "Predicting users' review helpfulness: the role of significant review and reviewer characteristics," *Soft Computing*, vol. 3, pp. 1–16, 2020.

[67] M. Luca and G. Zervas, "Fake it till you make it: Reputation, competition, and yelp review fraud," *Management Science*, vol. 62, no. 12, pp. 3412–3427, 2016.

[68] E. D. Wahyuni and A. Djunaidy, "Fake review detection from a product review using modified method of iterative computation framework," in *MATEC web of conferences*, vol. 58, p. 03003, EDP Sciences, 2016.

[69] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?," in *Seventh international AAAI conference on weblogs and social media*, 2013.

[70] H. Li, Z. Chen, A. Mukherjee, B. Liu, and J. Shao, "Analyzing and detecting opinion spam on a large-scale dataset via temporal and spatial patterns," in *ninth international AAAI conference on web and social Media*, vol. 62, pp. 3412–3427.

[71] H. Li, G. Fei, S. Wang, B. Liu, W. Shao, A. Mukherjee, and J. Shao, "Bimodal distribution and co-bursting in review spam detection," in *Proceedings of the 26th International Conference on World Wide Web*, vol. 62, pp. 1063–1072, 2017.

[72] S. Kc and A. Mukherjee, "On the temporal dynamics of opinion spamming: Case studies on yelp," in *Proceedings of the 25th International Conference on World Wide Web*, vol. 2, pp. 369–379, 2016.

[73] A. Pease, C. Lo, A. Earnest, V. Kiriakova, D. Liew, and S. Zoungas, "The efficacy of technology in type 1 diabetes: a systematic review, network meta-analysis, and narrative synthesis," *Diabetes technology & therapeutics*, vol. 22, no. 5, pp. 411–421, 2020.

[74] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*, vol. 22, pp. 985–994, 2015.

[75] B. Hooi, N. Shah, A. Beutel, S. Günnemann, L. Akoglu, M. Kumar, D. Makhija, and C. Faloutsos, "Birdnest: Bayesian inference for ratings-fraud detection," in *Proceedings of the 2016 SIAM International Conference on Data Mining*, vol. 2, pp. 495–503, SIAM, 2016.

[76] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, "Spotting fake reviews via collective positive-unlabeled learning," in *2014 IEEE international conference on data mining*, vol. 28, pp. 899–904, IEEE, 2014.

[77] L. Akoglu, R. Chandy, and C. Faloutsos, "Opinion fraud detection in online reviews by network effects," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, p. 899, 2013.

[78] S. K. Chauhan, A. Goel, P. Goel, A. Chauhan, and M. K. Gurve, "Research on product review analysis and spam review detection," in *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, vol. 17, pp. 390–393, IEEE, 2017.

[79] J. Li, C. Cardie, and S. Li, "Topicspam: a topic-model based approach for spam detection," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 9, pp. 217–221, 2013.

[80] A. Mukherjee, B. Liu, and N. Glance, "Spotting fake reviewer groups in consumer reviews," in *Proceedings of the 21st international conference on World Wide Web*, vol. 36, pp. 191–200, 2012.

[81] D. T. Hoang, N. T. Nguyen, B. Collins, and D. Hwang, "Decision support system for solving reviewer assignment problem," *Cybernetics and Systems*, vol. 37, pp. 1–19, 2021.

[82] S. Kumar, B. Hooi, D. Makhija, M. Kumar, C. Faloutsos, and V. Subrahmanian, "Rev2: Fraudulent user prediction in rating platforms," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, vol. 28, pp. 333–341, 2018.

[83] I. Moseley, A. Roy, A. Deluty, and J. A. Brewer, "Evaluating the quality of smartphone apps for overeating, stress, and craving-related eating using the mobile application rating scale," *Current Addiction Reports*, vol. 7, no. 3, pp. 260–267, 2020.

[84] Y. Lu, L. Zhang, Y. Xiao, and Y. Li, "Simultaneously detecting fake reviews and review spammers using factor graph model," in *Proceedings of the 5th annual ACM web science conference*, vol. 36, pp. 225–233, 2013.

[85] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting burstiness in reviews for review spammer detection," in *Seventh international AAAI conference on weblogs and social media*, vol. 6, p. 225, 2013.

[86] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," *arXiv preprint arXiv:1107.4557*, vol. 66, p. 33, 2011.

[87] A. Molavi Kakhki, C. Kliman-Silver, and A. Mislove, "Iolaus: Securing online content rating systems," in *Proceedings of the 22nd international conference on World Wide Web*, vol. 5, pp. 919–930, 2013.

[88] C. Xu and J. Zhang, "Towards collusive fraud detection in online reviews," in *2015 IEEE international conference on data mining*, vol. 19, pp. 1051–1056, IEEE, 2015.

[89] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, "Detecting product review spammers using rating behaviors," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, vol. 6, pp. 939–948, 2010.

[90] C. Xu, J. Zhang, K. Chang, and C. Long, "Uncovering collusive spammers in chinese review websites," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, vol. 4, pp. 979–988, 2013.

[91] G. Wang, S. Xie, B. Liu, and S. Y. Philip, "Review graph based online store review spammer detection," in *2011 IEEE 11th international conference on data mining*, vol. 29, pp. 1242–1247, IEEE, 2011.

[92] A. Mukherjee, A. Kumar, B. Liu, J. Wang, M. Hsu, M. Castellanos, and R. Ghosh, "Spotting opinion spammers using behavioral footprints," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, vol. 66, pp. 632–640, 2013.

[93] G. Wang, S. Xie, B. Liu, and P. S. Yu, "Identify online store review spammers via social review graph," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 3, no. 4, pp. 1–21, 2012.

[94] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S.-M.-R. Beheshti, E. Bertino, and N. Foo, "Collusion detection in online rating systems," in *Asia-Pacific Web Conference*, vol. 8, pp. 196–207, Springer, 2013.

[95] Z. Wang, T. Hou, D. Song, Z. Li, and T. Kong, "Detecting review spammer groups via bipartite graph projection," *The Computer Journal*, vol. 59, no. 6, pp. 861–874, 2016.

[96] Z. Wang, S. Gu, X. Zhao, and X. Xu, "Graph-based review spammer group detection," *Knowledge and Information Systems*, vol. 55, no. 3, pp. 571–597, 2018.

[97] S. Dhawan, S. C. R. Gangireddy, S. Kumar, and T. Chakraborty, "Spotting collective behaviour of online frauds in customer reviews," *arXiv preprint arXiv:1905.13649*, vol. 55, no. 3, pp. 571–597, 2019.

[98] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting burstiness in reviews for review spammer detection," in *Seventh international AAAI conference on weblogs and social media*, vol. 15, p. 57, 2013.

[99] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?," in *Seventh international AAAI conference on weblogs and social media*, vol. 2, p. 365, 2013.

[100] V. Gupta, A. Aggarwal, and T. Chakraborty, "Detecting and characterizing extremist reviewer groups in online product reviews," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 3, pp. 741–750, 2020.

[101] C. Hung and S.-J. Chen, "Word sense disambiguation based sentiment lexicons for sentiment classification," *Knowledge-Based Systems*, vol. 110, pp. 224–232, 2016.

[102] G. Liu, J. Guo, Y. Zuo, J. Wu, and R.-y. Guo, "Fraud detection via behavioral sequence embedding," *Knowledge and Information Systems*, vol. 10, pp. 1–24, 2020.

[103] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Detection of opinion spam based on anomalous rating deviation," *Expert Systems with Applications*, vol. 42, no. 22, pp. 8650–8657, 2015.

[104] A. P. Jadhav, W. Hacke, D. W. Dippel, C. Z. Simonsen, V. Costalat, J. Fiehler, G. Thomalla, M. Bendszus, T. Andersson, H. P. Mattle, *et al.*, "Select wisely: the ethical challenge of defining large core with perfusion

in the early time window," *Journal of NeuroInterventional Surgery*, vol. 13, no. 6, pp. 497–499, 2021.

[105] D. Kim, K. Park, and J.-H. Ahn, "Do verified consumer reviews affect sales? an empirical analysis of mixed review systems in the film industry," vol. 10, pp. 1–24, 2018.

[106] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction," *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 56–70, 2020.

[107] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *Journal of King Saud University-Computer and Information Sciences*, no. 2, p. 70, 2019.

[108] Z. Zhao and H. Liu, "Searching for interacting features in subset selection," *Intelligent Data Analysis*, vol. 13, no. 2, pp. 207–228, 2009.

[109] N. J. Gogtay and U. M. Thatte, "Principles of correlation analysis," *Journal of the Association of Physicians of India*, vol. 65, no. 3, pp. 78–81, 2017.

[110] J. R. Quinlan, "Discovering rules by induction from large collections of examples," *Expert systems in the micro electronics age*, 1979.

[111] A. G. Karegowda, A. Manjunath, and M. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 271–277, 2010.

[112] R. Kohavi and G. H. John, "The wrapper approach," in *Feature extraction, construction and selection*, vol. 22, pp. 33–50, Springer, 1998.

[113] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egyptian Informatics Journal*, vol. 19, no. 3, pp. 179–189, 2018.

[114] N. El Aboudi and L. Benhlima, "Review on wrapper feature selection approaches," in *2016 International Conference on Engineering & MIS (ICEMIS)*, vol. 29, pp. 1–5, IEEE, 2016.

[115] D. Ververidis and C. Kotropoulos, "Sequential forward feature selection with low computational cost," in *2005 13th European Signal Processing Conference*, vol. 17, pp. 1–4, IEEE, 2005.

[116] T. Joachims, "Svmlight: Support vector machine," *SVM-Light Support Vector Machine http://svmlight. joachims. org/, University of Dortmund*, vol. 19, no. 4, p. 25, 1999.

[117] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, vol. 398. John Wiley & Sons, 2013.

[118] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.

[119] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, p. 130, 2015.

[120] M. Ontivero-Ortega, A. Lage-Castellanos, G. Valente, R. Goebel, and M. Valdes-Sosa, "Fast gaussian naïve bayes for searchlight classification analysis," *Neuroimage*, vol. 163, pp. 471–479, 2017.

[121] A. Sharma, "Guided stochastic gradient descent algorithm for inconsistent datasets," *Applied Soft Computing*, vol. 73, pp. 1068–1080, 2018.

[122] H. Nguyen *et al.*, "Application of the k-nearest neighbors algorithm for predicting blast-induced ground vibration in open-pit coal mines: a case study," *Journal of Mining and Earth Sciences*, vol. 61, no. 6, pp. 22–29, 2020.

[123] Y. Shi, X. Song, and G. Song, "Productivity prediction of a multilateral-well geothermal system based on a long short-term memory and multi-layer perceptron combinational neural network," *Applied Energy*, vol. 282, p. 116046, 2021.

[124] J. Abellán-García, “Four-layer perceptron approach for strength prediction of uhpc,” *Construction and Building Materials*, vol. 256, p. 119465, 2020.

[125] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, “Prediction of undrained shear strength using extreme gradient boosting and random forest based on bayesian optimization,” *Geoscience Frontiers*, vol. 12, no. 1, pp. 469–477, 2021.