CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Lexicon Based Impact Analysis of Adverbs & Adjectives for Sentiment Analysis Evaluation

by

Umar Naseer

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

in the

Faculty of Computing
Department of Computer Science

2020

I dedicate my dissertation work to my parents, supervisor, and all other teachers. A special feeling of gratitude is for my father, the most unswerving man, I ever know in this world

CERTIFICATE OF APPROVAL

Lexicon Based Impact Analysis of Adverbs & Adjectives for Sentiment Analysis Evaluation

by

Umar Naseer

(MCS173016)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Basit Shahzad	NUML, Islamabad
(b)	Internal Examiner	Dr. Abdul Basit Siddiqui	CUST, Islamabad
(c)	Supervisor	Dr. M. Tanvir Afzal	CUST, Islamabad

Dr. M. Tanvir Afzal
Thesis Supervisor
May, 2020

Dr. Nayyer Masood
Head
Dept. of Computer Science
May, 2020

Dr. Muhammad Abdul Qadir
Dean
Faculty of Computing
May, 2020

Author's Declaration

I, **Umar Naseer** hereby state that my MS thesis titled "**Lexicon Based Impact Analysis of Adverbs & Adjectives for Sentiment Analysis Evaluation**" 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.

(Umar Naseer)

Registration No: MCS173016

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Lexicon Based Impact Analysis of Adverbs & Adjectives for Sentiment Analysis Evaluation**" 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.

(Umar Naseer)

Registration No: MCS173016

Acknowledgements

Firstly, I am very grateful to Almighty Allah who gives me this opportunity and courage to complete my MS thesis.

Then, I am very thankful to my supervisor **Dr. M. Tanvir Afzal** who helps me out throughout my whole MS Thesis and his kind guidance and support helped me a lot in completing my MS Thesis. I am so grateful to him that he shares his pearl of wisdom and his expertise that greatly assisted me in my Research Thesis.

I am also thankful to all peoples who helped me in my thesis.

Finally, I would like to thanks my parents for their support, guidance, love and patience. Without their support, love and it will not be possible for me to complete my MS thesis.

(Umar Naseer)
Registration No: MCS173016

Abstract

Examination of emotions is now an important part of social interaction in order to determine the meaning behind a collection of words. It is actually meant to track posts and conversations on social media, and then to find out how participants respond.Sentiment analysis is normally measure by counting positive and negative polarity bearing words.

The state-of-the-art techniques use different polarity attributes to identify sentiments among positive, negative and neutral. Once we conduct sentiment analysis on some reviews, we essentially check for the feelings in reviews and pick the terms of sentiment within those reviews. Such words are probably negative, positive or neutral. These sentiments bearing phrase have some score in lexicons (SentiWordNet), which are available to the public and are well known. We have performed sentiment analysis for evaluation of polarity bearing features and lexicon resources.

To evaluate the linguistic features a detailed methodology steps have been followed on comprehensive dataset. Dataset is consisting of product reviews which are retrieved by a crawler from amazon. These reviews are from 11 different product categories of amazon, and this data set includes star ranking of reviews given by reviewers, which express their emotions about product. This ranking is used as a benchmark for evaluation of results.

In this research we are evaluating three machine learning algorithms using lexicons with linguistic features, which are Adjectives, Adverbs and their distinct forms. These features are not only evaluating individually, classes are made by different combination of feature for evaluation. All classes are evaluated with machine learning algorithms using lexicon score by comparing with benchmark. At the last, results from all combination of linguistic features, lexicons have been analyzed with standard metrics; Precision, Recall and F1-Measure, using machine learning algorithms. Results are showing that combinations of forms of adverbs and adjectives have produced better results than using adjectives and adverbs as whole.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
DT	Decision Tree
GBC	Gradient Boosting Classifier
ML	Machine Learning
POS	Part of Speech
RF	Random Forest

Chapter 1

Introduction

1.1 Background

Sentiment Analysis is a set of method of examining and processing data in order to identify a subjective response. Basically, its a method of implementation in computer software, that detect and measure a general mood, group‘s opinions and emotions in online social media marketing, branding, product positioning and enterprise information sources.

What the other people think has always been an important piece of information for those who want to make decision. People‘s judgments about the products quality for buying, are governed by the opinions of other people. The internet gives the ability to peoples to interact, share, and collaborate through social networks, online communities, blogs, wikis and other online collaborative media. With the growing popularity of websites like eBay, TripAdvisor, Amazon and Epinion.com where peoples activity become central to most web applications, and their opinions gave birth to a collective intelligence that is often more listened than experts point of view.

The rapid growth of review sites in recent years has made greater research efforts in sentiment analysis. Because sentiment data and analysis have great importance for marketers, it really has meant to lets marketers about his product and brand reputation. It is also important to them who provide customer services and whose ultimate goal is profit.

The goal of sentiment analysis is to determine a persons general view on a given subject. The opinions expressed in product reviews provide valuable information to consumers as well as major online retailers like Amazon. And when a company recently released a new product and they want to assess its reception among consumers who use social media. Thus, Sentiment Analysis process started getting traction in 2010 and is now booming to such an extent that it has been dignified as a field of study, not a mere marketing tool.

Turney (Turney, 2002) identifies sentiments based on the sentiment orientation of reviews, use lexicon-based approach to extract sentiments. The lexicon-based approach is based on the assumption that the contextual sentiment orientation is the sum of the sentiment orientation of each word or phrase.

The approach generally uses a dictionary of sentiment words to identify and determine sentiment orientation (positive, negative or neutral). The dictionary is called the sentiment lexicon. The approach of using sentiment words (the lexicon) to determine sentiment orientations is called the lexicon-based approach to sentiment analysis (Taboada, 2006) (Ding, 2008). This approach is efficient and can be a key factor to analysis web users sentiment. It is thus applicable to our task as well.

Different researcher used different classification features for sentiment analysis. Much of the lexicon-based research has focused on using adjectives as indicators of the sentiment orientation of text. Some other approaches have also included the use of adverb and verb. Approaches used for sentiment analysis have made good progress, but a lot of challenges in this field still exist.

The adverb and adjectives are the messiest, and maybe likewise the most fascinating grammatical feature (Conlon et. al., 1992). Previous studies in NLP, however, has managed with adjectives & adverbs, language specialists have accomplished noteworthy part on this word classification. We accept particular forms of grammatical features can fill in as essential pointers to distinguish the identity of sentiment shown in a text.

In this manner, the entirety of the sub-forms of adjectives and adverbs must be joined and completely be assessed. The Amazon is one of the biggest internet business commercial centers wherein clients purchase various items on the web and express their assessment as reviews. In this study, we consolidate all sub-kinds of adjectives and adverbs and structure their every combination to distinguish the part of various polarity bearing POS shown in reviews about items bought from Amazon.

The approach generally uses a dictionary of sentiment words to identify and determine sentiment orientation (positive, negative or neutral). The dictionary is called the sentiment lexicon. The approach of using sentiment words (the lexicon) to determine sentiment orientations is called the lexicon-based approach to sentiment analysis (Taboada, 2006) (Ding, 2008). This approach is efficient and can be a key factor to analysis web user's sentiment. It is thus applicable to our task as well.

Different researcher used different classification features for sentiment analysis. Much of the lexicon-based research has focused on using adjectives as indicators of the sentiment orientation of text. Some other approaches have also included the use of adverb and verb. Approaches used for sentiment analysis have made good progress, but a lot of challenges in this field still exist.

1.2 Problem Statement

The state-of-the-art approaches use different polarity features for the classification of sentiments into positive, negative and neutral. Different approaches used different polarity features such as: adjectives, adverbs etc. Similarly, well-known lexicons (e.g. SentiWordNet) have been used to identify polarity scores. From the critical analysis of the literature, this research has identified the following research gap that, none of the contemporary state-of-the-art researches have exploited all forms of adjectives and adverbs and comprehensively evaluated their different combinations to identify the best performing POS in analyzing sentiments of product reviews. Since the polarity bearing features, adverbs and adjectives are the modifiers of verbsand noun respectively, therefore, we assume that their collective contribution can play a significant role in discerning the sentiment expressed in reviews of products on Amazon.

1.3 Purpose

The purpose of this thesis is investigating the utility of linguistic features and lexicons for detecting the sentiment. Sentiment analysis utilizes text analytic techniques that automatically detect the polarity of text. Polarity describes whether a review expresses a positive or negative sentiment, but sentiment analysis is much more than just subtracting the number of "positive" terms from the number of "negative" terms in a review in order to produce a score. For the classification of sentiments into positive, negative and neutral, different polarity features and combinations of their forms are used. So, for evaluation that which combination forms of POS as feature set play an important role and which MLA provide high accuracy of classifying polarity terms, a comprehensive and comparative study on "Machine Learning algorithms as well as forms of linguistic features" are performed.

1.4 Scope

The objective of this thesis to performed sentiment analysis for evaluation of polarity bearing features and lexicon resources which offer great businesses to those who spend an enormous amount of time and money to understand their customer opinions about their products and services.

1.5 Research Objectives

Performing sentiment analysis on the product reviews, these reviews represent the user's opinions for specific products. Normal user of a product posts their reviews in the form of short text usually contains few sentences. These sentences are comprised of some important words. As we know that in English parts of speech, a word can have different meanings depending on the structure of the sentence. Identifying the parts of speech that can present true meaning of the sentence is a challenging task. The parts of speech are used to estimate the sentiments of the user comments. Adverbs are important part of any sentence and hence needed to be analyzed their role in determining the true sentiments of user. The different types of adverbs should be identified and analyzed for determining the sentiment of the sentence. So, in this work we will identify and extract the different types of adverbs from the user review datasets and then estimate their importance in automatic classification of reviews in three sentiment classes i.e. positive, negative or neutral. For automatic classification different classifiers have been used in research community. It is yet to determine which classifier is the best to classify the reviews into the classes based on adverb features. The classifier works on some feature set. In this study we explore some very important features in the content (text) of reviews.

These features are adverbs. We explore different types of verbs that can be used to classify the reviews into positive or negative classes.

In case of classifiers we are interested in determining the performance of different classifiers that are used by research community for classification. We investigate how these classifiers work on the extracted feature set and which of them achieve high performance.

1.6 Research Questions

The following research questions have been identified during literature review;

Research Question 1:

What is the impact of parts of speech (adverb and adjective) on sentiment analysis?

Research Question 2:

What is the impact of different combinations of forms of adverbs and adjectives on the classification?

Research Question 3:

What is the best machine learning classifier for product review classification?

1.7 Limitation

Thousands of the people express their sentiment by for, chat, blogs and on social media. Due to growing information on the Web, this is highly unstructured and scattered. It's very critical to extract the structured and relevant information to measuring the sentiment. Extracting of knowledge refers to a Natural language processing NLP analysis. Analyzing Natural language is very difficult task because sentiment is essentially subjective from person to person, and can be unreasonable.

Chapter 2

Literature Review

Sentiment Analysis utilizes text analytical techniques that detect the polarity of text. The Liu (Liu, 2012), who covers the entire field of sentiment analysis in his book. Sentiment analysis has been done by using two main approaches: Lexicon-based Approach and Machine Learning Approach

- **Lexicon-based Approach** involves calculating orientation for a document from the sentiment orientation of words or phrases.
- **Machine Learning Approach** that is a text classification approach involves building classifiers from labeled instances of texts or sentences.

As described earlier, two distinct approaches to sentiment analysis. Lukasz (Lukasz Augustyniak, 2014) compared these two approaches in the domain of movie reviews. The goal of the lexicon-based method is to assign the sentiment orientation to the text. Its sentiment orientation is obtained by lexicon using information extraction from database with priory known sentiment. The goal of machine learning method is similar to lexicon-based approach but obtained in different way: in this sentiment analysis is performed by classification. In this method first feature words are collected from analyzed text and then classify sentiment polarity using classification algorithm of document. According to research evaluation, lexicon based approach is easily outperformed by classification approach.

Where hybrid techniques are the synthesis of the two other approaches. Lexical based techniques normally focus on univocal words like happy, sad, afraid etc. While statistical approaches used automated strategies to judge emotions on machine learning based analysis, and hybrid techniques collectively using both strategies to provide the outcomes of reviews that are not specifically stated but have some connection to the product.

Some studies are especially connected with our methodology as in Fang et al., (Pang and Lee, 2008) proposed a procedure which is utilized to sort the extremity dependent on grammatical forms (POS). Another methodology displayed by Hu and Liu gave a rundown of various words (for example both Positive and Negative words). The proposed rundown of words comprises of 2006 positive and 4783 negative words separately. These words depend on surveys gave online which is utilized to remove the emotional data for this examination. In addition, in a proposed book arrangement system Pang and Lee proposed how evacuate target sentences by extricating the emotional ones. As essentially, we should concentrate on abstract substance.

Method proposed by Gann et al. utilized token-based methodology with respect to data at twitter as the author allocated specific scores to each token. These scores being utilized to investigate, if a specific sentiment is good, bad or neutral. Some different strategies are likewise helpful like Liu theme displaying in which the author proposed a procedure of consequently distinguishing the highlights or parts of an item. Mining the methodologies for clear investigation is especially likely full of feeling to hear a thought of individuals' point of view.

2.1 Machine Learning Approach

In machine learning approaches, Bo Pang (Bo Pang, 2002) used movie reviews for experiment. They consider the problem of classifying review by overall sentiments for determining whether a review is positive or negative. They used Naive Bayes classification, Maximum Entropy classification, and Support Vector Machines to test the results. They performed experiment to achieve accuracies on the sentiment classification problem by comparing to standard topic-based categorization.

In a previous work, Wilson and Moore (Wilson & Moore, 2011) examining the utility of linguistic features for detecting the sentiments and also evaluating the usefulness of existing lexical resources. They give attention to features which capture information about the informal and creative language used in micro blogging. They used a supervised approach to the problem, for building training data they use unigram or bigram and also use the features representing information and Part Of Speech. They performed this experiment for answering the questions that is:

- How useful is the sentiment lexicon developed for formal text on the short and informal tweets?
- How much gain do we get from the domain-specific features?

According to him, Part Of Speech may not be useful because of poor quality of tagger results in micro blogging domain.

Another approach in which Zhang (Zhang, 2011) performed experiment in the domain of online Cantonese-written restaurant reviews. They are using machine learning techniques Nave Bayes and SVM to automatically classify user reviews as positive or negative. To examining the effects of the classifiers, they are choosing six feature arrangements through n-gram presence/frequency: unigram, unigram-frequency, bigram, bigram-frequency, trigram, and trigram-frequency. According to Zhang, Naive Bayes classifier achieves better accuracy than SVM.

Another research that is based on supervised learning method through this Zhao (Zhao, 2014) performed sentiment classification of news comments. For the outcome of sentiment classification of news comments, combined the feature selection methods (DF, IG, CHI, and MI) and feature representation methods (Presence, TF, and TF-IDF). After getting the features trained these features to different learning methods and analyzing the experiment result.

Kobayashi et al. (2005), proposed utilization of AI strategies to identify the client sentiments about various items accessible on different web-based business sites. The essential thought behind the proposed system is to separate the set trait esteem sets from the writings (for example client survey about a particular item). A Japanese web records dataset is utilized for directing investigations. SVM is utilized as a classifier for characterizing client sentiments into various classes.

2.2 Lexicon Based Approach

Lexicon based approaches, in which researchers has focused on classification of linguistic features for sentiment analysis. In earlier work much of the researches have focused on adjectives or adjective phrases as indicators of sentiment orientation of text (Hatzivassiloglou, 1997), (Hu, 2004). They were contributed in developing dictionaries. Some of others have also included the use of adverbs (Benamara , 2007) with adjectives in case of determining strength and also identifying sentiments by using weak opinion bearing words verbs (Kim, 2004), the exclusive use of verbs (Sokolova, 2008), takes all three (adjectives, verbs, and adverbs) into account for analyze sentiment (Subrahmanian, 2008)and two-word phrases combinations that included, mostly, Adjective + Noun, Adverb + Noun, and Adverb + Verb (Turney, 2002).

Benamara (Benamara, 2007) were introducing a concept of Adverb-Adjective-Combination AAC-based sentiment analysis. In this technique linguistic analysis of adverbs of degree is used for sentiment analysis. For the classification of adverbs of degree they proposed a methodology in which they defined a set of general axioms for scoring adverbs.

There are classifications of adverb of degree into five categories.

- Adverbs of affirmation (AFF): these include adverbs which show affirmation such as absolutely, certainly, exactly, totally, and so on.

- Adverbs of doubt (DOUBT): these include adverbs which show doubt such as possibly, roughly, apparently, seemingly, and so on.
- Strong intensifying adverbs (STRONG): these include adverbs which words show strong intensity such as astronomically, exceedingly, extremely, and immensely and so on.
- Weak intensifying adverbs (WEAK): these include adverbs which show weak intensity such as barely, scarcely, weakly, slightly, and so on.
- Negation and Minimizers (NEG): these include adverbs which words actually have a negative effect on sentiment such as "hardly".

They also proposed three alternative algorithms to assign a score to an adverb-adjective combination. According to Benamara adjectives and adverbs are better than only considering adjectives.

Another approach in which Taboada (Taboada M. , 2011) present a methodology to extract sentiment from text. For extracting sentiment by lexicons resources, they created their own lexicon resource and then conducted experiments to evaluate their lexicon with other lexicons. They are showing that lexicon which created by them are superior in term of performance. They are created Semantic Orientation CALculator for calculating the sentiment by using lexicons. Annotated words are used from lexicons with their semantic orientation (polarity and strength) and also incorporate with intensification and negation. This Semantic Orientation CAL is applied to the polarity classification task, the process of assigning a positive or negative label to a text that captures the texts sentiment towards its main subject matter.

Godbole (2007), built up a conclusion examination computerized instrument called "Serendio". It utilizes dictionary base technique that was created utilizing Serendio scientific classification. The proposed framework begins the assumption investigation process with the preprocessing of the dataset, the preprocessing steps, for example, expulsion of stop words, stemming, piecing and hash label recognition is done consequently. The extremity insightful grouping of tweets is done based on relevant direction of words present in certain tweet. In (Zhou et al. 2014) the creators attempted to take care of the issue of shortened forms and incorrect spelling typically present in a large portion of the tweets and surveys information. Regularly vocabulary-based strategy doesn't give any instrument to deal with such sorts of issues. To defeat the issue of condensing and spelling botches the creators proposed a methodology that stretches out the general vocabulary to oblige area explicit words and shortened forms. The terms in the all-inclusive vocabulary are naturally chosen dependent on their common data with emojis.

2.3 Hybrid Approach

R. Xia et al., (2010) built up a hybrid technique for sentiment examination. The proposed system consolidates both vocabulary and AI based methodologies for estimation examination. POS alongside their related and word-connection highlights are chosen structure vocabulary and afterward AI classifiers (Xia, 2011) (for example Naive Bayes, ME and SVM) are applied to decide the assumption of words. So as to accomplish better grouping outcomes tests were performed on the dataset utilizing distinctive blend, for example, fixed, weighted, meta classifiers and gathering mix procedures.

Raj Ganesh et all. (2018), presented a hybrid technique for slant examination. The methodology is a feedback-based proposal framework that utilizations slant investigation. The proposal framework deals with the surveys of client that characterize these audits into various classes and prescribes books to clients as indicated by the characterization classes. The proposed framework utilizes a separating based machine calculations calculation.

To conquer the issue of Rumors in tweets the authors (Gayathri and Narayanan, 2017) exhibited a crossover approach that utilises both vocabulary and machine learning calculations in order to gossipy tidbits in tweets. A model is created and prepared that utilizes grammatical features (POS) to distinguish and name explicit tweets that contains Rumors.

Pang and Lee (2008), look at that a significant piece of data gathering conduct has consistently been to discover what others think. With the developing accessibility and ubiquity of concluding rich assets, for example, online survey destinations and individual sites, new chances and difficulties emerge as individuals currently can, and do, effectively use data advances to search out and comprehend the assessments of others. The unexpected ejection of movement in area of conclusion mining and estimation examination, which manages the computational treatment of feeling, assumption, and subjectivity in content, has subsequently happened at any rate to some extent as an immediate reaction to the flood of enthusiasm for new frameworks that manage sentiments as a top of the line object. This study covers strategies and approaches that guarantee to straightforwardly empower feeling focused data looking for frameworks. Author's emphasis is on techniques that try to address the new difficulties raised by slant mindful applications, when contrasted with those that are as of now present in progressively conventional reality-based examination. We remember material for rundown of evaluative content and on more extensive issues in regards to protection, control, and financial effect that the improvement of assessment situated data get to administrations offers ascend to.

Cambero (2016), examine that sentiment analysis is becoming exponentially because of the significance of the computerization in mining, extricating and preparing data so as to decide the general assessment of an individual. The issue that this postulation proposes to deliver is to figure out what strategies are increasingly appropriate to remove emotional impressions progressively from Twitter. For live applications, since the sentiments gathered from Twitter are constrained to certain measure of characters and it will occur in a continuous situation, this gives a fascinating situation; we will test utilizing both the Machine Learning Approach and the Lexicon-based Approach, and afterward consolidate them with an end goal to expand the precision. So as to test the constant factor, I will execute a web administration to gather ongoing input from Twitter continuously, which will be later handled and dissected for precision and continuous execution.

Cassinelli and Chen (2009), address the issue of arranging reports by generally speaking notion into two classes (for example positive or negative) and into various classes (for example one to -ve stars). We apply AI systems to order an informational index of film surveys. Specifically, we utilize a boosting calculation. For deciding the extremity of an audit, we found that the calculation has a translation like past work in conclusion investigation, yet it accomplishes better exactness in a more eficient way. Comparative outcomes can be seen when we apply the procedures to the multi-class classification task. Without expressly utilizing the connections of difierent names during preparing, our classifier can find the supposition afinity between classifications.

Go et al. (2009), presented a novel methodology for naturally ordering the conclusion of Twitter messages. These messages are classified as either positive or negative concerning a question term. This is helpful for customers who need to inquire about the conclusion of items before buy, or organizations that need to screen the open feeling of their brands. There is no past research on grouping assumption of messages on microblogging administrations like Twitter. We present the aftereffects of AI calculations for arranging the slant of Twitter messages utilizing far off supervision. Our preparation information comprises of Twitter messages with emojis, which are utilized as loud marks. This kind of preparing information is copiously accessible and can be gotten through mechanized methods. We show that AI calculations (Naive Bayes, Maximum Entropy, and SVM) have exactness above 80% when prepared with emoji information. This paper likewise depicts the preprocessing steps required so as to accomplish high exactness. The primary commitment of this paper is utilizing tweets with emojis for far off administered learning.

Cunhaa et al. (2015), online life progressions and the fast increment in volume and unpredictability of information produced by Internet administrations are turning out to be testing innovatively, yet additionally as far as application regions.

Execution and accessibility of information handling are basic factors that should be assessed since customary information preparing components may not give sufficient help. Apache Hadoop with Mahout is a framework to store and process information on every scale, including various devices to disseminate preparing. It has been viewed as a viable instrument right now utilized by both little and huge organizations and enterprises, similar to Google and Facebook, yet in addition open and private social insurance establishments. Given its ongoing development and the expanding unpredictability of the related mechanical issues, an assortment of all-encompassing structure arrangements has been advanced for every particular application. To exhibit its worth, we will show its highlights, favorable circumstances and applications on wellbeing Twitter information. We show that large wellbeing social information can create significant data, important both for normal clients and professionals. Starter aftereffects of information investigation on Twitter well being information utilizing Apache Hadoop show the capability of the blend of these innovations.

Dass (2016), examine individuals' recognitions about the potential risks related with the nearness of hereditarily altered living beings (GMOs) in nourishment items. Author figured research questions and theories dependent on parameters, including age, sex, condition of home, and more to examine these recognitions. Author directed an online across the country study over the United States and enlisted members from the overall public to comprehend their discernments about dangers for GMOs and GM nourishments.

Gamon (2004), show that it is conceivable to perform programmed sentiment classification in the boisterous area of client input information. Author show that by utilizing huge element vectors in mix with highlight decrease, author can prepare straight help vector machines that accomplish high characterization exactness on information that present grouping difficulties in any event, for a human annotator. Author additionally show that, shockingly, the option of profound etymological investigation highlights to a lot of surface level word n-gram highlights contributes reliably to arrangement exactness right now.

Zhang and Ye (2008), propose that it is an errand of developing enthusiasm for public activity and scholarly research, which is to discover significant and opinionate archives as indicated by a client's question. One of the key issues is the manner by which to consolidate a record's opinionate score (the positioning score of to what degree it is abstract or goal) and theme significance score. Current answers for report positioning in supposition recovery are by and large specially appointed direct blend, which is shy of hypothetical establishment and cautious investigation. Right now, center around vocabulary-based sentiment recovery. A tale age model that brings together point importance and sentiment age by a quadratic blend is proposed right now. With this model, the importance-based positioning fills in as the weighting variable of the dictionary-based conclusion positioning capacity, which is basically unique in relation to the mainstream heuristic straight mix draws near. The impact of various assumption word references is likewise talked about. Other than the bound together age model, another commitment is that our work exhibits that in the conclusion recovery task, a Bayesian way to deal with consolidating various positioning capacities is better than utilizing a straight blend. It is additionally relevant to other outcome re-positioning applications in comparable situation.

Godbole et al. (2007), inspect that newspapers and online journals express assessment of news substances (individuals, places, things) while giving an account of ongoing occasions. Author present a framework that relegates scores showing positive or negative conclusion to each particular substance in the content corpus. Godboles framework comprises of an assumption identification stage, which partners communicated assessments with each important substance, and an estimation conglomeration and scoring stage, which scores every element comparative with others in a similar class. At last,author assess the significance of our scoring strategies over enormous corpus of news and websites.

Nandi and Agrawal (2016), discover that take different choices in people's day by day life consistently while shopping, contributing cash, managing others, picking our preferred genius, big names, choosing our pastors and these choices are being made based on our previous existence encounters. It is said by somebody "who gains from his own missteps is savvy, however who gains from others botch is virtuoso". Sentiment Mining empowers us to utilize others past encounters for taking right choices. In supposition examination author take surveys from informal organizations and procedure those audits in such a way, that we can comprehend author's assessment, which will help us in making techniques in future. In decisions huge measure of endeavors, time and cash spent. At the hour of races interpersonal organizations gets overwhelmed with the online conversations about ideological groups and political famous people, heaps of disputable conversations and discussions are held over informal communities. Every one of these conversations give us chance to utilize it as an asset for study and investigation. In social examination author are utilizing twitter as our information source and are applying cross breed approach of supposition investigation; it consolidates both the Lexical Dictionary based methodology with the highlights of Support vector AI classifier. Lexical methodology takes a shot at pack of words where test dataset words are coordinated with preset lexicon words for mining while SVM is a managed taking in classifier which concentrates highlights from test information and based on those highlights, grouping is finished.

Bendarkheili et al. (2019), look at that human feelings and choices are mostly roused by others' convictions and encounters which give them a solid foundation of anything they've heard or found out about. In like manner, removing and realizing others' opinion of an extraordinary subject or item have gotten critical for various kinds of specialist organizations just as the shoppers. Feeling mining is the zone of examining client's notions through the accessible surveys on the web and has the advantageous utilization of directing clients for internet shopping. Lamentably, not very many investigations have been accomplished for opinion examination in Persian web-based shopping and the current works have numerous constraints in their exhibitions, esp. in separating between a feeble and a solid state of mind in nostalgic sentences. This paper proposes another dictionary-based assessment digging technique for Persian internet shopping which considers the impact of intensifier descriptive words in removing the specific assessment of an audit. It has been applied on a genuine dataset separated from Digikala1 and has accomplished promising outcomes contrasted with those of master evaluators.

TABLE 2.1: Summary of Some Previous Work

Data set	**Method**	**Features**	**Paper's**
0.5 million Amazon reviews	ML Approach	Verb, Adverb, Adjective	(Kausar, 2020)
Datasets have been claimed and given by Facebook	Cloud machine learning	Nouns, verbs, adjectives, adverbs.	(Arulmurugan et al., 2019)
Performance reports	ML approach	Noun phrases	(Chiranjeevi et al., 2018)
Reviews automatically from the technology sites	lexicon-based and learning-based methods	Adjectives, Nouns, Verbs, and Adverbs	Lee 2017
160,000 tweets	ML Approach and the Lexicon-based Approach	Social media content in real-time	(Angel Cambero, 2016)
Movie reviews 1,000 positive and 1,000 negative reviews	ML Techniques	Adverb, noun and verb	Cassinelli and Wei Chen, 2009

Chapter 3

Proposed Methodology

In the literature review chapter, some significant researches are presented on the subject of sentiment analysis. There are some attempts to study sentiment analysis that is most active research area. It has been dignified as a field of study with the explosion of user-generated content in social media, discussion chat, blogs and reviews. Some researchers have been found to use different features to mine sentiment such as noun, adjective, verbs, adverbs and their different combinations. This has also been found to be important in the role of adverbs and adjectives in classifying sentiment. Few studies that measured one or two types of adverbs. The exploration of adverbs and adjectives of all kinds remains an open research question which has been answered in this thesis. This thesis explores some possible combinations of adverbs and adjectives to measure their impact on the classification of sentiments.

Examination of the sentiments has two primary approaches: Machine learning and lexicon-based. Machine based learning is supervised technique, that uses classification technique to classify text, because the classification of the data involves initial training. Lexicon-based method is unsupervised learning which uses lexicons of sentiment with words of opinion because it does not require any training to classify text. In this study, we have continued with a Hybrid approach for analysis of sentiments, as the aim of the thesis is to examine the sentiments of reviews for types evaluation of adverbs and adjectives and their best combinations towards sentiment analysis.

A detailed dataset containing reviews of different items on amazon is considered for research to evaluate the influence of adverbs, adjectives, and their various forms. The data set is pre-processed to remove the redundancies and then POS Tagger is used to apply a part of the speech tagging.

The SentiWordNet library (Baccianella et al., 2009) is used for various combinations that are processed after obtaining the adverbs and adjectives forms to obtain their ratings. In addition, reviews are categorized according to their ratings into five different classifications, such as strong negative, negative, neutral, positive and strong positives.

3.1 Data Collection

To analyze the impact of adverbs and adjectives and their different forms, we have formulated an in-depth methodology. All steps of data collection and the processing methodology are described as architecture diagram shown Figure 3.1.

Dataset for experiment is based on product reviews of Amazon. This dataset is crawled from Amazon generated by a crawler; this is about 11 different categories of products distinct in nature e.g. Cell Phones and its Accessories and downloaded in year 2020. Dataset is consisting of about 21.47 million reviews which also contain a review summary and ranking scale. Ranking of customer reviews is based on the sentiment of customers. For this, other customer gives their views that a customer review is helpful or not by voting. The overall helpfulness of all their reviews decides the rank of a reviewer, a result in the number of reviews they've received. A ranking scale contains five values, ranging from worst to excellent. These values are also referred to as 1 star to 5 star. Ranking scale is used as a benchmark for the evaluation of experiment results. Before starting sentiment analysis for evaluation, we need to do a linguistic analysis.

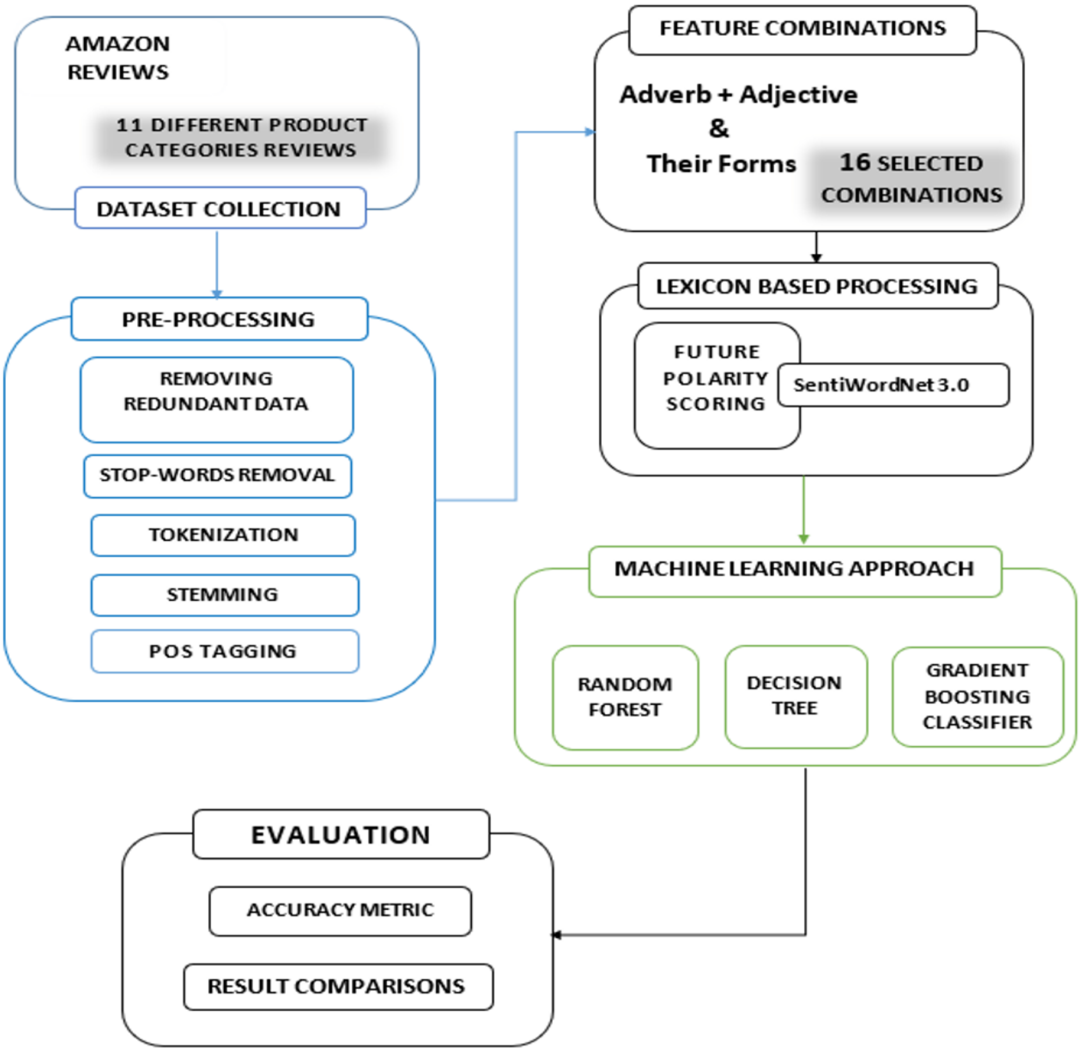

Figure 3.1: Research Methodology

3.2 Pre-Processing

The initial steps involve identification of a sentence boundary, like where the sentence ends. Thereafter, the text of a sentence is tagged and tokenized into single words. The words containing noise or stopwords, new lines tags ,white spaces, html tags, emoticons and special symbols have been executed.

3.2.1 Stopwords Removal

Stop words are usually extra words that aren't necessary for categorizing sentiment polarity. In our data set we remove all stop words which are beneficial for better accuracy.

3.2.2 Word Tokenization

That word is allocated with a token, and the word score from the SentiWordNet library is obtained based on that token.

3.2.3 Stemming

Porter's algorithm is designed to stem English-speaking texts, which was one of the most common stemming methods proposed in 1980. Porter Stemmers uses simple algorithms to determine in which order and when to use repair strategies as shown in Table 1.

Table 3.1: Stemming Strategies

Input	**Strip -ed**	**Affix Repair**
hoped	hop	hope (add -e if word is short)
hopped	hopp	hop (delete one if doubled)
hoping	hop	hope (add -e if word is short)

Stemming algorithm such as Porter Stemmer has been used to provide ways for search terms to find morphological variants.

1: Gets rid of plurals and -ed or -ing suffixes.

2: Turns terminal y to i when there is another vowel in the stem.

3: Maps double suffixes to single ones: -ization, -ational, etc.

4: Deals with suffixes, -full, -ness etc.

5: Takes off -ant, -ence, etc.

6: Removes a final -e.

3.2.4 POS Tagging

Part-Of-Speech (POS) is used to removal of ambiguity to make sense by making something clear and POS Tagging is the method of assigning a word to a specific part of the speech in a text. That is used to guide for the selection of linguistic features. With POS tags, linguistic features are easily identified. These features comprise of adjective, adverb and their distinct forms. These linguistic features are usually used as sentiment indicators. After the data collection, linguistic analysis is applied on the dataset in which POS tagger tags the words as corresponding to a particular POS. Linguistic feature (polarity words) are extracted from POS tagged words, which are used in Machine Learning based approach. Part-Of-Speech tags have been a popular choice for researchers to detect features that capture the sentiment.

The sentences of the reviews contain various parts of speech such as noun, adjective; verb and adverb. All POS are defined using Natural Language Tool Kit (NLTK) and labelled. As discussed earlier, the main ingredient of this study is adjective, adverbs and their forms have been omitted from the data set and therefore POS other than oriented ones. NLTK tagged adjectives and adverbs in the following forms:

Adjective (JJ): Its main syntactic role is to modification of a noun. For instance, a red hat.

Comparative adjectives (JJR): They compare differences between two objects modified by them. For example, larger, smaller, higher etc.

Superlative adjectives : They are used to compare multiple nouns or to describe higher level of comparison between entities. For example, she is the prettiest among other queens.

Adverb (RB): This modifies the verb using another adverb e.g. very, silently, much etc.

Superlative adverbs (RRS): It modifies general adverb, for example, best, longest and easiest etc.

Comparative adverbs (RBR): It modifies verbs along another adverb with comparison e.g. more, less and few etc.

Following is an example, which will help the readers to understand how adverbs and adjectives can be key features of text and how can they provide clue about whole context of review. Following is one of the reviews from data set.

Review:

"I was smacked to realize that the new office is renewable **annually** and as I **only** require basic office and would not gain from the upgraded office programmes. I looked **around** and found the 2013 which will do me for as long as my computer is alive and come to think of me as well. Compare but be sensible for home use - do you need the additional features and are you willing to pay **annually** for them. I wasn't and I am thrilled. I **already** have to make annual payments on other software I need such **as** protection but the annual cost **soon** mount up. Use your common sense. I do**n't** think many users of office know **how** to get the **most** out of it as home users unless they are studying or earning a living from the programme or using it **professionally**."

In above example, the adjectives are underlined and adverbs are in bold fonts. Now the question is that how the highlighted adjectives and adverbs can tackle the context of a review. The different forms of adjectives like new, renewable, upgraded, alive, additional, sensible and adverbs like annually, only, as-well, already, professionally are some general adjectives (JJ), superlative adjectives (JJS), general adverbs (RB) and superlative adverb (RBS). These words hold a particular meaning that differentiates them from routine words, therefore, their polarity score could be a great source of determining sentiment class of a review.

3.3 Lexicon-based Score Determination

Classification is achieved in lexicon-based methods by comparing thecharacteristics of a given text to sentiment lexicons whose sentiment values are earlier defined for their use.

To identify the sentiment of the review, acquired linguistic feature which are search by linguistical analysis are look up for polarity score in sentiment lexicons. Sentiment lexicons contain list of words used to express subjective feeling of people. Sentiment lexicons also contain sentiment score assign to the linguistic feature words. That is explaining how the terms used in the dictionary are positive, negative and objective.

To obtain polarity scores forSelected Part of Speech, we usedSentiWordNet 3.0 library. The SentiWordNet is a lexical tool that is explicitly designed for applications related to mining opinion or analysis of feelings. The SentiWordNet 3.0 is an update to SentiWordNet 1.0. The lexical resource is available freely. SentiWordNet assigns three sentiment numerical scores to POS. These scores are assigned on the basis of positivity, negativity and objectivity. The employed data set contains both positive words and negative words. Each POS in the reviews is assigned either negative or positive score. The positive values of the adjectives mentioned adjective forms are grouped and their average score is measured and all idiomatic phrases of the same term are also included in dictionary. Similarly, the negative results are merged for adjective forms, and their average score is calculated. These steps are applied on adverbs and its types in a same manner.

3.3.1 Sentence Scoring

The sentence score is calculated using the score of single words found in that particular sentence. The sentence score is calculated using equation (3.1). (Kausar et al., 2020)

$$senScore(s)\frac{1}{n} = \sum_{i=0}^{n} Pi \qquad \text{... (3.1)}$$

Where,

- senScore(S) is sentence graded.
- n Is the sum of the words.
- (Pi) polarity where, i is the word boundary

Below is an example explaining how to measure the sentence level ratings.

Sentence 1:

The Microsoft version 2013 office is very good and many things are enhanced especially the new style.

Explanation:

The words very and especially are general adverbs and words good, many and new are general adjectives. Such adjectives and adverbs earn SentiWordNet library polarity scores and average score is measured against this expression.

Sentence Score:

The total score of the above term will be positive because “SentiWordNet” will return positive score for the sentence.

3.3.2 Scoring of Review

The review score is calculated by considering the scores of all the sentences of the review. The review score is calculated using equation (3.2). (Kausar et al., 2020)

$$revScore(R) = \frac{1}{n}\sum_{i=0}^{n}(Si) \qquad \text{.... (3.2)}$$

Where,

- revScore (R) Presents review score.
- N is the sum of the sentences in a review.
- (Si) sentence present in a review where the sentence limit is i.

To classify reviews using adjectives, adverbs and their different forms, the reviews are passed to NLP tagger for tagging the said POS types. The forms of adverbs and adjectives are combined and their score is computed using SentiWordNet. First of all, sentence level score is computed and then review level scores is computed using SentiWordNet 3.0. The final obtained score the final scores is harnessed to classify review into any of the 5-star rating classes.

3.3.3 Star Rating of Review

On Amazon platform, each user provides with an option to give star rating to the product available on amazon. That star rating determines inclination of a user towards particular item based on his/her experience. The amazon even has star rates every time customer shares their opinions. The first phase to evaluate the review's 5-star rating is to figure out which frequencies are from the maximum to the minimum. In this regard, different researchers have distributed these ranges and assigned a particular description to each star rating (Pappas & Popescu-Belis, 2014; Lak & Turetken, 2014; Boon et al., 2012; Jang Jong 2011; Lee & Pang, 2005). The highly positive and highly negative scales, according to them, range from -1 to 1 respectively (Kincl et al. 2013; Mai et al. 2016; Zhang et al., 2010).

3.3.4 Feature Combinations

Three separate adverbial forms and three different kinds of adjectives were obtained in this analysis. Then, these forms are merged to measure the score for polarity. For combinations making the, the formula 2^{n-1} is used. To understand the behaviors, three distinct adverb forms are combined. Because three types of adjectives and adverbs exist each, therefore, according to the formula, 64 different combinations will be formed. But as from previous researches the distinct forms perform more effectively in pairs than single existence (Haider et al., 2018) (Das & Balabantaray, 2014). Since, pairs are used of three forms of adjectives and adverbs each, therefore, according to the formula, 16 different combinations will be formed. Those combinations are created as:

Table 3.2: Feature Sets

	Forms	**Symbols**
Set 1	Adverb	RB
	Comparative adverbs	RBR
	Superlative adverbs	RBS
Set 2	Adjective	JJ
	Comparative adjective	JJR
	Superlative adjectives	JJS

Table 3.3: Combination categories

No.	**Combinations Categories**
1	(RB+RBR+RBS+JJ+JJR+JJS)
2	(RB+RBR+RBS+JJ+JJR)
3	(RB+RBR+RBS+JJR+JJS)
4	(RBR+RBS+JJ+JJR)
5	(RBR+RBS+JJR+JJS)
6	(RB+RBS+JJ+JJR)
7	(RB+RBS+JJR+JJS)
8	(RB+RBR+JJ+JJR)
9	(RB+RBR+JJR+JJS)
10	(RB+RBR+RBS+JJ+JJS)
11	(RBR+RBS+JJ+JJS)
12	(RB+RBS+JJ+JJS)
13	(RB+RBR+JJ+JJS)
14	(RB+RBR+JJ+JJR+JJS)
15	(RB+RBS+JJ+JJR+JJS)
16	(RBR+RBS+JJ+JJR+JJS)

According to (Benamara , 2007) adjective and adverb performs better than adverb alone. So in our combinations we are ignoring the combinations have adjective or adverbs alone.

So, in the methodology explained in this chapter, we have explained a method to analyze individual and collective impact of different types of adjectives and adverbs. The polarity score of a review will specify the class of a review from any of the five classes explained above.

3.3.5 Classification

Each reviews a variable sequence of words and the sentiment of each review must be classified into above mentioned star rating classes (Ali et al., 2017) (Kim et al., 2016). The Large Amazon Review Dataset contains above 65% highly-polar reviews (good or bad) for training and testing. The problem is to determine whether a given review has a different sentiment depending on polarity of adverb of adjective features. Various methodologies have been practiced by different studies over the years starting from tree-based classifier to neural network-based approaches. Decision Tree, Random Forest, and Gradient Boosting Classifier have been chosen to determine the accuracy of results.

3.3.6 Score Aggregation

In score aggregating process, aggregate the scores of linguistic features (polarity bearing) words. For this purpose, we have made the classes through linguistic features which are considered in experiment (as mentioned in table 4) and with the combinations of features. There are sixteen classes and, in some classes, generic equation has been used as shown below:

3.4 Evaluation

In proposed approach, the purpose of evaluation is investigating the effectiveness of linguistic features as well as machine learning algorithms for detecting sentiment. Although all the existing techniques are effective but after critical reviewing the previous work, it has been concluded that all of the polarities bearing words and their forms and available machine learning algorithms have not been studied and compared on single comprehensive dataset. So, no one can say which one is more efficient than other. Evaluating all the techniques by using all combinations by aggregating scores and performing test on datasets so that we can analyze which combination and algorithm is more effective in sentiment analysis. For this purpose, classes for different linguistic feature with numerous possible combinations has been used for aggregating scores. In some classes generic equation as mentioned in previous section is also used for aggregating the score. This equation performs well in one of the scenarios which are helpful in research evaluation. At the last step of evaluation, aggregated scores have been compared with benchmark of each class individually. In this comparison if results are matched or unmatched then the results are also categorized in five ranks which are 1,2,3,4 and 5 that is equal to Negative, Weak Negative, Neutral, Weak Positive and Positive respectively. These results are evaluated by using standard measures to investigating each class against benchmark. These standard matrixes are Precision, Recall and F1-measure. Equation 3.3, 3.4 and 3.5 briefly explain these concepts.

Precision:

Percentage of selected items those are correct.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive + False Positive}} \quad \text{.... (3.3)}$$

Recall:

Percentage of correct items those are select.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad \text{.... (3.4)}$$

F1-Measure:

The harmonic mean of precision and recall is one factor that blends precision and recall.

$$\text{F1-Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{.... (3.5)}$$

True Negative: case was negative, and negatively predicted. True Positive: case has been positive and optimistic has been predicted. False negative: an positive case but negative predicted. False Positive: case was negative, but positive prediction.

This basic matrix is determined using Algorithms for Machine Learning. There are three different algorithms for machine learning which follow:

1. Random Forest
2. Decision Tree
3. Gradient Boost Classifier

By using Precision, Recall and F-measure we have done evaluation and comparison, detailed results have been discussed in chapter 4.

3.5 Tools

The following tools & techniques were used for implementing and evaluating the proposed methodology

- Natural language tool kit (NLTK) is used for tagging the reviews.

- Porter Stemmer is used to obtain the root of the term.
- Java & Python - are used for programming
- XAMP -Database is used for storing the data.
- Excel -MS Office is used for calculation and graphs
- SentiWordNet 3.0 -Lexicon is used for retrieve the score.

Chapter 4

Results and Analysis

This chapter describes the results and critical analysis extracted by proceeding the procedure discussed in the earlier chapter from research methodology. The purpose of this thesis is focusing on the extraction of sentiment of the reviews for investigating the utility of linguistic features as well as lexicon resources. In evaluation of polarity bearing features and machine learning algorithms for sentiment analysis, proposed approach is compared and evaluated against benchmark by the machine learning algorithms and some top results of this research are highlighted in this chapter.(Complete results are presented in Appendix A, Table A.1)

4.1 Analysis of Dataset

As we mentioned in previous chapter for the evaluation of linguistic feature as well as machine learning algorithms, amazon product reviews are used for sentiment analysis. In sentiment analysis, we used and investigate all forms of adverbs and adjectives selected possible combinations. In experiment we are precede with steps as mentioned in research methodology.

Experiment was performed on 21,470,250 product reviews from 11 different product categories, these categories have 2,110,984 distinct items which have 5 type reviews

Negative, Weak Negative, Neutral, Weak Positive and Positive. These reviews types have identified by voting of customers. Customers rank the product by giving the remarks about product and other customer who get to know about product features, price and others details, they read the reviews and give the feedback by voting. According to feedback, product stars are getting raise, ranking start from worst to best means from 1 to 5, 1 for negative, 2 for weak negative, 3 for neutral, 4 for weak positive and 5 for positive. These ranking are used as a benchmark for the evaluation of the results.

After getting the reviews linguistical analysis has on dataset because we require polarity bearing feature words for further process. We have applied part-of-speech tagger and extracted the required feature words which words have the sentiment of the overall contextual review for sentiment analysis.

TABLE 4.1: Dataset Details

No.	**Product Categories**	**No. of Distinct Products**	**No. of Distinct Reviews**
1	Amazon_Instant_Video	30648	583933
2	Apps_for_Android	61551	2638173
3	Automotive	331090	1373768
4	Beauty	259204	2023070
5	Cell_Phones_and_Accessories	346793	3447249
6	Digital_Music	279899	836006
7	Health_and_Personal_Care	263032	2982326
8	Movies_and_TV	208321	4607047
9	Musical_Instruments	84901	500176
10	Office_Products	134838	1243186
11	Pet_Supplies	110707	1235316
	Total:	**2110984**	**21470250**

4.2 Acquiring Score

By using these feature terms of the reviews, we are performed same experiment with SentiWordNet 3.0. SentiWordNet is limited to terms domain and do not take into account the relation between terms so it is just syntactic based (Erik

Cambria, 2010). Words after pre- processing and tokenization passed to scoring section and polarity from SentiWordNet is acquired.

4.3 Score Aggregation using formula

As discussed in chapter 3, for aggregating the score, we have made 16 classes to evaluate individual feature and the lexicons. In this section, results of lexicon are shown and how generic formula is applied on values. For this we are taking following review as an example:

Review:
"Samsung" new features are very nice, but speed is new phone's best trick.

Review Score:

TABLE 4.2: Score of polarity features

Lexicon Feature	**Adjective**	**Adverb**
Polarity Words	New, Best, Nice	Very
SentiWordNet	0.5,0.75,0.75	0.45

SentiWordNet:
Adjective average is 0.67 while adverb got 0.45 polarity. According to formula the as whole average of adjectives and adverb is 0.61. Which represents that the sentence is positive in nature.

After retrieving the score of lexicons, compared the results with benchmark for evaluation to results, by the machine learning algorithms. These steps performed on each class of combinations individually.

4.4 Results of Product Categories

To evaluating the research strategies, precision and recall are used as basic measures. In this section we have results of all 11 products and all our combination categories through this we are able to evaluate which features performed well over all for every product and which machine learning algorithm gives more accurate results.

*Note: the term Category in graphs represents the Combination Category, represented in Chapter 3 table 3.3

4.4.1 Product Category 1

This product category contains 30648 distinct product items. According to results of this product category combination category 13(RB + RBR +JJ + JJS) performed well having top F-Measure of 0.97 while using Random Forest and Gradient Boosting Classifier. Precision for this category is 0.95 having recall of 0.98. Figure 1 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

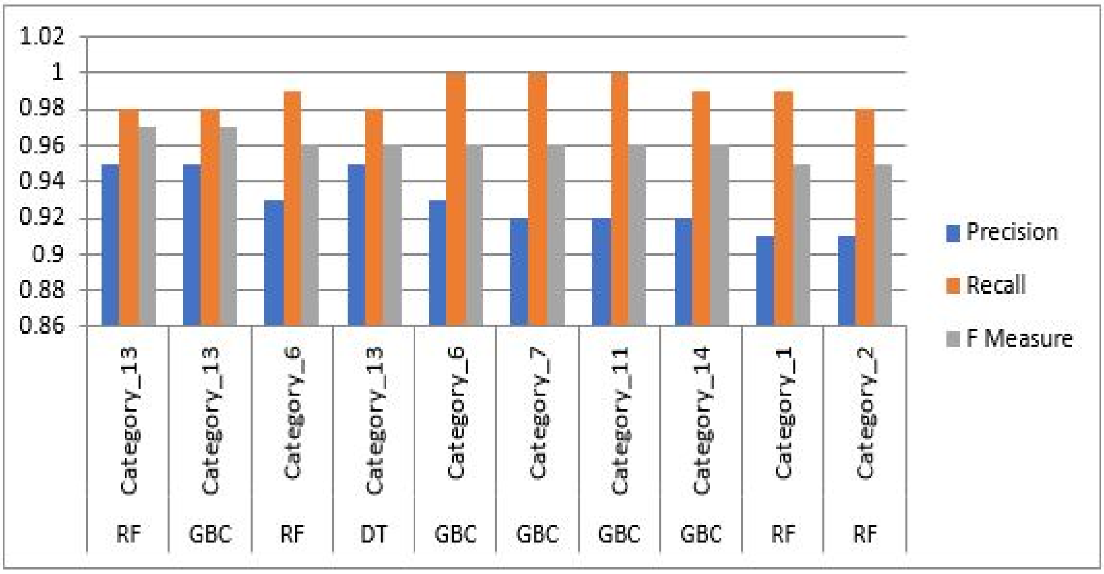

FIGURE 4.1: Top 10 Results of Product Category - Amazon Instant Video

4.4.2 Product Category 2

This product category contains 61551 distinct product items. According to results of this product category Gradient Boosting Classifier performed top with various combination categories like combination category 1, 11, 12, 15 & 16, performed well having top F-Measure of 0.92, while Random Forest have closer results to Gradient Boosting Classifier it also reaches F-Measure of 0.91 with various combination categories like combination category 1, 2, 9, 11 & 12. Figure 2 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

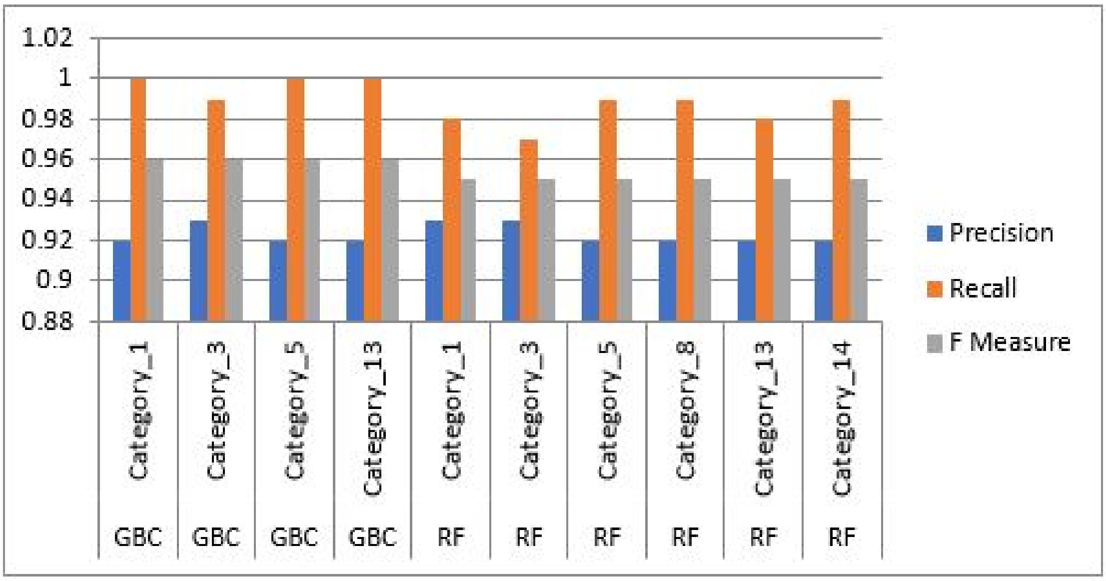

FIGURE 4.2: Top 10 Results of Product Category - Apps for Android

4.4.3 Product Category 3

This product category contains 331090 distinct product items. According to results of this product category Gradient Boosting Classifier performed top with various combination categories like combination category 1, 3,5 & 13 performed well having top F-Measure of 0.96, while Random Forest have closer results to Gradient Boosting Classifier it also reaches F-Measure of 0.95 with various combination categories like combination category 1, 3, 5, 8, 13 & 14. Figure 3 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

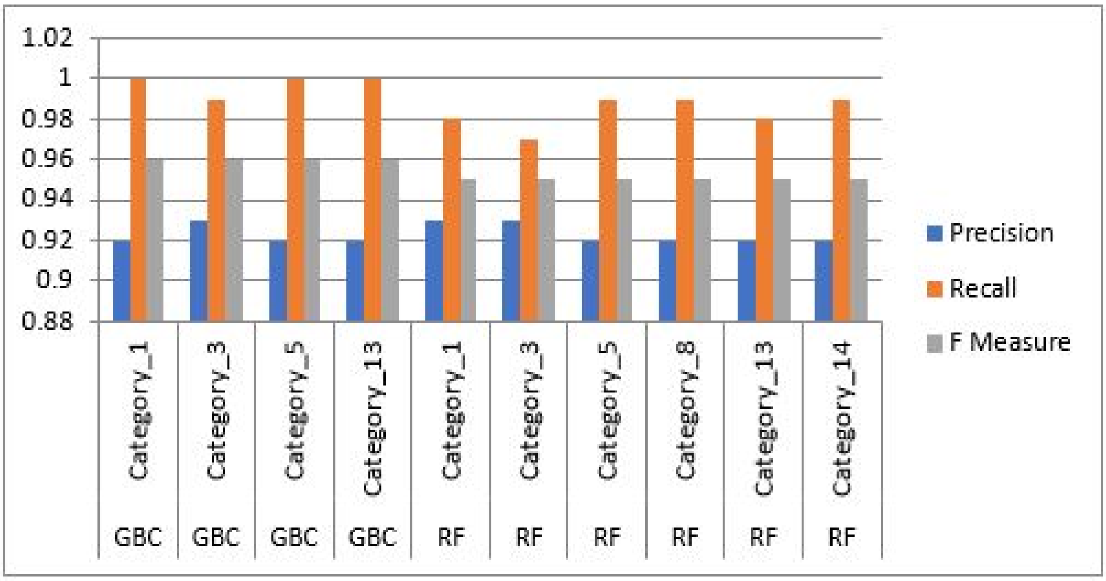

FIGURE 4.3: Top 10 Results of Product Category Automotive

4.4.4 Product Category 4

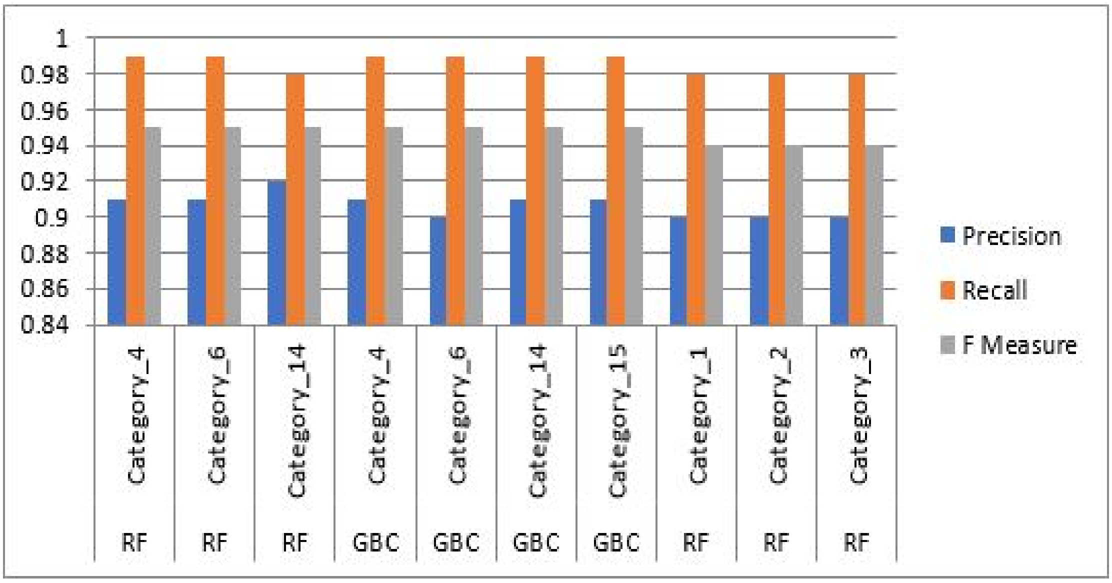

FIGURE 4.4: Top 10 Results of Product Category Beauty

This product category contains 259204 distinct product items. According to results of this product category Random Forest performed top 3 positions with various combination categories like combination category 4, 6 & 14 performed well having top F-Measure of 0.95, while Gradient Boosting Classifier have equal results to Random Forest it also reaches F-Measure of 0.95 with various combination categories like combination category 4, 6, 14 & 15. While Random Forest also got

F-Measure of 0.94 for combination categories 1, 2, & 3. Figure 4.4 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

4.4.5 Product Category 5

This product category contains 346793 distinct product items. According to results of this product category Random Forest performed top 10 positions with various combination categories like combination category 1, 2, 4, 5, 9, 10, 11, 14, 15 & 16 performed well. While among all these combinations category 16 got top F-Measure of 0.96, While all other mentioned categories got 0.95. Figure 5 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

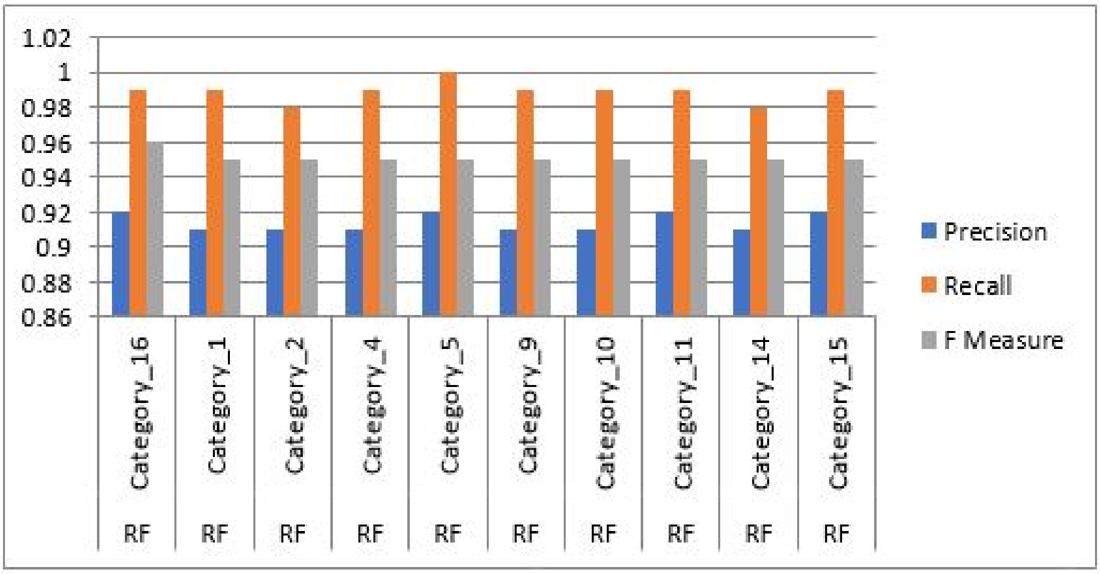

FIGURE 4.5: Top 10 Results of Product Category Cell Phones and Accessories

4.4.6 Product Category 6

This product category contains 279899 distinct product items. According to results of this product category Random Forest performed top 5 positions with various combination categories like combination category 5, 6, 7, 10 & 16 performed well having top F-Measure of 0.99, while Gradient Boosting Classifier have equal results to Random Forest it also reaches F-Measure of 0.99 with various combination categories like combination category 6, 10 & 16. While Random Forest also got F-Measure of 0.98 for combination categories 1 & 2. Figure 6 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

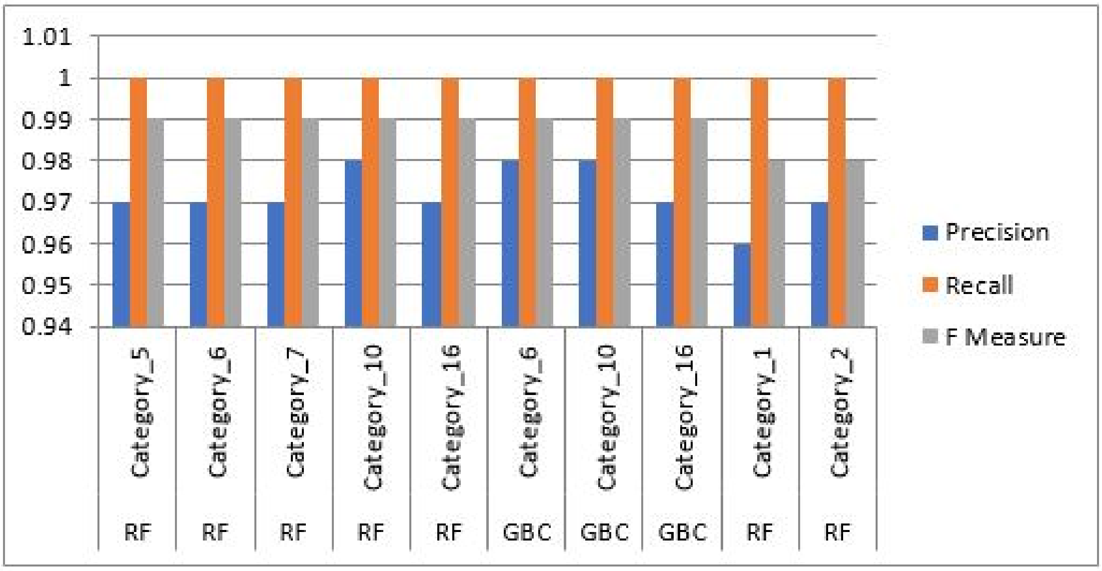

FIGURE 4.6: Top 10 Results of Product Category Digital Music

4.4.7 Product Category 7

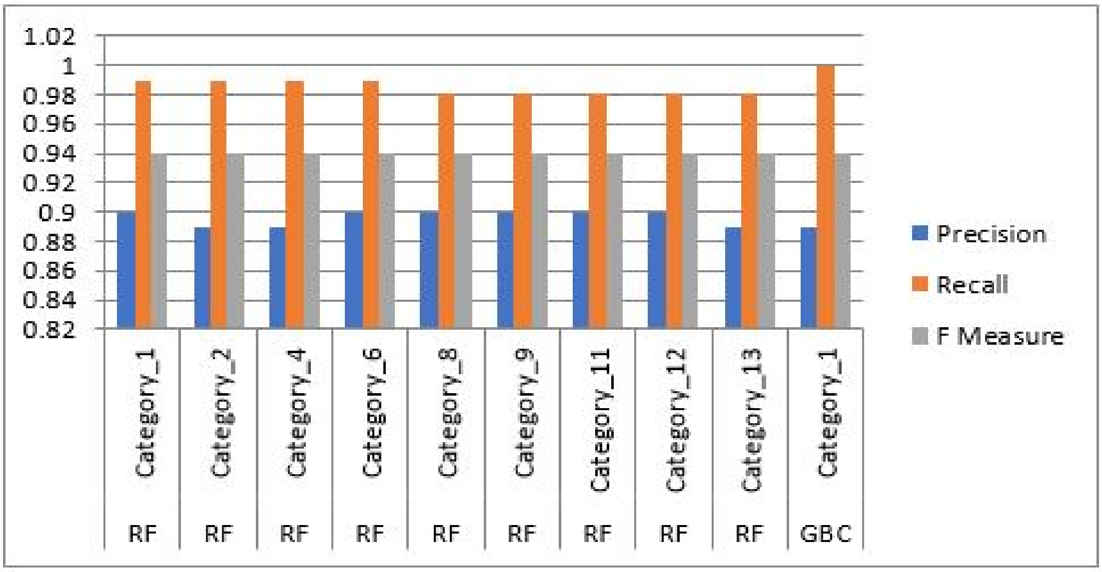

FIGURE 4.7: Top 10 Results of Product Category Health and Personal Care

This product category contains 263032 distinct product items. According to results of this product category Random Forest performed top 9 positions with various combination categories like combination category 1, 2, 4, 6, 8, 9, 11, 12 & 13 performed well. While Gradient Boosting Classifier also got same F Measure of 0.94 with combination category 1. Figure 7 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

4.4.8 Product Category 8

This product category contains 208321 distinct product items. According to results of this product category Random Forest performed top 6 positions with various combination categories like combination category 16, 3, 4, 5, 14 & 15 performed well. While Category 16 have top F-Measure of 0.97, while Gradient Boosting Classifier have equal results to other combination categories of Random Forest it also reaches F-Measure of 0.96 with various combination categories like combination category 3, 5, 6 & 10. Figure 8 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

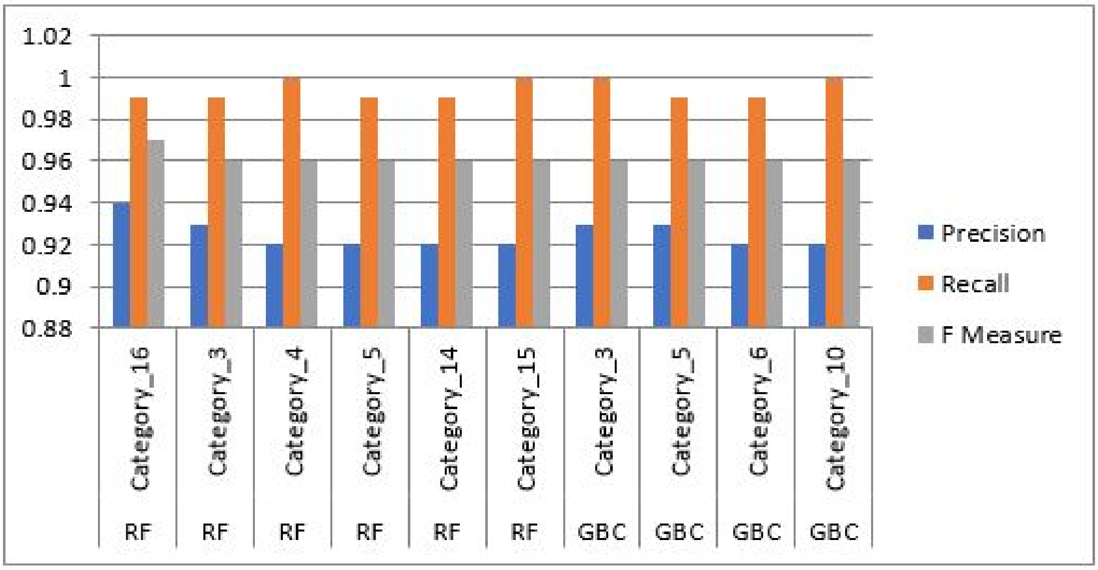

FIGURE 4.8: Top 10 Results of Product Category Movies and TV

4.4.9 Product Category 9

This product category contains 84901 distinct product items. According to results of this product category Random Forest and Gradient Boost Classifier got 4 positions each in top 10 while Decision Tree got 2 positions, all of them have same F-Measure of 0.97. Figure 9 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

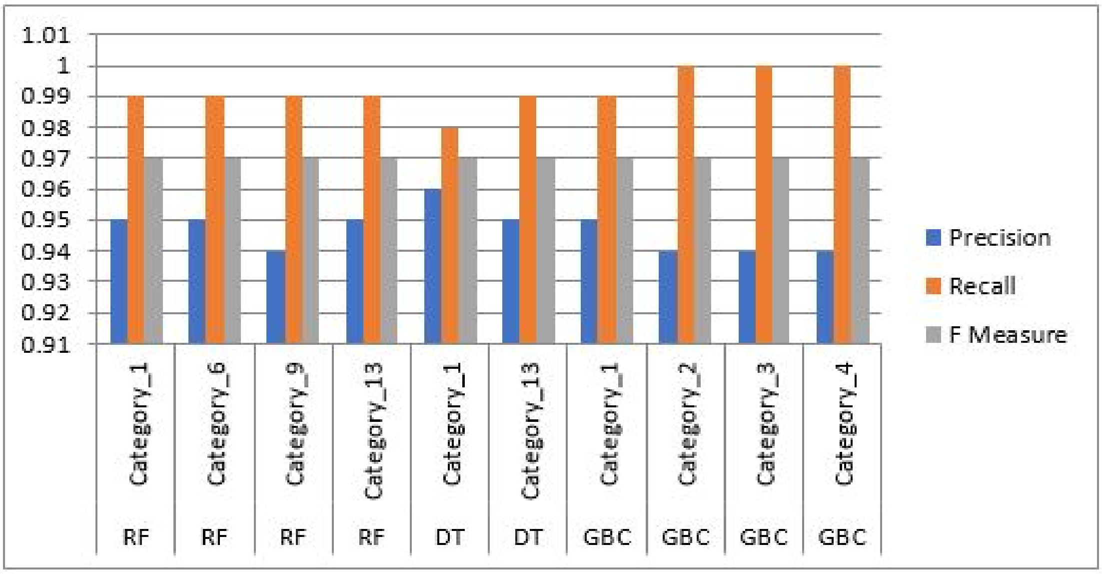

FIGURE 4.9: Top 10 Results of Product Category Musical Instruments

4.4.10 Product Category 10

This product category contains 134838 distinct product items. According to results of this product category Random Forest performed top 8 positions in top 10 while Decision Tree and Gradient Boost Classifier got 1 position each. Each of them got equal F-Measure of 0.97. Figure 10 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

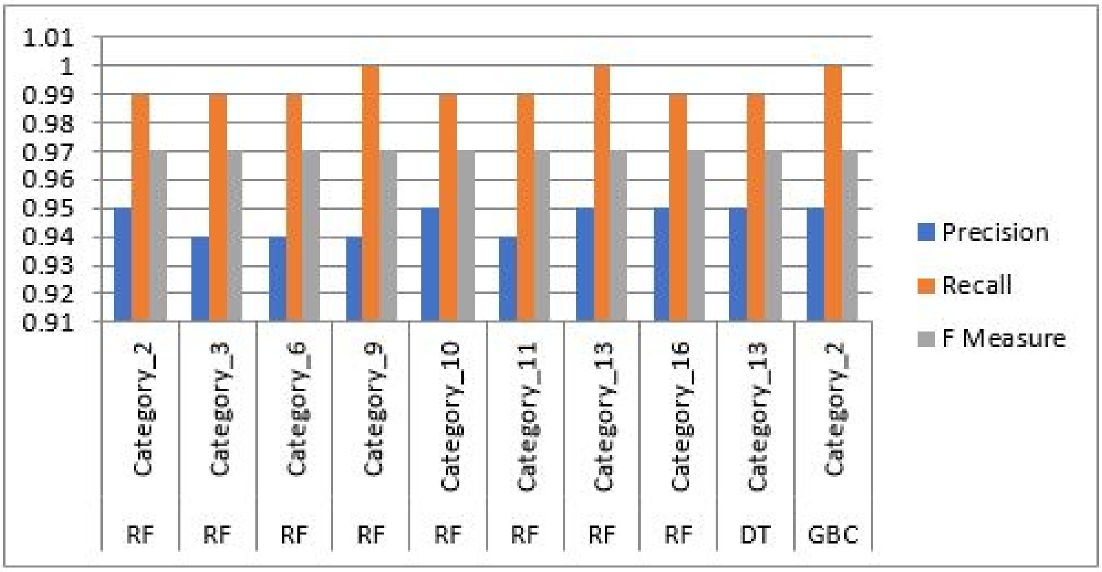

FIGURE 4.10: Top 10 Results of Product Category Office Products

4.4.11 Product Category 11

This product category contains 110707 distinct product items. According to results of this product category Random Forest performed top 8 positions in top 10 while Gradient Boost Classifier got top 2 positions have F-measure of 0.97 greater among others. Random Forest got F-Measure of 0.96s. Figure 11 shows precision, recall & F-Measure respectively, of top 10 results of this product category.

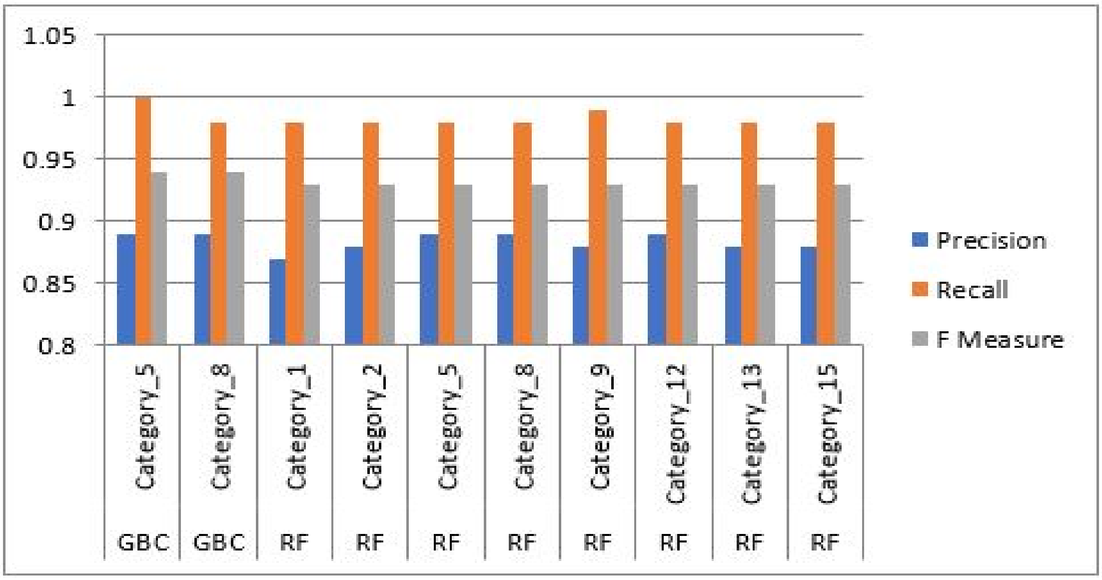

FIGURE 4.11: Top 10 Results of Product Category Pet Supplies

4.4.12 Best of All Products Data

Among All 11 product categories having 2110948 distinct product items and 21.47 million reviews, Random Forest performed best for correct predictions of sentiments. According to overall results of the 11 product categories Random Forest performed top 7 positions with various combination categories like combination category 1, 4, 5, 13 & 16 performed well having top F-Measure of 0.99 with combination category 5, while Gradient Boosting Classifier got 3 top positions with combination categories 1 & 13. Figure 12 shows precision, recall & F-Measure respectively, of top 10 results of all product category.

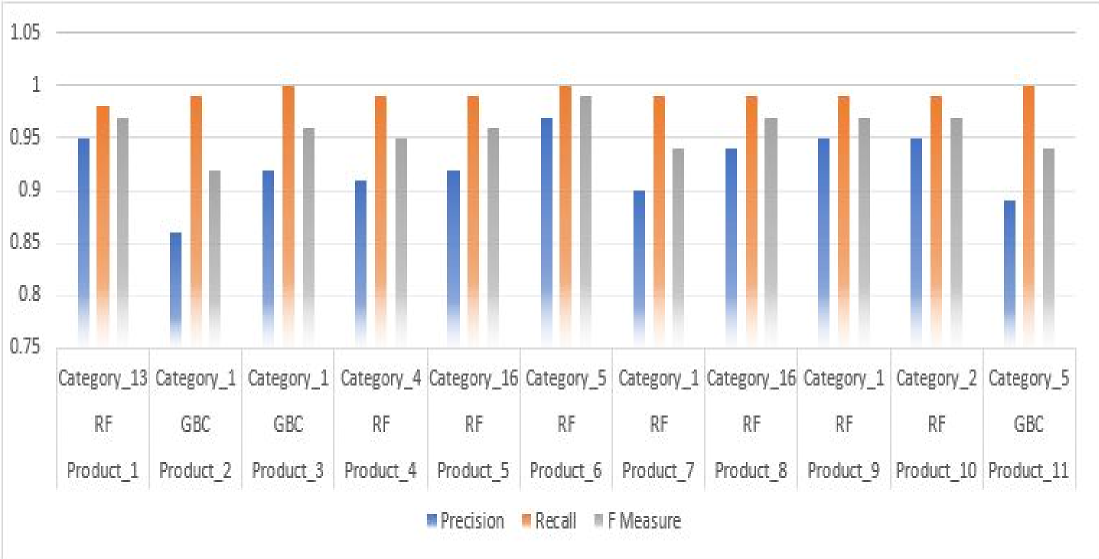

FIGURE 4.12: Top Results of all Product Categories

4.5 Results of Combinations Categories

To evaluating the research strategies, precision and recall are used as basic measures. The F-Measure value can be represented as a weighted harmonic mean of accuracy and recall, and is used to measure classifier accuracy.

In current section, we have results of all 16 selected combination categories for all 11 products datasets, through this we are able to evaluate which combinations performed well over all for every product and which machine learning algorithm gives more accurate results.

*Note: the term Category in graphs represents the Combination Category, represented in Chapter 3 table 3.3

4.5.1 Combination 1 - (RB+RBR+RBS+JJ+JJR+JJS)

In Adjective and Adverb along with all forms, in which we are using generic formula shown in chapter 3 for aggregating the score. According to (Benamara, 2007) adjective and adverb performs better than adverb alone, however, according to results, this combination shows better results while using Random Forest and Gradient Boosting Classifier as machine learning algorithms. Random Forest and Gradient Boosting Classifier shows precision of 0.96 and recall for this is maximum which lead their F-Measure to 0.98, which express that these combinations perform well among others. Figure 13 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

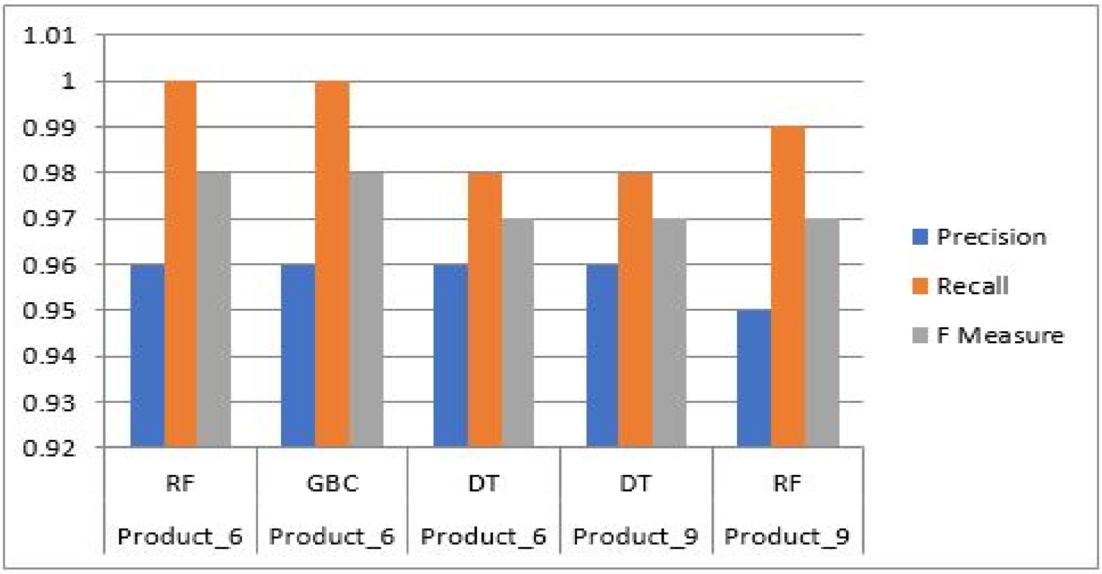

FIGURE 4.13: Top 5 Results of Combination 1

4.5.2 Combination 2 - (RB+RBR+RBS+JJ+JJR)

According to results, this combination shows equal top precision for all 3 classifiers which is 0.97, and recall for Random Forest and Gradient Boosting Classifier is maximum, while Decision Tree have recall 0.98. Figure 14 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

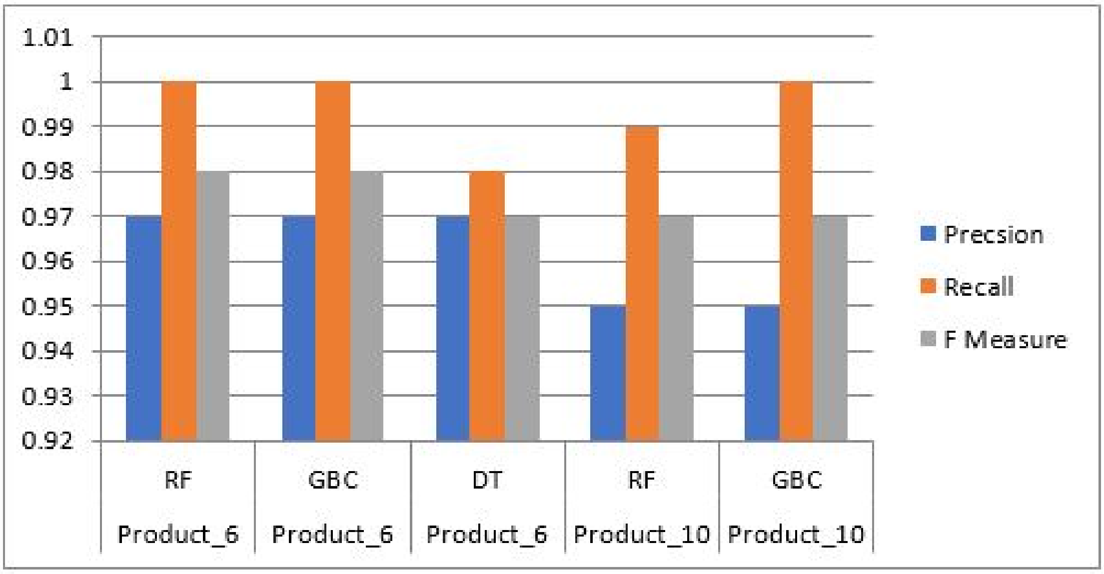

FIGURE 4.14: Top 5 Results of Combination 2

4.5.3 Combination 3 - (RB+RBR+RBS+JJR+JJS)

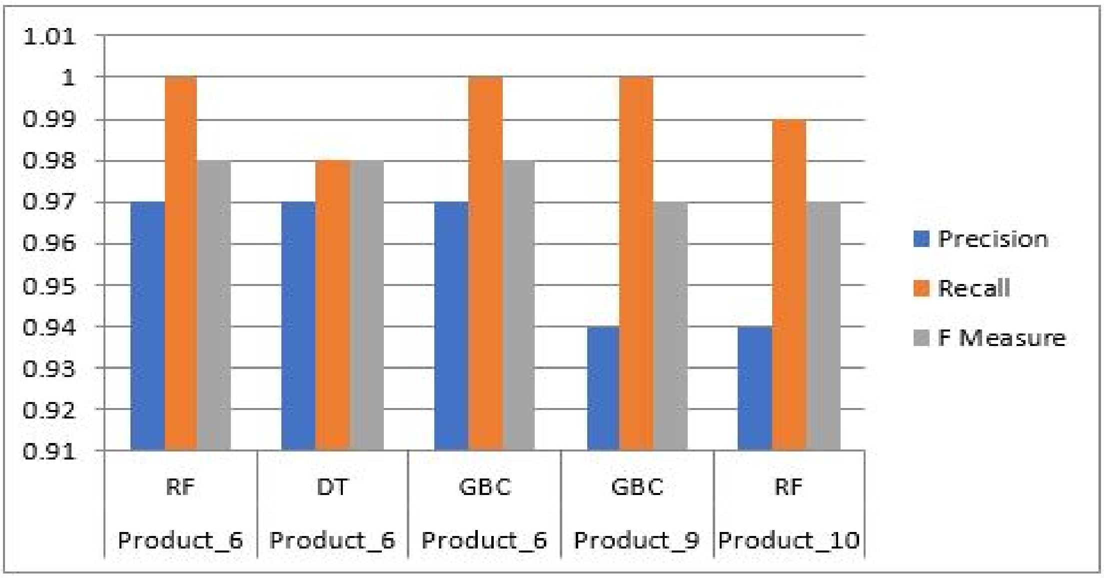

FIGURE 4.15: Top 5 Results of Combination 3

According to results, this combination shows equal top precision for all 3 classifiers which is 0.97, and recall for Random Forest and Gradient Boosting Classifier is maximum, while Decision Tree have recall 0.98. Figure 15 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

4.5.4 Combination 4 - (RBR+RBS+JJ+JJR)

According to results, this combination shows equal top F Measure for Random Forest and Gradient Boosting Classifier classifiers which is 0.98, and recall for Random Forest is maximum, while Decision Tree have F-Measure 0.97. Figure 16 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

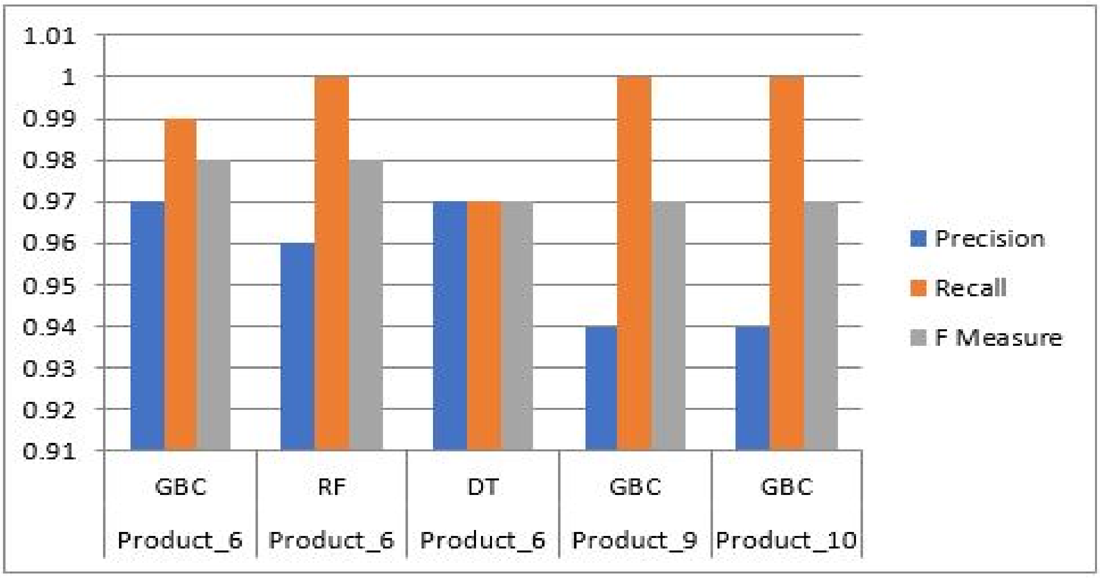

FIGURE 4.16: Top 5 Results of Combination 4

4.5.5 Combination 5 - (RBR+RBS+JJR+JJS)

According to results, this combination shows 0.99 F-Measure for Random Forest. While Gradient Boosting Classifier got 0.98 F-Measure and top F-Measure for Decision Tree is 0.97. Figure 17 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

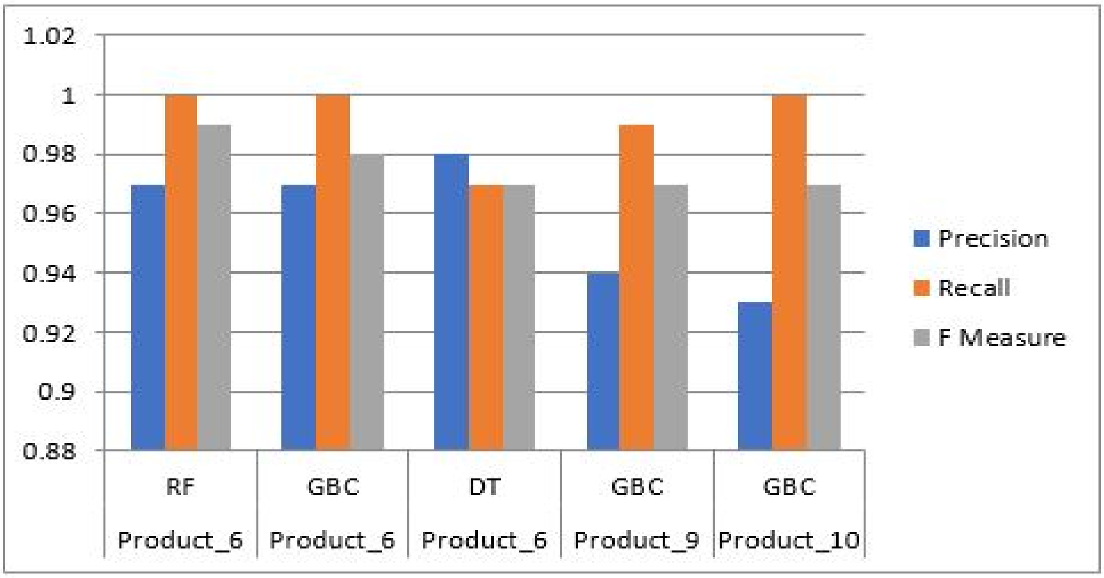

FIGURE 4.17: Top 5 Results of Combination 5

4.5.6 Combination 6 - (RB+RBS+JJ+JJR)

According to results, this combination shows 0.99 F-Measure for Gradient Boosting Classifier and Random Forest, while Decision Tree got 0.98 F-Measure for the same data set. Figure 18 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

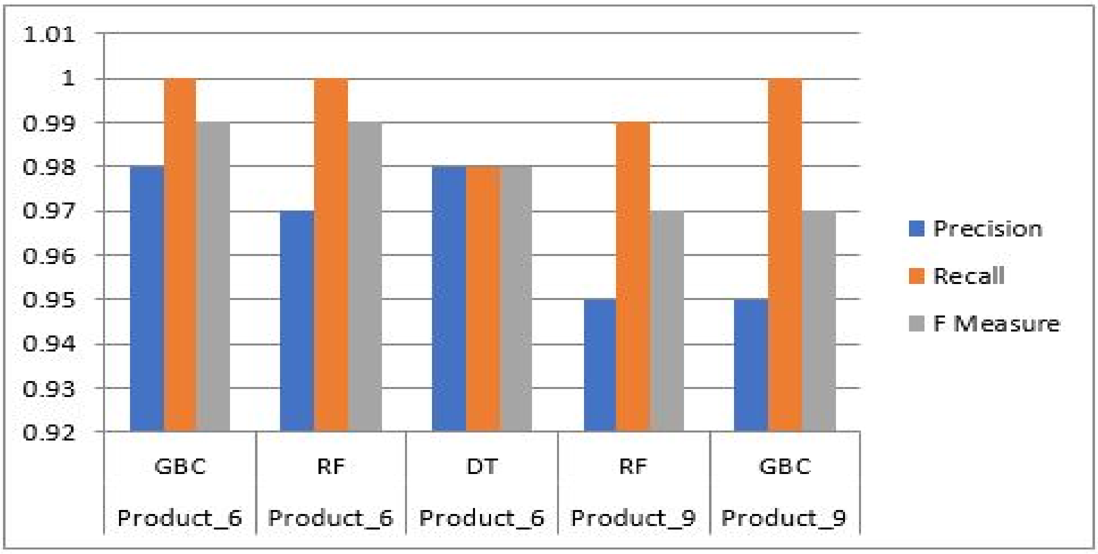

FIGURE 4.18: Top 5 Results of Combination 6

4.5.7 Combination 7 - (RB+RBS+JJR+JJS)

According to results, this combination shows 0.99 F-Measure for Random Forest. While Gradient Boosting Classifier and Decision Tree got 0.98 F-Measure for the same dataset. Figure 19 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

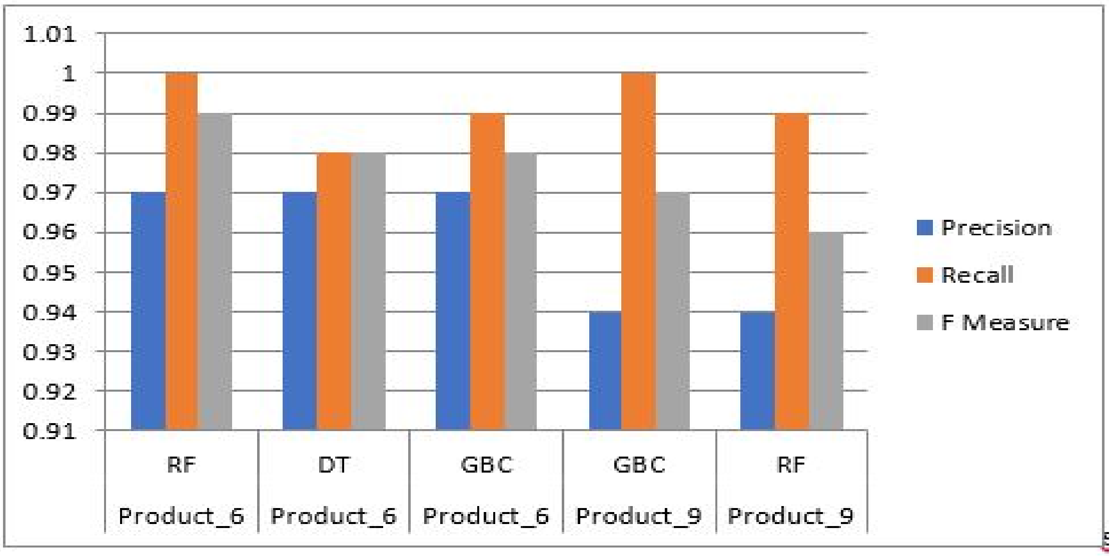

FIGURE 4.19: Top 5 Results of Combination 7

4.5.8 Combination 8 - (RB+RBR+JJ+JJR)

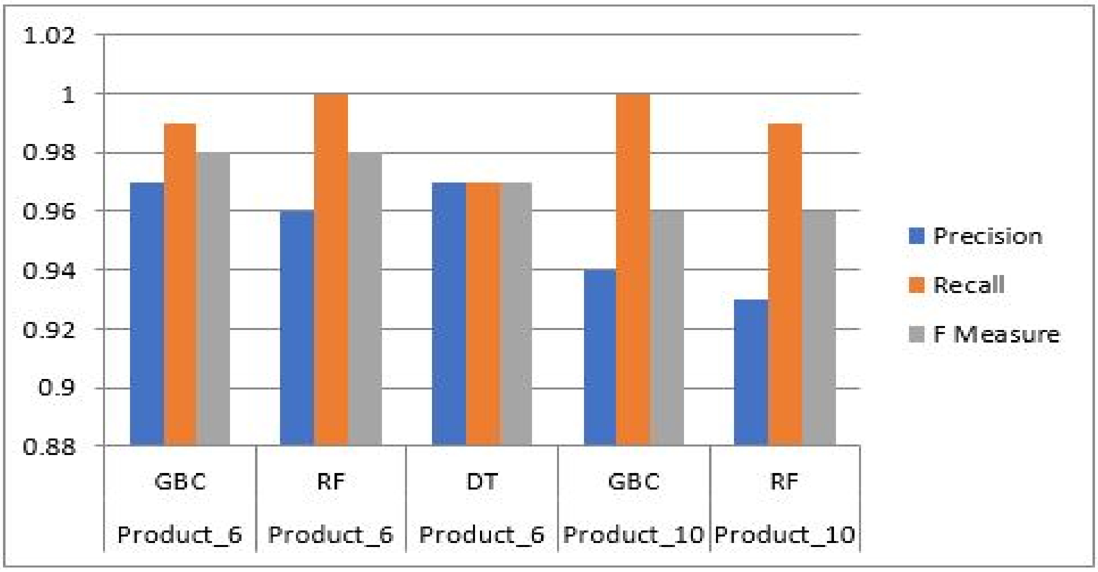

FIGURE 4.20: Top 5 Results of Combination 8

According to results, this combination shows 0.98 F-Measure for Gradient Boosting Classifier and Random Forest, while Decision Tree got 0.97 F-Measure for the same data set. Figure 20 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

4.5.9 Combination 9 - (RB+RBR+JJR+JJS)

According to results, this combination shows 0.98 F-Measure for GBC and Random Forest, while Decision Tree got 0.97 F-Measure for the same data set. Figure 21 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

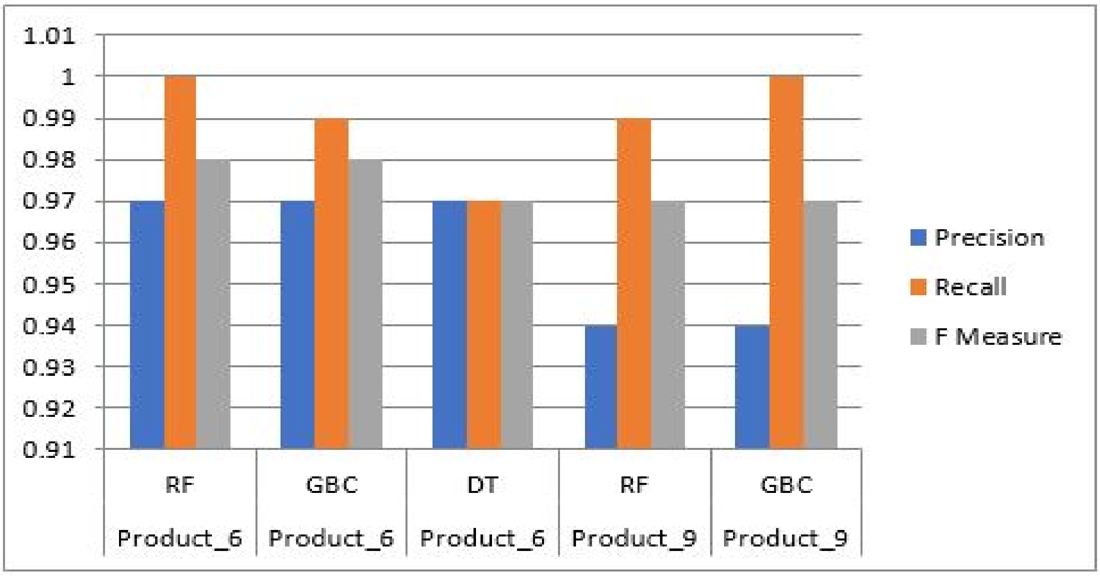

FIGURE 4.21: Top 5 Results of Combination 9

4.5.10 Combination 10 - (RB+RBR+RBS+JJ+JJS)

According to results, this combination shows 0.99 F-Measure for Gradient Boosting Classifier and Random Forest, while Decision Tree got 0.98 F-Measure for the same data set. Figure 22 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

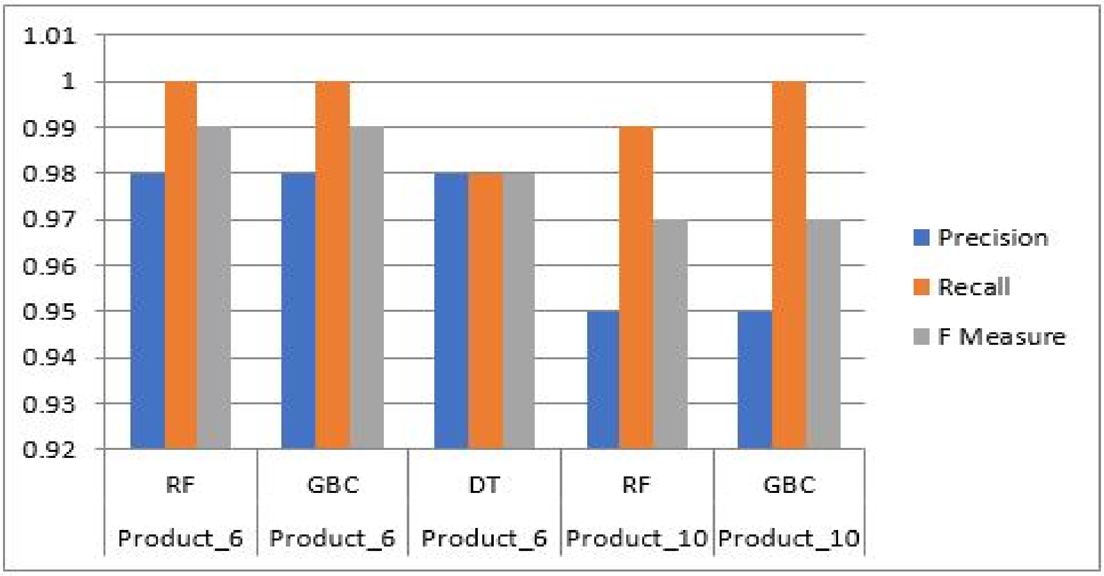

Figure 4.22: Top 5 Results of Combination 9

4.5.11 Combination 11 - (RBR+RBS+JJ+JJS)

According to results, this combination shows 0.98 F-Measure for GBC and Random Forest, while Decision Tree got 0.97 F-Measure for the same data set. Figure 23 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

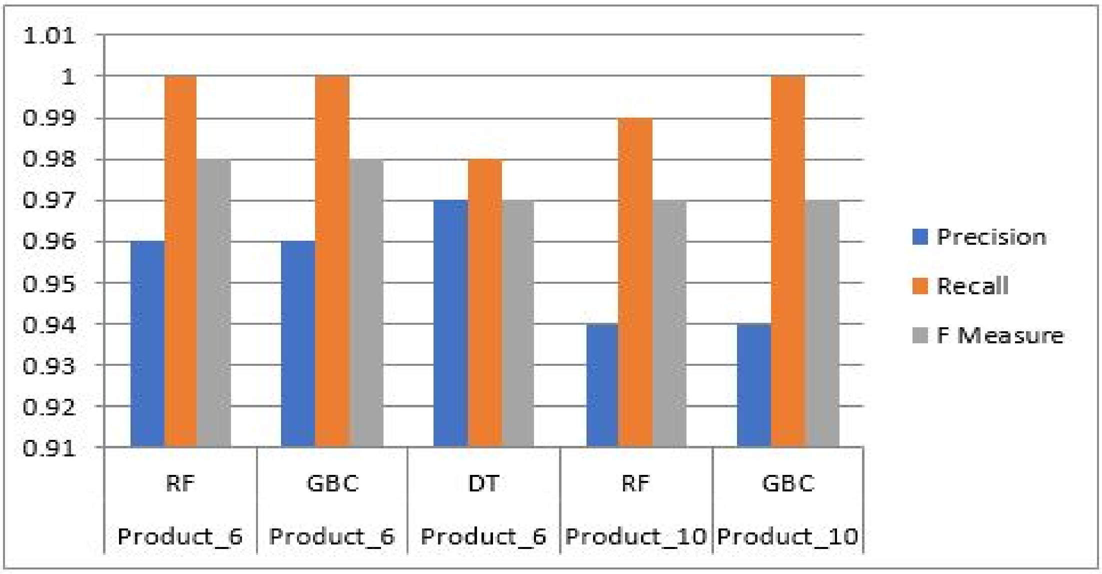

Figure 4.23: Top 5 Results of Combination 10

4.5.12 Combination 12 - (RB+RBS+JJ+JJS)

According to results, this combination shows 0.97 F-Measure for Gradient Boosting Classifier and Random Forest, while Decision Tree got 0.96 F-Measure for the same data set. Figure 24 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

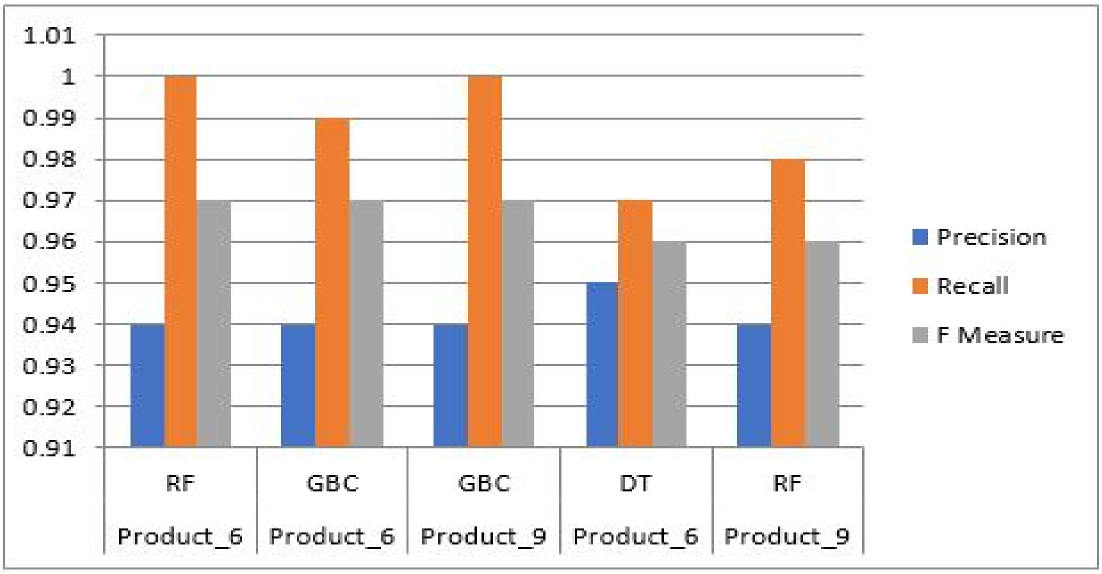

FIGURE 4.24: Top 5 Results of Combination 12

4.5.13 Combination 13 - (RB+RBR+JJ+JJS)

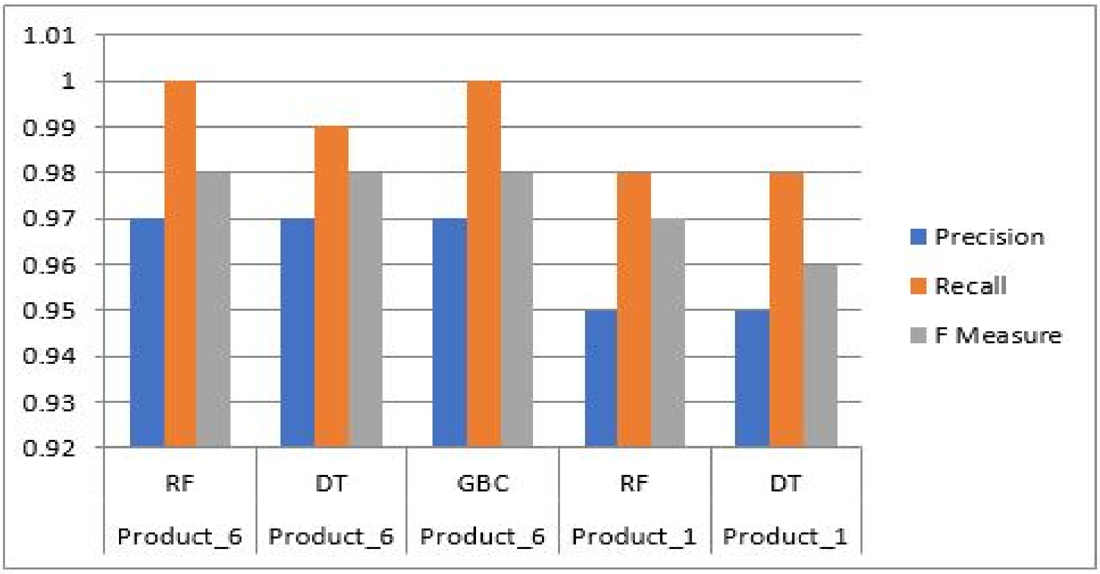

FIGURE 4.25: Top 5 Results of Combination 13

According to results, this combination shows 0.98 F-Measure for all three classifiers GBC, Random Forest and Decision Tree for the same data set. Figure 25 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

4.5.14 Combination 14 - (RB+RBR+JJ+JJR+JJS)

According to results, this combination shows 0.98 F-Measure for Gradient Boosting Classifier and Random Forest, while Decision Tree got 0.97 F-Measure for the same data set. Figure 26 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

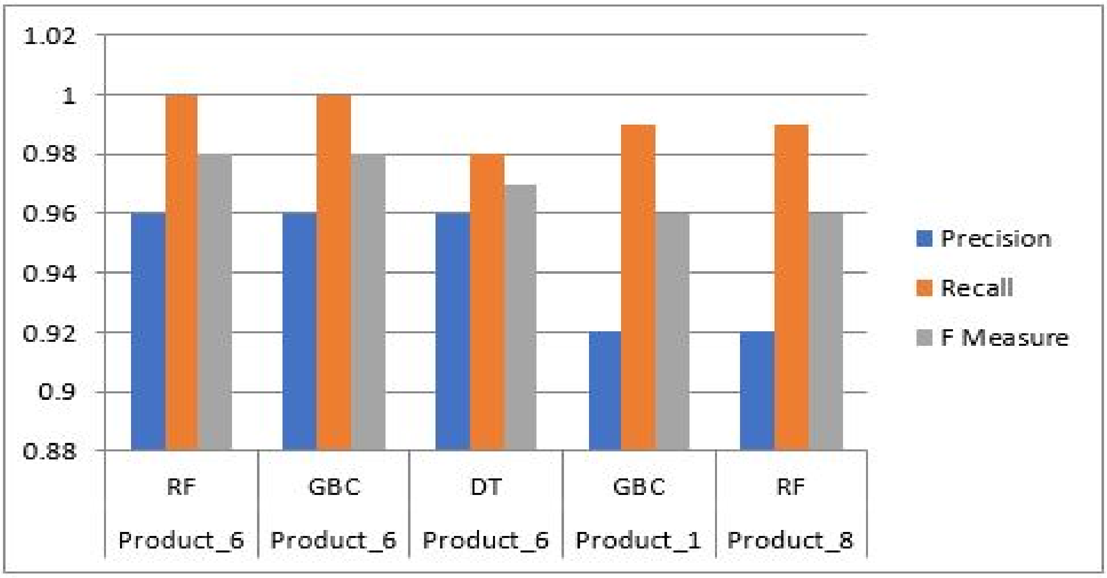

Figure 4.26: Top 5 Results of Combination 14

4.5.15 Combination 15 - (RB+RBS+JJ+JJR+JJS)

According to results, this combination shows 0.98 F-Measure for Gradient Boosting Classifier and Random Forest, while Decision Tree got 0.97 F-Measure for the same data set. Figure 27 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

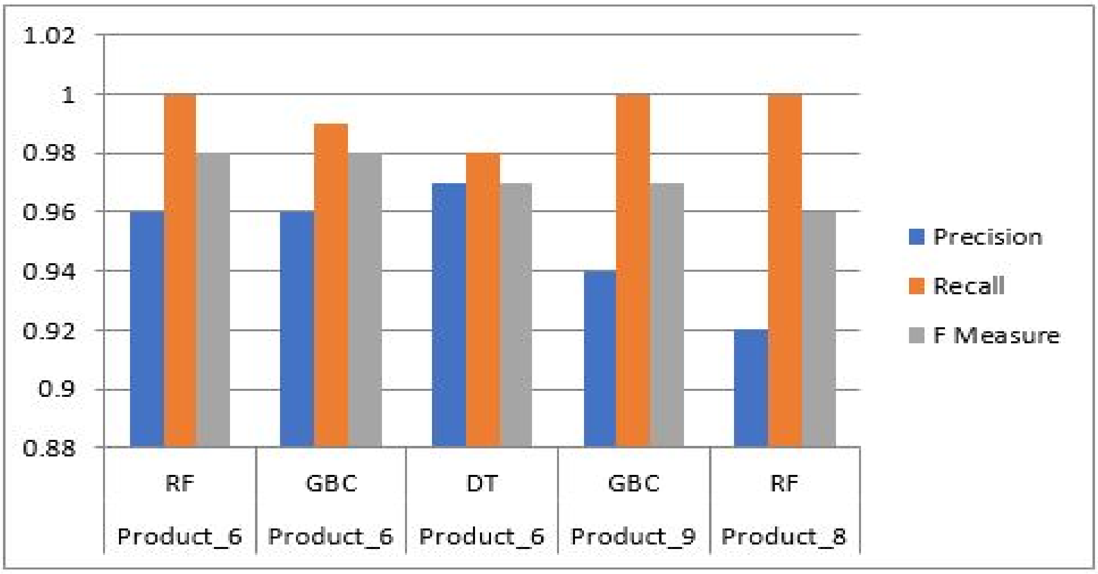

FIGURE 4.27: Top 5 Results of Combination 15

4.5.16 Combination 16 - (RBR+RBS+JJ+JJR+JJS)

According to results, this combination shows 0.99 F-Measure for Gradient Boosting Classifier and Random Forest, while Decision Tree got 0.98 F-Measure for the same data set. Figure 28 shows precision, recall & F-Measure respectively, of top 5 results of this combinations.

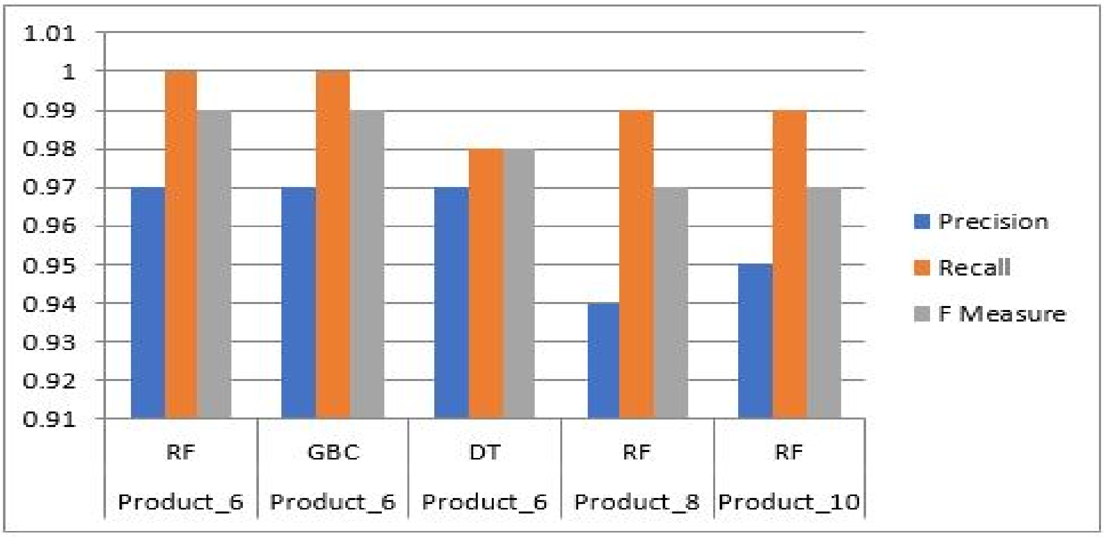

FIGURE 4.28: Top 5 Results of Combination 16

4.5.17 Best of All Combinations

While having look on results, we can clearly observer that Random Forest perform much better than other classifiers. And combination categories 5, 6, 7, 10 & 16 got highest F-Measure of 0.99. These Results are 81% in favor of Random Forest. Figure 29 shows precision, recall & F-Measure respectively, of top results of all combinations of categories.

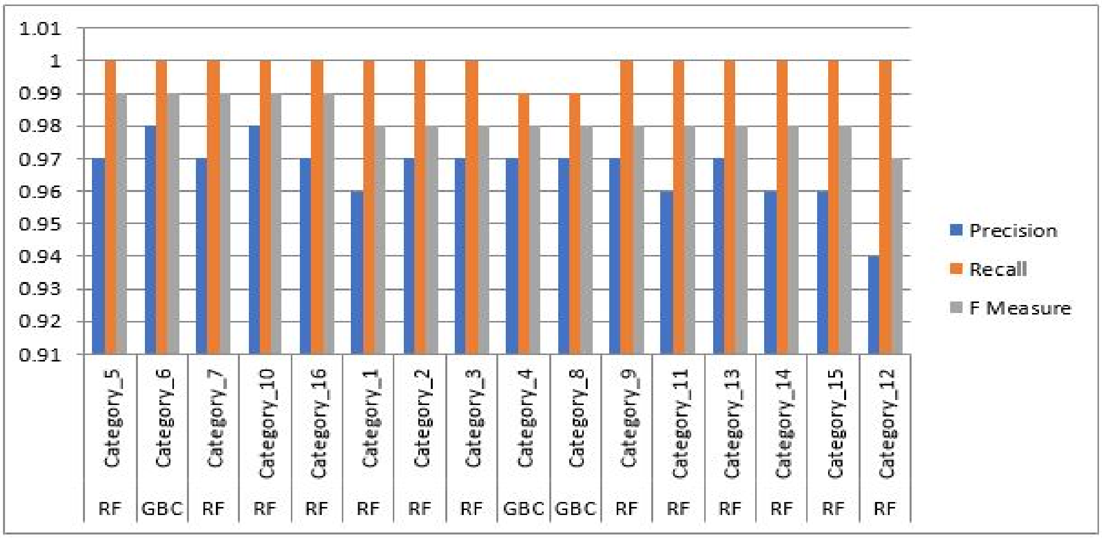

FIGURE 4.29: Top Results of All Combinations Categories

4.6 Results of Combinations According to MLA

In this section we compared each combination category among all categories, while having same dataset and machine learning algorithm. Points are allotted upon each successful comparison to combination having greater value of accuracy while if they got equal accuracy value the point is distributed equally to both of the competitors. Through this we are able to evaluate which combinations performed well over all for every product and which machine learning algorithm gives more accurate results.

*Note: the term Category in graphs represents the Combination Category, represented in Chapter 3 table 3.3

4.6.1 Random Forest

Random forests or random decision forests are an ensemble learning technique for classification, regression and other activities that works by building a variety of decision trees at the training time and outputting the class which is the class mode (classification) or mean prediction (regression) of the individual trees.

In this section we compared each combination category among all categories, while having same dataset and using Random Forest as machine learning algorithm. And we can see combination 13 (RB + RBR + JJ + JJS) got higher points as compared to other combination categories. Over all 165 comparison points distributed in this process and combination 13 leads with 101.5. while combination 1 (RB + RBR + RBS + JJ + JJR + JJS) and combination 6 (RB + RBS + JJ + JJR) got 96.5 and 95 points respectively. Figure 30 shows the comparison of categories for Random Forest.

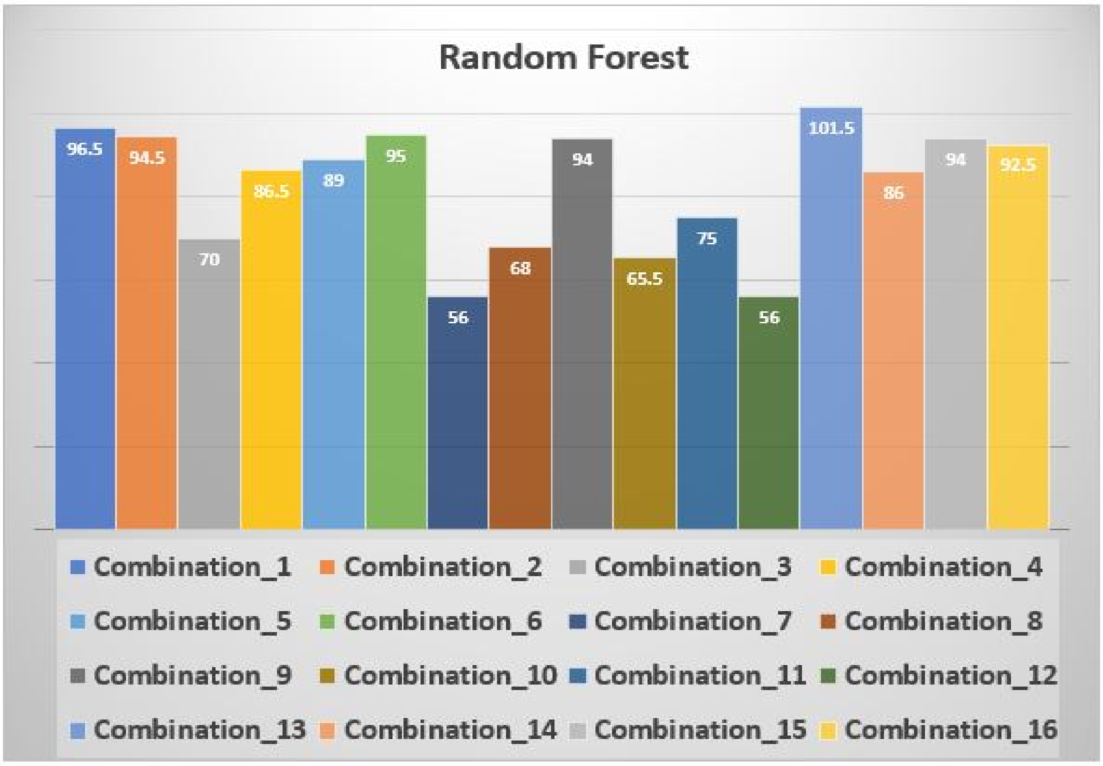

FIGURE 4.30: Comparison of Categories for Random Forest

4.6.2 Decision Tree

A decision tree is a flowchart-like structure where each inner node reflects a "test" on an attributes (e.g., whether a coin flip comes up heads or tails), each node presents the test result, and that each leaf node reflects a class mark (decision made after all attributes have been computed).

In this section we compared each combination category among all categories, while having same dataset and using Decision Tree as machine learning algorithm. And we can see combination 13 (RB + RBR + JJ + JJS) got higher points as compared to other combination categories. Over all 165 comparison points distributed in this process and combination 13 leads with 153. while combination 9 (RB + RBR + JJR + JJS) and combination 4 (RBR + RBS + JJ + JJR) got 102.5 and 97 points respectively. Figure 31 shows the comparison of categories for Decision Tree.

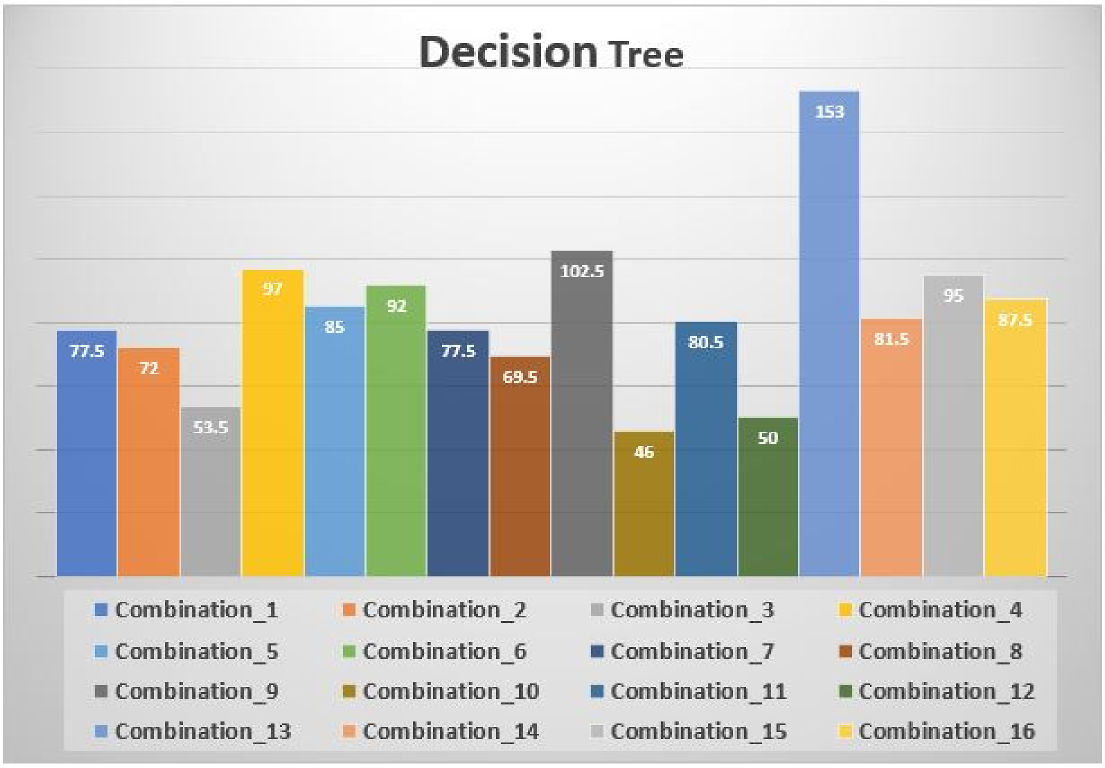

FIGURE 4.31: Comparison of Categories for Decision Tree

4.6.3 Gradient Boosting Classifier

Gradient boosting classifiers are a community of algorithms for machine learning that combine several weak learning models to establish a strong prediction model. Decision trees are typically used when stepping up gradients.

In this section we compared each combination category among all categories, while having same dataset and using Gradient Boosting Classifier as machine learning algorithm. And we can see combination 5 (RBR + RBS + JJR + JJS) got higher points as compared to other combination categories. Over all 165 comparison points distributed in this process and combination 5 leads with 102.5. while combination 6 (RB + RBS + JJ + JJR) and combination 15 (RB + RBS + JJ + JJR + JJS) got 100 and 96 points respectively. Figure 32 shows the comparison of categories for Gradient Boosting Classifier.

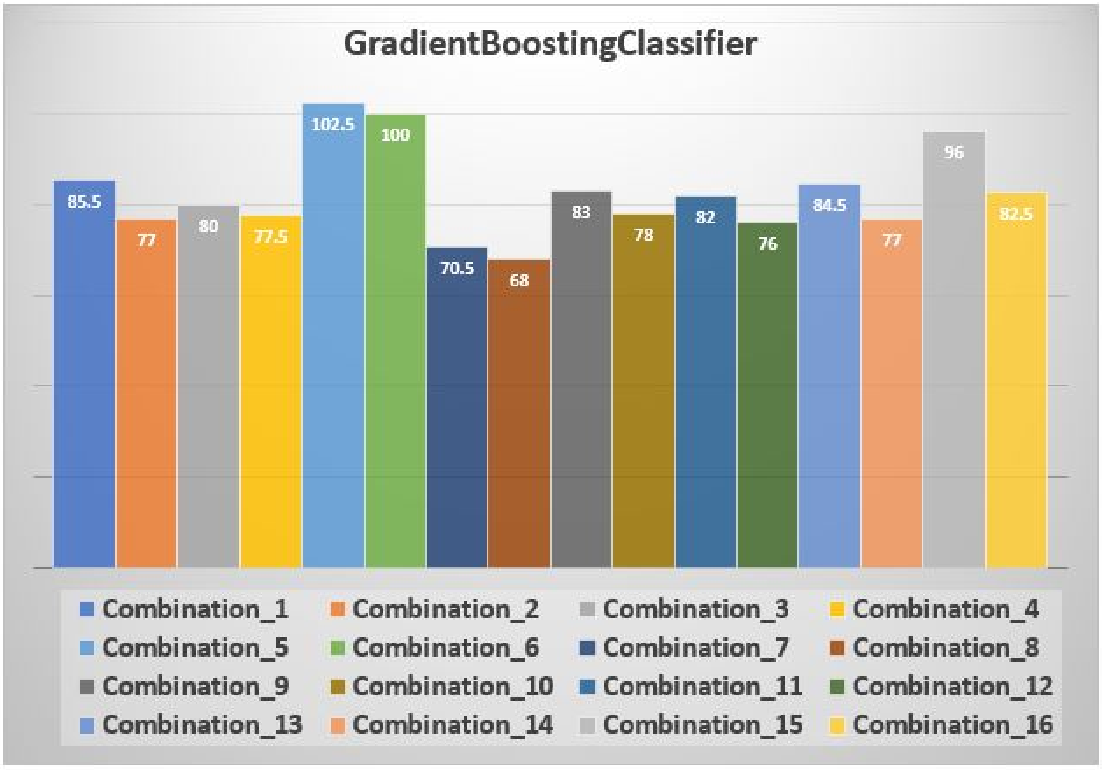

FIGURE 4.32: Comparison of Categories for Gradient Boosting Classifier

4.7 Comparison of Overall Combination Categories

In this section we compared each combination category among all categories, irrespective of machine learning algorithm, while having same dataset. Points are allotted upon each successful comparison to combination having greater value of accuracy while if they got equal accuracy value the point is distributed equally to both of the competitors. Through this we are able to evaluate which combinations performed well over all for every product and which machine learning algorithm gives more accurate results.

Form the results, we can see combination 13 (RB + RBR + JJ + JJS) got higher points as compared to other combination categories. Over all 495 comparison points distributed in this process and combination 13 leads with 339. While, combination 6 (RB + RBS + JJ + JJR) and combination 15 (RB + RBS + JJ + JJR + JJS) got 287and 285 points respectively. Figure 33 shows the overall comparison of categories.

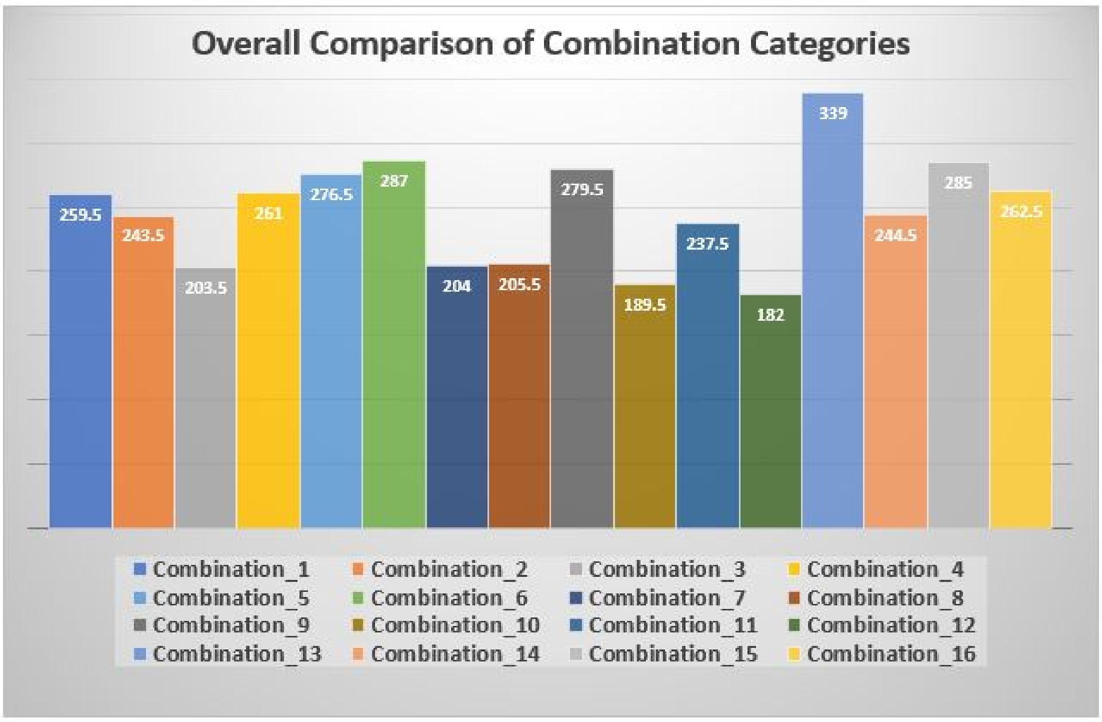

FIGURE 4.33: Overall Comparison of Categories

4.8 Comparison with Previous Studies

In a recent study (xu huahu et al., 2020) proposes a methodology of sentiment polarity classification for a large data collection of Instant Videos online reviews. In their research they use a comprehensive data set of five million Amazon online reviews. There are five classes (Strongly Negative, Negative, Neutral, Positive and Strongly Positive). They also consider three polarity features Verb, Adverb, Adjective and their combinations with their different senses in review-level categorization. The categorization experiments show promising results as the accuracy of the results is 81 percent better than many previous techniques whose average accuracy is 78 percent.

Figure 34 shows, by using Random Forest as a machine learning classifier, the previous researcher got precision of 0.95, while our combination categories got more accurate results. Combination 13 got 0.97 precision, combination 6 got 0.98 and combination 15 got 0.96 precision.

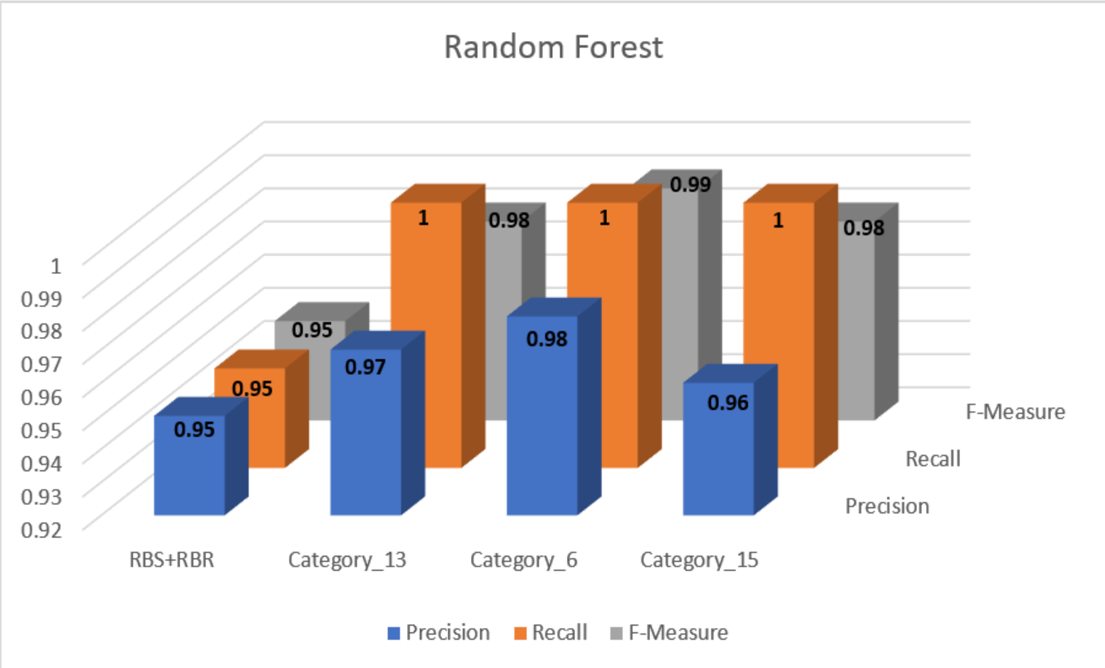

FIGURE 4.34: Comparisons of Top Results of Previous Research using Random Forest

Figure 35 shows, by using Decision Tree as a machine learning classifier, the previous researcher got precision of 0.95, while our combination categories got more accurate results. Combination 13 got 0.97 precision,combination 6 got 0.98 and combination 15 got 0.97 precision.

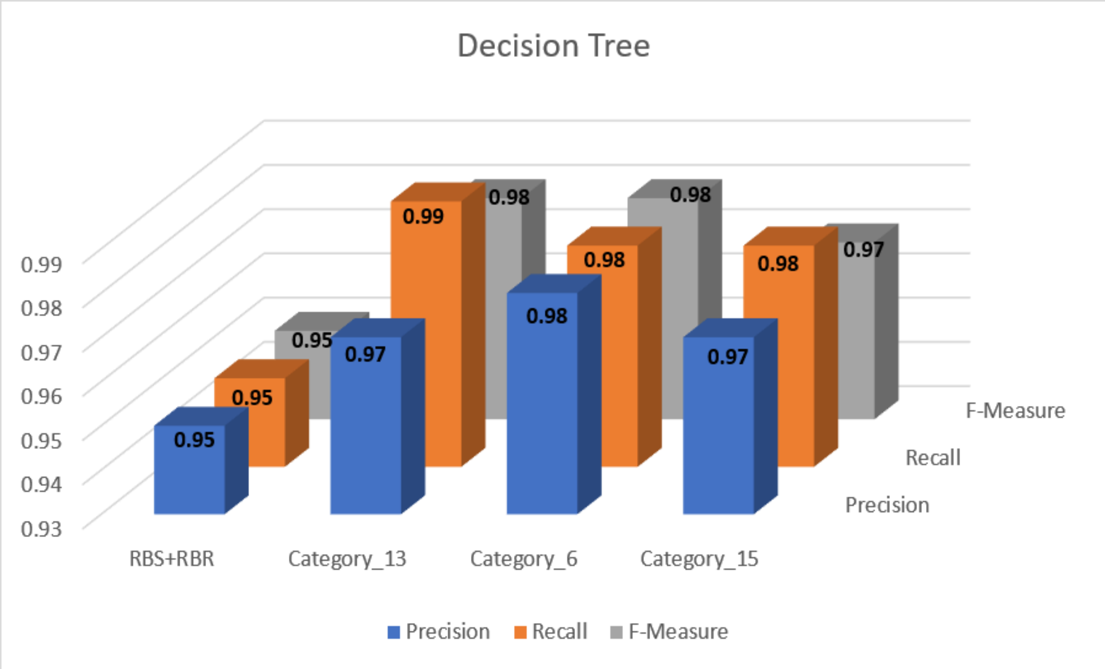

FIGURE 4.35: Comparisons of Top Results of Previous Research using Decision Tree

Figure 36 shows, by using Gradient Boosting Classifier as a machine learning classifier, the previous researcher got precision of 0.95, while our combination categories got more accurate results. Combination 13 got 0.97 precision, combination 6 got 0.98 and combination 15 got 0.96 precision.

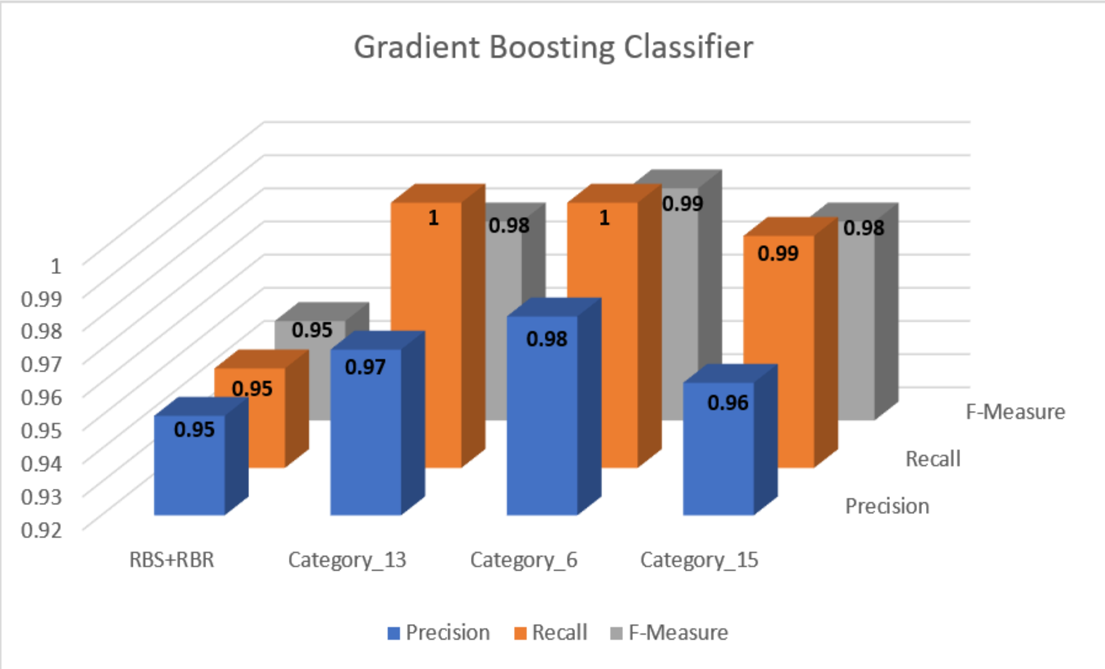

Figure 4.36: Comparisons of Top Results of Previous Research using Gradient Boosting Classifier

Chapter 5

Conclusion and Future Work

Analysis of sentiment is more than just a feature in a social analytics tool; it is a study field. This is an area that is still under investigation, as it has great dimensions. Because of the scope of this research, certain elements of linguistics continue to be discussed or not fully understood in the same way. This process getting traction in 2010, to understand how this analysis actually works, the implications of automating sentiment analysis, and what the future holds for sentiment analysis.

5.1 Conclusion

This thesis aimed to solve a sentiment analysis issue of reviews on amazon products for evaluating polarity bearing elements and lexicon resources. In order to determine the sentiment of the review, we extract the polarity bearing feature word from test dataset. The accuracy is on the extracted data by POS tagger is 95%, misclassified terms are expression and those words which dont have any sentiment score in lexicons. SentiWordNet is a popular lexicon used to identify reviews with a meaning. Sentiment has mainly two classifications positive and negative but neutral has also equal importance. SentiWordNet deal with neutral value whereas other lexicons like SenticNet discard concept of neutral values.

Using Random Forest as machine learning algorithm, combination 13 (RB + RBR + JJ + JJS) got higher F Measure as compared to other combination categories. As shown in figure 30, Over all 1320 comparison points distributed in the process and combination 13 leads with 101.5. while combination 1 (RB + RBR + RBS + JJ + JJR + JJS) and combination 6 (RB + RBS + JJ + JJR) got 96.5 and 95 points respectively. While using Decision Tree as machine learning algorithm, combination 13 (RB + RBR + JJ + JJS) got higher points as compared to other combination categories. As shown in figure 31, in distribution of 1320 comparison points combination 13 leads with 153. while combination 9 (RB + RBR + JJR + JJS) and combination 4 (RBR + RBS + JJ + JJR) got 102.5 and 97 points respectively. While having Gradient Boosting Classifier as machine learning algorithm, combination 5 (RBR + RBS + JJR + JJS) got higher points as compared to other combination categories. As shown in figure 32, it leads with 102.5. while combination 6 (RB + RBS + JJ + JJR) and combination 15 (RB + RBS + JJ + JJR + JJS) got 100 and 96 points respectively.

If we consider the positive reviews in this combination category 13 (RB + RBR + JJ + JJS), while using Random Forest and Decision Tree performed best at its F-Measure is up to 98%, and even its other combinations like combination category 6 (RB + RBS + JJ + JJR).

As shown in Figure 33, combination category 13 (RB + RBR + JJ + JJS), performs better than other combination categories as it got well margined lead in comparison points also has importance in sentiment analysis using Adjective and Adverb. In previous research most researchers have focus on Adjective and adverb as a whole (having all their forms included). According to them, adjective and adverbs combined together is more important feature which has more sentiment, but as a consequence of our results combined effect of some specific forms of adjectives and adverbs have more accuracy.

5.2 Future Work

In future work, we can make more classes by using those linguistic features which performed best in this experiment. By making hypothesis that which reviews has more positive terms are considered in class of "Adjective and Verb" and if reviews has more negative term than results are evaluated by using class of "Adverb and Verb", because negativity not dependent on "not" word and also intensity also not depend on adverb. Each feature has its importance for sentiment analysis.

Bibliography

Asmi, A., & Ishaya, T. (2012, October). Negation identification and calculation in sentiment analysis. In *The second international conference on advances in information mining and management* (pp. 1-7).

B. Liu, (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Found. Trends InformatioPang, B., Lee, L. (2006). Opin. Min. Sentim. Anal. Found. Trends Inf. Retrieval, 1(2), 91-231.

Benamara, F., Cesarano, C., Picariello, A., Recupero, D. R., & Subrahmanian, V. S. (2007). Sentiment analysis: Adjectives and adverbs are better than adjectives alone. *ICWSM*, 7, 203-206.

Bing Liu, L. Z. (2011). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. *HP Laboratories, Technical Report HPL-2011*, 89.

Bo Pang, L. L. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistics.

Barbosa, L., & Feng, J. (2010, August). Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd international conference on computational linguistics: posters* (pp. 36-44). Association for Computational Linguistics.

Benamara, F., Cesarano, C., Picariello, A., Recupero, D. R., & Subrahmanian, V. S. (2007). Sentiment analysis: Adjectives and adverbs are better than adjectives alone. *ICWSM*, 7, 203-206.

Cassinelli, A., & Chen, C. W. (2009). CS224N Final Project Boost up! Sentiment categorization with machine learning techniques, pp. 1-12, 2009.

Cambero, A. (2016). A Comparative Study of Twitter Sentiment Analysis Methods for Live Applications.

Cunha, J., Silva, C., & Antunes, M. (2015). Health twitter big bata management with hadoop framework. *Procedia Computer Science*, 64, 425-431.

Davidov, D., Tsur, O., & Rappoport, A. (2010, August). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd international conference on computational linguistics: posters* (pp. 241-249). Association for Computational Linguistics.

Dragut, E. C. (2014). The Role of Adverbs in Sentiment Analysis. *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore* (1929-2014) (pp. 38-41). Baltimore, Maryland USA: Association for Computational Linguistics.

F. Ali, D. Kwak, P. Khan, S. R. Islam, K. H. Kim, and K. Kwak, (2017). Fuzzy ontology-based sentiment analysis of transportation and city feature reviews for safe traveling. Transportation Research Part C: Emerging Technologies, 77, 33-48.

F. Ali, K.-S. Kwak, and Y.-G. Kim, (2016). Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification. Applied Soft Computing, 47, 235-250.

Farhadloo, M., & Rolland, E. (2016). Fundamentals of sentiment analysis and its applications. In *Sentiment Analysis and Ontology Engineering* (pp. 1-24). Springer, Cham.

Ferreira, L., Jakob, N., & Gurevych, I. (2008, August). A comparative study of feature extraction algorithms in customer reviews. In *2008 IEEE International Conference on Semantic Computing* (pp. 144-151). IEEE.

Gamallo, P., & Garcia, M. (2014, August). Citius: A naive-bayes strategy for sentiment analysis on english tweets. In *Proceedings of the 8th international Workshop on Semantic Evaluation* (SemEval 2014) (pp. 171-175).

Gamon, M. (2004, August). Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In *Proceedings of the 20th international conference on Computational Linguistics* (p. 841). Association for Computational Linguistics.

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(12), 2009.

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-Scale Sentiment Analysis for News and Blogs. *Icwsm*, 7(21), 219-222.

Hadi, A., & Chua, T.-S. (2012). Sentiment Classification Using the Meaning of Words. In *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence.*

Kamps, J., Marx, M., Mokken, R. J., & De Rijke, M. (2004, May). Using WordNet to measure semantic orientations of adjectives. In *LREC* (Vol. 4, pp. 1115-1118).

Kausar, S., Huahu, X., Shabir, M. Y., & Ahmad, W. (2019). A Sentiment Polarity Categorization Technique for Online Product Reviews. *IEEE Access.*

Kobayashi, N., Iida, R., Inui, K., & Matsumoto, Y. (2005). Opinion extraction using a learning-based anaphora resolution technique. In *Companion Volume to the Proceedings of Conference including Posters/Demos and tutorial abstracts.*

Kreutzer, J., & Witte, N. (2013). Opinion mining using sentiwordnet. *Uppsala University.*

Le, B., & Nguyen, H. (2015). Twitter sentiment analysis using machine learning techniques. *In Advanced Computational Methods for Knowledge Engineering* (pp. 279-289). Springer, Cham.

Li, Q. (2015). *Examining the accuracy of sentiment analysis by brand monitoring companies* (Bachelor's thesis, University of Twente).

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

Mohit, G. (2015). Adjective Intensity and Sentiment Analysis. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, (pp. 2520-2526). Portugal: Association for Computational Linguistics.

Musto, C., Semeraro, G., & Polignano, M. (2014, December). A Comparison of Lexicon-based Approaches for Sentiment Analysis of Microblog Posts. In *DART@ AI* IA* (pp. 59-68).

Nasukawa, T., & Yi, J. (2003, October). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture* (pp. 70-77). ACM.

Ohana, B., & Tierney, B. (2009). SSentiment classification of reviews using SentiWordNet. In *9th. it & t conference* (Vol. 13, pp. 18-30).

Ortega, J. L. (2016). *Social network sites for scientists: a quantitative survey.* Chandos Publishing.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.

Parikh, R., & Movassate, M. (2009). Sentiment analysis of user-generated twitter updates using various classification techniques. *CS224N Final Report*, 118.

Peddinti, V. M. K., &Chintalapoodi, P. (2011, August). Domain adaptation in sentiment analysis of twitter. In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence.*

R. Arulmurugan, K. R. Sabarmathi, and H. Anandakumar, (2017). Classification of sentence level sentiment analysis using cloud machine learning techniques. *Cluster Computing*, 1-11.

R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, (2019). Sentiment analysis on product reviews using machine learning techniques. In *Cognitive Informatics and Soft Computing* (pp. 639-647). Springer, Singapore.

Rajganesh, N., Asha, C., Keerthana, A. T., & Suriya, K. (2018). A Hybrid Feedback Based Book Recommendation System Using Sentiment Analysis. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, 3(3).

Schrauwen, S. (2010). Machine learning approaches to sentiment analysis using the dutchnetlog corpus. *Computational Linguistics and Psycholinguistics Research Center*, 30-34.

Smirnova, E., & Balog, K. (2011, April). A user-oriented model for expert finding. In *European Conference on Information Retrieval* (pp. 580-592). Springer, Berlin, Heidelberg.

Suchdev, R., Kotkar, P., Ravindran, R., &Swamy, S. (2014). Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach. *International Journal of Computer Applications*, 103(4).

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.

Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, (pp. 417-424). Canada.

Vohra, M. S., & Teraiya, J. (2013). Applications and challenges for sentiment analysis: A survey. *International journal of engineering research and technology*, 2(2), 1-6.

Wilson, T., & Moore, J. (2011). Twitter Sentiment Analysis: The Good the Bad and the OMG! Proceedings of the *Fifth International AAAI Conference on Weblogs and Social Media.*

Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*, 181(6), 1138-1152.

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. *HP Laboratories, Technical Report HPL-2011*, 89.

Zhang, Z., Ye, Q., Zhang, Z., & Li, Y. (2011). Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Systems with Applications*, 38(6), 7674-7682.

Zhao, Y. (2014). Sentiment Analysis on News Comments Based on Supervised Learning Method. *International Journal of Multimedia and Ubiquitous Engineering*, 333-346.

Zhang, M., & Ye, X. (2008, July). A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In *Proceedings of the 31st* *annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 411-418). ACM.

Zhou, Z., Zhang, X., & Sanderson, M. (2014, July). Sentiment analysis on twitter through topic-based lexicon expansion. In *Australasian Database Conference* (pp. 98-109). Springer, Cham.

Appendix A

Complete Results Data

Complete Results of All Data and Categories

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_1	RF	Category_1	0.91	0.99	0.95
Product_1	RF	Category_2	0.91	0.98	0.95
Product_1	RF	Category_3	0.91	0.98	0.95
Product_1	RF	Category_4	0.91	1	0.95
Product_1	RF	Category_5	0.91	0.99	0.95
Product_1	RF	Category_6	0.93	0.99	0.96
Product_1	RF	Category_7	0.92	0.99	0.95
Product_1	RF	Category_8	0.91	0.99	0.95
Product_1	RF	Category_9	0.9	0.99	0.94
Product_1	RF	Category_10	0.91	0.99	0.95
Product_1	RF	Category_11	0.92	0.98	0.95
Product_1	RF	Category_12	0.91	0.98	0.94
Product_1	RF	Category_13	0.95	0.98	0.97
Product_1	RF	Category_14	0.92	0.99	0.95
Product_1	RF	Category_15	0.91	0.99	0.95
Product_1	RF	Category_16	0.91	0.99	0.94
Product_1	DT	Category_1	0.92	0.94	0.93
Product_1	DT	Category_2	0.92	0.95	0.93
Product_1	DT	Category_3	0.92	0.95	0.93
Product_1	DT	Category_4	0.92	0.95	0.93
Product_1	DT	Category_5	0.92	0.93	0.93
Product_1	DT	Category_6	0.95	0.94	0.95
Product_1	DT	Category_7	0.92	0.96	0.94
Product_1	DT	Category_8	0.93	0.93	0.93
Product_1	DT	Category_9	0.92	0.96	0.94
Product_1	DT	Category_10	0.91	0.95	0.93
Product_1	DT	Category_11	0.93	0.93	0.93

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_1	DT	Category_12	0.92	0.95	0.93
Product_1	DT	Category_13	0.95	0.98	0.96
Product_1	DT	Category_14	0.93	0.95	0.94
Product_1	DT	Category_15	0.92	0.95	0.93
Product_1	DT	Category_16	0.92	0.94	0.93
Product_1	GBC	Category_1	0.91	1	0.95
Product_1	GBC	Category_2	0.91	1	0.95
Product_1	GBC	Category_3	0.91	0.99	0.95
Product_1	GBC	Category_4	0.91	0.99	0.95
Product_1	GBC	Category_5	0.91	0.99	0.95
Product_1	GBC	Category_6	0.93	1	0.96
Product_1	GBC	Category_7	0.92	1	0.96
Product_1	GBC	Category_8	0.91	1	0.95
Product_1	GBC	Category_9	0.9	1	0.95
Product_1	GBC	Category_10	0.91	1	0.95
Product_1	GBC	Category_11	0.92	1	0.96
Product_1	GBC	Category_12	0.91	1	0.95
Product_1	GBC	Category_13	0.95	0.98	0.97
Product_1	GBC	Category_14	0.92	0.99	0.96
Product_1	GBC	Category_15	0.91	0.99	0.95
Product_1	GBC	Category_16	0.91	1	0.95
Product_2	RF	Category_1	0.88	0.94	0.91
Product_2	RF	Category_2	0.86	0.96	0.91
Product_2	RF	Category_3	0.83	0.96	0.89
Product_2	RF	Category_4	0.85	0.96	0.9
Product_2	RF	Category_5	0.85	0.95	0.9
Product_2	RF	Category_6	0.84	0.95	0.89
Product_2	RF	Category_7	0.84	0.94	0.89
Product_2	RF	Category_8	0.84	0.96	0.9
Product_2	RF	Category_9	0.86	0.96	0.91
Product_2	RF	Category_10	0.82	0.94	0.88
Product_2	RF	Category_11	0.87	0.96	0.91
Product_2	RF	Category_12	0.87	0.95	0.91
Product_2	RF	Category_13	0.83	0.98	0.9
Product_2	RF	Category_14	0.85	0.95	0.9
Product_2	RF	Category_15	0.87	0.96	0.91
Product_2	RF	Category_16	0.87	0.95	0.91
Product_2	DT	Category_1	0.88	0.89	0.89
Product_2	DT	Category_2	0.87	0.91	0.89
Product_2	DT	Category_3	0.84	0.9	0.87
Product_2	DT	Category_4	0.86	0.91	0.89
Product_2	DT	Category_5	0.85	0.89	0.87
Product_2	DT	Category_6	0.85	0.89	0.87
Product_2	DT	Category_7	0.85	0.91	0.88
Product_2	DT	Category_8	0.86	0.9	0.88

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_2	DT	Category_9	0.86	0.9	0.88
Product_2	DT	Category_10	0.82	0.9	0.86
Product_2	DT	Category_11	0.9	0.92	0.91
Product_2	DT	Category_12	0.87	0.88	0.88
Product_2	DT	Category_13	0.83	0.96	0.89
Product_2	DT	Category_14	0.86	0.89	0.88
Product_2	DT	Category_15	0.88	0.9	0.89
Product_2	DT	Category_16	0.87	0.9	0.89
Product_2	GBC	Category_1	0.86	0.99	0.92
Product_2	GBC	Category_2	0.84	0.99	0.91
Product_2	GBC	Category_3	0.83	1	0.91
Product_2	GBC	Category_4	0.84	0.99	0.91
Product_2	GBC	Category_5	0.84	0.99	0.91
Product_2	GBC	Category_6	0.83	0.99	0.9
Product_2	GBC	Category_7	0.84	0.99	0.91
Product_2	GBC	Category_8	0.83	0.99	0.9
Product_2	GBC	Category_9	0.84	0.99	0.91
Product_2	GBC	Category_10	0.82	0.99	0.9
Product_2	GBC	Category_11	0.86	0.99	0.92
Product_2	GBC	Category_12	0.86	0.99	0.92
Product_2	GBC	Category_13	0.84	0.99	0.9
Product_2	GBC	Category_14	0.83	0.99	0.9
Product_2	GBC	Category_15	0.86	0.99	0.92
Product_2	GBC	Category_16	0.85	0.99	0.92
Product_3	RF	Category_1	0.93	0.98	0.95
Product_3	RF	Category_2	0.91	0.99	0.94
Product_3	RF	Category_3	0.93	0.97	0.95
Product_3	RF	Category_4	0.9	0.99	0.94
Product_3	RF	Category_5	0.92	0.99	0.95
Product_3	RF	Category_6	0.91	0.98	0.94
Product_3	RF	Category_7	0.9	0.99	0.94
Product_3	RF	Category_8	0.92	0.99	0.95
Product_3	RF	Category_9	0.9	0.99	0.94
Product_3	RF	Category_10	0.91	0.97	0.94
Product_3	RF	Category_11	0.9	0.98	0.94
Product_3	RF	Category_12	0.9	0.97	0.93
Product_3	RF	Category_13	0.92	0.98	0.95
Product_3	RF	Category_14	0.92	0.99	0.95
Product_3	RF	Category_15	0.92	0.99	0.95
Product_3	RF	Category_16	0.88	0.98	0.93
Product_3	DT	Category_1	0.93	0.93	0.93
Product_3	DT	Category_2	0.91	0.94	0.93
Product_3	DT	Category_3	0.93	0.93	0.93
Product_3	DT	Category_4	0.91	0.94	0.93
Product_3	DT	Category_5	0.93	0.95	0.94

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_3	DT	Category_6	0.92	0.93	0.93
Product_3	DT	Category_7	0.9	0.96	0.93
Product_3	DT	Category_8	0.94	0.95	0.94
Product_3	DT	Category_9	0.92	0.94	0.93
Product_3	DT	Category_10	0.91	0.93	0.92
Product_3	DT	Category_11	0.91	0.94	0.92
Product_3	DT	Category_12	0.91	0.93	0.92
Product_3	DT	Category_13	0.92	0.98	0.95
Product_3	DT	Category_14	0.92	0.95	0.93
Product_3	DT	Category_15	0.93	0.93	0.93
Product_3	DT	Category_16	0.89	0.92	0.91
Product_3	GBC	Category_1	0.92	1	0.96
Product_3	GBC	Category_2	0.9	1	0.94
Product_3	GBC	Category_3	0.93	0.99	0.96
Product_3	GBC	Category_4	0.89	1	0.94
Product_3	GBC	Category_5	0.92	1	0.96
Product_3	GBC	Category_6	0.91	1	0.95
Product_3	GBC	Category_7	0.9	0.99	0.94
Product_3	GBC	Category_8	0.91	1	0.95
Product_3	GBC	Category_9	0.9	1	0.95
Product_3	GBC	Category_10	0.9	0.99	0.95
Product_3	GBC	Category_11	0.9	1	0.94
Product_3	GBC	Category_12	0.89	1	0.94
Product_3	GBC	Category_13	0.92	1	0.96
Product_3	GBC	Category_14	0.91	1	0.95
Product_3	GBC	Category_15	0.91	1	0.95
Product_3	GBC	Category_16	0.88	1	0.94
Product_4	RF	Category_1	0.9	0.98	0.94
Product_4	RF	Category_2	0.9	0.98	0.94
Product_4	RF	Category_3	0.9	0.98	0.94
Product_4	RF	Category_4	0.91	0.99	0.95
Product_4	RF	Category_5	0.89	0.98	0.93
Product_4	RF	Category_6	0.91	0.99	0.95
Product_4	RF	Category_7	0.88	0.98	0.93
Product_4	RF	Category_8	0.9	0.98	0.93
Product_4	RF	Category_9	0.9	0.99	0.94
Product_4	RF	Category_10	0.89	0.99	0.93
Product_4	RF	Category_11	0.89	0.98	0.93
Product_4	RF	Category_12	0.9	0.97	0.93
Product_4	RF	Category_13	0.91	0.98	0.94
Product_4	RF	Category_14	0.92	0.98	0.95
Product_4	RF	Category_15	0.9	0.98	0.94
Product_4	RF	Category_16	0.91	0.97	0.94
Product_4	DT	Category_1	0.91	0.93	0.92
Product_4	DT	Category_2	0.91	0.91	0.91

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_4	DT	Category_3	0.91	0.93	0.92
Product_4	DT	Category_4	0.92	0.92	0.92
Product_4	DT	Category_5	0.9	0.92	0.91
Product_4	DT	Category_6	0.92	0.92	0.92
Product_4	DT	Category_7	0.89	0.95	0.92
Product_4	DT	Category_8	0.91	0.88	0.9
Product_4	DT	Category_9	0.91	0.93	0.92
Product_4	DT	Category_10	0.89	0.94	0.91
Product_4	DT	Category_11	0.91	0.92	0.92
Product_4	DT	Category_12	0.9	0.92	0.91
Product_4	DT	Category_13	0.91	0.97	0.94
Product_4	DT	Category_14	0.93	0.9	0.91
Product_4	DT	Category_15	0.92	0.94	0.93
Product_4	DT	Category_16	0.91	0.91	0.91
Product_4	GBC	Category_1	0.9	0.99	0.94
Product_4	GBC	Category_2	0.9	0.98	0.94
Product_4	GBC	Category_3	0.89	1	0.94
Product_4	GBC	Category_4	0.91	0.99	0.95
Product_4	GBC	Category_5	0.9	0.98	0.94
Product_4	GBC	Category_6	0.9	0.99	0.95
Product_4	GBC	Category_7	0.88	1	0.93
Product_4	GBC	Category_8	0.89	0.99	0.94
Product_4	GBC	Category_9	0.9	0.99	0.94
Product_4	GBC	Category_10	0.89	1	0.94
Product_4	GBC	Category_11	0.89	0.99	0.94
Product_4	GBC	Category_12	0.9	0.99	0.94
Product_4	GBC	Category_13	0.91	0.98	0.94
Product_4	GBC	Category_14	0.91	0.99	0.95
Product_4	GBC	Category_15	0.91	0.99	0.95
Product_4	GBC	Category_16	0.91	0.99	0.94
Product_5	RF	Category_1	0.91	0.99	0.95
Product_5	RF	Category_2	0.91	0.98	0.95
Product_5	RF	Category_3	0.91	0.98	0.94
Product_5	RF	Category_4	0.91	0.99	0.95
Product_5	RF	Category_5	0.92	1	0.95
Product_5	RF	Category_6	0.89	0.99	0.94
Product_5	RF	Category_7	0.9	0.98	0.94
Product_5	RF	Category_8	0.91	0.98	0.94
Product_5	RF	Category_9	0.91	0.99	0.95
Product_5	RF	Category_10	0.91	0.99	0.95
Product_5	RF	Category_11	0.92	0.99	0.95
Product_5	RF	Category_12	0.91	0.98	0.94
Product_5	RF	Category_13	0.9	0.99	0.94
Product_5	RF	Category_14	0.91	0.98	0.95
Product_5	RF	Category_15	0.92	0.99	0.95

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_5	RF	Category_16	0.92	0.99	0.96
Product_5	DT	Category_1	0.92	0.93	0.92
Product_5	DT	Category_2	0.92	0.93	0.92
Product_5	DT	Category_3	0.91	0.92	0.91
Product_5	DT	Category_4	0.93	0.93	0.93
Product_5	DT	Category_5	0.93	0.94	0.94
Product_5	DT	Category_6	0.9	0.95	0.92
Product_5	DT	Category_7	0.9	0.94	0.92
Product_5	DT	Category_8	0.92	0.93	0.92
Product_5	DT	Category_9	0.92	0.93	0.93
Product_5	DT	Category_10	0.91	0.94	0.92
Product_5	DT	Category_11	0.93	0.93	0.93
Product_5	DT	Category_12	0.92	0.93	0.92
Product_5	DT	Category_13	0.9	0.98	0.94
Product_5	DT	Category_14	0.93	0.93	0.93
Product_5	DT	Category_15	0.94	0.93	0.93
Product_5	DT	Category_16	0.93	0.93	0.93
Product_5	GBC	Category_1	0.9	0.99	0.95
Product_5	GBC	Category_2	0.91	1	0.95
Product_5	GBC	Category_3	0.9	0.99	0.95
Product_5	GBC	Category_4	0.91	1	0.95
Product_5	GBC	Category_5	0.91	1	0.95
Product_5	GBC	Category_6	0.89	1	0.94
Product_5	GBC	Category_7	0.9	1	0.95
Product_5	GBC	Category_8	0.91	0.99	0.95
Product_5	GBC	Category_9	0.91	0.99	0.95
Product_5	GBC	Category_10	0.91	1	0.95
Product_5	GBC	Category_11	0.92	0.99	0.95
Product_5	GBC	Category_12	0.91	0.99	0.95
Product_5	GBC	Category_13	0.9	0.99	0.94
Product_5	GBC	Category_14	0.91	0.99	0.95
Product_5	GBC	Category_15	0.92	0.99	0.95
Product_5	GBC	Category_16	0.92	0.99	0.95
Product_6	RF	Category_1	0.96	1	0.98
Product_6	RF	Category_2	0.97	1	0.98
Product_6	RF	Category_3	0.97	1	0.98
Product_6	RF	Category_4	0.96	1	0.98
Product_6	RF	Category_5	0.97	1	0.99
Product_6	RF	Category_6	0.97	1	0.99
Product_6	RF	Category_7	0.97	1	0.99
Product_6	RF	Category_8	0.96	1	0.98
Product_6	RF	Category_9	0.97	1	0.98
Product_6	RF	Category_10	0.98	1	0.99
Product_6	RF	Category_11	0.96	1	0.98
Product_6	RF	Category_12	0.94	1	0.97

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_6	RF	Category_13	0.97	1	0.98
Product_6	RF	Category_14	0.96	1	0.98
Product_6	RF	Category_15	0.96	1	0.98
Product_6	RF	Category_16	0.97	1	0.99
Product_6	DT	Category_1	0.96	0.98	0.97
Product_6	DT	Category_2	0.97	0.98	0.97
Product_6	DT	Category_3	0.97	0.98	0.98
Product_6	DT	Category_4	0.97	0.97	0.97
Product_6	DT	Category_5	0.98	0.97	0.97
Product_6	DT	Category_6	0.98	0.98	0.98
Product_6	DT	Category_7	0.97	0.98	0.98
Product_6	DT	Category_8	0.97	0.97	0.97
Product_6	DT	Category_9	0.97	0.97	0.97
Product_6	DT	Category_10	0.98	0.98	0.98
Product_6	DT	Category_11	0.97	0.98	0.97
Product_6	DT	Category_12	0.95	0.97	0.96
Product_6	DT	Category_13	0.97	0.99	0.98
Product_6	DT	Category_14	0.96	0.98	0.97
Product_6	DT	Category_15	0.97	0.98	0.97
Product_6	DT	Category_16	0.97	0.98	0.98
Product_6	GBC	Category_1	0.96	1	0.98
Product_6	GBC	Category_2	0.97	1	0.98
Product_6	GBC	Category_3	0.97	1	0.98
Product_6	GBC	Category_4	0.97	0.99	0.98
Product_6	GBC	Category_5	0.97	1	0.98
Product_6	GBC	Category_6	0.98	1	0.99
Product_6	GBC	Category_7	0.97	0.99	0.98
Product_6	GBC	Category_8	0.97	0.99	0.98
Product_6	GBC	Category_9	0.97	0.99	0.98
Product_6	GBC	Category_10	0.98	1	0.99
Product_6	GBC	Category_11	0.96	1	0.98
Product_6	GBC	Category_12	0.94	0.99	0.97
Product_6	GBC	Category_13	0.97	1	0.98
Product_6	GBC	Category_14	0.96	1	0.98
Product_6	GBC	Category_15	0.96	0.99	0.98
Product_6	GBC	Category_16	0.97	1	0.99
Product_7	RF	Category_1	0.9	0.99	0.94
Product_7	RF	Category_2	0.89	0.99	0.94
Product_7	RF	Category_3	0.86	0.99	0.92
Product_7	RF	Category_4	0.89	0.99	0.94
Product_7	RF	Category_5	0.89	0.98	0.93
Product_7	RF	Category_6	0.9	0.99	0.94
Product_7	RF	Category_7	0.89	0.98	0.93
Product_7	RF	Category_8	0.9	0.98	0.94
Product_7	RF	Category_9	0.9	0.98	0.94

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_7	RF	Category_10	0.89	0.97	0.93
Product_7	RF	Category_11	0.9	0.98	0.94
Product_7	RF	Category_12	0.9	0.98	0.94
Product_7	RF	Category_13	0.89	0.98	0.94
Product_7	RF	Category_14	0.88	0.99	0.93
Product_7	RF	Category_15	0.89	0.97	0.93
Product_7	RF	Category_16	0.89	0.98	0.93
Product_7	DT	Category_1	0.92	0.93	0.93
Product_7	DT	Category_2	0.91	0.94	0.92
Product_7	DT	Category_3	0.87	0.93	0.9
Product_7	DT	Category_4	0.92	0.94	0.93
Product_7	DT	Category_5	0.91	0.92	0.91
Product_7	DT	Category_6	0.91	0.92	0.91
Product_7	DT	Category_7	0.89	0.93	0.91
Product_7	DT	Category_8	0.91	0.93	0.92
Product_7	DT	Category_9	0.92	0.94	0.93
Product_7	DT	Category_10	0.89	0.93	0.91
Product_7	DT	Category_11	0.91	0.93	0.92
Product_7	DT	Category_12	0.91	0.93	0.92
Product_7	DT	Category_13	0.89	0.96	0.93
Product_7	DT	Category_14	0.91	0.92	0.91
Product_7	DT	Category_15	0.92	0.92	0.92
Product_7	DT	Category_16	0.91	0.92	0.92
Product_7	GBC	Category_1	0.89	1	0.94
Product_7	GBC	Category_2	0.89	0.99	0.94
Product_7	GBC	Category_3	0.86	1	0.92
Product_7	GBC	Category_4	0.89	1	0.94
Product_7	GBC	Category_5	0.89	0.99	0.94
Product_7	GBC	Category_6	0.89	0.99	0.94
Product_7	GBC	Category_7	0.89	1	0.94
Product_7	GBC	Category_8	0.89	1	0.94
Product_7	GBC	Category_9	0.89	1	0.94
Product_7	GBC	Category_10	0.89	1	0.94
Product_7	GBC	Category_11	0.9	1	0.94
Product_7	GBC	Category_12	0.89	1	0.94
Product_7	GBC	Category_13	0.89	0.99	0.94
Product_7	GBC	Category_14	0.87	1	0.93
Product_7	GBC	Category_15	0.89	1	0.94
Product_7	GBC	Category_16	0.88	0.99	0.93
Product_8	RF	Category_1	0.91	1	0.95
Product_8	RF	Category_2	0.91	1	0.95
Product_8	RF	Category_3	0.93	0.99	0.96
Product_8	RF	Category_4	0.92	1	0.96
Product_8	RF	Category_5	0.92	0.99	0.96
Product_8	RF	Category_6	0.92	0.99	0.95

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_8	RF	Category_7	0.91	0.99	0.95
Product_8	RF	Category_8	0.92	0.99	0.95
Product_8	RF	Category_9	0.9	1	0.95
Product_8	RF	Category_10	0.92	0.98	0.95
Product_8	RF	Category_11	0.92	0.99	0.95
Product_8	RF	Category_12	0.92	0.99	0.95
Product_8	RF	Category_13	0.91	0.99	0.95
Product_8	RF	Category_14	0.92	0.99	0.96
Product_8	RF	Category_15	0.92	1	0.96
Product_8	RF	Category_16	0.94	0.99	0.97
Product_8	DT	Category_1	0.92	0.94	0.93
Product_8	DT	Category_2	0.92	0.93	0.93
Product_8	DT	Category_3	0.93	0.94	0.93
Product_8	DT	Category_4	0.93	0.94	0.94
Product_8	DT	Category_5	0.94	0.94	0.94
Product_8	DT	Category_6	0.93	0.96	0.94
Product_8	DT	Category_7	0.91	0.94	0.93
Product_8	DT	Category_8	0.93	0.95	0.94
Product_8	DT	Category_9	0.92	0.95	0.93
Product_8	DT	Category_10	0.92	0.93	0.93
Product_8	DT	Category_11	0.92	0.94	0.93
Product_8	DT	Category_12	0.93	0.93	0.93
Product_8	DT	Category_13	0.91	0.97	0.94
Product_8	DT	Category_14	0.94	0.94	0.94
Product_8	DT	Category_15	0.93	0.95	0.94
Product_8	DT	Category_16	0.95	0.94	0.94
Product_8	GBC	Category_1	0.91	1	0.95
Product_8	GBC	Category_2	0.91	0.99	0.95
Product_8	GBC	Category_3	0.93	1	0.96
Product_8	GBC	Category_4	0.92	1	0.95
Product_8	GBC	Category_5	0.93	0.99	0.96
Product_8	GBC	Category_6	0.92	0.99	0.96
Product_8	GBC	Category_7	0.91	1	0.95
Product_8	GBC	Category_8	0.92	0.99	0.95
Product_8	GBC	Category_9	0.91	0.99	0.95
Product_8	GBC	Category_10	0.92	1	0.96
Product_8	GBC	Category_11	0.91	0.99	0.95
Product_8	GBC	Category_12	0.92	0.99	0.96
Product_8	GBC	Category_13	0.91	0.99	0.95
Product_8	GBC	Category_14	0.93	1	0.96
Product_8	GBC	Category_15	0.92	1	0.96
Product_8	GBC	Category_16	0.94	0.99	0.96
Product_9	RF	Category_1	0.95	0.99	0.97
Product_9	RF	Category_2	0.94	0.99	0.96
Product_9	RF	Category_3	0.94	0.98	0.96

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_9	RF	Category_4	0.93	0.99	0.96
Product_9	RF	Category_5	0.94	0.99	0.96
Product_9	RF	Category_6	0.95	0.99	0.97
Product_9	RF	Category_7	0.94	0.99	0.96
Product_9	RF	Category_8	0.92	0.99	0.95
Product_9	RF	Category_9	0.94	0.99	0.97
Product_9	RF	Category_10	0.93	0.99	0.96
Product_9	RF	Category_11	0.92	0.99	0.95
Product_9	RF	Category_12	0.94	0.98	0.96
Product_9	RF	Category_13	0.95	0.99	0.97
Product_9	RF	Category_14	0.92	1	0.96
Product_9	RF	Category_15	0.94	0.99	0.96
Product_9	RF	Category_16	0.93	0.98	0.95
Product_9	DT	Category_1	0.96	0.98	0.97
Product_9	DT	Category_2	0.95	0.95	0.95
Product_9	DT	Category_3	0.94	0.95	0.94
Product_9	DT	Category_4	0.94	0.96	0.95
Product_9	DT	Category_5	0.95	0.95	0.95
Product_9	DT	Category_6	0.96	0.95	0.95
Product_9	DT	Category_7	0.94	0.96	0.95
Product_9	DT	Category_8	0.93	0.94	0.94
Product_9	DT	Category_9	0.95	0.95	0.95
Product_9	DT	Category_10	0.93	0.96	0.94
Product_9	DT	Category_11	0.93	0.95	0.94
Product_9	DT	Category_12	0.95	0.92	0.93
Product_9	DT	Category_13	0.95	0.99	0.97
Product_9	DT	Category_14	0.94	0.96	0.95
Product_9	DT	Category_15	0.95	0.94	0.94
Product_9	DT	Category_16	0.94	0.92	0.93
Product_9	GBC	Category_1	0.95	0.99	0.97
Product_9	GBC	Category_2	0.94	1	0.97
Product_9	GBC	Category_3	0.94	1	0.97
Product_9	GBC	Category_4	0.94	1	0.97
Product_9	GBC	Category_5	0.94	0.99	0.97
Product_9	GBC	Category_6	0.95	1	0.97
Product_9	GBC	Category_7	0.94	1	0.97
Product_9	GBC	Category_8	0.92	0.99	0.96
Product_9	GBC	Category_9	0.94	1	0.97
Product_9	GBC	Category_10	0.93	1	0.96
Product_9	GBC	Category_11	0.93	1	0.96
Product_9	GBC	Category_12	0.94	1	0.97
Product_9	GBC	Category_13	0.95	0.99	0.97
Product_9	GBC	Category_14	0.92	1	0.96
Product_9	GBC	Category_15	0.94	1	0.97
Product_9	GBC	Category_16	0.93	1	0.96

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_10	RF	Category_1	0.92	1	0.95
Product_10	RF	Category_2	0.95	0.99	0.97
Product_10	RF	Category_3	0.94	0.99	0.97
Product_10	RF	Category_4	0.93	0.99	0.96
Product_10	RF	Category_5	0.93	1	0.96
Product_10	RF	Category_6	0.94	0.99	0.97
Product_10	RF	Category_7	0.93	0.99	0.96
Product_10	RF	Category_8	0.93	0.99	0.96
Product_10	RF	Category_9	0.94	1	0.97
Product_10	RF	Category_10	0.95	0.99	0.97
Product_10	RF	Category_11	0.94	0.99	0.97
Product_10	RF	Category_12	0.94	0.99	0.96
Product_10	RF	Category_13	0.95	1	0.97
Product_10	RF	Category_14	0.93	1	0.96
Product_10	RF	Category_15	0.93	0.99	0.96
Product_10	RF	Category_16	0.95	0.99	0.97
Product_10	DT	Category_1	0.92	0.96	0.94
Product_10	DT	Category_2	0.95	0.96	0.96
Product_10	DT	Category_3	0.94	0.96	0.95
Product_10	DT	Category_4	0.94	0.95	0.95
Product_10	DT	Category_5	0.95	0.96	0.95
Product_10	DT	Category_6	0.95	0.96	0.95
Product_10	DT	Category_7	0.93	0.97	0.95
Product_10	DT	Category_8	0.94	0.96	0.95
Product_10	DT	Category_9	0.95	0.96	0.96
Product_10	DT	Category_10	0.95	0.95	0.95
Product_10	DT	Category_11	0.95	0.96	0.95
Product_10	DT	Category_12	0.94	0.96	0.95
Product_10	DT	Category_13	0.95	0.99	0.97
Product_10	DT	Category_14	0.94	0.97	0.95
Product_10	DT	Category_15	0.94	0.95	0.95
Product_10	DT	Category_16	0.95	0.97	0.96
Product_10	GBC	Category_1	0.92	1	0.96
Product_10	GBC	Category_2	0.95	1	0.97
Product_10	GBC	Category_3	0.94	1	0.97
Product_10	GBC	Category_4	0.94	1	0.97
Product_10	GBC	Category_5	0.93	1	0.97
Product_10	GBC	Category_6	0.94	1	0.97
Product_10	GBC	Category_7	0.93	1	0.96
Product_10	GBC	Category_8	0.94	1	0.96
Product_10	GBC	Category_9	0.94	1	0.97
Product_10	GBC	Category_10	0.95	1	0.97
Product_10	GBC	Category_11	0.94	1	0.97
Product_10	GBC	Category_12	0.93	0.99	0.96
Product_10	GBC	Category_13	0.94	1	0.97

Continued on next page

Table A.1 – continued from previous page

Product	ML Classifier	Combination class	Precision	Recall	F-Measure
Product_10	GBC	Category_14	0.93	0.99	0.96
Product_10	GBC	Category_15	0.93	1	0.96
Product_10	GBC	Category_16	0.94	1	0.97
Product_11	RF	Category_1	0.87	0.98	0.93
Product_11	RF	Category_2	0.88	0.98	0.93
Product_11	RF	Category_3	0.87	0.96	0.91
Product_11	RF	Category_4	0.87	0.99	0.92
Product_11	RF	Category_5	0.89	0.98	0.93
Product_11	RF	Category_6	0.88	0.97	0.92
Product_11	RF	Category_7	0.87	0.98	0.92
Product_11	RF	Category_8	0.89	0.98	0.93
Product_11	RF	Category_9	0.88	0.99	0.93
Product_11	RF	Category_10	0.85	0.97	0.91
Product_11	RF	Category_11	0.88	0.96	0.92
Product_11	RF	Category_12	0.89	0.98	0.93
Product_11	RF	Category_13	0.88	0.98	0.93
Product_11	RF	Category_14	0.88	0.97	0.92
Product_11	RF	Category_15	0.88	0.98	0.93
Product_11	RF	Category_16	0.89	0.97	0.93
Product_11	DT	Category_1	0.88	0.89	0.89
Product_11	DT	Category_2	0.89	0.88	0.88
Product_11	DT	Category_3	0.86	0.92	0.89
Product_11	DT	Category_4	0.89	0.91	0.9
Product_11	DT	Category_5	0.89	0.92	0.91
Product_11	DT	Category_6	0.89	0.92	0.91
Product_11	DT	Category_7	0.87	0.92	0.89
Product_11	DT	Category_8	0.9	0.9	0.9
Product_11	DT	Category_9	0.9	0.93	0.91
Product_11	DT	Category_10	0.85	0.93	0.89
Product_11	DT	Category_11	0.89	0.93	0.91
Product_11	DT	Category_12	0.9	0.93	0.91
Product_11	DT	Category_13	0.88	0.97	0.92
Product_11	DT	Category_14	0.89	0.92	0.9
Product_11	DT	Category_15	0.89	0.93	0.91
Product_11	DT	Category_16	0.9	0.92	0.91
Product_11	GBC	Category_1	0.87	1	0.93
Product_11	GBC	Category_2	0.87	0.99	0.93
Product_11	GBC	Category_3	0.86	1	0.92
Product_11	GBC	Category_4	0.86	0.99	0.92
Product_11	GBC	Category_5	0.89	1	0.94
Product_11	GBC	Category_6	0.87	0.99	0.93
Product_11	GBC	Category_7	0.87	0.99	0.93
Product_11	GBC	Category_8	0.89	0.98	0.94
Product_11	GBC	Category_9	0.88	0.99	0.93
Product_11	GBC	Category_10	0.85	0.99	0.92

Continued on next page

Table A.1 – continued from previous page

Product	**ML Classifier**	**Combination class**	**Precision**	**Recall**	**F-Measure**
Product_11	GBC	Category_11	0.88	1	0.93
Product_11	GBC	Category_12	0.88	1	0.93
Product_11	GBC	Category_13	0.88	0.99	0.93
Product_11	GBC	Category_14	0.87	1	0.93
Product_11	GBC	Category_15	0.87	0.99	0.93
Product_11	GBC	Category_16	0.89	0.99	0.93