

Research Article

Hamza Aldabbas*, Abdullah Bajahzar, Meshrif Alruily, Ali Adil Qureshi, Rana M. Amir Latif, and Muhammad Farhan

Google Play Content Scraping and Knowledge Engineering using Natural Language Processing Techniques with the Analysis of User Reviews

<https://doi.org/10.1515/jisys-2019-0197>

Received Aug 08, 2019; accepted Dec 01, 2019

Abstract: To maintain the competitive edge and evaluating the needs of the quality app is in the mobile application market. The user's feedback on these applications plays an essential role in the mobile application development industry. The rapid growth of web technology gave people an opportunity to interact and express their review, rate and share their feedback about applications. In this paper we have scrapped 506259 of user reviews and applications rate from Google Play Store from 14 different categories. The statistical information was measured in the results using different of common machine learning algorithms such as the Logistic Regression, Random Forest Classifier, and Multinomial Naïve Bayes. Different parameters including the accuracy, precision, recall, and F1 score were used to evaluate Bigram, Trigram, and N-gram, and the statistical result of these algorithms was compared. The analysis of each algorithm, one by one, is performed, and the result has been evaluated. It is concluded that logistic regression is the best algorithm for review analysis of the Google Play Store applications. The results have been checked scientifically, and it is found that the accuracy of the logistic regression algorithm for analyzing different reviews based on three classes, i.e., positive, negative, and neutral.

Keywords: Text Mining, Semantic Analysis, Machine Learning, Natural Language Processing, Corpus, Google Play Store

1 Introduction

In natural language processing, classifying documents and strings into different categories is considered a vital task in the process. For organizing, the online information text classification gained an important role nowadays. In literature, the authors have used text classification of an email as spam for detecting user's sentiments of comments or tweets [1]. In text classification, it is difficult to conduct automatic tagging of customer queries, classification of blogs in different categories, and dealing with the small training dataset. More specifically, the learners find that text classification is extremely challenging for generalizing. In this

***Corresponding Author: Hamza Aldabbas:** Prince Abdullah bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, Salt-Jordan; Email: aldabbas@bau.edu.jo

Abdullah Bajahzar: Department of Computer Science and Information, College of Science at Zulfi, Majmaah University, Zulfi 11932, Saudi Arabia

Meshrif Alruily: Faculty of Computer and information sciences, Jouf University, Sakaka, Saudi Arabia

Ali Adil Qureshi: Department of Computer Science Khawaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan

Rana M. Amir Latif, Muhammad Farhan: Department of Computer Science COMSATS University Islamabad, Sahiwal Campus

research, different machine learning algorithms were used for Google Play categories, and classifications of text mining were used for android application reviews [2].

For mobile devices within the few clicks, the Google Play Store or application distribution platform that allows users to deploy, buy, and search software applications. In-text reviews, these platforms allow users to share their ratings and reviews about the application [3]. For example, for the specific application, they express their satisfaction or request a new feature. There is some information about the application in reviews, which is more useful for the analysis and application designer, such as documentation, feature reports, and bug reports of user experiences for a specific application of features. These reviews on the application can be present as the “Voice of the Users” that can be more helpful for the development effort and improve future release applications [4].

Several limitations prevent the development team and analysts from using the information in the reviews. Firstly, a considerable effort is required for analyzing many reviews. The latest analysis by the authors in [5] discovered that iPhone users usually put 22 reviews on average per day. A remarkably popular application like Facebook receives significantly more than 4000 reviews per day. Secondly, the standard of the reviews fluctuates extensively from useful ideas and advanced thoughts to insulting reviews. Third, a review typically contains a sentiment mix regarding different app features, which makes it hard, e.g., filtering positive and negative reviews or retrieve the exact reviews for specific features. The usefulness of this star rating from the reviews is restricted to progress teams as a score reflects a mean for that entire app and indeed will combine both positive and negative evaluation of their unique features [6].

Text mining, which is defined as the process by which high-quality information from text is derived, is also known as text data mining. High-quality information is a statistical pattern learning which is derived from the patterns and trends through different means [7]. In the text mining, the input text structuring process normally uses parsing in line with the addition of certain linguistic features that have already been derived and subsequently inserted into a database before output being eventually evaluated and interpreted [8]. Therefore, the high-quality in text mining usually represents the interest, relevance, and novelty combination. Typically, text mining task includes entity relation modeling, for example, learning about the relationship with named entities, document summarization, sentiment analysis, production of granular taxonomies, concept or entity extraction, text clustering, and text categorization [9].

For analyzing Google Play Store reviews semantics, we are comparing the different parameters with different machine learning algorithms and find the best algorithm that we can use for the analysis of semantic analysis of Google Play Store reviews. We calculate different parameters like accuracy, recall, F1 score, and precision with Bigram, Trigram, and N-gram [10]. In this sense, a bigram or diagram refers to two closed sequenced elements from a string of tokens such as typical words, syllables, or letters. A bigram is an n-gram for $n=2$. The distribution of frequency for every string bigram is mostly used for the text simple statistical analysis in many applications including the cryptography, speech recognition, and computational linguistics [11]. Trigrams, which are a case of the n-gram, are often used in natural processing of language to perform statistical analysis of texts, and to control and use ciphers and codes in cryptography. In the probability and computational linguistics fields, an n-gram is an adjacent sequence of n items from a specific sample of textual speech or content. As per each application, the items can be words base pairs, syllables, letters, or phonemes. The n-grams, which may also be called shingles in case of words, are typically collected from a speech corpus or text [12].

2 Literature Review

Downloading and employing mobile applications by billions of people around the world has increased rapidly these days due to the recent wide-spread of the easy-to-use stores such as the Apple, Google Play and Windows phone. It is believed that fragmentation relevant to mobile platforms such that of Apple iOS, Windows cell phone, and Android represents an absolute fascinating challenge in the progress of mobile apps. Not too long ago, businesses such as Adobe, IBM, and a growing network of programmers have advocated the development

of hybrid apps as a potential remedy to such trouble in the industry. Apps of the Hybrid phone are evolved steadily with their platforms and assembled on specifications of the web [13]. The authors, in this paper, evaluate the portable hybrid apps empirically for the aim of highlighting and investigating the potential and exceptional qualities of the openly offered hybrid apps as perceived by users and their related reviews. The analysis was conducted by mining 11,917 free applications and 3,041,315 reviews obtained from the Google Play Store and assessed according to the perspectives of the end-users. Consequently, the analysis built on an object and reproducible representation of the way by which the development of the hybrid mobile was performing "from the great outdoors" found in genuine reviews, thereby setting a foundation for prospective procedures and methods for establishing hybrid apps [14, 15].

User review is an essential part in the markets of the open mobile applications such as the Google Play Store. How is it possible to automatically combine countless reviews of users and produce a concrete sense from them? However, unfortunately, few analytic tools can provide into user reviews beyond simple summaries like user ratings histograms, [16]. This paper suggests the Wiscoma system, which may test hundred and thousands of users reviews along with opinions from mobile apps markets at about three distinct heights of depth. Authors suggest that their system can (a) find inconsistencies in reviews; (b) recognize causes why users dislike or like that specific app, supply a zoomable interactive perspective of users' review; and (c) present important insights into the whole app markets. This proposal applies to different types of apps that identify users significant preferences and concerns [17]. Results with the purposed system will be reported to the 32 GB dataset that is composed of over 13 million users' reviews for 171,493 Android applications in the Google Play Store. The author discusses how this proposed system can help mobile applications market operators such as end-users, individual app developers, and Google [18, 19].

Unlike services and products in Amazon.com, mobile apps are always evolving, with all new versions speedily changing the previous versions. Many app stores even now utilize an Amazon-style rating technique, which aggregates just about every rating ever assigned to an app into a store rating. The author mined 10000 mobile application store ratings from Google Play Store to examine the user's satisfaction level. Even though many applications rating designs were challenging to variate when these applications had gathered a considerable number of raters. The conclusion of this research that the current systems running in the market cannot analyze the user satisfaction levels that can discourage developers from improving the quality of the application [20].

Now, using apps has increased together with the rising craze in the direction of mobiles. The end-users will prefer mobile phones to get several types of mobile app for different purposes. The user will download the app by checking the number of downloads of that particular app [21]. What would be the reviews and ratings? What would be the comments? Users download mobile applications. In the mobile application market, the fraud ranking of the application is an illegal activity that is used to push up the mobile application in the list of the popularity of the application. The application developer uses this fake mechanism periodically in the different application development process [22].

Research on mining user reviews in mobile application stores has progressed in the last couple of decades. Most of the suggested methods count on optimizing the meta description of reading user reviews to different kinds of educational user prerequisites along with uninformative suggestions. Determined by the essential characteristics of reviews regularly produces high-dimensional variations. That raises the intricacy of the classifier also may cause overfitting issues. Authors suggest a publication recruitment tactic for apps inspection classification [23].

3 Methodology for Google Play Content Scraping and Knowledge Engineering

In the process of classification, starting with the scraping of reviews on applications. On Google Play Store using the App ID request for scrap, the reviews of that specific application scrap several pages with reviews and ratings of the applications. In the next step, apply Bigram, Trigram, and N-gram on reviews by using a

python language often used in information retrieval and text mining. After applying Bigram, Trigram, and N-gram extract different features of each application. By using a python, a different algorithm is used for the classification of Naïve Bayes Multinomial, Random forest, and logistic regression, in addition to setting the different parameters such as Precision, Recall, Accuracy, and F1 score and finding the statistics of these parameters. After analyzing and testing, this statistical information analyzes which algorithm has a maximum Precision, Recall, Accuracy, and F1 score information and analyzes that is best for the analysis of reviews classification, as shown in Figure 1.

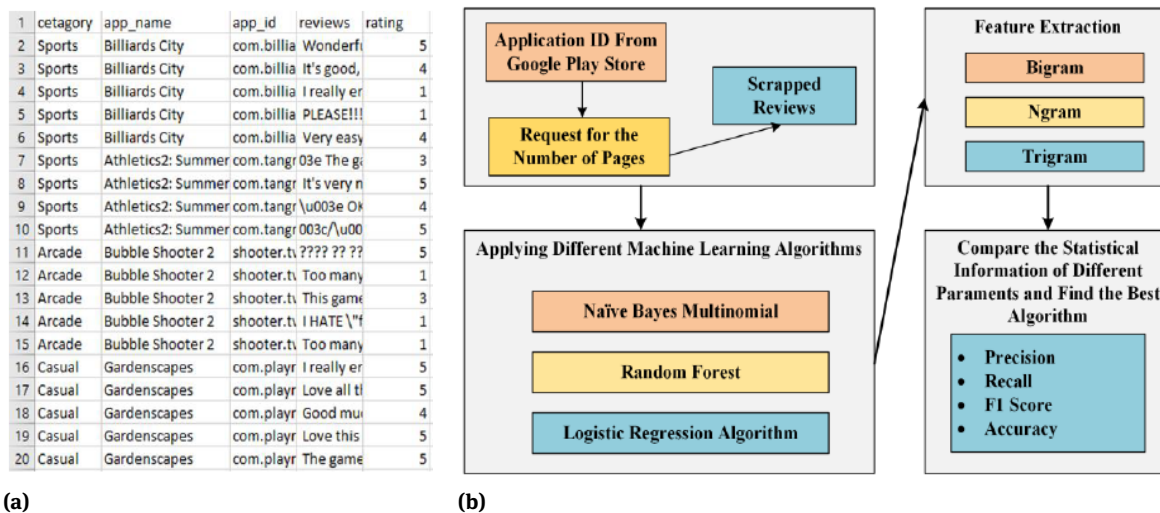


Figure 1: The view of Dataset and System Architecture (a) Sample screenshot of the dataset (b) methodology diagram of reviews analysis of Google Play Store

4 Data Collection Process

Mobile applications are part of our lives. According to a report, half a million applications were introduced in 2011, and in October 2012, 0.675 million applications were available on the Google Play Store. In our daily life that people use Android apps mostly. Now a day's Android app is being used by every one of us; people use different Android apps, like messaging, social media, gaming, and browsers. This online marketplace provides mobile users with both free and paid access to over a million mobile applications, also refers as "mobile apps" On the Google Play Store website, users can choose from over a million mobile apps for various datasets with predefined categories. Data collection always plays a vital role in every research, and the validity and accuracy of the dataset is also a significant part of any dataset collection process. In this research, we have scraped the hundreds and thousands of user's reviews and ratings of different applications of different categories, as shown in Figure 1. In the start, we have selected different categories of Google Play Store. After choosing different 14 categories of Google Play Store, different scrap application of each category that is shown in Table 1. These categories of applications are Action, Arcade, Card, Communication, Finance, Health and Fitness, Photography, Shopping, Sports, Video Player Editor, Weather, Casual, Medical, and Racing. We have scraped 506259 reviews from 14 different categories of Google Play Store application, as indicated in Table 1.

Table 1: Detailed description of a dataset statistics scraped from Google Play Store

Application Category	Total Reviews	Application Category	Total Reviews
Action	47116	Shopping	43370
Arcade	36521	Sports	33770
Card	37761	Video Player Editor	20791
Communication	30010	Weather	27334
Finance	28233	Casual	52572
Health and Fitness	34425	Medical	24012
Photography	41450	Racing	42394

5 Basic Text Pre-processing of reviews

1. Lower casing
2. Stop words removal
3. Frequent words removal
4. Rare words removal
5. Spelling correction
6. Tokenization
7. Stemming
8. Lemmatization

6 Results and Discussion

This section addresses the evaluation of the scraped dataset by using different machine learning algorithm like Logistics Regression Algorithm, Naïve Bayes Multinomial, and Random Forest Algorithm. The Bigram, Trigram, and N-gram to were evaluated to find out the best algorithm on the basis of precision, recall, accuracy, and F1 score

6.1 Logistics Regression Algorithm for Bigram, Trigram, N-gram

Logistic regression is the statistical model used to model a binary dependent variable. This model is estimating the population parameter (which is a quantity entering into the probability distribution of a statistic of the logistic model). The logistic regression algorithm has been applied in the form of binomial regression. We have scraped 506259 reviews from 14 different categories of Google Play Store application. We have applied a logistic regression algorithm on different population parameter concerning Bigram, Trigram, and N-gram. Find the accuracy of classification of each category application and in statistical information find precision, recall, and F1 score these all parameters we use to measure the accuracy of the dataset is shown in Table 2. Figure 2 views Bar chart of Logistics Regression Algorithm for different precision, recall, f1, and accuracy by using Bigram, Trigram, and N-gram.

6.2 Naïve Bayes Multinomial for Bigram, Trigram, N-gram

Naïve Bayes Multinomial used for classification which with the high dimensional dataset. In this algorithm, certain features are dependent on the occurrence of other features. This model is fast to make predictions. We have scraped 506259 reviews from 14 different categories of Google Play Store application. We have

Table 2: Statistical information of Logistic Regression Tree for different parameters by using Bigram, Trigram, N-gram

Category of Applications	Bigram of Logistic Regression Tree				Trigram of Logistic Regression Tree				N-gram of Logistic Regression Tree			
	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy
Medical	0.69	0.77	0.70	0.7659	0.69	0.76	0.70	0.7642	0.68	0.76	0.69	0.7613
Action	0.62	0.71	0.64	0.7107	0.61	0.70	0.63	0.7026	0.60	0.69	0.61	0.6939
Arcade	0.66	0.75	0.68	0.7504	0.65	0.75	0.67	0.7480	0.65	0.75	0.67	0.7463
Card	0.64	0.71	0.65	0.7138	0.63	0.71	0.64	0.7083	0.61	0.70	0.63	0.7009
Casual	0.69	0.77	0.71	0.7680	0.69	0.77	0.70	0.7650	0.67	0.76	0.69	0.7597
Communication	0.54	0.62	0.56	0.6166	0.53	0.61	0.55	0.6106	0.53	0.61	0.54	0.606
Finance	0.58	0.65	0.59	0.6452	0.56	0.63	0.57	0.6328	0.55	0.62	0.56	0.6211
Health & fitness	0.78	0.83	0.80	0.8323	0.79	0.83	0.80	0.8341	0.78	0.83	0.80	0.8294
Photography	0.66	0.73	0.68	0.7340	0.66	0.73	0.67	0.7309	0.65	0.73	0.66	0.7256
Racing	0.67	0.74	0.68	0.7449	0.66	0.74	0.66	0.7374	0.65	0.73	0.66	0.7343
Shopping	0.60	0.68	0.60	0.6751	0.60	0.67	0.59	0.6659	0.60	0.66	0.59	0.6632
Sports	0.58	0.65	0.60	0.6548	0.57	0.64	0.58	0.6449	0.56	0.64	0.57	0.6409
Video players & editors	0.59	0.69	0.62	0.6931	0.58	0.68	0.61	0.6849	0.58	0.69	0.61	0.6854
Weather	0.61	0.68	0.62	0.6838	0.61	0.68	0.61	0.6798	0.60	0.68	0.61	0.6772

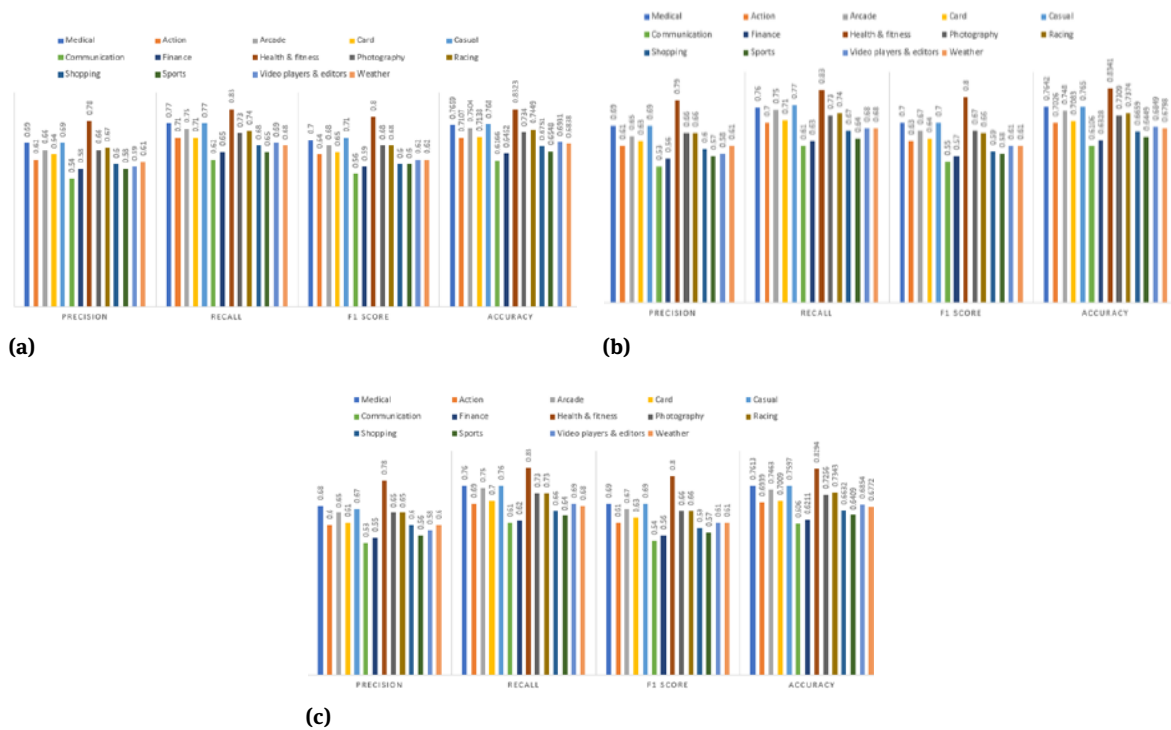


Figure 2: Bar chart visualization of Logistic Regression Tree for different parameters by using (a) Bigram (b) Trigram (c) N-gram

applied a Naïve Bayes Multinomial algorithm on different population parameter concerning Bigram, Trigram, and N-gram. Find the accuracy of classification of each category application and in statistical information find precision, recall, and F1 score these all parameters we use to measure the accuracy of the dataset is shown in Table 3. Figure 3 views Bar chart of Naïve Bayes Multinomial for different precision, recall, f1, and accuracy by using Bigram, Trigram, and N-gram.

6.3 Random Forest Algorithm for Bigram, Trigram, N-gram

Random forest classifier is majorly used for decision tree. Many decision trees can develop on the bases of a random selection of datasets and variables. We have scraped 506259 reviews from 14 various categories found in Google Play Store application. The random forest algorithm has been applied on different population parameter concerning Bigram. Find the accuracy of classification of each category application and in statistical information find precision, recall, and F1 score these all parameters we use to measure the accuracy of the dataset is shown in Table 4. Figure 4 views Bar chart of Random Forest Algorithm for different precision, recall, f1, and accuracy by using Bigram, Trigram, and N-gram.

6.4 Comparison of Different Machine Learning Algorithms using Bigram

This online marketplace provided free and paid access to users. On the Google Play Store, users can choose from over a million apps from various predefined categories. In this research, we have scraped 506259 reviews from 14 different categories of Google Play Store application. Evaluated the results by using different machine learning algorithms like Naïve Bayes Multinomial, Random Forest, and Logistic Regression algorithm on different parameters concerning Bigram, Trigram, and N-gram. That can check the semantics of reviews about some applications form users that their reviews are good, bad, normal, and so on. Calculated to Bigram, Tri-

Table 3: Statistical information of Naïve Bayes Multinomial for different parameters by using Bigram, Trigram, N-gram

Category of Applications	Bigram of Naïve Bayes Multinomial				Trigram of Naïve Bayes Multinomial				N-gram of Naïve Bayes Multinomial			
	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy
Medical	0.70	0.76	0.72	0.7567	0.70	0.75	0.71	0.7546	0.69	0.76	0.70	0.7555
Action	0.64	0.70	0.65	0.6987	0.62	0.69	0.64	0.6924	0.60	0.69	0.61	0.6856
Arcade	0.66	0.73	0.68	0.7315	0.65	0.73	0.68	0.7325	0.64	0.74	0.67	0.7369
Card	0.63	0.71	0.66	0.7059	0.63	0.70	0.65	0.7042	0.62	0.70	0.63	0.7018
Casual	0.70	0.75	0.72	0.7507	0.69	0.75	0.71	0.7545	0.69	0.76	0.71	0.7579
Communication	0.57	0.63	0.59	0.629	0.55	0.62	0.57	0.6173	0.54	0.61	0.55	0.6123
Finance	0.59	0.63	0.60	0.6272	0.58	0.63	0.59	0.6261	0.56	0.61	0.57	0.6123
Health & fitness	0.79	0.82	0.80	0.8172	0.79	0.82	0.80	0.8192	0.79	0.82	0.80	0.8224
Photography	0.66	0.72	0.68	0.7224	0.66	0.72	0.68	0.7239	0.65	0.72	0.67	0.7224
Racing	0.67	0.73	0.69	0.7341	0.66	0.73	0.68	0.7324	0.65	0.73	0.66	0.7301
Shopping	0.61	0.66	0.62	0.6563	0.61	0.66	0.62	0.6567	0.59	0.65	0.59	0.6484
Sports	0.60	0.66	0.62	0.6555	0.60	0.66	0.61	0.6558	0.56	0.64	0.57	0.6384
Video players & editors	0.62	0.69	0.64	0.6883	0.60	0.68	0.63	0.6844	0.60	0.68	0.62	0.6849
Weather	0.62	0.68	0.64	0.6824	0.62	0.68	0.64	0.6842	0.61	0.68	0.62	0.6827

Table 4: Statistical information of Random Forest for different parameters by using Bigram, Trigram, N-gram

Category of Applications	Bigram of Random Forest				Trigram of Random Forest				N-gram of Random Forest			
	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy
Medical	0.68	0.75	0.70	0.7538	0.67	0.75	0.69	0.7492	0.66	0.75	0.69	0.7455
Action	0.60	0.69	0.63	0.6896	0.60	0.68	0.62	0.6833	0.58	0.68	0.61	0.6756
Arcade	0.65	0.74	0.68	0.7374	0.64	0.74	0.67	0.7352	0.64	0.73	0.67	0.7340
Card	0.62	0.69	0.64	0.6919	0.59	0.67	0.61	0.6746	0.59	0.68	0.62	0.6815
Casual	0.66	0.74	0.69	0.7442	0.67	0.74	0.69	0.7414	0.65	0.74	0.68	0.7385
Communication	0.54	0.61	0.56	0.606	0.52	0.60	0.54	0.600	0.51	0.59	0.54	0.593
Finance	0.54	0.60	0.56	0.6048	0.55	0.61	0.57	0.6141	0.53	0.60	0.56	0.5992
Health & fitness	0.75	0.82	0.78	0.8172	0.74	0.81	0.77	0.8149	0.74	0.82	0.77	0.8166
Photography	0.65	0.72	0.67	0.7193	0.63	0.71	0.66	0.7123	0.64	0.72	0.67	0.7179
Racing	0.65	0.73	0.67	0.7254	0.64	0.72	0.66	0.7206	0.62	0.72	0.65	0.7152
Shopping	0.59	0.66	0.60	0.6611	0.59	0.66	0.60	0.6576	0.57	0.65	0.59	0.6484
Sports	0.57	0.63	0.59	0.6332	0.56	0.62	0.57	0.6220	0.56	0.63	0.58	0.6267
Video players & editors	0.60	0.68	0.62	0.6796	0.59	0.67	0.61	0.6685	0.59	0.68	0.61	0.6767
Weather	0.58	0.65	0.61	0.6505	0.59	0.66	0.61	0.6586	0.58	0.65	0.60	0.6523

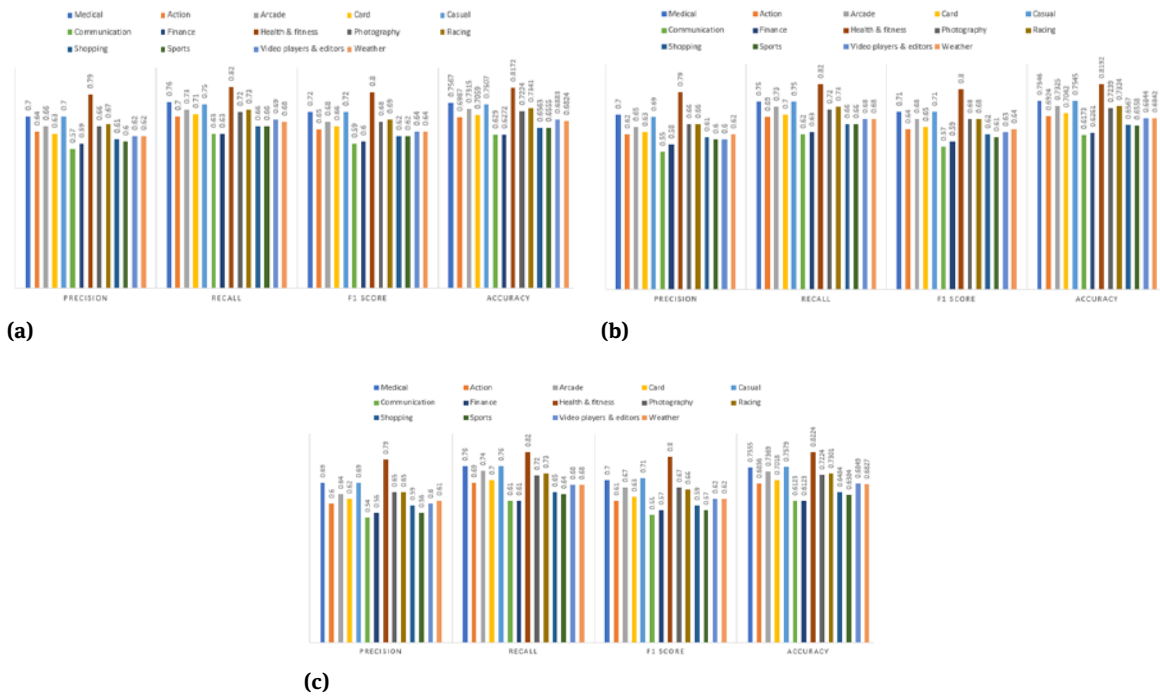


Figure 3: Bar chart visualization of Naïve Bayes Multinomial for different parameters by using (a) Bigram (b) Trigram (c) N-gram

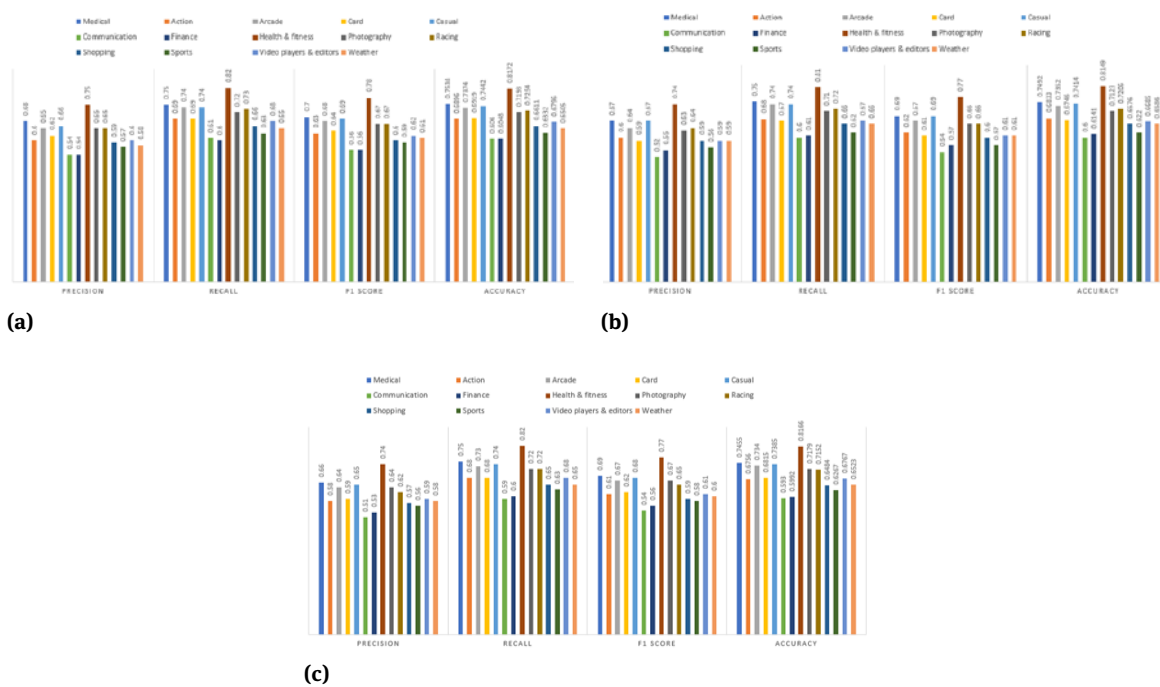


Figure 4: Bar chart visualization of Random Forest for different parameters by using (a) Bigram (b) Trigram (c) N-gram

gram, and N-gram with different parameters like accuracy, precision, recall, and F1 score, the concluded results were compared to the statistical result of the algorithms. Visualized these statistical results in the form of a bar chart, as shown in Figure 5 to Figure 7. After comparison, analyzed that the logistic regression algo-

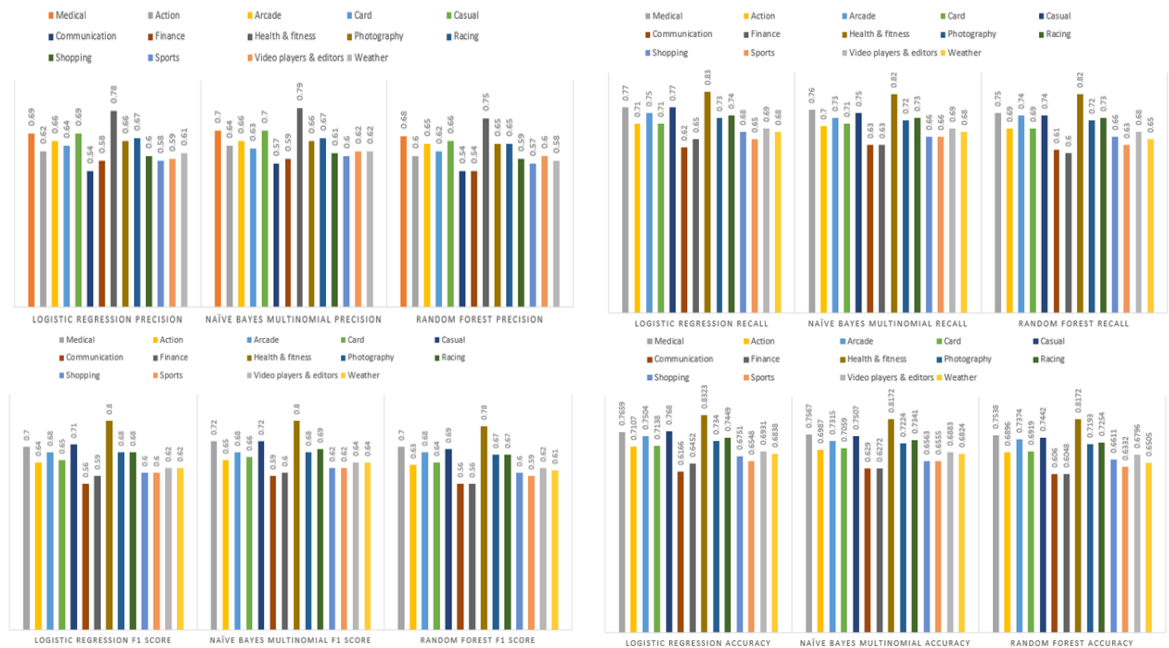


Figure 5: Bar chart Visualization of concluded results with different machine learning algorithms for different parameters by using Bigram

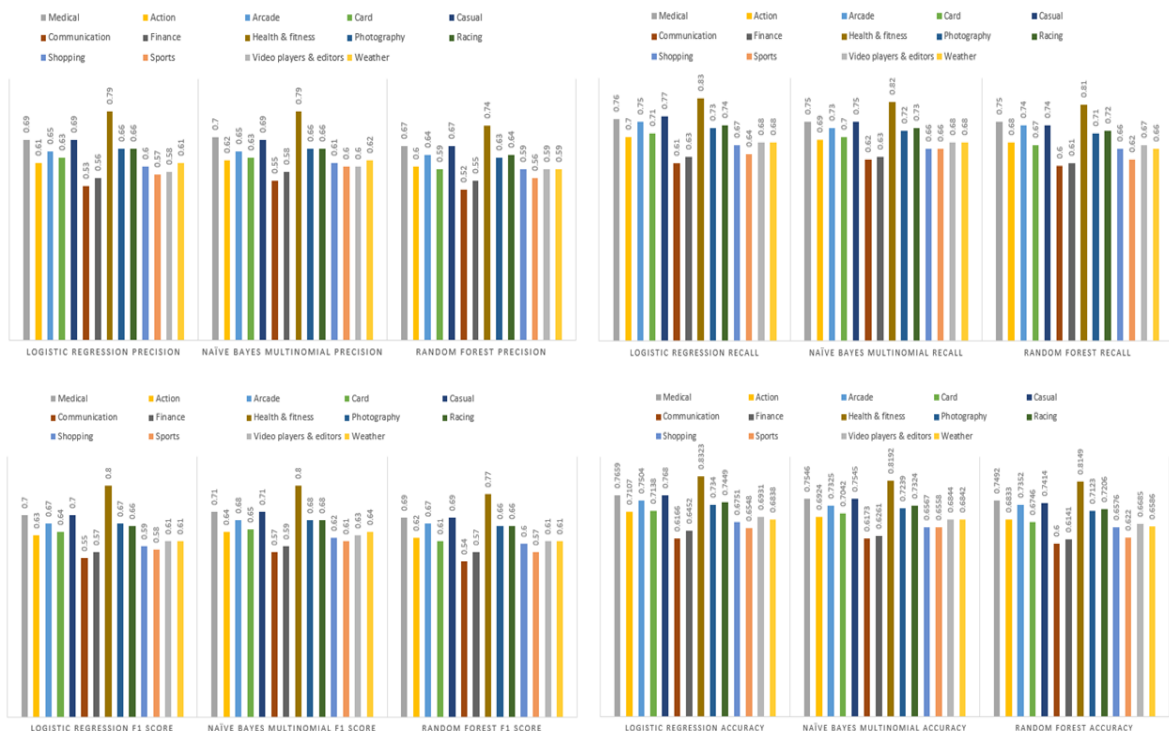


Figure 6: Bar chart Visualization of concluded results with different machine learning algorithms for different parameters by using Trigram

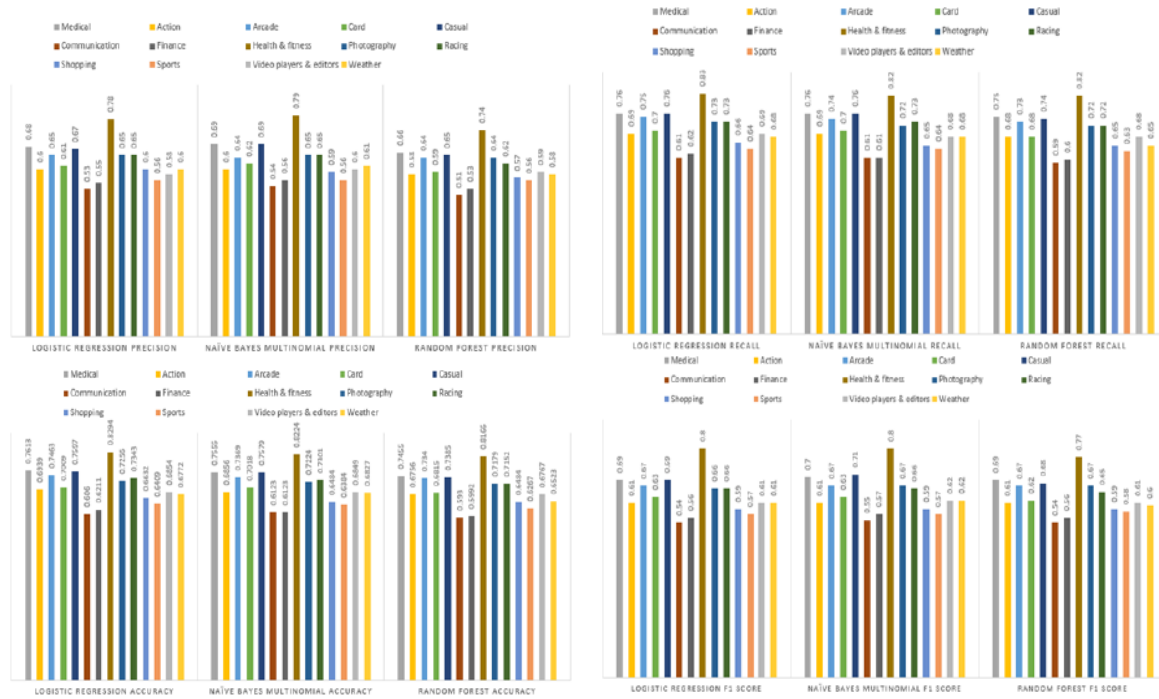


Figure 7: Bar chart Visualization of concluded results with different machine learning algorithms for different parameters by using N-gram

rithm is the best algorithm for checking the semantic analysis of any Google application users' reviews, as shown in Table 5 to Table 7.

Table 5: Different machine learning algorithms for different parameters by using Bigram

Application Category	Logistic Regression Precision	Naïve Bayes Multinomial Precision	Random Forest Precision	Logistic Regression Recall	Naïve Bayes Multinomial Recall	Random Forest Recall	Logistic Regression F1 Score	Naïve Bayes Multinomial F1 Score	Random Forest F1 Score	Logistic Regression Accuracy	Naïve Bayes Multinomial Accuracy	Random Forest Accuracy
Medical	0.69	0.70	0.68	0.77	0.76	0.75	0.70	0.72	0.70	0.7659	0.7567	0.7538
Action	0.62	0.64	0.60	0.71	0.70	0.69	0.64	0.65	0.63	0.7107	0.6987	0.6896
Arcade	0.66	0.66	0.65	0.75	0.73	0.74	0.68	0.68	0.68	0.7504	0.7315	0.7374
Card	0.64	0.63	0.62	0.71	0.71	0.69	0.65	0.66	0.64	0.7138	0.7059	0.6919
Casual	0.69	0.70	0.66	0.77	0.75	0.74	0.71	0.72	0.69	0.7680	0.7507	0.7442
Communication	0.54	0.57	0.54	0.62	0.63	0.61	0.56	0.59	0.56	0.6166	0.629	0.606
Finance	0.58	0.59	0.54	0.65	0.63	0.60	0.59	0.60	0.56	0.6452	0.6272	0.6048
Health & fitness	0.78	0.79	0.75	0.83	0.82	0.82	0.80	0.80	0.78	0.8323	0.8172	0.8172
Photography	0.66	0.66	0.65	0.73	0.72	0.72	0.68	0.68	0.67	0.7340	0.7224	0.7193
Racing	0.67	0.67	0.65	0.74	0.73	0.73	0.68	0.69	0.67	0.7449	0.7341	0.7254
Shopping	0.60	0.61	0.59	0.68	0.66	0.66	0.60	0.62	0.60	0.6751	0.6563	0.6611
Sports	0.58	0.60	0.57	0.65	0.66	0.63	0.60	0.62	0.59	0.6548	0.6555	0.6332
Video players & editors	0.59	0.62	0.60	0.69	0.69	0.68	0.62	0.64	0.62	0.6931	0.6883	0.6796
Weather	0.61	0.62	0.58	0.68	0.68	0.65	0.62	0.64	0.61	0.6838	0.6824	0.6505

Table 6: Different machine learning algorithms for different parameters by using Trigram

Application Category	Logistic Regression Precision	Naïve Bayes Multinomial Precision	Random Forest Precision	Logistic Regression Recall	Naïve Bayes Multinomial Recall	Random Forest Recall	Logistic Regression F1 Score	Naïve Bayes Multinomial F1 Score	Random Forest F1 Score	Logistic Regression Accuracy	Naïve Bayes Multinomial Accuracy	Random Forest Accuracy
Medical	0.69	0.70	0.67	0.76	0.75	0.75	0.70	0.71	0.69	0.7659	0.7546	0.7492
Action	0.61	0.62	0.60	0.70	0.69	0.68	0.63	0.64	0.62	0.7107	0.6924	0.6833
Arcade	0.65	0.65	0.64	0.75	0.73	0.74	0.67	0.68	0.67	0.7504	0.7325	0.7352
Card	0.63	0.63	0.59	0.71	0.70	0.67	0.64	0.65	0.61	0.7138	0.7042	0.6746
Casual	0.69	0.69	0.67	0.77	0.75	0.74	0.70	0.71	0.69	0.7680	0.7545	0.7414
Communication	0.53	0.55	0.52	0.61	0.62	0.60	0.55	0.57	0.54	0.6166	0.6173	0.600
Finance	0.56	0.58	0.55	0.63	0.63	0.61	0.57	0.59	0.57	0.6452	0.6261	0.6141
Health & fitness	0.79	0.79	0.74	0.83	0.82	0.81	0.80	0.80	0.77	0.8323	0.8192	0.8149
Photography	0.66	0.66	0.63	0.73	0.72	0.71	0.67	0.68	0.66	0.7340	0.7239	0.7123
Racing	0.66	0.66	0.64	0.74	0.73	0.72	0.66	0.68	0.66	0.7449	0.7324	0.7206
Shopping	0.60	0.61	0.59	0.67	0.66	0.66	0.59	0.62	0.60	0.6751	0.6567	0.6576
Sports	0.57	0.60	0.56	0.64	0.66	0.62	0.58	0.61	0.57	0.6548	0.6558	0.6220
Video players & editors	0.58	0.60	0.59	0.68	0.68	0.67	0.61	0.63	0.61	0.6931	0.6844	0.6685
Weather	0.61	0.62	0.59	0.68	0.68	0.66	0.61	0.64	0.61	0.6838	0.6842	0.6586

Table 7: Different machine learning algorithms for different parameters by using N-gram

Application Category	Logistic Regression Precision	Naïve Bayes Multinomial Precision	Random Forest Precision	Logistic Regression Recall	Naïve Bayes Multinomial Recall	Random Forest Recall	Logistic Regression F1 Score	Naïve Bayes Multinomial F1 Score	Random Forest F1 Score	Logistic Regression Accuracy	Naïve Bayes Multinomial Accuracy	Random Forest Accuracy
Medical	0.68	0.69	0.66	0.76	0.76	0.75	0.69	0.70	0.69	0.7613	0.7555	0.7455
Action	0.60	0.60	0.58	0.69	0.69	0.68	0.61	0.61	0.61	0.6939	0.6856	0.6756
Arcade	0.65	0.64	0.64	0.75	0.74	0.73	0.67	0.67	0.67	0.7463	0.7369	0.7340
Card	0.61	0.62	0.59	0.70	0.70	0.68	0.63	0.63	0.62	0.7009	0.7018	0.6815
Casual	0.67	0.69	0.65	0.76	0.76	0.74	0.69	0.71	0.68	0.7597	0.7579	0.7385
Communication	0.53	0.54	0.51	0.61	0.61	0.59	0.54	0.55	0.54	0.606	0.6123	0.593
Finance	0.55	0.56	0.53	0.62	0.61	0.60	0.56	0.57	0.56	0.6211	0.6123	0.5992
Health & fitness	0.78	0.79	0.74	0.83	0.82	0.82	0.80	0.80	0.77	0.8294	0.8224	0.8166
Photography	0.65	0.65	0.64	0.73	0.72	0.72	0.66	0.67	0.67	0.7256	0.7224	0.7179
Racing	0.65	0.65	0.62	0.73	0.73	0.72	0.66	0.66	0.65	0.7343	0.7301	0.7152
Shopping	0.60	0.59	0.57	0.66	0.65	0.65	0.59	0.59	0.59	0.6632	0.6484	0.6484
Sports	0.56	0.56	0.56	0.64	0.64	0.63	0.57	0.57	0.58	0.6409	0.6384	0.6267
Video players & editors	0.58	0.60	0.59	0.69	0.68	0.68	0.61	0.62	0.61	0.6854	0.6849	0.6767
Weather	0.60	0.61	0.58	0.68	0.68	0.65	0.61	0.62	0.60	0.6772	0.6827	0.6523

6.5 Semantic Analysis of Google Play Store Applications Reviews using Logistic Regression Algorithm

After checking the different population parameter, analyze that the logistic regression algorithm is the best algorithm having the highest accuracy. In this section, we performed analysis and classify all reviews in different classes positive, negative, and neutral. Set target value if the value of the comment is positive, it is equal to 1 if the review is negative, it is equal to 0. Also, analyze the neutral class with the confidence rate if the confidence rate is between the 0 and 1 then classify this to neutral class. Different parameters in our dataset like the category of application, Application Name, Application ID, Reviews, and rating, as shown in Figure 8. However, for checking the semantics of each review, these parameters are more enough.

1	Cetagory	App_Name	App_ID	Reviews	Rating
2	Sports	Billiards City	com.billiards.city.pool.nation.club	Wonderfull App. Completed all 1020 levels, Can'	5
3	Sports	Billiards City	com.billiards.city.pool.nation.club	It's good, I like the gameplay. Please change up t	4
4	Sports	Billiards City	com.billiards.city.pool.nation.club	I really enjoyed this game until I saw one of the i	1
5	Sports	Billiards City	com.billiards.city.pool.nation.club	PLEASE!!! Get rid of the odd shaped tables and g	1
6	Communication	Hangouts Dialer	com.google.android.apps.hangoutsdialer	st from the notification menu\u003c\u003e Oth	3
7	Communication	Hangouts Dialer	com.google.android.apps.hangoutsdialer	Wish i found it earlier!!!	5
8	Communication	Hangouts Dialer	com.google.android.apps.hangoutsdialer	sage\u003c\u003e I love it i dont even have to p	5
9	Communication	Hangouts Dialer	com.google.android.apps.hangoutsdialer	Wanted to make video calling as moto g phone d	1
10	Arcade	Leps World 2	at.ner.lepsWorld2	It is a good time disaster	5
11	Arcade	Leps World 2	at.ner.lepsWorld2	It is so nise i have never seen before	5
12	Arcade	Leps World 2	at.ner.lepsWorld2	I played it totally more than 3 times	5
13	Arcade	Leps World 2	at.ner.lepsWorld2	Is awesome cool game love it some time you lov	5
14	Video Players & Editors	Youtube	com.google.android.youtube	Excellent App	5
15	Video Players & Editors	Youtube	com.google.android.youtube	Very nice	5
16	Video Players & Editors	Youtube	com.google.android.youtube	Very good	4
17	action	WARSHIP BATTLE	com.joycity.warshipbattle	The best at all features this game... Very nice....	5
18	action	WARSHIP BATTLE	com.joycity.warshipbattle	I love this game good work	5
19	action	WARSHIP BATTLE	com.joycity.warshipbattle	I like it	4
20	Weather	NOAA Weather Radar & Alerts	com.apalon.weatherradar.free	Just better than the rest, period.	5
21	Weather	NOAA Weather Radar & Alerts	com.apalon.weatherradar.free	Great tool for the road or around town.	5
22	photography	Sweet Selfie	com.cam001.selfie	It's a very good app	5

Figure 8: Sample screenshot of the original dataset that scrapped

7 Data Preparation Step

7.1 HTML Decoding

To convert HTML encoding into text, and in the start or ending up in the text field as '&' '\amp' & 'quot.'

7.2 Data Preparation 2: '#' hashtag

“#” carries import information that must deal with is necessary.

7.3 Uniform Recourse Locator (URL) links

In this step URL must be removed.

7.4 UTF-8 BOM (Byte Order Mark)

For characters patterns like “\xef\xbf\xbd,” these are UTF-8 BOM. It is a sequence of bytes (EF BB BF) which helps the reader identify a file encoded in UTF-8.

7.5 Hashtag / Numbers

A hashtag text can refer to the useful information on the comment. It is possible that it is tough to remove the whole text together by using the “#” or with a number or with any other unique character needs to accommodate.

7.6 Negation Handling

~ is the factor that is not suitable in the review remove them.

7.7 Tokenizing and Joining

Parse the whole comment into small pieces and then merge again. After applying the above rules on cleaning, the reviews cleaned formed of reviews.

7.8 Find Null Entries from the Reviews

In order to remove the noises and inconstant from data, the null value needs to be removed.

```
<class 'pandas.core.frame.DataFrame'>  
Int64Index: 400000 entries, 0 to 399999  
Data columns (total of 2 columns):  
text → 399208 non-null object  
target → 400000 non-null int64  
dtypes: int64(1), object (1)  
memory usage: 9.2 + MB
```

7.9 Negative and Positive Words Dictionary

By using word cloud corpus, we have created a dictionary contains a positive and negative words on the basis of words occurrence in a text to get the idea of what kind of words are frequent in the corpus, as shown in Figure 9.

7.10 The Semantic Analysis of Reviews using Logistic Regression Algorithm

In the result, we classified all reviews into three different classes and we checked the confidence rate of each rate that how much that comment is positive, negative, and neutral. Set the target value equal to 0 to 1 and check the confidence value in that ratio and check the class of the review using the logistic regression algorithm, as shown in Figure 10.



Figure 9: (a) Positive (b) Negative word dictionary by using the word cloud corpus

1	Reviews	Tags	Confidence	Target
2	wonderfull app completed all levels can not wait for more levels level has bug but you can get around it to complete it by not tou	Positive	0.98	1
3	it good like the gameplay please change up the music as it gets repetitive after the nd level and you can hear the track loop sta	Positive	0.833	1
4	really enjoyed this game until saw one of the adverts throughout the ngame with man and woman spooning it is hardly appropria	Negative	0.997	0
5	please get rid of the odd shaped tables and go back to the classic table please fix level where the balls and cue stick disappear	Negative	0.999	0
6	very easy game to play and has actually given me pointers on how to play nreal game of pool like where to hit the cue ball and	Positive	0.966	1
7	not going to lie just started playing this game probably about hours ago and am thoroughly addicted it pretty awesome game m	Positive	0.956	1
8	level the lower left cushion let ball disappear into nothingness nafter striking the ball it can be seen at different edges of the scre	Neutral	0.546	0
9	people balls do not stop at the same time it wont let the ball go in the pocket that has hand in it the table control the game que	Negative	0.911	1
10	the gameplay is fun graphics are good and there are lot of levels however the number of ads are insane is there paid version whe	Negative	0.687	1
11	great game until you get above level than it starts shooting balls off screen where can not see them or play anymore really have	Negative	0.998	1
12	just wanted to thank the supervisors creator those in charge whatever powers that be that after days of my complaint am no lon	Positive	0.608	1
13	as pool player this is good app find it helps me with using different english techniques control of the ball as well as learning varic	Positive	0.998	1
14	just wanted basic billiards game there was this redundant tutorial level up system that was totally unnecessary and map thing to	Negative	1	0
15	installed this game and could not even play it because the game would not load and would shut down waste of time and really v	Negative	0.998	0
16	really enjoyed the older version alot more than now some of these odd ball ntables are more aggravating than enjoyable do not k	Negative	0.639	1
17	really like the game up to certain point when the normal billiard table is changed to crooked one now its hard to play the game s	Positive	0.795	1
18	pretty decent game but has way too many ads now made me watch second video every time wanted to retry table no thanks wlf	Negative	0.977	1
19	level and til now no problem but then on this levels the balls keep disappearing under the table and from there there no way to s	Negative	0.993	1
20	first had trouble getting pass level learned it amp now on level great game finish there are levels total will go back and replay so	Negative	0.522	1

Figure 10: Final sentiment analysis results on Google Play reviews using a logistic regression algorithm

8 Conclusion and Future Work

Hundreds and thousands of apps uploaded by developers and downloaded by users are on the Google Play Store. Users use these applications for their specific purpose, and they have their personal experiences. Users download and use these applications and express the application's experience in the form of comments or reviews and give the applications a 0-5 scale rating. We have scraped 506259 reviews for 14 different categories of Google Play Store applications in this research work. We have analyzed the class of the reviews that may be positive, negative, and neutral. We have checked the application semantics with different algorithms of machine learning. We have used three different machine learning algorithms, such as Logistic Regression Algorithm, Random Forest, and Multinomial Naïve Bayes. Evaluate Bigram, Trigram, and N-gram with various parameters such as precision, accuracy, recall, and F1 score, and compared the statistical results of these algorithms. After contrast, we have evaluated that the Logistic regression algorithm is the most active algorithm with a high precision score, and we can use this machine-learning algorithm to test the user reviews.

In the future we will increase the number of categories of applications and number of reviews. We will compare the accuracy of the logistic regression algorithm with other different algorithms. We will generate the clusters and check the relationship between application reviews and ratings that can help to more accurately analyze each application.

References

- [1] Y. Goldberg, "Neural network methods for natural language processing," *Synthesis Lectures on Human Language Technologies*, vol. 10, no. 1, pp. 1-309, 2017.
- [2] N. Genc-Nayebi and A. Abran, "A systematic literature review: Opinion mining studies from mobile app store user reviews," *Journal of Systems and Software*, vol. 125, pp. 207-219, 2017.
- [3] E. Cambria, B. Schuller, Y. Xia, and B. White, "New avenues in knowledge bases for natural language processing," *Knowledge-Based Systems*, vol. 108, no. C, pp. 1-4, 2016.
- [4] R. Agerri, X. Artola, Z. Beloki, G. Rigau, and A. Soroa, "Big data for Natural Language Processing: A streaming approach," *Knowledge-Based Systems*, vol. 79, pp. 36-42, 2015.
- [5] Y. Man, C. Gao, M. R. Lyu, and J. Jiang, "Experience report: Understanding cross-platform app issues from user reviews," in *2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE)*, 2016, pp. 138-149: IEEE.
- [6] T. Denoeux, "Logistic regression, neural networks and dempster-shafer theory: a new perspective," *Knowledge-Based Systems*, 2019.
- [7] C. Gao, Y. Zhao, R. Wu, Q. Yang, and J. Shao, "Semantic trajectory compression via multi-resolution synchronization-based clustering," *Knowledge-Based Systems*, 2019.
- [8] K. Santo, S. S. Richtering, J. Chalmers, A. Thiagalingam, C. K. Chow, and J. Redfern, "Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps," *JMIR mHealth and uHealth*, vol. 4, no. 4, p. e132, 2016.
- [9] P. Barlas, I. Lanning, and C. Heavey, "A survey of open source data science tools," *International Journal of Intelligent Computing and Cybernetics*, vol. 8, no. 3, pp. 232-261, 2015.
- [10] R. Zeng and P. M. Greenfield, "Cultural evolution over the last 40 years in China: Using the Google Ngram Viewer to study implications of social and political change for cultural values," *International Journal of Psychology*, vol. 50, no. 1, pp. 47-55, 2015.
- [11] C. Li, Y. Liu, and L. Zhao, "Using external resources and joint learning for bigram weighting in ilp-based multi-document summarization," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, pp. 778-787.
- [12] A. Pagán, H. I. Blythe, and S. P. Liversedge, "Parafoveal preprocessing of word initial trigrams during reading in adults and children," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 42, no. 3, p. 411, 2016.
- [13] C. Mellish and X. Sun, "The semantic web as a linguistic resource: Opportunities for natural language generation," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 2005, pp. 77-87: Springer.
- [14] J. Schütte, R. Fedler, and D. Titze, "Condroid: Targeted dynamic analysis of android applications," in *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*, 2015, pp. 571-578: IEEE.
- [15] B. Yu, Z.-b. Xu, and C.-h. Li, "Latent semantic analysis for text categorization using neural network," *Knowledge-Based Systems*, vol. 21, no. 8, pp. 900-904, 2008.
- [16] R. Liu and X. Zhang, "Generating machine-executable plans from end-user's natural-language instructions," *Knowledge-Based Systems*, vol. 140, pp. 15-26, 2018.
- [17] M. Gomez, "Towards Improving the Quality of Mobile Apps by Leveraging Crowdsourced Feedback," Université Lille 1; Inria Lille-Nord Europe, 2016.
- [18] M. Gómez, R. Rouvov, M. Monperrus, and L. Seinturier, "A recommender system of buggy app checkers for app store moderators," in *Proceedings of the Second ACM International Conference on Mobile Software Engineering and Systems*, 2015, pp. 1-11: IEEE Press.
- [19] G. V. Georgiev and D. D. Georgiev, "Enhancing user creativity: Semantic measures for idea generation," *Knowledge-Based Systems*, vol. 151, pp. 1-15, 2018.
- [20] I. J. M. Ruiz, M. Nagappan, B. Adams, T. Berger, S. Dienst, and A. E. Hassan, "Examining the rating system used in mobile-app stores," *IEEE Software*, vol. 33, no. 6, pp. 86-92, 2015.
- [21] T. Xu, Q. Peng, and Y. Cheng, "Identifying the semantic orientation of terms using S-HAL for sentiment analysis," *Knowledge-Based Systems*, vol. 35, pp. 279-289, 2012.
- [22] N. Milosevic, A. Dehghantanha, and K.-K. R. Choo, "Machine learning aided Android malware classification," *Computers & Electrical Engineering*, vol. 61, pp. 266-274, 2017.
- [23] J. Atkinson, A. Ferreira, and E. Aravena, "Discovering implicit intention-level knowledge from natural-language texts," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, 2008, pp. 249-262: Springer.