 doi:10.20944/preprints202003.0249.v1

Article

Machine Learning Algorithm's Measurement and Analytical Visualization of User's Reviews for Google Play Store

Abdul Karim[1], Azhari Azhari[1], Samir Brahim Belhaouri[2, *] and Ali Adil Qureshi[3]

[1] Department of Computer Science and Electronics, University Gadjah Mada, Yogyakarta, Indonesia; abdullkarim@mail.ugm.ac.id; arisn@ugm.ac.id
[2] Division of Information & Computer Technology, College of Science & Engineering, Hamad Bin Khalifa University, Doha, Qatar ; sbelhaouari@hbku.edu.qa
[3] Department of Computer Science, Khwaja Fareed University of Engineering & Information Technology, Rahim Yar Khan, Pakistan; shah87061@gmail.com
* Correspondence: sbelhaouari@hbku.edu.qa

Abstract: The fact is quite transparent that almost everybody around the world is using android apps. Half of the population of this planet is associated with messaging, social media, gaming, and browsers. This online marketplace provides free and paid access to users. On the Google Play store, users are encouraged to download countless of applications belonging to predefined categories. In this research paper, we have scrapped thousands of users reviews and app ratings. We have scrapped 148 apps' reviews from 14 categories. We have collected 506259 reviews from Google play store and subsequently checked the semantics of reviews about some applications form users to determine whether reviews are positive, negative, or neutral. We have evaluated the results by using different machine learning algorithms like Naïve Bayes, Random Forest, and Logistic Regression algorithm. we have calculated Term Frequency (TF) and Inverse Document Frequency (IDF) with different parameters like accuracy, precision, recall, and F1 and compared the statistical result of these algorithms. We have visualized these statistical results in the form of a bar chart. In this paper, the analysis of each algorithm is performed one by one, and the results have been compared. Eventually, We've discovered that Logistic Regression is the best algorithm for a review-analysis of all Google play store. We have proved that Logistic Regression gets the speed of precision, accuracy, recall, and F1 in both after preprocessing and data collection of this dataset.

Keywords: machine learning; preprocessing; semantic analysis; text mining; TF/IDF; scraping; Google Play Store

---

1. Introduction

The essential task of natural language processing is the classification of text strings or documents into different categories that are the part of this process, which depends upon the content of the string. Text classification has a variety of applications, including detection of user sentiments on comments or tweets, classification of an email as spam. Presently, text classification has gained vital importance in organizing online information [1]. User reviews and the mobile program ecosystem have an abundance of information regarding expectations and user experience. Programmers and app store regulators can leverage the data to better understand their audience. App stores enable users to search for, buy and install programs that are mobile and give comments in the form of evaluations and reviews. The rapid increase in the number of applications and complete app store earning has accelerated opinion aggregation studies and app store data mining. There are now some academic studies centered on user testimonials and mobile program stores, in addition to studies analyzing online product reviews. In this article, we used various algorithms, and text classification techniques with android app reviews[2]. Application distribution platforms, or app stores, allow users to search, buy, and deploy software apps for mobile devices with a few clicks. These platforms also allow users to share their opinion about the app in text reviews, where they can, e.g., express their satisfaction with a specific app feature or request a new feature [3]. Recent empirical studies have shown that app store reviews include information that is useful to analysts and documentation of user experiences with specific app features. This feedback can represent the "voice of the users" and be used to drive the development effort and improve forthcoming releases [4]. The main points of this research are the following:

1. We have scrapped recent android application reviews by using the scrapping technique.
2. The raw data we have scrapped from the Google Play store and we collect this data in chunks and normalized the dataset for our analysis.
3. We have compared the accuracy of various machine learning algorithms and find the best algorithm according to the results.
4. Algorithms can check the polarity of sentiment based on review is positive, negative, and neutral as well as we can prove this using the word cloud corpus.

Text mining also referred to as text data mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text. High-quality information is typically derived through the devising of patterns and trends using statistical pattern learning [5]. Text mining usually involves the process of structuring the input text parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database, deriving patterns within the structured data, and finally evaluation and interpretation of the output [6]. 'High quality' in text mining usually refers to some combination of relevance, novelty, and interest. Typical text mining tasks include text categorization, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarization, and entity relation modeling, i.e., learning relations between named entities [7].

There are several limitations which prevent analysts and development teams from using the information in the reviews. First, app stores include many reviews, which require a significant effort to be analyzed. A recent study has confirmed that iOS users submit on average, 22 reviews per day per app [8]. Top-rated apps, such as Facebook get more than 4000 reviews per day. Second, the quality of the reviews varies widely, from helpful advice to sardonic comments. Third, a review is nebulous, concerning different app features, making it challenging to filter positive and negative feedback or retrieve the feedback for specific features. The usefulness of the star ratings in the reviews is limited for development teams since a rating represents an average for the whole app and can combine both positive and negative evaluations of the single features [9].

In linguistics, semantic analysis is the process of relating syntactic structures, from the levels of phrases, clauses, sentences, and paragraphs writing. It also involves removing features

 doi:10.20944/preprints202003.0249.v1

specific to linguistic and cultural contexts, to the extent that such a project is possible [10]. The elements of idiom and figurative speech, being cultural, are often also converted into relatively invariant meanings in semantic analysis. Semantics, although related to pragmatics, is distinct in that the former deals with word or sentence choice in any given context, while pragmatics consider the unique meaning derived from context or tone. In different terms repetition, semantics is about universally coded meaning, and pragmatics, the meaning encoded in words that are then interpreted by an audience [11].

In information retrieval, TF/IDF, short for Term Frequency-Inverse Document Frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in searches of information retrieval, text mining, and user modeling. The TF/IDF value increases proportionally to the number of times a word appears in the document and the number of documents in the corpus that contain the word, which helps to adjust for the fact that some words appear more frequently in general [12]. TF/IDF is one of the most popular term-weighting schemes today; 83% of text-based recommender systems in digital libraries use TF/IDF. Using of common words in TF/IDF e.g., articles receive a significant weight even if they contribute no real information about common words. In TF/IDF, the more familiar a word is in the corpus, the smaller weight it receives. Thus, common words like articles receive small weights but rare words, that it is assumed to carry more information, receive larger weights [13].

Beautiful Soup is a Python library for pulling data out of HTML and XML files. It works with the favorite parser to provide idiomatic ways of navigating, searching and modifying the parse tree [14]. RE-module is the core of text processing. The RE-module provides sophisticated ways to create and use regular expressions [15]. A regular expression is a kind of formula that specifies patterns in text strings. The name "regular expression" comes from the earlier mathematical treatment of "regular sets. We are still stuck with the phrase. Regular expressions give us a simple way to specify a set of related strings by describing the pattern they have in common. We write a pattern to summarize some set of matching strings. This pattern string can be compiled into an object that efficiently determines if and where a given string matches the pattern [16, 17].

2. Literature Review

In this paper, authors propose a framework that allows developers to filter, summarize, and analyze user reviews written about applications. Author extract automatically relevant features from reviews of apps (e.g., information about functionalities, bugs, and requirements) and analyze the sentiment associated with each of them. In this research three main building blocks (i) topic modeling, (ii) sentiment analysis and (iii) summarization interface, are discussed. The topic modeling block aims to find semantic topics from textual comments, extracting the features based on the most relevant words of each topic. The sentiment analysis block detects the sentiment associated with each discovered feature [18]. The summarization interface provides developers an intuitive visualization of the features (i.e., topics) along with their associated sentiment, providing more valuable information than a 'star rating'. Our evaluation shows that the topic modeling block can organize the information provided by users into subcategories that facilitate the understanding of features that may positive, negative, and neutral impact on the overall evaluation of the application.

 doi:10.20944/preprints202003.0249.v1

Regarding user satisfaction, authors can observe that, despite the star rating being a good measure of evaluation, the Sentiment Analysis technique is more precise in capturing the sentiment transmitted by the user using comment [19].

Authors discussed Sentiment Analysis of App Reviews. In this study, result shows sentiment analysis of App Reviews approaches that is helpful for the developers. With these approaches, the developer can accumulate, filter, and examine user reviews. In this research, use of simple language techniques to recognize fine-grained app features in the reviews. In this extract, the user gives an opinion about the recognized features by giving a typical score to all reviews [20]. By using topic modeling techniques, authors can group fine-grained features into more comfortable and meaningful features. Authors compared the result and analyzed the 7 applications taken from the Apple app store and Google play store. In this way, the app developer can systematically examine the user opinion about a single feature and filter view. Authors got the accuracy of up to 91% and recall up to 73% [21].

Text mining techniques have been recently employed to classify and summarize user reviews on mobile application stores. However, due to the inherently diverse and unstructured nature of user-generated online textual data, text-based review mining techniques often produce excessively complicated models that are prone to overfitting. In this paper, authors proposed approach, based on frame semantics, for review mining [22]. Semantic frames help to generalize from the raw text (individual words) to more abstract scenarios (contexts). This representation of text is expected to enhance the predictive capabilities of review mining techniques and reduce the chances of overfitting [23]. First, authors investigate the performance of semantic frames in classifying informative user reviews into various categories of software requests about maintenance. Second, the authors propose and evaluate the performance of multiple summarization algorithms in generating concise and representative summaries of informative reviews. Three different datasets of app store reviews, sampled from a broad range of application domains, have been used to conduct experimental analysis [24]. The results have shown that semantic frames can enable an efficiently quick and precise review classification process. However, in reviewing summarization tasks, our deductions claim that text-based summarization generates more comprehensive summaries than frame-based summarization. In closing, authors have introduced MARC 2.0, a review classification and summarization suite that implements the algorithms investigated in the analysis [25].

Nowadays, the use of apps has increased with the use of advance mobile technology. User prefer to use mobile phone for mobile application as compare to other gadgets. Users already downloaded different mobile applications in their mobile phones and they uses these application and left reviews about it [26]. In the mobile app market, fallacious ranking points may lead to pushing up mobile apps in the popularity list. Indeed, it turns more periodic for app developers to use fake mechanism. The paper has, at this moment, proposed semantic analysis of app review for fraud detection in mobile apps. Firstly, authors have proposed to detect the misrepresentation by excavating the active periods correctly, also called as leading sessions, of the mobile apps [27]. Authors have an intention to inspect two types of evidence: ranking-based review and -based and use natural language processing (NLP) to get action words. Next, authors have agreed to convert review to ratings and finally perform pattern analysis on the session with app data gathered from the app store. So, the paper has proposed an approach to validate its effectiveness and show the scalability of the detection algorithm [28].

 doi:10.20944/preprints202003.0249.v1

User review is a crucial component of open mobile app markets such as the Google Play Store. The question arises: How do authors automatically summarize millions of user reviews and make sense out of them? Unfortunately, beyond simple summaries such as histograms of user ratings, few analytic tools can provide insights into user reviews [29]. In this paper, authors proposed a system "Wiscom" that can analyze millions of user ratings and comments in mobile app markets at three different levels of detail. This system is able to (a) discover inconsistencies in reviews; (b) identify reasons why users like, or dislike a given app, and provide an interactive, zoomable view of how users' reviews evolve over time; and (c) provide valuable insights into the entire app market, identifying users' significant concerns and preferences of different types of apps. Results using a techniques and are reported on a 32GB dataset consisting of over 13 million user reviews of 171,493 Android apps in the Google Play Store [30]. Three operator such as Google as well as individual app developers and end-users [31]. The products on "Amazon.com", the mobile apps are continuously evolving, with newer versions rapidly superseding the older ones. Many app stores still use an Amazon-style rating system, which aggregates every rating ever assigned to an app into one store rating [32]. To examine whether the store rating captures the fickle user-satisfaction levels regarding new app versions, researchers mined the store ratings of more than 10,000 mobile apps in Google Play, every day for a year. Even though many apps' version ratings rose or fell, their store rating was resilient to fluctuations once they had gathered a substantial number of raters. The conclusion is that current store ratings are not dynamic enough to capture changing user satisfaction levels. This resilience is a significant problem that can discourage developers from improving app quality [33].

3. The methodology of Analytical Measurement and Visualization of Users' Reviews

In this methodology for classification is started with the scraping of reviews on applications. On Google Play store using the AppID request for scrape the reviews of that specific application scrape several pages with reviews and rating of the applications. We have scarped this dataset for classifying the user reviews that is a positive, negative, or neutral review. After scraping the bulk raw reviews, the next step of preprocessing of those reviews. In preprocessing different steps, we normalize our reviews after preprocessing. These steps are involves removing a special character, remove a single character, remove a single character from the start, subtracting multiple spaces with single spaces and remove prefixed, then converting data into lowercase and at the end of this stop words and stemming is performed. These are some significant steps for refining our reviews. After refining reviews, the bag of words approach is performed. In next step apply TF (Term Frequency) on reviews by using a python language after that we apply TF/IDF (Term Frequency-Inverse Document Frequency), its often used in information retrieval and text mining. After applying TF/IDF, feature extraction performs on each application. By using python, we use different algorithm for classification Naïve Bayes, Random Forest, and Logistic Regression and check the different parameters like accuracy, precision, recall, and F1-score and find the statistical information of these parameters. After analyzing and testing from statistical information we get the result about which algorithm has a maximum accuracy, precision, recall, and F1-score information, and we can access which algorithm is best for analyzing of reviews for classification as shown in Figure. 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

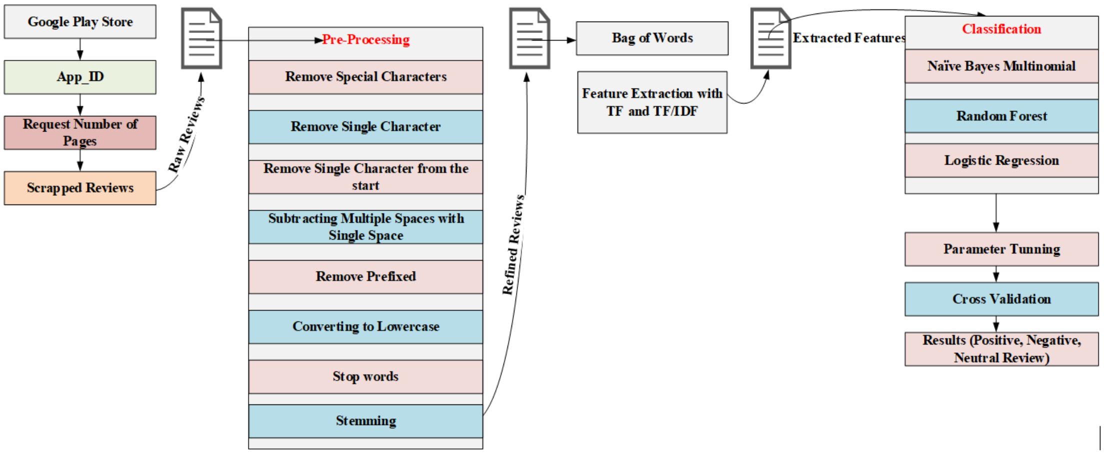


Figure 1. Flow of Google Play Store Application Reviews Classification

4. The Methodology of Data Collection Process

The advancement in technology, Mobile applications can become a part of our daily life. Half-million applications were introduced in 2011, and in October 2012, 0.675 million applications were accessible on the Google Play store. Now a day's Android app is being used by lots of peoples; people use different Android apps, like messengers, social media, games and browsers. This online marketplace provides free and paid access for mobile users 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 an important task in every research, and the validity and accuracy of the dataset is also a significant part of any dataset collection process. In this research, scrape the thousands of users review and rating of different applications based on the different categories, as shown in Figure. 2. We select 14 categories of Google play store, different scrape application of each category, as 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 scrape the thousands of reviews and ratings of application, and convert these data into a .CSV file format. After this we applying preprocessing for removing special characters, remove a single character, remove a single character from the start, subtracting multiple spaces with single spaces, remove prefixed, converting data into lowercase, stop words and then stemming technique on data in .CSV file. Then evaluating results by using different machine learning algorithm and find the best algorithm for classification. We have download 148 apps that appeared in 14 categories from Google play store fetch several reviews and enter the required pages according to the reviews. We Collect total of 506259 reviews in the Google play store website, as shown in Figure. 2., to fetch the data, in first step we use the requests library We use Python's Scikit-Learn Library (Pedregosa , Varoquaux , Gramfort , Michel , Thirion , Grisel & Vanderplas, 2011) for machine learning because this library provides machine learning algorithms like classification, regression, clustering, model validation etc. [34]. The requests library allows the user to send HTTP/1.1 requests using Python to add content like headers. This library allows users to process response data in python. Then use the Re-library for text processing. A regular expression is a unique sequence of characters that help the user match or find other strings or sets of strings, using a specific syntax held in a pattern. After using the Re-library, use the Beautiful Soup library. Beautiful Soup library is used to extract data from the HTML and XML files. This library works quickly and saves the programmer's time, as shown in Table. 1.

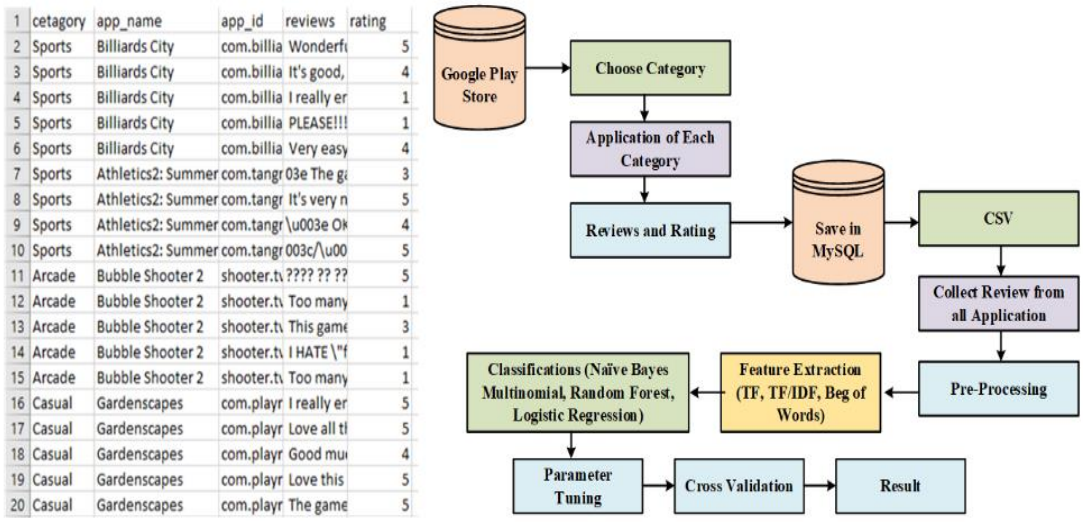


Figure 2. Methodology diagram of data collection and sample shot of dataset

Table 1. Detail measurements of dataset scrapped

Action		Arcade		Card	
App Name	**Reviews**	**App Name**	**Reviews**	**App Name**	**Reviews**
Bush Rush	4001	29 Card Game	4001	Angry Bird Rio	4481
Gun Shot Fire War	3001	Blackjack 21	1601	Bubble Shooter 2	4001
Metal Soldiers	4001	Blackjack	4481	Jewels Legend	4001
N.O.V.A Legacy	4364	Callbreak Multiplayer	4001	Lep World 2	3001
Real Gangster Crime	4001	Card Game 29	3066	Snow Bros	3001
Shadow Fight 2	4481	Card Words Kingdom	4472	Sonic Dash	4481
Sniper 3D Gun Shooter	4481	Gin Rummy	3616	Space Shooter	4401
Talking Tom Gold Run	4001	Spider Solitaire	2801	Subway Princes Runner	3001
Temple Run 2	3001	Teen Patti Gold	4481	Subway Surfers	4481
Warship Battle	4001	World Series of poker	4001	Super Jabber Jump 3	2912
Zombie Frontier 3	4001				
Zombie Hunter	3782				

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

King					
Communication		**Finance**		**Health and Fitness**	
App Name	**Reviews**	**App Name**	**Reviews**	**App Name**	**Reviews**
Dolphin Browser	3001	bKash	4481	Home Workout - No Equipment	4481
Firefox Browser	3001	CAIXA	1220	Home Workout for Men	1163
Google Duo	3001	CAPTETEB	697	Lose Belly Fat In 30 Days	4481
Hangout Dialer	3001	FNB Banking App	2201	Lose It! - Calorie Counter	4001
KakaoTalk	3001	Garanti Mobile Banking	859	Lose Weight Fat In 30 Days	4481
LINE	3001	Monobank	605	Nike+Run Club	1521
Messenger Talk	3001	MSN Money-Stock Quotes & News	3001	Seven - 7 Minutes Workout	4425
Opera Mini Browser	3001	Nubank	1613	Six Pack In 30 Days	3801
UC Browser Mini	3001	PhonePe-UPI Payment	4001	Water Drink Reminder	4481
WeChat	3001	QIWI Wallet	1601	YAZIO Calorie Counter	1590
		Yahoo Finance	3001		
		YapiKredi Mobile	1952		
		Stock	3001		
Photography		**Shopping**		**Sports**	
App Name	**Reviews**	**App Name**	**Reviews**	**App Name**	**Reviews**
B612 - Beauty & Filter Camera	4001	AliExpress	4481	Billiards City	4481
BeautyCam	4001	Amazon for Tablets	4481	Real Cricket 18	3001
BeautyPlus	4001	Bikroy	4481	Real Football	3001
Candy Camera	4481	Club Factory	4001	Score! Hero	3001
Google Photos	4481	Digikala	4001	Table Tennis 3D	3001
HD Camera	4001	Divar	4001	Tennis	3001
Motorola Camera	4001	Flipkart Online Shopping App	4481	Volleyball Champions 3D	3001
Music Video Maker	4001	Lazada	4481	World of Cricket	4481
Sweet Selfie	4481	Myntra Online Shopping App	4481	Pool Billiards Pro	4001
Sweet Snap	4001	Shop clues	4481	Snooker Star	2801
Video Player Editor		**Weather**		**Casual**	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

App Name	Reviews	App Name	Reviews	App Name	Reviews
KIneMaster	1441	NOAA Weather Radar & Alerts	3601	Angry Bird POP	4481
Media Player	2713	The Weather Channel	4001	BLUK	3281
MX Player	3001	Transparent Weather & Clock	1441	Boards King	4481
Power Director Video Editor App	1641	Weather & Clock Weight for Android	4481	Bubble Shooter	4481
Video Player All Format	1041	Weather & Radar - Free	3601	Candy Crush Saga	4481
Video Player KM	3001	Weather Forecast	1681	Farm Heroes Super Saga	4481
Video Show	1321	Weather Live Free	1721	Hay Day	4481
VivaVideo	4190	Weather XL PRO	1401	Minion Rush	4481
You Cut App	1241	Yahoo Weather	4361	My Talking Tom	4481
YouTube	1201	Yandex. Weather	1045	Pou	4481
				Shopping Mall Girl	4481
				Gardens capes	4481

Medical		Racing	
App Name	**Reviews**	**App Name**	**Reviews**
Anatomy Learning	2401	Asphalt Nitro	4481
Diseases & Dictionary	3201	Beach Buggy Racing	4481
Disorder & Diseases Dictionary	2401	Bike Mayhem Free	4481
Drugs.com	2401	Bike Stunt Master	2745
Epocrates	1001	Dr. Driving 2	4481
Medical Image	1423	Extreme Car Driving	4481
Medical Terminology	1448	Hill Climb Racing 2	3801
Pharmapedia Pakistan	4134	Racing Fever	4481
Prognosis	2401	Racing in Car 2	4481
WikiMed	3201	Trial Xtreme 4	4481

 doi:10.20944/preprints202003.0249.v1

5. Results and Experiments

The results have been evaluated by fetching the reviews of different categories. Perform a series of steps which predict the sentiment reviews of different categories. The usage of Python's Scikit-Learn Library because this library provides different features like classification, Regression, clustering, and model validation. Different methods and features that are perform are as follows: Import the data, Feature Extraction, Convert the text into Numbers, Training and testing sets, Training text classification model and predicting sentiments and Evaluating model.

6. Analytical Measurement and Visualization After Preprocessing

These are the statistical information of different algorithm on the base of the different parameters after preprocessing; compare and find the best algorithm that uses for the analysis and classification of reviews.

6.1 Naïve Bayes Multinomial

Naïve Bayes is used for classification. It assumes that the occurrence of a specific feature is independent of the occurrence of other features. It is fast to make models and make predictions. We have scraped 148 apps reviews form 14 categories from Google play store. There are 40 reviews on one page, we have collected a total of 506259 reviews from Google play store applications. Apply the Naïve Bayes algorithm for classification on that dataset of reviews and find different information on different parameters concerning TF and TF/IDF. Find the accuracy of classification of each category application and in statistical information find precision, recall, and F1 score these all parameters use to measure the accuracy of the dataset is shown in Table. 2. Also, bar chart visualization of Naïve Bayes algorithm in which series1 shows the accuracy of Naïve Bayes algorithm, series2 shows the precision, series3 shows the recall and series4 shows the F1 score measurement as shown in Figure. 3.

Table 2. Statistical information of Naïve Bayes Multinomial algorithm on TF and TF/IDF bases after preprocessing

Application Category	Naïve Bayes Multinomial							
	TF				TF/IDF			
	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Sports	0.602	0.359	0.316	0.315	0.594	0.341	0.227	0.203
Communication	0.587	0.333	0.332	0.304	0.597	0.297	0.301	0.254
Action	0.691	0.334	0.294	0.288	0.686	0.297	0.231	0.215
Arcade	0.725	0.283	0.231	0.235	0.737	0.319	0.191	0.168
Video players & editors	0.676	0.331	0.306	0.294	0.67	0.314	0.233	0.215
Weather	0.662	0.329	0.261	0.266	0.642	0.301	0.194	0.168

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

Card	0.689	0.31	0.285	0.276	0.68	0.28	0.227	0.209
Photography	0.696	0.367	0.327	0.31	0.705	0.362	0.276	0.248
Shopping	0.667	0.358	0.341	0.321	0.678	0.299	0.316	0.289
Health & fitness	0.788	0.273	0.212	0.218	0.811	0.208	0.194	0.177
Finance	0.532	0.301	0.287	0.266	0.557	0.284	0.258	0.226
Casual	0.73	0.334	0.285	0.288	0.745	0.334	0.205	0.182
Medical	0.745	0.359	0.272	0.279	0.753	0.338	0.204	0.181
Racing	0.718	0.357	0.278	0.285	0.72	0.331	0.218	0.201

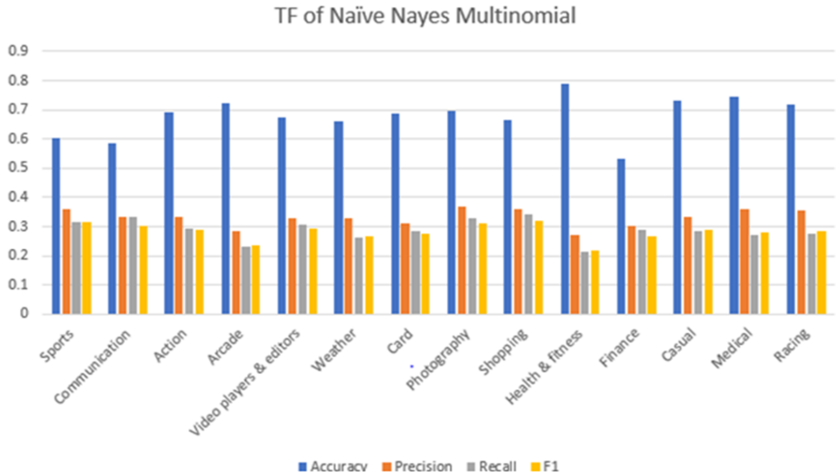


Figure 3(a). Bar chat visualization of TF Naïve Bayes Multinomial algorithm after preprocessing

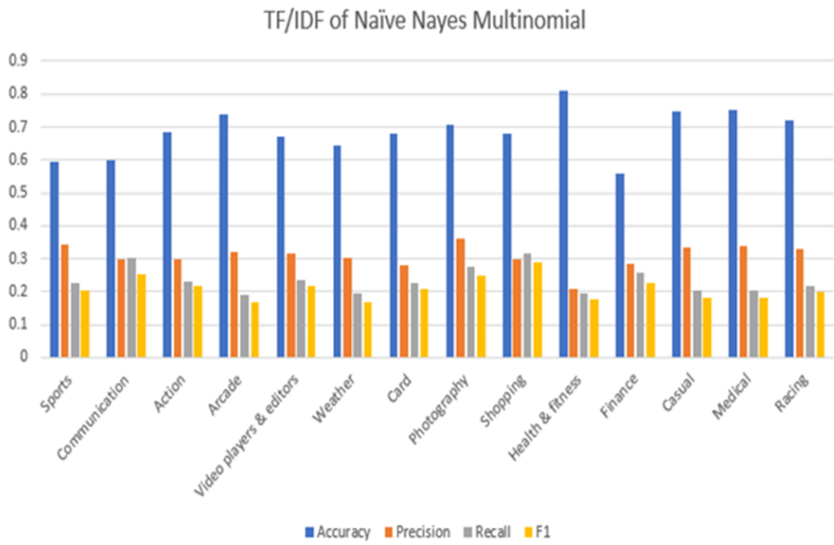


Figure 3(b). Bar chat visualization of TF/IDF Naïve Bayes Multinomial algorithm after preprocessing

6.2 Random Forest Algorithm

Random Forests Classifier is the class of all methods that are designed explicitly for decision tree. Random forest randomly decided based on a random selection of data and a random selection of variables. Random Forest algorithm, check the classification on reviews and apply different

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

information on different parameters concerning TF and TF/IDF. Find the accuracy of classification of each category application and in statistical information find precision, recall, and F1 score these all parameters use to measure the accuracy of the dataset is shown in Table. 3. Also, bar chart visualization of Random Forest algorithm in which series1 shows the accuracy of Random Forest algorithm, series2 shows the precision, series3 shows the recall and series4 shows the F1 score measurement as shown in Figure. 4.

Table 3. Statistical information of Random Forest algorithm on TF and TF/IDF bases after preprocessing

Application Category	Random Forest							
	TF				TF/IDF			
	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Sports	0.585	0.34	0.312	0.308	0.589	0.344	0.308	0.304
Communication	0.544	0.314	0.313	0.294	0.545	0.307	0.312	0.288
Action	0.683	0.338	0.31	0.308	0.691	0.347	0.306	0.302
Arcade	0.721	0.32	0.27	0.274	0.729	0.334	0.262	0.269
Video players &editors	0.664	0.347	0.313	0.304	0.664	0.34	0.304	0.295
Weather	0.632	0.285	0.243	0.248	0.638	0.305	0.252	0.255
Card	0.665	0.312	0.285	0.279	0.673	0.321	0.284	0.277
Photography	0.683	0.353	0.32	0.312	0.69	0.352	0.315	0.301
Shopping	0.648	0.354	0.333	0.324	0.653	0.359	0.33	0.316
Health & fitness	0.765	0.324	0.248	0.254	0.779	0.315	0.235	0.24
Finance	0.517	0.309	0.291	0.27	0.52	0.31	0.293	0.27
Casual	0.728	0.341	0.284	0.292	0.732	0.342	0.274	0.28
Medical	0.729	0.33	0.28	0.285	0.739	0.336	0.265	0.271
Racing	0.714	0.359	0.317	0.319	0.724	0.37	0.306	0.311

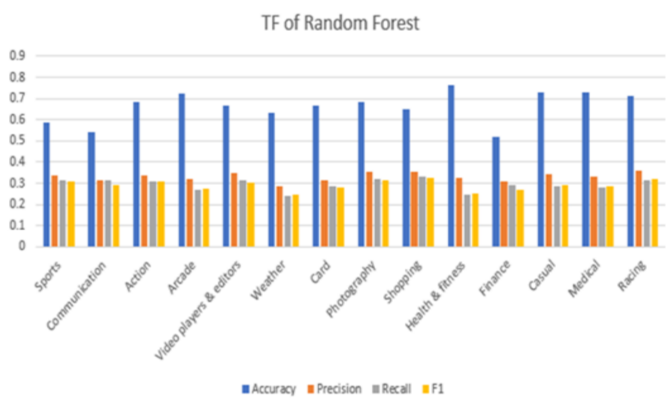


Figure 4(a). Bar chat visualization of TF Random Forest algorithm after preprocessing

 doi:10.20944/preprints202003.0249.v1

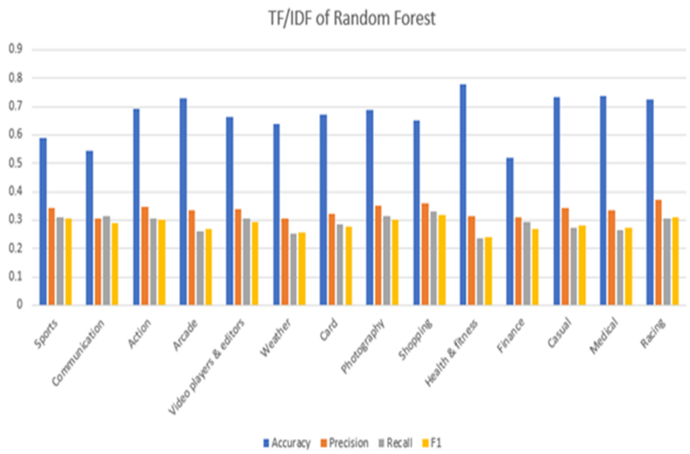


Figure 4(b). Bar chat visualization of TF/IDF Random Forest algorithm after preprocessing

6.2 Logistic Regression Algorithm

According to the statistics, the logistic product can be reliable statistical version, in which its essential type that runs on the logistic functionality to simulate a binary determining factor; lots complex extensions where exist. Back in Regression investigation, Logistic Regression will be estimating the parameters of the logistic version; it is an application for binomial Regressions. Apply the Logistic Regression algorithm for classification on dataset for reviews and find different information on different parameters concerning TF and TF/IDF. Find the accuracy of classification of each category application and in statistical information find precision, recall, and F1 score these all parameters use to measure the accuracy of the dataset is shown in Table. 4. Also, bar chart visualization of Logistic Regression algorithm in which series1 shows the accuracy of Logistic Regression algorithm, series2 shows the precision, series3 shows the recall and series4 shows the F1 score measurement as shown in Figure. 5.

Table 4. Statistical information of Logistic Regression algorithm on TF and TF/IDF bases after preprocessing

Application Category	**Logistic Regression**							
	TF				TF/IDF			
	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Sports	0.622	0.414	0.343	0.343	0.621	0.404	0.319	0.315
Communication	0.585	0.349	0.329	0.32	0.599	0.352	0.327	0.301
Action	0.707	0.395	0.312	0.313	0.71	0.38	0.299	0.293
Arcade	0.744	0.353	0.262	0.266	0.747	0.351	0.25	0.252
Video players &	0.684	0.37	0.306	0.304	0.687	0.352	0.289	0.276

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

editors								
Weather	0.667	0.379	0.288	0.299	0.667	0.421	0.262	0.265
Card	0.696	0.379	0.301	0.305	0.698	0.344	0.283	0.271`
Photography	0.703	0.391	0.321	0.315	0.71	0.405	0.311	0.297
Shopping	0.67	0.407	0.342	0.336	0.682	0.444	0.332	0.315
Health & fitness	0.796	0.38	0.278	0.295	0.801	0.391	0.23	0.235
Finance	0.592	0.352	0.311	0.303	0.593	0.353	0.298	0.276
Casual	0.747	0.381	0.29	0.302	0.753	0.364	0.277	0.28
Medical	0.754	0.401	0.277	0.288	0.759	0.459	0.244	0.245
Racing	0.737	0.428	0.312	0.318	0.74	0.401	0.295	0.297

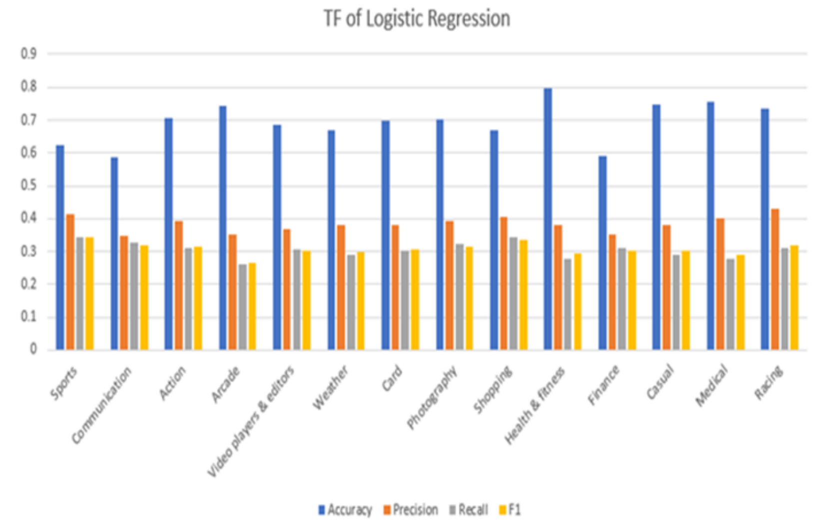


Figure 5(a). Bar chat visualization of TF Logistic Regression algorithm after preprocessing

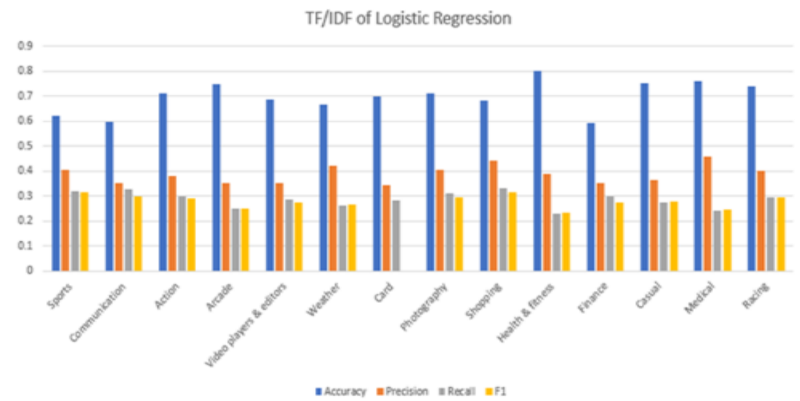


Figure 5(b). Bar chat visualization of TF/IDF Logistic Regression algorithm after preprocessing

7. Different Machine Learning Algorithm Comparison After Preprocessing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

Google Play Store is an online market place that provided free and paid access to users. Google Play store, users can choose from over a million apps from various predefined categories. In this research, scrapped thousands of users' review and app ratings. We evaluated the results by using different machine learning algorithms like Naïve Bayes, Random Forest, and Logistic Regression algorithm that can check the semantics of reviews about some applications form users that their reviews are good, bad, normal and so on. Calculated Term Frequency (TF) and Inverse Document Frequency (IDF) with different parameters like accuracy, precision, recall, and F1 score after the preprocessing of the Raw reviews in the concluded results compared the statistical result of these algorithms. Visualized these statistical results in the form of a bar chart, as shown in Figure. 6. After comparison, analyzed that the Logistic Regression algorithm is the best algorithm for checking the semantic analysis of any Google application users' reviews on both TF and TF/IDF bases. As in sports category in TF base, showed that Logistic Regression algorithm has 0.622% accuracy, 0.414% precision, 0.343% recall and 0.343% F1 score and the statistical information with another category of application as shown in Table. 10. Also, in TF/IDF base showed that Logistic Regression algorithm has 0.621% accuracy, 0.404% precision, 0.319% recall and 0.315% F1 score and the statistical information with another category of application is shown in Table. 5.

Table 5. Different machine learning algorithm comparison of on TF based after preprocessing

Application Category	Naïve Bayes Accuracy	Random Forest Accuracy	Logistic Regression Accuracy	Naïve Bayes Precision	Random Forest Precision	Logistic Regression Precision	Naïve Bayes Recall	Random Forest Recall	Logistic Regression Recall	Naïve Bayes F1 score	Random Forest F1	Logistic Regression F1 score
Sports	0.602	0.585	0.622	0.359	0.34	0.414	0.316	0.312	0.343	0.315	0.308	0.343
Communication	0.587	0.544	0.585	0.333	0.314	0.349	0.332	0.313	0.329	0.304	0.294	0.32
Action	0.691	0.683	0.707	0.334	0.338	0.395	0.294	0.31	0.312	0.288	0.308	0.313
Arcade	0.725	0.721	0.744	0.283	0.32	0.353	0.231	0.27	0.262	0.235	0.274	0.266
Video players & editors	0.676	0.664	0.684	0.331	0.347	0.37	0.306	0.313	0.306	0.294	0.304	0.304
Weather	0.662	0.632	0.667	0.329	0.285	0.379	0.261	0.243	0.288	0.266	0.248	0.299
Card	0.689	0.665	0.696	0.31	0.312	0.379	0.285	0.285	0.301	0.276	0.279	0.305
Photography	0.696	0.683	0.703	0.367	0.353	0.391	0.327	0.32	0.321	0.31	0.312	0.315
Shopping	0.667	0.648	0.67	0.358	0.354	0.407	0.341	0.333	0.342	0.321	0.324	0.336

Health & fitness	0.788	0.765	0.796	0.273	0.324	0.38	0.212	0.248	0.278	0.218	0.254	0.295
Finance	0.532	0.517	0.592	0.301	0.309	0.352	0.287	0.291	0.311	0.266	0.27	0.303
Casual	0.73	0.728	0.747	0.334	0.341	0.381	0.285	0.284	0.29	0.288	0.292	0.302
Medical	0.745	0.729	0.754	0.359	0.33	0.401	0.272	0.28	0.277	0.279	0.285	0.288
Racing	0.718	0.714	0.737	0.357	0.359	0.428	0.278	0.317	0.312	0.285	0.319	0.318

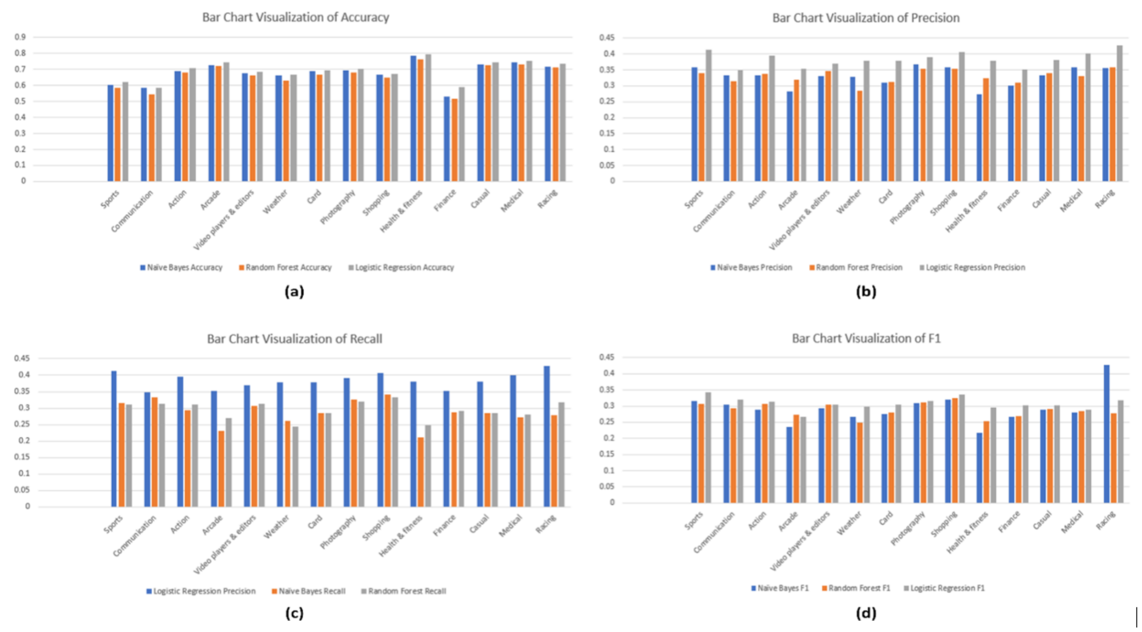


Figure 6. Bar chat visualization of different machine learning algorithm comparison on TF based after preprocessing

Table 6. Different machine learning algorithm comparison of on TF/IDF based after preprocessing

Application Category	Naïve Bayes Accuracy	Random Forest Accuracy	Logistic Regression Accuracy	Naïve Bayes Precision	Random Forest Precision	Logistic Regression Precision	Naïve Bayes Recall	Random Forest Recall	Logistic Regression Recall	Naïve Bayes F1 score	Random Forest F1 score	Logistic Regression F1 score
Sports	0.594	0.589	0.621	0.341	0.344	0.404	0.227	0.308	0.319	0.203	0.304	0.315
Communication	0.597	0.545	0.599	0.297	0.307	0.352	0.301	0.312	0.327	0.254	0.288	0.301
Action	0.686	0.691	0.71	0.297	0.347	0.38	0.231	0.306	0.299	0.215	0.302	0.293

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

Arcade	0.737	0.729	0.747	0.319	0.334	0.351	0.191	0.262	0.25	0.168	0.269	0.252
Video players & editors	0.67	0.664	0.687	0.314	0.34	0.352	0.233	0.304	0.289	0.215	0.295	0.276
Weather	0.642	0.638	0.667	0.301	0.305	0.421	0.194	0.252	0.262	0.168	0.255	0.265
Card	0.68	0.673	0.698	0.28	0.321	0.344	0.227	0.284	0.283	0.209	0.277	0.271`
Photography	0.705	0.69	0.71	0.362	0.352	0.405	0.276	0.315	0.311	0.248	0.301	0.297
Shopping	0.678	0.653	0.682	0.299	0.359	0.444	0.316	0.33	0.332	0.289	0.316	0.315
Health & fitness	0.811	0.779	0.801	0.208	0.315	0.391	0.194	0.235	0.23	0.177	0.24	0.235
Finance	0.557	0.52	0.593	0.284	0.31	0.353	0.258	0.293	0.298	0.226	0.27	0.276
Casual	0.745	0.732	0.753	0.334	0.342	0.364	0.205	0.274	0.277	0.182	0.28	0.28
Medical	0.753	0.739	0.759	0.338	0.336	0.459	0.204	0.265	0.244	0.181	0.271	0.245
Racing	0.72	0.724	0.74	0.331	0.37	0.401	0.218	0.306	0.295	0.201	0.311	0.297

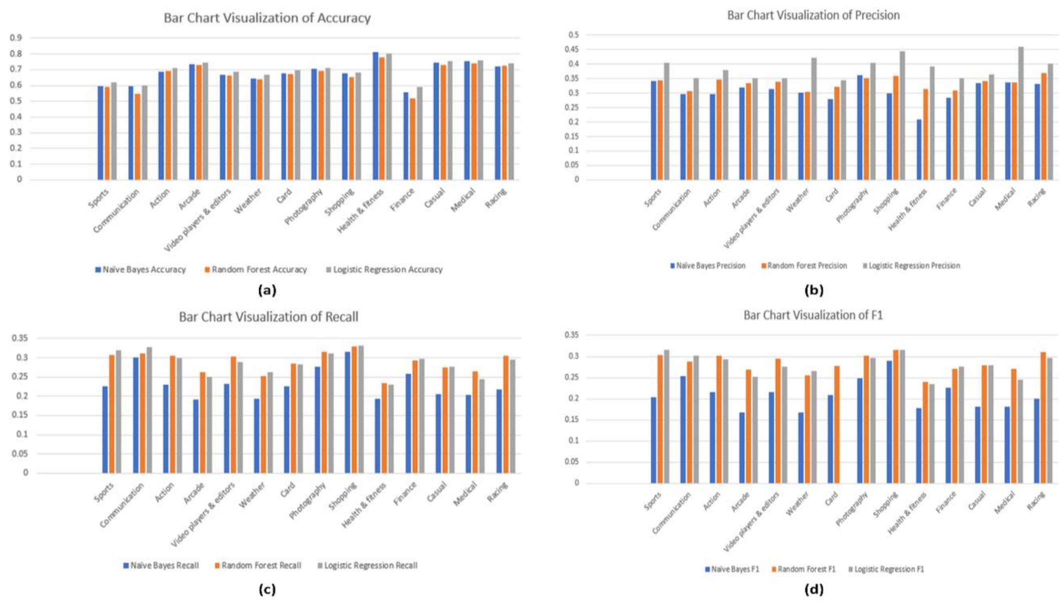


Figure 7. Bar chat visualization of different machine learning algorithm comparison on TF/IDF based after preprocessing

 doi:10.20944/preprints202003.0249.v1

8. Analytical Measurement and Visualization Without Preprocessing of Dataset

These are the statistical information of different algorithm on the base of the different parameters after data collection; compare and find the best algorithm that uses for the analysis and classification of reviews.

8.1 Naïve Bayes Multinomial

Naïve Bayes is commonly used classifcation algorithm. Naïve Bayes assumes that the occurrence of a specific feature is independent of the occurrence of other features. It is fast to make models and make predictions. Apply the Naïve Bayes algorithm for classification on that dataset of reviews and find different information on different parameters concerning TF and TF/IDF. Find the accuracy of classification of each category application and in statistical information find precision, recall, and F1 score these all parameters use to measure the accuracy of the dataset is shown in Table 7. Also, bar chart visualization of Naïve Bayes algorithm in which series1 shows the accuracy of Naïve Bayes algorithm, series2 shows the precision, series3 shows the recall and series4 shows the F1 score measurement as shown in Figure. 8.

Table 7. Statistical information of Naïve Bayes Multinomial algorithm on TF and TF/IDF based without preprocessing of the dataset

Application Category	Naïve Bayes Multinomial							
	TF				TF/IDF			
	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Sports	0.607	0.368	0.335	0.334	0.593	0.328	0.221	0.194
Communication	0.584	0.334	0.337	0.311	0.597	0.292	0.302	0.255
Action	0.689	0.336	0.303	0.297	0.683	0.329	0.227	0.208
Arcade	0.724	0.273	0.221	0.227	0.737	0.33	0.191	0.168
Video players & editors	0.681	0.346	0.325	0.312	0.669	0.281	0.229	0.208
Weather	0.669	0.327	0.276	0.281	0.641	0.306	0.19	0.161
Card	0.689	0.282	0.305	0.272	0.68	0.29	0.223	0.205
Photography	0.691	0.366	0.335	0.317	0.707	0.367	0.279	0.25
Shopping	0.663	0.364	0.352	0.333	0.679	0.365	0.319	0.292
Health & fitness	0.788	0.277	0.218	0.225	0.811	0.2	0.194	0.176
Finance	0.536	0.312	0.297	0.277	0.554	0.261	0.257	0.226
Casual	0.727	0.338	0.304	0.306	0.745	0.31	0.204	0.181
Medical	0.749	0.348	0.29	0.298	0.753	0.38	0.203	0.18
Racing	0.717	0.351	0.289	0.297	0.718	0.359	0.214	0.195

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

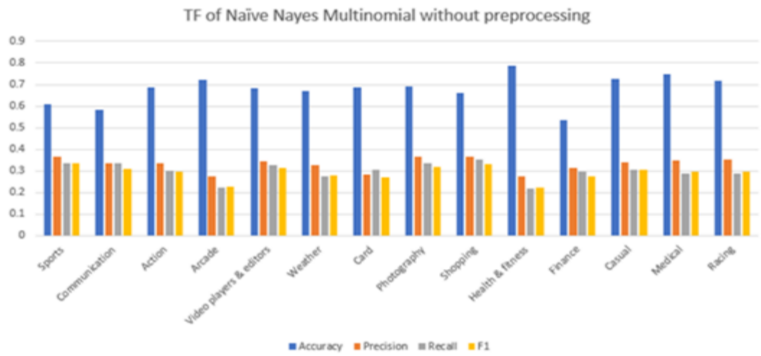


Figure 8(a). Bar chat visualization of TF Naïve Bayes Multinomial algorithm based without preprocessing of data

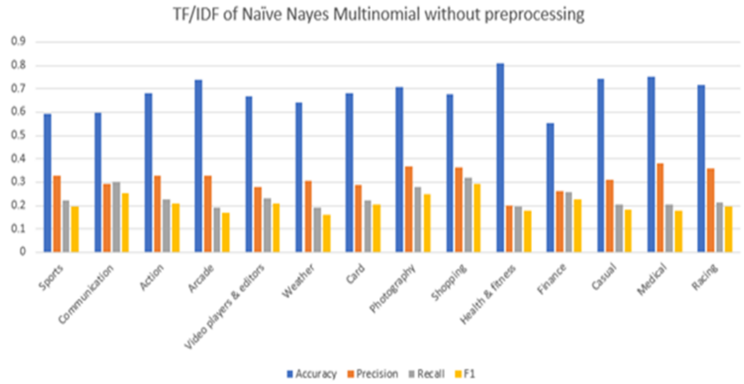


Figure 8(b). Bar chat visualization of TF/IDF Naïve Bayes Multinomial algorithm based without preprocessing of data

8.2 Random Forest Algorithm

Random Forests Classifier is the class of all methods that are designed explicitly for decision tree. It develops a lot of decision tree based on a random selection of data and a random selection of variables. Apply the Random Forest algorithm for classification on that dataset of reviews and find different information on different parameters concerning TF and TF/IDF. Find the accuracy of classification of each category application and in statistical information find precision, recall, and F1 score these all parameters use to measure the accuracy of the dataset is shown in Table. 8. Also, bar chart visualization of Random Forest algorithm in which series1 shows the accuracy of Random Forest algorithm, series2 shows the precision, series3 shows the recall and series4 shows the F1 score measurement as shown in Figure. 9.

Table 8. Statistical information of Random Forest algorithm on TF and TF/IDF based without preprocessing of the dataset

Application Category	Random Forest							
	TF				TF/IDF			
	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
Sports	0.589	0.359	0.311	0.314	0.595	0.352	0.307	0.309

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

Communication	0.559	0.321	0.314	0.296	0.555	0.32	0.309	0.29
Action	0.686	0.35	0.307	0.308	0.695	0.369	0.306	0.307
Arcade	0.725	0.338	0.265	0.275	0.729	0.336	0.256	0.265
Video players & editors	0.665	0.351	0.311	0.308	0.665	0.335	0.306	0.298
Weather	0.641	0.335	0.273	0.282	0.647	0.323	0.255	0.264
Card	0.666	0.325	0.28	0.281	0.669	0.322	0.274	0.273
Photography	0.689	0.372	0.328	0.321	0.69	0.363	0.318	0.308
Shopping	0.654	0.37	0.333	0.325	0.653	0.361	0.328	0.315
Health & fitness	0.778	0.299	0.215	0.22	0.788	0.354	0.22	0.225
Finance	0.532	0.311	0.292	0.276	0.529	0.317	0.294	0.272
Casual	0.735	0.345	0.273	0.284	0.739	0.346	0.267	0.277
Medical	0.737	0.342	0.276	0.284	0.743	0.351	0.259	0.268
Racing	0.719	0.361	0.311	0.317	0.726	0.383	0.307	0.317

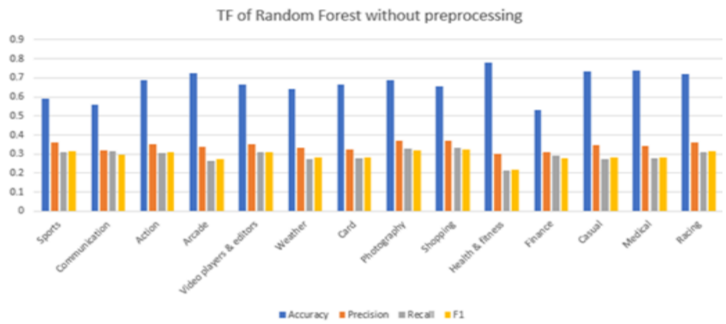


Figure 9 (a). Bar chat visualization of TF Random Forest algorithm based without preprocessing of data

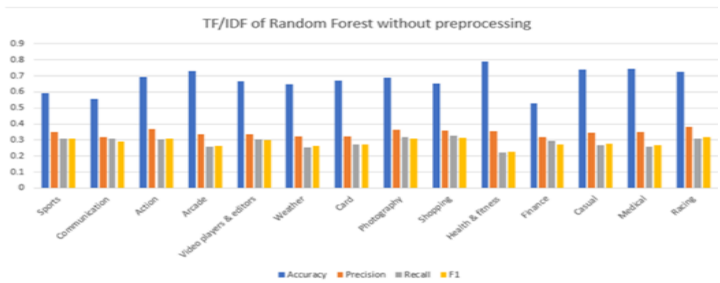


Figure 9 (b). Bar chat visualization of TF Random Forest algorithm based without preprocessing of data

8.3 Logistic Regression Algorithm

In statistics, the logistic product can be a trusted statistical version which, in its essential type that runs on the logistic functionality to simulate a binary determining factor; lots complex extensions exist. Back in Regression investigation, Logistic Regression will be estimating the parameters of the logistic version; it is an application of both binomial Regressions. Apply the Logistic Regression algorithm for classification on that dataset of reviews and find different information on different parameters concerning TF and TF/IDF. Find the accuracy of classification of

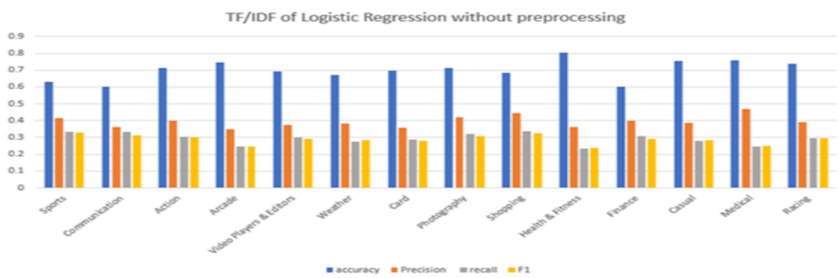
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

each category application and in statistical information find precision, recall, and F1 score these all parameters use to measure the accuracy of the dataset is shown in Table. 9. Also, bar chart visualization of Logistic Regression algorithm in which series1 shows the accuracy of Logistic Regression algorithm, series2 shows the precision, series3 shows the recall and series4 shows the F1 score measurement as shown in Figure. 10.

Table 9. Statistical information of Logistic Regression algorithm on TF and TF/IDF based without preprocessing of the dataset

Application Category	Logistic Regression							
	TF				TF/IDF			
	accuracy	Precision	recall	F1 score	accuracy	Precision	recall	F1 score
Sports	0.623	0.416	0.35	0.353	0.629	0.416	0.331	0.328
Communication	0.588	0.355	0.334	0.326	0.602	0.361	0.334	0.312
Action	0.71	0.405	0.32	0.324	0.712	0.398	0.304	0.299
Arcade	0.744	0.369	0.271	0.278	0.747	0.349	0.246	0.247
Video Players & Editors	0.69	0.39	0.323	0.323	0.693	0.375	0.3	0.291
Weather	0.674	0.386	0.303	0.316	0.674	0.384	0.275	0.282
Card	0.697	0.373	0.306	0.31	0.699	0.359	0.288	0.281
Photography	0.707	0.403	0.332	0.328	0.714	0.422	0.322	0.309
Shopping	0.674	0.411	0.351	0.346	0.686	0.444	0.339	0.324
Health & Fitness	0.794	0.369	0.28	0.295	0.803	0.363	0.232	0.239
Finance	0.595	0.363	0.319	0.314	0.604	0.401	0.308	0.29
Casual	0.747	0.385	0.3	0.314	0.755	0.385	0.28	0.285
Medical	0.757	0.41	0.295	0.31	0.759	0.468	0.246	0.249
Racing	0.738	0.419	0.317	0.325	0.74	0.391	0.295	0.297

Figure 10 (a). Bar chat visualization of TF Logistic Regression algorithm based without preprocessing of data



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

Figure 10 (b). Bar chat visualization of TF/IDF Logistic Regression algorithm based without preprocessing of data

9. Different Machine Learning Algorithm Comparison Without Preprocessing of Dataset

Using the Google Play store, users can choose from over a million apps from various predefined categories. Evaluate the results by using different machine learning algorithms like Naïve Bayes, Random Forest, and Logistic Regression algorithm that can check the semantics of reviews about some applications form users that their reviews are good, bad, normal and so on. Calculate Term Frequency (TF) and Inverse Document Frequency (IDF) with different parameters like accuracy, precision, recall, and F1 score after the data collection of the Raw reviews in the concluded results compared the statistical result of these algorithms. Visualized these statistical results in the form of a bar chart, as shown in Figure. 11. After comparison, analyze that the Logistic Regression algorithm is the best algorithm to check the semantic analysis of any Google application users' reviews on both TF and TF/IDF bases. As in sports category in TF base, show that Logistic Regression algorithm has 0.623% accuracy, 0.416% precision, 0.35% recall, and 0.353% F1 score and the statistical information with another category of application is shown in Table. 15. Also, in TF/IDF base show that Logistic Regression algorithm has 0.629% accuracy, 0.416% precision, 0.331% recall, and 0.328% F1 score and the statistical information with another category of application is shown in Table. 10.

Table 10. Different machine learning algorithm comparison on TF based without preprocessing of the dataset

Application Category	Naïve Bayes Accuracy	Random Forest Accuracy	Logistic Regression Accuracy	Naïve Bayes Precision	Random Forest Precision	Logistic Regression Precision	Naïve Bayes Recall	Random Forest Recall	Logistic Regression Recall	Naïve Bayes F1 score	Random Forest F1 score	Logistic Regression F1 score
Sports	0.607	0.589	0.623	0.368	0.359	0.416	0.314	0.314	0.35	0.334	0.314	0.353
Communication	0.584	0.559	0.588	0.334	0.321	0.355	0.296	0.296	0.334	0.311	0.296	0.326
Action	0.689	0.686	0.71	0.336	0.35	0.405	0.308	0.308	0.32	0.297	0.308	0.324
Arcade	0.724	0.725	0.744	0.273	0.338	0.369	0.275	0.275	0.271	0.227	0.275	0.278
Video players & editors	0.681	0.665	0.69	0.346	0.351	0.39	0.308	0.308	0.323	0.312	0.308	0.323
Weather	0.669	0.641	0.674	0.327	0.335	0.386	0.282	0.282	0.303	0.281	0.282	0.316

Card	0.689	0.666	0.697	0.282	0.325	0.373	0.281	0.281	0.306	0.272	0.281	0.31
Photography	0.691	0.689	0.707	0.366	0.372	0.403	0.321	0.321	0.332	0.317	0.321	0.328
Shopping	0.663	0.654	0.674	0.364	0.37	0.411	0.325	0.325	0.351	0.333	0.325	0.346
Health & fitness	0.788	0.778	0.794	0.277	0.299	0.369	0.22	0.22	0.28	0.225	0.22	0.295
Finance	0.536	0.532	0.595	0.312	0.311	0.363	0.276	0.276	0.319	0.277	0.276	0.314
Casual	0.727	0.735	0.747	0.338	0.345	0.385	0.284	0.284	0.3	0.306	0.284	0.314
Medical	0.749	0.737	0.757	0.348	0.342	0.41	0.284	0.284	0.295	0.298	0.284	0.31
Racing	0.717	0.719	0.738	0.351	0.361	0.419	0.317	0.317	0.317	0.297	0.317	0.325

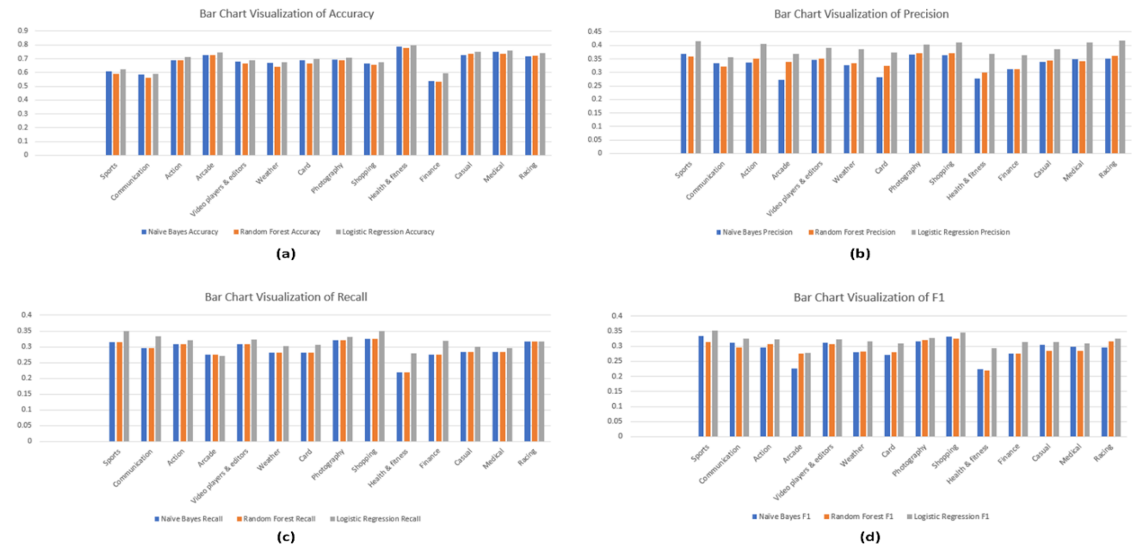


Figure 11. Bar chat visualization of different machine learning algorithm comparison on TF based without preprocessing of data

Table 11. Different machine learning algorithm comparison on TF/IDF based without preprocessing of the dataset

Application Category	**Naïve Bayes Accuracy**	**Random Forest Accuracy**	**Logistic Regression Accuracy**	**Naïve Bayes Precision**	**Random Forest Precision**	**Logistic Regression Precision**	**Naïve Bayes Recall**	**Random Forest Recall**	**Logistic Regression Recall**	**Naïve Bayes F1 score**	**Random Forest F1 score**	**Logistic Regression F1 score**
Sports	0.593	0.595	0.629	0.328	0.352	0.416	0.221	0.307	0.331	0.194	0.309	0.328

Communication	0.597	0.555	0.602	0.292	0.32	0.361	0.302	0.309	0.334	0.255	0.29	0.312
Action	0.683	0.695	0.712	0.329	0.369	0.398	0.227	0.306	0.304	0.208	0.307	0.299
Arcade	0.737	0.729	0.747	0.33	0.336	0.349	0.191	0.256	0.246	0.168	0.265	0.247
Video players & editors	0.669	0.665	0.693	0.281	0.335	0.375	0.229	0.306	0.3	0.208	0.298	0.291
Weather	0.641	0.647	0.674	0.306	0.323	0.384	0.19	0.255	0.275	0.161	0.264	0.282
Card	0.68	0.669	0.699	0.29	0.322	0.359	0.223	0.274	0.288	0.205	0.273	0.281
Photography	0.707	0.69	0.714	0.367	0.363	0.422	0.279	0.318	0.322	0.25	0.308	0.309
Shopping	0.679	0.653	0.686	0.365	0.361	0.444	0.319	0.328	0.339	0.292	0.315	0.324
Health & fitness	0.811	0.788	0.803	0.2	0.354	0.363	0.194	0.22	0.232	0.176	0.225	0.239
Finance	0.554	0.529	0.604	0.261	0.317	0.401	0.257	0.294	0.308	0.226	0.272	0.29
Casual	0.745	0.739	0.755	0.31	0.346	0.385	0.204	0.267	0.28	0.181	0.277	0.285
Medical	0.753	0.743	0.759	0.38	0.351	0.468	0.203	0.259	0.246	0.18	0.268	0.249
Racing	0.718	0.726	0.74	0.359	0.383	0.391	0.214	0.307	0.295	0.195	0.317	0.297

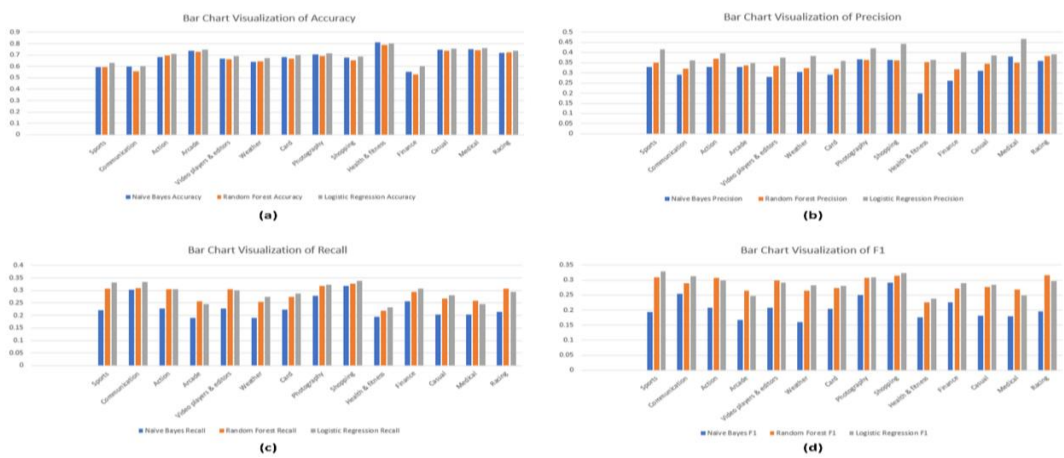


Figure 12. Bar chat visualization of different machine learning algorithm comparison on TF/IDF based without preprocessing of data

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

10. Semantic Analysis of Google Play Store Applications Reviews Using Logistic Regression Algorithm

After checking the different paraments, analyze that Logistic Regression algorithm is the best algorithm having the highest accuracy. In this section, 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, and 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. 13. However, for checking the semantics of each review, these parameters are more enough. That is why we select only reviews of all application.

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	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 13. Sample screenshot of the original dataset that scrapped

11. Data Preparation and Cleaning of Reviews steps

HTML Decoding

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

Data Preparation 2: '# ' Mention

"#" carries import information that must deal with is necessary.

URL Links

 doi:10.20944/preprints202003.0249.v1

Remove all URL’s that appears in reviews remove them.

UTF-8 BOM (Byte Order Mark)

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

Hashtag / Numbers

Hashtag text can provide useful information about the comment. It might be a bit tough to get rid of all the text together with the “#” or with a number or with any other unique character needs to deal.

Negation Handling

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

Tokenizing and Joining

Parse the whole comment into small pieces/segments and then merge again. After applying the above rules on cleaning, the reviews cleaned form of reviews, as shown in Figure. 14.

1	Reviews
2	wonderfull app completed all levels can not wait for more levels level has bug but you can get around it to complete
3	it good like the gameplay please change up the music as it gets repetitive after the nd level and you can hear the t
4	really enjoyed this game until saw one of the adverts throughout the ngame with man and woman spooning it is ha
5	please get rid of the odd shaped tables and go back to the classic table please fix level where the balls and cue st
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
7	not going to lie just started playing this game probably about hours ago and am thoroughly addicted it pretty awes
8	level the lower left cushion let ball disappear into nothingness nafter striking the ball it can be seen at different edge
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 tl
10	the gameplay is fun graphics are good and there are lot of levels however the number of ads are insane is there pai
11	great game until you get above level than it starts shooting balls off screen where can not see them or play anymo
12	just wanted to thank the supervisors creator those in charge whatever powers that be that after days of my compla
13	as pool player this is good app find it helps me with using different english techniques control of the ball as well as
14	just wanted basic billiards game there was this redundant tutorial level up system that was totally unecessary and
15	installed this game and could not even play it because the game would not load and would shut down waste of tim
16	really enjoyed the older version alot more than now some of these odd ball ntables are more aggravating than enjoy
17	really like the game up to certain point when the normal billiard table is changed to crooked one now its hard to pla
18	pretty decent game but has way too many ads now made me watch second video every time wanted to retry table
19	level and til now no problem but then on this levels the balls keep disappearing under the table and from there there
20	first had trouble getting pass level learned it amp now on level great game finish there are levels total will go back a

Figure 14. Sample screenshot of the cleaning dataset after preprocessing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

12. Find Null Entries from Reviews

It seems there are about 700-800 null entries in the reviews column of the dataset. Which might happen during the cleaning process to remove the null entries with using commands as shown below.

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

13. Negative and Positive Words Dictionary

By using word cloud corpus, made negative and positive words dictionary based on the occurrence of words in a sentence to get the idea of what kind of words are frequent in the corpus as shown in Figure. 16.

(a) (b)

Figure 15. (a) Positive (b) Negative word dictionary by using the word cloud corpus

14. Semantic Analysis of Google Play Store Applications Reviews Using Logistic Regression Algorithm

In the result, classify all reviews three different classes and can check the confidence rate of each rate that how much that comment is positive, negative, and neutral. Set the target value equal to 0 to1 and check the confidence value in that ratio and check the class of the review using the Logistic Regression algorithm, as shown in Figure. 16.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0249.v1

1	Reviews	Tags	Confidance	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 unecessary 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 I	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 wl	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 w	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 16. Final sentiment analysis results on Google Play reviews using Logistic Regression Algorithm

15. Conclusion and Future Work

On the Google Play store, users may download more than one million applications from different categorized groups. In this research, we have built hundreds and thousands of user's application reviews. We have 14 categories and download 148 app reviews. Accumulate 506259 reviews out of Google play store. Assessed the outcome using unique machine learning algorithms such as Naïve Bayes, Random Forest, and Logistic Regression algorithm, which will assess the semantics of application users' reviews are equally positive, negative, and neutral. Calculate Time-Frequency (TF) and Inverse Document Frequency (IDF) using various parameters such as precision, accuracy, recall, and F1 score and about the statistical effect of those calculations. Using only TF, it's not an issue if a word is common or not. Thus, common words like, e.g., articles receive a large weight even if they contribute no real information. In TF/IDF, the more common a word is in the corpus, the smaller weight it receives. Thus, common words like articles receive small weights but rare words, that are assumed to carry more information, receive larger weights. Visualize these statistical results in the form of a bar chart. The study of each algorithm has been conducted one by one and the results can be compare. The evaluated results are shows that Logistic Regression could be an optimal algorithm to get a review of this Google Play store tool. Logistic Regression got the optimal speed of precision, accuracy, recall, and F1 score in equally earlier and right after preprocessing of their dataset. Some statistical results as in sports category in TF base after preprocessing the Logistic Regression algorithm has 0.622% accuracy, 0.414% precision, 0.343% recall and 0.343% F1 score and in TF/IDF based Logistic Regression algorithm has 0.621% accuracy, 0.404% precision, 0.319% recall and 0.315% F1 score. Also, the sports category in TF base after data collection the Logistic Regression algorithm has 0.623% accuracy, 0.416% precision, 0.35% recall and 0.353% F1 score and in TF/IDF based Logistic Regression algorithm has 0.629% accuracy, 0.416% precision, 0.331% recall and 0.328% F1 score and the statistical information with another category of applications analyze in concluded table below that shows the authenticity of this analysis. The

 doi:10.20944/preprints202003.0249.v1

section performs analysis and classifies 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, and it is equal to 0. Analyze the neutral class with the confidence rate if the confidence rate is between the 0 and 1 then classify this to neutral class.

In the future, increase the category of applications and increase the number of reviews. Compare the Logistic Regression algorithm accuracy results with different algorithms. Generate clusters and check the relationship between reviews and ratings of the application that can analyze each application more precisely.

Refrences

[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] M. M. S. Missen *et al.*, "OpinionML—Opinion Markup Language for Sentiment Representation," *Symmetry,* vol. 11, no. 4, p. 545, 2019.

[5] C. Gao, Y. Zhao, R. Wu, Q. Yang, and J. Shao, "Semantic trajectory compression via multi-resolution synchronization-based clustering," *Knowledge-Based Systems,* 2019.

[6] 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.

[7] 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.

[8] 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.

[9] J. R. Saura and D. R. Bennett, "A Three-Stage method for Data Text Mining: Using UGC in Business Intelligence Analysis," *Symmetry,* vol. 11, no. 4, p. 519, 2019.

[10] F. Benedetti, D. Beneventano, S. Bergamaschi, and G. Simonini, "Computing inter-document similarity with context semantic analysis," *Information Systems,* vol. 80, pp. 136-147, 2019.

[11] J. S. Sachs, "Recognition memory for syntactic and semantic aspects of connected discourse," *Perception & Psychophysics,* vol. 2, no. 9, pp. 437-442, 1967.

[12] T. Zhang and S. S. Ge, "An Improved TF-IDF Algorithm Based on Class Discriminative Strength for Text Categorization on Desensitized Data," in *Proceedings of the 2019 3rd International Conference on Innovation in Artificial Intelligence*, 2019, pp. 39-44: ACM.

[13] Y. Zhang, W. Ren, T. Zhu, and E. Faith, "MoSa: A Modeling and Sentiment Analysis System for Mobile Application Big Data," *Symmetry,* vol. 11, no. 1, p. 115, 2019.

[14] L. Richardson, "Beautiful Soup Documentation," ed: Crummy, http://www. crummy. com/software/BeautifulSoup/bs4/doc/, accessed May, 2015.

[15] C. Chapman and K. T. Stolee, "Exploring regular expression usage and context in Python," in *Proceedings of the 25th International Symposium on Software Testing and Analysis*, 2016, pp. 282-293: ACM.

[16] H. Chivers, "Optimising Unicode regular expression evaluation with previews," *Software: Practice and Experience,* vol. 47, no. 5, pp. 669-688, 2017.

[17] M. G. Coutinho, M. F. Torquato, and M. A. Fernandes, "Deep Neural Network Hardware Implementation Based on Stacked Sparse Autoencoder," *IEEE Access,* 2019.

[18] A. Di Sorbo *et al.*, "What would users change in my app? summarizing app reviews for recommending software changes," in *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, 2016, pp. 499-510: ACM.

[19] M. Colhon, Ş. Vlăduţescu, and X. Negrea, "How Objective a Neutral Word Is? A Neutrosophic Approach for the Objectivity Degrees of Neutral Words," *Symmetry,* vol. 9, no. 11, p. 280, 2017.

[20] A. Al-Subaihin *et al.*, "App store mining and analysis," in *Proceedings of the 3rd International Workshop on Software Development Lifecycle for Mobile*, 2015, pp. 1-2: ACM.

[21] S. Panichella, A. Di Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall, "Ardoc: App reviews development oriented classifier," in *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, 2016, pp. 1023-1027: ACM.

[22] J. Tröger *et al.*, "Exploitation vs. exploration—computational temporal and semantic analysis explains semantic verbal fluency impairment in Alzheimer's disease," *Neuropsychologia,* vol. 131, pp. 53-61, 2019.

[23] K. Kansal and A. Subramanyam, "Hdrnet: Person Re-Identification Using Hybrid Sampling in Deep Reconstruction Network," *IEEE Access,* vol. 7, pp. 40856-40865, 2019.

[24] N. Jha and A. Mahmoud, "Using frame semantics for classifying and summarizing application store reviews," *Empirical Software Engineering,* pp. 1-34, 2018.

[25] N. Jha and A. Mahmoud, "Using frame semantics for classifying and summarizing application store reviews," 2018.

[26] N. M. Puram and K. R. Singh, "Semantic Analysis of App Review for Fraud Detection using Fuzzy Logic," 2018.

[27] Y. Wang *et al.*, "Multi-objective workflow scheduling with Deep-Q-network-based Multi-agent Reinforcement Learning," *IEEE Access,* 2019.

[28] N. M. Puram and K. Singh, "An Implementation to Detect Fraud App Using Fuzzy Logic."

[29] W. Yang, J. Li, Y. Zhang, Y. Li, J. Shu, and D. Gu, "APKLancet: tumor payload diagnosis and purification for android applications," in *Proceedings of the 9th ACM symposium on Information, computer and communications security*, 2014, pp. 483-494: ACM.

[30] 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, 2016.

[31] C. Yang, M. Jiang, B. Jiang, W. Zhou, and K. Li, "Co-Attention Network with Question Type for Visual Question Answering," *IEEE Access,* 2019.

[32] J.-H. Huh, "Big data analysis for personalized health activities: machine learning processing for automatic keyword extraction approach," *Symmetry,* vol. 10, no. 4, p. 93, 2018.

[33] S. Guo, R. Chen, and H. Li, "Using knowledge transfer and rough set to predict the severity of Android test reports via text mining," *Symmetry,* vol. 9, no. 8, p. 161, 2017.

[34] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.