REVIEWS ANALYSIS OF UNLOCKED MOBILE PHONES IN AMAZON USING RAPID MINER

Fatima Ijaz[*1]
Department of Computer Science
Riphah International University
Lahore,Pakistan
fatimaijaz084@gmail.com

Aneela Akhtar
Department of Computer Science
GC University
Faisalabad, Pakistan
aneelaakhter766@gmail.com

Maryam Shafique
Department of Computer Science
GC University
Faisalabad , Pakistan
Maryamshafique112@gmail.com

Yasir Ali
Department of Computer Science
GC University
Faisalabad , Pakistan
chyasir070@gmail.com

Abstract

Consumers beforehand retail a phone ensure review on the way to acquire an improved considerate of the expedient and this scheme derive a most advantageous resolution intended for this. Appearing in this paradigm, all characteristic of a cell phone be rate base lying on community view and a global ranking designed for all categories. Amazon is solitary of the biggest internet vendor, which make manner intended for the majority community review on top of their products and consequently we collect records on behalf of reaction analysis commencing Amazon. Inside this research, we consider the occurrence of dissimilar machine learning techniques and we assess our model accurateness. Reviews and rating analysis conduct in rapid miner and improve the accuracy, and text mining for obtains the true and false reviews. Our research provides high accuracy then previous researches.

Keywords— Amazon reviews, regression, classification, text mining, decision tree, generalized linear model, gradient boosted trees, random forest, support vector machine, naïve bayes, and fast large margin.

1. INTRODUCTION

Purchase merchandise is a communication flanked by two entities, customers and commerce owner. Customers frequently utilize review in the direction of construct decision regarding what harvest toward purchase, whereas business, lying on the supplementary worker, not solitary desire in the direction of be bought their harvest excluding too desire toward obtain criticism within conditions of customer reviews. Consumer's reviews concerning purchase goods mutual lying on the internet contain huge collision. Individual quality be usually prearranged en route for construct decision base on top of analyze and in receipt of the benefit of further customer understanding and opinion for the reason that others frequently encompass a enormous impertinence happening our viewpoint, behaviors', opinion of actuality, and the choice we construct. Consequently, we inquire others intended for their criticism at any time we are decide resting on liability impressive. Furthermore, this actuality apply not merely in the direction of customers except also headed for organization and institution. In the previous some years, purchaser traditions of express their opinion and manner contain distorted according toward change within community network, essential community and additional social media communities. Discover great amount of facts commencing amorphous figures at the network have turn out to be an significant confront appropriate on the way to its significance in dissimilar area of living. In the direction of permit superior in sequence withdrawal the machine learning algorithms are functional resting on figures. We examine within this investigate ranking scrutiny of mobile phone review in use commencing the Amazon website, and how these review assist customers on the way to include confidence to facilitate they have finished the accurate conclusion concerning their purchase. Too, the explore inside this occupation aim headed for assist company appreciate their customers' response just before uphold their goods /services or boost them. Inside adding up, charitable them insight concerning them inside provided that offer resting on specific goods on the way to enhance their profits and purchaser fulfillment.

In this research we conduct an analysis on reviews and ratings of unlocked mobile phones in Amazon for this purpose used the rapid miner studio. The unlocked mobile phones in Amazon dataset are given from kaggle. The five machine learning algorithms applied on items file for regression and 4 machine learning algorithms applied on reviews file for classification then find the accuracy. Conduct text mining and obtain the true and false reviews.

The structure starts with the part1 with both the fundamentals of the content delivery system as Description. Part 2 contains the problems statements. Segment 3 consists on previous papers literatures. Segment 4 I will discuss the methodology. The development of the current scheme and results is shown in Part 5. Part 6 of the current system shall have the summary and future products.

1. Problem statement

A moment ago, electronic business websites utilize of the Internet have amplified toward the summit to facilitate customers rely lying on them intended for trade and advertising. Because these websites provide customers the aptitude on the way to inscribe commentary concerning dissimilar goods and services, enormous amount of review encompass become obtainable. Accordingly, the require intended for toward analyse individuals review toward appreciate customers' feedbacks have augmented intended

for together vendor and customers. Conversely, it is hard toward glance at every the feedbacks intended for an exacting point particularly meant for the accepted substance by means of a lot of remarks. This is supposed headed for assist company progress their goods and too facilitate prospective buyer construct enhanced decision whilst trade goods.

3. Literature review

[2] The study obtainable in this document be conceded absent as follow: the statistics was pre-processed, previous to organism rehabilitated as of manuscript in the direction of vector depiction by means of a variety of characteristic removal techniques such because container -of- language, TF-IDF, Glove, and word2vec. We learn the presentation of dissimilar machine learning algorithms, such because logistic regression, stochastic gradient descent, naive Bayes and convolutional neural network in accumulation, we adding our model with accurateness, F1score, precision, recall and log loss occupation. Furthermore, we are relevant Emerald method in the direction of impart diagnostic reason intended for the review being classified since optimistic, pessimistic or unbiased.

[7} Opinion psychiatry of products review, and request trouble, has lately turned out to be extremely accepted within wording drawing out and computational linguistics study. At this point, we wish for to learning the association connecting the Amazon manufactured goods review and the ranking of the goods given by the consumers. We make use of together conventional machine learning algorithms counting Naive Bayes analysis, Support Vector Machines, K nearest neighbor technique and deep neural networks such as Recurrent Neural Network (RNN), Recurrent Neural Network (RNN). By means of compare these consequences, we might acquire a improved considerate of the these algorithms. They might too act as a addition in the direction of additional deception score discovery method.
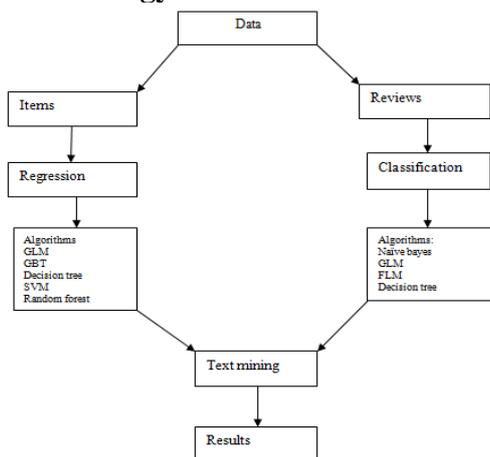
4. Methodology

Figure1: process flow diagram

4.1 Data collection

The dataset use on behalf of this paper is obtained from Kaggle. The records composed from Amazon it is concerning unlocked mobile phones. It contains of 400,000 reviews, two data files that are items and reviews and the item files contains 10 columns which are: 1) asin, 2) Product Fame: it depict the name of the corporation, 3)title, 4) url, 5)image, 6) Price: the charge of the mobile; 7)

Rating: star ranking which the costumer give toward the product; 8) Reviews: the users opinion about each product;9) total review 10) original price. And the reviews file contain 8 columns which are: 1) asin, 2) name 3) rating 4) date 5) verified, 6) title 7) body 8) Review Votes: the Amount of customers who nominated the review. Furthermore, we used item data file for rating analysis and reviews data file used to sentiment analysis.

4.2 Review analysis of unlocked mobile phones in Amazon

The reviews could assist companies get better their products as well as help potential buyers to create the correct choice. A consumer already trade a phone checks review on the way to obtain an enhanced considerate of the mechanism and this estimate originate a mainly constructive clarification for this. For analysis the reviews I have select the unlocked phone in Amazon dataset and choose the reviews data file to classification. This file upload the rapid miner studio then applied the reviews analysis steps. These steps are: load data, select task, prepare target, select inputs, model types and results. Four machine learning algorithms are apply for review analysis that are generalized linear model, fast large margin, decision tree and naïve bayes.

4.3 Naïve bayes

Naïve bayes simulator most likely shows the reviews the true reviews are 94% and false are 6%. The important factors for true texts are seven which contains two basic instances that are support true and contradicts true. The performance of naïve bayes algorithm described the profit. The profits from model are 15,515, profits for best option (true) is: 15,515 and the gain are 0.
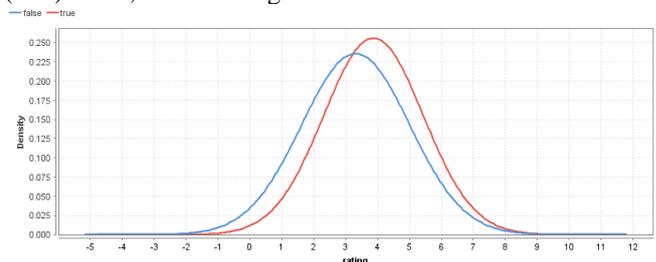
Figure2: naïve bayes sentiment rating analysis

The above figure shows the naïve bayes sentiment rating. The red curved line represents the true reviews and the blue curved line represents the false reviews ratings. On the X axis put the rating and the Y axis shows the density. The naïve bayes model shows the costs for wrong predications. The costs are shows as negative numbers and benefits or gain are shown as positive numbers. **Cost matrix** shows the predicted false and predicted true values, it have two attributes "true false" and "true true" . Predicted false for true false is 1 and for true true is -1. Predicted true for true false is -1 and for true ture is 1. **Confusion matrix** predicted true for true false is 1912 and for true true is 17427. The class predication is 90.11%, and class recall for true false is 0.0, class recall for true true values is 100.0%.

4.4 Generalized linear model

The simulator most likely in generalized linear model the true reviews are 78% and the false reviews are 22%. The important factors for true values are seven that are; body: described, body: recommend and body: compare etc. **profits** from model is 15,512 and profits for best option (true) is

15,514. The gain value is 2. **Confusion matrix** the predication false for true false is 0 and for true true values is 1. The predication true for true false is 1913 and for true true values is 17426. The class predication is 90.11%. The class recalls true false values is 0.00% and true true values is 99.99%.

4.5 Fast large margin

In the fast large margin the true values is 89% and the false values is 11%. The important factors for true values are seven that are: helpful votes, rating, and body: over price etc. **optimal parameters** have two factors that are c and error rate, the c = 0.001 and the error rate is 9.9%. If the c is 1000 then the error rate is 10.1%.

4.6 Decision tree

The decision tree simulator most likely true value is 90% and the false value are 10%. The optimal parameters show the error rate for parameters. The maximal depth is 2. If the error rate of maximal depth 2, 4, 7 or 10 then the error rate remains same that is 9.9%.

Figure3: decision tree model

4.7 Text mining of review data

Figure4: review analysis

The above figure4 represents the total overall reviews of unlocked mobile phones in Amazon. Analysis the document in texts and give the sentiment reviews analysis. The true document texts are 90.10% and the false reviews value is 9.90%. In the general statistics use the name, value, and body: ability instances and give the values mimum is 0, maximum is 0.738, average is 0.0 and the standard deviation value is 0.09.

Figure5: word cloud of texts reviews

The above figure shows the texts document reviews in world cloud form. It first generates tokens and then numbers of documents for true / false values. In the end it gives the weights by correlation that contains on attributes and their weights.

4.8 Rating analysis of unlocked mobile phones in Amazon

Companies want to sell their products but also desire in the direction of obtains reaction inside conditions of customer review. Consumer's reviews about purchased products shared on the internet contain huge collision. Rating is very important factor for any company to improve business. In this research analysis the unlocked mobile phones rating in Amazon, for this purpose use the item data files. And apply the six algorithms that are generalized linear model, decision tree, gradient boosted trees, random forest and support vector machine for regression. The result shows in root means square error, absolute error, relative error, standard deviation, squared error, correlation and relative error lenient.

Figure6:rating distribution

The above figure represents the rating distribution. Rating is distributed in 5 elements that are 1 to 5.

In this research the top 10 mobile companies are choose for the reviews number analysis. In the following shows the 10 companies reviews numbers.

Figure7: top 10 companies based on review number

In the research some factors are described that are used to make predications: that are titled: wife, titled: old, titled: computer, titled: mini, titled: charger etc. every algorithm has its own important factors. The following figure shows the important factors for predication.

Figure8: important factor for predication

In texts choose the text attributes then generate token in this way make a document. Token is equal to the number of documents. After generating documents to apply training

and scoring of on total number of document. The result is shown in world cloud form the high scoring words are big. The following figure represents the training and scoring of texts ratings. In last produce the weights by correlation in tabular form.
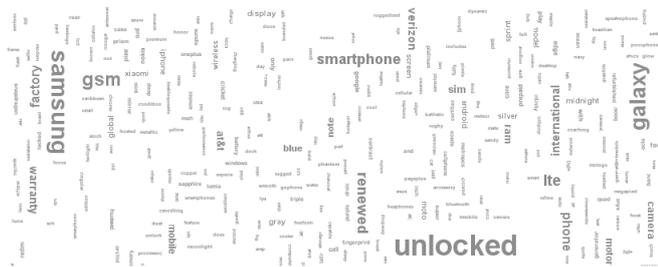

Figure9: training and scoring of text ratings

The decision tree is used to make production model for this purpose used brand that contain top 10 mobile phone companies. The other algorithm make predication chart. The flowing figure shows the production model of decision tree.

Figure10: decision tree production model

5. Results and discussion

The rapidminer give result through statistics that contains on brands summary for this purpose choose the 10 unlocked mobile phone brands and its distinct values. Count the values and shows the ratings in percentages form. The following table shows the top 10 mobile phone rating in Amazon.

Brand	count	percentage
Samsung	364	48.32%
Motorola	105	14.66%
Apple	63	8.80%
Xiaomi	46	6.42%
Nokia	44	6.15%
Google	38	5.31%
Huawei	32	4.47%
Sony	27	3.77
One plus	10	1.40
ASUS	5	0.70

Table1: brand summary

the reviews performance algorithms results is represent through accuracy, classification error, AUC, precision, recall, F measure, sensitivity and specificity. The algorithms are used for this purpose are: naïve bayes, generalized linear model, fast large margin and decision tree the accuracy and other charactistic of all algorithms are same instead AUC. The specificity of all techniques remains 0. The training time of naïve bayes model is 124ms, generalized linear model is 230ms, fast large margin is 432ms and the scoring time of naïve bayes model is 12s, generalized linear model is 13s and the fast large margin is 13s. The following table contains the performance algorithms results of reviews analysis.

Model	Accuracy	classification	AUC	Precision	Recall	F measure	Sensitivity
Naïve bayes	90.1	9.9%	50.0	90.1	100.0	94.8	100.0
Generalized linear model	90.1	9.9%	72.8	90.1	100.0	94.8	100.0
Fast large margin	90.1	9.9%	64.3	90.1	100.0	94.8	100.0
Decision tree	90.1	9.9%	50.0	90.1	100.0	94.8	100.0

Table2: algorithm results of Review analysis

The performance of rating analysis of unlocked mobile phones in Amazon the results are shown through regression characteristics that are root means square error (RMSE), absolute error, relative error, standard deviation, squared error and relative error lenient for this purpose 5 algorithms are used that are naïve bayes, decision tree, generalized linear model, random forest, gradient boosted tree and support vector machine.

Models	RMSE	Absolute error	Relative error lenient	Squared error	Correlation	Standard deviation	Relative error	predication
Generalized linear model	0.7	0.5	14.1%	0.6	0.18	1.3	14.1	3.76
Decision tree	0.66	0.47	11.8%	0.44	0.30	1.2	11.9	3.63
Random forest	0.67	0.47	11.8%	0.45	0.36	1.3	11.9	3.73
Gradient boosted tree	0.62	0.42	10.6%	0.39	0.36	1.5	10.6	3.79
Support vector machine	0.66	0.44	11.0%	0.44	0.33	1.6	11.1	3.74

Table3: algorithms result of rating analysis

5.1 Comparative study

In this part our work is compared with other associated mechanism. This comparative analysis is based on accuracy. The following table represents the comparison;

Paper title	Year	Dataset	Accuracy
Amazon reviews, business analytics with sentiment analysis.[4]	2016	Review of cell phone & accessories	80.11%
Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning. [3]	2013	reviews of books	84.44%
Mining comparative opinions from customer reviews for competitive intelligence.[5]	2011	Customer product reviews	61.00%
Amazing: A sentiment mining & Retrieval System.[5]	2009	E commerce reviews	87.60%
"Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews"[6]	2016	Review on Camera	80.00%

Table4: previous researches comparison

Distinctive researches scheduled within the table have conduct dissimilar pre-processing stepladder and characteristic descent process. As in our study we united in the direction of manage every the ancestry process and preprocessing ladder and choose the preeminent accuracy from it. In Our research the mobile phones reviews accuracy is 90.1%.

Conclusion

In this research, the unlocked mobile phones in Amazon dataset are used for ratings and reviews analysis. For this aim the 4 algorithms are apply on reviews data file and analysis the reviews given accuracy is 90.1%. and 5 algorithms are apply on items data file to analysis and count the rating, the gradient boosted tree is best option because that's RMSE is 0.62 and accuracy is high then other algorithms. This analysis is conduct on rapid miner studio. In future we apply clustering techniques and other AI algorithms for best outcome.

References

suggestion gadget using K-Menus Clustering and the K-NN algorithm. Facts BX books are derived from facts. The device is imposing strange programming language. It is seen that when the device is implemented in Python programming language, the RMSE fee of the predicted technique goes far beyond the previous method. Likewise, it has been seen that the RMSE cost of the proposed gadget is meeting the same fees as the present approach, but without a lower number of commenters. The proposed system is find accuracy through root mean square error (RMSE) that is 0.3. The proposed career can be further expanded by using additional information units. The concept of critical analysis can be used to illustrate the performance of the book's recommendation system in destiny, so the model can be adjusted to accommodate further conditions. In destiny, there may be an interest in a male or female feature that is hidden in consumer advice.

References

[1] Alam, M. S., Hossain, S. B., & Awal, S. B. (2019). Feature Based Mobile Phone Rating Using Sentiment Analysis and Machine Learning Approaches. BRAC.

[2] Alghamdi, N. S. (2019). A Comparison of Sentiment Analysis Methods on Amazon Reviews of Mobile Phones. *(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019* , (1-10).

[3] Elli, M. S., & Wang, Y.-F. (2016). Amazon Reviews, business analytics with sentiment analysis.

[4] Haque, T. U., & Saber, N. N. (2018). Sentiment Analysis on Large Scale Amazon Product Reviews. *2018 IEEE International Conference on Innovative Research and Development (ICIRD)* , (1-7).

[5] Rain, C. ((2013)). Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning. Swarthmore College.

[6] Shaikh, Tahura, & DeepaDeshpande. (2016). Feature Selection Methods in Sentiment Analysis and Sentiment Classification of Amazon Product Reviews.

[7] Tan, W. (2018). SentimentAnalysisforAmazonReviews. (1-5).

[8] YeYiran. (2019). Aspect-basedSentimentAnalysisonmobilephone reviewswithLDA . *ACM* , (1-5).

[9] Bhatt, A., Patel, A., & Chheda, H. (2015). Amazon Review Classification and Sentiment Analysis. *(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6)* , (1-4).

[10] Xu, & Kaiquan. (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. *Decision support systems 50.4* , 743-754.