

The User Analysis of Amazon Using Artificial Intelligence at Customer Churn

Mohammed Ali Alzahrani

Information Technology Department, Computing and ICT Faculty, Taif University, Taif, Saudi Arabia

Email: marzahrani@tu.edu.sa

How to cite this paper: Alzahrani, M.A. (2024) The User Analysis of Amazon Using Artificial Intelligence at Customer Churn. *Journal of Data Analysis and Information Processing*, 12, 40-48.
<https://doi.org/10.4236/jdaip.2024.121003>

Received: November 30, 2023

Accepted: February 5, 2024

Published: February 8, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Customer churns remains a key focus in this research, using artificial intelligence-based technique of machine learning. Research is based on the feature-based analysis four main features were used that are selected on the basis of our customer churn to deduct the meaning full analysis of the data set. Data-set is taken from the Kaggle that is about the fine food review having more than half a million records in it. This research remains on feature based analysis that is further concluded using confusion matrix. In this research we are using confusion matrix to conclude the customer churn results. Such specific analysis helps e-commerce business for real time growth in their specific products focusing more sales and to analyze which product is getting outage. Moreover, after applying the techniques, Support Vector Machine and K-Nearest Neighbour perform better than the random forest in this particular scenario. Using confusion matrix for obtaining the results three things are obtained that are precision, recall and accuracy. The result explains feature-based analysis on fine food reviews, Amazon at customer churn Support Vector Machine performed better as in overall comparison.

Keywords

Customer Churn, Machine Learning, Amazon Fine Food Reviews, Data Science, Artificial Intelligence

1. Introduction

Artificial intelligence is making everything easy, a few years back online shopping was a dream but in the last two decades, and it is invading the actual physical markets. The balloon of data science is touching its extreme level as the need of the era and market depends upon artificial intelligence-based solutions. Data science is keen on preparing real-time solutions based on applied science [1].

Companies turn towards e-commerce and online product selling to grow business and generate revenue.

Step-by-step processes are conducted in the first phase, the data set cleaning process is initiated, in the second phase features are selected for Amazon customer churn based on the data set, and in the third step, the machine learning algorithms are implemented. The three main algorithms are used in this task, support vector machine, k-nearest neighbor, and random forest [2]. The results were further obtained using a confusion matrix in percentage.

2. Literature Review

E-commerce is the largest growing thing on the internet using all the possible directions that are involved technically. Machine learning, business intelligence, and artificial intelligence-based solutions are a few of the best solutions developed to generate leads in e-commerce. The world's business hubs just turned after the COVID-19 breakthrough into online smart places. People overall recommend online market places more over the regular shopping markets. The trend is getting changed even in the developed world and underdeveloped countries, all just turned towards online market places [3].

Developing such online marketplaces can build the economy, can overcome the current fear of the pandemic, and develop more reliable technical smart markets to generate business leads. Developing such solution, which compete in the latest tech-based problems, especially highly involvement of artificial intelligence-based solutions. Recommending items based on the search history of the person browsing, is one of example, telling the store owner what products are generating revenue and what products are still just filling up storage space. Adding more features liking responding to the voice of the customer is a new artificial intelligence-based solution [4]. Generating bill online and sending the items or products directly to the delivery address without the involvement of the human being is what Amazon is achieving.

All aspects are part of the things either a part of business intelligence or a part of the artificial intelligence or some of the tech involved to progress fast. Customer churn is the process of realizing how many customers are still customers and how many customers are left buying the products. In terms of the business, it always matters if a platform's regular customer left its place of buying. To analyze such a process to know how many customers are left buying the product is known as customer churn [5]. The ratio of the customer that stops buying the product in a certain period is considered as the customer churn impacts the e-commerce platforms. In E-commerce platforms, Amazon comes as the leading one.

Machine learning is developing clear difference in the history of e-commerce by setting up new solutions to the problems noticed in online shopping. Dealing and manipulating the data sets and concluding some clear decision that helps in real-world problems. Machine Learning problems get resolved using linear re-

gression imparting and generating solutions that were hard to make manually, especially for unstructured data. Structured data can be handled by machine learning techniques and algorithms. Feature-based analysis helps in developing better decisions on the structured data. Machine learning paves the future by providing the optimum solutions to the problem [6]. The issues machine learning can handle in e-commerce are managing the whole platform related to sales and purchases, following up with the customers noticing the issues customer face while purchasing the product. Understanding the scientific reasoning of big data and the analytics generated using machine learning algorithms.

Optimized big data or getting patterns generated using the data of the products helps in growing the product is what artificial intelligence and business intelligence are created. Taking care of the customer's experience with the product helps increase the overall business. Improving customer concern problems and first analyzing them to create a clear terminology [7]. The terminology may involve artificial intelligence and machine learning to develop solutions in applied science. The telecom sector is the one example of that in the telecom sector feedback deducts information about user experience. That user experience feedback helps in improving services similarly, artificial intelligence and machine learning can find out meaningful information to help grow more businesses and also improve services [8].

3. Literature Review Matrix

As shown below **Table 1**, the matrix explains about the related works already published.

4. Data Set

The analysis of Amazon in terms of customer churn using artificial intelligence and machine learning algorithms would conduct using a data set that is gathered from "Kaggle" as a platform. Data-set collected to best suit the work overall [9]. This data set adds its abilities to the whole process by allowing it to process through the chosen machine learning techniques. Various data set platforms were checked to obtain the most appropriate data set. The data should be used to conduct the research while searching the data set that holds text as a property is the main key focus before choosing the data set.

Fine Food Reviews

The Data-set is selected as fine food reviews, the data set holds half a million records that are collected on the based on last past ten years of customer reviews published on Amazon up till 2012 [10]. Actually, it holds so many records so while using this data set one would remain unable to see duplication.

5. Methodology

The data set is obtained from the Kaggle, in the first step of the implementation

Table 1. Literature review matrix of published work.

Author Date	Authors Name	Methods: Study, types	Major Findings	Limitations Gaps	Implications for future research & practice
2021 [1]	Lalwani, Mishra, Chadha, & Sethi	The method used in this work is the gravitational search algorithm, used after preprocessing the data set.	It tackles the customer churn problem in the telecommunication sector, using machine learning algorithms.	Different fractions can be applied to the data set, and improvement in accuracy can be obtained.	Can be used for
2021 [2]	Gopal, & MohdNawi	Extended Support Vector Machine, KNN, PSO were discussed.	Customer identification was focused in this recent work according to customer churn.	The analytics were works on the unstructured data only.	Feature extraction and choosing require more attention.
2019 [3]	Çelik, & Osmanoglu	Different machine learning and regression models were explained.	Calculating effective price analysis keeping in view about customer churn.	General research on cost effective model is conducted.	This can be conducted specifically for a price-effective market benefiting the customers.
2019 [4]	Ahmad, Jafar, & Aljoumaa	Random Forest, Decision Tree, Gradient Boosted Machine and XGBOOST	Telecommunication sector requires a prediction about their customers. Based on social network analysis	The deficiencies were observed using feature-based analysis.	More betterment can be observed using it in an applied science area.
2020 [5]	Baker, Baugh, & Sammon	Statistical methods are used to find customer churn facts.	The features that are used were so proper if these features were selected to obtain any firm customer churn the growth will take place.	Artificial intelligence and machine learning are not used.	This can be implemented using artificial intelligence predictive models.

the data set is cleaned using data cleaning technique, empty records were removed from the data set, stemming is performed onto the text of the data set, stop words were removed, these all steps were performed in the very first phase [11]. Data is cleaned to perform machine learning algorithms to conclude some results by applying the feature-based analysis on customer churn. Mainly, three steps are considered in implementing the work, the first is data cleaning, the second is the feature for customer churn, and the third one is the machine learning algorithms. Different algorithms are used to analyze the better results and performance overall [12]. Major focus is that it remains attached with the machine learning techniques as this work stays attached as a machine learning solution. As shown below **Figure 1**, it chooses upon the best of the machine learning techniques such as SVM, KNN and Random Forest.

5.1. Data Cleaning

The step-by-step data set will be cleaned, as shown below in **Figure 2** the first phase link will be removed, Empty Records will be removed, and the data set is



Figure 1. Step by step processing.

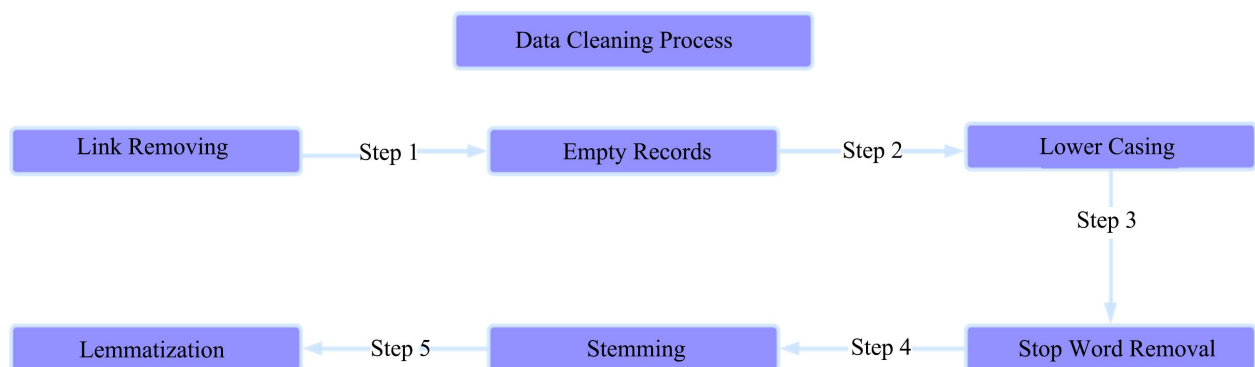


Figure 2. Data cleaning process.

fetches and considers in the cleaning process the empty records from the data set are removed in this phase. Text will be converted to the lower casing; in this step the text is converted to the lower space from the upper space. While cleaning the data-set before applying the machine learning model stop words are also required to be removed, stemming is the process, that converts the word to its accurate form or base state, it helps the machine to understand the meaning of

the words and sentences more accurately [13]. Lemmatization is the phenomenon that converts the lemma from other forms of the word to the base one.

5.2. Features for Customer Churn

Collection of the selected features is the most important thing within this process as it helps to produce better output in terms of a productive manner. Keep on checking that Amazon holds their previous customers the research is conducted based on the data set known as fine food review.

The First feature is unique reviews calculation, the second feature is positive feedback, the third feature is repetitive customers, and the fourth feature is repetitive customers based on a good review. This feature based analysis will help in maintaining the customer churn. The feature-based customer analysis produces better results of the customer churn than deploying algorithms without features [14]. So, the above key features are used to conclude better results and the betterment in customer churn. As shown below in **Figure 3**, selected features followed by the machine learning techniques. These key features will generate the optimum path toward successful research-based analysis using machine learning algorithms that are, SVM, KNN, and Random Forest.

Mainly, there were various huge data sets available but to focus it for customer churn mainly these four things are focused firstly unique reviews, positive feedback, positive feedback of repetitive customers, and repetitive customers. These features help in concluding the better customer churn.

5.3. Machine Learning Algorithms

The First step is considered as data cleaning of the data set fine food review well followed with the data cleaning steps. The second step is considered as the feature selection based on customer churn concerning Amazon. The three main algorithms are a key focus and are considered as third step in concluding the results. The methodology is deployed in three main parts, and the third part is the machine learning algorithm, Support Vector Machine is the first to be considered as it clearly explains the percentages, and K-Nearest Neighbor finds out the productive outputs in terms of results [15]. Furthermore, the third one is considered the random forest as it also performs better in the machine learning-based decisions.

5.4. Confusion Matrix

As shown below in **Table 2**, the accuracy obtained using feature-based analysis the confusion matrix is used based on the following conditions [16].

$$\text{Accuracy (\%)} = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) * 100 \quad (1)$$

$$\text{Precision (\%)} = \left(\frac{TN}{N} \right) * 100 \quad (2)$$

$$\text{Recall (\%)} = \left(\frac{TP}{P} \right) * 100 \quad (3)$$

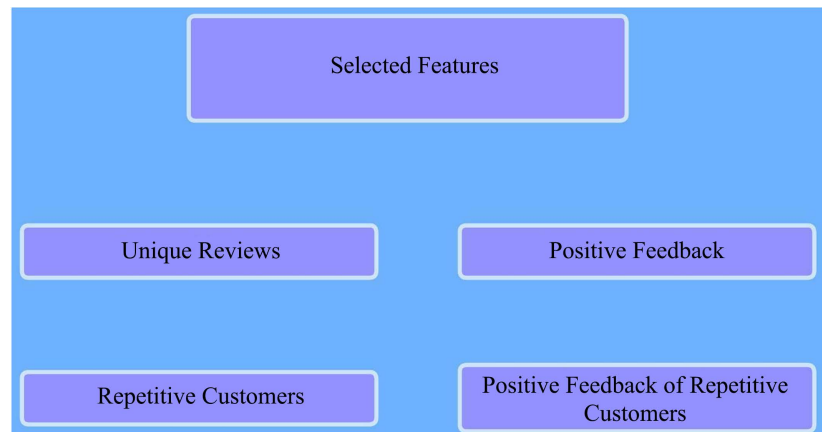


Figure 3. Selected features.

Table 2. Confusion matrix criteria.

Positive	The Attribute is accurate against the feature.
Negative	The Attribute is not accurate against the feature.
True Positive (TP)	Positive instances of feature are marked as positive.
False Positive (FP)	Negative instances of feature are marked as positive.
True Negative (TN)	Negative instances of feature are marked as negative.
False Negative (FN)	Positive instances of feature are marked as negative.

Table 3. Techniques used.

Methods	KNN	SVM	Random Forest
Accuracy	90	93	87
Precision	0.84	0.86	0.82
Recall	0.91	0.89	0.91

6. Results

As shown below in **Table 3**, the overall performance of all three methods remains progressive, even random forest performed better and make it 87% accuracy while the training data-set was at 30% and the testing data set was at 70. Similarly, on the same proportion, KNN performed 90% and the SVM performed up to the 93% accuracy with the same proportion of the training data and testing data.

Both algorithms yielded for the better results, the overall accuracy of SVM is provided better than the KNN and random forest. The overall results are observed and concluded SVM method is a far better technique as it holds proven results [17].

7. Conclusion

Machine learning is initially proving easy in deciding to the structured data, ef-

fectively. The feature-based analysis is far better in terms of concluding the circumstance regarding the data-set of fine food reviews from Amazon keeping customer churn in the view. The features that were selected are based on customer churn, which relates to the customer churn. The feedback of the customer is analyzed and considered to conclude results. While comparing three machine learning techniques SVM provided up with the 93% accuracy, KNN concluded up to 90% and the random forest achieved 87%. Overall performance of SVM remains better leading it to a better solution in term of terms feature-based analysis.

Acknowledgements

I acknowledge the efforts of Mr. Muhammad Ehtisham in this work in data collection and data analysis that make it complete.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Lalwani, P., Mishra, M.K., Chadha, J.S. and Sethi, P. (2021) Customer Churn Prediction System: A Machine Learning Approach. *Computing*, **104**, 271-294. <https://doi.org/10.1007/s00607-021-00908-y>
- [2] Gopal, P. and MohdNawi, N.B. (2021, December) A Survey on Customer Churn Prediction Using Machine Learning and Data Mining Techniques in E-Commerce. 2021 *IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Brisbane, 8-10 December 2021, 1-8. <https://doi.org/10.1109/CSDE53843.2021.9718460>
- [3] Çelik, O. and Osmanoglu, U.O. (2019) Comparing to Techniques Used in Customer Churn Analysis. *Journal of Multidisciplinary Developments*, **4**, 30-38.
- [4] Ahmad, A.K., Jafar, A. and Aljoumaa, K. (2019) Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform. *Journal of Big Data*, **6**, Article 28. <https://doi.org/10.1186/s40537-019-0191-6>
- [5] Baker, S.R., Baugh, B. and Sammon, M.C. (2020) Measuring Customer Churn and Interconnectedness. *National Bureau of Economic Research*, Working Paper 27707. <https://doi.org/10.3386/w27707>
- [6] Lu, J. (2020, April) Artificial Intelligence and Business Innovation. 2020 *International Conference on E-Commerce and Internet Technology (ECIT)*, Zhangjiajie, 22-24 April 2020, 237-240. <https://doi.org/10.1109/ECIT50008.2020.00061>
- [7] Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y.K., D'Ambra, J. and Shen, K.N. (2021) Algorithmic Bias in Data-Driven Innovation in the Age of AI. *International Journal of Information Management*, **60**, 102387. <https://doi.org/10.1016/j.ijinfomgt.2021.102387>
- [8] Libai, B., Bart, Y., Gensler, S., Hofacker, C.F., Kaplan, A., Kötterheinrich, K. and Kroll, E.B. (2020) Brave New World? On AI and the Management of Customer Relationships. *Journal of Interactive Marketing*, **51**, 44-56. <https://doi.org/10.1016/j.intmar.2020.04.002>

- [9] Saran Kumar, A. and Chandrakala, D. (2016) A Survey on Customer Churn Prediction Using Machine Learning Techniques. *International Journal of Computer Applications*, **975**, 8887.
- [10] Yarkareddy, S., Sasikala, T. and Santhanalakshmi, S. (2022, January) Sentiment Analysis of Amazon Fine Food Reviews. *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, 20-22 January 2022, 1242-1247. <https://doi.org/10.1109/ICSSIT53264.2022.9716410>
- [11] Sharedalal, R. (2019) Amazon Fine Food Reviews-Design and Implementation of An Automated Classification System. <https://etda.libraries.psu.edu/catalog/16238rns26>
- [12] Ahmed, H.M., Javed Awan, M., Khan, N.S., Yasin, A. and Faisal Shehzad, H.M. (2021) Sentiment Analysis of Online Food Reviews Using Big Data Analytics. *Elementary Education Online*, **20**, 827-836.
- [13] Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G. and Chatzisavvas, K.C. (2015) A Comparison of Machine Learning Techniques for Customer Churn Prediction. *Simulation Modelling Practice and Theory*, **55**, 1-9. <https://doi.org/10.1016/j.simpat.2015.03.003>
- [14] Kumar, S. and Kumar, M. (2019, May) Predicting Customer Churn Using Artificial Neural Network. In: Macintyre, J., Iliadis, L., Maglogiannis, I. and Jayne, C., Eds., *EANN2019: Engineering Applications of Neural Networks*, Springer, Cham, 299-306. https://doi.org/10.1007/978-3-030-20257-6_25
- [15] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., *et al.* (2019) Machine Learning and the Physical Sciences. *Reviews of Modern Physics*, **91**, 045002. <https://doi.org/10.1103/RevModPhys.91.045002>
- [16] Krstinić, D., Braović, M., Šerić, L. and Božić-Štulić, D. (2020) Multi-Label Classifier Performance Evaluation with Confusion Matrix. *Computer Science and Information Technologies*, **10**, 1-14. <https://doi.org/10.5121/csit.2020.100801>
- [17] Jiang, T., Gradus, J.L. and Rosellini, A.J. (2020) Supervised Machine Learning: A Brief Primer. *Behavior Therapy*, **51**, 675-687. <https://doi.org/10.1016/j.beth.2020.05.002>