

## Amazon Product Sentiment Analysis using RapidMiner

Nur Hasifah A Razak<sup>1</sup>, Muhammad Firdaus Mustapha<sup>2\*</sup>, Nur Ami rah Marzuki<sup>3</sup>, Nur Saidatul Sa'adiah Tajul Othamany<sup>4</sup>

<sup>1,2,3,4</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Kelantan, Bukit Ilmu, 18500 Machang, Kelantan, Malaysia

\* Corresponding author : mdfirdaus@uitm.edu.my

Received: 28 September 2022; Accepted: 02 November 2022; Available online (in press): 16 November 2022

### ABSTRACT

*Nowadays, online reviews from customers have created significance for any business especially when it comes to Amazon website. This research predicts the customer reviews based on three main categories; health and beauty, toys and games and electronics. The reviews are classified whether as positive, negative, or neutral. Sentiment Analysis is a data analysis concept in which a collection of reviews is considered, and those reviews are analyzed, processed, and recommended to the user. The dataset use in this research is collected from the Dataworld website. The research presented in this paper was carried out initially; the reviews must be pre-processed in order to remove the unwanted data before being converted from text to vector representation using a range of feature extraction techniques such as TF-IDF. After that, the dataset is classified using Naive Bayes, Decision Tree and Random Forest algorithms. The accuracy, precision and recall were implemented as performance measures in order to evaluate the performance sentiment classification for the given reviews. The result shows that Decision Tree is the best classifier with the highest accuracy for the health and beauty, and electronic categories. For the toys and games category, the best classifier with the highest accuracy is Random Forest.*

**Keywords:** Decision Tree, Naive Bayes, Random Forest, Sentiment Analysis

## 1 INTRODUCTION

In the era of the Covid-19 pandemic, the rapidity of online sales was undeniably great. It is because all transactions nowadays are done via online. E-commerce (EC) is becoming a new attraction to the entire world community. According to Cambridge Dictionary [1], e-commerce is an activity of buying and selling goods and services on the Internet, and e-commerce has rapidly grown. There are various types of e-commerce, such as Amazon, Shopee, eBay, etc. From e-commerce, customers can not only buy the product, but they also can see numerous reviews or feedbacks from other customers on the same product. The feedback can be about the product, shop services, or the delivery process. A large number of reviews makes it difficult to read and analyze it. Reviews contain two parts; good reviews and bad reviews. As is known, reviews from buyers are very important to sellers in growing sales of their goods. Sellers with the best reputation usually have a high increase in sales volume [2].

Sentiment analysis (SA) is introduced to ensure an easy review analysis process for customers. Sentiment analysis is the process of extracting subjective information from the text to help businesses understand the social sentiment of their brand, product, or service [3]. It is done by monitoring online conversations. The customer needs to be able to understand what feedback is so that they can make an informed decision about whether to buy from or not the products. The main purpose of this paper is to investigate the effects of positive or negative reviews on attitudes towards a product from Amazon.com. Classifying text into positive, negative, or neutral can be a helpful way to simplify and understand the text. Recent research has focused on sentiment analysis in content containing personal opinions. For instance, Thelwall et al. [4] have studied the concept of informality and its impact on sentiment classification. They specifically dealt with the issue of identifying slang usage and documents that lack the uniform vocabulary and spelling that a movie review database would have. To provide readers greater insight into the positive or negative sentiment, they made an effort to categorize opinions about various product components and display these individually.

This research presents three supervised machine learning algorithms which are Naïve Bayes (NB), Decision Tree (DT), and Random Forest (RF) to classify an opinion document; comparison with three distinct Amazon review datasets. Three different Amazon review datasets are health and beauty, electronics, and toys and games. This research also spots positive reviews, neutral reviews, and negative reviews with the use of this method.

## 2 LITERATURE REVIEW

Many studies have been conducted to analyze the sentiments of product reviews on Amazon. In the following, these related researches are reviewed, in terms of pre-processing techniques, feature extraction methods, methodology, and evaluation metrics. The information is gathered from a variety of sources, including Twitter and sites like consumer reviews and product comments. For example, Kumar et al. [5] concentrated on mining evaluations from the Amazon website for the three most well-known mobile phone manufacturers, the Redmi Note 3, Samsung J7, and Apple iPhone 5S. Using the Amazon API, 21,500 reviews on English-language Amazon have provided feedback, and 3,000 of those reviews were chosen at random for the experiment. Aljuhani et al. [6] [7] conducted a study on a mobile phone product dataset from Amazon that consists more than 400,000 consumer reviews. Al Amrabi et al. [8] also examined more than 1,000 reviews from Amazon.

In the pre-processing step, there are several steps and methods were conducted to enhance the quality of textual data, such as stop-word removal, tokenization, stemming, lemmatization, and Part of speech (POS) tagging. For example, Bansal and Srivastava [7] removed stop words to reduce the number of words, converted the words to lowercase, and removed both whitespace and punctuation. Tokenization is a process to separate a sentence into meaningful tokens (symbols, phrases, or words) by removing punctuation marks. The stemming process is for utilizing the roots of the word only, while lemmatization is a process of grouping the forms of a word into a single canonical form [9]. POS tagging is crucial for natural language processing to recognize different parts of speech in the text.

Fang and Zhan [10] have done sentiment analysis on product review data that used a Naive Bayes classifier for extracting subjective content and confronting polarity categorization problems, concentrating on each sentence level and document level. A general process is proposed with an adequate process description for the sentiment polarity category. For analysis, they used online product reviews on Amazon datasets. Each level of experimental categorization sentence-level and

state-level was carried out with encouraging results. The advantage of their study is to produce medium accuracy in comparison to the SVM classifier, which helps the customer to create a fast purchase decision and gives additional value to the classifier.

Bansal and Srivastava [7] used a continuous bag of words (CBOW) and skip-gram methods with four different classes of algorithms; Naïve Bayes, Support Vector Machine (SVM), Logistic Regression (LR), and Random Forest [11] to classify the customer reviews. From their experiments, Random Forest collected the highest accuracy (91%). On the other hand, word2vec [12], CBOW, and skip-gram [13] models were used to represent features.

In terms of results, Zheng et al. [14] attained the highest accuracy (81.20%) when using SVM and weighted unigram as features. Besides, Rathor et al. [15] reported that random forest scored the highest accuracy (90.66%) when used with CBOW as features. Therefore, the purpose of this study is to classify the positive, neutral, and negative reviews of buyers over three different products on Amazon.

### 3 METHODOLOGY

An overview of the suggested sentiment analysis approach for Amazon reviews for three different product kinds is given in this section. Figure 1 illustrates the steps of the present project, from data collection to model assessment for each categorization.

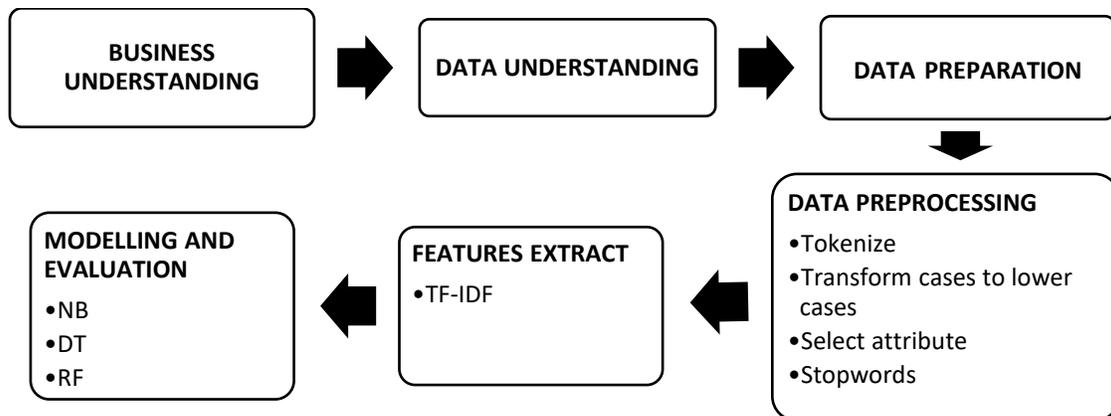


Figure 1 : The general sentiment analysis process for three different product categories' Amazon reviews.

#### 3.1 Step 1: Business Understanding

This study is based on a machine-learning algorithm analysis of the sentiment value of the standard dataset. To evaluate the review categorization algorithms, this study used the original dataset of Amazon reviews. Although there are many types of items available on Amazon.com, this study will specialize in three datasets: the Health and Beauty dataset, the Electronics dataset, and the Toys and Games dataset because all these three products have the most reviews overall. The datasets were

collected from the Dataworld [16]. The question being asked is, "How are the review sentiments for a distinct sort category at Amazon.com?" Finding out how sentiment analysis applying Naive Bayes, Decision Tree, and Random Forest works is another goal. The three datasets that were gathered are represented in Table 1.

Table 1: Number of reviews for three types of category dataset

Category	Reviews
Health and Beauty	12,071
Electronic	13,995
Toys and Games	1,676

### 3.2 Step 2: Data Understanding

The unprocessed data were acquired from the 28,333-record DataWorld website. A total of 24 columns are there, including id, date added, date updated, name, brand, categories, primary categories, etc. The review did purchase, the review does recommend, the review id, and the review number helpful all have null values in the raw dataset. A few columns from the Original Amazon Consumer Product Review dataset are shown in Figure 2. There are several columns have null and improperly formatted values.

Abc Datafiniti!Amazon!Consumer!Revi Reviews.Date	Abc Datafiniti!Amazon!Consumer!Revi reviews.dateSeen	TjF Datafiniti!Amazon!Consumer!Revi reviews.didPurchase	TjF Datafiniti!Amazon!Consumer!Revi reviews.doRecommend	# Datafiniti!Amazon!Consumer!Revi Reviews.Id	# Datafiniti!Amazon!Consum reviews.numHelpful
2017-03-02T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2016-08-31T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2016-12-06T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2016-03-28T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2017-04-22T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2016-02-17T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2017-04-17T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2017-04-16T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2016-04-07T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2017-05-24T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2016-11-21T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2017-02-26T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2017-01-05T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null
2016-11-15T00:00:00.000Z	2017-08-28T00:00:00Z	null	null	null	null

Figure 2: Sample of raw data

### 3.3 Step 3: Data Preparation

Data cleaning is used to remove some of the blank columns that would hinder the analysis process because the original dataset is difficult to model and contains missing values. Excel was used to

reduce the raw data's 24 columns to 5 columns. The five columns that will be applied in the analysis process are the id, primary categories (health and beauty, electronics, and toys and games), review rating, review text, and reviews username. Using a tableau filter's inclusion and excluding functions, null values and duplicates were eliminated. The data that has been normalized and utilized for analysis are displayed in Figure 3. It contains fewer columns than the original dataset from that time.

Id	Primary Categories	Reviews.Rating	Reviews.Text	Reviews.Username
AVpgNzjwLJeJML43Kpxn	Health & Beauty	3	I order 3 of them and one of t...	Byger yang
AVpgNzjwLJeJML43Kpxn	Health & Beauty	4	Bulk is always the less expen...	ByMG
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	Well they are not Duracell but...	BySharon Lambert
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	Seem to work as well as nam...	Bymark sexson
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	These batteries are very long...	Bylinda
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	Bought a lot of batteries for ...	ByPainter Marlow
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	ive not had any problame wit...	ByAmazon Customer
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	Well if you are looking for che...	ByLynn Pendegrass
AVpgNzjwLJeJML43Kpxn	Health & Beauty	3	These do not hold the amoun...	ByK.R.
AVpgNzjwLJeJML43Kpxn	Health & Beauty	4	AmazonBasics AA AAA batte...	BySilverLane
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	I find amazon basics batterie...	ByTXcust
AVpgNzjwLJeJML43Kpxn	Health & Beauty	3	When I first started getting t...	Byd'lsigny
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	Use it for my fish tank's light ...	ByFabiola Sarabia
AVpgNzjwLJeJML43Kpxn	Health & Beauty	5	just got em so I can't really c...	ByDavyo

Figure 3: Processed data for analysis

### 3.4 Step 4: Data Pre-processing

The reviews were tokenized, their spelling was verified, and all terms were changed to lowercase during the pre-processing stage. Stopwords are frequent words that need to be filtered out before the classifier is trained. The data was cleaned of stop words such as "a, an, you, with, etc." Based on the customer review content, each review in the dataset was classified as either positive, negative, or neutral. The process of choosing a subset of essential attributes to be used in the creation of a model is known as attribute selection in machine learning, sometimes known as feature selection. The classification accuracy and performance may be improved greatly by attribute selection. The experiment's chosen characteristics included review text and scores. After that, the dataset was divided into 20% for testing and evaluation and 80% for training. Figure 4 shows the sentiment analysis data for the health and beauty dataset.

Row No.	score	review text
1	negative	Lots of claims in the description and lots of good reviews here, so I don't know what went wrong, but it appears these batteri...
2	negative	WORST batteries EVER. I thought there would be no appreciable difference between these and Duracell. I was wrong. Mous...
3	negative	Worst performance in AA alkaline batteries we have ever seen. We bought these for use in our manufacturing company, and...
4	negative	Over half were dead or real close to dead and ones that did work lose power really fast. Very disappointed in Amazon on the...
5	negative	These are horrible! Don't waste your money. These batteries die faster than any I've ever tried. A few hours in a Wii remote a...
6	negative	Worst batteries ever. After m Duracell's in my wireless alarm keypad died after over a year I thought I would try these Amazon...
7	negative	Worst batteries ever. After m Duracell's in my wireless alarm keypad died after over a year I thought I would try these Amazon...
8	negative	DO NOT BUY. All batteries arrived completely dead. I thought it might be a bad batch. Ordered more, all arrived dead.
9	negative	I was a little surprised how large a portion of these batteries measured as dead or low out of the box. I did not notice at first ...
10	negative	Batteries were horrible. As soon as I would put a pair in my Amazon Fire stick remote, I'd get a message that batteries were I...
11	negative	These are cheap batteries. To make matters worse, Amazon sends old DEAD batteries mixed in with newer ones and ship t...
12	negative	batteries suck! brand new out of box dead or close too. Change smoke alarms every month and remote batteries every 2 we...
13	negative	WHAT A HORRENDOUS PRODUCT. I THOUGHT THAT BEING FROM AMAZON IT WOULD BE GOOD. WRONG!!!! IT'S STILL ...
14	negative	I am very disappointed in these. I put them in my remote and within 24 hrs they are completely dead. For shame, Amazon. F...
15	negative	These batteries are only OK. They don't last as long as Duracell and when they die, it's usually instantly. One moment they s...

ExampleSet (9,657 examples, 1 special attribute, 1 regular attribute)

Figure 4: Sentiment analysis data for health and beauty dataset

### 3.5 Step 5: Features Extract

The study of how natural (human) language is handled computationally is known as natural language processing (NLP). In this phase, the Term Frequency-Inverse Document Frequency (TF-IDF) approach is used to vectorize the documents from the pre-processing stage. Each technique produces a matrix that displays all of the dataset's documents as vectors. These vectors may be put into the Naïve Bayes classifier, decision trees, and random forests machine learning algorithms to create classification models for the project. Figure 5 shows the result from TF-IDF vector creator for health and beauty dataset.

Row No. ↑	score	aa	aaa	aaaa	aaas	aas	abg	abi	ability
1	negative	0	0	0	0	0	0	0	0
2	negative	0	0	0	0	0	0	0	0
3	negative	0.191	0	0	0	0	0	0	0
4	negative	0	0.232	0	0	0	0	0	0
5	negative	0	0	0	0	0	0	0	0
6	negative	0	0	0	0	0	0	0	0
7	negative	0	0	0	0	0	0	0	0
8	negative	0	0	0	0	0	0	0	0
9	negative	0	0	0	0	0	0	0	0
10	negative	0	0	0	0	0	0	0	0
11	negative	0	0	0	0	0	0	0	0
12	negative	0	0	0	0	0	0	0	0
13	negative	0	0	0	0	0	0	0	0
14	negative	0	0	0	0	0	0	0	0

ExampleSet (9,657 examples, 1 special attribute, 4,997 regular attributes)

Figure 5: Result of the TF-IDF for Health and Beauty Dataset.

### 3.6 Step 6: Modelling and Evaluation

#### 3.6.1 Modelling

The feature vector, TF-IDF, is formed before the decision tree is used as a classification procedure. According to Pavan Vadapalli, a decision tree is a prediction model or solution that facilitates decisions. Decision tree can provide precise inferences by using designs, design models, or representations that follow a tree-like structure. The datasets for three different categories: health & beauty, toys & games and electronics are generated into RapidMiner's operator Decision Tree. Based on Figure 6, the statistic indicates that it is a negative document more than 0.307 frequently includes the phrase "dead." If not, the term "dead" is in positive documents less than or equal to 0.307. Combining all these facts, a model that can label any document is created. The second algorithm to compare the accuracy for the three datasets is Naive Bayes. Naive Bayes is the fastest and simplest classification algorithm for a large chunk of data [17]. Based on Figure 7, this study evaluates the simple distribution for the label attribute (score). For class negative, the probability is 0.086, class positive is 0.775, and class neutral is 0.138. Overall, it can be stated that the review for the health and beauty product is positive. A supervised classification technique is the random forest algorithm. It is a decision tree algorithm-based ensemble learning approach [2], [11]. This ensemble approach combines the forecasts of a few base estimators built with a decision tree algorithm to better respond over a single estimator. A forest, or collection of categorization trees, is what Random Forest grows. Each tree provides its category forecast as one vote in the classification of fresh data. The category with the most votes wins, according to the forest. The accuracy results in the random forest increase with the number of trees. Figures 8 shows the random forest description.

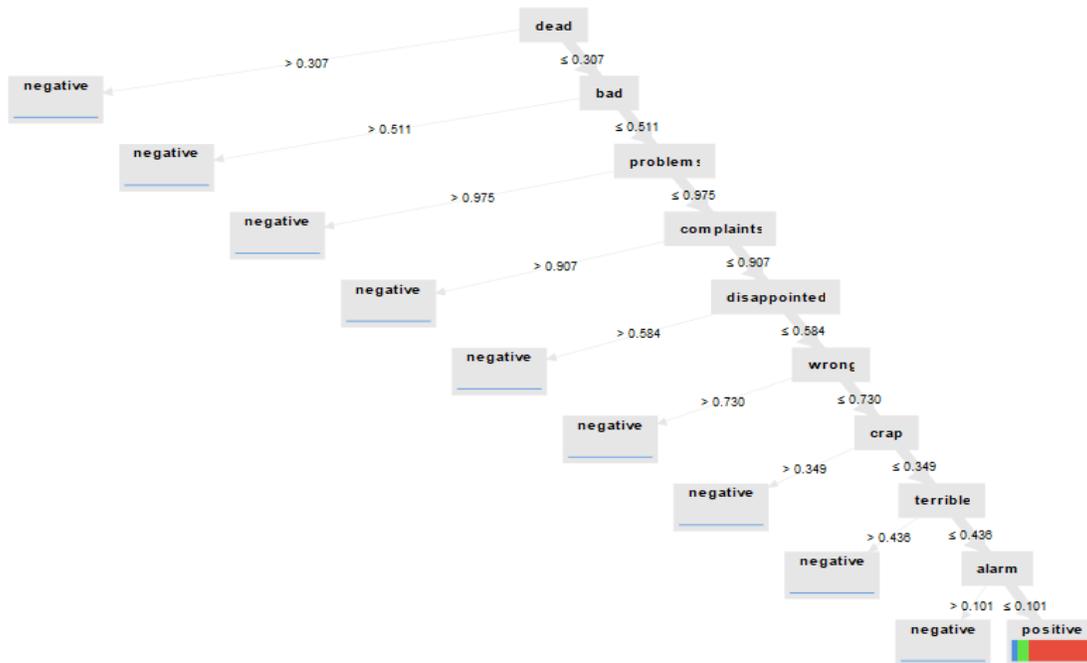


Figure 6: Decision tree model for health and beauty dataset

## SimpleDistribution

Distribution model for label attribute score

Class negative (0.086)  
5434 distributions

Class neutral (0.138)  
5434 distributions

Class positive (0.775)  
5434 distributions

Figure 7: Naïve Bayes model for health and beauty dataset

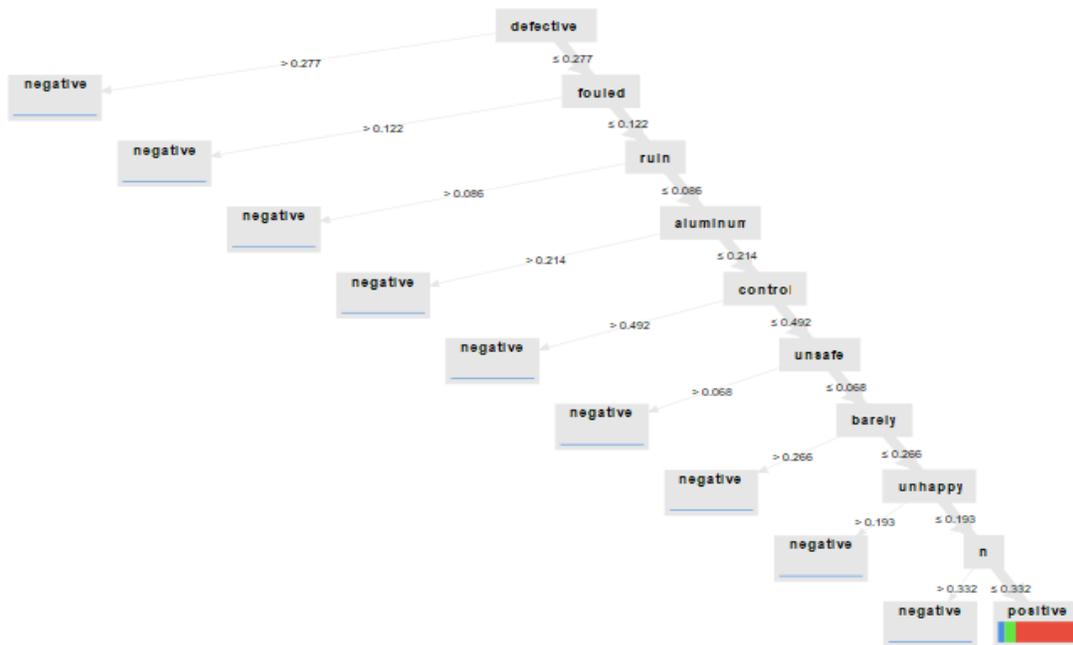


Figure 8: Random Forest model for health and beauty

### 3.6.2 Applying Model

After collecting the decision tree, the dataset's accuracy will be projected. Decision Tree and Apply Model were two of the process operators used in this process. After the process was completed, accuracy score and the prediction for the three different product reviews will be displayed. According to the Figure 9, the prediction (score) column was predicted by the system using RapidMiner, while the score column was based on the real data. The confidence (positive, negative, neutral) column

shows the accuracy value for the score (positive, negative, neutral). Repeat the process using different types of algorithms, Naïve Bayes and Random Forest.

Row No.	score	prediction(s...	confidence{...	confidence{...	confidence{...	aa	aaa	aaaaa	aaas
1	negative	negative	0.979	0	0.021	0	0	0	0
2	negative	positive	0.075	0.140	0.785	0	0	0	0
3	negative	positive	0.075	0.140	0.785	0	0	0	0
4	negative	positive	0.075	0.140	0.785	0	0	0	0
5	negative	positive	0.075	0.140	0.785	0	0	0	0
6	negative	positive	0.075	0.140	0.785	0	0	0	0
7	negative	positive	0.075	0.140	0.785	0	0	0	0
8	negative	positive	0.075	0.140	0.785	0	0	0	0
9	negative	positive	0.075	0.140	0.785	0	0	0	0
10	negative	negative	0.979	0	0.021	0	0	0	0
11	negative	positive	0.075	0.140	0.785	0	0.102	0	0
12	negative	positive	0.075	0.140	0.785	0	0	0	0
13	negative	negative	0.979	0	0.021	0	0	0	0
14	negative	positive	0.075	0.140	0.785	0.103	0	0	0

Figure 9: Prediction Analysis Using Decision Tree for Health and Beauty Dataset

### 3.6.3 Training and Validation

Cross validation is used for training and validation. Inside the cross-validation process, decision trees, apply models, and performance were applied. It is divided into two sets of data, the training set, and the testing set. The performance and application model operators were used for the testing set while the decision tree operators were used for the training set. The number of folds is 10 which means the dataset was divided into 10 parts. Figure 10 illustrates the training and testing process. Repeat the process using different algorithm, Naïve Bayes and Random Forest.

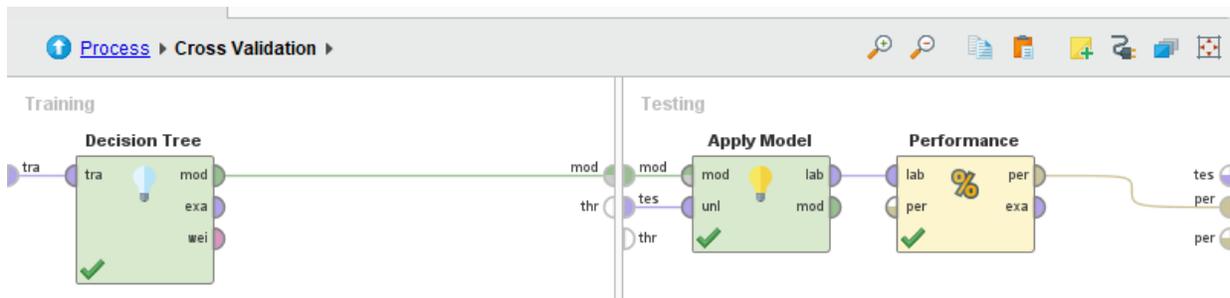


Figure 10: Testing and Training process for Decision Tree

### 3.6.4 Evaluate the Model

At this point, evaluation of the model is done to know how accurate the decision tree model is. The result for evaluation of model is obtained by combining the performance operator with the same operators from the preceding phase. But, since it just provided the output of the specified dataset for

testing, the result produced is not the final result. To determine the true accuracy, a validation process must be carried out.

#### 4 RESULTS AND DISCUSSION

Exploratory Data Analysis (EDA) is a way to visualize and interpret information that is hidden in rows and column formats [18]. In addition, by using various visualizations, it allows us to gain the most possible understanding of the dataset. The Health and Beauty reviews dataset, the Electronics reviews dataset, and the Toys and Games reviews dataset were used to identify the sentiment of three datasets. In this part, the experimental results from three distinct supervised machine learning methods are provided. One method for assessing a classifier's effectiveness is the confusion matrix. There are six possible results for a given set of a classifier and a document: true negative, false negative, true neutral, false neutral, true positive, and false positive. The confusion matrix is obtained by implementing Naïve Bayes, Decision Tree, and Random Forest algorithms. The summary of the research observations is illustrated in Figure 11. Comparison of the accuracy of several classifiers on the Health and Beauty reviews dataset and the Electronic dataset, the Decision Tree algorithm fared better than the Naïve Bayes and Random Forest algorithms while for the Toys and Games dataset, the Random Forest algorithm performed better than the Naïve Bayes and Decision Tree algorithms. To the datasets of Amazon product reviews, this study used three supervised machine learning algorithms which are Naïve Bayes, Decision Tree, and Random Forest. This study found that fully trained supervised machine learning systems might provide incredibly helpful classifications of the sentiment analysis of reviews (Negative, Neutral, Positive).

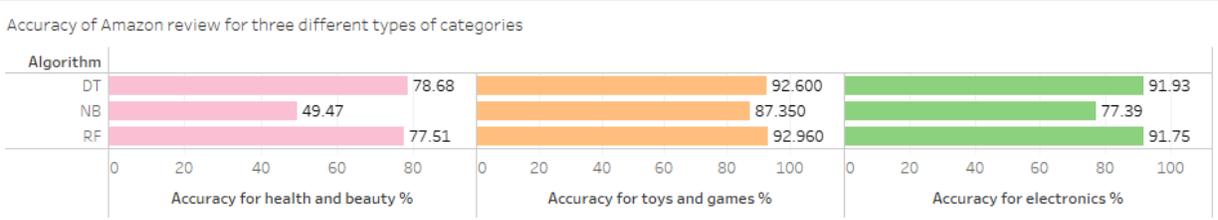


Figure 11: Graphs compared the accuracy for three types of products on Amazon .com using three different kinds of learning algorithms

From the analyzation of the accuracy for the three different algorithm, decision tree proved to be the most accurate algorithm for the electronics dataset and the health and beauty dataset, successfully classifying 78.68% of the health and beauty dataset's items and 91.93% of the electronics reviews dataset. Random Forest is the best method to analyse the Toys and Games dataset since it accurately categorized the data in 92.96% of cases. As contrast to Health and Beauty and Electronics, Toys and Games have fewer datasets (1,676 datasets), so Random Forest is a good approach for small data as it generates decision trees with bootstrap. Each decision tree receives a sample of data with replacement, increasing the likelihood that a suitable model will be created even when the data is minimal. Decision-tree classifiers SLIQ and SPRINT have been shown to attain good accuracy, efficiency, and compactness for very large datasets [19]; the latter has substantially superior computational characteristics for large datasets. Health and Beauty and Electronics both have a large number of data-12,071 datasets and 13,995 datasets, so the Decision Tree is the most suitable

algorithm to give the highest accuracy because it is simple and requires less effort to understand an algorithm.

Table 2 displays the accuracy utilizing Decision tree results from cross-validation that included training and testing. The accuracy of the Health and Beauty dataset, which includes reviews that are positive, negative, or neutral, is shown in Table 2. The model was able to predict 15.13 out of 100 for negative reviews and 0.00 out of 100 for neutral reviews from the class recall row. The predicted value for a positive review was 99.83 out of 100.

Table 2: Accuracy using cross validation for Health and Beauty Dataset

Accuracy: 78.68 % +/- 0.34% (micro average: 78.68%)

	<b>True negative</b>	<b>True neutral</b>	<b>True positive</b>	<b>Class precision</b>
<b>pred. negative</b>	158	4	16	88.76%
<b>pred. neutral</b>	0	0	0	0.00%
<b>pred. positive</b>	886	1667	9340	78.53%
<b>class recall</b>	15.13%	0.00%	99.83%	

Table 3 displays the accuracy utilizing Decision tree results from cross-validation that included training and testing. The accuracy of the Toys and Games dataset, which includes reviews that are positive, negative, or neutral, is shown in Table 3. The model was able to predict 4.29 out of 100 for negative reviews and 0.00 out of 100 for neutral reviews from the class recall row. The predicted value for a positive review was 99.42 out of 100.

Table 3: Accuracy using cross validation for Toys and Games Dataset

Accuracy: 92.60 % +/- 0.98% (micro average: 92.60%)

	<b>True negative</b>	<b>True neutral</b>	<b>true positive</b>	<b>class precision</b>
<b>pred. negative</b>	3	0	7	30.00%
<b>pred. neutral</b>	0	0	2	0.00%
<b>pred. positive</b>	67	48	1549	93.09%
<b>class recall</b>	4.29%	0.00%	99.42%	

Table 4 displays the accuracy utilizing Decision tree results from cross-validation that included training and testing. The accuracy of the Electronic dataset, which includes reviews that are positive, negative, or neutral is displayed in Table 4. The model was able to predict 1.50 out of 100 for negative reviews and 5.10 out of 100 for neutral reviews from the class recall row. The predicted value for a positive review was 99.92 out of 100.

Table 4: Accuracy using cross validation for Electronic Dataset

Accuracy: 91.93%

	<b>true negative</b>	<b>true neutral</b>	<b>true positive</b>	<b>class precision</b>
<b>pred. negative</b>	2	0	1	66.67%
<b>pred. neutral</b>	0	5	1	83.33%
<b>pred. positive</b>	131	93	2566	91.97%
<b>class recall</b>	1.50%	5.10%	99.92%	

## 5 CONCLUSION

In conclusion, this research presented the sentiment analysis of Amazon's review for three categories of products; Health and Beauty, Toys and Games and Electronics using different types of machine learning algorithms which are Naive Bayes, Decision Tree and Random Forest. These three algorithms were used to classify the dataset to find the accuracy, precision and recall from the performance measures. The methods were proposed in order to get the accuracy of the review classification for the three categories. The highest accuracy for Health and Beauty, and Electronic categories are by using the Decision Tree algorithm with accuracy of 78.63% and 91.93% respectively meanwhile for the Toys and Games is by using Random Forest algorithm with an accuracy of 92.96%. Therefore, it can be concluded that Decision Tree is the best algorithm for Health and Beauty, and Electronic categories with the large number of data as a Decision Tree is fast and works well with large data sets compared to Toys and Games category that have a less number of data that suitable using the Random Forest algorithm.

## ACKNOWLEDGEMENT

Highly appreciation to Universiti Teknologi MARA, Kelantan Branch, Malaysia for financial support and to all reviewers for their constructive comments.

## REFERENCES

- [1] "E-commerce," *Cambridge Dictionary*. [Online]. Available: <https://dictionary.cambridge.org/dictionary/english/e-commerce> (accessed Mar. 05, 2022).
- [2] M. A. Fauzi, "Random forest approach fo sentiment analysis in Indonesian language," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [3] S. Gupta. "Sentiment Analysis: Concept, Analysis and Applications." 2018. <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications->

6c94d6f58c17 (accessed Mar. 05, 2022).

- [4] G. Paltoglou, S. Gobron, M. Skowron, M. Thelwall, and D. Thalmann, "Sentiment analysis of informal textual communication in cyberspace," in *Engage (Springer LNCS State-of-the-Art Survey)*, 2010, pp. 13–25.
- [5] K. L. S. Kumar, J. Desai, and J. Majumdar, "Opinion mining and sentiment analysis on online customer review," in *2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 2016, pp. 1–4, doi: 10.1109/ICCIC.2016.7919584.
- [6] S. A. Aljuhani and N. F. Saleh, "A Comparison of Sentiment Analysis Methods on Amazon Reviews of Mobile Phones," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, pp. 608–617, 2019, doi: 10.14569/IJACSA.2019.0100678.
- [7] B. Bansal and S. Srivastava, "Sentiment classification of online consumer reviews using word vector representations," *Procedia Comput. Sci.*, vol. 132, pp. 1147–1153, 2018, doi: <https://doi.org/10.1016/j.procs.2018.05.029>.
- [8] Y. Al Amrani, M. Lazaar, and K. E. El Kadiri, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis," *Procedia Comput. Sci.*, vol. 127, pp. 511–520, 2018, doi: <https://doi.org/10.1016/j.procs.2018.01.150>.
- [9] A. Jivani, "A Comparative Study of Stemming Algorithms," *Int. J. Comp. Tech. Appl.*, vol. 2, no. 6, pp. 1930–1938, 2011.
- [10] X. Fang and J. Zhan, "Sentiment analysis using product review data," *J. Big Data*, vol. 2, no. 1, p. 5, 2015, doi: 10.1186/s40537-015-0015-2.
- [11] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [12] Y. Goldberg and O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method," *CoRR*, vol. abs/1402.3, 2014, doi: <https://doi.org/10.48550/arXiv.1402.3722>.
- [13] D. Guthrie, B. Allison, W. Liu, L. Guthrie, and Y. Wilks, "A Closer Look at Skip-gram Modelling," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, May 2006, pp. 1222–1225, [Online]. Available: [http://www.lrec-conf.org/proceedings/lrec2006/pdf/357\\_pdf.pdf](http://www.lrec-conf.org/proceedings/lrec2006/pdf/357_pdf.pdf).
- [14] L. Zheng, H. Wang, and S. Gao, "Sentimental feature selection for sentiment analysis of Chinese online reviews," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 1, pp. 75–84, 2018, doi: 10.1007/s13042-015-0347-4.
- [15] A. S. Rathor, A. Agarwal, and P. Dimri, "Comparative Study of Machine Learning Approaches for Amazon Reviews," *Procedia Comput. Sci.*, vol. 132, pp. 1552–1561, 2018, doi: <https://doi.org/10.1016/j.procs.2018.05.119>.

- [16] "Consumer reviews of Amazon products." *Data.world*, 2022. <https://data.world/datafiniti/consumer-reviews-of-amazon-products> (accessed Mar. 01, 2022).
- [17] H. Dhaduk. "Performing Sentiment Analysis With Naive Bayes Classifier!" *Analytics Vidhya*, 2022. <https://www.analyticsvidhya.com/blog/2021/07/performing-sentiment-analysis-with-naive-bayes-classifier/> (accessed Mar. 05, 2022).
- [18] A. S. M. AlQahtani, "Product Sentiment Analysis for Amazon Reviews," *Int. J. Comput. Sci. Inf. Technol.*, vol. 13, no. 3, pp. 15–30, 2021, doi: 10.5121/ijcsit.2021.13302.
- [19] M. Mehta, J. Rissanen, and R. Agrawal, "MDL-Based Decision Tree Pruning," in *First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, 1995, pp. 216–221.