

## **Sentiment Analysis on Users' Satisfaction for Mobile Banking Apps in Malaysia**

Misinem<sup>1</sup>, Tri Basuki Kurniawan<sup>2,\*</sup>, Mohd Zaki Zakaria<sup>3</sup>, Muhammad Aqil Azfar Bin Uzailee<sup>3</sup>

<sup>1</sup> Faculty of Vocational, University of Bina Darma, Palembang, Indonesia

<sup>2</sup> Faculty of Computer Science, University of Bina Darma, Palembang, Indonesia

<sup>3</sup> Faculty of Computer & Mathematics Sciences, University Technology Mara, Malaysia

\***Email:** tribasukikurniawan@binadarma.ac.id

### **Abstract**

A service offered by a bank or other financial institution that enables its users to execute financial transactions remotely using a mobile device, such as a smartphone or tablet, is known as mobile banking. The bank provides many mobile banking apps in Malaysia, such as CIMBClicks, Maybank2U, etc. Data classification has become the region of interest in Data Mining research as it indicates the arising of technology's application that leads to increment data usage. In this research, collection data and experiments are conducted, and results are shown. The preprocessing is discussed and explained in quite detail, and the results are analyzed and discussed further. The dataset was compared with five algorithms: Linear Regression, Naïve Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM). The best accuracy result shows 94.37% by the Decision Tree algorithm, and Naïve Bayes obtained the worst outcome at 66%.

### **Keywords**

Sentiment Analysis, Mobile Banking, Mobile Application, User Satisfaction

### **Introduction**

Mobile banking can be defined as a service provided by a bank or other financial institution that allows its users to conduct financial transactions remotely using a mobile device such as a smartphone or tablet.

The bank provides many mobile banking apps in Malaysia, such as CIMBClicks, Maybank2U, etc. Data classification has become the region of interest in Data Mining research as it becomes indicates arising of technology's application that leads to increment data usage (Flores, Icoy, Pena, & Gorro, 2018) (Aldiabat, Al-Gasaymeh, & Rashid, 2019) (Nair, Prabhu, Aditya, Durgalashmi, & Prabhu, 2021).

In their study, Rao et al. stated that some of us visit websites to check our mail, do online account transfers, or even merely converse. However, we are unaware that we are being watched while doing these activities (Rao, Cheng, & Dhillon, 2014).

Alzahrin, in his research, highlighted in the New Straits Time online portal as Chief Statistician Datuk Seri Dr. Uzir Mahidin said in the statement that households across Malaysia with mobile phones and computers rapidly rose to 98.1% and 74.1%, respectively, compared to 97.9% and 67.6% in 2015" (Alzahrin Alias, 2018). He said on the Statistics Department's website, "Individuals using internet aged 15 years and above in Malaysia rose by nine percentage points to 80.1 per cent in 2017, from 71.1 per cent in 2015.

Due to trust issues, some people still do not have good exposure to the benefits of using an online Mobile Banking Application (MBA). Inadequate skills or lack of trust when using mobile banking apps are the main reasons that prevent users from utilizing the MBA in their daily lives.

Nowadays, users do not feel safe using Mobile Banking Application (MBA) to do the transaction, as numerous hacking cases happened this year. According to The Stars Online, the statement from CIMB Bank Berhad highlighted that "its online banking remains secure, and all users' transactions continued to be protected." The mobile and wireless market has turned out to be one of the quickest developing markets in the world over the past couple of years. The convenience of mobile banking is that the banks will undertake the banking transactions outside working hours and can be accessed from anywhere. It has become the customer preference.

There are many reasons why some customers prefer the traditional bank transaction method (Xuan, Khoa, & Kha, 2020). Some Malaysian do not know how to use mobile banking applications. It becomes more crucial to ascertain the customers' perception of the overall service quality and their satisfaction with the current mobile banking services (N., H., & Paramashivaiah, 2013).

In (N. et al., 2013) used the method of distributing questionnaires and interviews to get the data from the user. The influential antecedent in business based on the internet is the user or customer experience because the end user often pays for most new services and products, as shown in Figure 1.

$$Y = \alpha + \beta_1 \text{BankingNeeds} + \beta_2 \text{CoreServices} + \beta_3 \text{Convenience} + \beta_4 \text{RiskPrivacy} + \beta_5 \text{ConsCont} + \beta_6 \text{ProbResol} + \beta_7 \text{CostSaved} + \beta_8 \text{InterestPolicy} + \beta_9 \text{BnkChrg} + \beta_{10} \text{FeatureAvailability} + \epsilon$$

Where the dependent variable is Y = Overall satisfaction of the customers, and the independent variables are:  
 Banking Needs = Banking Needs,  
 Core Services = Core Services of the bank, Convenience = Consumer Convenience being provided,  
 Risk Privacy = Risk & Privacy involved, Cons Cont = Consumers' continuation factors, Prob Resol = Resolution of consumers' problems, Cost Saved = Saving of Cost of consumers,  
 Interest Policy = Interest-related policies adopted,  
 Bnk Chrg = Charges levied by the bank,  
 Feature Availability = Features Available to satisfy online consumers,  $\alpha$  = Intercept, and  $\epsilon$  = Error term.

Figure 1. Proposed Regression Model

Then, the regression model was applied, and these are the result obtained at the end of the research, as shown in Table 1.

Table 1. Result of the multiple Regression Analysis-Coefficients

Variable	Unstandardized Coefficients		Standardized Coefficients $\beta$	T	Significance
	B	Standard Error			
Constant	3.31	0.069		45.013	0.000
Banking Needs	0.683	0.069	0.512	80.642	0.000
Core Services	0.629	0.069	0.509	8.619	0.000
Convenience	0.217	0.069	0.179	3.059	0.003
Risk Privacy	0.204	0.069	0.125	2.827	0.000
Cons Cont	0.141	0.069	0.108	1.882	0.000
Prob Resol	0.269	0.069	0.222	3.679	0.000
Cost Saved	0.211	0.069	0.194	3.539	0.000
Interest Policy	0.059	0.069	0.042	0.739	0.458
Bnk Chrg	0.123	0.069	0.091	1.603	0.108
Feature Availability	0.149	0.069	0.119	2.068	0.039

Based on Table 1, it can be concluded that this research found that elements including banking needs, core services, followed problem resolution, cost savings, convenience and risk, and privacy concerns can impact users' happiness when using internet banking.

In the beginning, banks faced mobile banking challenges in the early part of the decade when the first smartphones hit the market in 2007. Customers thought it was hard to see their financial information on their little phones' screens; that was basic at the turn of the 21st century.

Banks are also adopting Artificial Intelligence, such as IBM Watsons, to recommend products and care for customer queries and invest in alternative security identification, such as face ID, facial recognition, and voice ID. HSBC Bank Malaysia Bhd said that the growing

importance of digital innovation would shape the next wave of economic development as ASEAN turns 50 this year.

Then, Kaur and Mangat highlighted a few essential things before the researcher started the sentiment extraction. Firstly, the researcher needs to determine the objective and subjective text (Kaur & Mangat, 2017). The subjective text holds only the sentiments, while objective text only contains information.

### Methodology

In the first phase, the researcher will list the factors that affect the users' satisfaction when using mobile banking apps based on the previous research. The second phase is knowledge acquisition by doing a literature review to understand the proposed project. Continued with the next phase is Analysis. In this phase, the researcher will find suitable techniques for sentiment analysis to measure the users' satisfaction. The second last phase is designing a prototype by using Rapid Miner and Python. Lastly, the documentation phase involves writing the project's full report, as shown in Figure 2.

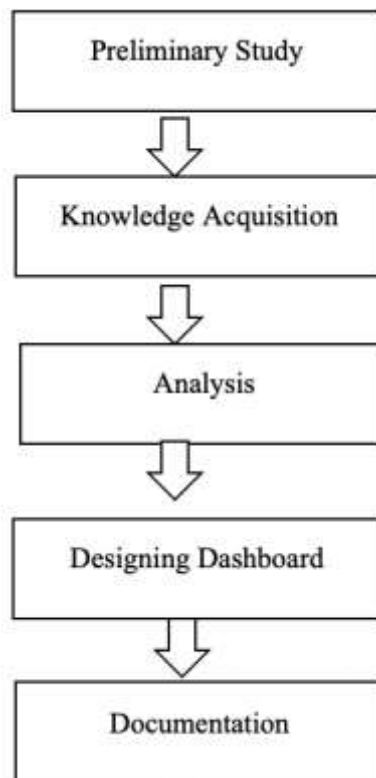


Figure 2. Process in Research Methodology Phase

The preliminary study is the collection of information phases. Figure 2 shows the reason why this phase has been done to review all the related theories and data that are relevant to be used in this research. The activities involved in this phase are to list the possible factors that affect users' satisfaction when using Mobile Banking Apps in Malaysia and do some research about the suitable

method that can be used to figure out the opportunities for better improvement by implementing Artificial Intelligence.

To find suitable techniques to develop the prototype of users' satisfaction when using mobile banking apps, the researcher needs to review the literature by reading the previous research. Many techniques can predict the users' satisfaction, but each technique gives different output to every problem, as shown in Figure 3.

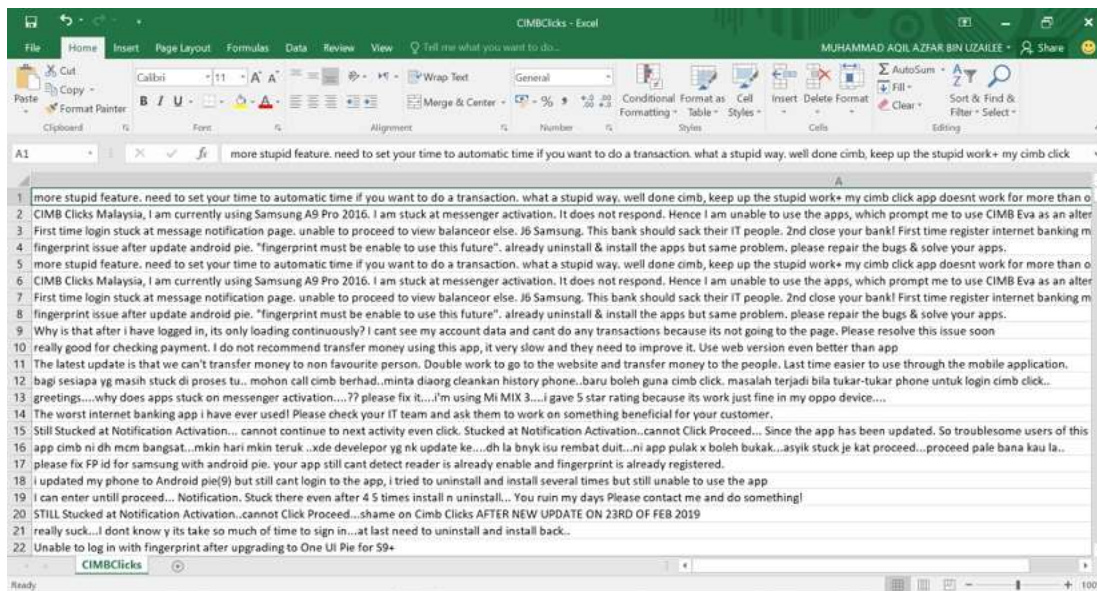


Figure 3. Sample collection of data set CIMBClicks

Google Play Store is one of the platforms that the researcher used to get the information in this project. The data was collected using Data Miner. Data Miner is one of the data scrapper extensions installed in Google Chrome. The web scrapping will extract all the vital and valuable information to develop this project. The researcher used Microsoft Excel to gather the raw data from a review of the mobile banking applications. Figure 3 illustrates the data collection open by Microsoft Excel in CSV format.

One of the major application fields for machine learning algorithms is Text Analysis. However, in the raw data, a sequence of symbols cannot be fed directly to the algorithms as most expect numerical feature vectors with fixed sized rather than raw text documents with variable lengths. The method that is being used in this project is *Bag of Words*. This method is implemented to transform real-world data which are incomplete, inconsistent, and lack understanding to be understandable. The Data Pre-Processing involved cleaning, normalization, extraction, and selection. At the end of this phase, the researcher will obtain cleaned data.

According to (Rouse, 2011), Tokenization is replacing sensitive data with a unique identification symbol that holds all important information about the data without jeopardizing security. For this project, the researcher does the Tokenization in the data set of CIMBClick's reviews from the Google Play Store, as shown in Figure 4.

```
0 [Everytime, I, want, to, make, a, transaction,...
1 [this, apps, was, great, before, the, update, ...
2 [Why, ca, n't, the, scan, fingerprint, pop-up,...
3 [no, point, updating, ..., now, need, to, chan...
4 [Everytime, I, want, to, make, a, transaction,...
5 [this, apps, was, great, before, the, update, ...
6 [Why, ca, n't, the, scan, fingerprint, pop-up,...
7 [no, point, updating, ..., now, need, to, chan...
8 [Useless, app, ,, cant, even, login, ., Also, ...
9 [This, app, is, really, suck, ,, I, ca, n't, l...
10 [The, fingerprint, to, view, quick, balance, i...
11 [This, latest, version, comes, with, new, rest...
12 [very, very, troublesome, ., info, not, clear,...
13 [The, CIMB, app, developer, team, needs, to, l...
14 [I, tried, to, login, but, the, incorrect, sec...
15 [for, security, purpose, ,, i, want, a, 2, fac...
16 [CIMB, CLICKS, ,, may, I, know, what, wrong, w...
17 [From, 4, stars, to, no, star, very, disappoin...
18 [Cant, even, log, in, after, the, latest, upda...
19 [my, cimb, clicks, get, suspended, because, I,...
20 [One, of, the, worst, banking, apps, in, Malay...
21 [after, updated, a, latest, version, ,, need, ...
22 [at, last, ,, fingerprint, authentication, for...
23 [The, most, unstable, app, i, have, installed,...
24 [Great, job, adding, fingerprint, authenticati...
25 [This, apps, is, driving, me, mad, ., I, ca, n...
26 [This, banking, application, was, one, of, the...
27 [Seriously, ,, it, 's, been, years, I, 'm, try...
28 [Used, to, love, this, app, but, it, 's, does,...
```

Figure 4. Output after the Tokenization process

One of the major processes in data pre-processing is to remove useless data. Useless data is referred to as stop words in natural language processing. Stop words is commonly used word such as "a", "the", "an" and so on that has been programmed to ignore by the search engine. Next, Figure 5 illustrates the output generated after removing stop words in the data. Then, Figure 6 shows the output generated as soon the punctuation is successfully removed from the data.

Stemming is one of the important processes in this project. It can be defined as reducing a word to its word stem that affixes to prefixes and suffixes or the roots or words, also known as a lemma. Besides, stemming is part of linguistic studies in Artificial Intelligence (AI) and morphology information extraction and retrieval.

Lemmatization is the process of grouping together the inflected form of a word, so it can be analyzed as a single item, identified by the word's lemma or dictionary form, as shown in Figure 7. After all data pre-processing is finished, the clean data is saved in CSV File entitled "CleanDataset.csv". Figure 8 illustrates the data in Excel.

Bag-of-Words is simplifying model of representation used in information retrieval and natural language processing. A text is represented as the bag of words in this model. Besides, the bag of words models has also been used for computer vision. According to (Brownlee, 2017) it is simple to implement and understand, and it has been seen with great success in problems such as document classification and language modeling. Figure 9 shows the generated output of the process, which is the Bag-of-Words model.

```
0      Everytme I want make transaction , sure I get...
1      apps great update , I would rate 5 stars , 's ...
2      Why ca n't scan fingerprint pop-up appear sele...
3      point updating ... need change password , syst...
4      Everytme I want make transaction , sure I get...
5      apps great update , I would rate 5 stars , 's ...
6      Why ca n't scan fingerprint pop-up appear sele...
7      point updating ... need change password , syst...
8      Useless app , cant even login . Also sure app ...
9      This app really suck , I ca n't login . Everti...
10     The fingerprint view quick balance usable Sams...
11     This latest version comes new restrictions . S...
12     troublesome . info clear . n't guys think type...
13     The CIMB app developer team needs learn Mayban...
14     I tried login incorrect secure word always app...
15     security purpose , want 2 factor authenticatio...
16     CIMB CLICKS , may I know wrong app , I ca n't ...
17     From 4 stars star disappointed I needed change...
18     Cant even log latest updates ... keep promptin...
19     cimb clicks get suspended I try put wrong pass...
20     One worst banking apps Malaysia . Ca n't login...
21     updated latest version , need change new passw...
22     last , fingerprint authentication XPocoF1 , ov...
23     The unstable app installed On huawei p20 . Oft...
24     Great job adding fingerprint authentication 's...
25     This apps driving mad . I ca n't log since Feb...
26     This banking application one best app banking ...
27     Seriously , 's years I 'm trying get app work ...
28     Used love app 's n't get better . Out 10 I log...
29     Eversince update , I ca n't login thru desktop...
```

Figure 5. Output Removing Stop Words

```
0      everytime i want make transaction sure i get ...
1      apps great update i would rate 5 stars s suc...
2      why ca nt scan fingerprint popup appear select...
3      point updating need change password system f...
4      everytime i want make transaction sure i get ...
5      apps great update i would rate 5 stars s suc...
6      why ca nt scan fingerprint popup appear select...
7      point updating need change password system f...
8      useless app cant even login also sure app su...
9      this app really suck i ca nt login evertime ...
10     the fingerprint view quick balance usable sams...
11     this latest version comes new restrictions st...
12     troublesome info clear nt guys think type vi...
13     the cimb app developer team needs learn mayban...
14     i tried login incorrect secure word always app...
15     security purpose want 2 factor authentication...
16     cimb clicks may i know wrong app i ca nt top...
17     from 4 stars star disappointed i needed change...
18     cant even log latest updates keep prompting s...
19     cimb clicks get suspended i try put wrong pass...
20     one worst banking apps malaysia ca nt login k...
21     updated latest version need change new passwo...
22     last fingerprint authentication xpocof1 over...
23     the unstable app installed on huawei p20 ofte...
24     great job adding fingerprint authentication s ...
25     this apps driving mad i ca nt log since febru...
26     this banking application one best app banking ...
27     seriously s years i m trying get app work it...
28     used love app s nt get better out 10 i login ...
29     eversince update i ca nt login thru desktop s...
```

Figure 6. Output after removing punctuation

Out [17]:

	Review	NoMissingValue	tokenize	RemoveStopWord	Caps	Additional	Additionalk	Stemming	Lemmatizer
0	Everytime I want to make a transaction, for su...	Everytime I want to make a transaction, for su...	Everytime, I, want, to, make, a, transaction,	Everytime I want make transaction sure get	everytime i want make transaction sure i get	everytime i want make transaction sure i get	everytime, i, want, make, transacti, sure, i, g,	everytime, i, want, make, transacti, sure, i, g,	
1	this apps was great before the update I would ...	this apps was great before the update I would ...	[this, apps, was, great, before, the, update,	apps great update i would rate 5 stars s...	apps great update i would rate 5 stars s...	apps great update i would rate 5 stars s...	[apps, great, update, i, would, rate, 5, star, s,	[apps, great, update, i, would, rate, 5, star, s,	
2	Why can't the scan fingerprint pop-up appear I...	Why can't the scan fingerprint pop-up appear I...	[why, ca, n't, the, scan, fingerprint, pop-up,	Why ca n't scan fingerprint pop-up appear sele...	why ca n't scan fingerprint pop-up appear select...	why ca n't scan fingerprint pop-up appear select...	[why, ca, n't, scan, fingerprint, pop-up, appear,	[why, ca, n't, scan, fingerprint, pop-up, appear,	
3	no point updating, now need to change password...	no point updating, now need to change password...	[no, point, updating, now, need, to, chan,	point updating need change password syst...	point updating need change password system I...	point updating need change password system I...	[point, update, need, change, password, syst,	[point, update, need, change, password, system,	

Figure 7. Output after Lemmatization

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Review	NoMissingValue	tokenize	RemoveSt	Caps	Additional	Additional	Stemming	Lemmatizer										
2	0	Everytime	Everytime	everytime	everytime	Everytime	Everytime	everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime
3	1	this	apps	this	apps	this	apps	this	apps	this	apps	this	apps	this	apps	this	apps	this	apps
4	2	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't
5	3	no	point	no	point	no	point	no	point	no	point	no	point	no	point	no	point	no	point
6	4	Everytime	Everytime	everytime	everytime	Everytime	Everytime	everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime	Everytime
7	5	this	apps	this	apps	this	apps	this	apps	this	apps	this	apps	this	apps	this	apps	this	apps
8	6	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't	Why	can't
9	7	no	point	no	point	no	point	no	point	no	point	no	point	no	point	no	point	no	point
10	8	Unless	ap	Unless	ap	Unless	ap	Unless	ap	Unless	ap	Unless	ap	Unless	ap	Unless	ap	Unless	ap
11	9	This	app	This	app	This	app	This	app	This	app	This	app	This	app	This	app	This	app
12	10	The	finger	The	finger	The	finger	The	finger	The	finger	The	finger	The	finger	The	finger	The	finger
13	11	This	latest	This	latest	This	latest	This	latest	This	latest	This	latest	This	latest	This	latest	This	latest
14	12	very	very	very	very	very	very	very	very	very	very	very	very	very	very	very	very	very	very
15	13	The	CIMB	The	CIMB	The	CIMB	The	CIMB	The	CIMB	The	CIMB	The	CIMB	The	CIMB	The	CIMB
16	14	I	tried	I	tried	I	tried	I	tried	I	tried	I	tried	I	tried	I	tried	I	tried
17	15	for	secure	for	secure	for	secure	for	secure	for	secure	for	secure	for	secure	for	secure	for	secure
18	16	CIMB	CLIC	CIMB	CLIC	CIMB	CLIC	CIMB	CLIC	CIMB	CLIC	CIMB	CLIC	CIMB	CLIC	CIMB	CLIC	CIMB	CLIC
19	17	From	4	stars	From	4	stars	From	4	stars	From	4	stars	From	4	stars	From	4	stars
20	18	Can't	even	Can't	even	Can't	even	Can't	even	Can't	even	Can't	even	Can't	even	Can't	even	Can't	even
21	19	my	amb	my	amb	my	amb	my	amb	my	amb	my	amb	my	amb	my	amb	my	amb
22	20	One	of	the	One	of	the	One	of	the	One	of	the	One	of	the	One	of	the

Figure 8. Clean Data set in Excel

0	([52.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
1	([41.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
2	([37.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
3	([54.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
4	([52.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
5	([41.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
6	([37.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
7	([54.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
8	([15.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
9	([63.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
10	([14.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0...
11	([59.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
12	([54.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
13	([54.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
14	([16.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
15	([17.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
16	([10.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
17	([8.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
18	([66.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
19	([40.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
20	([18.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
21	([61.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
22	([5.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
23	([55.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
24	([11.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
25	([15.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
26	([14.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
27	([12.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...
28	([9.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
29	([13.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0...
7534	([3.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0...

Figure 9. Output Bag-of-Words



Feature Extraction is different from Feature Selection. Feature selection is the former consists of changing arbitrary data, such as images or text, into numerical features usable for machine learning. The latter is one of the machine learning techniques applied to these features, as shown in Figure 10.

	01	020	03	032200	0362047788	05	0thnx	10	100	1000	...	0'	0%	0%0	0%	0%0	0%	0%	0%	0%	
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

Figure 10. The output of Feature Extraction

The third phase in the research methodology is Analysis. This phase is vital to ensure the researcher understands the project better by extracting the information in the first phase. The first activity involved in the analysis process is gathering the information in the preliminary study and discussing further with the supervisor about the domain in this project. All the information gathered in the first phase was revised to ensure the domain in this research is crystal clear. Next, the second activity involved in this phase is discussing with the supervisor regarding the topic and domain of the research.

The fourth phase in this research is designing the dashboard. In this phase, the researcher will develop the prototype architecture. The architecture illustrates how the prototype produces the output: users' satisfaction with using MBA, as shown in Figures 11 and 12.

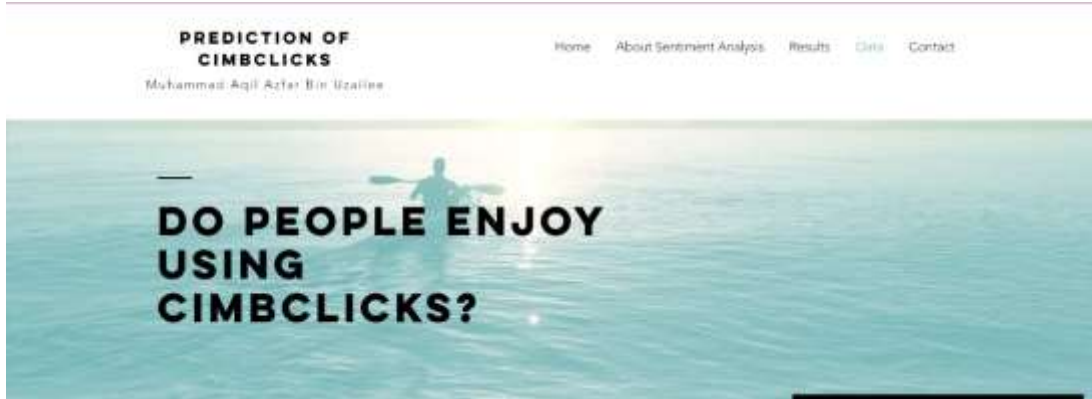


Figure 11. User Interface for Website in Prediction Review

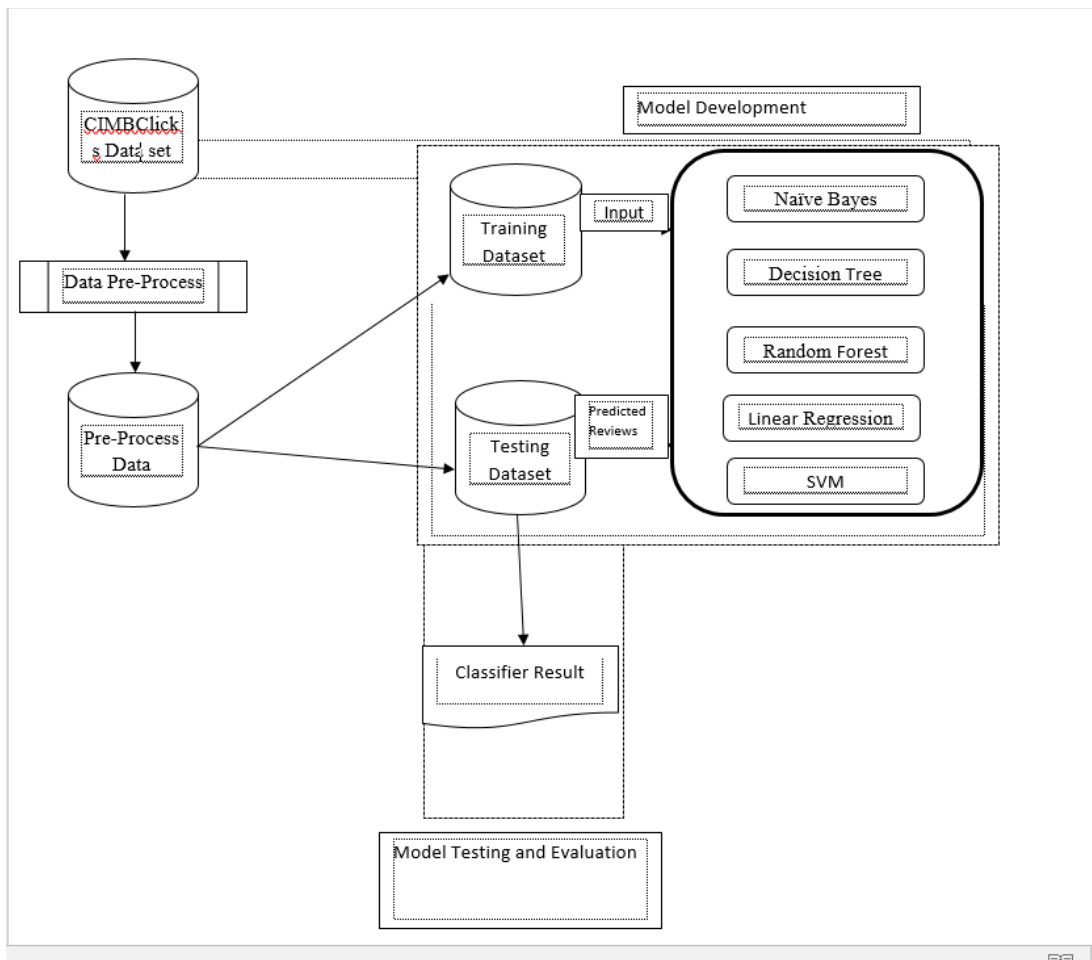


Figure 12. System Architecture for Project

Figure 12 shows the activities involved by the researcher will show all the processes involved in developing the prototype detail.

## Results and Discussion

The scrapped dataset was 7454 reviews from Google Play Store websites about CIMBClicks mobile application. The researcher scrapped the reviews from Google Play Store by using the Data Miner extension. The data will be saved in Excel file format .csv after scrapping, as shown in Figure 13.

Review	
0	Everytime I want to make a transaction, for su...
1	this apps was great before the update,I would ...
2	Why can't the scan fingerprint pop-up appear t...
3	no point updating...now need to change passwor...
4	Everytime I want to make a transaction, for su...

Figure 13. Output Read Data File CSV

There are five classifiers used in this project to generate the output, which is a prediction of the CIMBClicks reviews, which are Linear Regression, Naïve Bayes, Decision Tree, Random Forest, and Support Vector Machine (SVM). Each of the classifiers produces a different output of accuracy.

The Linear Regression algorithm is running, and the result was obtained as shown in Figure 14. Linear Regression is a common and basic used type of predictive Analysis. In Linear Regression, the connections are demonstrated utilizing direct indicator works whose obscure model parameters are assessed from the information.

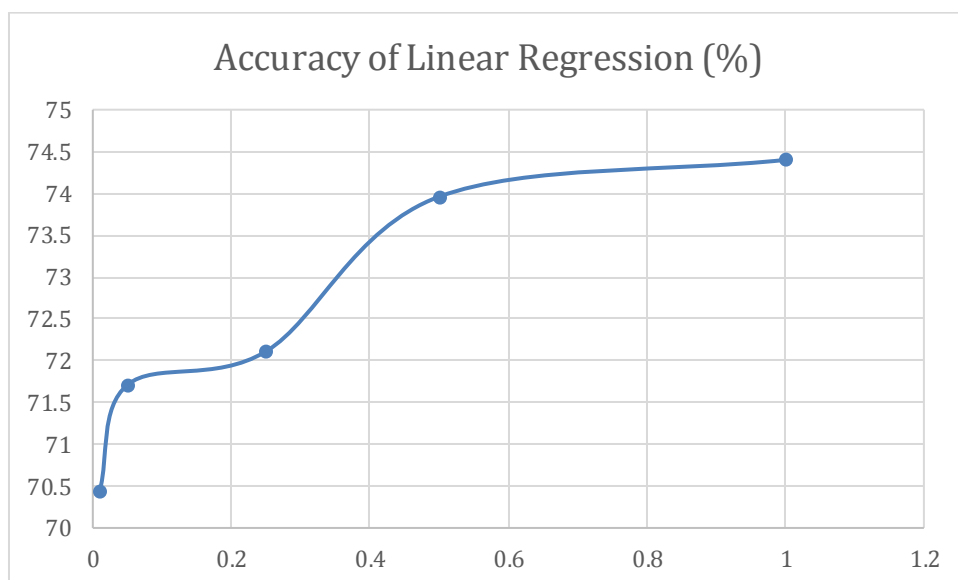


Figure 14. Accuracy graph for Linear Regression

Figure 22 illustrates the graph of accuracy for the Linear Regression model in this project. The X-axis represents the value of the constant, c, and the Y-axis represents the accuracy for each value constant inserted into the formula to train the data.

Naïve Bayes is a family of simple "probabilistic classifiers" based on applying Bayesian's Theorem between the features with strong assumptions in machine learning. Since the 1960s, Naïve Bayes has been extensively studied. The result for Naïve Bayes obtained, as shown in Figure 15.

	precision	recall	f1-score	support
0	0.50	1.00	0.67	1794
1	1.00	0.49	0.65	3500
avg / total	0.83	0.66	0.66	5294

Final Accuracy: 0.6601813373630525

Figure 15. Accuracy for Naïve Bayes

Figure 15 shows the output after running the model using the Naïve Bayes classifier to predict the reviews of CIMBClicks from Google Play Store. The final accuracy achieved by Naïve Bayes is 66.01%.

The decision Tree is one technique that supports decision tools that uses a tree-like model of decisions that includes chance, the event outcome, utility, and resource cost. The algorithm was running, and the result was obtained, as shown in Figure 16.

	precision	recall	f1-score	support
0	0.86	0.99	0.92	1794
1	1.00	0.92	0.96	3500
avg / total	0.95	0.94	0.94	5294

Final Accuracy: 0.943709860219116

Figure 16. Accuracy of Decision Tree

Figure 16 shows the generated accuracy of the Decision Tree model that predicts reviews of the CIMBClicks data set. The final accuracy of the model is 94.37 per cent which is the highest accuracy among the other classifiers in this project.

Random Forest is an ensemble learning method for regression, classification, or other tasks that operates by constructing outputting the class and constructing a multitude of decision trees that is the mean predicting of the individual trees or mode of the classes. The Random Forest result is shown in Figure 17.

	precision	recall	f1-score	support
0	0.86	0.97	0.91	1794
1	0.98	0.92	0.95	3500
avg / total	0.94	0.94	0.94	5294

Final Accuracy: 0.9369097091046468

Figure 17. Accuracy of Random Forest

Figure 17 illustrates the result of predicting the reviews of CIMBClicks by implementing the Random Forest classifier in this model. The accuracy of Random Forest generates is 93.69% after the training process, and it is the second highest among the other classifiers used in the project.

Support Vector Machine (SVM) is one of the supervised learning machine models that is associated learning algorithm that use to analyze the data used for regression analysis and classification. Then, the result of the SVM algorithm was obtained and shown in Figure 18.

	precision	recall	f1-score	support
0	0.82	0.80	0.81	1776
1	0.90	0.91	0.91	3518
avg / total	0.87	0.88	0.88	5294

Final Accuracy: 0.8755194559879108

Figure 18 Accuracy of Random Forest

Figure 18 shows the generated output by running the SVM model in this project. The accuracy of this model is 87.55% and the third highest accuracy among the other classifiers, as shown in Table 2 and the graph of comparison results in Figure 19.

Classifier	Accuracy (%)
Linear Regression	74.40
Naïve Bayes	66.01
Decision Tree	94.37
Random Forest	93.69
Support Vector Machine (SVM)	87.55

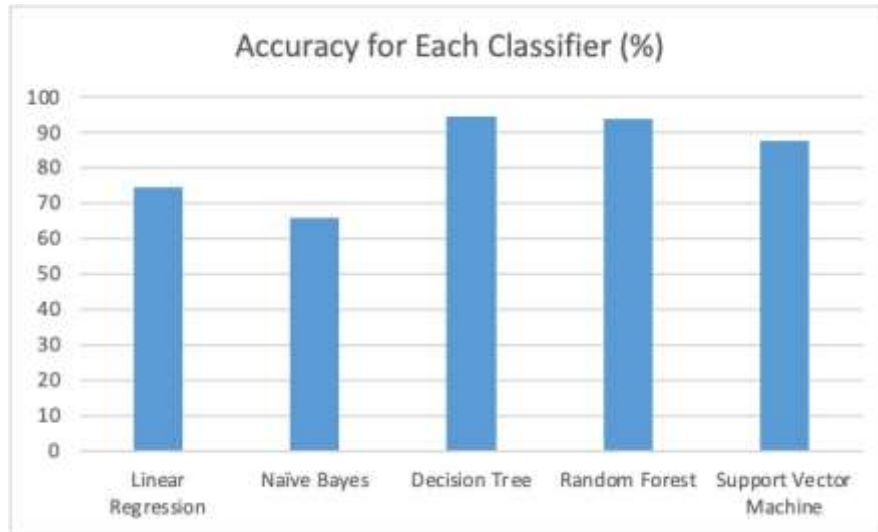


Figure 19. Bar Chart for Each Classifier

Data Visualization is a general term that explains any effort to help people understand the significance of the data by placing it in a visual context. Trends, patterns, correlations, and so on that might go undetected in text-based data can be recognized and exposed easier with data visualization software. Confusion Matrix and Word Cloud are used to visualize the data for this project.

Word Cloud, known as tag clouds or text clouds, work simply, which is the bolder and bigger appearance in word cloud as the more specific word appear in a source of textual data.

There are a few instances when word clouds excel, which understands how users feel about Malaysia's Mobile Banking Application (MBA). Text cloud visualization can turn users' feedback from a pile of information to something immediately valuable to the developer that drives positively in the future. Besides, the researcher can identify new terms to target. Using a word cloud makes the researcher aware of potential keywords to target that the site content already uses, as shown in Figure 20.



Figure 20. Word Cloud of Data set

Figure 20 shows the code using Python to generate the Word Cloud to visualize to the user what terms are frequently used by the user to give feedback to the developer, whether it is contributing to positive, negative, or neutral feedback. Figure 29 shows the Word Cloud generate from the data set consisting of reviews CIMBClick's users.

## References

- Aldiabat, K., Al-Gasaymeh, A., & Rashid, A. (2019). The Effect of Mobile Banking Application on Customer Interaction in the Jordanian Banking Industry. *International Journal of Interactive Mobile Technologies (IJIM)*, 13, 37. <https://doi.org/10.3991/ijim.v13i02.9262>
- Alzahrin Alias. (2018). Malaysia's Internet penetration is now 85.7 percent. In *New Straits Times*. Retrieved from <https://www.nst.com.my/business/2018/03/346978/malaysias-internet-penetratio-now-857-cent>
- Brownlee, J. (2017). *A Gentle Introduction to the Bag-of-Words Model*. Retrieved from Machine Learning Mastery: [machinelearningmastery.com/gentle-introduction-bag-words-model/](http://machinelearningmastery.com/gentle-introduction-bag-words-model/).
- Flores, A. C., Icoy, R. I., Pena, C. F., & Gorro, K. D. (2018). An Evaluation of SVM and Naive {B}ayes with SMOTE on Sentiment Analysis Data Set. *IEEE*, 1–4.
- Kaur, H., & Mangat, V. and Nidhi. (2017). A Survey of Sentiment Analysis Techniques. *IEEE*, 921–925.
- N., M., H., M., & Paramashivaiah, D. P. (2013). A Study on Customer Satisfaction towards Online Banking services with reference to Bangalore city. *Acme Intellectuals International Journal of Research in Management*, 1–18.
- Nair, A. B., Prabhu, K. S., Aditya, B. R., Durgalashmi, C. V., & Prabhu, A. S. (2021). Study on the Usage of Mobile Banking Application during COVID-19 Pandemic. *Webology*, 18(SpecialIssue2), 190–207. <https://doi.org/10.14704/WEB/V18SI02/WEB18066>
- Rao, A. A., Cheng, L. F., & Dhillon, J. S. (2014). A Preliminary Study on Online Data Privacy Frameworks. *IEEE*, 15–20.
- Rouse, M. (2011). *tokenization*. Retrieved from <https://searchsecurity.techtarget.com/definition/tokenization>
- Xuan, L. T. T., Khoa, T. T., & Kha, N. T. (2020). Reasons for customers reluctance to use electronic payments – A study in Ho Chi Minh City. *Hcmcoujs - Economics and Business Administration*, 9(2). <https://doi.org/10.46223/hcmcoujs.econ.en.9.2.155.2019>