CHAPTER 4
ANALYSIS AND DESIGN

4.1. Analysis

This project uses Support Vector Machine (SVM) algorithm for analyzing and predicting sentiment wheter the result is positive or negative. The programming language used in this project is Python, through the file IPython notebook (ipynb) editor which different than ordinary Python (py) file although all of them use Python for programming language. This file is done by Jupyter Notebook which are installed on Anaconda pack along with Spyder text editor. After the main dataset obtained from Kaggle dataset and prediction dataset obtained from data crawling, the main dataset used to make sentiment analyzer and the prediction dataset used to predicting sentiment determining if it is positive or negative. This is the example of main dataset and prediction dataset in below:

Table 4.1. Main Dataset (After remove asin, name, title, verified column)

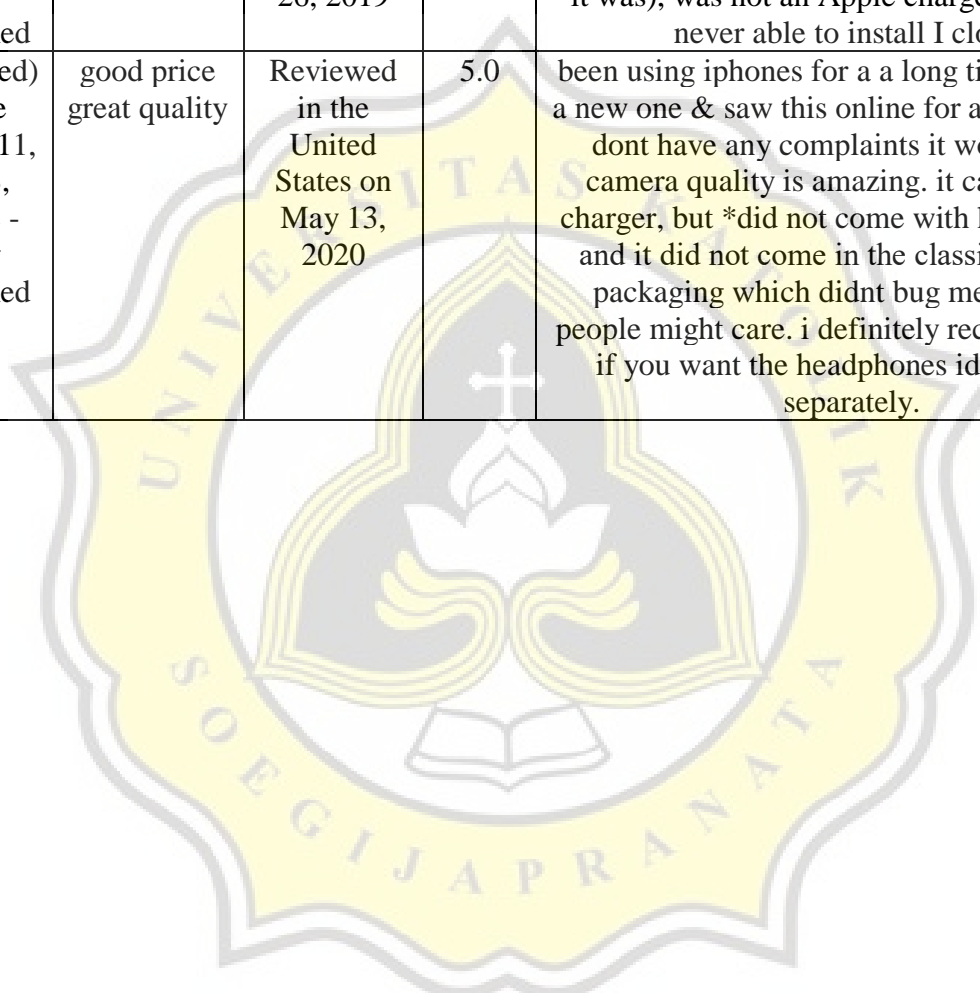| overall	reviewTime	reviewText	helpfulVotes
3	2005	I had the Samsung A600 for awhile which is absolute doo doo. You can read my review on it and detect my rage at the stupid thing. It finally died on me so I used this Nokia phone I bought in a garage sale for $1. I wonder y she sold it so cheap?... Bad: ===> I hate the menu. It takes forever to get to what you want because you have to scroll endlessly. Usually phones have numbered categories so u can simply press the # and get where you want to go. ===> It's a pain to put it on silent or vibrate. If you're in class and it rings, you have to turn it off immediately. There's no fast way to silence the damn thing. Always remember to put it on silent! I learned that the hard way. ===> It's so true about the case. It's a mission to get off and will break ur nails in the process. Also, you'll damage the case each time u try. For some reason the phone started giving me problems once I did succeed in opening it. ===> Buttons could be a bit bigger. Vibration could be stronger. Good: ===> Reception is not too shabby. I was using it in the elevator which is a remarkable feat considering my old phone would lose service by simply putting it in my pocket. ===> Compared to my old Samsung, this phone works	1.0

		quite well. The ring tones are loud enough to hear and the phone actually charges quickly and has great battery life. It doesn't heat up like a potatoe in the oven either during long phone convos. ===> Nice bright, large screen. ===> Cute ways to customize it. Scroll bar can be set to purple, pink, aqua, orange, etc. Overall: Okay phone. It serves its purpose but definitely pales in comparison to these new phones coming out from Sprint. Why get so so when you can get great?	
1	2004	Due to a software issue between Nokia and Sprint this phone's text messaging capabilities don't work with Sprint's system and won't until a software patch comes out ""some time in the next few months"". You will have to spend at least 1 hour with Sprint's award winning customer service team to find someone who will admit this to you. The problem is that Nokia designed their phones so that incoming messages are retrieved quickly and then viewed ""offline"" the way most providers work. Sprint, however, likes to have people hook up to their server first and then stay connected, burning minutes while they check their inbox, compose a reply and wait for the Sprint server to respond so they can send it out. Innovation in money-making at its finest.	17.0
5	2003	This is a great, reliable phone. I also purchased this phone after my samsung A460 died. The menu is easily comprehendable and speed dialing is available for around 300 numbers. Voice dialing is also a nice feature, but it takes longer than speed dialing. The only thing that bothers me is the games...Nokia seems to have taken snake (1 and 2) off their phones. There is a skydiving game, bowling, and tennis (like pong). The ringers are very nice, and a feature is available to choose a different ringer for each person calling. However, ringtones are not available online to download to this phone. You're pretty much stuck with what you have. There are vibrating ringtones and regular (midi) polyphonic tones. All they need are covers in a reasonable price range...	5.0
3	2004	Love the Phone, BUT...!","I love the phone and all, because I really did need one, but I didn't expect the price of the bill when I received one. Also, I've had my phone for a little over two months now and still have yet to receive my free accessories that were supposed to come with the phone. Every time I call the company, they keep telling me to wait a couple of weeks, and that I should be receiving it shortly. Other than that, I do love	1.0

		the phone and all that I am able to do with it; and I'm not just talking about making the phone calls! :)	
4	2005	Great phone service and options, lousy case!","The phone has been great for every purpose it offers, except the day i bought it-i couldnt get the case off. You can take the case off to put your own pictures in the jaket which is super cool, but when i took it back to the store-an employee said they were meant to be hard to get off. Well, he could barely even get it off- he came close to snapping the casing in half. I was never able to get it off after that which isnt THAT big of a deal, but it got dirty under the clear casing and it looks really dirty now. MAKE SURE YOU CAN GET YOUR CASE ON AND OFF! Oh and you have to charge it ALL off the time.	1.0

Table 4.2. Prediction dataset

product	title	date	rating	body
(Renewed) Apple iPhone 11, 64GB, Purple - Fully Unlocked	The iPhone review	Reviewed in the United States on January 25, 2020	5.0	My iPhone came with a big scuff on the corner (aluminum frame) and a small (not really noticible) scratch on the front and back of the phone. I got the 64 gb product red unlocked iPhone 11 which was, I believe a Verizon unlocked phone. My phone came in the iPhone 11 apple box with a original apple charger and original apple brick. When I set up my phone I just put in my T-Mobile sim and the phone started working with my data. I did not need to get it activated or anything.
(Renewed) Apple iPhone 11, 64GB, Purple - Fully Unlocked	My phone was damaged	Reviewed in the United States on December 28, 2019	1.0	My phone re start over and over again, that's so sad because the i phone looks like new☺
(Renewed) Apple iPhone 11, 64GB, Purple - Fully Unlocked	Highly Recommend	Reviewed in the United States on April 3, 2020	5.0	I received my phone a day early and it came exactly as I imagined. I was a little nervous buying a refurbished phone on the internet, but I truly have no complaints. The phone came in perfect, like new condition and has been working great with my carrier and in general.Came with a charger and charging block as mentioned in the description. It also came with a screen protector that was already put on the screen! That was an awesome and much appreciated touch. There's also 60 day

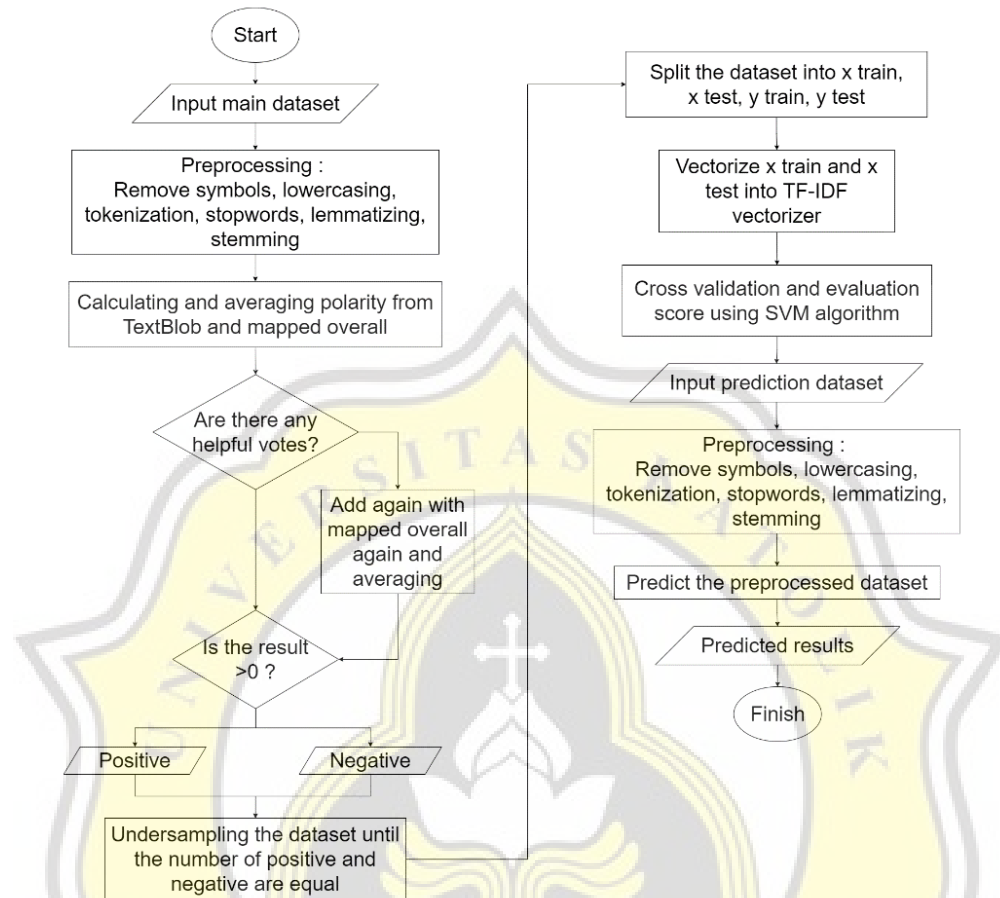|				warranty and free returns just in case something were to happen. I'm incredibly pleased. Would definitely recommend to anyone!
(Renewed) Apple iPhone 11, 64GB, Purple - Fully Unlocked	Misled,	Reviewed in the United States on December 26, 2019	1.0	Very disappointed with this purchase: Was same price as a brand new Apple I-Phone. This phone had 2 deep scratches on screen, did not come with earphones, was not an Apple charging cord, (although it had a sticker saying it was), was not an Apple charger. Also was never able to install I cloud.
(Renewed) Apple iPhone 11, 64GB, Purple - Fully Unlocked	good price great quality	Reviewed in the United States on May 13, 2020	5.0	been using iphones for a a long time. i needed a new one & saw this online for a good price. i dont have any complaints it works great. camera quality is amazing. it came with a charger, but *did not come with headphones* and it did not come in the classic "iPhone" packaging which didnt bug me but other people might care. i definitely recommend but if you want the headphones id buy them separately.

4.2. Design

Figure 4.1 Flowchart of analyzing and predicting sentiment result

After get the main dataset from Kaggle and the prediction dataset from crawling the dataset, the main dataset inputed for making prediction analyzer. Then, it passed through by preprocessing such as removing symbols, lowercasing, tokenization, stopwords, lemmatizing, and stemming. After that, the preprocessed datas are processed by TextBlob library for displaying the polarity results which ranged from -1 until 1 and the overall of the preprocessed datas which have 1, 2, 3, 4, and 5 overalls are mapped into -1, -0.5, 0, 0.5, and 1 suited into TextBlob polarity. If there are any helpful votes on them, the mapped overalls are added in this sentiment value. Then, all of sentiment values are calculated and then average them into a sentiment mean. If the result are >0, then sentiment is positive. But if the result are <=0, the result are negative.

Before the preprocessed data is splitted, It processed through undersampling because of the larger number of positives than negatives so the positive sentiments are sampelled into equal

23

number of negatives. Then, the preprocessed data is splitted into x train, x test, y train, and y test which the test size of them compared into 0,1 – 0,5 which are the common test size used in train test split and in random state of them are 42 which the common number random sampling used. After splitted, the splitted datasets are vectorize using Term Frequency–Inverse Document Frequency (TF-IDF) calculation. Subsequently, the vectorized of them are processing 5 and 10 cross validation and evaluating the scores such as accuracy, f1, precision, and recall using SVM algorithm which compared based from its kernels. Besides that, there are some confusion matrix which contains true positive, true negative, false positive, and false negative.

After evaluating and cross validating, the prediction dataset inputted to analyzing and predicting wheter each of its datas are classified as positive or negative. Before analyzing and predicting, it passed through by preprocessing such as removing symbols, lowercasing, tokenization, stopwords, lemmatizing, and stemming. This result of it are displaying positive or negative after automatically predicted.