

## CHAPTER 4

### ANALYSIS AND DESIGN

#### 4.1. Analysis

This research uses SVM as the classification algorithm. The data used is 3000 twitter data that has gone through preprocessing and sentiment determination using the lexicon based method. When determining the sentiment of a sentence, the determination is made by adding up the n polarity scores of words in the sentence. The results of the classification will be tested using the k-folding method with a total of k = 5.

##### 4.1.1. Crawling data

The data is taken from Twitter with message specifications, favorite count, retweet count, created at, username, and followers. The data for retweets is not taken. Then the data is saved in CSV format.

**Table 4.1. Example Data**

no	Message	Favorite Count	Retweet Count	Created At	Username	Followers
0	Jokowi Utamakan Pemerataan dukung pemindahan IKN <a href="https://t.co/NQB8lvhoOP">https://t.co/NQB8lvhoOP</a>	0	0	2022-06-12 02:19:15	clarke	848
1	"ASN hingga TNI Pindah ke IKN Nggak Perlu Beli Rumah Sudah Disediakan!...Jokowi Utamakan Pemerataan <a href="https://t.co/No48j4wmUf">https://t.co/No48j4wmUf</a> ."	0	0	2022-06-12 02:18:03	Ronnie Adi	5
2	"Perencanaan Pembangunan IKN Jokowi Utamakan Pemerataan <a href="https://t.co/9ZjZf6c6ah">https://t.co/9ZjZf6c6ah</a> "	0	0	2022-06-12 02:17:38	Baim Saputra	1
3	"Pembangunan IKN Menggerakkan Banyak Sektor Ekonomi Jokowi Utamakan Pemerataan <a href="https://t.co/RG0VqQhRi7">https://t.co/RG0VqQhRi7</a> "	0	0	2022-06-12 02:17:36	Baim Saputra	1

4	"IKN Nusantara Diyakini Bakal Gerakkan Banyak Sektor Ekonomi Jokowi Utamakan Pemerataan <a href="https://t.co/JShzIfI80F">https://t.co/JShzIfI80F</a> "	0	0	2022-06-12 02:17:33	Baim Saputra	1
5	"Kawasan IKN Nusantara Bakal Ditanami Tumbuhan Endemik Dari 34 Provinsi di Indonesia Jokowi Utamakan Pemerataan <a href="https://t.co/x8flJJJanO1">https://t.co/x8flJJJanO1</a> "	0	0	2022-06-12 02:17:31	Baim Saputra	1
6	"3 Desain Terbaik Istana Wakil Presiden hingga Legislatif IKN Segera Diungkapkan Jokowi Utamakan Pemerataan <a href="https://t.co/98u1eAEJtm">https://t.co/98u1eAEJtm</a> "	0	0	2022-06-12 02:17:28	Baim Saputra	1
7	"#SonoraPagi Perusahaan telekomunikasi PT Telkom Indonesia (Telkom) berencana membangun pusat data (data center) di beberapa kota di Tanah Air. Salah satunya adalah Ibu Kota Negara (IKN) baru. <a href="https://t.co/PszPxI6Qvd">https://t.co/PszPxI6Qvd</a> "	0	0	2022-06-12 02:17:27	sonorafm96.7	24978
9	"@VIVAcoid Program utang Program mobil esemka Program IKN biar gk mangkrak Program buzerRp"	0	0	2022-06-12 02:16:52	aa' huda	1
10	Beban Jakarta yang Berat Menjadi Urgensi Pemindahan IKN. Jokowi Utamakan Pemerataan <a href="https://t.co/uRG7Uhl3nY">https://t.co/uRG7Uhl3nY</a>	1	0	2022-06-12 02:16:40	Datuk Rajo Viral	795

#### 4.1.2. Preprocessing

In this process, the data that has been obtained is processed to remove characters that can interfere with the sentiment determination process such as Unicode, emoticons, URLs, don't forget to also do lowering case so as not to interfere with the TF-IDF

calculation because capital letters can change the value of the word. Furthermore, the data that is still in the form of a sentence is converted into a token in the Tokenize process. The library used in this process is Corpus NLTK. This token is then filtered from the stopword using the Sastrawi library. Finally, this data is destemming to remove suffixes on tokens that can obscure the meaning of the word

**Table 4.2. Preprocessing Example Data**

Hasil	Token Result
0,jokowi utamakan pemerataan dukung pemindahan ikn ,"	'jokowi', 'utama', 'perata', 'dukung', 'pindah', 'ikn']"
1,asn hingga tni pindah ke ikn nggak perlu beli rumah sudah disediakan jokowi utamakan pemerataan ,"	'asn', 'tni', 'pindah', 'ikn', 'nggak', 'beli', 'rumah', 'sedia', 'jokowi', 'utama', 'perata']"
2,perencanaan pembangunan ikn jokowi utamakan pemerataan ,"	'rencana', 'bangun', 'ikn', 'jokowi', 'utama', 'perata']"
3,pembangunan ikn menggerakkan banyak sektor ekonomi jokowi utamakan pemerataan ,"	'bangun', 'ikn', 'gerak', 'sektor', 'ekonomi', 'jokowi', 'utama', 'perata']"
4,ikn nusantara diyakini bakal gerakkan banyak sektor ekonomi jokowi utamakan pemerataan ,"	'ikn', 'nusantara', 'yakin', 'gerak', 'sektor', 'ekonomi', 'jokowi', 'utama', 'perata']"
5,kawasan ikn nusantara bakal ditanami tumbuhan endemik dari provinsi di indonesia jokowi utamakan pemerataan ,"	'kawasan', 'ikn', 'nusantara', 'tanam', 'tumbuh', 'endemik', 'provinsi', 'indonesia', 'jokowi', 'utama', 'perata']"
6, desain terbaik istana wakil presiden hingga legislatif ikn segera diumumkan jokowi utamakan pemerataan ,"	'desain', 'baik', 'istana', 'wakil', 'presiden', 'legislatif', 'ikn', 'umum', 'jokowi', 'utama', 'perata']"
7, perusahaan telekomunikasi pt telkom indonesia telkom berencana membangun	'usaha', 'telekomunikasi', 'pt', 'telkom', 'indonesia', 'telkom', 'rencana', 'bangun', 'pusat',

pusat data data center di beberapa kota di tanah air salah satunya adalah ibu kota negara ikn baru ,"	'data', 'data', 'center', 'kota', 'tanah', 'air', 'salah', 'satu', 'kota', 'negara', 'ikn']"
8,jokowi utamakan pemerataan serta pemindahan ikn demi indonesia yang lebih maju ,"	'jokowi', 'utama', 'perata', 'pindah', 'ikn', 'indonesia', 'maju']"
9, program utang program mobil esemka program ikn biar gk mangkrak program buzerrp,"	'program', 'utang', 'program', 'mobil', 'esemka', 'program', 'ikn', 'biar', 'gk', 'mangkrak', 'program', 'buzerrp']"
10,beban jakarta yang berat menjadi urgensi pemindahan ikn jokowi utamakan pemerataan ,"	'beban', 'jakarta', 'berat', 'urgensi', 'pindah', 'ikn', 'jokowi', 'utama', 'perata']"
11,dukung ikn nusantara jokowi utamakan pemerataan ,"	'dukung', 'ikn', 'nusantara', 'jokowi', 'utama', 'perata']"
12,dukung ikn nusantara jokowi utamakan pemerataan ,"	'dukung', 'ikn', 'nusantara', 'jokowi', 'utama', 'perata']"

#### 4.1.3. Determine Sentiment

In determine sentiment, the Indonesian lexicon is used which contains weights of sentiment on a scale of +5 to -5, separated into positive and negative lexicon. If the polarity is added, it will be considered neutral. Tokens that have been obtained previously are processed into the lexicon. Here is the following results :

**Table 4.3. Sentiment Result**

<b>Label</b>	<b>Count</b>
Positive	1674
Negative	1005
Neutral	321

Total	3000
-------	------

**Table 4.4. Example Data of Sentiment Result**

Hasil	Coba token	Polarity score	polarity
jokowi utamakan pemerataan dukung pemindahan ikn	"['jokowi', 'utama', 'perata', 'dukung', 'pindah', 'ikn']",	4	positive
asn hingga tni pindah ke ikn nggak perlu beli rumah sudah disediakan jokowi utamakan pemerataan ,	"['asn', 'tni', 'pindah', 'ikn', 'nggak', 'beli', 'rumah', 'sedia', 'jokowi', 'utama', 'perata']",	-1	negative
perencanaan pembangunan ikn jokowi utamakan pemerataan ,	"['rencana', 'bangun', 'ikn', 'jokowi', 'utama', 'perata']",	0	neutral
pembangunan ikn menggerakkan banyak sektor ekonomi jokowi utamakan pemerataan ,	"['bangun', 'ikn', 'gerak', 'sektor', 'ekonomi', 'jokowi', 'utama', 'perata']",	2	positive
ikn nusantara diyakini bakal gerakkan banyak sektor ekonomi jokowi utamakan pemerataan ,	"['ikn', 'nusantara', 'yakin', 'gerak', 'sektor', 'ekonomi', 'jokowi', 'utama', 'perata']",	5	positive
kawasan ikn nusantara bakal ditanami tumbuhan endemik dari provinsi di indonesia jokowi utamakan pemerataan ,	"['kawasan', 'ikn', 'nusantara', 'tanam', 'tumbuh', 'endemik', 'provinsi', 'indonesia', 'jokowi', 'utama', 'perata']",	2	positive
desain terbaik istana wakil presiden hingga legislatif ikn	"['desain', 'baik', 'istana', 'wakil', 'presiden', 'legislatif', 'ikn', 'umum',	12	positive

segera diumumkan jokowi utamakan pemerataan ,	'jokowi', 'utama', 'perata"]",		
perusahaan telekomunikasi pt telkom indonesia telkom berencana membangun pusat data data center di beberapa kota di tanah air salah satunya adalah ibu kota negara ikn baru ,	"['usaha', 'telekomunikasi', 'pt', 'telkom', 'indonesia', 'telkom', 'rencana', 'bangun', 'pusat', 'data', 'data', 'center', 'kota', 'tanah', 'air', 'salah', 'satu', 'kota', 'negara', 'ikn']",	-13	negative
jokowi utamakan pemerataan serta pemindahan ikn demi indonesia yang lebih maju ,	"['jokowi', 'utama', 'perata', 'pindah', 'ikn', 'indonesia', 'maju']",	0	neutral
program utang program mobil esemka program ikn biar gk mangkrak program buzerrp,	"['program', 'utang', 'program', 'mobil', 'esemka', 'program', 'ikn', 'biar', 'gk', 'mangkrak', 'program', 'buzerrp']",	-3	negative
bebani jakarta yang berat menjadi urgensi pemindahan ikn jokowi utamakan pemerataan ,	"['bebani', 'jakarta', 'berat', 'urgensi', 'pindah', 'ikn', 'jokowi', 'utama', 'perata']",	-5	negative

#### 4.1.4. TF-IDF

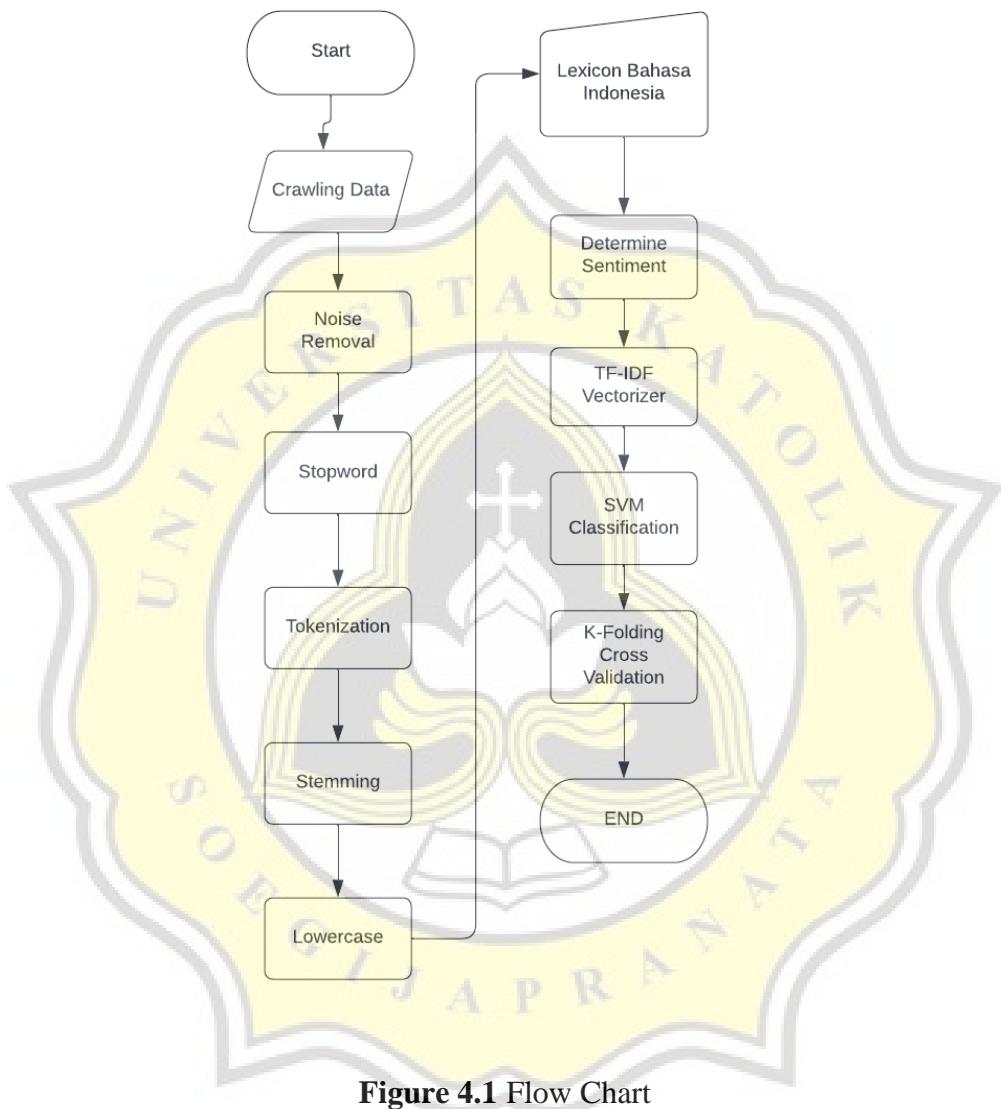
At this stage the word represented in the form vector and TF-IDF. Use TF-IDF method in progress weighting can produce a vector with many terms so that every word can be recognized which counts as a single feature.

#### 4.1.5. SVM Classification

In this section, to tested lexicon based method accuration in determining the sentiment of an opinion tweet, author use SVM algorithm for classification. In the classification process, it is evaluated and validated using the K-fold cross validation method. So the dataset will be divided into 5 parts with 4/5 parts used for the training

process and 1/5 of the part is used for the testing process. The iteration takes 5 times with variations of training and testing data using a combination of 5 parts data.

## 4.2. Design



**Figure 4.1** Flow Chart