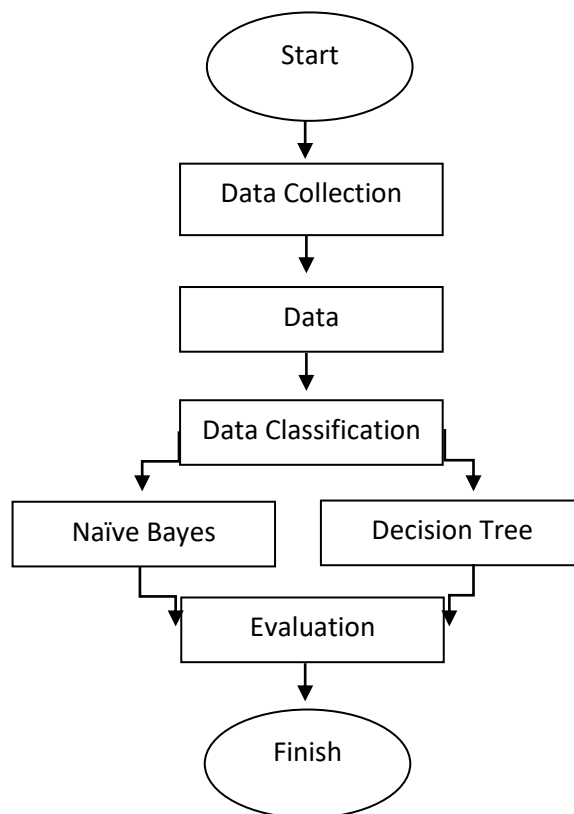


BAB III METODOLOGI PENELITIAN

3.1 Tahapan Metode Penelitian

Tahapan metode penelitian diperlukan sebagai kerangka dan panduan proses penelitian, sehingga rangkaian proses penelitian dapat dilakukan secara terarah, teratur dan sistematis. Berikut metode yang diusulkan dalam penelitian ini ditunjukkan pada gambar 3.1



Gambar 3.1 Tahapan Metode Penelitian

3.2 Metode Pengumpulan Data

Tahap pertama adalah proses pengumpulan data pada *Twitter* atau disebut *crawling* data. *Crawling* data pada *Twitter* yaitu suatu proses mengunduh data berupa user atau *tweet* dari server *Twitter* dengan bantuan *Application Programming Integration* (API) (Eka Sembodo et al., 2016). Pada proses ini kata kunci yang digunakan yaitu “kenaikan tiket Borobudur” dengan total data sebanyak 591 *tweet*.

3.3 Preprocessing Data

Preprocessing adalah salah satu tahap yang paling penting di dalam *text mining* (Alasadi & Bhaya, 2017) yang digunakan untuk melakukan perubahan data mentah menjadi data yang siap digunakan dan lebih terstruktur dengan pembersihan dan penyeragaman data karena data yang dikumpulkan biasanya masih berupa data kotor. Adapun tahap *preprocessing* yang diterapkan dalam penelitian ini adalah *normalisasi*, *case folding*, *cleansing*, *tokenization*, *stopword removal*, *transform cases*, dan *labelling*.

1. Normalisasi

Normalisasi adalah proses yang bertujuan untuk memperbaiki kata yang memiliki kesalahan penulisan ataupun pengejaan serta kata yang dituliskan dengan disingkat.

2. Case Folding

Case folding adalah teknik guna mengganti huruf kapital yang ditemukan pada data sebagai huruf kecil seluruhnya.

3. Cleansing

Cleansing adalah proses yang bertujuan untuk menghilangkan berbagai informasi yang tidak dibutuhkan dalam proses analisis sentimen baik berupa link (*http*, *https*, *pic.twitter*), *hashtag*, *username* (dituliskan *@username*), maupun karakter spesial lainnya untuk memperoleh hasil analisis yang lebih baik.

4. Tokenization

Tokenization adalah proses pemecahan setiap kalimat yang terdapat pada data menjadi potongan-potongan kata. Caranya adalah dengan menjadikan spasi sebagai acuan untuk memisah setiap katanya.

5. Stopword Removal

Stopword Removal ialah kata yang tidak bermakna penting, seperti kata “di”, “dan”, “dengan”, “oleh”, dan sebagainya. Sehingga, *stopword removal* ialah proses guna meniadakan kata yang tidak memiliki makna berharga terhadap data yang akan digunakan.

6. Transform Cases

Transform Cases adalah proses untuk merubah seluruh huruf menjadi huruf kecil atau kapital semua.

7. Labelling

Labelling merupakan pemrosesan terhadap hasil data *stopword removal*, dimana hasilnya diberikan perhitungan *polarity* dari komentar yang diambil, sehingga mendapatkan klasifikasi yaitu label positif, negatif, dan label netral.

3.4 Klasifikasi Data

Setelah seluruh tahapan sebelumnya selesai dilakukan, selanjutnya akan dilakukan proses klasifikasi sentimen terhadap data yang sudah disiapkan. Klasifikasi yang dilakukan pada penelitian dengan dua macam proses klasifikasi, yang pertama adalah klasifikasi sentimen dengan metode *Naive Bayes* dan yang kedua adalah klasifikasi sentiment dengan metode *Decision Tree*.

3.4.1 Metode Naïve Bayes

Pada tahap ini dilakukan perhitungan menggunakan sampel data yang sudah dibagi menjadi data latih yang dapat dilihat pada tabel 3.1 berikut ini.

Tabel 3.1 Data Latih Hasil Preprocessing

No	Candi	Tiket	Borobudur	Harga	Masuk	Kenaikan	Kelas
1	1	1	1	0	0	1	Netral
2	1	1	1	1	0	2	Netral
3	1	1	1	1	0	0	Negatif
4	1	1	0	1	0	1	Positif
5	1	2	1	2	2	0	Negatif
6	1	1	1	1	0	0	Negatif
7	1	1	1	0	1	0	Netral
8	1	1	1	0	1	0	Netral
9	0	0	0	0	0	0	Netral
10	2	1	1	0	0	0	Positif

Setelah ditentukan data latih, kemudian dilakukan perhitungan menggunakan persamaan (1).

Menentukan probabilitas kelas:

$$P(\text{positif}) : \frac{2}{10} = 0,2$$

$$P(\text{negatif}) : \frac{3}{10} = 0,3$$

$$P(\text{netral}) : \frac{5}{10} = 0,5$$

Kemudian dilanjutkan menghitung peluang terhadap kata-kata yang bersumber dari data latih menggunakan persamaan (2)

Menentukan probabilitas kategori:

Perhitungan kelas positif:

$$P(\text{Candi} = \text{Candi}|\text{Positif}) = \frac{2}{2} = 1$$

$$P(\text{Tiket} = \text{Tiket}|\text{Positif}) = \frac{2}{2} = 1$$

$$P(\text{Borobudur} = \text{Borobudur}|\text{Positif}) = \frac{1}{2} = 0,5$$

$$P(\text{Harga} = \text{Harga}|\text{Positif}) = \frac{1}{2} = 0,5$$

$$P(\text{Masuk} = \text{Masuk}|\text{Positif}) = \frac{0}{2} = 0$$

$$P(\text{Kenaikan} = \text{Kenaikan}|\text{Positif}) = \frac{1}{2} = 0,5$$

Perhitungan kelas Negatif:

$$P(\text{Candi} = \text{Candi}|\text{Negatif}) = \frac{3}{3} = 1$$

$$P(\text{Tiket} = \text{Tiket}|\text{Negatif}) = \frac{3}{3} = 1$$

$$P(\text{Borobudur} = \text{Borobudur}|\text{Negatif}) = \frac{3}{3} = 1$$

$$P(\text{Harga} = \text{Harga}|\text{Negatif}) = \frac{3}{3} = 1$$

$$P(\text{Masuk} = \text{Masuk}|\text{Negatif}) = \frac{1}{3} = 0,33$$

$$P(\text{Kenaikan} = \text{Kenaikan}|\text{Negatif}) = \frac{0}{3} = 0$$

Perhitungan kelas Netral:

$$P(\text{Candi} = \text{Candi}|\text{Netral}) = \frac{4}{5} = 0,8$$

$$P(\text{Tiket} = \text{Tiket}|\text{Netral}) = \frac{4}{5} = 0,8$$

$$P(\text{Borobudur} = \text{Borobudur}|\text{Netral}) = \frac{4}{5} = 0,8$$

$$P(\text{Harga} = \text{Harga}|\text{Netral}) = \frac{1}{5} = 0,2$$

$$P(\text{Masuk} = \text{Masuk}|\text{Netral}) = \frac{2}{5} = 0,4$$

$$P(\text{Kenaikan} = \text{Kenaikan}|\text{Netral}) = \frac{2}{5} = 0,4$$

Menguji hasil klasifikasi Naïve Bayes:

$$P(\text{Main}|\text{Positif}) = P(\text{Main}|\text{Positif}) * P(\text{Candi}=\text{Positif}) * P(\text{Tiket}=\text{Positif})$$

$$* P(\text{Borobudur}=\text{Positif}) * P(\text{Harga}=\text{Positif}) * P(\text{Masuk}=\text{Positif}) *$$

$$P(\text{Kenaikan}=\text{Positif})$$

$$= 0,2 * (1 * 1 * 0,5 * 0,5 * 0,5)$$

$$= 0$$

$$\begin{aligned}
P(\text{Main}|\text{Negatif}) &= P(\text{Main}|\text{Negatif}) * (P(\text{Candi}=\text{Negatif}) * P(\text{Tiket}=\text{Negatif}) \\
&\quad * P(\text{Borobudur}=\text{Negatif}) * P(\text{Harga}=\text{Negatif}) * P(\text{Masuk}=\text{Negatif}) \\
&\quad * P(\text{Kenaikan}=\text{Negatif}) \\
&= 0,3 * (1 * 1 * 1 * 1 * 0,33 * 0) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
P(\text{Main}|\text{Netral}) &= P(\text{Main}|\text{Netral}) * (P(\text{Candi}=\text{Netral}) * P(\text{Tiket}=\text{Netral}) \\
&\quad * P(\text{Borobudur}=\text{Netral}) * P(\text{Harga}=\text{Netral}) * P(\text{Masuk}=\text{Netral}) * \\
&\quad P(\text{Kenaikan}=\text{Netral}) \\
&= 0,5 * (0,8 * 0,8 * 0,8 * 0,2 * 0,4 * 0,4) \\
&= 0,008192
\end{aligned}$$

Dari hasil perhitungan peluang dokumen terhadap kelas diatas, diketahui bahwa peluang dokumen terhadap kelas positif sama dengan kelas negatif dengan nilai 0 dan kelas netral yaitu dengan nilai 0,008192. Maka dapat diambil kesimpulan bahwa sampel data latih diklasifikasikan sebagai sentimen Netral.

3.4.2 Metode Decision Tree

Pada tahap ini dilakukan perhitungan menggunakan sampel data yang dapat dilihat pada tabel 3.2 berikut ini.

Tabel 3.2 Nilai Data Latih

	Nilai	Jumlah(S)	Positif(Si)	Negatif(Si)	Netral(Si)
Total		10	2	3	5
Candi	1	9	2	3	4
	0	1	0	0	1
Tiket	1	9	2	3	4
	0	1	0	0	1
Borobudur	1	8	1	3	4
	0	2	1	0	1

	Nilai	Jumlah(S)	Positif(Si)	Negatif(Si)	Netral(Si)
Harga	1	5	1	3	1
	0	5	1	0	4
Masuk	1	3	0	1	2
	0	7	2	2	3
Kenaikan	1	3	1	0	2
	0	7	1	3	3

$$Entropy(Total) = -\frac{2}{10} \log_2 \frac{2}{10} + -\frac{3}{10} \log_2 \frac{3}{10} + -\frac{5}{10} \log_2 \frac{5}{10} = 1,485475297$$

Setelah ditentukan data latih, kemudian dilakukan perhitungan menggunakan persamaan (3). Perhitungan information gain pada setiap atribut data latih untuk menentukan simpul akar:

1) Candi

Perhitungan entropy atribut “Candi” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{0}{10} \log_2 \frac{0}{10} + -\frac{0}{10} \log_2 \frac{0}{10} + -\frac{1}{10} \log_2 \frac{1}{10} = 0$$

$$Entropy(1) = -\frac{2}{10} \log_2 \frac{2}{10} + -\frac{3}{10} \log_2 \frac{3}{10} + -\frac{4}{10} \log_2 \frac{4}{10} = 0,514246535$$

$$Gain (Total, Candi) = 1,485475297 - \frac{9}{10} + \frac{1}{10} = 0,122653$$

2) Tiket

Perhitungan entropy atribut “Tiket” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{0}{10} \log_2 \frac{0}{10} + -\frac{0}{10} \log_2 \frac{0}{10} + -\frac{1}{10} \log_2 \frac{1}{10} = 0$$

$$Entropy(1) = -\frac{2}{10} \log_2 \frac{2}{10} + -\frac{3}{10} \log_2 \frac{3}{10} + -\frac{4}{10} \log_2 \frac{4}{10} = 0,514246535$$

$$Gain (Total, Tiket) = 1,485475297 - \frac{9}{10} + \frac{1}{10} = 0,122653$$

3) Borobudur

Perhitungan entropy atribut “Borobudur” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{1}{10} \log_2 \frac{1}{10} + -\frac{0}{10} \log_2 \frac{0}{10} + -\frac{1}{10} \log_2 \frac{1}{10} = 0$$

$$Entropy(1) = -\frac{1}{10} \text{Log}_2 \frac{1}{10} + -\frac{3}{10} \text{Log}_2 \frac{3}{10} + -\frac{4}{10} \text{Log}_2 \frac{4}{10} = 1,382053726$$

$$Gain (Total, Borobudur) = 1,485475297 - \frac{8}{10} + \frac{2}{10} = 0,379832$$

4) Harga

Perhitungan entropy atribut “Harga” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{1}{10} \text{Log}_2 \frac{1}{10} + -\frac{0}{10} \text{Log}_2 \frac{0}{10} + -\frac{4}{10} \text{Log}_2 \frac{4}{10} = 0$$

$$Entropy(1) = -\frac{1}{10} \text{Log}_2 \frac{1}{10} + -\frac{3}{10} \text{Log}_2 \frac{3}{10} + -\frac{1}{10} \text{Log}_2 \frac{1}{10} = 1,185475297$$

$$Gain (Total, Harga) = 1,485475297 - \frac{5}{10} + \frac{5}{10} = 0,892738$$

5) Masuk

Perhitungan entropy atribut “Masuk” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{2}{10} \text{Log}_2 \frac{2}{10} + -\frac{2}{10} \text{Log}_2 \frac{2}{10} + -\frac{3}{10} \text{Log}_2 \frac{3}{10} = 1,449860916$$

$$Entropy(1) = -\frac{0}{10} \text{Log}_2 \frac{0}{10} + -\frac{1}{10} \text{Log}_2 \frac{1}{10} + -\frac{2}{10} \text{Log}_2 \frac{2}{10} = 0$$

$$Gain (Total, Masuk) = 1,485475297 - \frac{3}{10} + \frac{7}{10} = 0,470573$$

6) Kenaikan

Perhitungan entropy atribut “Kenaikan” dengan nilai 0 dan 1:

$$Entropy(0) = -\frac{1}{10} \text{Log}_2 \frac{1}{10} + -\frac{3}{10} \text{Log}_2 \frac{3}{10} + -\frac{3}{10} \text{Log}_2 \frac{3}{10} = 1,374372166$$

$$Entropy(1) = -\frac{0}{10} \text{Log}_2 \frac{0}{10} + -\frac{1}{10} \text{Log}_2 \frac{1}{10} + -\frac{2}{10} \text{Log}_2 \frac{2}{10} = 0$$

$$Gain (Total, Kenaikan) = 1,485475297 - \frac{3}{10} + \frac{7}{10} = 0,523415$$

Setelah dilakukan perhitungan menggunakan persamaan (3). Maka hasil perhitungan dapat dilihat pada tabel 3.3 berikut ini.

Tabel 3.3 Perhitungan Information Gain untuk Simpul Akar

Perhitungan Node Akar									
Total Entropy: 1,485475297									
Atribut	Pos (0)	Neg (0)	Net (0)	Entropy (0)	Pos (1)	Neg (1)	Net (1)	Entropy (1)	Gain
Candi	0	0	1	0	2	3	4	1,514246535	0,122653
Tiket	0	0	1	0	2	3	4	1,514246535	0,122653
Borobudur	1	0	1	0	1	3	4	1,382053726	0,379832
Harga	1	0	4	0	1	3	1	1,185475297	0,892738
Masuk	2	2	3	1,449860916	0	1	2	0	0,470573
Kenaikan	1	3	3	1,374372166	1	0	2	0	0,523415
Max Gain: 0,892738				Node: Harga					

3.5 Evaluasi

Evaluasi hasil dari pengujian data menggunakan tabel *Confussion Matrix* yaitu accuracy. *Confussion Matrix* merepresentasikan prediksi dan kondisi sebenarnya (aktual) dari data yang dihasilkan oleh algoritma yang digunakan. Akurasi digunakan untuk evaluasi karena untuk mengetahui rasio prediksi benar dengan data keseluruhan. Evaluasi hasil pengujian data dapat menggunakan tabel confusion matrix seperti pada tabel 3.4 berikut ini.

Tabel 3.4 Confusion Matrix

Kelas	Aktual Positif	Aktual Netral	Aktual Negatif
Prediksi Positif	True Positive (TP)	False Positive1 (FP1)	False Positive2 (FP2)
Prediksi Netral	False Neutral1 (FNt1)	True Neutral (TNt)	False Neutral2 (FNt2)
Prediksi Negatif	False Negative1 (FNg1)	False Negative2 (FNg2)	True Negative (TNg)

Keterangan:

- True Negative* / Aktual Negatif = data aktual yang berada di kelas negative dan model telah memprediksi negatif

- b. *True Positive* / Aktual Positif = data aktual yang berada di kelas positif dan model telah memprediksi positif
- c. *True Neutral* / Aktual Netral = data aktual yang berada di kelas netral dan model telah memprediksi netral
- d. *False Negative* / Prediksi Negatif = data aktual yang berada di kelas positif atau netral, namun model telah memprediksi negatif
- e. *False Positive* / Prediksi Positif = data aktual yang berada di kelas negatif atau netral, namun model telah memprediksi positif
- f. *False Neutral* / Prediksi Netral = data aktual yang berada di kelas negative atau positif, namun model telah memprediksi netral

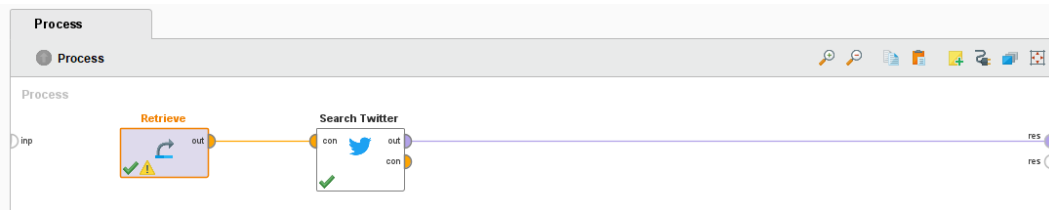
Dari keenam kemungkinan *confusion matrix* pada table 3.4 kemudian dapat dilakukan perhitungan nilai akurasi menggunakan persamaan (10) berikut ini.

$$\frac{TP + TNg + TNt}{TP + FNt2 + \dots + FNt1 + TNg} \quad (10)$$

BAB IV HASIL DAN PEMBAHASAN

4.1 Fase Pengumpulan Data

Tahap pertama adalah proses pengumpulan data pada *Twitter* atau disebut *crawling data*. *Crawling data* pada *Twitter* yaitu suatu proses mengunduh data berupa user atau *tweet* dari server *Twitter* dengan bantuan *Application Programming Integration (API)* (Eka Sembodo et al., 2016). Pada proses ini kata kunci yang digunakan yaitu “kenaikan tiket Borobudur” dengan total data sebanyak 591 *tweet*. Proses *crawling data* pada tahapan pengumpulan data dapat dilihat pada gambar 4.1 dan 4.2 berikut ini:



Gambar 4.1 Proses Crawling Data Twitter

Row No.	Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text
1	?	pemerintah te...	?	?	?	?	?	?
2	Jun 14, 2022 ...	detikcom	69183155	?	-1	in	<a href="https...	Menteri PU
3	?	?	?	?	?	?	?	?
4	Jun 20, 2022 ...	mnctrijaya su...	1438732688...	?	-1	in	<a href="http://...	Gubernur .
5	Jun 20, 2022 ...	Binsar Ritonga	1321801888...	?	-1	in	<a href="http://...	Lah drpd n
6	Jun 20, 2022 ...	Eka	3771801192	?	-1	in	<a href="http://...	RT @gald:
7	Jun 19, 2022 ...	@??	1417332125...	?	-1	in	<a href="http://...	RT @Jero
8	Jun 19, 2022 ...	Sugeng Novi...	488563266	amakmal_37	2204919582	in	<a href="http://...	@amakma
9	Jun 19, 2022 ...	Nur Hasanud...	1410146144...	?	-1	in	<a href="http://...	RT @kolor
10	Jun 19, 2022 ...	CSIRT Bulel...	1113798518...	?	-1	in	<a href="https...	Tiket Masu
11	Jun 19, 2022 ...	Achmad N	1487349455...	?	-1	in	<a href="http://...	RT @kolor
12	Jun 19, 2022 ...	Mochall	1488375140...	?	-1	in	<a href="http://...	RT @kolor

Gambar 4.2 Hasil Crawling Data Twitter

4.2 Fase Preprocessing Data

Preprocessing adalah salah satu tahap yang paling penting di dalam *text mining* (Alasadi & Bhaya, 2017) yang digunakan untuk melakukan perubahan data mentah menjadi data yang siap digunakan dan lebih terstruktur dengan pembersihan dan penyeragaman data karena data yang dikumpulkan biasanya masih berupa data kotor. Adapun tahap preprocessing yang diterapkan dalam penelitian ini adalah *cleansing*, *tokenization*, *transform cases*, *stopword removal*, *filter token (by length)*, dan *labelling*.

4.2.1 Cleansing

Cleansing adalah proses yang bertujuan untuk menghilangkan berbagai informasi yang tidak dibutuhkan dalam proses analisis sentimen baik berupa link (*http*, *https*, *pic.twitter*), *hashtag*, *username* (dituliskan *@username*), angka, maupun karakter spesial lainnya untuk memperoleh hasil analisis yang lebih baik seperti pada tabel 4.1 berikut ini.

Tabel 4.1 Perbandingan Teks Sebelum dan Sesudah dilakukan Proses *Cleansing*

Proses	Teks
Teks sebelum dilakukan cleansing	RT @Bambang_DP: Setelah melihat video ini, sepertinya tiket masuk candi borobudur menjadi masuk akal. https://t.co/atjgGnUdCT
Teks sesudah dilakukan cleansing	Setelah melihat video ini sepertinya tiket masuk candi borobudur menjadi masuk akal

Sumber: Hasil Penelitian (2023)

4.2.2 Tokenization

Tokenization adalah proses pemecahan setiap kalimat yang terdapat pada data menjadi potongan-potongan kata. Caranya adalah dengan menjadikan