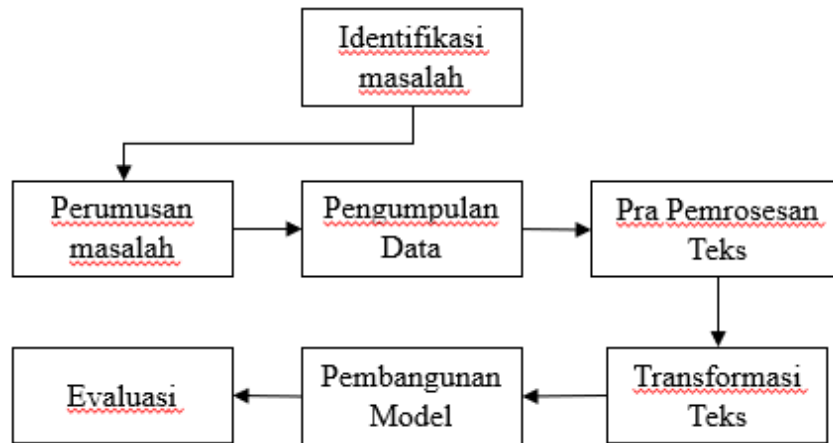


BAB 3

METODE PENELITIAN

3.1 Kerangka berpikir

Pada penelitian ini, dijalankan proses penelitian berdasarkan kerangka berpikir pada Gambar 3.1 berikut



Gambar 3.1 Kerangka berpikir

3.1.1 Identifikasi masalah

Teknologi yang semakin maju berbuah pada lahirnya berbagai sarana bagi masyarakat luas untuk melakukan komunikasi satu sama lain, salah satunya adalah munculnya aplikasi media sosial. Pengguna media sosial memiliki pengalaman yang beragam dalam penggunaan aplikasi tersebut. Pengalaman pengguna ini dapat menjadi acuan bagi perusahaan pengembang aplikasi dalam menentukan pilihan untuk mengembangkan aplikasi. Untuk mengetahui pilihan yang tepat berdasarkan umpan balik dan pengalaman yang dialami oleh pengguna aplikasi, dibutuhkan teknologi *text mining* melalui pendekatan AI yang disebut *supervised learning* (pembelajaran yang diawasi).

Dengan teknologi ini, perusahaan pengembang aplikasi dapat menggunakan model klasifikasi, ekstraksi fitur, dan pemrosesan bahasa asli untuk menganalisis sentimen dari pengguna aplikasi. Ketika sentimen telah di analisis, perusahaan dapat mengetahui secara ilmiah bagaimana sentimen pengguna aplikasi dalam menggunakan aplikasi. Setelah sentimen pengguna diketahui, perusahaan dapat menentukan pilihan terbaik untuk menentukan pengembangan aplikasi.

3.1.2 Perumusan masalah

Teknologi *text mining* yang menggunakan pendekatan *supervised learning* membutuhkan data tekstual yang akan di proses. Selain data tekstual, diperlukan juga data label yang berguna sebagai “pengawas” dalam pembangunan model klasifikasi untuk *text mining*. Dari banyaknya model klasifikasi, penelitian ini berfokus pada dua algoritma pembentuk model klasifikasi, yakni algoritma *Naïve Bayes* dan *Decision Tree*. Maka dalam penelitian ini, akan dilakukan pembangunan model klasifikasi menggunakan algoritma *Naïve Bayes* dan *Decision Tree*.

3.1.3 Pengumpulan data

Data yang dikumpulkan berupa data ulasan pengguna aplikasi media sosial Twitter yang didapatkan dari kolom ulasan Google Play Store.

3.1.4 Pra pemrosesan teks

Sebelum data di proses lebih lanjut, dilakukan persiapan terhadap data, dimana data akan dibersihkan, dirapikan, serta dipersiapkan untuk dilanjutkan ke tahap selanjutnya. Pembersihan, perapian, serta persiapan ini, meliputi penghapusan karakter, tokenisasi kata, pengembalian data ke bentuk dasar, hingga penghilangan kata *stopword*.

3.1.5 Transformasi teks

Setelah dibersihkan, data melalui tahapan transformasi teks, dimana data yang sebelumnya berbentuk teks akan diubah menjadi angka. Pengubahan ini berdasarkan penghitungan bobot menggunakan pengeksrak fitur TF-IDF.

3.1.6 Pembangunan model

Setelah data menjadi berbentuk angka, data dapat digunakan untuk pembangunan model klasifikasi. Dalam pembangunan model ini, digunakan dua algoritma *data mining* yaitu algoritma *Naïve Bayes* dan *Decision Tree*.

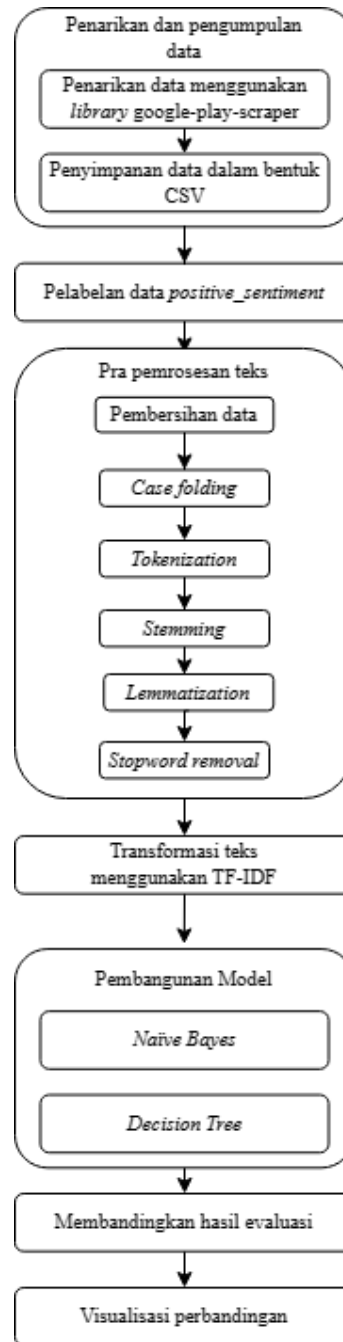
3.1.7 Evaluasi

Setelah model klasifikasi dibangun, kedua model akan dibandingkan hasil evaluasi performanya, baik dari segi akurasi, presisi, *recall*, *f1-score*, dan luas area dibawah kurva ROC (*Receiving Operating Characteristic*).

3.2 Metode penelitian

Pada penelitian ini, akan dilakukan proses *text mining* melalui enam tahapan yakni penarikan dan pengumpulan data, pelabelan data, pra pemrosesan teks, transformasi teks,

pembangunan model, dan evaluasi. Tahap-tahap proses *text mining* ini akan dijelaskan dengan *flowchart* yang lebih rinci pada Gambar 3.2 berikut.



Gambar 3.2. *Flowchart* Penelitian

3.2.1 Penarikan dan pengumpulan data

Dalam proses ini, penarikan dan pengumpulan data ulasan aplikasi Twitter yang didapatkan dari toko aplikasi Google Play Store. Untuk mendapatkan dan mengumpulkan data tersebut, proses ini menggunakan bahasa pemrograman *python*

dan memasang *library* `google_play_scraper`. *Library* ini dapat melakukan *scraping* pada laman toko aplikasi Google Play Store. Proses *scraping* ini dapat mengambil seluruh data yang terdapat di laman Google Play Store, namun pada penelitian ini hanya dibutuhkan data ulasan dari aplikasi Twitter.

Untuk memfokuskan proses *scraping* menggunakan *function* `reviews_all` untuk melakukan pengambilan data ulasan aplikasi Twitter dengan beberapa batasan. Batasannya yaitu data yang diambil merupakan data ulasan yang ditulis di negara Indonesia, menggunakan bahasa Inggris, dan disortir berdasarkan waktu penulisan ulasan terbaru. Penyortiran ini berguna untuk proses penelitian untuk mengetahui data ulasan secara historis. Setelah mendapatkan data, data tersebut disimpan kedalam sebuah *dataframe* Pandas. Menggunakan Pandas, dilakukan *query* terhadap data untuk mendapatkan ulasan yang hanya ditulis dari tahun 2020 hingga 2023.

Data yang disimpan yaitu `'reviewId'`, `'content'`, `'score'`, `'year'`. Kolom `'reviewId'` kolom `'year'` hanya akan digunakan sebagai identitas data, sedangkan kolom `'content'` yang berisikan ulasan akan menjadi kolom yang digunakan untuk *text mining* dengan kolom `'score'` yang digunakan sebagai pelabelan data. Setelah itu kemudian disimpan data yang telah diambil menggunakan *function* `to_csv` yang mengubah data yang semula berbentuk *dataframe* menjadi CSV. Tujuan penyimpanan ini untuk mencadangkan data supaya data tidak hilang ketika komputer mengalami kendala selama penelitian.

3.2.2 Pelabelan data

Pelabelan data dilakukan berdasarkan skor *rating* ulasan yang tersedia di kolom *score* sebagai referensi sentimen. *Rating* ulasan yang bernilai 1 dan 2 dikelompokkan kedalam kelompok ulasan dengan sentimen negatif. Sedangkan *rating* ulasan yang bernilai 4 dan 5 dikelompokkan kedalam kelompok ulasan dengan sentimen positif (Fransiska, S., et al., 2020). Ulasan yang memiliki nilai *rating* 3 dikelompokkan sebagai ulasan dengan sentimen netral, pada penelitian ini sentimen netral tidak dibutuhkan, sehingga ulasan dengan sentimen netral tidak diproses.

Dalam pelabelan ini, ulasan yang memiliki sentimen positif akan diberi label *positive_sentiment* (1). Sedangkan, ulasan yang memiliki sentimen negatif akan diberi label *positive_sentiment* (0). Setelah data selesai diberi label data dalam penelitian ini selanjutnya akan dinyatakan sebagai data ulasan.

3.2.3 Pra pemrosesan teks

Data ulasan selanjutnya akan melalui tahap pra pemrosesan. Tujuan dari tahapan ini yaitu untuk memastikan data siap untuk di transformasi menjadi data numerik yang dapat digunakan untuk membangun model klasifikasi. Dalam tahap ini, data ulasan akan melalui pembersihan data, *case folding*, *tokenization*, *stemming*, *lemmatization*, dan *stopword removal*.

3.2.3.1 Pembersihan data (*Data Cleansing*)

Tahap ini merupakan tahap pertama dari pra pemrosesan teks. Pada tahap ini, data ulasan akan dibersihkan dari karakter-karakter yang tidak diperlukan dalam analisis sentimen. Karakter-karakter ini merupakan karakter yang berupa bukan huruf seperti tanda tanya, tanda seru, tanda pagar, *tag HTML*, *emoticon*, dan juga *link URL* atau laman *website*.

Tabel 3.1 Contoh tahap pembersihan data

Sebelum pembersihan	Setelah pembersihan
<i>Very bad app.I CAN'T LOG IN OR CREATE MY ACCOUNT</i>	<i>Very bad app I CANT LOG IN OR CREATE MY ACCOUNT</i>
<i>Very poor Twitter control team .they can ban our leaders Twitter account. Shame on you . I'm sadly to say .very bad team 😞😞😞</i>	<i>Very poor Twitter control team they can ban our leaders Twitter account Shame on you Im sadly to say very bad team</i>
<i>I really enjoy using the app.</i>	<i>I really enjoy using the app</i>
<i>I think is good and useful app... Otherwise there is fake account and they're publishing fake news and stuff</i>	<i>I think is good and useful app Otherwise there is fake account and theyre publishing fake news and stuff</i>
<i>Noticed the bad reviews for twitter but personally I've had no issues</i>	<i>Noticed the bad reviews for twitter but personally Ive had no issues</i>
<i>It also help getting across to our leaders by expressing our concerns without moving out of the comfort of our rooms... Thanks a lot to Twitter....</i>	<i>It also help getting across to our leaders by expressing our concerns without moving out of the comfort of our rooms Thanks a lot to Twitter</i>
<i>The app is really good. Works great all of the time.</i>	<i>The app is really good Works great all of the time</i>

<i>It's nice app ☐ I am very happy ☺</i>	<i>Its nice app I am very happy</i>
<i>I can't even log in to my account every time I try to log in it says suspicious activity. Seriously new updates fix the things but it is doing the opposite</i>	<i>I cant even log in to my account every time I try to log in it says suspicious activity Seriously new updates fix the things but it is doing the opposite</i>
<i>The new video player is utterly bad.</i>	<i>The new video player is utterly bad</i>

3.2.3.2. Case folding

Setelah data ulasan tidak memiliki karakter yang bukan berupa huruf, dilakukan proses *case folding* untuk mengubah huruf kapital pada data ulasan menjadi huruf kecil. Tujuan dari proses ini untuk membuat data menjadi lebih konsisten.

Tabel 3.2 Contoh tahap *case folding*

Sebelum Case Folding	Setelah Case Folding
<i>Very bad app I CANT LOG IN OR CREATE MY ACCOUNT</i>	<i>very bad app i cant log in or create my account</i>
<i>Very poor Twitter control team they can ban our leaders Twitter account Shame on you I'm sadly to say very bad team</i>	<i>very poor twitter control team they can ban our leaders twitter account shame on you im sadly to say very bad team</i>
<i>I really enjoy using the app</i>	<i>i really enjoy using the app</i>
<i>I think is good and useful app Otherwise there is fake account and theyre publishing fake news and stuff</i>	<i>i think is good and useful app otherwise there is fake account and theyre publishing fake news and stuff</i>
<i>Noticed the bad reviews for twitter but personally Ive had no issues</i>	<i>noticed the bad reviews for twitter but personally ive had no issues</i>
<i>It also help getting across to our leaders by expressing our concerns without moving out of the comfort of our rooms Thanks a lot to Twitter</i>	<i>it also help getting across to our leaders by expressing our concerns without moving out of the comfort of our rooms thanks a lot to twitter</i>
<i>The app is really good Works great all of the time</i>	<i>the app is really good works great all of the time</i>

<i>Its nice app I am very happy</i>	<i>its nice app i am very happy</i>
<i>I cant even log in to my account every time I try to log in it says suspicious activity Seriously new updates fix the things but it is doing the opposite</i>	<i>i cant even log in to my account every time i try to log in it says suspicious activity seriously new updates fix the things but it is doing the opposite</i>
<i>The new video player is utterly bad</i>	<i>the new video player is utterly bad</i>

3.2.3.2 Tokenization

Setelah data ulasan menjadi konsisten yang berisikan huruf kecil semua, data ulasan yang berupa kalimat akan dipecah-pecah menjadi sekumpulan kata-kata. Proses pemecahan ini disebut sebagai *tokenization* atau tokenisasi. Hasil dari proses ini berbentuk kata-kata yang akan disebut sebagai *token-token*.

Tabel 3.3 Contoh tahap *tokenisasi*

Sebelum Tokenization	Setelah Tokenization
<i>very bad app i cant log in or create my account</i>	<i>very, bad, app, i, cant, log, in, or, create, my, account</i>
<i>very poor twitter control team they can ban our leaders twitter account shame on you im sadly to say very bad team</i>	<i>very, poor, twitter, control, team, they, can, ban, our, leaders, twitter, account, shame, on, you, im, sadly, to, say, very, bad, team</i>
<i>i really enjoy using the app</i>	<i>i, really, enjoy, using, the, app</i>
<i>i think is good and useful app otherwise there is fake account and theyre publishing fake news and stuff</i>	<i>i, think, is, good, and, useful, app, otherwise, there, is, fake, account, and, theyre, publishing, fake, news, and, stuff</i>
<i>noticed the bad reviews for twitter but personally ive had no issues</i>	<i>noticed, the, bad, reviews, for, twitter, but, personally, ive, had, no, issues</i>
<i>it also help getting across to our leaders by expressing our concerns without moving out of the comfort of our rooms thanks a lot to twitter</i>	<i>it, also, help, getting, across, to, our, leaders, by, expressing, our, concerns, without, moving, out, of, the, comfort, of, our, rooms, thanks, a, lot, to, twitter</i>

<i>the app is really good works great all of the time</i>	<i>the, app, is, really, good, works, great, all, of, the, time</i>
<i>its nice app i am very happy</i>	<i>its, nice, app, i, am, very, happy</i>
<i>i cant even log in to my account every time i try to log in it says suspicious activity seriously new updates fix the things but it is doing the opposite</i>	<i>i, cant, even, log, in, to, my, account, every, time, i, try, to, log, in, it, says, suspicious, activity, seriously, new, updates, fix, the, things, but it, is, doing, the, opposite</i>
<i>the new video player is utterly bad</i>	<i>the, new, video, player, is, utterly, bad</i>

3.2.3.3 Stemming

Token-token yang dihasilkan dari tahap tokenisasi, jika token atau kata memiliki imbuhan pada proses ini akan dikembalikan ke bentuk dasar dari kata tersebut. Hal ini bertujuan untuk membantu proses deteksi dan penghapusan *stopword*.

Tabel 3.4 Contoh tahap *stemming*

Sebelum Stemming	Setelah Stemming
<i>very, bad, app, i, cant, log, in, or, create, my, account</i>	<i>ver, bad, app, i, cant, log, in, or, creat, my, account</i>
<i>very, poor, twitter, control, team, they, can, ban, our, leaders, twitter, account, shame, on, you, im, sadly, to, say, very, bad, team</i>	<i>ver, poor, twitter, control, team, they, can, ban, our, leader, twitter, account, shame, on, you, im, sad, to, say, ver, bad, team</i>
<i>i, really, enjoy, using, the, app</i>	<i>i, real, enjoy, us, the, app</i>
<i>i, think, is, good, and, useful, app, otherwise, there, is, fake, account, and, theyre, publishing, fake, news, and, stuff</i>	<i>i, think, is, good, and, use, app, otherwis, there, is, fake, account, and, theyr, publish, fake, news, and, stuff</i>
<i>noticed, the, bad, reviews, for, twitter, but, personally, ive, had, no, issues</i>	<i>notic, the, bad, review, for, twitter, but, person, ive, had, no, issu</i>
<i>it, also, help, getting, across, to, our, leaders, by, expressing, our, concerns,</i>	<i>it, also, help, get, across, to, our, leader, by, express, our, concern,</i>

<i>without, moving, out, of, the, comfort, of, our, rooms, thanks, a, lot, to, twitter</i>	<i>without, mov, out, of, the, comfort, of, our, room, thank, a, lot, to, twitter</i>
<i>the, app, is, really, good, works, great, all, of, the, time</i>	<i>the, app, is, real, good, work, great, all, of, the, time</i>
<i>its, nice, app, i, am, very, happy</i>	<i>it, nice, app, i, am, very, happy</i>
<i>i, cant, even, log, in, to, my, account, every, time, i, try, to, log, in, it, says, suspicious, activity, seriously, new, updates, fix, the, things, but it, is, doing, the, opposite</i>	<i>i, cant, even, log, in, to, my, account, every, time, i, try, to, log, in, it, say, suspici, activ, serious, new, updat, fix, the, thing, but it, is, do, the, opposit</i>
<i>the, new, video, player, is, utterly, bad</i>	<i>the, new, video, play, is, utter, bad</i>

3.2.3.4 Lemmatization

Tahap stemming memunculkan masalah baru yaitu ketika token atau kata memiliki perubahan setelah diberi imbuhan, tahap *stemming* hanya akan memotong imbuhan dari kata tersebut, seperti contoh “*using*” menjadi “*us*”. Untuk menyelesaikan masalah ini, token-token harus melalui proses *lemmatization*.

Tabel 3.5 Contoh tahap lemmatization

Sebelum Lemmatization	Setelah Lemmatization
<i>ver, bad, app, i, cant, log, in, or, creat, my, account</i>	<i>very, bad, app, i, cant, log, in, or, creat, my, account</i>
<i>ver, poor, twitter, control, team, they, can, ban, our, leader, twitter, acount, shame, on, you, im, sad, to, say, ver, bad, team</i>	<i>very, poor, twitter, control, team, they, can, ban, our, leader, twitter, account, shame, on, you, im, sad, to, say, very, bad, team</i>
<i>i, real, enjoy, using, the, app</i>	<i>i, real, enjoy, use, the, app</i>
<i>i, think, is, good, and, use, app, otherwis, there, is, fake, account, and, theyr, publish, fake, news, and, stuff</i>	<i>i, think, is, good, and, use, app, otherwise, there, is, fake, account, and, they, publish, fake, news, and, stuff</i>
<i>notic, the, bad, review, for, twitter,</i>	<i>notice, the, bad, review, for, twitter,</i>

<i>but, person, ive, had, no, issu</i>	<i>but, person, ive, had, no, issue</i>
<i>it, also, help, get, across, to, our, leader, by, express, our, concern, without, mov, out, of, the, comfort, of, our, room, thank, a, lot, to, twitter</i>	<i>it, also, help, get, across, to, our, leader, by, express, our, concern, without, move, out, of, the, comfort, of, our, room, thank, a, lot, to, twitter</i>
<i>the, app, is, real, good, work, great, all, of, the, time</i>	<i>the, app, is, real, good, work, great, all, of, the, time</i>
<i>it, nice, app, i, am, very, happy</i>	<i>it, nice, app, i, am, very, happy</i>
<i>i, cant, even, log, in, to, my, account, every, time, i, try, to, log, in, it, say, suspici, activ, serious, new, updat, fix, the, thing, but it, is, do, the, opposit</i>	<i>i, cant, even, log, in, to, my, account, every, time, i, try, to, log, in, it, say, suspicion, active, serious, new, update, fix, the, thing, but it, is, do, the, opposite</i>
<i>the, new, video, play, is, utter, bad</i>	<i>the, new, video, play, is, utter, bad</i>

3.2.3.5 Stopword Removal

Setelah semua token telah dihapus imbuhanannya dan dikembalikan menjadi bentuk kata dasar, dilakukan penghapusan *stopword*. *Stopword* merupakan kata-kata yang lazim digunakan dalam percakapan sehari-hari namun tidak memiliki kepentingan dalam proses *text mining*.

Tabel 3.6 Contoh tahap *stopword removal*

Sebelum Stopword Removal	Setelah Stopword Removal
<i>very, bad, app, I, cant, log, in, or, creat, my, account</i>	<i>bad, app, log, creat, account</i>
<i>very, poor, twitter, control, team, they, can, ban, our, leader, twitter, account, shame, on, you, im, sad, to, say, very, bad, team</i>	<i>poor, twitter, control, team, ban, leader, twitter, account, shame, sad, bad, team</i>
<i>I, real, enjoy, using, the, app</i>	<i>enjoy, use, app</i>
<i>i, think, is, good, and, use, app, otherwise, there, is, fake, account, and, they, publish, fake, news, and,</i>	<i>good, use, app, fake, account, publish, fake, news, stuff</i>

<i>stuff</i>	
<i>notice, the, bad, review, for, twitter, but, person, ive, had, no, issue</i>	<i>notice, bad, review, twitter, person, issue</i>
<i>it, also, help, get, across, to, our, leader, by, express, our, concern, without, move, out, of, the, comfort, of, our, room, thank, a, lot, to, twitter</i>	<i>leader, express, concern, comfort, room, lot, twitter</i>
<i>the, app, is, real, good, work, great, all, of, the, time</i>	<i>app, real, good, work, great, time</i>
<i>it, nice, app, i, am, very, happy</i>	<i>nice, app, happy</i>
<i>i, cant, even, log, in, to, my, account, every, time, i, try, to, log, in, it, say suspici, active, serious, new, update, fix, the, thing, but, it, is, do, the, opposite</i>	<i>log, account, time, log, suspicion, active, update, fix, thing, opposite</i>
<i>the, new, video, play, is, utter, bad</i>	<i>new, video, play, utter, bad</i>

3.2.4 Transformasi teks

Setelah dilakukan pra pemrosesan teks, selanjutnya ke tahap transformasi teks. Pada tahap ini, data ulasan yang sebelumnya berbentuk teks akan diubah menjadi data numerik. Transformasi ini bertujuan supaya data lebih memungkinkan untuk digunakan dalam pembangunan model.

Pada tahap ini, digunakan penghitungan bobot menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Metode ini menghitung frekuensi kata dan pendistribusiannya dalam semua koleksi dokumen (Darren, et al., 2022).

Tabel 3.7 Data sebelum transformasi TF-IDF

Indeks	Data
0	<i>bad, app, log, creat, account</i>
1	<i>poor, twitter, control, team, ban, leader, twitter, account, shame, sad, bad, team</i>
2	<i>enjoy, use, app</i>
3	<i>good, use, app, fake, account, publish, fake, news, stuff</i>

Indeks	Data
4	<i>notice, bad, review, twitter, person, issue</i>
5	<i>leader, express, concern, comfort, room, lot, twitter</i>
6	<i>app, real, good, work, great, time</i>
7	<i>nice, app, happy</i>
8	<i>log, account, time, log, suspicion, active, update, fix, thing, opposite</i>
9	<i>new, video, play, utter, bad</i>

3.2.5. Pembangunan model

Tahap selanjutnya yaitu pembangunan model klasifikasi menggunakan *Naïve Bayes* dan *Decision Tree*.

3.2.5.1. *Naïve Bayes*

Dalam pembangunan model, digunakan dokumen dari data yang telah di transformasi dengan dokumen pada indeks 0 dan indeks 1 menjadi data latih, dan dokumen dengan indeks 2 menjadi data uji. Dalam pembuatan kelas sentimen positif, digunakan dokumen pada indeks 0. Berikut Tabel 3.9 menggambarkan hasil pembuatan kelas sentimen positif.

Tabel 3.9 Data kelas sentimen positif

Tokens	TF-IDF
	Dokumen Indeks 0
account	0.8
active	0
app	1
bad	0.8
ban	0
comfort	0
concern	0
control	0
creat	0.2
enjoy	0
express	0
fake	0
fix	0
good	0
great	0
happy	0
issue	0
leader	0
log	0.4
lot	0
new	0
news	0
nice	0
notice	0
opposite	0
person	0
play	0
poor	0
publish	0
real	0
review	0

room	0
sad	0
shame	0
suspicion	0
stuff	0
team	0
time	0
thing	0
twitter	0
utter	0
update	0
use	0
work	0
video	0
SUM	3.2

Sedangkan, dalam pembuatan kelas sentimen negatif, digunakan dokumen pada indeks 1. Berikut Tabel 3.10 menggambarkan hasil pembuatan kelas sentimen negatif.

Tabel 3.10 Data kelas sentimen negatif

Tokens	TF-IDF
	Dokumen Indeks 1
account	0.3336
active	0
app	0
bad	0.3336
ban	0.0834
comfort	0
concern	0
control	0.0834
creat	0
enjoy	0
express	0
fake	0
fix	0
good	0
great	0
happy	0
issue	0
leader	0.1668
log	0
lot	0
new	0
news	0

nice	0
notice	0
opposite	0
person	0
play	0
poor	0.0834
publish	0
real	0
review	0
room	0
sad	0.0834
shame	0.0834
suspicion	0
stuff	0
team	0.167
time	0
thing	0
twitter	0.2502
utter	0
update	0
use	0
work	0
video	0
SUM	1.6682

Berdasarkan tahap pembagian kelas sentimen positif dan negatif, penelitian ini memiliki jumlah data dokumen latih sebanyak 2 buah, dan frekuensi dokumen dalam masing-masing kelas baik kelas sentimen negatif maupun kelas sentimen positif, terdapat 1 buah. Maka penghitungan data uji berdasarkan rumus 2.7, bagi kelas sentimen positif adalah seperti berikut.

$$P(v_{positif}) = \frac{|1|}{|2|} = 0.5$$

Dengan diketahuinya hasil penghitungan data uji maka untuk menghitung probabilitas pada kelas sentimen positif ditunjukkan pada Tabel 3.11 sebagai berikut.

Tabel 3.11 Penghitungan probabilitas kelas sentimen positif

Kata	Probabilitas
account	$P(a_{account} v_{positif}) = \frac{0.8 + 1}{3.2 + 45 } = 0.03734439834$
active	$P(a_{active} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
app	$P(a_{app} v_{positif}) = \frac{1 + 1}{3.2 + 45 } = 0.04149377593$

bad	$P(a_{bad} v_{positif}) = \frac{0.8 + 1}{3.2 + 45 } = 0.03734439834$
ban	$P(a_{ban} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
comfort	$P(a_{comfort} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
concern	$P(a_{concern} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
control	$P(a_{control} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
creat	$P(a_{creat} v_{positif}) = \frac{0.2 + 1}{3.2 + 45 } = 0.02489626556$
enjoy	$P(a_{enjoy} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
express	$P(a_{express} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
fake	$P(a_{fake} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
fix	$P(a_{fix} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
good	$P(a_{good} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
great	$P(a_{great} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
happy	$P(a_{happy} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
issue	$P(a_{issue} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
leader	$P(a_{leader} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
log	$P(a_{log} v_{positif}) = \frac{0.4 + 1}{3.2 + 45 } = 0.02904564315$
lot	$P(a_{lot} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
new	$P(a_{new} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$
news	$P(a_{news} v_{positif}) = \frac{0 + 1}{3.2 + 45 } = 0.02074688797$

nice	$P(a_{nice} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
notice	$P(a_{notice} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
opposite	$P(a_{opposite} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
person	$P(a_{person} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
play	$P(a_{play} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
poor	$P(a_{poor} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
publish	$P(a_{publish} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
real	$P(a_{real} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
review	$P(a_{review} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
room	$P(a_{room} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
sad	$P(a_{sad} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
shame	$P(a_{shame} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
suspicion	$P(a_{suspicion} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
stuff	$P(a_{stuff} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
team	$P(a_{team} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
time	$P(a_{time} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
thing	$P(a_{thing} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
twitter	$P(a_{twitter} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
utter	$P(a_{utter} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$

update	$P(a_{update} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
use	$P(a_{use} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
work	$P(a_{work} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$
video	$P(a_{video} v_{positif}) = \frac{0+1}{3.2+ 45 } = 0.02074688797$

Sedangkan penghitungan data uji bagi kelas sentimen negatif adalah seperti berikut.

$$P(v_{negatif}) = \frac{|1|}{|2|} = 0.5$$

Dengan diketahuinya hasil penghitungan data uji maka untuk menghitung probabilitas pada kelas sentimen negatif ditunjukkan pada Tabel 3.12 sebagai berikut.

Tabel 3.12 Penghitungan probabilitas kelas sentimen negatif

Kata	Probabilitas
account	$P(a_{account} v_{negatif}) = \frac{0.34+1}{1.6+ 45 } = 0.02875536481$
active	$P(a_{active} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
app	$P(a_{app} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
bad	$P(a_{bad} v_{negatif}) = \frac{0.34+1}{1.6+ 45 } = 0.02875536481$
ban	$P(a_{ban} v_{negatif}) = \frac{0.08+1}{1.6+ 45 } = 0.02317596567$
comfort	$P(a_{comfort} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
concern	$P(a_{concern} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
control	$P(a_{control} v_{negatif}) = \frac{0.08+1}{1.6+ 45 } = 0.02317596567$
creat	$P(a_{creat} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$

enjoy	$P(a_{enjoy} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
express	$P(a_{express} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
fake	$P(a_{fake} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
fix	$P(a_{fix} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
good	$P(a_{good} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
great	$P(a_{great} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
happy	$P(a_{happy} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
issue	$P(a_{issue} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
leader	$P(a_{leader} v_{negatif}) = \frac{0.16+1}{1.6+ 45 } = 0.02489270386$
log	$P(a_{log} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
lot	$P(a_{lot} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
new	$P(a_{new} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
news	$P(a_{news} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
nice	$P(a_{nice} v_{positif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
notice	$P(a_{notice} v_{positif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
opposite	$P(a_{opposite} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
person	$P(a_{person} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
play	$P(a_{play} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
poor	$P(a_{poor} v_{negatif}) = \frac{0.08+1}{1.6+ 45 } = 0.02317596567$

publish	$P(a_{publish} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
real	$P(a_{real} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
review	$P(a_{review} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
room	$P(a_{room} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
sad	$P(a_{sad} v_{negatif}) = \frac{0.08+1}{1.6+ 45 } = 0.02317596567$
shame	$P(a_{shame} v_{negatif}) = \frac{0.08+1}{1.6+ 45 } = 0.02317596567$
suspicion	$P(a_{suspicion} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
stuff	$P(a_{stuff} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
team	$P(a_{team} v_{negatif}) = \frac{0.16+1}{1.6+ 45 } = 0.02489270386$
time	$P(a_{time} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
thing	$P(a_{thing} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
twitter	$P(a_{twitter} v_{negatif}) = \frac{0.25+1}{1.6+ 45 } = 0.02682403433$
utter	$P(a_{utter} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
update	$P(a_{update} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
use	$P(a_{use} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
work	$P(a_{work} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$
video	$P(a_{video} v_{negatif}) = \frac{0+1}{1.6+ 45 } = 0.02145922747$

Nilai probabilitas ini dapat digunakan untuk mengukur kelas sentimen pada data uji yang akan digunakan, yaitu dokumen pada indeks 2. Tabel 3.13 menunjukkan data yang akan menjadi data uji.

Tabel 3.13 Data Uji *Naïve Bayes*

Data uji
<i>enjoy, use, app</i>

Pengujian data untuk menghitung probabilitas data uji berisikan sentimen positif dijalankan menggunakan probabilitas kelas sentimen positif pada Tabel 3.11 dan menggunakan rumus 2.3 seperti berikut.

$$\begin{aligned}
 P(\text{Data uji} | V_{\text{positif}}) &= P(a_{\text{enjoy}} | v_{\text{positif}}) * P(a_{\text{use}} | v_{\text{positif}}) * P(a_{\text{app}} | v_{\text{positif}}) \\
 &\quad * P(V_{\text{positif}}) \\
 P(\text{Data uji} | V_{\text{positif}}) &= 0.04149377593 * 0.02074638797 * 0.04149377593 * 0.5 \\
 &= 1.786 * 10^{-5}
 \end{aligned}$$

Selanjutnya pengujian data untuk menghitung probabilitas data uji berisikan sentimen negatif dijalankan menggunakan probabilitas kelas sentimen negatif pada Tabel 3.12 dan menggunakan rumus 2.3 seperti berikut.

$$\begin{aligned}
 P(\text{Data uji} | V_{\text{negatif}}) &= P(a_{\text{enjoy}} | v_{\text{negatif}}) * P(a_{\text{use}} | v_{\text{negatif}}) * P(a_{\text{app}} | v_{\text{negatif}}) \\
 &\quad * P(V_{\text{negatif}}) \\
 P(\text{Data uji} | V_{\text{negatif}}) &= 0.02145922747 * 0.02145922747 * 0.02145922747 * 0.5 \\
 &= 4.941 * 10^{-6}
 \end{aligned}$$

Berdasarkan penghitungan, diketahui bahwa nilai probabilitas kelas sentimen positif memiliki nilai sebesar $1.786 * 10^{-5}$, sedangkan nilai probabilitas kelas sentimen negatif memiliki nilai sebesar $4.941 * 10^{-6}$. Hal ini menunjukkan bahwa nilai probabilitas kelas sentimen positif lebih tinggi dibandingkan dengan nilai probabilitas kelas sentimen negatif. Maka, dapat disimpulkan bahwa data uji memiliki sentimen positif.

3.2.5.2. *Decision Tree*

Decision Tree atau pohon keputusan merupakan algoritma klasifikasi yang menggunakan penggambaran pohon untuk menyatakan keputusan. Dalam pembuatan pohon, digunakan Tabel 3.8 sebagai tabel informasi.

Dari Tabel 3.8 tersebut, digunakan kolom token sebagai atribut dengan nilai-nilai TF-IDF nya sebagai kelas-kelas terpisah apakah token tersebut > 0 atau $= 0$, sehingga didapatkan 45 atribut. Atribut-atribut ini nantinya akan diukur nilai *gain* menggunakan rumus 2.9 serta nilai *entropy* dari masing-masing atribut menggunakan rumus 2.8.

3.2.5.2.1. Pencarian nilai *entropy* dari atribut *account*

Pertama dilakukan klasifikasi atribut *account* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0 . Pengelompokkan akan dijelaskan pada Tabel 3.14 berikut.

Tabel 3.14. Pengelompokkan kelas pada atribut *account*

<i>account</i>	label	jumlah
> 0	0	3
> 0	1	1
$= 0$	0	1
$= 0$	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6.

$$Entropy(account > 0) = -\frac{1}{4} \log_2 \left(\frac{1}{4} \right) - \frac{3}{4} \log_2 \left(\frac{3}{4} \right) = 0.81$$

$$Entropy(account = 0) = -\frac{1}{6} \log_2 \left(\frac{1}{6} \right) - \frac{5}{6} \log_2 \left(\frac{5}{6} \right) = 0.65$$

Kemudian, ditentukan nilai *entropy* dari atribut *account* menggunakan rumus 2.8.

$$Entropy_{account} = \frac{4}{10} (0.81) + \frac{6}{10} (0.65) = 0.584$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9.

$$Gain(account > 0) = 0.81 - 0.584 = 0.226$$

$$Gain(account = 0) = 0.65 - 0.584 = 0.067$$

3.2.5.2.2. Pencarian nilai *entropy* dari atribut *active*

Pertama dilakukan klasifikasi atribut *active* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0 . Pengelompokkan akan dijelaskan pada Tabel 3.15

Tabel 3.15 Pengelompokkan kelas pada atribut *active*

<i>active</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(active > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(active = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *active* menggunakan rumus 2.8

$$Entropy_{active} = \frac{1}{10} (0) + \frac{9}{10} (0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(active > 0) = 0 - 0.828 = -0.828$$

$$Gain(active = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.3. Pencarian nilai *entropy* dari atribut *app*

Pertama dilakukan klasifikasi atribut *app* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.16.

Tabel 3.16. Pengelompokkan kelas pada atribut *app*

<i>app</i>	label	jumlah
> 0	0	1
> 0	1	4
= 0	0	3
= 0	1	2

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(app > 0) = -\frac{1}{5} \log_2 \left(\frac{1}{5} \right) - \frac{4}{5} \log_2 \left(\frac{4}{5} \right) = 0.73$$

$$Entropy(app = 0) = -\frac{3}{5} \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} \log_2 \left(\frac{2}{5} \right) = 0.97$$

Kemudian, ditentukan nilai *entropy* dari atribut *app* menggunakan rumus 2.8

$$Entropy_{app} = \frac{4}{10}(0.73) + \frac{6}{10}(0.97) = 0.874$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (app > 0) = 0.73 - 0.874 = -0.144$$

$$Gain (app = 0) = 0.97 - 0.874 = 0.096$$

3.2.5.2.4. Pencarian nilai *entropy* dari atribut *bad*

Pertama dilakukan klasifikasi atribut *bad* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.17

Tabel 3.17 Pengelompokkan kelas pada atribut *bad*

<i>bad</i>	label	jumlah
> 0	0	3
> 0	1	1
= 0	0	1
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(bad > 0) = -\frac{1}{4}LOG_2\left(\frac{1}{4}\right) - \frac{3}{4}LOG_2\left(\frac{3}{4}\right) = 0.81$$

$$Entropy (bad = 0) = -\frac{1}{6}LOG_2\left(\frac{1}{6}\right) - \frac{5}{6}LOG_2\left(\frac{5}{6}\right) = 0.65$$

Kemudian, ditentukan nilai *entropy* dari atribut *bad* menggunakan rumus 2.8

$$Entropy_{bad} = \frac{4}{10}(0.81) + \frac{6}{10}(0.65) = 0.584$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (bad > 0) = 0.81 - 0.584 = 0.226$$

$$Gain (bad = 0) = 0.65 - 0.584 = 0.067$$

3.2.5.2.5. Pencarian nilai *entropy* dari atribut *ban*

Pertama dilakukan klasifikasi atribut *ban* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.18

Tabel 3.18 Pengelompokkan kelas pada atribut *ban*

<i>ban</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(ban > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(ban = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *ban* menggunakan rumus 2.8

$$Entropy_{ban} = \frac{1}{10} (0) + \frac{9}{10} (0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(ban > 0) = 0 - 0.828 = -0.828$$

$$Gain(ban = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.6. Pencarian nilai *entropy* dari atribut *comfort*

Pertama dilakukan klasifikasi atribut *comfort* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.19

Tabel 3.19 Pengelompokkan kelas pada atribut *comfort*

<i>comfort</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(comfort > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy (comfort = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *comfort* menggunakan rumus 2.8

$$Entropy_{comfort} = \frac{1}{10} (0) + \frac{9}{10} (0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (comfort > 0) = 0 - 0.891 = -0.891$$

$$Gain (comfort = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.7. Pencarian nilai *entropy* dari atribut *concern*

Pertama dilakukan klasifikasi atribut *concern* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.20

Tabel 3.20 Pengelompokkan kelas pada atribut *concern*

<i>concern</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(concern > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy (concern = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *concern* menggunakan rumus 2.8

$$Entropy_{concern} = \frac{1}{10} (0) + \frac{9}{10} (0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (concern > 0) = 0 - 0.891 = -0.891$$

$$Gain (concern = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.8. Pencarian nilai *entropy* dari atribut *control*

Pertama dilakukan klasifikasi atribut *control* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.21

Tabel 3.21 Pengelompokkan kelas pada atribut *control*

<i>control</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(control > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(control = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *control* menggunakan rumus 2.8

$$Entropy_{control} = \frac{1}{10} (0) + \frac{9}{10} (0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(control > 0) = 0 - 0.828 = -0.828$$

$$Gain(control = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.9. Pencarian nilai *entropy* dari atribut *creat*

Pertama dilakukan klasifikasi atribut *creat* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.22

Tabel 3.22 Pengelompokkan kelas pada atribut *create*

<i>creat</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(creat > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(creat = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *creat* menggunakan rumus 2.8

$$Entropy_{creat} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(creat > 0) = 0 - 0.828 = -0.828$$

$$Gain(creat = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.10. Pencarian nilai *entropy* dari atribut *enjoy*

Pertama dilakukan klasifikasi atribut *enjoy* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.23

Tabel 3.23 Pengelompokkan kelas pada atribut *enjoy*

<i>comfort</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(enjoy > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(enjoy = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *enjoy* menggunakan rumus 2.8

$$Entropy_{enjoy} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(enjoy > 0) = 0 - 0.891 = -0.891$$

$$Gain(enjoy = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.11. Pencarian nilai *entropy* dari atribut *express*

Pertama dilakukan klasifikasi atribut *express* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.24

Tabel 3.24 Pengelompokkan kelas pada atribut *express*

<i>express</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(express > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(express = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *express* menggunakan rumus 2.8

$$Entropy_{express} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(express > 0) = 0 - 0.891 = -0.891$$

$$Gain(express = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.12. Pencarian nilai *entropy* dari atribut *fake*

Pertama dilakukan klasifikasi atribut *fake* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.25

Tabel 3.25 Pengelompokkan kelas pada atribut *fake*

<i>fake</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(fake > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(fake = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *fake* menggunakan rumus 2.8

$$Entropy_{fake} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(fake > 0) = 0 - 0.891 = -0.891$$

$$Gain(fake = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.13. Pencarian nilai *entropy* dari atribut *fix*

Pertama dilakukan klasifikasi atribut *fix* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.26

Tabel 3.26 Pengelompokkan kelas pada atribut *fix*

<i>fix</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(fix > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(fix = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *fix* menggunakan rumus 2.8

$$Entropy_{fix} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(fix > 0) = 0 - 0.828 = -0.828$$

$$Gain(fix = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.14. Pencarian nilai *entropy* dari atribut *good*

Pertama dilakukan klasifikasi atribut *good* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.27

Tabel 3.27 Pengelompokkan kelas pada atribut *good*

<i>good</i>	label	jumlah
> 0	0	0
> 0	1	2
= 0	0	4
= 0	1	4

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(good > 0) = -\frac{2}{2} \text{LOG}_2 \left(\frac{2}{2} \right) = 0$$

$$Entropy(good = 0) = -\frac{4}{8} \text{LOG}_2 \left(\frac{4}{8} \right) - \frac{4}{8} \text{LOG}_2 \left(\frac{4}{8} \right) = 1$$

Kemudian, ditentukan nilai *entropy* dari atribut *good* menggunakan rumus 2.8

$$Entropy_{good} = \frac{2}{10} (0) + \frac{8}{10} (0.5) = 0.8$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(good > 0) = 0 - 0.8 = -0.8$$

$$Gain(good = 0) = 1 - 0.8 = 0.2$$

3.2.5.2.15. Pencarian nilai *entropy* dari atribut *great*

Pertama dilakukan klasifikasi atribut *great* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.28

Tabel 3.28 Pengelompokkan kelas pada atribut *great*

<i>great</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(great > 0) = -\frac{1}{1} \text{LOG}_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(great = 0) = -\frac{4}{9} \text{LOG}_2 \left(\frac{4}{9} \right) - \frac{5}{9} \text{LOG}_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *great* menggunakan rumus 2.8

$$Entropy_{great} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(great > 0) = 0 - 0.891 = -0.891$$

$$Gain(great = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.16. Pencarian nilai *entropy* dari atribut *happy*

Pertama dilakukan klasifikasi atribut *happy* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.29

Tabel 3.29 Pengelompokkan kelas pada atribut *happy*

<i>happy</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(happy > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(happy = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *happy* menggunakan rumus 2.8

$$Entropy_{happy} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(happy > 0) = 0 - 0.891 = -0.891$$

$$Gain(happy = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.17. Pencarian nilai *entropy* dari atribut *issue*

Pertama dilakukan klasifikasi atribut *issue* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.30

Tabel 3.30 Pengelompokkan kelas pada atribut *issue*

<i>issue</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(issue > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(issue = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *issue* menggunakan rumus 2.8

$$Entropy_{issue} = \frac{1}{10} (0) + \frac{9}{10} (0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(issue > 0) = 0 - 0.891 = -0.891$$

$$Gain(issue = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.18. Pencarian nilai *entropy* dari atribut *leader*

Pertama dilakukan klasifikasi atribut *leader* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.31

Tabel 3.31. Pengelompokkan kelas pada atribut *leader*

<i>leader</i>	label	jumlah
> 0	0	1
> 0	1	1
= 0	0	3
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(leader > 0) = -\frac{1}{2} \log_2 \left(\frac{1}{2} \right) - \frac{1}{2} \log_2 \left(\frac{2}{2} \right) = 1$$

$$Entropy(leader = 0) = -\frac{3}{8} \log_2 \left(\frac{3}{8} \right) - \frac{5}{8} \log_2 \left(\frac{5}{8} \right) = 0.96$$

Kemudian, ditentukan nilai *entropy* dari atribut *leader* menggunakan rumus 2.8

$$Entropy_{leader} = \frac{2}{10}(1) + \frac{8}{10}(0.96) = 0.868$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (leader > 0) = 1 - 0.868 = 0.132$$

$$Gain (leader = 0) = 0.96 - 0.868 = 0.092$$

3.2.5.2.19. Pencarian nilai *entropy* dari atribut *log*

Pertama dilakukan klasifikasi atribut *log* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.32

Tabel 3.32. Pengelompokkan kelas pada atribut *log*

<i>log</i>	label	jumlah
> 0	0	2
> 0	1	0
= 0	0	2
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(log > 0) = -\frac{2}{2} \log_2 \left(\frac{2}{2} \right) = 0$$

$$Entropy (log = 0) = -\frac{2}{8} \log_2 \left(\frac{2}{8} \right) - \frac{6}{8} \log_2 \left(\frac{6}{8} \right) = 0.81$$

Kemudian, ditentukan nilai *entropy* dari atribut *log* menggunakan rumus 2.8

$$Entropy_{log} = \frac{2}{10}(0) + \frac{8}{10}(0.81) = 0.648$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (log > 0) = 1 - 0.614 = 0.386$$

$$Gain (log = 0) = 0.96 - 0.614 = 0.346$$

3.2.5.2.20. Pencarian nilai *entropy* dari atribut *lot*

Pertama dilakukan klasifikasi atribut *lot* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.33

Tabel 3.33 Pengelompokkan kelas pada atribut *lot*

<i>lot</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(\text{lot} > 0) = -\frac{1}{1} \text{LOG}_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(\text{lot} = 0) = -\frac{4}{9} \text{LOG}_2 \left(\frac{4}{9} \right) - \frac{5}{9} \text{LOG}_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *happy* menggunakan rumus 2.8

$$Entropy_{lot} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(\text{lot} > 0) = 0 - 0.891 = -0.891$$

$$Gain(\text{lot} = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.21. Pencarian nilai *entropy* dari atribut *new*

Pertama dilakukan klasifikasi atribut *new* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.34

Tabel 3.34 Pengelompokkan kelas pada atribut *new*

<i>new</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(\text{new} > 0) = -\frac{1}{1} \text{LOG}_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(\text{new} = 0) = -\frac{3}{9} \text{LOG}_2 \left(\frac{3}{9} \right) - \frac{6}{9} \text{LOG}_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *new* menggunakan rumus 2.8

$$Entropy_{new} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (new > 0) = 0 - 0.828 = -0.828$$

$$Gain (new = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.22. Pencarian nilai *entropy* dari atribut *news*

Pertama dilakukan klasifikasi atribut *news* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.35

Tabel 3.35 Pengelompokkan kelas pada atribut *news*

<i>news</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan Tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(news > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy (news = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *news* menggunakan rumus 2.8

$$Entropy_{news} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (news > 0) = 0 - 0.891 = -0.891$$

$$Gain (news = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.23. Pencarian nilai *entropy* dari atribut *nice*

Pertama dilakukan klasifikasi atribut *nice* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.36

Tabel 3.36 Pengelompokkan kelas pada atribut *nice*

<i>nice</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(nice > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(nice = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *nice* menggunakan rumus 2.8

$$Entropy_{nice} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(nice > 0) = 0 - 0.891 = -0.891$$

$$Gain(nice = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.24. Pencarian nilai *entropy* dari atribut *notice*

Pertama dilakukan klasifikasi atribut *notice* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.37.

Tabel 3.37 Pengelompokkan kelas pada atribut *notice*

<i>notice</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(notice > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(notice = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *notice* menggunakan rumus 2.8

$$Entropy_{notice} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (notice > 0) = 0 - 0.891 = -0.891$$

$$Gain (notice = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.25. Pencarian nilai *entropy* dari atribut *opposite*

Pertama dilakukan klasifikasi atribut *opposite* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.38

Tabel 3.38 Pengelompokkan kelas pada atribut *opposite*

<i>opposite</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(opposite > 0) = -\frac{1}{1}LOG_2\left(\frac{1}{1}\right) = 0$$

$$Entropy (opposite = 0) = -\frac{3}{9}LOG_2\left(\frac{3}{9}\right) - \frac{6}{9}LOG_2\left(\frac{6}{9}\right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *fix* menggunakan rumus 2.8

$$Entropy_{opposite} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhi dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (opposite > 0) = 0 - 0.828 = -0.828$$

$$Gain (opposite = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.26. Pencarian nilai *entropy* dari atribut *person*

Pertama dilakukan klasifikasi atribut *person* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.39

Tabel 3.39 Pengelompokkan kelas pada atribut *person*

<i>person</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(person > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(person = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *person* menggunakan rumus 2.8

$$Entropy_{person} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(person > 0) = 0 - 0.891 = -0.891$$

$$Gain(person = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.27. Pencarian nilai *entropy* dari atribut *play*

Pertama dilakukan klasifikasi atribut *play* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.40

Tabel 3.40 Pengelompokkan kelas pada atribut *play*

<i>play</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(play > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(play = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *play* menggunakan rumus 2.8

$$Entropy_{play} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(play > 0) = 0 - 0.828 = -0.828$$

$$Gain(play = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.28. Pencarian nilai *entropy* dari atribut *poor*

Pertama dilakukan klasifikasi atribut *poor* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.41

Tabel 3.41 Pengelompokkan kelas pada atribut *poor*

<i>poor</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(poor > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(poor = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *poor* menggunakan rumus 2.8

$$Entropy_{poor} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(poor > 0) = 0 - 0.828 = -0.828$$

$$Gain(poor = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.29. Pencarian nilai *entropy* dari atribut *publish*

Pertama dilakukan klasifikasi atribut *publish* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.42

Tabel 3.42 Pengelompokkan kelas pada atribut *publish*

<i>publish</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(publish > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(publish = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *publish* menggunakan rumus 2.8

$$Entropy_{publish} = \frac{1}{10} (0) + \frac{9}{10} (0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(publish > 0) = 0 - 0.891 = -0.891$$

$$Gain(publish = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.30. Pencarian nilai *entropy* dari atribut *real*

Pertama dilakukan klasifikasi atribut *real* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.43

Tabel 3.43 Pengelompokkan kelas pada atribut *real*

<i>real</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(real > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(real = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *real* menggunakan rumus 2.8

$$Entropy_{real} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (real > 0) = 0 - 0.891 = -0.891$$

$$Gain (real = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.31. Pencarian nilai *entropy* dari atribut *review*

Pertama dilakukan klasifikasi atribut *review* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.44

Tabel 3.44 Pengelompokkan kelas pada atribut *review*

<i>review</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(review > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy (review = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *review* menggunakan rumus 2.8

$$Entropy_{review} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (review > 0) = 0 - 0.891 = -0.891$$

$$Gain (review = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.32. Pencarian nilai *entropy* dari atribut *room*

Pertama dilakukan klasifikasi atribut *room* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.45

Tabel 3.45 Pengelompokkan kelas pada atribut *room*

<i>room</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(room > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(room = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *room* menggunakan rumus 2.8

$$Entropy_{room} = \frac{1}{10} (0) + \frac{9}{10} (0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(room > 0) = 0 - 0.891 = -0.891$$

$$Gain(room = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.33. Pencarian nilai *entropy* dari atribut *sad*

Pertama dilakukan klasifikasi atribut *sad* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.46

Tabel 3.46 Pengelompokkan kelas pada atribut *sad*

<i>sad</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(sad > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(sad = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *sad* menggunakan rumus 2.8

$$Entropy_{sad} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(sad > 0) = 0 - 0.828 = -0.828$$

$$Gain(sad = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.34. Pencarian nilai *entropy* dari atribut *shame*

Pertama dilakukan klasifikasi atribut *shame* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.47

Tabel 3.47 Pengelompokkan kelas pada atribut *shame*

<i>shame</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(shame > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(shame = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *shame* menggunakan rumus 2.8

$$Entropy_{shame} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(shame > 0) = 0 - 0.828 = -0.828$$

$$Gain(shame = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.35. Pencarian nilai *entropy* dari atribut *suspicion*

Pertama dilakukan klasifikasi atribut *suspicion* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.48

Tabel 3.48 Pengelompokkan kelas pada atribut *suspicion*

<i>suspicion</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(suspicion > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(suspicion = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *suspicion* menggunakan rumus 2.8

$$Entropy_{suspicion} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(suspicion > 0) = 0 - 0.828 = -0.828$$

$$Gain(suspicion = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.36. Pencarian nilai *entropy* dari atribut *stuff*

Pertama dilakukan klasifikasi atribut *stuff* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.49

Tabel 3.49 Pengelompokkan kelas pada atribut *stuff*

<i>stuff</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan Tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(stuff > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(stuff = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *stuff* menggunakan rumus 2.8

$$Entropy_{stuff} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (stuff > 0) = 0 - 0.891 = -0.891$$

$$Gain (stuff = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.37. Pencarian nilai *entropy* dari atribut *team*

Pertama dilakukan klasifikasi atribut *team* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.50

Tabel 3.50 Pengelompokkan kelas pada atribut *team*

<i>team</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(team > 0) = -\frac{1}{1}LOG_2\left(\frac{1}{1}\right) = 0$$

$$Entropy (team = 0) = -\frac{3}{9}LOG_2\left(\frac{3}{9}\right) - \frac{6}{9}LOG_2\left(\frac{6}{9}\right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *team* menggunakan rumus 2.8

$$Entropy_{team} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (team > 0) = 0 - 0.828 = -0.828$$

$$Gain (team = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.38. Pencarian nilai *entropy* dari atribut *time*

Pertama dilakukan klasifikasi atribut *time* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.51

Tabel 3.51 Pengelompokkan kelas pada atribut *time*

<i>time</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(time > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(time = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *time* menggunakan rumus 2.8

$$Entropy_{time} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(time > 0) = 0 - 0.891 = -0.891$$

$$Gain(time = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.39. Pencarian nilai *entropy* dari atribut *thing*

Pertama dilakukan klasifikasi atribut *thing* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.52

Tabel 3.52 Pengelompokkan kelas pada atribut *thing*

<i>thing</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(thing > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(thing = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *thing* menggunakan rumus 2.8

$$Entropy_{thing} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (thing > 0) = 0 - 0.828 = -0.828$$

$$Gain (thing = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.40. Pencarian nilai *entropy* dari atribut *twitter*

Pertama dilakukan klasifikasi atribut *twitter* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.53

Tabel 3.53 Pengelompokkan kelas pada atribut *twitter*

<i>twitter</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(twitter > 0) = -\frac{1}{3}LOG_2\left(\frac{1}{3}\right) - \frac{2}{3}LOG_2\left(\frac{2}{3}\right) = 0.91$$

$$Entropy (twitter = 0) = -\frac{3}{7}LOG_2\left(\frac{3}{7}\right) - \frac{4}{7}LOG_2\left(\frac{4}{7}\right) = 0.98$$

Kemudian, ditentukan nilai *entropy* dari atribut *twitter* menggunakan rumus 2.8

$$Entropy_{twitter} = \frac{3}{10}(0.91) + \frac{7}{10}(0.98) = 0.959$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (twitter > 0) = 0.91 - 0.959 = 0.049$$

$$Gain (twitter = 0) = 0.98 - 0.959 = 0.021$$

3.2.5.2.41. Pencarian nilai *entropy* dari atribut *utter*

Pertama dilakukan klasifikasi atribut *utter* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.54

Tabel 3.54 Pengelompokkan kelas pada atribut *utter*

<i>utter</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(utter > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(utter = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *utter* menggunakan rumus 2.8

$$Entropy_{utter} = \frac{1}{10} (0) + \frac{9}{10} (0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(utter > 0) = 0 - 0.828 = -0.828$$

$$Gain(utter = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.42. Pencarian nilai *entropy* dari atribut *update*

Pertama dilakukan klasifikasi atribut *update* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.55

Tabel 3.55 Pengelompokkan kelas pada atribut *update*

<i>update</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(update > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(update = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *update* menggunakan rumus 2.8

$$Entropy_{update} = \frac{1}{10}(0) + \frac{9}{10}(0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (update > 0) = 0 - 0.828 = -0.828$$

$$Gain (update = 0) = 0.92 - 0.828 = 0.092$$

3.2.5.2.43. Pencarian nilai *entropy* dari atribut *use*

Pertama dilakukan klasifikasi atribut *use* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.56

Tabel 3.56 Pengelompokkan kelas pada atribut *use*

<i>use</i>	label	jumlah
> 0	0	0
> 0	1	2
= 0	0	4
= 0	1	4

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(use > 0) = -\frac{2}{2}LOG_2\left(\frac{2}{2}\right) = 0$$

$$Entropy (use = 0) = -\frac{4}{8}LOG_2\left(\frac{4}{8}\right) - \frac{4}{8}LOG_2\left(\frac{4}{8}\right) = 1$$

Kemudian, ditentukan nilai *entropy* dari atribut *use* menggunakan rumus 2.8

$$Entropy_{use} = \frac{2}{10}(0) + \frac{8}{10}(0.5) = 0.8$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (use > 0) = 0 - 0.8 = -0.8$$

$$Gain (use = 0) = 1 - 0.8 = 0.2$$

3.2.5.2.44. Pencarian nilai *entropy* dari atribut *work*

Pertama dilakukan klasifikasi atribut *work* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.57

Tabel 3.57 Pengelompokkan kelas pada atribut *work*

<i>work</i>	label	jumlah
> 0	0	0
> 0	1	1
= 0	0	4
= 0	1	5

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(work > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(work = 0) = -\frac{4}{9} \log_2 \left(\frac{4}{9} \right) - \frac{5}{9} \log_2 \left(\frac{5}{9} \right) = 0.99$$

Kemudian, ditentukan nilai *entropy* dari atribut *work* menggunakan rumus 2.8

$$Entropy_{work} = \frac{1}{10}(0) + \frac{9}{10}(0.99) = 0.891$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain(work > 0) = 0 - 0.891 = -0.891$$

$$Gain(work = 0) = 0.99 - 0.891 = 0.099$$

3.2.5.2.45. Pencarian nilai *entropy* dari atribut *video*

Pertama dilakukan klasifikasi atribut *video* yang memiliki nilai TF-IDF = 0 dan nilai TF-IDF > 0. Pengelompokkan akan dijelaskan pada Tabel 3.58

Tabel 3.58 Pengelompokkan kelas pada atribut *video*

<i>video</i>	label	jumlah
> 0	0	1
> 0	1	0
= 0	0	3
= 0	1	6

Berdasarkan tabel klasifikasi diatas, dapat ditentukan nilai *entropy* dari kelas masing masing menggunakan rumus 2.6

$$Entropy(video > 0) = -\frac{1}{1} \log_2 \left(\frac{1}{1} \right) = 0$$

$$Entropy(video = 0) = -\frac{3}{9} \log_2 \left(\frac{3}{9} \right) - \frac{6}{9} \log_2 \left(\frac{6}{9} \right) = 0.92$$

Kemudian, ditentukan nilai *entropy* dari atribut *video* menggunakan rumus 2.8

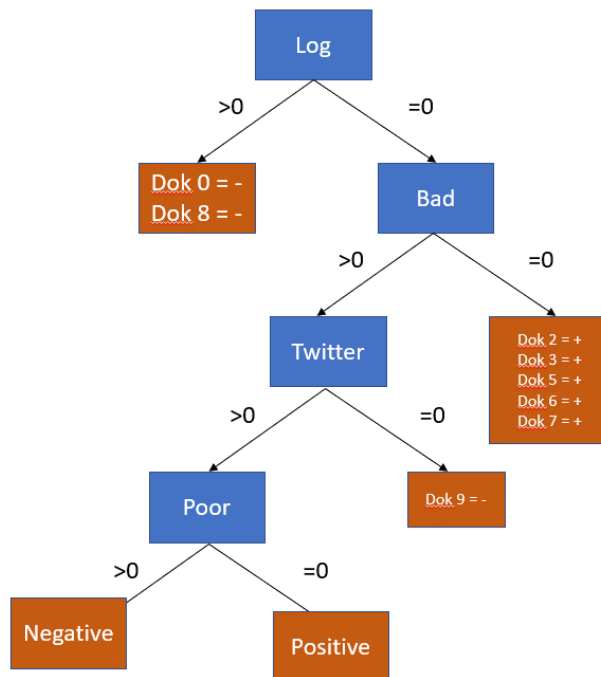
$$Entropy_{video} = \frac{1}{10} (0) + \frac{9}{10} (0.92) = 0.828$$

Terakhir dicari nilai *gain* berdasarkan setiap kelas menggunakan rumus 2.9

$$Gain (video > 0) = 0 - 0.828 = -0.828$$

$$Gain (video = 0) = 0.92 - 0.828 = 0.092$$

Berdasarkan perhitungan nilai *gain*, dibuat pohon keputusan (*decision tree*) yang ditentukan *root* dari pohon berdasarkan nilai *gain* terbesar. Berikut Gambar 3.3 menunjukkan pohon keputusan



Gambar 3.3 Hasil pohon keputusan

Berdasarkan pohon keputusan, ditentukan aturan prediksi, yakni “apabila data memiliki kata *bad*, *twitter*, dan *poor* maka data tersebut memiliki sentimen negatif”.

Selanjutnya melakukan pengujian hasil aturan prediksi dengan data uji yang telah dilakukan pra pemrosesan sehingga menghasilkan data seperti berikut.

Tabel 3.59 Data uji *decision tree*

Data uji
<i>poor, twitter, control, team, ban, leader,</i> <i>twitter, account, shame, sad, bad, team</i>

Berdasarkan aturan prediksi yang dibuat, data uji memiliki sentimen negatif dikarenakan data ini memenuhi syarat untuk menjadi data yang memiliki sentimen negatif.

3.2.6 Evaluasi

Dalam pembuatan evaluasi ini, digunakan *confusion matrix* untuk mengetahui nilai positif sejati atau *True Positive* (TP), positif palsu atau *Fake Positive* (FP), negatif palsu atau *Fake Negative* (FN), negatif sejati atau *True Negative* (TN). Berikut Tabel 3.60 menunjukkan contoh *confusion matrix*

Tabel 3.60 Contoh *confusion Matrix*

	Data asli		
	Matrix	Positif	Negatif
	Positif	TP	FP
	Negatif	FN	TN

Dimana:

TP = *True Positive* (jumlah data positif yang diprediksi sebagai positif)

FP = *Fake Positive* (jumlah data negatif yang diprediksi sebagai positif)

FN = *Fake Negative* (jumlah data positif yang diprediksi sebagai data negatif)

TN = *True Negative* (jumlah data negatif yang diprediksi sebagai data negatif).

Digunakan sampel yang telah dibuat dengan model prediksi *Naïve Bayes* dengan hasil prediksi dan nilai label nya seperti Tabel 3.61 berikut.

Tabel 3.61 Sampel hasil prediksi *Naïve Bayes*

<u>Indeks</u>	<u>Label</u>	<u>Prediksi</u>
0	Negatif	Negatif
1	Negatif	Negatif
2	Positif	Positif
3	Positif	Positif
4	Positif	Positif
5	Positif	Negatif
6	Positif	Positif
7	Positif	Positif
8	Negatif	Positif
9	Negatif	Negatif

Digunakan sampel yang telah dibuat dengan model prediksi *Decision Tree* dengan hasil prediksi dan nilai label nya seperti Tabel 3.62 berikut.

Tabel 3.62 Sampel hasil prediksi *Decision Tree*

<u>Indeks</u>	<u>Label</u>	<u>Prediksi</u>
0	Negatif	Negatif
1	Negatif	Positif
2	Positif	Positif
3	Positif	Negatif
4	Positif	Positif
5	Positif	Positif
6	Positif	Positif
7	Positif	Positif
8	Negatif	Negatif
9	Negatif	Positif

3.2.6.1 Confusion Matrix

Penghitungan *confusion matrix* pada hasil prediksi *Naïve Bayes* dengan menggunakan sampel pada Tabel 3.61 didapatkan *confusion matrix* dengan hasil seperti Tabel 3.63 berikut.

Tabel 3.63 *Confusion matrix* dari sampel *Naïve Bayes*

	Data asli		
	Matrix	Positif	Negatif
	Positif	5	1
	Negatif	1	3

Berdasarkan tabel diatas, dapat diketahui bahwa untuk data yang diprediksi dengan algoritma *Naïve Bayes* memiliki nilai TP sejumlah 5, nilai FP sejumlah 1, nilai FN sejumlah 1, dan nilai TN sejumlah 3.

Penghitungan *confusion matrix* pada hasil prediksi *Decision Tree* dengan menggunakan sampel pada Tabel 3.62 didapatkan *confusion matrix* seperti Tabel 3.64 berikut.

Tabel 3.64 *Confusion matrix* dari sampel *Decision Tree*

Hasil prediksi	Data asli		
	Matrix	Positif	Negatif
	Positif	5	2
	Negatif	1	2

Berdasarkan gambar, dapat diketahui bahwa untuk data yang diprediksi dengan algoritma *Decision Tree* memiliki nilai TP sejumlah 5, nilai FP sejumlah 2, nilai FN sejumlah 1, dan nilai TN yang sejumlah 2.

Dengan mengetahui nilai-nilai diatas, dapat dihitung evaluasi performa pembuatan model seperti nilai akurasi, presisi, *recall*, serta *F1-Score*. Selain itu, dapat juga digunakan untuk melakukan penggambaran kurva ROC (*Receiving Operating Characteristic*) yang dapat dihitung luas area dibawah kurva tersebut. Penghitungan luas area dibawah kurva ROC disebut juga sebagai AUC-ROC (*Area Under Curve-Receiving Operating Characteristic*).

3.2.6.2 Nilai akurasi

Nilai akurasi merupakan pengukuran yang mengukur tingkat akurasi algoritma dalam melakukan pembangunan model prediksi yang diuji menggunakan data uji. Perhitungan akurasi akan menggunakan rumus 2.10. Berikut perhitungan akurasi dari model prediksi *Naïve Bayes*.

$$Akurasi = \frac{5 + 3}{5 + 1 + 1 + 3}$$

$$Akurasi = \frac{8}{10}$$

$$Akurasi = 80\%$$

Selanjutnya, perhitungan akurasi dari model prediksi *Decision Tree*

$$Akurasi = \frac{5 + 2}{5 + 1 + 2 + 2}$$

$$Akurasi = \frac{7}{10}$$

$$Akurasi = 70\%$$

3.2.6.3 Nilai presisi

Nilai presisi mengukur frekuensi data positif (TP) terhadap seluruh data yang diprediksi sebagai positif (TP dan FP). Perhitungan presisi akan menggunakan rumus 2.11. Berikut perhitungan nilai presisi dari model prediksi *Naïve Bayes*.

$$Presisi = \frac{5}{5 + 1}$$

$$Presisi = \frac{5}{6}$$

$$Presisi = 0.834 = 83.4\%$$

Selanjutnya, perhitungan presisi dari model prediksi *Decision Tree*.

$$Presisi = \frac{5}{5 + 2}$$

$$Presisi = \frac{5}{7}$$

$$Presisi = 0.714 = 71.4\%$$

3.2.6.4 Nilai *recall*

Nilai *recall* mengukur frekuensi data positif (TP) terhadap seluruh data yang sejatinya bernilai positif (TP dan FN). Perhitungan *recall* akan menggunakan rumus 2.12. Berikut perhitungan *recall* dari model *Naïve Bayes*.

$$Recall = \frac{5}{5 + 1}$$

$$Recall = \frac{5}{6}$$

$$Recall = 0.834 = 83.4\%$$

Selanjutnya, perhitungan *recall* dari model *Decision Tree*.

$$Recall = \frac{5}{5 + 1}$$

$$Recall = \frac{5}{6}$$

$$Recall = 0.834 = 83.4\%$$

3.2.6.5 Nilai *F1-Score*

Nilai *F1-Score* mengukur frekuensi harmonik dari nilai *recall* dan nilai presisi. Perhitungan *F1-Score* akan menggunakan rumus 2.13. Berikut perhitungan nilai *F1-Score* dari model prediksi *Naïve Bayes*.

$$F1 - Score = 2 * \frac{0.834 * 0.834}{0.834 + 0.834}$$

$$F1 - Score = 0.834 = 83.4\%$$

Selanjutnya perhitungan *F1-Score* dari model prediksi *Decision Tree*.

$$F1 - Score = 2 * \frac{0.834 * 0.714}{0.834 + 0.714}$$

$$F1 - Score = 0.769 = 76.9\%$$

3.2.6.6 Penghitungan luas area dibawah kurva ROC (AUC-ROC)

Sebelum perhitungan luas dilakukan, dibuat kurva ROC terlebih dahulu. Kurva ini dibentuk dengan menggunakan pemetaan titik dari hasil pasangan *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada setiap batasan klasifikasi. Batasan berada pada *range* 0 hingga 1. Dengan semakin besar nilai batasan, model prediksi akan bekerja lebih baik dalam melakukan prediksi pada nilai negatif. Sebaliknya, semakin kecil nilai batasan, model prediksi akan bekerja lebih baik pada nilai positif.

3.2.6.6.1 Penghitungan dengan sampel dari model prediksi Naïve Bayes

3.2.6.6.1.1 Penghitungan FPR dan TPR berdasarkan batasan yang ditentukan

Berdasarkan dataset sampel, ditentukan 4 batasan klasifikasi, dengan contoh proses klasifikasi seperti berikut

- Batas 0

Pada batas ini, semua nilai positif diklasifikasi seluruhnya secara benar, dan nilai negatif diklasifikasikan seluruhnya secara salah. Hal ini dapat dilihat dengan hasil prediksi menunjukkan semua nilai positif, meskipun label menyatakan nilai negatif. Berikut Tabel 3.65 hasil klasifikasi pada batas 0.

Tabel 3.65 Hasil klasifikasi batas 0 *Naïve Bayes*

Indeks	Label	Prediksi
0	Negatif	Positif
1	Negatif	Positif
2	Positif	Positif
3	Positif	Positif
4	Positif	Positif
5	Positif	Positif
6	Positif	Positif
7	Positif	Positif
8	Negatif	Positif
9	Negatif	Positif

Hal ini membuat nilai TP = 6, nilai FP = 4, nilai FN = 0, serta nilai TN = 0

Pemetaan titik menggunakan nilai FPR dan TPR, dengan penghitungan TPR merujuk pada rumus 2.14

$$TPR = \frac{6}{(6 + 0)} = 1$$

Penghitungan FPR dihitung merujuk pada rumus 2.15

$$FPR = \frac{4}{(4 + 0)} = 1$$

Maka pemetaan (FPR, TPR) pada batasan ini (1, 1)

- Batas 0.5

Pada batas ini, 50% dari dataset nilai negatif nya diklasifikasikan secara tepat, dengan nilai positif yang diklasifikasikan secara salah, namun 50% lainnya masih mengklasifikasi nilai positif secara tepat, dengan nilai negatif yang diklasifikasi secara salah. Berikut Tabel 3.66 hasil klasifikasi pada batas 0.5.

Tabel 3.66 Hasil klasifikasi batas 0.5 *Naïve Bayes*

Indeks	Label	Prediksi
0	Negatif	Negatif
1	Negatif	Negatif
2	Positif	Negatif
3	Positif	Negatif
4	Positif	Negatif
5	Positif	Positif
6	Positif	Positif
7	Positif	Positif
8	Negatif	Positif
9	Negatif	Positif

Hal ini membuat nilai TP = 3, nilai FP = 2, nilai FN = 3, nilai TN = 2

Pemetaan titik menggunakan nilai FPR dan TPR, dengan penghitungan TPR merujuk pada rumus 2.14

$$TPR = \frac{3}{(3 + 3)} = 0.5$$

Penghitungan FPR dihitung merujuk pada rumus 2.15

$$FPR = \frac{2}{(2 + 2)} = 0.5$$

Maka pemetaan (FPR, TPR) pada batasan ini (0.5, 0.5)

- Batas 0.8

Pada batas ini, 80% dari dataset nilai negatif nya diklasifikasikan secara tepat, dengan nilai positif yang diklasifikasikan secara salah

namun 20% lainnya masih mengklasifikasi nilai positif secara tepat, dengan nilai negatif yang diklasifikasi secara salah. Berikut Tabel 3.67 hasil klasifikasi pada batas 0.8.

Tabel 3.67 Hasil klasifikasi batas 0.8 *Naïve Bayes*

Indeks	Label	Prediksi
0	Negatif	Negatif
1	Negatif	Negatif
2	Positif	Negatif
3	Positif	Negatif
4	Positif	Negatif
5	Positif	Negatif
6	Positif	Negatif
7	Positif	Negatif
8	Negatif	Positif
9	Negatif	Positif

Hal ini membuat nilai TP = 0, nilai FP = 2, nilai FN = 6, nilai TN = 2
 Pemetaan titik menggunakan nilai FPR dan TPR, dengan penghitungan
 TPR merujuk pada rumus 2.14

$$TPR = \frac{0}{(0 + 6)} = 0$$

Penghitungan FPR dihitung merujuk pada rumus 2.15

$$FPR = \frac{2}{(2 + 2)} = 0.5$$

Maka pemetaan (FPR, TPR) pada batasan ini (0, 0.5)

- Batas 1

Pada batas ini, semua nilai positif diklasifikasi seluruhnya secara salah, dan nilai negatif diklasifikasikan seluruhnya secara benar. Hal ini dapat dilihat dengan hasil prediksi menunjukkan semua nilai positif, meskipun label menyatakan nilai negatif. Berikut Tabel 3.68 menunjukkan hasil klasifikasi pada batas 1.

Tabel 3.68 Hasil klasifikasi batas 1 *Naïve Bayes*

Indeks	Label	Prediksi
0	Negatif	Negatif
1	Negatif	Negatif
2	Positif	Negatif
3	Positif	Negatif
4	Positif	Negatif
5	Positif	Negatif
6	Positif	Negatif
7	Positif	Negatif
8	Negatif	Negatif
9	Negatif	Negatif

Hal ini membuat nilai TP = 0, nilai FP = 0, nilai FN = 6, dan nilai TN = 4

Pemetaan titik menggunakan nilai FPR dan TPR, dengan penghitungan TPR merujuk pada rumus 2.14

$$TPR = \frac{0}{(0 + 6)} = 0$$

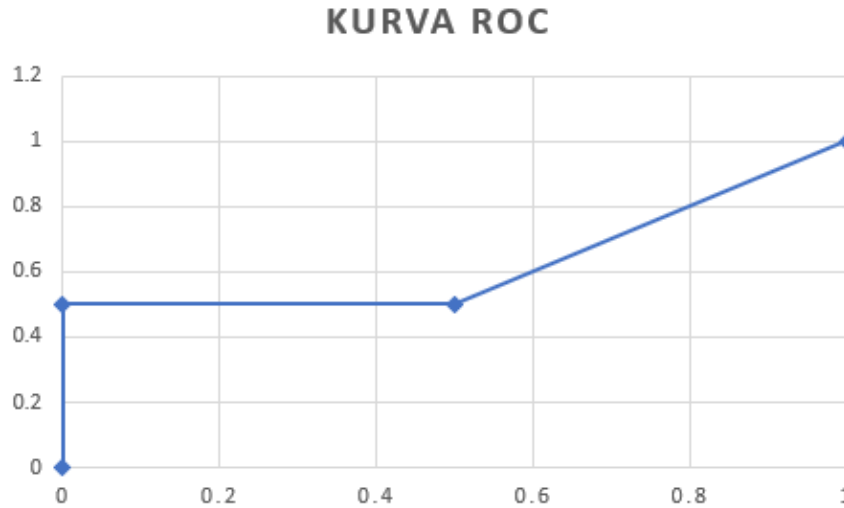
Penghitungan FPR dihitung merujuk pada rumus 2.15

$$FPR = \frac{0}{(0 + 4)} = 0$$

Maka pemetaan (FPR, TPR) pada batasan ini (0, 0)

3.2.6.6.1.2 Pemetaan titik TPR dan FPR

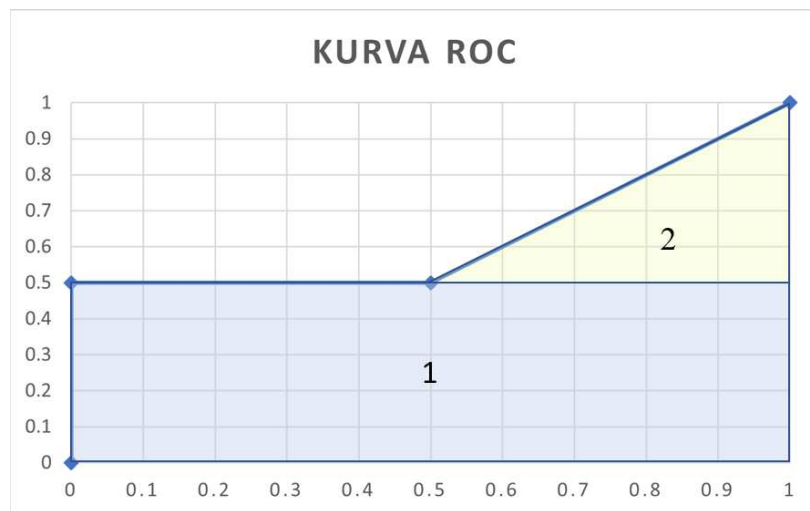
Berdasarkan penghitungan FPR dan TPR, diketahui himpunan (TPR, FPR), yakni berisikan {(1, 1), (0.5, 0.5), (0, 0.5), (0, 0)}. Pembuatan kurva ROC dapat dilakukan dengan melakukan pemetaan terhadap himpunan tersebut, dengan hasil seperti Gambar 3.4 berikut.



Gambar 3.4 Kurva ROC algoritma *Naïve Bayes*

3.2.6.6.1.3 Penghitungan AUC-ROC

Penghitungan nilai AUC-ROC dapat ditentukan dengan menghitung luas dari area dibawah kurva ROC, perhitungan luas digambarkan dengan Gambar 3.5 berikut.



Gambar 3.5 Penghitungan luas area dibawah kurva ROC *Naïve Bayes*

Kurva membentuk 2 bangun, dimana bangun 1 berupa persegi panjang dengan panjang 1 satuan dan lebar 0.5 satuan. Bangun 2 berupa segitiga siku-siku dengan alas 0.5 satuan dan tinggi 0.5 satuan.

Penghitungan luas bangun 1, dapat menggunakan rumus penghitungan luas persegi panjang seperti berikut.

$$\text{Luas 1} = \text{panjang} * \text{lebar}$$

$$\text{Luas 1} = 1 * 0.5$$

$$\text{Luas 1} = 0.5 \text{ satuan}$$

Selanjutnya penghitungan luas bangun 2 seperti berikut.

$$\text{Luas 2} = (0.5) * \text{alas} * \text{tinggi}$$

$$\text{Luas 2} = (0.5) * 0.5 * 0.5$$

$$\text{Luas 2} = 0.125 \text{ satuan}$$

Total luas seluruh bangun dapat ditentukan dengan menjumlahkan luas bangun 1 dengan luas bangun 2.

$$\text{Luas total} = \text{Luas 1} + \text{Luas 2}$$

$$\text{Luas total} = 0.500 + 0.125$$

$$\text{Luas total} = 0.625 \text{ satuan}$$

Maka, nilai AUC-ROC atau luas dari area dibawah kurva ROC (*Receiver Operating Characteristic*) adalah 0.625.

3.2.6.6.2 Penghitungan dengan sampel dari model prediksi *Decision Tree*

3.2.6.6.2.1 Penghitungan FPR dan TPR berdasarkan batasan yang ditentukan

Berdasarkan dataset sampel, ditentukan 4 batasan klasifikasi, dengan contoh proses klasifikasi seperti berikut

- Batas 0

Pada batas ini, semua nilai positif diklasifikasi seluruhnya secara benar, dan nilai negatif diklasifikasikan seluruhnya secara salah. Hal ini dapat dilihat dengan hasil prediksi menunjukkan semua nilai positif, meskipun label menyatakan nilai negatif. Berikut Tabel 3.69 menunjukkan hasil klasifikasi pada batas 0 menggunakan algoritma *Decision Tree*.

Tabel 3.69 Hasil klasifikasi batas 0 *Decision Tree*

Indeks	Label	Prediksi
0	Negatif	Positif
1	Negatif	Positif
2	Positif	Positif
3	Positif	Positif
4	Positif	Positif
5	Positif	Positif
6	Positif	Positif
7	Positif	Positif
8	Negatif	Positif
9	Negatif	Positif

Hal ini membuat nilai TP = 6, nilai FP = 4, nilai FN = 0, dan nilai TN = 0

Pemetaan titik menggunakan nilai FPR dan TPR, dengan penghitungan TPR merujuk pada rumus 2.14

$$TPR = \frac{6}{(6 + 0)} = 1$$

Penghitungan FPR dihitung merujuk pada rumus 2.15

$$FPR = \frac{4}{(4 + 0)} = 1$$

Maka pemetaan (FPR, TPR) pada batasan ini (1, 1)

- Batas 0.3

Pada batas ini, 30% dari dataset nilai negatif nya diklasifikasikan secara tepat, dengan nilai positif yang diklasifikasikan secara salah, namun 70% lainnya masih mengklasifikasi nilai positif secara tepat, dengan nilai negatif yang diklasifikasi secara salah. Berikut Tabel 3.70 menunjukkan hasil klasifikasi pada batas 0.3 menggunakan algoritma *Decision Tree*.

Tabel 3.70 Hasil klasifikasi batas 0.3 *Decision Tree*

Indeks	Label	Prediksi
0	Negatif	Negatif
1	Negatif	Negatif
2	Positif	Negatif
3	Positif	Positif
4	Positif	Positif
5	Positif	Positif
6	Positif	Positif
7	Positif	Positif
8	Negatif	Positif
9	Negatif	Positif

Hal ini membuat nilai TP = 5, nilai FP = 2, nilai FN = 1, nilai TN = 2

Pemetaan titik menggunakan nilai FPR dan TPR, dengan penghitungan TPR merujuk pada rumus 2.14

$$TPR = \frac{5}{(5 + 1)} = 0.8$$

Penghitungan FPR dihitung merujuk pada rumus 2.15

$$FPR = \frac{2}{(2 + 2)} = 0.5$$

Maka pemetaan (FPR, TPR) pada batasan ini (0.5, 0.8)

- Batas 0.6

Pada batas ini, 60% dari dataset nilai negatif nya diklasifikasikan secara tepat, dengan nilai positif yang diklasifikasikan secara salah namun 40% lainnya masih mengklasifikasi nilai positif secara tepat, dengan nilai negatif yang diklasifikasi secara salah. Berikut Tabel 3.71 menunjukkan hasil klasifikasi pada batas 0.6 menggunakan algoritma *Decision Tree*.

Tabel 3.71 Hasil klasifikasi batas 0.6 *Decision Tree*

Indeks	Label	Prediksi
0	Negatif	Negatif
1	Negatif	Negatif
2	Positif	Negatif
3	Positif	Negatif
4	Positif	Negatif
5	Positif	Negatif
6	Positif	Positif
7	Positif	Positif
8	Negatif	Positif
9	Negatif	Positif

Hal ini membuat nilai TP = 2, nilai FP = 2, nilai FN = 4, nilai TN = 2

Pemetaan titik menggunakan nilai FPR dan TPR, dengan penghitungan

TPR merujuk pada rumus 2.14

$$TPR = \frac{2}{(2 + 4)} = 0.3$$

Penghitungan FPR dihitung merujuk pada rumus 2.15

$$FPR = \frac{2}{(2 + 2)} = 0.5$$

Maka pemetaan (FPR, TPR) pada batasan ini (0.5, 0.3)

- Batas 1

Pada batas ini, semua nilai positif diklasifikasi seluruhnya secara salah, dan nilai negatif diklasifikasikan seluruhnya secara benar. Hal ini dapat dilihat dengan hasil prediksi menunjukkan semua nilai positif, meskipun label menyatakan nilai negatif. Berikut Tabel 3.72 menunjukkan hasil klasifikasi pada batas 1 menggunakan algoritma *Decision Tree*.

Tabel 3.72 Hasil klasifikasi batas 1 *Decision Tree*

Indeks	Label	Prediksi
0	Negatif	Negatif
1	Negatif	Negatif
2	Positif	Negatif
3	Positif	Negatif
4	Positif	Negatif
5	Positif	Negatif
6	Positif	Negatif
7	Positif	Negatif
8	Negatif	Negatif
9	Negatif	Negatif

Hal ini membuat nilai TP = 0, nilai FP = 0, nilai FN = 6, dan nilai TN = 4

Pemetaan titik menggunakan nilai FPR dan TPR, dengan penghitungan TPR merujuk pada rumus 2.14

$$TPR = \frac{0}{(0 + 6)} = 0$$

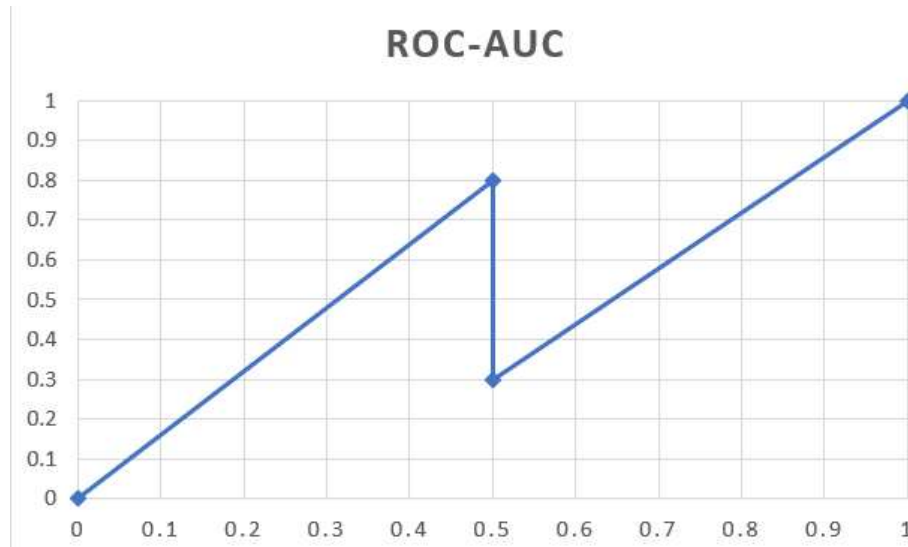
Penghitungan FPR dihitung merujuk pada rumus 2.15

$$FPR = \frac{0}{(0 + 4)} = 0$$

Maka pemetaan (FPR, TPR) pada batasan ini (0, 0)

3.2.6.6.2.2 Pemetaan titik TPR dan FPR

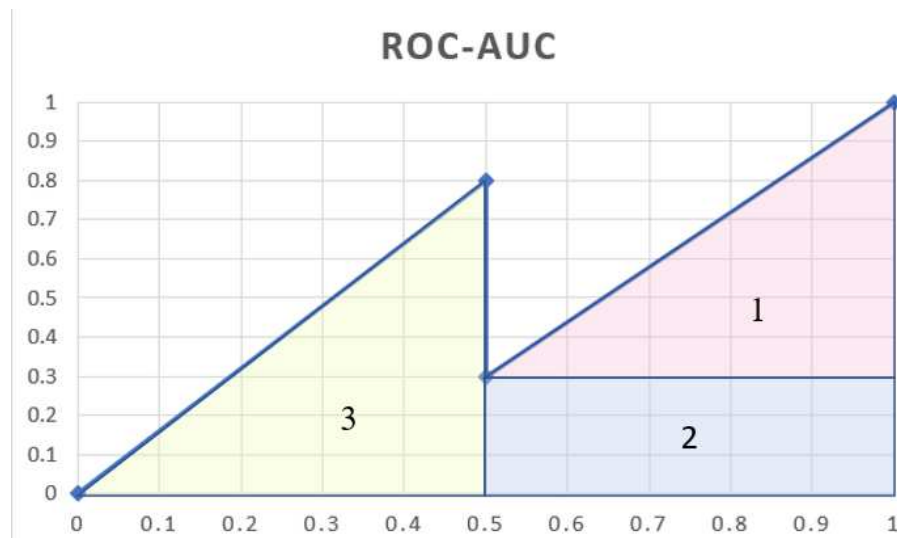
Berdasarkan penghitungan FPR dan TPR, diketahui himpunan (TPR, FPR), yakni berisikan {(1, 1), (0.5, 0.8), (0.3, 0.5), (0, 0)}. Pembuatan kurva ROC dapat dilakukan dengan melakukan pemetaan terhadap himpunan tersebut, dengan hasil seperti Gambar 3.6 berikut.



Gambar 3.6 Kurva ROC model prediksi *Decision Tree*

3.2.6.6.2.3 Penghitungan AUC-ROC

Penghitungan nilai AUC-ROC dapat ditentukan dengan menghitung luas dari area dibawah kurva ROC, perhitungan luas digambarkan dengan Gambar 3.7 berikut.



Gambar 3.7 Penghitungan luas area dibawah kurva ROC *Decision Tree*

Kurva membentuk 3 bangun, dimana bangun 1 berupa segitiga siku-siku dengan alas persegi panjang 0.5 satuan dan tinggi sebesar 0.7 satuan. Bangun 2 merupakan persegi panjang dengan panjang 0.5 satuan dan lebar 0.3 satuan.

Bangun 3 berupa segitiga siku-siku dengan alas 0.5 satuan dan tinggi 0.8 satuan.

Penghitungan luas bangun 1, seperti berikut.

$$\text{Luas 1} = (0.5) * \text{alas} * \text{tinggi}$$

$$\text{Luas 1} = (0.5) * 0.5 * 0.7$$

$$\text{Luas 1} = 0.175 \text{ satuan}$$

Penghitungan luas bangun 2 seperti berikut.

$$\text{Luas 2} = \text{panjang} * \text{lebar}$$

$$\text{Luas 2} = 0.5 * 0.3$$

$$\text{Luas 1} = 0.15 \text{ satuan}$$

Selanjutnya penghitungan luas bangun 3 seperti berikut.

$$\text{Luas 2} = (0.5) * \text{alas} * \text{tinggi}$$

$$\text{Luas 2} = (0.5) * 0.5 * 0.8$$

$$\text{Luas 2} = 0.2 \text{ satuan}$$

Total luas seluruh bangun dapat ditentukan dengan menjumlahkan luas bangun 1 dengan luas bangun 2.

$$\text{Luas total} = \text{Luas 1} + \text{Luas 2} + \text{Luas 3}$$

$$\text{Luas total} = 0.175 + 0.150 + 0.200$$

$$\text{Luas total} = 0.525 \text{ satuan}$$

Maka, nilai AUC-ROC atau luas dari area dibawah kurva ROC (*Receiver Operating Characteristic*) adalah 0.525.

3.2.6.7 Perbandingan Hasil Evaluasi

Berikut Tabel 3.73 menjelaskan tentang perbandingan antara nilai akurasi, presisi, *recall*, *F1 Score*, dan nilai AUC ROC antara algoritma *Naïve Bayes* dan *Decision Tree*.

Tabel 3.73 Perbandingan hasil evaluasi performa

<u>Evaluasi Performa</u>	<u>Algoritma Naïve Bayes</u>	<u>Algoritma Decision Tree</u>
<u>nilai akurasi</u>	80%	70%
<u>nilai presisi</u>	83.4%	71.4%
<u>nilai recall</u>	83.4%	83.4%
<u>nilai F1 Score</u>	83.4%	76.9%
<u>nilai AUC-ROC</u>	0.625	0.525

Berdasarkan Tabel 3.73, dapat disimpulkan bahwa algoritma *Naïve Bayes* memiliki performa yang lebih unggul dalam melakukan klasifikasi dibandingkan algoritma *Decision Tree*. Baik secara nilai akurasi, presisi, *recall*, *f1-score*, maupun AUC-ROC, algoritma *Naïve Bayes* memberikan hasil yang lebih tinggi dibandingkan dengan *Decision Tree*.

3.2.7 Visualisasi

Terakhir dilakukan visualisasi untuk memberikan penggambaran perbandingan antara kedua algoritma. Teknik visualisasi *bar* dari *library python matplotlib* digunakan, dimana diagram batang akan dibuat dengan nama algoritma yang dibandingkan menjadi sumbu X, sedangkan sumbu Y berisikan skala 0 hingga 1, dengan jarak 0.1. Diagram batang digunakan untuk memperjelas perbedaan antara kedua algoritma. Penjelasan visualisasi dilakukan pada bab 4.