

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

ANALISIS SENTIMEN OPINI MASYARAKAT PADA DRAMA KOREA MENGGUNAKAN METODE NAÏVE BAYES

SKRIPSI



Oleh:

NAMA : Fitriani

NIM : 701190027

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTHAN THAHA
SAIFUDDIN JAMBI
2023**

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

HALAMAN JUDUL

ANALISIS SENTIMEN OPINI MASYARAKAT PADA DRAMA KOREA MENGGUNAKAN METODE NAÏVE BAYES

SKRIPSI



Oleh:

NAMA : Fitriani

NIM : 701190027

**Skripsi ini ditulis untuk memenuhi
Sebagian persyaratan untuk mendapatkan gelar Sarjana**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTHAN THAHA
SAIFUDDIN JAMBI**

2023

ii

HALAMAN PERSETUJUAN



KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SULTHAN THAHA SAIFUDDIN JAMBI
FAKULTAS SAINS DAN TEKNOLOGI
Jalan Lintas Jambi-Muaro Bulian Km. 16 Simpang Sungai Duren Kab. Muaro Jambi
36363 Email: fst.uinjambi.ac.id

NOTA DINAS

Jambi, 12 Juni 2023

Kepada,
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sulthan Thaha Saifuddin di Jambi

Assalamualaikum.wr.wb

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan :

Judul : Analisis Sentimen Opini Masyarakat Pada Drama Korea Menggunakan Metode Naive Bayes
Nama : Fitriani
NIM : 701190027
Jurusan : Sistem Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Sulthan Thaha Saifuddin untuk diujikan dalam Sidang Munaqasyah.

Wassalamualaikum.wr.wb

Pembimbing I,



Utami Mizani Putri, S.T., M.Kom
NIP. 198610162022032001

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
 2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sultha Jambi
 2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sultha Jambi



KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SULTHAN THAHA SAIFUDDIN JAMBI
FAKULTAS SAINS DAN TEKNOLOGI
Jalan Lintas Jambi-Muaro Bulian Km. 16 Simpang Sungai Duren Kab. Muaro Jambi
36363 Email: fst.uinjambi.ac.id

NOTA DINAS

Jambi, 12 Juni 2023

Kepada,
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sulthan Thaha Saifuddin di Jambi

Assalamualaikum.wr.wb

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan :

Judul : Analisis Sentimen Opini Masyarakat Pada Drama Korea Menggunakan Metode Naive Bayes
Nama : Fitriani
NIM : 701190027
Jurusan : Sistem Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Sulthan Thaha Saifuddin untuk diujikan dalam Sidang Munaqasyah.

Wassalamualaikum.wr.wb

Pembimbing 11



Mhd. Theo Ari Bangsa, S.Kom., M.Sc
NIDN. 2007089205

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sultha Jambi
 2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sultha Jambi

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI
SULTHAN THAHA SAIFUDDIN JAMBI
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Lintas Jambi Ma. Bulian Km. 16 Sei. Duren Kec. Jaluko Kab. Muaro Jambi 36363
Telp/Fax : (0741) 584118 Web : <https://uinjambi.ac.id> email : mail@uinjambi.ac.id

PENGESAHAN SKRIPSI / TUGAS AKHIR Nomor : B- /D.VII/PP.009/08/2023

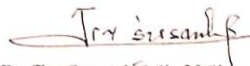
Skrripsi/ Tugas Akhir dengan Judul : "ANALISIS SENTIMEN OPINI MASYARAKAT PADA DRAMA KOREA MENGGUNAKAN METODE NAÏVE BAYES"

Yang dipersiapkan dan disusun oleh

Nama : Fitriani
NIM : 701190027
Telah dimunaqasahkan pada : 26 Juni 2023
Nilai Munaqasah : 85,46

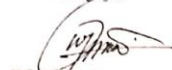
Dan dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sulthan Thaha Saifuddin Jambi.

TIM MUNAQASAH :
Ketua Sidang



Dr. Try Susanti, S.Si., M.Si
NIP. 197603032005012005

Penguji I



Efitra, M. Kom
NIP. 199112262019031013

Penguji II



Yerix Karandiani, M. Kom
NIDN. 2014068504

Pembimbing I



Utami Mizan Putri, M. Kom
NIP. 198610162022032001

Pembimbing II



Mhd. Theo Ari Bangsa, M. Sc
NIDN. 2007089205

Sekretaris Sidang



Mhd. Theo Ari Bangsa, M. Sc
NIDN. 2007089205

Jambi, 08-08- 2023

Dekan Fakultas Sains dan Teknologi



Iskandar K. S. A., M.Pd., M.S.I., M.H., Ph.D
NIP.197512242009121001

PERSEMBAHAN

Skripsi ini kupersembahkan kepada kedua orang tuaku yang tercinta yang telah mengasuhku mulai dari lahir hingga dewasa sekarang ini, semoga kedua orang tuaku selalu mendapat rahmat dari Allah Subhanahuwa Ta'ala, Amin. Dan terima kasih juga kepada Saudara/i kandung yang tercinta, Terimakasih atas dukungan dan do'anya sehingga saya dapat menyelesaikan studi pendidikan di perguruan tinggi ini, sahabat-sahabat seperjuangan kukhususnya Program Studi Sistem Informasi, umumnya Fakultas Sains dan Teknologi UIN Sulthan Thaha Saifuddin Jambi serta orang-orang yang mencintai ilmu pengetahuan. Dan terima kasih juga kepada para bapak ibu dosen khususnya dosen pembimbing satu dan dosen pembimbing dua. Terimakasih untuk semua yang telah membantuku dalam penyelesaian skripsi ini, semoga Allah SubhanahuwaTa'ala selalu memberikan taufiq dan hidayah kepada kita semua. Amin yaa Robbal 'Alamin.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sulthan Thaha Saifuddin Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sulthan Thaha Saifuddin Jambi

Moto :

“Adakah sama orang-orang yang mengetahui dengan orang-orang yang tidak mengetahui? Sesungguhnya orang yang berakallah yang dapat menerima pelajaran” (Q.S Az-Zumar ayat 9)

“Hidup adalah pilihan, maka aku memilih untuk memperjuangkan dan mendapatkan semua yang ku inginkan” (Fitriani)

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sulthan Thaha Saifuddin Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sulthan Thaha Saifuddin Jambi

HALAMAN KEASLIAN NASKAH

PERNYATAAN KEASLIAN NASKAH

Yang bertandatangan dibawah ini :

Nama : Fitriani
Nim : 701190027
Jurusan : Sistem Informasi

Menyatakan bahwa skripsi yang berjudul :

ANALISIS SENTIMENM OPINI MASYARAKAT PADA DRAMA KOREA MENGUNAKAN METODE NAIVE BAYES

Merupakan Skripsi yang saya susun sebagai syarat untuk memperoleh gelar Sarjana dari Fakultas Sains dan Teknologi Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi adalah hasil karya sendiri.

Adapun bagian-bagian tertentu dalam penulisan skripsi yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah dan etika penulisan ilmiah.

Apabila dikemudian hari ditemukan seluruh atau sebagian skripsi bukan hasil karya sendiri menerima sanksi sesuai dengan peraturan dan perundang-undangan yang berlaku.

Jambi, 2 Agustus 2023



Fitriani
NIM.701190027

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
 2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

KATA PENGANTAR

Puji dan Syukur Penulis panjatkan kepada Allah Subhanahuwa Ta'ala dan salam atas Nabi Muhammad Shalallahu 'alaihiwasallam, pembawa risalah pencerahan bagi manusia.

Penulisan skripsi ini dimaksudkan untuk memenuhi salah satu syarat akademik guna mendapatkan gelar Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Sains dan Teknologi UIN Sulthan Thaha Saifuddin Jambi.

Dalam penyusunan skripsi ini banyak ditemukan kesulitan-kesulitan dan hambatan-hambatan, namun berkat inayah Allah SWT, serta bantuan dari berbagai pihak segala kesulitan dan hambatan tersebut dapat diatasi, sehingga skripsi ini dapat diselesaikan. Untuk itu, penulis mengucapkan terima kasih kepada yang terhormat:

1. Bapak Prof. DR. H. Su'aidi, MA.Ph.D. selaku Rektor UIN Sulthan Thaha Saifuddin Jambi
2. Bapak Iskandar, S.Pd., M.Pd., M.SI., Ph.D. Selaku Dekan Fakultas Sains dan Teknologi UIN Sulthan Thaha Saifuddin Jambi.
3. Ibu Dr. Try Susanti, S.Si., M.Si. selaku Ketua Program Studi Sistem Informasi Fakultas Sains dan Teknologi.
4. Ibuk Utami Mizani Putri, S.T., M.Si Selaku Pembimbing I dan Bapak M.Theo Ari Bangsa, M. Cs. Selaku Pembimbing II.
5. Bapak-bapak dan ibu-ibu dosen serta staf Fakultas Sains dan Teknologi UIN Sulthan Thaha Saifuddin Jambi.
6. Ayah, Ibu, dan saudara-saudaraku yang telah mendukung dan memberikan motivasi.
7. Rekan-rekan seperjuangan Angkatan 2019 di Fakultas sains dan teknologi Sulthan Thaha Saifuddin Jambi.
8. Almamaterku.
9. Seluruh member BTS yaitu Kim Namjoon, Kim Seokjin, Min Yoongi,

Jung Hoseok, Park Jimin, Kim Taehyung, Jeon Jungkook yang secara tidak langsung Menjadi healing terbaik dalam pembuatan skripsi.

Akhirnya kepada semua pihak, penulis sertakan do'a semoga Allah SWT membalas pahala kebaikan yang telah diberikan agar berlipat ganda dan berkenan untuk mengabulkannya, amin ya rabbal 'alamin. Semoga skripsi ini akan memberikan banyak manfaat bagi yang membacanya.

Jambi, 2023
Penulis

Fitriani
701190027

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
 2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

ABSTRAK

Drama korea terus berkembang dan digemari oleh banyak kalangan dengan menyajikan berbagai *gendre* seperti *gendre action*. Terdapat banyak opini masyarakat mengenai topik ini di berbagai media sosial seperti twitter. Penelitian ini dilakukan untuk mengetahui opini masyarakat dan mengetahui bahwa metode naïve bayes dapat digunakan dalam mengklasifikasi analisis sentiment drama korea *gendre action*. Pengumpulan data dilakukan dengan *crawling* menggunakan kata kunci *Taxi Driver, Weak Hero Class, Vincenzo, The Uncanny Counter, Duty After School* pada twitter sebanyak 2,250 data dengan rentan waktu 2021-2023 setelah melauai tahap cleansing dan pre prosesing pada aplikasi rapidminer, data yang tersisa 1,815. Dalam pengujian menggunakan cross validation jumlah data latih yang berbeda yaitu 1200 dan 1300. Pengujian pertama dengan data latih 1200 didapatkan akurasi 78,95%, sedangkan dengan data latih 1300 didapatkan akurasi sebesar 79,12%. Peneliti juga melakukan pengujian secara manual menggunakan *confution matrix* dengan perbandingan data 65% : 35% yaitu menggunakan data latih 1,176 dan data model 639 didapatkan akurasi 80,12% serta pengujian manual ke dua menggunakan data latih 1,302 dan data model 513 dengan perbandingan 72% : 28% mendapatkan akurasi sebesar 85,76%. Sehingga menunjukkan bahwa semakin besar data latih maka semakin besar pula akurasi yang akan di dapatkan. Pengujian selanjutnya menggunakan *split validation* dengan perbandingan data 50% : 50% didapatkan akurasi sebesar 75,63% dan dilakukan pengujian dengan perbandingan data 70% : 30% didapatkan hasil akurasi sebesar 77,12. Peneliti juga melakukan pengujian dengan menambahkan data baru untuk melihat hasil prediksi, dari 10 data baru didapatkan 6 benar dan 4 data salah. Dari jumlah data keseluruhan didapatkan komentar positif lebih besar yaitu 71% dibandingkan komentar negatif hanya 29%, menunjukkan bahwa sentimen positif lebih cenderung dibandingkan sentimen negatif, sehingga drakor *gendre action* lebih banyak diminati oleh masyarakat dalam jangka waktu 2021-2023.

Kata kunci : Drama Korea, Data Mining, Naïve Bayes Clasifier

ABSTRACT

Korean drama continues to grow and is loved by many groups by presenting various genres, such as the action genre. There are many public opinions on this topic on various social media sites, such as Twitter. This research was conducted to find out public opinion and to know that the Nave Bayes method can be used in classifying the sentiment analysis of Korean drama genre action. Data collection was carried out by crawling using the keywords Taxi Driver, Weak Hero Class, Vincenzo, The Uncanny Counter, and Duty After School on Twitter. There were as many as 2,250 records with a timeframe of 2021–2023. After going through the cleansing and pre-processing stages on the RapidMiner application, the remaining data was 1,815 records. In testing using cross-validation, the number of different training data points is between 1200 and 1300. The first test with 1200 training data obtained an accuracy of 78.95%, while the second test with 1300 training data obtained an accuracy of 79.12%. The researcher also conducted manual testing using the confusion matrix with a data ratio of 65%:35%, using 1.176 training data and 639 model data to obtain an accuracy of 80.12%, and a second manual test using 1.302 training data and 513 model data with a 72% ratio:28% to obtain an accuracy of 85.76%. So it shows that the greater the training data, the greater the accuracy that will be obtained. Subsequent tests using split validation with a data comparison of 50%: 50% obtained an accuracy of 75.63%, and testing was carried out with a data comparison of 70%: 30% to obtain an accuracy of 77.12. The researcher also tested by adding new data to see the prediction results; out of 10 new data points, 6 were correct and 4 were incorrect. From the total amount of data, positive comments were greater, namely 71% compared to only 29% negative comments, indicating that positive sentiments are more likely than negative sentiments, so that the genre action drama is more in demand by the public in the 2021–2023 period.

Keywords : *Korean drama, data mining, naïve bayes classifier*

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
PERSEMBAHAN.....	iv
MOTO.....	v
HALAMAN KEASLIAN NASKAH.....	vi
KATA PENGANTAR.....	viii
ABSTRAK.....	x
ABSTRACT.....	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL.....	iv
DAFTAR LAMPIRAN.....	v
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB II.....	5
LANDASAN TEORI.....	5
2.1 Tinjauan Pustaka.....	5
2.1.1 Analisis Sentimen.....	5
2.1.2 Drama Korea.....	5
2.1.3 Naive Bayes.....	6
2.1.4 Teks Mining.....	7

2.1.5 Rapid Miner	7
2.1.6 Confution Matrix.....	8
2.1.7 Ekstraksi Fitur TF-IDF.....	9
2.1.8 Naive Bayes Classifier	11
a. Menentukan probabilitas class	12
b. Menguji hasil klasifikasi NBC	12
c. Menentukan probabilitas kategori	14
2.2 Penelitian Terdahulu	15
BAB III	17
METODOLOGI PENELITIAN.....	17
3.1 Metode Pengumpulan Data	17
3.2 Tahapan Penelitian	18
3.3 Pengambilan Data Pada Twitter (Crawling)	18
3.4 Preprocessing Data	20
a. Cleansing Retweet.....	21
b. Cleansing Url	22
c. Cleansing Hastag.....	23
d. Cleansing Simbol	24
e. Filter Example	25
f. Remove Duplicate.....	26
3.5 Pelabelan Data Manual	27
3.6 Klasifikasi Metode Naïve Bayes.....	28
3.7 Ekstraksi Fitur TF-IDF.....	29
a. Tokenize	30
b. Case Folding	31
c. Stopword	32
d. Filter Token.....	33
3.8 Pemodelan Metode Naïve Bayes	33
3.9 Pengujian Dan Evaluasi	34
a. Cross Validation	34

b. Split Validation	35
3.10 Confution matrix	35
a. Akurasi	35
b. Presissi	36
c. Recall.....	36
3.16 Alat dan Bahan	37
3.16.1 Alat yang digunakan dalam penelitian.....	37
3.16.2 Bahan yang digunakan dalam penelitian.....	37
BAB IV	38
HASIL DAN PEMBAHASAN.....	38
4.1 Ekstraksi Fitur TF-IDF.....	38
a. Tokenisasi.....	38
b. Pembobotan Kata	39
c. Worcloud	40
4.2 Hasil Pemodelan Menggunakan Metode NBC	41
a. Pengujian Data Latih 1200	41
b. Pengujian data latih 1300	42
c. Perhitungan manual	42
4.3 split validation.....	45
4.4 Hasil percobaan data baru	46
4.5 Pembahasan Dan Evaluasi	46
BAB V	48
KESIMPULAN DAN SARAN.....	49
5.1 Kesimpulan	49
5.2 Saran.....	49
DAFTAR PUSTAKA	51
LAMPIRAN-LAMPIRAN.....	54

DAFTAR TABEL

Tabel 2.1 confusion matrix.....	8
Tabel 2.2 perhitungan manual TF IDF.....	10
Tabel 2.3 contoh perhitungan manual NBC.....	11
Tabel 2.4 Probabilitas.....	13
Tabel 2.5 Hasil Pengujian Klasifikasi NBC.....	14
Tabel 2.6 Tinjauan pustaka	15
Tabel 3.1 data crawling	19
Tabel 3.2 Cleansing retweet	21
Tabel 3.3 Cleansing URL.....	22
Tabel 3.4 Cleansing hastag	23
Tabel 3.5 Cleansing simbol.....	23
Tabel 3.6 Pelabelan Data.....	25
Tabel 3.7 Tokenize.....	27
Tabel 3.8 Case Folding	30
Tabel 3.9 Stopword	31
Tabel 4.1 perhitungan manual 65% : 35%	33
Tabel 4.2 perhitungan manual 72% : 28%	42
Tabel 4.3 Perbandingan Pengujian Data	44
Tabel 4.4 Percobaan tambah data baru	46

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambai
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambai

DAFTAR GAMBAR

Gambar 3.1 pengambilan token	17
Gambar 3.2 Tahapan Penelitian	18
Gambar 3.3 proses crawling data	19
Gambar 3.4 Prosessing Data	21
Gambar 3.5 Remove Duplicate	26
Gambar 3.6 Klasifikasi Metode Naïve Bayes	29
Gambar 3.7 Prosessing Data	30
Gambar 3.8 Pemodelan Metode Naïve Bayes.....	34
Gambar 3.9 Pengujian Dan Evaluasi	35
Gambar 4.1 Hasil TF-IDF	38
Gambar 4.2 tokenisasi	39
Gambar 4.3 Pembobotan Kata	39
Gambar 4.4 wordcloud.....	40
Gambar 4.5 proses dokumen NBC	41
Gambar 4.6 pengujian 1200 data latih	41
Gambar 4.7 pengujian 1300 data latih	42
Gambar 4.8 pengujian data 50% : 50%	45
Gambar 4.9 pengujian data 70% : 30%	45
Gambar 4.10 diagram hasil analisis sentimen.....	47

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

DAFTAR LAMPIRAN

Lampiran 1. Tampilan awal rapidminer	54
Lampiran 2. Akses token	54
Lampiran 3. Data hasil proses	55
Lampiran 4. Proses pengolahan data.....	55
Lampiran 5. proses penggabungan keseluruhan data crawling.....	56
Lampiran 6. Hasil penggabungan data.....	56
Lampiran 7. Hasil cleansing.....	57
Lampiran 8. Pelabelan manual di excel	57
Lampiran 9. Data latih.....	58
Lampiran 10. Pelabelan otomatis data model	58
Lampiran 11. Proses wourdcloud.....	59
Lampiran 12. Sentimen manual 639 data.....	59
Lampiran 13. Sentimen manual 513 data.....	60
Lampiran 14. Hasil prediksi penambahan data baru	60
Lampiran 15. Kalimat stopword.....	61

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

DAFTAR RIWAYAT HIDUP



A. Identitas Diri

Nama : Fitriani
Jenis Kelamin : Perempuan
Tempat/Tanggal Lahir : Sebatik 12 September 2021
Alamat : Mayang Mangurai, Perumahan Harjita Village
Email : fitriani612@gmail.com
Hp : 085609031465

B. Riwayat Pendidikan

1. SDN 69/X Nipah Panjang Tahun 2013
2. MTSN Kuala Tungkal 1 Nipah Panjang Tahun 2016
3. SMAN 3 Tanjung Jabung Timur Tahun 2019

Jambi, 11 Agustus 2023

Fitriani

NIM. 701190027

Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Drama Korea telah dikenal luas berkat daya cipta dan imajinasi para senimannya, yang terus berkembang. Drama Korea menarik inspirasi dari masalah sehari-hari, kehidupan, dan budaya, serta sangat digemari oleh masyarakat hampir di seluruh Asia bahkan seluruh dunia, baik di kalangan remaja maupun dewasa, terutama di kalangan kaum hawa. Dengan menyajikan tema dan *Gendre* yang beragam (Herawati et al., 2021). Penggemar drama korea di Indonesia memiliki kenaikan di beberapa tahun terakhir begitupula dengan perkembangan pada industri drakor, sehingga dapat mendorong pertumbuhan ekonomi kreatif di Indonesia seperti yang dilakukan oleh beberapa produser Indonesia yang tertarik dan membuat drakor dengan berbagai tema maupun gendre seperti cinta pertama dan *hangout* (Dicky, 2023). Menurut Antinasari pada penelitiannya mengatakan bahwa review film dapat digunakan sebagai alat untuk menilai film. Masyarakat bisa *me-review* film yang akan mereka tonton dan industri film bisa memanfaatkannya untuk mengetahui bagaimana tanggapan masyarakat terhadap film-film yang mereka buat (Antinasari et al., 2017). Dengan begitu terdapat banyak opini masyarakat pada berbagai platform internet yang sulit untuk di baca secara manual satu per satu, maka di perlukan analisis sentiment dengan data yang akan di ambil pada aplikasi twitter (Ratnawati, 2018).

Twitter adalah situs media sosial yang dirancang untuk memfasilitasi pengiriman pesan cepat antara teman, keluarga, dan rekan kerja. Aktivitas penggunaan Twitter ini juga dapat mencakup teks, gambar, video, dan tautan. (Yahyadi & Latifah, 2022). Twitter memudahkan dalam melakukan analisis dikarenakan maksimal 140 kata sehingga kebanyakan opini langsung masuk ke intinya juga memudahkan untuk mendapatkan nilai akurasi yang tinggi (Bhatia et al., 2018). Kemudian data yang di ambil dari twitter akan dilakukan analisis sentimen untuk menentukan apakah seseorang memiliki kesan positif atau negatif terhadap sesuatu dengan melihat pemikiran atau pendapatnya. (Gifari et al., 2022).

Analisis sentimen dikenal sebagai penambangan opini yang digunakan untuk meneliti bagaimana orang-orang menilai suatu layanan, peristiwa, isu ataupun suatu individu dan masih banyak lagi yang lainnya, analisis sentiment memiliki karakteristik yang subyektif sehingga kumpulan pendapat orang-orang perlu dianalisis karena suatu ulasan mengandung beragam opini dari sudut pandang yang berbeda. (Bhatia et al., 2018). Analisis sentiment juga digunakan untuk memperoleh berbagai sumber data dari twitter dan untuk mendapatkan sudut pandang dari pengguna, selain itu analisis sentimen juga untuk mengetahui, mengekstrak, dan menjalankan informasi tekstual secara otomatis sehingga menemui keterangan sentimental dalam mengungkapkan pemikiran, (Gifari et al., 2022). Maka peneliti memilih metode Naive Bayes Classifier (NBC)

Dalam menganalisis sentiment masyarakat, peneliti memilih metode naive bayes dikarenakan metode tersebut memiliki performa yang lebih baik. Metode Naive Bayes Classifier merupakan pengklasifikasi probabilistik langsung

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

berdasarkan penerapan teorema Bayes (dari statistik Bayesian) dengan asumsi independen yang kuat (naif) (Sinaga, 2017). Naïve bayes clasification sering digunakan dalam analisis sentimen dan text mining karena kecepatan dan kesederhanaannya dalam mengolah data, cocok digunakan pada jumlah data yang besar. (Krisdiyanto, 2021) Penelitian yang dilakukan oleh (Devita et al., 2018) membuktikan bahwa dengan data latih/*training* yang sedikit dapat menghasilkan akurasi yang maksimal pada metode naïve bayes, dengan 40 data uji menghasilkan akurasi 70% sedangkan metode KNN hanya mendapatkan akurasi 40%. Akan tetapi semakin tinggi akurasi yang dihasilkan maka semakin banyak pula data *training* yang digunakan, dengan kata lain jumlah data *training* sangat berpengaruh terhadap kinerja sistem dalam mendapatkan akurasi. (Ratnawati, 2018). Dalam penelitian ini menggunakan jumlah data latih yang lebih besar dibandingkan data model sehingga mendapatkan akurasi yang maksimal.

Berdasarkan uraian diatas, maka penulis tertarik melakukan penelitian tugas akhir dengan judul penelitian “**Analisis Sentiment Opini Masyarakat Pada Drama Korea Menggunakan Metode Naïve Bayes.**”

1.2 Rumusan Masalah

Bagaimana analisis sentiment mengukur opini masyarakat mengenai drama korea?

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

1. Data yang diteliti hanya data berbahasa Indonesia pada twitter dengan *keyword* : Taxi Driver, Weak Hero Class, Vincenzo, The Uncanny Counter, Duty After School
2. Metode yang digunakan adalah naïve bayes classifier
3. Data yang digunakan hanya 1,815 data pada tahun 2021-2023

1.4 Tujuan Penelitian

Untuk mengetahui opini masyarakat dan mengetahui bahwa metode naïve bayes dapat digunakan dalam mengklasifikasi analisis sentiment pada twitter studi kasus drama korea *gendre action*

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberi manfaat berupa

1. Memberikan informasi sentiment positif maupun negatif pada drama korea.
2. Mengetahui performa dari metode naïve bayes classifier dalam melakukan analisis sentiment terhadap drama korea.
3. Menjadi rekomendasi untuk produser dalam membuat film bergendre kriminal dan *action*.
4. Menjadi referensi bagi penonton.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 Analisis Sentimen

Analisis sentimen adalah teknik penggalian, pemahaman, dan pengolahan data secara otomatis dalam bentuk teks yang tidak terstruktur untuk menghasilkan data sentimen dari sebuah kalimat opini atau pernyataan. (Arsi & Waluyo, 2021)

Analisis sentimen, sering dikenal sebagai penambangan opini, adalah cabang studi yang meneliti bagaimana perasaan orang tentang berbagai objek, termasuk barang dan jasa, bisnis tertentu, kandidat politik, kepedulian sosial, peristiwa terkini, dan subjek, serta karakteristiknya. Meskipun analisis sentimen lebih umum digunakan, penambangan opini dan analisis sentimen sering digunakan di kalangan akademisi. Karena fakta bahwa, untuk pertama kalinya dalam sejarah manusia, sejumlah besar data opini telah ditangkap dalam bentuk digital, kemunculan dan perluasan bidang ini sejalan dengan media sosial di Web, seperti blog, blog mikro, Twitter, dan jejaring sosial. Salah satu bidang studi yang paling aktif dalam pemrosesan bahasa alami sejak awal tahun 2000-an adalah analisis sentimen. (Bhatia et al., 2018)

2.1.2 Drama Korea

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

Drama Korea adalah serial drama yang dibuat dan diproduksi di negara Korea dan disiarkan di beberapa Negara. Dengan menyuguhkan berbagai macam cerita yang berbeda-beda menjadi daya tarik tersendiri bagi para penontonnya. Ini menunjukkan seberapa cepat Korea, khususnya Korea Selatan, berkembang di industri hiburan. Drama Korea ini disajikan dengan episode biasanya berkisar antara 16 hingga 32. Drama Korea biasanya berdurasi 40 hingga 1 jam dengan mencakup berbagai genre, termasuk misteri, komedi, aksi, dan cinta. alur cerita yang menarik akan membuat penonton tertarik dan memungkinkan mereka untuk menikmati sepenuhnya hubungan dramatis antara para *actor*. Aplikasi yang dirancang untuk mengakses drama Korea, seperti drakor.id, viu, netflix, iflix, hoox, dan lainnya, membuat K-Drama terlalu mudah diakses. (Aprillia & Listiani, 2019)

2.1.3 Naive Bayes

Naive Bayes cepat dan akurat. Dalam formulasi naive bayes, kelas dokumen ditentukan oleh kata dan frekuensi kemunculannya. Metode langsung untuk mengungkapkan, menggunakan, dan memperoleh informasi probabilistik disediakan oleh naive bayes. Itu dapat menghasilkan hasil yang luar biasa. Pada banyak kumpulan data, orang sering menemukan bahwa naive bayes bersaing dan terkadang bahkan mengalahkan pengklasifikasi yang lebih canggih. (Džeroski, 2008)

Teks paling sering dikategorikan menggunakan kategorisasi tekstual NBC karena cepat dan mudah digunakan. Karena kecepatannya yang luar biasa, lebih

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

baik memilih pengklasifikasi Naive Bayes (NBC). Salah satu algoritma yang banyak digunakan dalam proses analisis sentimen adalah metode klasifikasi naive bayes. Oleh karena itu, lebih tepat menggunakan algoritma Naive Bayes Multinomial Classifier dalam penyelidikan ini. lebih akurat daripada pendekatan lain dan memproses data dengan cepat.(Informatika, 2022)

2.1.4 Teks Mining

Tujuan dari text mining adalah untuk mengekstrak data dari kumpulan dokumen. Melalui pemrosesan, kategorisasi, dan analisis sejumlah besar data tidak terstruktur, penambangan teks mampu menghasilkan informasi. Untuk mengekstraksi informasi yang bermakna dari kumpulan makalah dengan sumber, penambangan teks digunakan. Data teks dengan struktur tidak terstruktur. Analisis perasaan yang menentukan apakah suatu pernyataan positif atau buruk secara emosional dapat dibuat dengan proses pencarian informasi dalam text mining. Dokumen yang tidak terstruktur atau sebagian terstruktur menambang objek teks. Ekstraksi data yang diperlukan secara otomatis dan berhasil dari berbagai dokumen adalah penambangan teks.(Verawardina et al., 2021)

2.1.5 Rapid Miner

Perangkat lunak terbuka disebut Rapid Miner (sumber terbuka). Alat untuk analisis prediktif, penambangan teks, dan penambangan data disebut Rapid Miner. Berbagai metode deskriptif dan preskriptif digunakan oleh Rapid Miner. dalam memberikan pengetahuan kepada pengguna sehingga mereka dapat membuat

keputusan terbesar. Sekitar 500 karyawan data mining bekerja untuk Rapid Miner, termasuk karyawan input, output, preprocessing, dan visualisasi. Rapid Miner adalah perangkat lunak mandiri yang dapat digunakan untuk menganalisis data dan menambangnya untuk digunakan di produk lain. Penggunaan bahasa pemrograman Java memungkinkan Rapid Miner berjalan di semua sistem operasi. (Ardiansyah & Walim, 2018).

2.1.6 Confution Matrix

Confution matrix merupakan metode *evaluasi* perhitungan *akurasi* pada data mining yang berisi informasi klasifikasi yang dipredkiksi oleh sebuah sistem. Cara kerja *confution matrix* yaitu dengan membandingkan hasil prediksi nilai klasifikasi dengan data asli. Dibutuhkan ukuran untuk mengetahui seberapa tepat akurasi hasil prediksi dari sistem. (Hendrastuty et al., 2021). Dalam hasil klasifikasi confutin matrix terdapat 4 istilah yaitu seperti tabel dibawah ini:

Tabel 2.1 confution matrix

Predicted Value	Actual Value	
	TP (True Positive)	FP (False Positive)
	FN (False Negative)	TN (True Negative)

Keterangan :

1. True Positive (TP)

Hasil prediksi menyatakn positif dan hasil yang sebenarnya juga positif

2. True Negative (TN)

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi

@ Hak cipta milik UINSutha Jambi



UNIVERSITAS ISLAM NEGERI
SULTHAN THAHA SAIFUDDIN
J A M B I

Hasil prediksi menyatakan negatif dan hasil yang sebenarnya juga menyatakan negatif

3. False Positive (FP)

Hasil prediksi menyatakan positif akan tetapi hasil yang sesungguhnya menyatakan negatif

4. False Negative (FN)

Hasil prediksi menyatakan negatif akan tetapi hasil yang sesungguhnya menyatakan positif

2.1.7 Ekstraksi Fitur TF-IDF

Proses ekstraksi fitur melibatkan pemberian nomor numerik untuk setiap frekuensi kemunculan kata dalam dokumen teks. Pendekatan pembobotan kata salah satunya disebut TF-IDF (*Term Frequency-Inverse Document Frequency*), yang menggabungkan *term Frequency* dan *Document Frequency* menjadi satu metode pembobotan. Algoritma TF-IDF adalah suatu algoritma yang berdasarkan nilai statistik menunjukkan kemunculan suatu kata di dalam dokumen. TF (*Term Frequency*) menyatakan banyaknya suatu kata muncul dalam sebuah dokumen. DF (*Document Frequency*) menyatakan banyaknya dokumen yang mengandung suatu kata dalam satu segmen publikasi

$$IDF = \log \left(\frac{N}{DF(w)} \right)$$

$$TF - IDF (w,d) = TF (w,d) \times IDF (w)$$

Keterangan:

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi



Hak Cipta Dilindungi Undang-Undang:
 1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
 2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

TF-IDF(w,d) : bobot suatu kata dalam keseluruhan dokumen

W : suatu kata (word)

D : suatu dokumen (document)

TF(w,d) : frekuensi kemunculan sebuah kata w dalam dokumen d

IDF(w) : inverse DF dari kata w

N : jumlah keseluruhan dokumen

DF(w) : jumlah dokumen yang mengandung kata w

Berikut merupakan contoh perhitungan manual TF IDF

Kata kunci : ilmu semantik

Dokumen 1(d 1) : program logika dan semantik

Dokumen 2 (d 2) : ilmu antar individu

Dokumen 3 (d 3) : dalam program ilmu terdapat transfer ilmu semantik

Tabel 2.2 perhitungan manual TF IDF

Token	Kk	Tf			df	D/df	IDF(log D/df)	W			
		D1	D2	D3				Kk	D1	D2	D3
Program	0	1	0	1	2	1,5	0,176	0	0,176	0	0,176
Logika	0	1	0	0	1	3	0,447	0	0,447	0	0
Semantik	1	1	0	1	2	1,5	0,176	0,176	0,176	0	0,176
Ilmu	1	0	1	2	2	1,5	0,176	0,176	0	0,176	0,352
Individu	0	0	1	0	1	1,5	0,447	0	0	0,447	0
Transfer	0	0	0	1	1	3	0,447	0	0	0	0,447

Mencari perangkian : D1 : 0,176

D2 : 0,176

D3 : 0,528

2.1.8 Naive Bayes Classifier

Langkah berikut melibatkan klasifikasi data menggunakan algoritma Naïve Bayes. Proses klasifikasi Nave Bayes adalah sebagai berikut:

Rumus Naive Bayes Classifier (NBC) :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Keterangan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Tabel 2.3 contoh perhitungan manual naïve bayes

No	Cuaca	Suhu	Kelembapan	Berangin	Main
1.	Cerah	Panas	Tinggi	Salah	Tidak
2.	Berawan	Panas	Tinggi	Benar	Tidak
3.	Hujan	Panas	Tinggi	Salah	Ya
4.	Hujan	Sejuk	Tinggi	Salah	Ya
5.	Hujan	Dingin	Normal	Salah	Ya
6.	Hujan	Dingin	Normal	Benar	Ya
7.	Berawan	Dingin	Normal	Benar	Ya
8.	Cerah	Sejuk	Tinggi	Salah	Tidak
9.	Cerah	Dingin	Normal	Salah	Ya
10.	Hujan	Sejuk	Normal	Salah	Ya

11.	Cerah	Sejuk	Normal	Benar	Ya
12.	Berawan	Sejuk	Tinggi	Benar	Ya
13.	Berawan	Panas	Normal	Salah	Ya
14.	Hujan	Sejuk	Tinggi	Benar	Tidak

a. Menentukan probabilitas class

$$Ya \rightarrow \frac{10}{14} = 0,71$$

$$Tidak \rightarrow \frac{4}{14} = 0,29$$

b. Menentukan probabilitas kategori

Ya

$$p(\text{cuaca=cerah} \mid ya) = 2/10 = 0,2$$

$$P(\text{cuaca=berawan} \mid ya) = 4/10 = 0,4$$

$$P(\text{cuaca hujan} \mid ya) = 4/10 = 0,4$$

$$P(\text{suhu panas} \mid ya) = 2/10 = 0,2$$

$$P(\text{suhu sejuk} \mid ya) = 4/10 = 0,4$$

$$P(\text{suhu dingin} \mid ya) = 4/10 = 0,4$$

$$P(\text{kelembaban tinggi} \mid ya) = 3/10 = 0,3$$

$$P(\text{kelembaban normal} \mid ya) = 7/10 = 0,7$$

$$P(\text{berangin salah} \mid ya) = 6/10 = 0,6$$

$$P(\text{berangin benar} \mid ya) = 4/10 = 0,4$$

Tidak

$$p(\text{cuaca=cerah} \mid \text{Tidak}) = 3/4 = 0,75$$

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

$$P(\text{cuaca=berawan} \mid \text{Tidak})=0/4=0$$

$$P(\text{cuaca hujan} \mid \text{Tidak})=1/4=0,25$$

$$P(\text{suhu panas} \mid \text{Tidak})=2/4=0,5$$

$$P(\text{suhu sejuk} \mid \text{Tidak})= 2/4=0,5$$

$$P(\text{suhu dingin} \mid \text{Tidak})= 0/4=0$$

$$P(\text{kelembaban tinggi} \mid \text{Tidak})=4/4=1$$

$$P(\text{kelembaban normal} \mid \text{Tidak})=0/4=0$$

$$P(\text{berangin salah} \mid \text{Tidak})= 2/4=0,5$$

$$P(\text{berangin benar} \mid \text{Tidak})= 2/4=0,5$$

Tabel 2.4 Probabilitas

Kategori/atribut	Subset	Ya	Tidak
Cuaca	Cerah	0,2	0,75
	Berawan	0,4	0
	Hujan	0,4	0,25
Suhu	Panas	0,2	0,5
	Sejuk	0,4	0,5
	Dingin	0,4	0
Kelembapan	Tinggi	0,3	1
	Normal	0,7	0
Berangin	Salah	0,6	0,5
	Benar	0,4	0,5

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

c. Menguji hasil klasifikasi NBC

$$P(c | F1, F2, F3, \dots, Fn) = p(c) \prod_{i=1}^n P(Fi | C)$$

Contoh data nomor 1

$$\begin{aligned} P(\text{main} | \text{ya}) &= P(\text{main} | \text{ya}) * p(\text{cuaca}=\text{cerah}) * p(\text{suhu}=\text{panas}) * p(\text{kelembapan} \\ &= \text{tinggi}) * p(\text{berangin}=\text{salah}) \\ &= 0,71 * (0,2 * 0,2 * 0,3 * 0,6) \\ &= \mathbf{0,005112} \end{aligned}$$

$$\begin{aligned} P(\text{main} | \text{Tidak}) &= P(\text{main} | \text{Tidak}) * p(\text{cuaca}=\text{cerah}) * p(\text{suhu}=\text{panas}) * p(\text{kele} \\ &= \text{mbapan}=\text{tinggi}) * p(\text{berangin}=\text{salah}) \\ &= 0,29 * (0,75 * 0,5 * 1 * 0,5) \\ &= \mathbf{0,054375} \end{aligned}$$

Tabel 2.5 Hasil Pengujian Klasifikasi NBC

No	Ya	Tidak	Fakta	Klasifikasi	Prediksi
1	0,005112	0,054375	Tidak	Tidak	Sesuai
2	0,003408	0,054375	Tidak	Tidak	Sesuai
3	0,010224	0	Ya	Ya	Sesuai
4	0,020448	0,018125	Ya	Ya	Sesuai
5	0,047712	0	Ya	Ya	Sesuai
6	0,031808	0	Ya	Ya	Sesuai
7	0,031808	0	Ya	Ya	Sesuai
8	0,010224	0,054375	Tidak	Tidak	Sesuai

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

9	0,023856	0	Ya	Ya	Sesuai
10	0,047712	0	Ya	Ya	Sesuai
11	0,015904	0	Ya	Ya	Sesuai
12	0,013632	0	Ya	Ya	Sesuai
13	0,023856	0	Ya	Ya	Sesuai
14	0,013632	0,029	Tidak	Tidak	Sesuai

Pengujian:

Cuaca : hujan

Suhu : panas

Kelembapan : tinggi

Berangin : benar

Main ?

$$P(\text{main} \mid \text{ya}) = 0,71 * (0,4 * 0,2 * 0,3 * 0,4) = 0,006816$$

$$P(\text{main} \mid \text{tidak}) = 0,29 * (0,25 * 0,5 * 1 * 0,5) = 0,018125$$

2.2 Penelitian Terdahulu

Penelitian-penelitian yang sangat berkaitan dan menginspirasi penelitian. Dasar-dasar penelitian sebelumnya yang menjadi tinjauan pustaka pada penelitian ini dirangkum dalam tabel.

Tabel 2.6 Tinjauan pustaka

Nama	Penelitian	Hasil
(Devita et al., 2018)	Perbandingan Kinerja Metode Naive Bayes Dan	Metode Naive Bayes mengungguli metode K-

	K-Nearest Tentang Klasifikasi Artikel Bahasa Indonesia	Nearest Neighbor dalam hal performa. Telah ditunjukkan bahwa 28 jurnal dari Indonesia diklasifikasikan menggunakan pendekatan Naive Bayes dari 40 data uji. Jika dibandingkan, pendekatan K-Nearest Neighbor hanya dapat mengkategorikan 16 jurnal Indonesia dari 40 data uji.
(Ratnawati, 2018)	Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter	Hasil Akurasi yang lebih tinggi berarti sistem berhasil mengklasifikasikan data, yang artinya. Fold kedua menghasilkan akurasi tertinggi, dengan 90%, presisi 92%, penarikan kembali 90%, dan ukuran-f 90%.
(Ulfa et al., 2018)	Twitter Sentiment Analysis using Naive Bayes Classifier with Mutual Information Feature Selection	Hasil tanpa fitur mutual information (MI) 96,2% sedangkan menggunakan MI mendapatkan akurasi sebesar 97,9%.
(Kurniawan et al., 2019)	Analisis Sentimen Opini Film Menggunakan Metode Naive Bayes dan Lexicon Based Features Arifin	Nilai akurasi, presisi, recall, dan ukuran yang diperoleh ketika pendekatan Naive Bayes dan fitur berbasis leksikon diuji untuk analisis sentimen film masing-masing adalah 0,8, 0,8, 0,8, dan 0,8.
(Fransiska & Yolanda, 2019)	Analisis Sentimen Twitter Untuk Review Film Menggunakan Algoritma Naive Bayes Classifier (Nbc) Pada Sentimen R Programming	Lebih banyak sentimen negatif akan ditampilkan dalam hasil analisis sentimen daripada sentimen positif atau netral.
(Gifari et al., 2022)	Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine	Akurasi yang dihasilkan berpengaruh pada data training yang digunakan. Didapatkan akursi 0,85% dengan menggunakan data latih yang besar

Hak Cipta Dilindungi Undang-Undang:

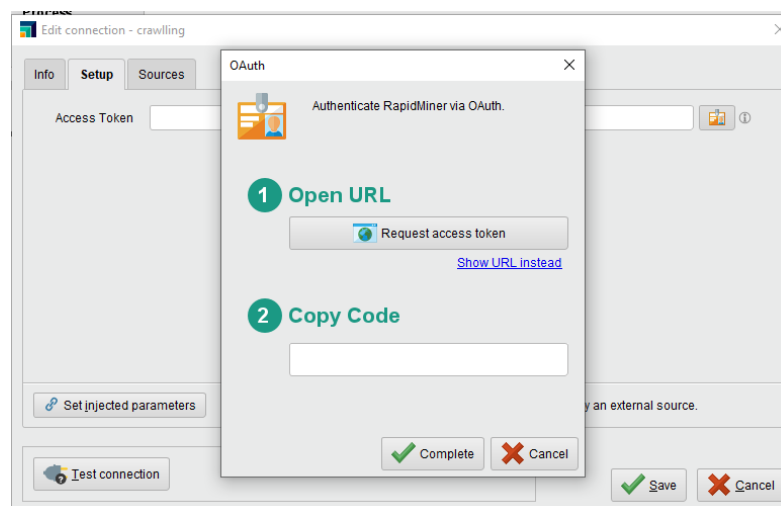
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

BAB III

METODOLOGI PENELITIAN

3.1 Metode Pengumpulan Data

Data yang digunakan dalam penelitian sebagai bagian dari proses pengumpulan data. Peneliti mengumpulkan data berupa ulasan dan komentar pada Twitter mengenai Drama Korea menggunakan rapid miner. Metode *crawling* digunakan untuk mengekstrak data tweet dari jejaring *social* twitter. Data tersebut hanya yang berbahasa Indonesia mengenai drama korea yang bergendre *action*. hal pertama yang harus dilakukan adalah menghubungkan rapid miner dengan akun twitter yang telah dibuat sebelumnya untuk mendapatkan akses token dari aplikasi twitter, sehingga data dari twitter dapat di *crawling* dengan memasukan *keyword* pada aplikasi rapidminer. Gambar dibawah ini merupakan proses akses token pada aplikasi rapid miner yang didapat dari twitter dengan menekan tombol open url.



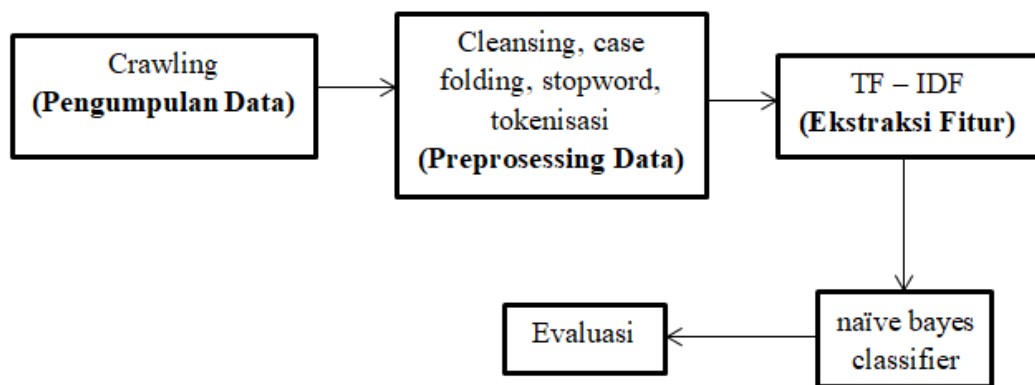
Gambar 3.1 pengambilan token

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

3.2 Tahapan Penelitian

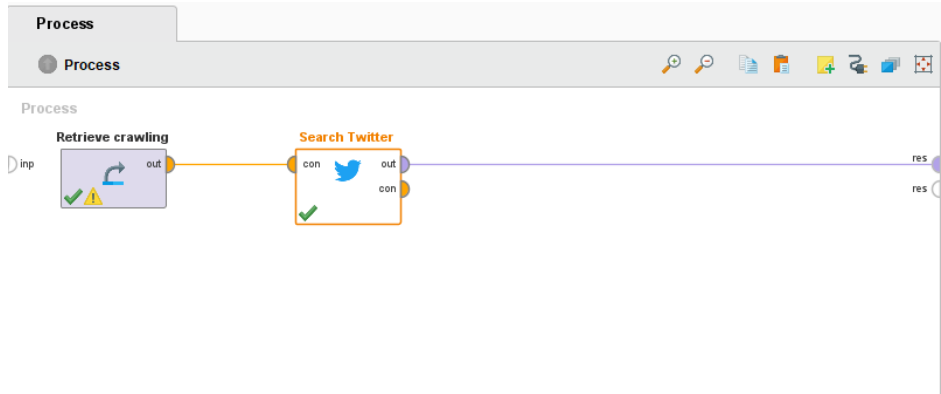
Pada tahapan penelitian ini dilakukan dengan beberapa proses yaitu, metode pengumpulan data dengan cara (*crawling*), Preprocessing Data dengan beberapa tahapan yaitu (*cleasnsing, case folding, stopword, tokenisasi*), kemudian ekstraksi fitur dengan menggunakan rumus (TF-IDF), setelah itu dilakukan penerapan naive bayes classifier. Berikut gambar proses tahapan penelitian.



Gambar 3.2 Tahapan Penelitian

3.3 Pengambilan Data Pada Twitter (Crawling)

Tahapan crawling pada twitter dilakukan dengan memasukan beberapa keyword yaitu, Taxi Driver, Weak Hero Class, Vincenzo, The Uncanny Counter, Duty After School pada operator search twitter. Gambar dibawah ini merupakan proses crawling data dimana data yang berhasil didapatkan sebanyak 2,250 data mentah.



Gambar 3.3 proses crawling data

Data yang didapatkan merupakan data recent dan popular dari seluruh data mulai 2021 hingga 05 April 2023. Tabel dibawah merupakan hasil dari *crawling* data pada aplikasi twitter.

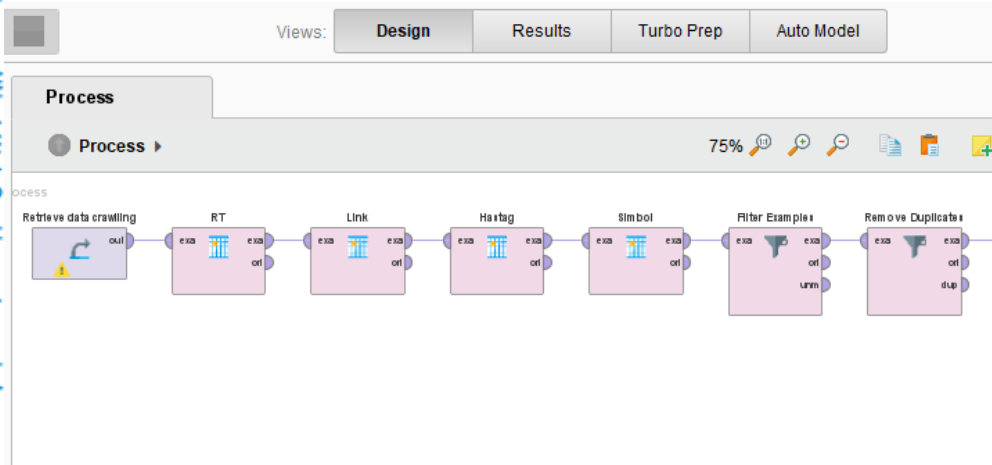
Tabel 3.1 data *crawling*

No.	Data Crawling
1.	RT@yuciellyw kalau kamu butuhnya yang ringan ringan di netflix ada love to hate you terus kalau nyarinya yang rada rada berat di netflix ada kingdom ini latarnya sageuk atau gak yang seru tapi menegangkan the uncanny counter terus kalau nyarinya sad ending#uncannycounter
2.	ga akan bosan buat kasih recomend mouse ? klo kelen suka drakor yg bikin mikir, sama taxi driver ? y'all kudu nonton krna casenya base on reality jdi kewren binggo pokoknya mahh https://t.co/OdkEc68VvA
3.	suka bgt sama drakor Taxi Driver. action-nya ga yg tembak2an, lebih ke h2h. nontonnya jdi lebih realistik aja lawannya di bikin tumbang sm tonjokan. soalnya kalo organisasi sipil gt kayanya kureng kalo di ceritain sampe punya/pake gun(?) anw sopir taksi satu ini meresahkan yh?? https://t.co/857j5O9i9Q
4.	Banyak Makna yang bisa kita ambil dari Drama Taxi Driver ini lebih singkatnya dalam Drama ini menunjukkan betapa pentingnya keadilan sosial dalam masyarakat tanpa memandang kelas sosial
5.	Habis nonton Taxi Driver season 1, ternyata seseru itu dan bbrp kasus di season 1 diambil dari kasus di dunia nyata yang sangat terkenal. Lanjut season 2 dan dipastikan bbrp kasus di season 2 juga diambil dari kasus di dunia nyata.
6.	Terkesan dengan Duty After School bagian 1 hingga tak sabar menunggu bagian keduanya, selesiannya malah begitu. Di sisi lain gue paham, tapi di sisi lainnya gue tak puas dan menyayangkannya. Sama seperti Weak Hero

	Class 1, seharusnya gue tidak menonton selesai menit akhirnya.
7.	Sembari nunggu episode lanjutannya Duty After School. Nonton Weak Hero Class 1 dulu deh. Wml, semoga masuk ke seleraku https://t.co/0r0gLYeh8X
8.	RT@0xtashaa BAGUS BGTT STIAP EPS ITU BEDA MASALAH, agak mirip kek taxi driver. cm bedanya kl tomorrow itu 'malaikat maut' yg mau nylamatin org yg pgn bunuh diri biar gajadi bundir.
9.	dulu tuh waktu awal2 nge drakor taun 2018/2019 gua selalu nyuruh org2 buat nntn bya dr romantic,krn emg gua sesuka itu sm tu drakor,sampe akhirnya tahta nya ke geser dikit sm vincenzo dan the uncanny counter wkwkwk #Vincenzo #uncannycounter
10.	RT@tanyarlfe nonton weak hero class 1 deh nder, aku sempet depresi jg dan gampang nangis. km nonton itu emosi dan mental km bener2 dimainin dampaknya emosi diri sendiri bisa kembali normal. walaupun harus nangis kejer, tp itu titik poin sembuhnya nder.
....
2,250	Yeyyy selesai taxi driver 2, over all 8,7/10?? berani bgt kasus2nya, penyelesaiannya mantep, yg paling penting adlh tidak monoton scene2 berantemnya jd aku sih yes. Di season 2 kasus2nya jauh lebih berat drpd season 1. Di season 1 kayak banyak bgt kasusnya ya? Di yg ini cm bbrp

3.4 Preprocessing Data

Tahapan selanjutnya adalah processing data. Pada tahapan ini peneliti melakukan pembersihan data yang telah di *crawling* karena data yang didapatkan masih berupa data mentah dan memerlukan cleansing untuk mempermudah dalam melakukan analisis sentiment. Ada beberapa tahapan dalam melakukan preprocessing data yaitu cleansing retweet, URL, hastag, dan symbol, tujuannya untuk menyiapkan data untuk dilakukan analisis dengan cara menghapus beberapa hal yang tidak diperlukan.



Gambar 3.4 Processing Data

Gambar diatas merupakan tahapan preprocessing data atau cleansing data dengan menggunakan operator replace yang tersedia pada aplikasi rapid miner untuk menghilangkan cleansing retweet, URL, hastag, dan symbol pada twitter yang telah di crawling karena tidak diperlukan pada analisis sentiment.

a. Cleansing Retweet

Tabel dibawah ini merupakan cleansing retweet (RT) dimana seluruh data tweet yang mengandung retweet (RT) akan dihilangkan karena tidak digunakan atau diperlukan pada analisis sentiment serta tidak memiliki pengaruh.

Tabel 3.2 Cleansing retweet

Sebelum Cleansing Retweet	Sesudah Cleansing Retweet
RT@Oxtashaa BAGUS BGTT STIAP EPS ITU BEDA MASALAH, agak mirip kek taxi driver. cm bedanya kl tomorrow itu 'malaikat maut' yg mau nylamatin org yg pgn bunuh diri biar gajadi bundir.	Oxtashaa BAGUS BGTT STIAP EPS ITU BEDA MASALAH, agak mirip kek taxi driver. cm bedanya kl tomorrow itu 'malaikat maut' yg mau nylamatin org yg pgn bunuh diri biar gajadi bundir.
RT@tanyarlfe nonton weak hero class 1 deher, aku sempet depresi jg dan gampang nangis. km nonton itu emosi	tanyarlfe nonton weak hero class 1 deher, aku sempet depresi jg dan gampang nangis. km nonton itu emosi

Hak Cipta Dilindungi Undang-Undang:
 1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
 2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

dan mental km bener2 dimainin dampaknya emosi diri sendiri bisa kembali normal. walaupun harus nangis kejer, tp itu titik poin sembuhnya nder.	dan mental km bener2 dimainin dampaknya emosi diri sendiri bisa kembali normal. walaupun harus nangis kejer, tp itu titik poin sembuhnya nder.
RT@kdrama_menfess gue ga bisa nonton action thriller pokoknya yang ada darah darah nya. trus nonton doctor romantic 2, loh kok kuat (tp ttp gue skip kalo emg banyak darahnya). akhirnya nonton taxi driver sampe ngikutin on going season 2 ?	kdrama_menfess gue ga bisa nonton action thriller pokoknya yang ada darah darah nya. trus nonton doctor romantic 2, loh kok kuat (tp ttp gue skip kalo emg banyak darahnya). akhirnya nonton taxi driver sampe ngikutin on going season 2 ?
RT @ChangeToku @ZoelHelmiLubis1 Betuuullll.. ya Allah.. hbs liat taxi driver yg brutal kyk gt, eh liat Vidio David d aniaya, ngeriiii.. tryta d real life g jauh beda dg drakor.. ?????	Betuuullll.. ya Allah.. hbs liat taxi driver yg brutal kyk gt, eh liat Vidio David d aniaya, ngeriiii.. tryta d real life g jauh beda dg drakor.. ?????
RT@yuciellyw kalau kamu butuhnya yang ringan ringan di netflix ada love to hate you terus kalau nyarinya yang rada rada berat di netflix ada kingdom ini latarnya sageuk atau gak yang seru tapi menegangkan the uncanny counter terus kalau nyarinya sad ending #uncannycounter	yuciellyw kalau kamu butuhnya yang ringan ringan di netflix ada love to hate you terus kalau nyarinya yang rada rada berat di netflix ada kingdom ini latarnya sageuk atau gak yang seru tapi menegangkan the uncanny counter terus kalau nyarinya sad ending #uncannycounter

b. Cleansing Url

Tabel dibawah merupakan contoh sebelum dan sesudah dilakukan proses cleansing url dimana seluruh data tweet yang mengandung URL akan dihapus karena tidak diperlukan.

Tabel 3.3 Cleansing URL

Sebelum Cleansing URL	Setelah Cleansing URL
suka bgt sama drakor Taxi Driver. action-nya ga yg tembak2an, lebih ke h2h. nontonnya jdi lebih realistis aja lawannya di bikin tumbang sm tonjokan. soalnya kalo organisasi sipil gt kayanya kureng kalo di ceritain sampe punya/pake gun(?) anw sopir	suka bgt sama drakor Taxi Driver. action-nya ga yg tembak2an, lebih ke h2h. nontonnya jdi lebih realistis aja lawannya di bikin tumbang sm tonjokan. soalnya kalo organisasi sipil gt kayanya kureng kalo di ceritain sampe punya/pake gun(?) anw sopir

taksi satu ini meresahkan yh?? https://t.co/857j5O9i9Q	taksi satu ini meresahkan yh??
Ada pula Yoo In Soo, yang dikenal dengan perannya dalam drama Korea <i>Alchemy of Souls</i> dan <i>The Uncanny Counter 2</i> . Sang aktor berperan sebagai Bang Sam Shik, troublemaker desa yang jatuh cinta dengan Mi Joo sejak kecil. https://t.co/uQVxSzFcS4	Ada pula Yoo In Soo, yang dikenal dengan perannya dalam drama Korea <i>Alchemy of Souls</i> dan <i>The Uncanny Counter 2</i> . Sang aktor berperan sebagai Bang Sam Shik, troublemaker desa yang jatuh cinta dengan Mi Joo sejak kecil.
Sembari nunggu episode lanjutannya <i>Duty After School</i> . Nonton <i>Weak Hero Class 1</i> dulu deh. Wml, semoga masuk ke seleraku https://t.co/0r0gLYeh8X	Sembari nunggu episode lanjutannya <i>Duty After School</i> . Nonton <i>Weak Hero Class 1</i> dulu deh. Wml, semoga masuk ke seleraku
suka bgt sama drakor <i>Taxi Driver</i> . action-nya ga yg tembak2an, lebih ke h2h. nontonnya jdi lebih realistis aja lawannya di bikin tumbang sm tonjokan. soalnya kalo organisasi sipil gt kayanya kureng kalo di ceritain sampe punya/pake gun(?) anw sopir taksi satu ini meresahkan yh?? https://t.co/857j5O9i9Q	suka bgt sama drakor <i>Taxi Driver</i> . action-nya ga yg tembak2an, lebih ke h2h. nontonnya jdi lebih realistis aja lawannya di bikin tumbang sm tonjokan. soalnya kalo organisasi sipil gt kayanya kureng kalo di ceritain sampe punya/pake gun(?) anw sopir taksi satu ini meresahkan yh??
•kdm• aku baru selesai <i>Weak Hero Class 1</i> . storynya manis bgt yaaa sebenarnya, tapi ujungnya bikin mewek. btw ini salah satu fav scene aku??? https://t.co/E54KLD2k0y	•kdm• aku baru selesai <i>Weak Hero Class 1</i> . storynya manis bgt yaaa sebenarnya, tapi ujungnya bikin mewek. btw ini salah satu fav scene aku???

c. Cleansing Hastag

Tabel dibawah ini merupakan cleansing Hastag dimana seluruh data tweet yang mengandung Hastag akan dihilangkan karena tidak digunakan atau diperlukan pada analisis sentiment serta tidak memiliki pengaruh

Tabel 3.4 *Cleansing hastag*

Sebelum <i>Cleansing Hastag</i>	Setelah <i>Cleansing Hastag</i>
yuciellyw kalau kamu butuhnya yang ringan ringan di netflix ada love to hate you terus kalau nyarinya yang rada rada	yuciellyw kalau kamu butuhnya yang ringan ringan di netflix ada love to hate you terus kalau nyarinya yang rada rada

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi

berat di netflix ada kingdom ini latarnya sageuk atau gak yang seru tapi menegangkan the uncanny counter terus kalau nyarinya sad ending#uncannycounter	berat di netflix ada kingdom ini latarnya sageuk atau gak yang seru tapi menegangkan the uncanny counter terus kalau nyarinya sad ending
kdrama_menfess vincenzo, taxi driver sama the glory si keren banget bales dendamnya menurutku #vincenzo #taxidriver #theglory	kdrama_menfess vincenzo, taxi driver sama the glory si keren banget bales dendamnya menurutku
dulu tuh waktu awal2 nge drakor taun 2018/2019 gua selalu nyuruh org2 buat nntn bya dr romantic,krn emg gua sesuka itu sm tu drakor,sampe akhirnya tahta nya ke geser dikit sm vincenzo dan the uncanny counter wkwkwk #Vincenzo #uncannycounter	dulu tuh waktu awal2 nge drakor taun 2018/2019 gua selalu nyuruh org2 buat nntn bya dr romantic,krn emg gua sesuka itu sm tu drakor,sampe akhirnya tahta nya ke geser dikit sm vincenzo dan the uncanny counter wkwkwk
ccandylan dii viu kemarinn aku pake buaat marathon dr romantic 2, di viu ada taxi driver, duty after school, terus ada kemarin baru release film gore project wolf hunting! menurutku rame rame ajasihh #drromantic2 #taxidriver #dutyafterschool	ccandylan dii viu kemarinn aku pake buaat marathon dr romantic 2, di viu ada taxi driver, duty after school, terus ada kemarin baru release film gore project wolf hunting! menurutku rame rame ajasihh
Orang orang kok bisa anjhier ngerekomendasiin nntn taxi driver. Awalnya sie oke ya seru banget gila sie buat yang gak terlalu suka romantic scene lu cocok sie karna scene nya dikit dan ya gak terlalu romantic b aja, tapi klo lu gk terlalu suka sikapat'an mending jangan nntn dah #taixidriver	Orang orang kok bisa anjhier ngerekomendasiin nntn taxi driver. Awalnya sie oke ya seru banget gila sie buat yang gak terlalu suka romantic scene lu cocok sie karna scene nya dikit dan ya gak terlalu romantic b aja, tapi klo lu gk terlalu suka sikapat'an mending jangan nntn dah

d. Cleansing Simbol

Tabel dibawah ini merupakan cleansing Simbol dimana seluruh data tweet yang mengandung Simbol akan dihilangkan karena tidak digunakan atau diperlukan pada analisis sentiment serta tidak memiliki pengaruh.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

Tabel 3.5 *Cleansing* simbol

Sebelum <i>Cleansing</i> Simbol	Setelah <i>Cleansing</i> Simbol
Terkesan dengan Duty After School bagian 1 hingga tak sabar menunggu bagian keduanya, selesainya malah begitu. Di sisi lain gue paham, tapi di sisi lainnya gue tak puas dan menyayangkannya. Sama seperti Weak Hero Class 1, seharusnya gue tidak menonton selesaian menit akhirnya.	Terkesan dengan Duty After School bagian 1 hingga tak sabar menunggu bagian keduanya selesainya malah begitu Di sisi lain gue paham tapi di sisi lainnya gue tak puas dan menyayangkannya. Sama seperti Weak Hero Class 1 seharusnya gue tidak menonton selesaian menit akhirnya
Yeyyy selesaii taxi driver 2, over all 8,7/10?? berani bgt kasus2nya, penyelesaiannya mantep, yg paling penting adlh tidak monoton scene2 berantemnya jd aku sih yes. Di season 2 kasus2nya jauh lebih berat drpd season 1. Di season 1 kayak banyak bgt kasusnya ya? Di yg ini cm bbrp	Yeyyy selesaii taxi driver 2 over all 8710 berani bgt kasus2nya penyelesaiannya mantep, yg paling penting adlh tidak monoton scene2 berantemnya jd aku sih yes Di season 2 kasus2nya jauh lebih berat drpd season 1 Di season 1 kayak banyak bgt kasusnya ya Di yg ini cm bbrp
dulu tuh waktu awal2 nge drakor taun 2018/2019 gua selalu nyuruh org2 buat nntn bya dr romantic,krn emg gua sesuka itu sm tu drakor,sampe akhirnya tahta nya ke geser dikit sm vincenzo dan the uncanny counter wkwkwk	dulu tuh waktu awal2 nge drakor taun 20182019 gua selalu nyuruh org2 buat nntn bya dr romantic krn emg gua sesuka itu sm tu drakor sampe akhirnya tahta nya ke geser dikit sm vincenzo dan the uncanny counter wkwkwk
suka bgt sama drakor Taxi Driver. action-nya ga yg tembak2an, lebih ke h2h. nontonnya jdi lebih realistis aja lawannya di bikin tumbang sm tonjokan. soalnya kalo organisasi sipil gt kayanya kureng kalo di ceritain sampe punya/pake gun(?) anw sopir taksi satu ini meresahkan yh??	suka bgt sama drakor Taxi Driver actionnya ga yg tembak2an lebih ke h2h nontonnya jdi lebih realistis aja lawannya di bikin tumbang sm tonjokan. soalnya kalo organisasi sipil gt kayanya kureng kalo di ceritain sampe punya pake gun anw sopir taksi satu ini meresahkan yh
Sinopsis Taxi Driver 2 Ceritakan Aksi Balas Dendam, Masuk Daftar Drakor dengan Rating Tinggi Awal Tahun	Sinopsis Taxi Driver 2 Ceritakan Aksi Balas Dendam Masuk Daftar Drakor dengan Rating Tinggi Awal Tahun

e. Filter Example

Filter example merupakan salah satu fitur yang terdapat pada aplikasi rapid miner yang berfungsi untuk menghilangkan data kosong (*missing value*) di



tanda dengan “?” akan dihapus, sehingga data crawling yang awalnya 2,250 menjadi 2,142 data setelah proses filter example dilakukan.

f. Remove Duplicate

Remove Duplicate berfungsi untuk menghapus file yang duplikat atau sama dengan cara membandingkan data yang telah di *crawling* secara keseluruhan, jika 2 atau lebih komentar yang sama terdapat pada suatu dataset maka akan terhapus secara permanen dengan menyisahkan satu data atau komentar yang sama tersebut sehingga data yang didapatkan tidak memiliki duplikat atau kesamaan dengan data lain. Data yang awalnya berjumlah 2,142 setelah proses filter example, kini menjadi 1,865 data setelah dilakukan remove duplikat.

Row No.	Text
1	Mmmm Terakhir namatin taxi driver 2 dan blm ada drama yang bikin excited lagi
2	kdramenfess vincenzo taxi driver sama the glory si keren banget bales dendamnya menurutku
3	ccandylan dii viu kemarin aku pake buaat marathon dr romantic 2 di viu ada taxi driver duty after school terus ada kemarin baru release film gore proj...
4	taxi driver seru bgt jirrr
5	kdramenfess Vincenzo the glory dan taxi driver mantap sih bls dendamnya
6	kdramenfess Taxi driver 1 jga gk kalah seru
7	gue nonton taxi driver dulu baru nonton w two world jadinya kek gbs ngebayangin pak jang taxi driver yg super baik itu mukanya jadi karakter jahat TT
8	kdramenfess drakor where stars land - staff airport jujur ini yg buat aku nonton sih km temanya yg beda guess what...
9	yook semangat lanjutin taxi driver 2 nya ga nambah2 lg ini ngestuck eps 6 wkwk
10	Nonton taxi driver yang dibayangkan gue mukanya wonwoo gara2 mirip bgt sama lee jehoon
11	Akhirnya mood untuk nonton drakor kembali lagi baru nyelesain kokdu abis ini taxi driver 2 sisa 5 episode ehehehehe
12	taxi driver seru juga
13	nafaaisya Ya allah podo siissss aku tamat nonton taxi driver rasane hampa wkwkwkwksk
14	Ini ada agc duty sama taxi driver Sinting bgt wkwkwk bkin pucing

Gambar 3.5 Remove Duplicate

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

3.5 Pelabelan Data Manual

Pelabelan manual dilakukan pada 1200 dan 1300 dari 1,815 data dimana data diberi label positif dan negatif sebelum pelabelan otomatis menggunakan rapid miner, untuk mengatasi kesalahan pelabelan maka dilakukan majority rule yaitu pengambilan keputusan melalui voting terbanyak sehingga diperlukan 3 orang dalam melabeli data. Jika dalam satu komentar terdapat 2 orang yang berpendapat positif maka sentimennya positif begitupula sebaliknya dan jika dari ke 3 orang tersebut berpendapat positif maka sudah pasti komentar tersebut memiliki sentimen positif begitupula sebaliknya, seperti tabel dibawah ini.

Keterangan :

P1 : pelabel 1

P2 : pelabel 2

P3 : pelabel 3

Sentimen : hasil prediksi sentimen berdasarkan majority rule

Tabel 3.6 Pelabelan data manual

Teks	P1	P2	P3	Sentimen
kraramamenfess vincenzo taxi driver sama the glory si keren banget bales dendamnya menurutku	p	p	p	Positif
gamalatuconsina iya yg itu kek dia ngomong aja aku dah kesel bgt nah kebetulan abis selesai happiness pas itu kan tayang taxi driver wah kesalnya berlanjut tuh	n	n	n	Negatif
kumparan jadi inget kasus di drakor taxi driver yg polisi dimanipulasi kematiannya dibuat seolah2 bunuh diri	n	p	n	Negatif
ccandylan dii viu kemarinn aku pake buaat marathon dr romantic 2 di viu ada taxi driver	p	p	p	Positif

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

duty after school terus ada kemarin baru release film gore project wolf hunting menurutku rame rame ajasihh				
sumpah gweh nangis banget sampe sesegukan nonton weak hero class kek anjir nyesek bgt bgt smwaa gara2 cowo	n	p	n	Negatif
ekonomark oh sama the uncanny counter pemerannya sejeong itu seru bgt jugaak d dann the first responder juga super seruu btw itu rata2 drakor yg rada berat tapi comedy juga d	p	p	p	Positif
anjir kelar juga namatin taxi driver season 2 masih banyak banget teka-teki yg belum keungkap pls season 3 rainbow crew harus tetep sama dan lengkap anggotanya	p	p	n	Positif
kDRAMAMENFESS murid das sma weak hero class bisa digabung aja gk biar di weak hero bisa merasakan kesolidannya	p	p	p	Positif
kDRAMAMENFESS duty after school taxi driver 12 akoh otw mau nonton bora deborah setelah drama panjang menegangkan itu fyuhhh	p	p	p	Positif
oalahpah kalo diriku sih lanjutmeskipun rasa penasarannya udah berkurang gara-gara spoiler yang ada dimana-mana selain itu belaum tertarik nonton series lain setelah namatin taxi driver 2	p	n	p	Positif

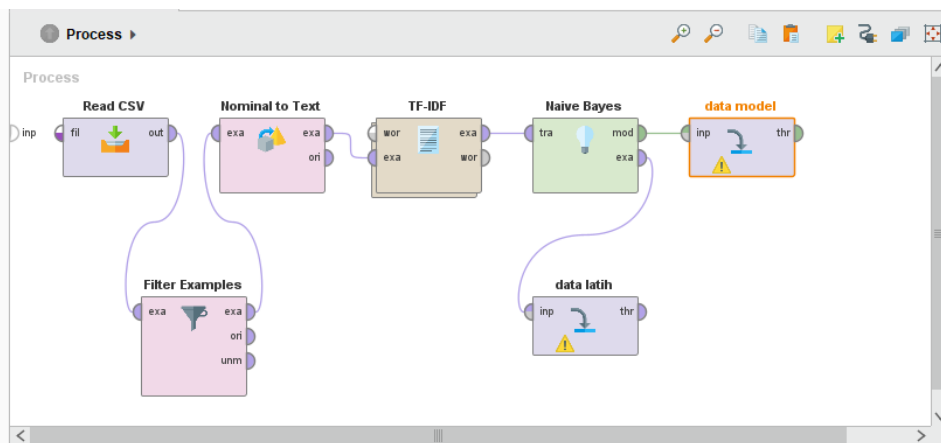
3.6 Klasifikasi Metode Naïve Bayes

Pada tahapan ini metode naïve bayes berperan untuk membagi data model dan data latih sehingga diperlukan beberapa operator yaitu Filter example karena pada proses pembangunan model pada data yang memiliki label yaitu 1200 dan 1300 yang dilabeli secara manual. Operator nominal to teks digunakan untuk memproses data filter example yang awalnya nominal menjadi teks karena metode pemrosesan data TF-IDF membutuhkan data berupa teks bukan nominal Serta mempelajari dan mengetahui apakah suatu tweet berupa sentiment positif atau

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

negatif, sehingga bisa menghasilkan model klasifikasi untuk menentukan sentiment dari tweet lain yang belum memiliki sentiment. Kemudian model naïve bayes disimpan pada operator store yang diberi nama data model dan data hasil pemrosesan juga disimpan yang diberi nama data latih untuk mentraining model klasifikasi.



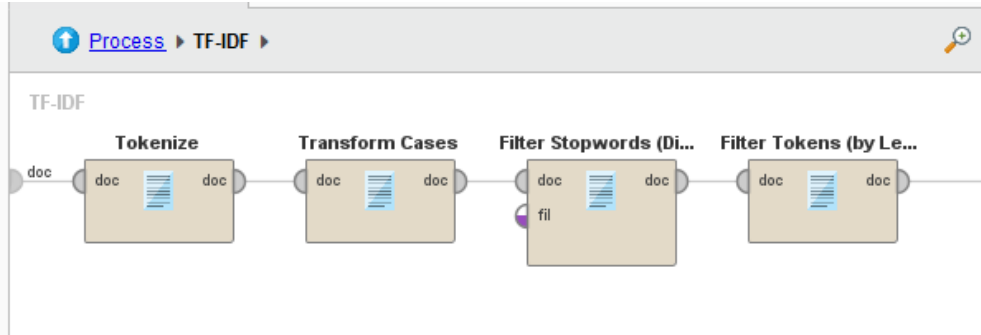
Gambar 3.6 Klasifikasi Metode Naïve Bayes

3.7 Ekstraksi Fitur TF-IDF

merupakan proses yang didalamnya terdapat beberapa operator yaitu *tokenize*, *transform cases*, *filter stopwords*, dan *filter tokens* dengan menggunakan operator tambahan yang bisa didapatkan dari google. Proses ini di masukan kedalam operator proses dokumen from data.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunan Jember
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunan Jember



Gambar 3.7 Processing Data

a. *Tokenize*

Tokenize merupakan proses pemisahan kalimat menjadi kata atau bagian yang berbeda yang awalnya panjang menjadi suatu kata per kata yang lebih kecil atau sering disebut sebagai token. *Tokenize* berguna untuk mempermudah dalam menentukan sentiment pada setiap kata sehingga panggilan makna disetiap kalimat menjadi lebih mudah.

Tabel 3.7 *Tokenize*

Sebelum <i>tokenize</i>	Setelah <i>tokenize</i>
Drakor Taxi Driver korban atau orang terzholimi bisa balas dendam lewat jasa taksi mewah mereka kaya gladiator atau agent gitu	Drakor – Taxi – Driver – korban - atau orang – terzholimi – bisa - balas dendam – lewat – jasa – taksi - mewah mereka – kaya – gladiator – atau – agent – gitu
Gue tu lg bosen drama bgtttt tp jd semangat gara2 ada Duty After School baru aja tamat rasanya kek gamau pisah ajaaa lg bosen sama taxi driver yg s2 ini malah bosenin menurutku timbang yg s1 ini mau lanjut aja males bgt	Gue – tu – lg – bosen – drama – bgtttt – tp – jd – semangat – gara2 – ada – Duty – After – School – baru – aja – tamat – rasanya – kek – gamau –pisah – ajaaa – lg – bosen – sama – taxi – driver – yg – s2 – ini – malah – bosenin menurutku – timbang – yg – s1 – ini mau – lanjut – aja – males – bgt
•kdm• aku baru selesai weak hero class 1 storynya manis bgt yaaa sebenarnya tapi ujungnya bikin mewek btw ini	•kdm• – aku – baru – selesai – weak hero – class – 1 – storynya – manis bgt – yaaa – sebenarnya – tapi ujungnya –

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suttha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suttha Jambi

salah satu fav scene aku	bikin – mewek – btw – ini salah – satu – fav – scene – aku
ga akan bosan buat kasih recomend mouse klo kelen suka drakor yg bikin mikir sama taxi driver yall kudu nonton krna casenya base on reality jdi kewren bingo pokoknya mahh	ga – akan – bosan – buat – kasih recommend – mouse – klo – kelen – suka – drakor – yg – bikin – mikir – sama – taxi – driver – yall – kudu nonton – krna – casenya – base – on reality – jdi – kewren – bingo pokoknya – mahh
ga akan bosan buat kasih recomend mouse klo kelen suka drakor yg bikin mikir sama taxi driver yall kudu nonton krna casenya base on reality jdi kewren bingo pokoknya mahh	ga – akan – bosan – buat - kasih recommend – mouse – klo – kelen – suka – drakor – yg – bikin – mikir – sama – taxi – driver – yall – kudu – nonton – krna – casenya – base - on reality – jdi – kewren – bingo – pokoknya – mahh

b. Case Folding

Case folding merupakan tahapan processing mengubah huruf besar menjadi kecil karena data yang di dapatkan dari hasil *crawling* tidak terstruktur terhadap penggunaan huruf capital pada suatu kalimat ataupun data, sehingga perlu dilakukan case folding untuk mengubah semua huruf pada dokumen disetiap data menjadi huruf kecil yang dapat dilakukan dengan menggunakan operator transform cases yang tersedia di rapid miner

Tabel 3.8 *Case folding*

Sebelum Case Folding	Setelah Case Folding
Terkesan dengan Duty After School bagian 1 hingga tak sabar menunggu bagian keduanya selesiannya malah begitu Di sisi lain gue paham tapi di sisi lainnya gue tak puas dan menyayangkannya Sama seperti Weak Hero Class 1 seharusnya gue tidak menonton selesaian menit akhirnya	terkesan dengan duty after school bagian 1 hingga tak sabar menunggu bagian keduanya selesiannya malah begitu di sisi lain gue paham tapi di sisi lainnya gue tak puas dan menyayangkannya sama seperti weak hero class 1 seharusnya gue tidak menonton selesaian menit akhirnya
Yeeyyy selesaii taxi driver 2 over all	yeeyyy selesaii taxi driver 2 over all

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

8710 berani bgt kasus2nya penyelesaiannya mantep yg paling penting adlh tidak monoton scene2 berantemnya jd aku sih yes Di season 2 kasus2nya jauh lebih berat drpd season 1 Di season 1 kayak banyak bgt kasusnya ya Di yg ini cm bbrp	8710 berani bgt kasus2nya penyelesaiannya mantep yg paling penting adlh tidak monoton scene2 berantemnya jd aku sih yes di season 2 kasus2nya jauh lebih berat drpd season 1 di season 1 kayak banyak bgt kasusnya ya di yg ini cm bbrp
Oxtashaa BAGUS BGTT STIAP EPS ITU BEDA MASALAH agak mirip kek taxi driver cm bedanya kl tomorrow itu malaikat maut yg mau nylamatin org yg pgn bunuh diri biar gajadi bundir	Oxtashaa bagus bgtt stiap eps itu beda masalah agak mirip kek taxi driver cm bedanya kl tomorrow itu malaikat maut yg mau nylamatin org yg pgn bunuh diri biar gajadi bundir
kDRAMAMENFESS Doctor Cha lucu, relatable dan refreshing Dr Romantic 3 series wajib buat penikmat Drakor Oasis klo suka yg sedih dan super angst Taxi Driver 2 thrilling dan seru queenmaker tipe drama politic yg bnyk aksi2 pinternya sama the Good Bad Mother	kDRAMAMENFESS doctor cha lucu, relatable dan refreshing dr romantic 3 series wajib buat penikmat drakor oasis klo suka yg sedih dan super angst taxi driver 2 thrilling dan seru queenmaker tipe drama politic yg bnyk aksi2 pinternya sama the good bad mother
Gue tu lg bosen drama bgttttt tp jd semangat gara2 ada Duty After School baru aja tamat rasanya kek gamau pisah ajaaa lg bosen sama taxi driver yg s2 ini malah bosenin menurutku timbang yg s1 ini mau lanjut aja males bgt	gue tu lg bosen drama bgttttt tp jd semangat gara2 ada duty after school baru aja tamat rasanya kek gamau pisah ajaaa lg bosen sama taxi driver yg s2 ini malah bosenin menurutku timbang yg s1 ini mau lanjut aja males bgt

c. Stopword

Stopword merupakan proses pengurangan kata yang memiliki makna kurang berarti pada suatu dokumen seperti kata “yang” “apa” dan masih banyak lagi. Dengan menggunakan operator filter stopword serta stopword bahasa dengan jumlah 758 kata tidak penting yang harus di download terlebih dahulu pada aplikasi google, tabel dibawah ini merupakan beberapa contoh kata stopword.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suntho Jambi

Tabel 3.9 Stopword

ada	Dari	apa	Ke
akankah	Di	banyak	Tadi
akhiri	Dia	baru	Tapi
Aku	Itu	cuma	Waktu
Andalah	Kalau	dan	Yang

d. Filter Token

Filter token dilakukan sama seperti *stopword* yaitu menghapus kata yang tidak memiliki makna penting akan tetapi pada filter token kita menentukan mana yang dihapus dengan memasukkan *min character* 4 dan *max character* 25 sehingga kata yang memiliki kurang dari 4 dan kata lebih dari 25 akan dihapus karena tidak diperlukan.

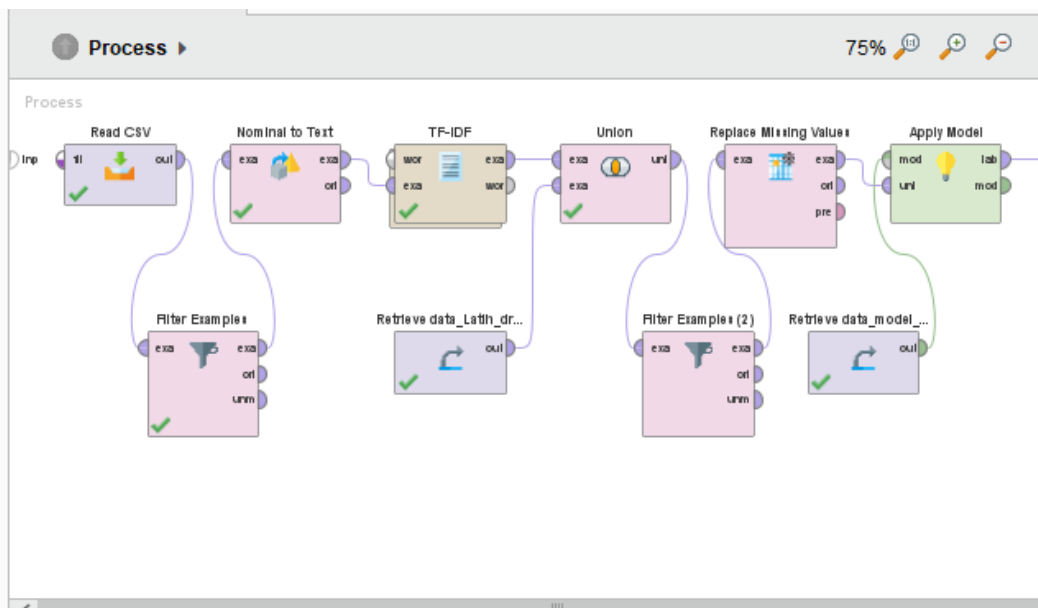
3.8 Pemodelan Metode Naïve Bayes

Pada proses ini hampir sama dengan proses sebelumnya dimana menggunakan operator *read CSV* untuk memasukkan data dilanjutkan dengan menggunakan filter example “*is missing*” data yang diambil adalah data kosong yang belum diberi label. Kemudian menggunakan *nominal to teks* dan lanjut ke pemrosesan data *TF-IDF* yang berisi *tokenize*, *transform cases*, *stopword* dan filter token. operator *union* yang berfungsi untuk menggabungkan data latih dan data model karena memiliki atribut yang berbeda serta diperlukan Filter example

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

untuk membuang data latih. Data model yang telah di simpan sebelumnya di sambungkan pada operator apply model pada bagian “mod” dan hasil dari replace missing values dihubungkan juga pada operator apply model pada bagian “unl” berupa data yang belum memiliki label sehingga pada operator apply model didapatkan hasil sentimen berupa positif maupun negatif dari keseluruhan data.



Gambar 3.8 Pemodelan Metode Naïve Bayes

3.9 Pengujian Dan Evaluasi

a. Cross Validation

K- fold cross validation merupakan Metodologi tambahan data mining yang berupaya mendapatkan hasil paling akurat. Teknik ini di mana percobaan dijalankan k kali untuk satu model dengan parameter yang sama. Tahapan yang dilakukan dalam pengujian model klasifikasi meliputi pengujian dan evaluasi. Dengan menggunakan validasi silang *k-fold* dengan $k = 5$, pengujian dilakukan.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

Lima divisi digunakan untuk memisahkan data. Data tersebut terdiri dari 4 data *training* dan 1 data *testing* yang keduanya memiliki nomor yang sama. (Theo et al., 2020)



Gambar 3.9 Pengujian Dan Evaluasi

b. Split Validation

Split validation merupakan metode klasifikasi pada naïve bayes, yang bertugas membagi data menjadi dua bagian yaitu data latih dan data model. Data latih digunakan sebagai data pelatihan sedangkan data model adalah data yang bertugas sebagai pengujian.

3.10 Confution matrix

Hasil klasifikasi biner pada suatu dataset direpresentasikan oleh matriks berukuran 2x2 yang disebut dengan confution matrix. Kinerja klasifikasi dapat dihitung dengan menggunakan beberapa rumus umum. Hasil akurasi, presisi dan recall akan didapatkan dalam bentuk presentai. (Andika et al., 2019)

a. Akurasi

Akurasi merupakan tingkat prediksi atau pengukuran yang benar, dibawah ini merupakan rumus perhitungan akurasi.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} * 100\%$$

b. Presisi

Presisi merupakan prediksi jumlah keseluruhan data yang relevan atau akurat yang telah ditentukan oleh sistem, dibawah ini merupakan rumus dari presisi.

$$\text{Presisi} = \frac{TP}{FP+TN} * 100\%$$

c. Recall

Recall merupakan prediksi jumlah keseluruhan data relevan atau akurat yang terdapat pada koleksi data, dibawah ini merupakan rumus recall.

$$\text{Recall} = \frac{TP}{FN+TP} * 100\%$$

keterangan :

TP : *True* Positif

TN : *True* Negatif

FP : *Fals* Positif

FN : *Fals* Negatif

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

3.16 Alat dan Bahan

3.16.1 Alat yang digunakan dalam penelitian ini yaitu:

Hardware:

PC/Laptop dengan spesifikasi *processor* Intel (R) Core i3 RAM 4 GB,
System operasi Windows 10 64 bit.

Software:

1. Aplikasi Rapid Miner
2. Pengumpulan data menggunakan aplikasi Twitter.

3.16.2 Bahan yang digunakan dalam penelitian ini yaitu:

Data / komentar yang terdapat pada aplikasi Twitter.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

BAB IV

HASIL DAN PEMBAHASAN

4.1 Ekstraksi Fitur TF-IDF

Setelah melalui seluruh proses TF-IDF yang menghitung kemunculan suatu kata dalam data twitter, maka didapatkan data seperti di bawah ini. Jika nilainya 0 artinya kata tersebut tidak muncul pada pada suatu komentar, apabila memiliki nilai seperti pada kata besok 0,213 berarti muncul pada suatu komentar.

ing	berular	berusaha	berwarna	besok	best	betapa	beugh	beut	beycz	b
	0	0	0	0.213	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0.257	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0

Gambar 4.1 Hasil TF-IDF

a. Tokenisasi

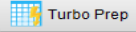
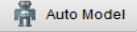
Seperti gambar dibawah ini merupakan hasil dari tokenisasi berupa *wordlist* dimana kata yang tidak diperlukan dihilangkan, pada proses ini juga dapat dilihat seberapa sering suatu kata muncul dalam keseluruhan dokumen seperti pada kata nonton memiliki total *accurences* sebesar 898 dan pada tabel dokumen *occurences* dengan jumlah 801 yang artinya kata nonton muncul sebanyak 898 kali pada 801 dokumen pada keseluruhan data.

Word	Attribute Name	Total Occurences ↓	Document Occurences
nonton	nonton	898	801
https	https	645	634
hero	hero	398	396
weak	weak	398	396
class	class	394	390
tanyarlfes	tanyarlfes	380	380
season	season	316	276
drama	drama	268	259
counter	counter	264	262
uncanny	uncanny	262	262
kdrama	kdrama	233	233
drakor	drakor	223	213
seru	seru	222	210
menfess	menfess	215	215
nder	nder	177	172

Gambar 4.2 tokenisasi

b. Pembobotan kata

Pada proses ini tidak jauh beda dari output dari output dari proses diatas hanya saja pada proses ini sudah tidak berupa wordlist tapi *example* dengan jumlah 4755. Dapat dilihat pada gambar dibawah ini.

Open in  Turbo Prep  Auto Model

Row No.	word	in documents ↓	total
2982	nonton	801	898
1644	https	634	645
1588	hero	396	398
4547	weak	396	398
750	class	390	394
4155	tanyarlfes	380	380
3709	season	276	316
787	counter	262	264
4431	uncanny	262	262
1090	drama	259	268
2007	kdrama	233	233
2611	menfess	215	215
1086	drakor	213	223
3832	seru	210	222
2837	nder	172	177

ExampleSet (4,755 examples, 0 special attributes, 3 regular attributes)

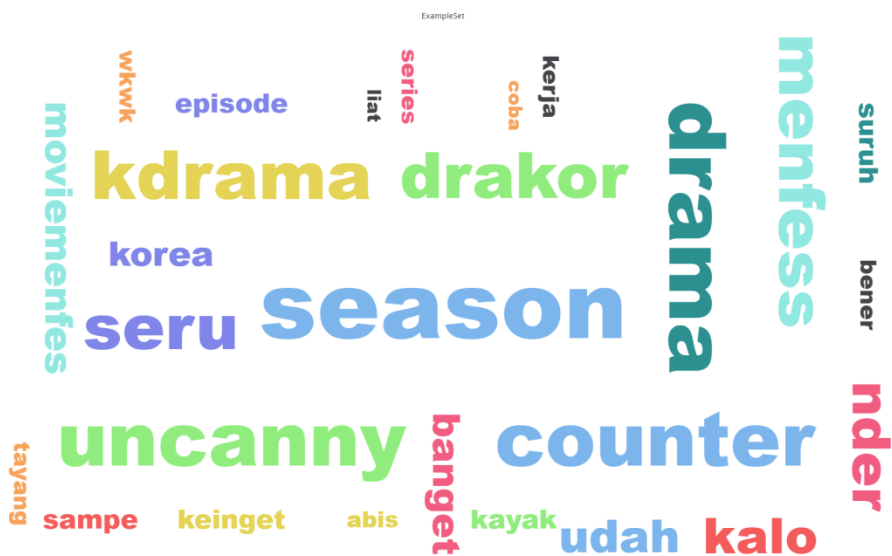
Gambar 4.3 Pembobotan Kata

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

c. Wordcloud

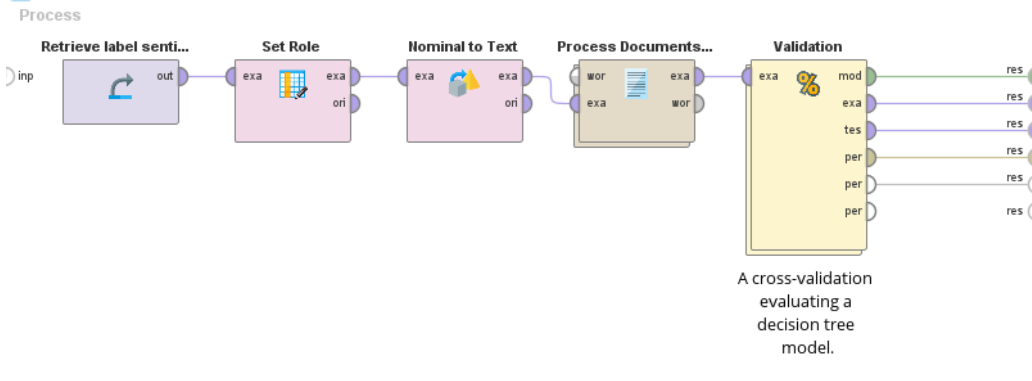
Wordcloud menampilkan frekuensi kemunculan kata yang sangat penting dilakukan dalam analisis sentimen, dimanana *wordcloud* hanya menampilkan data teks dalam bentuk gambar dan memiliki warna sehingga mudah dimengerti dan membantu untuk menemukan kata- kata yang lebih sering muncul dalam suatu data. Semakin sering suatu kata tersebut muncul maka semakin besar pula ukurannya dan begitu pula sebaliknya, semakin jarang kata itu muncul maka akan semakin kecil pula ukuran gambar kata tersebut. Karena ada banyak data maka dipilih 25 data teratas yang paling sering muncul untuk melihat kata apa yang paling populer pada dataset yang telah di crawling.



Gambar 4.4 *Wordcloud*

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suntho Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suntho Jambi



Gambar 4.5 proses dokumen NBC

a. Pengujian data latih 1200

accuracy: 78.95% +/- 2.13% (micro average: 78.95%)

	true positif	true negatif	class precision
pred. positif	1164	90	92.82%
pred. negatif	292	269	47.95%
class recall	79.95%	74.93%	

Gambar 4.6 pengujian 1200 data latih

Hasil perhitungan metode naïve bayes dengan data latih 1200 dan data model 615 mendapatkan hasil true positif 1164, true negatif 269, fals positif 90, fals negatif 292 memiliki akurasi sebesar 78,95%, presisi 92,82%, recall 79,95% seperti yang terdapat pada gambar 4.6.

b. Pengujian data latih 1300

accuracy: 79.12% +/- 1.96% (micro average: 79.12%)

	true positif	true negatif	class precision
pred. positif	1200	85	93.39%
pred. negatif	294	236	44.53%
class recall	80.32%	73.52%	

Gambar 4.7 pengujian 1300 data latih

Sedangkan dengan data latih 1300 dan data model 515 mendapatkan hasil true positif 1200, true negatif 236, fals positif 85 dan fals negatif 296 serta akurasi sebesar 79,95%, presisi 93,39%, recall 80,32% dengan menggunakan metode naïve bayes, sehingga membuktikan bahwa semakin tinggi jumlah data latih maka semakin tinggi pula akurasi yang akan dihasilkan.

c. Perhitungan manual

Dilakukan pengujian sebanyak 2 kali untuk melihat perbandingan seberapa besar akurasi yang didapatkan dengan data model yang berbeda.

Pengujian manual pertama menggunakan data latih 1,176 dari total keseluruhan data 1,815 diambil 639 data model dengan perbandingan 65% : 35% untuk dilakukan pengujian sehingga didapatkan hasil true positif berjumlah 499, true negatif 13, fals positif 85 dan fals negatif 42 data.

Tabel 4.1 perhitungan manual 65% : 35%

	True positif	True negatif
Pred. positif	499	85
Pred. negatif	42	13
Total Data	639	

Pengujian pertama pada perhitungan manual menggunakan rumus confusion matrix, dapat dilihat dibawah ini yang menunjukkan bahwa setelah dilakukan perhitungan manual dengan rumus confusion matrix maka didapatkanlah hasil akurasi sebesar 80,12%, presisi 85,44% dan recall 92,23% seperti dibawah ini.

$$\begin{aligned} \text{Akurasi} &= \frac{499+13}{499+13+85+42} * 100\% \\ &= \frac{512}{639} * 100\% \\ &= 80,12\% \end{aligned}$$

$$\begin{aligned} \text{Presisi} &= \frac{499}{85+499} * 100\% \\ &= \frac{499}{584} * 100\% \\ &= 85,44\% \end{aligned}$$

$$\begin{aligned} \text{Recal} &= \frac{499}{42+499} * 100\% \\ &= \frac{499}{541} * 100\% \\ &= 92,23\% \end{aligned}$$

Pengujian ke 2 menggunakan data latih sebesar 1,302 dari keseluruhan data 1,815 digunakan 513 sebagai data model dengan perbandingan 72% : 28% dari total data keseluruhan sehingga didapatkan true positif 429, true negatif 11, fals positif 40 dan fals negatif 33.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

Tabel 4.2 perhitungan manual 72% : 28%

	True Positif	True Negatif
Pred. Positif	429	40
Pred. Negatif	33	11
Total Data	513	

Pengujian pertama pada perhitungan manual menggunakan rumus confusion matrix, dapat dilihat dibawah ini yang menunjukkan bahwa setelah dilakukan perhitungan manual dengan rumus confusion matrix maka didapatkanlah hasil akurasi sebesar 85,76%, presisi 91,47% dan recall 92,85% seperti dibawah ini.

$$\begin{aligned} \text{Akurasi} &= \frac{429+11}{429+11+40+33} * 100\% \\ &= \frac{440}{513} * 100\% \\ &= 85,76\% \end{aligned}$$

$$\begin{aligned} \text{Presisi} &= \frac{429}{40+429} * 100\% \\ &= \frac{429}{469} * 100\% \\ &= 91,47\% \end{aligned}$$

$$\begin{aligned} \text{Recal} &= \frac{429}{33+429} * 100\% \\ &= \frac{429}{462} * 100\% \\ &= 92,85\% \end{aligned}$$

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

4.3 Split Validation

Pada proses split validation dilakukan dengan menggunakan 2 kali pengujian dengan jumlah data yang berbeda, pengujian pertama dengan perbandingan data 50% : 50% dari total keseluruhan 1,815 data dan didapatkan hasil true positif 575, true negatif 111, fals positif 49 dan fals negatif 172, maka didapatkan akurasi sebesar 75,63%, presisi 92,15% dan recall 76,97% seperti gambar dibawah ini.

accuracy: 75.63%

	true positif	true negatif	class precision
pred. positif	575	49	92.15%
pred. negatif	172	111	39.22%
class recall	76.97%	69.38%	

Gambar 4.8 pengujian data 50% : 50%

Pengujian ke 2 menggunakan split validation dilakukan dengan perbandingan data 70% : 30% dari data keseluruhan 1,815 sehingga didapatkan hasil true positif 342, true negatif 78, fals positif 18, fals negatif 106 dengan akurasi sebesar 77,12%, presisi 95,00% dan recall 76,34% seperti gambar dibawah ini.

accuracy: 77.21%

	true positif	true negatif	class precision
pred. positif	342	18	95.00%
pred. negatif	106	78	42.39%
class recall	76.34%	81.25%	

Gambar 4.9 pengujian data 70% : 30%

Berdasarkan pada pengujian diatas menunjukkan bahwa semakin besar data latih maka semakin besar pula nilai akurasi yang didapatkan, dibuktikan pada perhitungan manual beserta pengujian otomatis *cross validation* dan *split validation* menggunakan rapidminer.

4.4 Perbandingan Pengujian Menggunakan Metode NBC

Tabel 4.3 Perbandingan Pengujian Data

Perbandingan pengujian	accuracy	precision	recall
Pengujian otomatis cross validation 1200 data latih	78,95%	92,82%	79,95%
Pengujian otomatis cross validation 1300 data latih	79,95%	93,39%	80,32%
Pengujian manual 65% : 35%	80,12%	85,44%	92,23%
Pengujian manual 72% : 28%	85,76%	91,47%	92,85%
pengujian otomatis split validation data 50% : 50%	75,63%,	92,15%	76,97%
Pengujian otomatis split validation data 70% : 30%	77,12%,	95,00%	76,34%

Berdasarkan pada tabel 4.3 perbandingan pengujian data, diketahui bahwa menggunakan cross validation dengan data latih 1300 mendapatkan akurasi yang maksimal yaitu 79,95%, sedangkan pada pengujian manual akurasi tertinggi terdapat pada perbandingan data 72% : 28% dengan akurasi 85,76%, pada pengujian menggunakan split validation didapatkan akurasi tertinggi 77,12% menggunakan perbandingan data 70% : 30%.

4.5 Hasil percobaan data baru

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

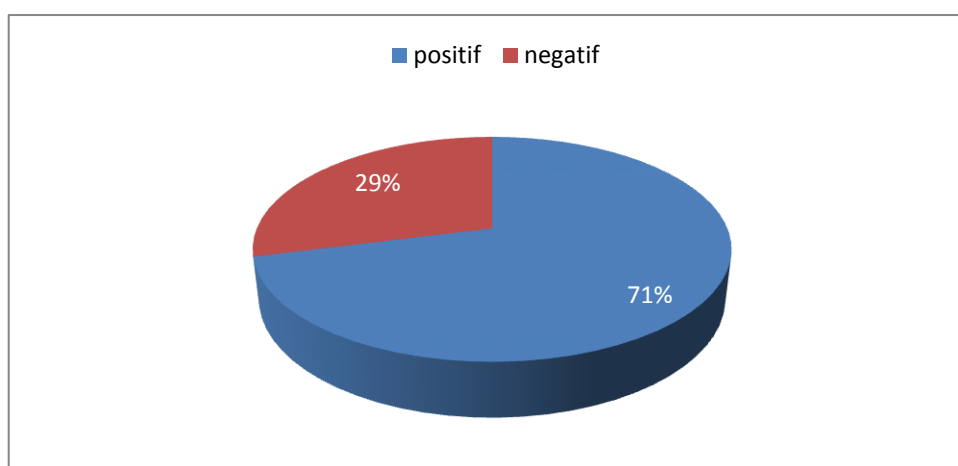
2. Dilarang memperbanyak sebagaian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutaha Jambi

Selanjutnya peneliti melakukan percobaan dengan memasukan data baru sebanyak 10 komentar kedalam data hasil crawling untuk melihat bagaimana hasil prediksi yang akan dihasilkan, dapat dilihat pada tabel 4.4

Tabel 4.4 Percobaan tambah data baru

No.	Text	Prediksi	Hasil
1.	aku suka banget nonton drama korea gendre action	Positif	Benar
2.	healing terbaik setelah lelah beraktifitas seharian adalah nonton drakor	Positif	Benar
3.	alur drama nya kurang menarik	Positif	Benar
4.	aktor dan aktrisnya tidak bisa acting natural jadi terlihat lebay	Positif	Salah
5.	saya tidak jadi nonton karena kurang suka dengan pemeran utamanya	Positif	Salah
6.	kalian harus nonton drakor yang ini ceritanya seru banget	Negatif	Salah
7.	saya tidak suka menonton drakor	Positif	Salah
8.	aku sudah nonton berkali kali tetap ngak bikin bosan	Positif	Benar
9.	alur ceritanya seru dan selalu bikin penasaran disetiap akhir episode	Positif	Benar
10.	bosan nonton drakor terus	Negatif	Benar

Berdasarkan tabel diatas didapatkan hasil prediksi 6 data dengan prediksi benar dan 4 data salah dari 10 data baru yang ditambahkan.



Gambar 4.10 diagram hasil analisis sentimen

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi

2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

Berdasarkan gambar diatas menunjukkan bahwa opini masyarakat mengenai drama korea dengan *gendre action*. Didapatkan komentar positif lebih besar yaitu 71% dibandingkan komentar negatif hanya 29%, sehingga menunjukkan bahwa sentimen positif lebih tinggi dibanding negatif.

4.6 Pembahasan

Dapat dilihat dari hasil diatas menjawab tujuan serta manfaat penelitian, dimana metode naïve bayes dapat digunakan dalam melakukan analisis sentimen dengan studi kasus drama korea menggunakan data latih 1300 didapatkan jumlah akurasi sebesar 79,12% menggunakan *cross validation* dari keseluruhan data 1,815 komentar. Peneliti juga menggunakan perhitungan manual serta menggunakan *split validation* untuk melihat apakah metode naïve bayes dapat digunakan dengan berbagai percobaan yang berbeda. Dari hasil tersebut dapat memberikan informasi mengenai seberapa besar komentar positif dan negatif pada drama korea *gendre action* dengan data yang didapatkan pada tahun 2021 - 2023 menunjukkan komentar positif lebih dominan dibandingkan dengan komentar negatif, sehingga dapat dijadikan rekomendasi baik bagi penonton dalam menentukan *gendre* film yang mereka pilih maupun produser dalam membuat film *gendre action* karena banyaknya komentar positif dibanding negatif mengenai opini masyarakat, seperti hasil yang telah didapatkan oleh peneliti.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunan Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunan Jambi

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa analisis sentiment mampu mengukur opini masyarakat mengenai drama korea menggunakan metode naïve bayes. Berdasarkan dari hasil pengklasifikasian sistem dengan percobaan pengujian data latih 1200 didapatkan akurasi 78,95% sedangkan dengan data latih 1300 didapatkan akurasi sebesar 79,12%. Pengujian menggunakan *cross validation*, peneliti juga melakukan pengujian secara manual menggunakan *confusion matrix* dengan perbandingan data 65% : 35% yaitu menggunakan data latih 1,176 dan data model 639 didapatkan akurasi 80,12% serta pengujian manual ke dua menggunakan data latih 1,302 dan data model 513 dengan perbandingan 72% : 28% mendapatkan akurasi sebesar 85,76%. Sehingga menunjukkan bahwa semakin besar data latih maka semakin besar pula akurasi yang akan di dapatkan. Pengujian selanjutnya menggunakan split validation dengan perbandingan data 50% : 50% didapatkan akurasi sebesar 75,63% dan dilakukan pengujian dengan perbandingan data 70% : 30% didapatkan hasil acurasi sebesar 77,12.

Pengujian menggunakan aplikasi rapidminer mendapatkan akurasi 79,12% dengan keseluruhan data 1815, komentar positif lebih besar yaitu 71% dibandingkan komentar negatif hanya 29%.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi

2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

5.2 Saran

Berdasarkan hasil penelitian ini masih terdapat ketidaksempurnaan sehingga peneliti mengharapkan beberapa hal untuk penelitian selanjutnya:

1. Dapat menggunakan data crawling yang lebih besar sehingga mendapatkan akurasi yang maksimal
2. Dapat melakukan perbandingan metode untuk melihat perbandingan akurasi yang akan didapatkan
3. Dapat dalam melakukan pelabelan manual dibantu oleh pakar bahasa indonesia

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sulthan Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sulthan Jambi

DAFTAR PUSTAKA

- Andika, L.A., Amalia, P. & Azizah, N., 2019, Analisis Sentimen Masyarakat Terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 Pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier, , 2, 1, 34–41.
- Antinasari, P., Perdana, R.S. & Fauzi, M.A., 2017, Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku, , 1, 12, 1733–1741.
- Aprillia, N.H.P. & Listiani, E., 2019, Pengaruh Drama Korea Terhadap Perilaku Imitasi Remaja, *Prosiding Manajemen Komunikasi*, 0, 0, 340–350. <https://karyailmiah.unisba.ac.id/index.php/mankom/article/view/16672>,.
- Ardiansyah, D. & Walim, W., 2018, Algoritma C4.5 Untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa Smp Dengan Menggunakan Aplikasi Rapid Miner, *Jurnal Inkofar*, 1, 2, 5–12.
- Arsi, P. & Waluyo, R., 2021, Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (Svm) Sentiment Analysis On The Discussion Of Relocating I Ndongesia ' S Capital City Using The Support Vector Machine (Svm), , 8, 1, 147–156.
- Bhatia, S., Sharma, M. & Bhatia, K.K., 2018, Sentiment Analysis And Mining Of Opinions, *Studies In Big Data*, 30, May, 503–523.
- Devita, R.N., Herwanto, H.W. & Wibawa, A.P., 2018, Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia, *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5, 4, 427.
- Dicky, 2023, Popularitas Dan Perkembangan Drama Korea Di Indonesia, *Belitong Info*. <https://www.belitonginfo.com/2023/03/11/drakor-Indo-Perkembangan-Dan-Popularitas-Drama-Korea-Di-Indonesia/#Gsc.Tab=0>, Diakses 11 May 2023.
- Džeroski, S., 2008, *Data Mining*,
- Fransiska, S. & Yolanda, 2019, Analisis Sentimen Twitter Untuk Review Film

Menggunakan Algoritma Naive Bayes Classifier (Nbc) Pada Sentimen R Programming, *Jurnal Siliwangi*, 5, 2, 68–71.

Gifari, O.I., Adha, M., Hendrawan, I.R., Freddy, F. & Durrand, S., 2022, Analisis Sentimen Review Film Menggunakan Tf-Idf Dan Support Vector Machine, , 2, 1, 36–40.

Hendrastuty, N., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N., Isnain, A.R., Rahman Isnain, A., Yanti Rahmadhani, A., Styawati, S., Hendrastuty, N. & Isnain, A.R., 2021, Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine, *Jurnal Informatika: Jurnal Pengembangan It*, 6, 3, 150–155. [Http://Situs.Com](http://Situs.Com),.

Herawati, N., Safitri, S.A. & Tandiyayu, D.C., 2021, Drama Korea Dan Penggunaan Media Ilegal Di Kalangan Remaja, *Jurnal Sosial-Politika*, 2, 2, 109–117.

Informatika, J.T., 2022, Jurnal Teknik Informatika, Vol. 14, No. 3, Agustus 2022, , 14, 3, 135–139.

Krisdiyanto, T., 2021, Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan Ppkm Pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers, *Jurnal Coreit: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 7, 1, 32.

Kurniawan, A., Adinugroho, S. & Features, B., 2019, Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes Dan Lexicon Based Features, , 3, 9, 8335–8342.

Ratnawati, F., 2018, Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter, *Inovtek Polbeng - Seri Informatika*, 3, 1, 50.

Sinaga, A., 2017, Implementasi Sentiment Analysis Untuk Menentukan Tingkat Popularitas Tujuan Wisata, , , November, 24–25.

Theo, M., Bangsa, A., Priyanta, S. & Suyanto, Y., 2020, Aspect-Based Sentiment Analysis Of Online Marketplace Reviews Using Convolutional Neural

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunthah Jambi

Network, , 14, 2, 123–134.

Ulfa, M.A., Irmawati, B. & Husodo, A.Y., 2018, Twitter Sentiment Analysis Using Naïve Bayes Classifier With Mutual Information Feature Selection, *Journal Of Computer Science And Informatics Engineering (J-Cosine)*, 2, 2, 106–111.

Verawardina, U., Edi, F. & Watrianthos, R., 2021, Analisis Sentimen Pembelajaran Daring Pada Twitter Di Masa Pandemi Covid-19 Menggunakan Metode Naïve Bayes, , 5, 157–163.

Yahyadi, A. & Latifah, F., 2022, Analisis Sentimen Twitter Terhadap Kebijakan Ppkm Di Tengah Pandemi Covid-19 Menggunakan Mode Lstm, *Journal Of Information System, Applied, Management, Accounting And Research*, 6, 2, 464–471. [Http://Journal.Stmikjayakarta.Ac.Id/Index.Php/Jisamar](http://Journal.Stmikjayakarta.Ac.Id/Index.Php/Jisamar),.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

LAMPIRAN-LAMPIRAN

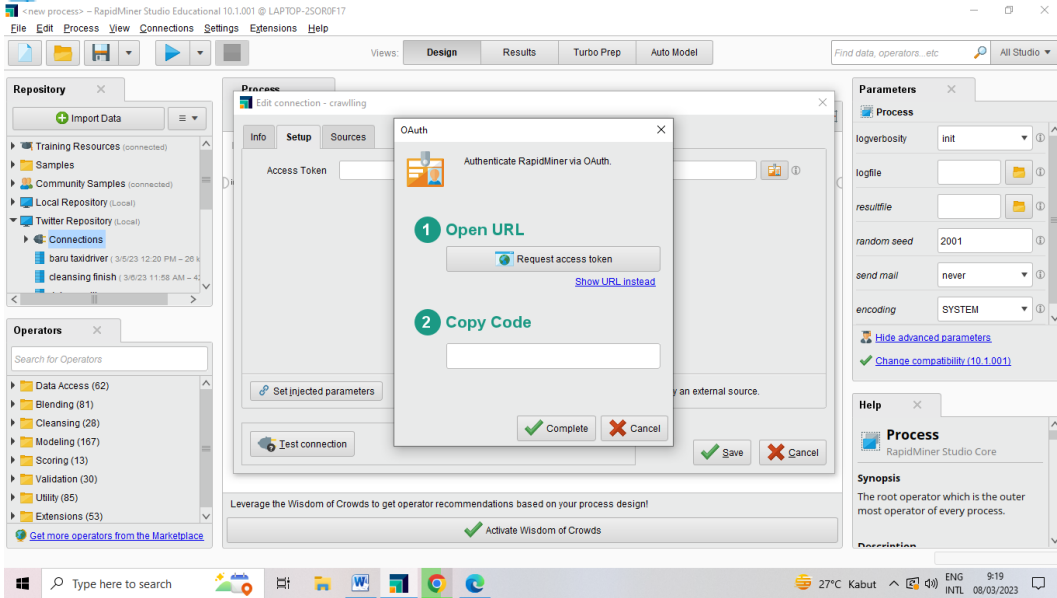


Lampiran 1. Tampilan awal rapidminer

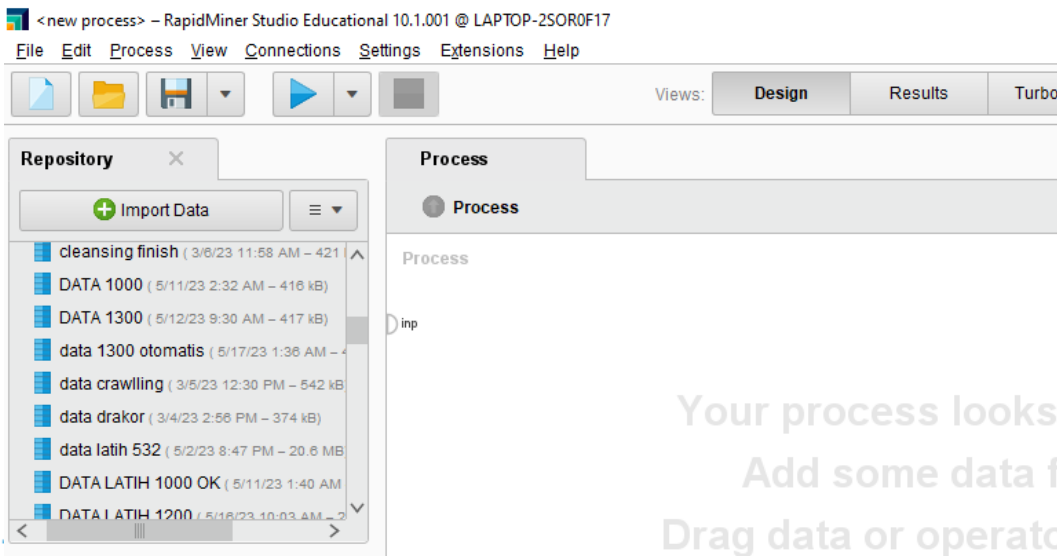
Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sutha Jambi
 2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

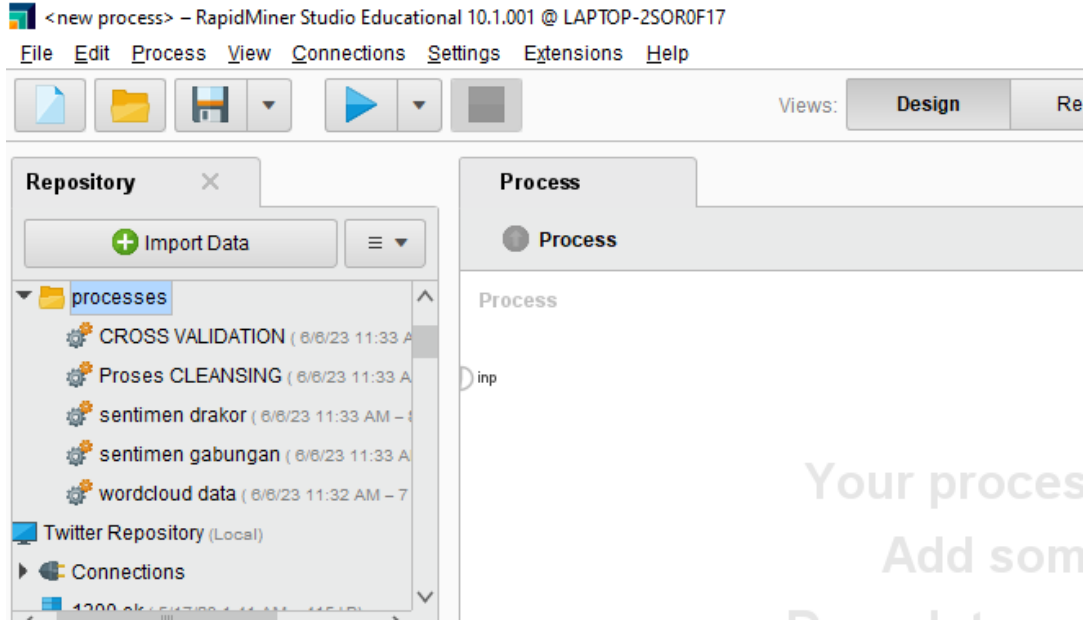


Lampiran 2. Akses token

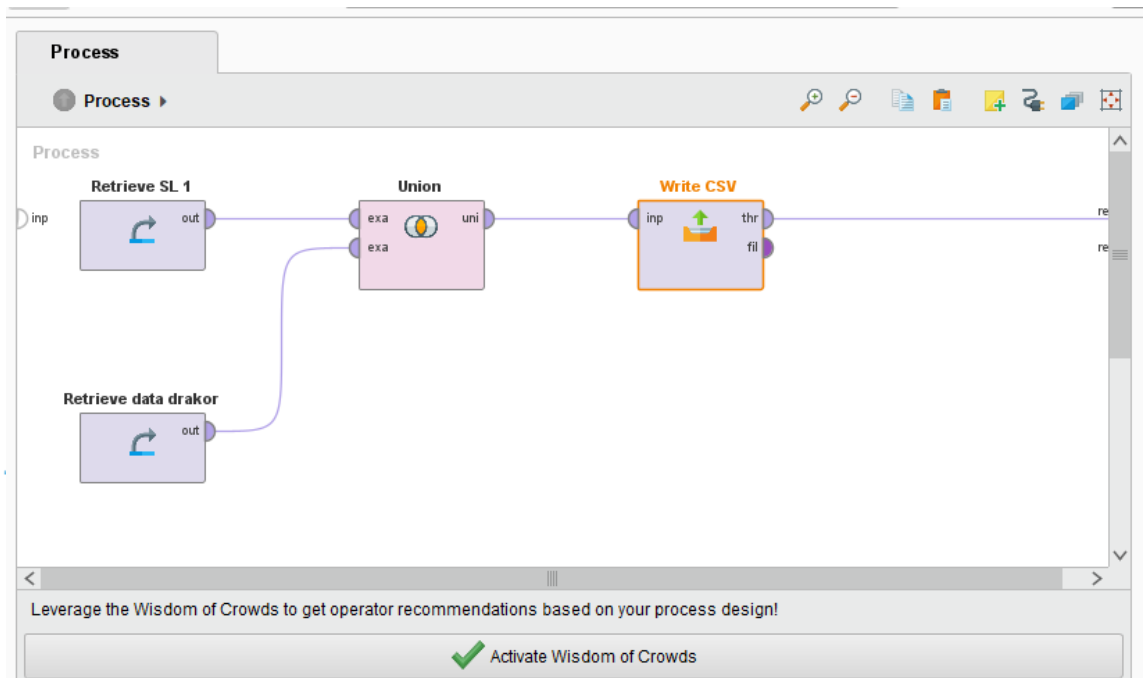


Lampiran 3. Data hasil proses

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
 2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi



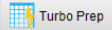

Lampiran 4. Proses pengolahan data



Lampiran 5. proses penggabungan keseluruhan data crawling

Row No.	Text
1	Mmmm... Terakhir namatin taxi driver 2 dan blm ada drama yang bikin excited lagi??? https://t.co/KgICq0iVWF
2	@kdrama_menfess vincenzo, taxi driver sama the glory si keren banget bales dendamnya menurutku
3	@ccandylan di vii kemarin aku pake buaat marathon dr romantic 2, di vii ada taxi driver, duty after school, terus ada kemarin baru release film gore project...
4	taxi driver seru bgt jiiirr
5	@kdrama_menfess Vincenzo, the glory dan taxi driver mantap sih bls dendamnya
6	@kdrama_menfess Taxi driver 1 jga gk kalah seru
7	gue nonton taxi driver dulu baru nonton w two world jadinya kek gbs ngebayangin pak jang (taxi driver) yg super baik itu mukanya jadi karakter jahat T^T
8	RT @EASYEXPRESS0: @kdrama_menfess drakor where stars land - staff airport. jujur ini yg buat aku nonton sih krn temanya yg beda. guess what...
9	yook semangat lanjutin taxi driver 2 nya?? ga nambah2 lg ini ngestuck eps 6 wkwk
10	Nonton taxi driver yang dibayangkan gue mukanya wonwoo gara2 mirip bgt sama lee jehoon ?
11	Akhirnya mood untuk nonton drakor kembali lagi, baru nyelesain kokdu, abis ini taxi driver 2 sisa 5 episode ehehehehe
12	taxi driver seru juga
13	@nafaaaisya Ya allah podo sissss aku tamat nonton taxi driver rasane hampa wkwwkwkksk
14	Ini ada agc duty sama taxi driver. Sinting bgt wkwwkw bkin pucing

Lampiran 6. Hasil penggabungan data

Open in  Turbo Prep  Auto Model Filter (1,865 / 1,865 examples):

Row No.	Text
1	Mmmm Terakhir namatin taxi driver 2 dan blm ada drama yang bikin excited lagi...
2	kdramamenfess vincenzo taxi driver sama the glory si keren banget bales dendamnya menurutku
3	ccandylan di vii kemarin aku pake buaat marathon dr romantic 2 di vii ada taxi driver duty after school terus ada kemarin baru release film gore project wolf...
4	taxi driver seru bgt jiiirr
5	kdramamenfess Vincenzo the glory dan taxi driver mantap sih bls dendamnya
6	kdramamenfess Taxi driver 1 jga gk kalah seru
7	gue nonton taxi driver dulu baru nonton w two world jadinya kek gbs ngebayangin pak jang taxi driver yg super baik itu mukanya jadi karakter jahat TT
8	kdramamenfess drakor where stars land - staff airport. jujur ini yg buat aku nonton sih krn temanya yg beda guess what...
9	yook semangat lanjutin taxi driver 2 nya ga nambah2 lg ini ngestuck eps 6 wkwk
10	Nonton taxi driver yang dibayangkan gue mukanya wonwoo gara2 mirip bgt sama lee jehoon
11	Akhirnya mood untuk nonton drakor kembali lagi, baru nyelesain kokdu abis ini taxi driver 2 sisa 5 episode ehehehehe
12	taxi driver seru juga
13	nafaaaisya Ya allah podo sissss aku tamat nonton taxi driver rasane hampa wkwwkwkksk
14	Ini ada agc duty sama taxi driver Sinting bgt wkwwkw bkin pucing

ExampleSet (1,865 examples, 0 special attributes, 1 regular attribute)

Lampiran 7. Hasil cleansing

Hak Cipta Dilindungi Undang-Undang:

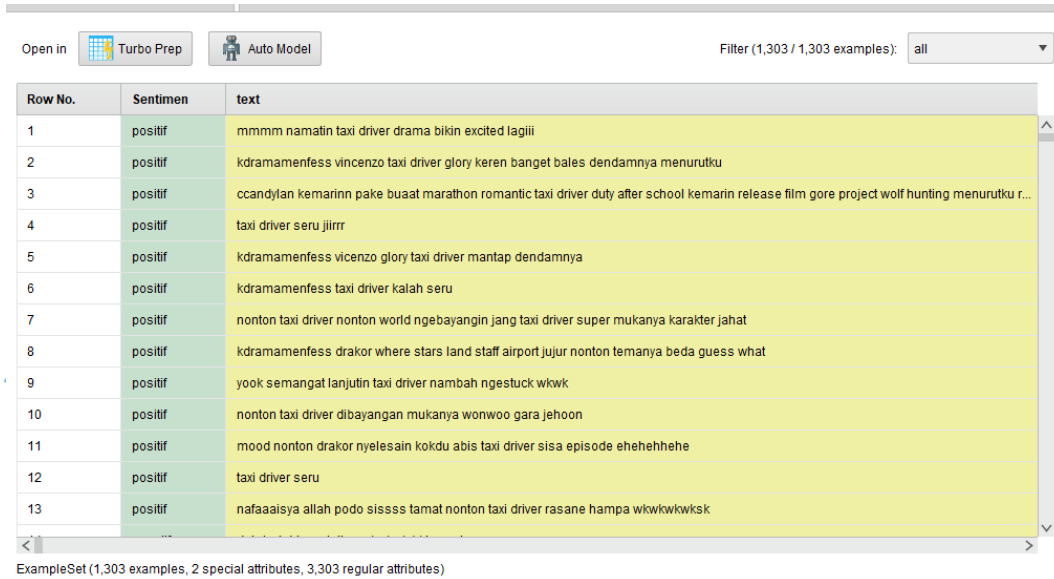
- Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - Pengutipan tidak merugikan kepentingan yang wajar UIN Sultha Jambi
- Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sultha Jambi

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sunha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

	A	B	C	D	E
1	Text	fitri	kurnia	ahmad	Sentimen
2	Mmmm Terakhir namatin taxi driver 2 dan blm ada drama yang bikin excited lagi	positif	positif	positif	positif
3	kdramenfess vincenzo taxi driver sama the glory si keren banget bales dendam	negatif	positif	positif	positif
4	ccandylan di vii kemarin aku pake buat marathon dr romantic 2 di vii ada taxi dr	positif	positif	positif	positif
5	taxi driver seru bgt jiirr	positif	positif	positif	positif
6	kdramenfess Vincenzo the glory dan taxi driver mantap sih bls dendamnya	positif	positif	positif	positif
7	kdramenfess Taxi driver 1 jga gk kalah seru	positif	positif	positif	positif
8	gue nonton taxi driver dulu baru nonton w two world jadinya kek gbs ngebayangin	positif	positif	positif	positif
9	kdramenfess drakor where stars land - staff airport jujur ini yg buat aku nonton	positif	positif	positif	positif
10	yook semangat lanjutin taxi driver 2 nya ga nambah2 lg ini ngestuck eps 6 wkwk	positif	positif	positif	positif
11	Nonton taxi driver yang dibayangkan gue mukanya wonwoo gara2 mirip bgt sama le	positif	positif	positif	positif
12	Akhirnya mood untuk nonton drakor kembali lagi baru nyelesain kokdu abis ini taxi	positif	positif	positif	positif
13	taxi driver seru juga	positif	positif	positif	positif
14	nafaaaisya Ya allah podo sissss aku tamat nonton taxi driver rasane hampa wkwkwk	positif	positif	positif	positif
15	Ini ada agc duty sama taxi driver Sinting bgt wkwkwk bkin pucing	negatif	negatif	negatif	negatif
16	abis tamatin taxi driver 2 SERUUU SGT DITUNGGU S3	positif	positif	positif	positif
17	Mau nonton apalagi ini ya taxi driver udah kelar	positif	positif	positif	positif
18	kamar raph abis mandi trus mam sambil nonton taxi driver 2 NIKMAT ya soalnya bs	positif	positif	positif	positif
19	snowzeax taxi driver sm aos bnyk bgt kya gini	positif	positif	positif	positif
20	kdramenfess yg unik dan berkesan tentu aja kim doki dari taxi driver dan han ge	positif	positif	positif	positif
21	gamalatuconsina lya yg itu kek dia ngomong aja aku dah kesel bgt Nah kebetulan a	negatif	negatif	negatif	negatif
22	adelyine Walaupun seriesanime yang setengah jam tetep bosan Taxi Driver 2 seru	negatif	negatif	negatif	negatif
23	aveothic Sebenarnya ga terlalu suka nonton karena gampang bosan Tapi last time r	negatif	negatif	negatif	negatif
24	yayayaiye kayaknya di taxi driver 3 bakal ada jeon wonwoo sbg taxi driver yg baru	positif	positif	positif	positif
25	kan gue jd nonton taxi driver BYE OL	positif	positif	positif	positif

Lampiran 8. Pelabelan manual di excel



Row No.	Sentimen	text
1	positif	mmmm namatin taxi driver drama bikin excited lagi
2	positif	kdramenfess vincenzo taxi driver glory keren banget bales dendamnya menurutku
3	positif	ccandylan kemarin pake buat marathon romantic taxi driver duty after school kemarin release film gore project wolf hunting menurutku r...
4	positif	taxi driver seru jiirr
5	positif	kdramenfess vincenzo glory taxi driver mantap dendamnya
6	positif	kdramenfess taxi driver kalah seru
7	positif	nonton taxi driver nonton world ngebayangin jang taxi driver super mukanya karakter jahat
8	positif	kdramenfess drakor where stars land staff airport jujur nonton temanya beda guess what
9	positif	yook semangat lanjutin taxi driver nambah ngestuck wkwk
10	positif	nonton taxi driver dibayangkan mukanya wonwoo gara2 jehoon
11	positif	mood nonton drakor nyelesain kokdu abis taxi driver sisa episode ehehehehe
12	positif	taxi driver seru
13	positif	nafaaaisya allah podo sissss tamat nonton taxi driver rasane hampa wkwkwkwk

ExampleSet (1,303 examples, 2 special attributes, 3,303 regular attributes)

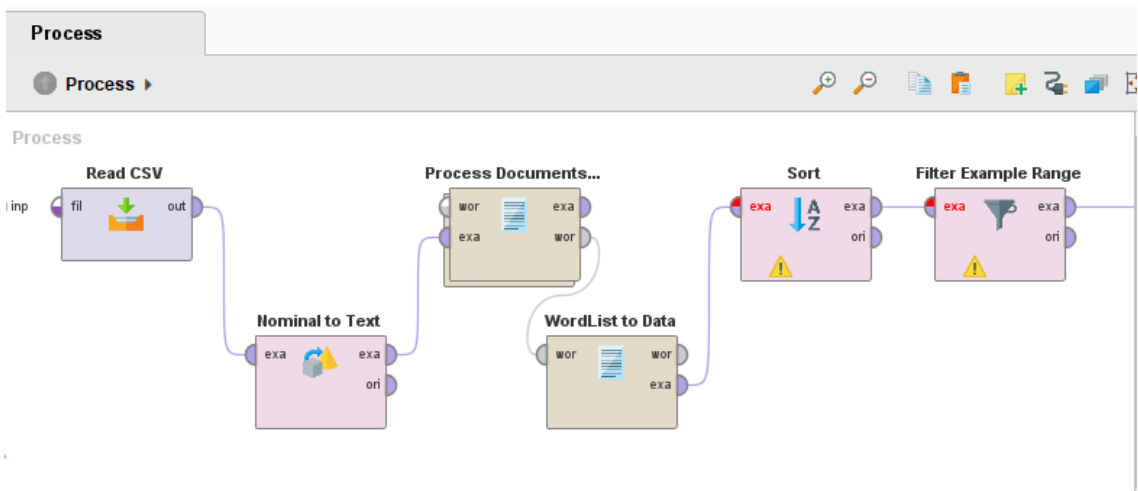
Lampiran 9. Data latih

Open in Turbo Prep Auto Model Filter (1,814 / 1,814 examples): all

Row No.	Sentimen	prediction(S...	confidence(...	confidence(...	text	aaaa	aaaaaaaa	aaaaaaaaaaaa	aaaaaaaa
832	?	positif	1	0	uncanny cou...	0	0	0	0
833	?	positif	1	0	sagipridee ta...	0	0	0	0
834	?	negatif	0	1	cepatlah taya...	0	0	0	0
835	?	positif	1	0	claramdhni k...	0	0	0	0
836	?	positif	1	0	uncanny cou...	0	0	0	0
837	?	positif	1	0	kdramamenf...	0	0	0	0
838	?	positif	1	0	onyourleocat ...	0	0	0	0
839	positif	positif	1	0	mmmm nam...	0	0	0	0
840	positif	positif	1	0	kdramamenf...	0	0	0	0
841	positif	positif	1	0	ccandylan ke...	0	0	0	0
842	positif	positif	1	0	taxi driver ser...	0	0	0	0
843	positif	positif	1	0	kdramamenf...	0	0	0	0
844	positif	positif	1	0	kdramamenf...	0	0	0	0
845	positif	positif	1	0	nonton taxi dr...	0	0	0	0

ExampleSet (1,814 examples, 5 special attributes, 4,229 regular attributes)

Lampiran 10. Pelabelan otomatis data model



Lampiran 11. Proses wordcloud

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sulha Jambi
 2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sulha Jambi

	text	Sentimen	TP	TN	FP	FN
1	text					
2	hayyujin iyakann libur	positif	P			
3	kelar nonton taxi driver endingnya greget wkwk	negatif		N		
4	nctdreamina denger suara jaemin langsung taxi driver wkwk	positif	P			
5	seru bgtt taxi driver	positif	P			
6	taxi driver tamat bingung nnton	positif	P			
7	nonton taxi driver episode ketar ketir	negatif			P	
8	tamatin taxi driver tamatin tugas	positif	P			
9	kmrn binge watch taxi driver sampe lupa tugas	positif	P			
10	wkwkwkwk anime wanpis kimestu yaiba tengoku daimyoku jigokuraku stone vinland saga eden ze	positif	P			
11	wonu rambut cepak jelmaan doki taxi driver	positif	P			
12	lebaran dirumah sakit diisi marathon movie dinetflix quite western front taxi driver kill boksoon n	positif	P			
13	jehoon beranda tiktok mulu njir cakep banget penasaran taxi driver seru	positif	P			
14	taxi driver keknya emosional sedi efek	positif				N
15	butuh taxi driver jabanin dehkh saking serunya	positif	P			
16	kikufLOWa nnattawin adik wang cerita taxi driver	positif	P			
17	esom cakep taxi driver ampunnn	positif	P			
18	maubngasih spoiler taxi driver doki iwir	positif	P			
19	bingung nontoj drama taxi driver tamat hampa	positif	P			
20	jaeqha hyung taxi driver season wvakakkakak	positif	P			
21	taxi driver netflix	positif	P			
22	minseo lulusan taxi driver	negatif			P	
23	rekomendasiin drakor kasusnya beda beda taxi driver first responders thank before	positif	P			
24	bosen drama bgttttt semangat gara duty after school tamat gamau pisah ajaaa bosen taxi driver bo	positif				N
25	saran drakor ramee polll nonton taxi driver	positif	P			

Lampiran 12. Sentimen manual 639 data

	text	Sentimen	TP	TN	FP	FN
1	text					
2	hayyujin iyakann libur	positif	P			
3	kelar nonton taxi driver endingnya greget wkwk	negatif		N		
4	nctdreamina denger suara jaemin langsung taxi driver wkwk	positif	P			
5	seru bgtt taxi driver	positif	P			
6	taxi driver tamat bingung nnton	positif	P			
7	nonton taxi driver episode ketar ketir	negatif			P	
8	tamatin taxi driver tamatin tugas	positif	P			
9	kmrn binge watch taxi driver sampe lupa tugas	positif	P			
10	wkwkwkwk anime wanpis kimestu yaiba tengoku daimyoku jigokuraku stone vinland saga eden ze	positif	P			
11	wonu rambut cepak jelmaan doki taxi driver	positif	P			
12	lebaran dirumah sakit diisi marathon movie dinetflix quite western front taxi driver kill boksoon n	positif	P			
13	jehoon beranda tiktok mulu njir cakep banget penasaran taxi driver seru	positif	P			
14	taxi driver keknya emosional sedi efek	positif				N
15	butuh taxi driver jabanin dehkh saking serunya	positif	P			
16	kikufLOWa nnattawin adik wang cerita taxi driver	positif	P			
17	esom cakep taxi driver ampunnn	positif	P			
18	maubngasih spoiler taxi driver doki iwir	positif	P			
19	bingung nontoj drama taxi driver tamat hampa	positif	P			
20	jaeqha hyung taxi driver season wvakakkakak	positif	P			
21	taxi driver netflix	positif	P			
22	minseo lulusan taxi driver	negatif			P	
23	rekomendasiin drakor kasusnya beda beda taxi driver first responders thank before	positif	P			
24	bosen drama bgttttt semangat gara duty after school tamat gamau pisah ajaaa bosen taxi driver bo	positif				N
25	saran drakor ramee polll nonton taxi driver	positif	P			

Lampiran 13. Sentimen manual 513 data

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penerbitan, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sulha Jambi
 2. Dilarang memperbanyak sebagai dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sulha Jambi

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian dan atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Sultha Jambi
2. Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sultha Jambi

positif	..	suka banget nonton drama korea gendre action
positif	..	healing terbaik lelah beraktifitas sehari nonton drakor
positif	..	alur drama menarik
positif	..	aktor aktrisnya acting natural lebay
positif	..	nonton suka pemeran utamanya
negatif	..	nonton drakor ceritanya seru banget
positif	..	suka menonton drakor
positif	..	nonton berkali kali ngak bikin bosan
positif	..	alur ceritanya seru bikin penasaran disetiap episode
negatif	..	bosan nonton drakor

Lampiran 14. Hasil prediksi penambahan data baru

	A	B	C	D	E	F	G	H
1	ada	apalagi	banyak	berapa	bermula	bukan	di	dikatakannya
2	adalah	apatah	bapak	berapakah	bersama	bukankah	dia	dikerjakan
3	adanya	artinya	baru	berapalah	bersama-sama	bukanlah	diakhiri	diketahui
4	adapun	asal	bawah	berapapun	bersiap	bukannya	diakhirinya	diketuainya
5	agak	asalkan	beberapa	berarti	bersiap-siap	bulan	dialah	dikira
6	agaknya	atas	begini	berawal	bertanya	bung	diantara	dilakukan
7	agar	atau	beginian	berbagai	bertanya-tanya	cara	diantaranya	dilalui
8	akan	ataukah	beginikah	berdatangan	berturut	caranya	diberi	dilihat
9	akankah	ataupun	beginilah	beri	berturut-turut	cukup	diberikan	dimaksud
10	akhir	awal	begitu	berikan	berturut	cukupkah	diberikannya	dimaksudkan
11	akhiri	awalnya	begitukah	berikut	berujar	cukuplah	dibuat	dimaksudkannya
12	akhirnya	bagai	begitulah	berikutnya	berupa	cuma	dibuatnya	dimaksudnya
13	aku	bagaimana	begitupun	berjumlah	besar	dahulu	didapat	diminta
14	akulah	bagaimana	bekerja	berkali-kali	betul	dalam	didatangkan	dimintai
15	amat	bagaimanakah	belakang	berkata	betulkah	dan	digunakan	dimisalkan
16	amatlah	bagaimanapun	belakangan	berkehendak	biasa	dapat	diibaratkan	dimulai
17	anda	bagi	belum	berkeinginan	biasanya	dari	diibaratkannya	dimulailah
18	andalah	bagian	belumlah	berkenaan	bila	daripada	diingat	dimulainya
19	antar	bahkan	benar	berlainan	bilakah	datang	diingatkan	dimungkinkan
20	antara	bahwa	benarkah	berlalu	bisa	dekat	diinginkan	dini
21	antaranya	bahwasanya	benarlah	berlangsung	bisakah	demi	dijawab	dipastikan
22	apa	baik	berada	berlebihan	boleh	demikian	dijelaskan	diperbuat
23	apaan	bakal	berakhir	bermacam	bolehkah	demikianlah	dijelaskannya	diperbuatnya
24	apabila	bakalan	berakhirilah	bermacam-macam	bolehlah	dengan	dikarenakan	dipergunakan
25	apakah	balik	berakhirnya	bermaksud	buat	depan	dikatakan	diperkirakan

Lampiran 15. Kalimat stopwoard