

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 TERHADAP PINDAHNYA IBU KOTA INDONESIA DENGAN MENGGUNAKAN KLASIFIKASI NAÏVE BAYES

SKRIPSI



Oleh:

KURNIA ARDIANSYAH LUBIS

701190050

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI**

UNIVERSITAS ISLAM NEGERI SULTHAN THAHASAIFUDDIN 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 TERHADAP PINDAHNYA IBU KOTA INDONESIA DENGAN MENGGUNAKAN KLASIFIKASI NAÏVE BAYES

SKRIPSI



Oleh:

KURNIA ARDIANSYAH LUBIS

701190050

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

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI**

UNIVERSITAS ISLAM NEGERI SULTHAN THAHASAIFUDDIN JAMBI

2023

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, 05 Juni 2023

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

Assalamualaikum.wr.wb

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

Judul : Analisis Sentiment Opini Masyarakat Terhadap Pindahnya Ibu Kota Indonesia Dengan Menggunakan Klasifikasi Naïve Bayes

Nama : Kurnia Ardiansyah lubis
NIM : 701190050
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,

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

Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian 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 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 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, 01 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 Sentiment Opini Masyarakat Terhadap Pindahnya Ibu Kota Indonesia Dengan Menggunakan Klasifikasi Naïve Bayes

Nama : Kurnia Ardiansyah lubis
NIM : 701190050
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 II,

Andreo Yudertha, ST., M. Eng
NIP. 198907262020121006

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-654 /D. VII/PP.009/08/2023

Skripsi/ Tugas Akhir dengan Judul : "Analisis Sentiment Opini Masyarakat Terhadap Pindahanya Ibu kota Indonesia dengan menggunakan klasifikasi Naive Bayes"

Yang dipersiapkan dan disusun oleh

Nama : Kurnia Ardiansyah Lubis

NIM : 701190050

Telah dimunaqasahkan pada : 26 Juni 2023

Nilai Munaqasah : 89,19

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


Yefix Ramadhani, M. Kom
NIDN.2014068504

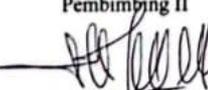
Penguji II


Mutamassikin, M. Kom
NIP.199004092019031014

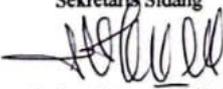
Pembimbing I


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

Pembimbing II


Andreo Yudertha, ST., M. Eng
NIP. 198907262020121006

Sekretaris Sidang


Andreo Yudertha, ST., M. Eng
NIP. 198907262020121006

Jambi, 14-08- 2023

Dekan Fakultas Sains dan Teknologi



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

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

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 yakni abang dan kakak, Terimakasih atas dukungan dan do'anya sehingga saya dapat menyelesaikan studi pendidikan di perguruan tinggi ini, sahabat-sahabat seperjuangan khususnya Program Studi Sistem Informasi, umumnya Fakultas Sains dan Teknologi UIN Sulthan Thaha Saifuddin Jambi serta orang-orang yang mencintai ilmu pengetahuan. Dan terimah 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 :

لَا يُكَلِّفُ اللَّهُ نَفْسًا إِلَّا وُسْعَهَا

Allah tidak membebani seseorang itu melainkan sesuai dengan kesanggupannya (Q.S. Al Baqarah : 286).

“Wahai orang-orang yang beriman, Jauhilah banyak dari prasangka, sesungguhnya sebagian prasangka itu dosa dan janganlah kamu mencari-cari kesalahan orang lain dan janganlah ada di antara kamu yang menggunjing sebagian yang lain (Surat Al Hujarat ayat 12)

Ketika kamu memutuskan untuk melangkah, itu artinya kamu siap untuk terjatuh dan ketika kamu sudah pandai melangkah ingatla kamu pernah terjatuh. “Tidak ada usaha yang sia-sia yang sia-sia itu hidupmu yang enggan berusaha. Mencoba memang tidak membuatmu 100% berhasil tetapi tidak mencoba sudah dipastikan gagal”.

Kurnia Ardiansyah I

- 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

PERNYATAAN KEASLIAN NASKAH

PERNYATAAN KEASLIAN NASKAH

Yang bertandatangan dibawah ini:

Nama : Kurnia Ardiansyah Lubis

NIM : 701190050

Jurusan : Sistem Informasi

Menyatakan bahwa skripsi yang berjudul:

ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP PINDAHNYA IBU KOTA INDONESIA DENGAN MENGGUNAKAN KLASIFIKASI NAÏVE BAYES

Merupakan skripsi yang saya susun sebagai syarat untuk memperoleh gelar Sarjana dari Fakultas Sains dan Teknologi UIN 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 saya sendiri atau terindikasi adanya unsur plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi sesuai dengan peraturan dan perundang- undangan yang berlaku.

Jambi, Juni 2023



Kurnia Ardiansyah Lubis
Kurnia Ardiansyah Lubis

701190050

Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian 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 atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sultha 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. Bapak MHD. Theo Ari Bangsa, M. Cs. Selaku Pembimbing I dan Bapak Andreo Yudertha, ST., M. Eng. 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.

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 Sulthan Thaha Saifuddin Jambi

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

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, Juni 2022
Penulis

Kurnia Ardiansyah Lubis
701190050

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



ABSTRACT

Berdasarkan isu yang di bahas pemerintah pada tahun 2019, timbulah beragam tweet yang bermunculan seperti #ibukotapindah, #pemindahanibukota. Dari banyaknya respon masyarakat pengguna Twitter terhadap isu yang menjadi salah satu trending topik tersebut maka perlu diadakannya analisis sentimen untuk mengetahui apakah masyarakat Indonesia merespon isu ini dengan respon positif, netral, atau malah negatif. Sehingga dengan adanya penelitian ini kita bisa melihat gambaran pendapat masyarakat terkait kebijakan pemerintah yang merencanakan pemindahan ibu kota negara Indonesia dan juga bisa jadi bahan evaluasi pemerintah dalam pengambilan keputusan, dengan adanya analisis sentimen ini memudahkan kita untuk melihat informasi data di media sosial yang besar dengan cepat. Pengumpulan data pada penelitian ini adalah dengan melakukan penarikan data set dari media sosial. Data set yang di tarik merupakan tweet dari masyarakat dan tweet itu akan di kelompokkan dengan label positif dan negatif menggunakan metode Naïve Bayes Clasifier (NBC). Analisis dengan metode Naïve Bayes di mulai dengan crawling data dari Twitter pada tanggal 6 maret 2023 sebanyak 4533 data mentah dengan 2 kata kunci IKN dan IBU KOTA PINDAH. Setelah crawling selanjutnya pembersihan data atau preprocessing yang terdiri dari cleansing, stopword, dan casefolding sehingga menyisakan data yang bersih sebanyak 1449 data. data di bagi menjadi dua yaitu data latih dan data uji data latih digunakan untuk memberikan pengetahuan kepada rapid miner dalam menggolongkan suatu data sehingga data uji tersebut dapat terprediksi oleh aplikasi Rapid Miner, dalam pengujian manual pengujian pertama (69%:31%) di dapatkan Acuraccy 74.61% dan yang kedua (83%:17%) didapatkan Acuraccy 76.30% selanjutnya pengujian oleh sistem dengan perbandingan (50%:50%) di dapatkan Acuraccy 70.30% dan yang kedua (70%:30%) data uji, didapatkan Acuraccy sebesa 73.79%. Atas kebijakan pemerintah yang ingin memindahkan ibu kota indonesia menimbulkan beragam respon positif dan negatif dari masyarakat, sentiment yang paling banyak yaitu sentiment positif sebesar 1073 (74%), sentiment negatif ditemukan sebanyak 376 (26%). Saran untuk penelitian selanjutnya untuk menambahkan sentiment netral pada penelitian selanjutnya.

Kata kunci : Ibu kota Indonesia, *Data Mining*, *Naive Bayes*

ABSTRACT

Based on the issues discussed by the government in 2019, various tweets emerged, such as #capitalmove and #relocationofthecity. From the many responses of the Twitter user community to the issue, which is one of the trending topics, it is necessary to conduct a sentiment analysis to find out whether the Indonesian people respond to this issue with a positive, neutral, or even negative response. So that with this research we can see an overview of public opinion regarding government policies that plan to relocate the capital city of Indonesia and can also be used as material for evaluating the government in making decisions, this sentiment analysis makes it easier for us to see data and information on large social media quickly. Collecting data for this study is done by pulling data sets from social media. The data set that is pulled is tweets from the community, and the tweets will be grouped with positive and negative labels using the Naive Bayes Classifier (NBC) method. Analysis using the Naive Bayes method began with crawling data from Twitter on March 6, 2023, including as many as 4533 raw files with the keywords *IKN* and *MOVING CAPITAL CITY*. After crawling, data cleaning or preprocessing consists of cleansing, stopwords, and casefolding, leaving 1449 clean records. The data is divided into two categories: training data and test data. Training data is used to provide knowledge to the Rapid Miner in classifying data so that the test data can be predicted by the Rapid Miner application. In the first manual test (69%: 31%), you get an accuracy of 74.61%, and in the second (83%: 17%), you obtain an accuracy of 76.30%. Then, testing by the system with a comparison (50%: 50%) obtained an accuracy of 70.30%, and the second (70%: 30%) test data obtained an accuracy of 73.79%. Regarding the government's policy of wanting to move the capital city of Indonesia, it elicited various positive and negative responses from the public; the most frequent sentiments were positive sentiments of 1073 (74%), and negative sentiments of 376 (26%). Suggestions for further research to add neutral sentiment to further research.

Kata kunci : *Capital of Indonesia, Data mining, Naive Bayes*

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

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN.....	ii
NOTA DINAS 1	ii
NOTA DINAS 2	iii
HALAMAN PENGESAHAN.....	iv
PERSEMBAHAN.....	v
MOTO.....	vi
PERNYATAAN KEASLIAN NASKAH.....	vii
KATA PENGANTAR	viii
ABSTRACT.....	x
<i>ABSTRACT</i>	xi
DAFTAR ISI.....	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL.....	xv
DAFTAR LAMPIRAN.....	xvi
DAFTAR RIWAYAT HIDUP.....	xvii
BAB I	1
PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
BAB II	7
LANDASAN TEORI.....	7
2.1 Kajian Pustaka.....	7
2.1.1 Analisis.....	7
2.1.2 Analisis Sentiment.....	7
2.1.3 <i>Naïve Bayes Clasifier (NBC)</i>	8
2.1.4 <i>Text Mining</i>	11

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

2.1.5	Twitter	11
2.1.6	<i>Term Frecuency-Inverse Document Frecuency (TF-IDF)</i>	12
2.1.7	Confusion Matrik.....	14
2.2	<i>Analisis Sentiment</i> Twitter	15
2.3	Penelitian Yang Relevan	17
BAB III.....		21
METODE PENELITIAN.....		21
3.1	Tahapan Penelitian	21
3.1.1	Data <i>Crawling</i>	21
3.1.2	<i>Preprocesing</i>	24
3.1.3	Ekstrasi Fitur.....	30
3.1.4	Pelabelan Manual	31
3.1.5	<i>Clasificasion</i> Naïve Bayes	32
3.2	Evaluasi	36
BAB IV		38
HASIL DAN PEMBAHASAN.....		38
4.1	Pembahasan dan hasil dari Ekstrasi Fitur	38
4.2	Hasil permodelan Metode <i>Naïve Bayes</i>	42
4.3	Hasil pengujian <i>Split Validation</i> oleh sistem	44
4.4	Hasil pengujian <i>Cross Validation</i>	45
4.5	Hasil percobaan data baru	46
4.6	Hasil Evaluasi.....	47
BAB V.....		49
KESIMPULAN DAN SARAN.....		49
5.1	Kesimpulan	49
4.6	Saran.....	50
DAFTAR PUSTAKA		51
PAMPIRAN.....		53

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 sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suttha Jambi

DAFTAR GAMBAR

Gambar 2.1 Contoh kata-kata <i>stopword</i>	17
Gambar 3.1 Tahapan Penelitian Menggunakan Klasifikasi Naïve Bayes.....	21
Gambar 3.2 Tahapan pengkoneksian Rapid Miner ke Twitter	22
Gambar 3.3 Tahapan <i>Crawling</i>	23
Gambar 3.4 Tahapan pembersihan data	24
Gambar 3.5 Membuat model Naïve Bayes pada Rapid Miner	33
Gambar 3.6 Data tahap <i>Filter Examples</i> dengan pengaturan (<i>sentiment, is not missing</i>).....	34
Gambar 3.7 Proses pelabelan dengan model dan data latih Naïve Bayes.....	34
Gambar 3.8 Data tahap <i>Filter Examples</i> dengan pengaturan (<i>sentiment, no missing</i>)	35
Gambar 3.9 Ilustrasi <i>Split Validation</i>	37
Gambar 4.1 Proses TF-IDF di Rapid Miner.....	38
Gambar 4.2 Proses di dalam operator TF-IDF.....	39
Gambar 4.3 Gambar hasil dari data yang di <i>Tokenize</i>	39
Gambar 4.4 Contih kata-kata hasil pembobotan TF-IDF.....	40
Gambar 4.5 kata-kata yang sudah di <i>Stopword</i> dan <i>Filter Token</i>	41
Gambar 4.6 <i>Wordcloud</i>	42
Gambar 4.7 Pengujian data 50% : 50%	44
Gambar 4.8 Pengujian data 70% : 30%	45
Gambar 4.9 Pengujian <i>Cross Validation</i>	46
Gambar 4.10 Jumlah <i>sentiment</i> positif dan negatif.....	46

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

DAFAR TABEL

Tabel 2.1 Contoh data sampel perhitungan metode <i>Naïve Bayes</i>	8
Tabel 2.2 Probalitas kategori.....	10
Tabel 2.3 Hasil klasifikasi <i>Naïve Bayes</i>	11
Tabel 2.4 Proses TF-IDF.....	13
Tabel 2.5 Proses <i>Cleansing</i>	15
Tabel 2.6 Proses <i>Case Folding</i>	16
Tabel 2.7 Proses <i>Tokenizing</i>	16
Tabel 2.8 Tabel pembandingan peneliti terdahulu.....	19
Tabel 3.1 Tampilan beberapa data mentah hasil <i>Crawling</i>	23
Tabel 3.2 <i>Cleansing RT</i>	25
Tabel 3.3 <i>Cleansing URL</i>	26
Tabel 3.4 <i>Cleansing HASTAGE</i>	27
Tabel 3.5 <i>Cleansing MENTION</i>	28
Tabel 3.6 <i>Cleansing SIMBOL</i>	29
Tabel 3.7 Data pelabelan manual	31
Tabel 3.8 Ilustrasi skema <i>5-fold Cross Validation</i>	37
Tabel 4.1 Hasil uji data uji 449 dan data latih 1000.....	43
Tabel 4.2 Hasil uji data uji 249 dan data latih 1200.....	44
Tabel 4.3 Percobaan tambah data baru	46
Tabel 4.4 Hasil Evaluasi	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 logo dari aplikasi Rapid Miner.....	53
Lampiran 2 Tampilan awal Rapid Miner	53
Lampiran 3 Tampilan proses-proses pada Rapid Miner	54
Lampiran 4 Tampilan data model dan data latih pada Rapid Miner.....	54
Lampiran 5 Tampilan beberapa data kotor yang di tarik dari Twitter	55
Lampiran 6 Tampilan beberapa data yang sudah melalui tahap <i>Preprocessing</i> ...	55
Lampiran 7 Data pelabelan 3 orang	56
Lampiran 8 Gambaran untuk mencari nilai <i>accuracy</i> manual dalam data.....	56
Lampiran 9 Tahapan <i>Split Validation</i> pada Rapid Miner	57
Lampiran 10 Tahapan <i>Cross Validation</i> Pada Rapid Miner	57
Lampiran 11 Kamus <i>Stopword</i>	57

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

DAFTAR RIWAYAT HIDUP

A. Identitas Diri



Nama : Kurnia Ardiansyah Lubis
Jenis Kelamin : Laki-Laki
Tempat/Tanggal Lahir : Pauh / 26 Oktober 2001
Alamat : Sarolangun
Email : kurniaardiansyah81@gmail.com
Hp : 082289894870

B. Riwayat Pendidikan

1. MIN 3 Sarolangun Tahun 2013
2. MTS Negeri 6 Sarolangun Tahun 2016
3. SMA Negeri 3 Sarolangun Tahun 2019

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Provinsi DKI Jakarta adalah ibu kota dari Negara Republik Indonesia. Dipilihnya Jakarta sebagai ibu kota. Provinsi DKI Jakarta menanggung fungsi yang sangat besar salah satunya fungsi ekonomi dan juga sosial politik yang sangat besar (Prabowo, 2021).

Menurut Bappenas sebanyak 56,56% dari total penduduk Indonesia yang tinggal atau menetap di pulau Jawa. Sedangkan 7,33% saja penduduk yang berada di pulau Kalimantan Indonesia. Dari data persentase tersebut terlihat jelas bahwa pulau Jawa sangat padat penduduk, maka pemerintah berinisiatif untuk memindahkan ibu kota ke Kalimantan. Selain itu kontribusi ekonomi terhadap Produk Domesti Bruto pulau Jawa masih menjadi pusat perekonomian di Indonesia. Pemerintah juga mempertimbangkan ketersediaan air bersih dan konversi lahan yang kurang merata (Rochmanto & Hermanto, 2020). Hampir setiap tahun terjadi banjir yang besarnya bervariasi. Banjir yang terjadi tahun 2007 merupakan yang terbesar, hampir mencakup 70% wilayah Jakarta. Akibat banjir 2007 tersebut menimbulkan pemikiran atau gagasan untuk memindahkan Ibu Kota Jakarta (Yahya, 2018).

Perencanaan tentang pemindahan Ibu kota sebenarnya sudah dibahas beberapa tahun yang lalu, namun pada tahun 2019 rencana ini baru mulai terealisasikan. Adapun tujuan utama dari pemerintah mengenai rencana

pemindahan ini yaitu agar pembangunan di Indonesia merata, mengurangi beban Ibu kota di Jakarta serta mengubah pola pikir masyarakat Indonesia yang mulanya Jawa sentris menjadi Indonesia sentris. Dari rencana pemerintah atas pemindahan ibukota ini menimbulkan banyak pro dan kontra bagi penduduk Indonesia di berbagai media sosial. Masyarakat berpendapat dari anggaran pemindahan Ibu kota baru lebih baik digunakan untuk hal yang lebih penting seperti mengurangi bencana *covid-19* yang terjadi di Indonesia (Agustina & Hendry, 2021).

Atas rencana pemerintah tersebut untuk memindahkan ibu kota, hal ini menimbulkan reaksi positif dan negatif dari masyarakat Indonesia melalui beberapa media sosial seperti Twitter dan Facebook, tidak hanya masyarakat beberapa toko publik pun juga memberikan pro dan kontra terhadap rencana pemerintah tersebut. Beberapa pihak mengharapkan kepada pemerintah untuk lebih mengkaji lagi rencana tersebut dari waktu, tempat dan mekanisme pemindahannya. Banyak pihak yang keberatan dan tidak setuju atas pemerintah untuk memindahkan ibu kota Indonesia. Reaksi atau respon yang berbeda tersebut perlu dilihat dengan gambaran umum bagaimana pendapat dan respon yang diberikan masyarakat apakah menyetujui kebijakan pemerintah atas pemindahan ibu kota atau tidak.

Salah satu platform media sosial seperti Twitter dapat kita perhatikan khusus karena para pemakainya sangat dengan mudah mendapatkan maupun mengunggah informasi tentang pendapat mereka tentang masalah tertentu melalui pesan publik yang disebut tweet. Selain sebuah informasi yang diberikan secara sukarela oleh pemakai, tweeter juga dapat menyimpan informasi tentang lokasi



letak pemakai dan berisi link, emoji, dan tagar untuk membantu pemakai menampilkan emosi mereka dengan lebih baik dan merupakan sumber informasi yang berguna. Diluar itu, pejabat pemerintah dan tokoh politik menggunakan Twitter untuk memberi informasikan kepada publik tentang kegiatan mereka (Yulita et al., 2021).

Pengguna Twitter yang banyak dengan hak yang bebas berpendapat di negara Indonesia, berbagai informasi atau isu politik umumnya menjadi trending topik di Twitter, dan salah satu isu yang sempat gempar dan trending topic adalah rencana pemerintah atas pemindahan ibu kota negara. Dari isu yang di bahas pemerintah pada tahun 2019, timbulah beragam tweet yang bermunculan seperti *#ibukotapindah*, *#pemindahanibukota*. Dari banyaknya respon masyarakat pengguna Twetter terhadap isu yang menjadi salah satu trending topik tersebut maka perlu diadakannya analisis sentimen untuk mengetahui apakah masyarakat Indonesia merespon isu ini dengan respon positif, netral, atau malah negatif. Sehingga dengan adanya penelitian ini kita bisa melihat gambaran pendapat masyarakat terkait kebijakan pemerintah yang merencanakan pemindahan ibu kota negara Indonesia dan juga bisa jadi bahan evaluasi pemerintah dalam pengambilan keputusan, dengan adanya analisis sentimen ini memudahkan kita untuk melihat informasi data di media sosial yang besar dengan cepat.

Analisis sentimen adalah cabang ilmu pada *data mining* yang di pakai untuk menganalisis, mengolah dan memperoleh data tekstual dari entitas semacam layanan, produk, individu, organisasi, dan topik tertentu. Analisis ini berguna buat mendapatkan sebetuk informasi dari sebuah himpunan data yang ada (Kevin

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



et al., 2020). Pada analisis sentimen terdapat beberapa metode klasifikasi *mechine learning* seperti *Naïve Bayes*, *Support Vector Machine*, *Logistic Resresion* dan *Lexion Based*, yang digunakan untuk mencari hasil yang terbaik. Namun pada penelitian ini peneliti memilih untuk menggunakan metode *Naïve Bayes Clasifier* (NBC) dikarenakan metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Membandingkan data latih dan data uji yang berdasarkan probalitas kata kuncinya merupakan metode klasifikasi teks Naive Bayes Classifier (Krisdiyanto, 2021). Berdasarkan pemaparan latar belakang di atas maka peneliti tertarik untuk meneliti lebih dalam tentang **“ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP PINDAHNYA IBU KOTA INDONESIA DENGAN MENGGUNAKAN KLASIFIKASI NAÏVE BAYES”**

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah di paparkan sebelumnya, maka rumusan masalah yang akan di angkat di penelitian ini adalah bagaimana mengklasifikasikan sentimen publik terhadap data tweet pengguna Twitter menggunakan metode *Naïve Bayes Clasifier (NBC)* .

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan penelitian ini adalah bagaimana mengklasifikasikan sentimen opini masyarakat berdasarkan tweet pengguna Twitter menggunakan metode *Naïve Bayes Clasifier (NBC)*.

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

1.4 Manfaat Penelitian

Manfaat dilakukan penelitian ini adalah sebagai berikut :

1. Memberikan pengetahuan mengenai klasifikasi opini masyarakat sehubungan dengan pindahnya ibu kota Indonesia menggunakan metode *Naïve Bayes*.
2. Melihat arah dampak kebijakan pemerintah sehubungan dengan pindahnya ibu kota bagi masyarakat umum.
3. Melakukan klasifikasi opini kedalam kelas positif, netral dan negatif dalam jumlah besar dan cepat.
4. Hasil dari analisis *sentimen* dapat digunakan sebagai bahan evaluasi dan pembelajaran.

Manfaat penelitian ini bagi kampus adalah sebagai berikut :

1. Dengan adanya penelitian ini dapat mengembangkan materi pengajaran.
2. Dengan adanya penelitian ini dapat meningkatkan reputasi kampus melalui hasil penelitian.

1.5 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Hasil dari penelitian ini merupakan data yang berupa komentar positif dan negatif.
2. Untuk data tweet yang di ambil hanya data yang menggunakan bahasa Indonesia.

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian 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 atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi



3. Analisis sentimen yang di buat adalah analisis opini masyarakat sehubungan dengan pindahnya ibu kota Indonesi4. Jumlah data mentah yang di ambil dari Twitter sebanyak 4533 data .
5. Sumber data yang digunakan atau di ambil ialah data Twitter.

@ Hak cipta milik UIN Sutha Jambi

State Islamic University of Sulthan Thaha Saifuddin Jambi



UNIVERSITAS ISLAM NEGERI
SULTHAN THAHA SAIFUDDIN
J A M B I

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

BAB II

LANDASAN TEORI

2.1 Kajian Pustaka

2.1.1 Analisis

Analisis adalah fungsi yang melibatkan angka Operasi seperti menganalisis, memisahkan, menyortir, mengkategorikan dan mengelompokkan kembali menurut kriteria tertentu, kemudian mencari hubungannya dan menafsirkan maknanya. Dengan kata lain analisisnya banyak kegiatan, fungsi dan proses yang saling berhubungan untuk memecahkan masalah atau memecahkan komponen sehingga lebih menjadi akurat atau ditail dan di gabungan kembali lalu tarik kesimpulan (Syahidin & Adnan, 2022).

2.1.2 Analisis *Sentimen*

Analisis sentimen adalah cabang ilmu pada data *mining* yang di pakai untuk menganalisis, mengolah dan memperoleh data tekstual dari entitas semacam layanan, produk, individu, organisasi, dan topik tertentu. Analisis ini berguna supaya mendapatkan bentuk informasi dari sebuah himpunan data yang ada (Kevin et al., 2020). Analisis *sentimen* merupakan salah satu ilmu di dalam bidang *text mining* yang memepelajari analisis opini, sikap emosional, evaluasi mereka di tuangkan dalam bentuk teks (Wiwik, 2022).

2.1.3 Naïve Bayes Clasifier (NBC)

Metode *Naive Bayes* adalah metode klasifikasi yang digunakan dalam analisis sentimen pada *text mining*. Metode ini berpotensi baik untuk klasifikasi. *Naive Bayes* banyak digunakan dalam teknik klasifikasi khususnya di Twitter dengan menggunakan berbagai metode seperti *Unigram Naive Bayes*, *Multinomial Naive Bayes*, dan *Maximum Entropy Classification*, adapun dalam penelitian ini peneliti menggunakan *Multinomial Naive Bayes* dimana sebuah dokumen dapat dikategorikan bertema olahraga, politik, teknologi atau lain-lain berdasarkan frekuensi kata-kata yang muncul pada dokumen. Ciri utama dari klasifikasi *Naive Bayes* adalah untuk mendapatkan hipotesis yang kuat tentang suatu kondisi atau peristiwa (Samsir, et al.,2021). Digunakan persamaan :

$$P_i (Y|X) = \frac{p(x|y) p(y)}{p(x)}$$

Menunjukkan Y suatu kelas spesifik, X merupakan data pada kelas yang belum diketahui, sedangkan $P(Y|X)$ adalah probabilitas hipotesis berdasarkan kondisi, sedangkan $P(Y)$ dan $P(X|Y)$ merupakan probabilitas sebelumnya dari kelas berdasarkan kondisi hipotesis, sedangkan $P(X)$ merupakan probabilitas Y. Adapun contoh cara kerja manual metode *Naive Bayes* adalah sebagai berikut:

Tabel 2.1 Contoh data sampel perhitungan metode *Naive Bayes*

No	Cuaca	Suhu	Kelembapan	Berangin	Main
1	Cerah	Panas	Tinggi	Salah	Tidak
2	Cerah	Panas	Tinggi	Benar	Tidak
3	Berawan	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

Tabel 2.1 (lanjutan)

8	Cerah	Sejuk	Tinggi	Salah	Tidak
9	Cerah	Dingin	Normal	Salah	Ya
10	Hujan	Sejuk	Normal	Salah	Ya

1. Menentukan probabilitas class

$$\mathbf{Ya} \quad 7/10 \quad 0,7$$

$$\mathbf{Tidak} \quad 3/10 \quad 0,3$$

2. Menentukan probabilitas kategori

Ya Cuaca

$$P(\text{Cuaca} = \text{Cerah} | \text{ya}) = 1/7 = 0,14$$

$$P(\text{Cuaca} = \text{Berawan} | \text{ya}) = 2/7 = 0,28$$

$$P(\text{Cuaca} = \text{Hujan} | \text{ya}) = 4/7 = 0,57$$

Kelembapan

$$P(\text{Kelembapan} = \text{Tinggi} | \text{ya}) = 2/7 = 0,28$$

$$P(\text{Kelembapan} = \text{Normal} | \text{ya}) = 5/7 = 0,71$$

Suh

$$P(\text{Suhu} = \text{Panas} | \text{ya}) = 1/7 = 0,14$$

$$P(\text{Suhu} = \text{Sejuk} | \text{ya}) = 2/7 = 0,28$$

$$P(\text{Suhu} = \text{Dingin} | \text{ya}) = 4/7 = 0,57$$

Berangin

$$P(\text{Berangin} = \text{Salah} | \text{ya}) = 5/7 = 0,71$$

$$P(\text{Berangin} = \text{Benar} | \text{ya}) = 2/7 = 0,28$$

Tidak Cuaca

$$P(\text{Cuaca} = \text{Cerah} | \text{tidak}) = 3/3 = 1$$

$$P(\text{Cuaca} = \text{Berawan} | \text{tidak}) = 0/3 = 0$$

$$P(\text{Cuaca} = \text{Hujan} | \text{tidak}) = 0/3 = 0$$

Kelembapan

$$P(\text{Kelembapan} = \text{Tinggi} | \text{tidak}) = 3/3 = 1$$

$$P(\text{Kelembapan} = \text{Normal} | \text{tidak}) = 0/3 = 0$$

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

**Suhu**

$$P(\text{Suhu} = \text{Panas} | \text{tidak}) = 2/3 = 0,66$$

$$P(\text{Suhu} = \text{Sejuk} | \text{tidak}) = 1/3 = 0,33$$

$$P(\text{Suhu} = \text{Dingin} | \text{tidak}) = 0/3 = 0$$

Berangin

$$P(\text{Berangin} = \text{Salah} | \text{tidak}) = 2/3 = 0,66$$

$$P(\text{Berangin} = \text{Benar} | \text{tidak}) = 1/3 = 0,33$$

Tabel 2.2 Probabilitas kategori

KATEGORI/ATRIBUT	SUBSET	Ya	Tidak
Cuaca	Cerah	0,14	1
	Berawan	0,28	0
	Hujan	0,57	0
Suhu	Panas	0,14	0,66
	Sejuk	0,28	0,33
	Dingin	0,57	0
Kelembapan	Tinggi	0,28	1
	Normal	0,71	0
Berangin	Salah	0,71	0,66
	Benar	0,28	0,33

3. Menguji hasil klasifikasi *Naïve Bayes Clasification*

Contoh data no.1

$$P(\text{Main} | \text{Ya}) = P(\text{Main} | \text{Ya}) * (P(\text{Cuaca} = \text{Cerah}) * P(\text{Suhu} = \text{Panas})$$

$$* P(\text{Kelembapan} = \text{Tinggi}) * P(\text{Berangin} = \text{Salah}))$$

$$= 0,7 * (0,14 * 0,14 * 0,28 * 0,71)$$

$$= 0,002727536$$

$$P(\text{Main} | \text{Tidak}) = P(\text{Main} | \text{Tidak}) * (P(\text{Cuaca} = \text{Cerah}) * P(\text{Suhu} = \text{Panas})$$

$$* P(\text{Kelembapan} = \text{Tinggi}) * P(\text{Berangin} = \text{Salah}))$$

$$= 0,3 * (1 * 0,66 * 1 * 0,66)$$

$$= 0,13068$$

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian 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 atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sultha Jambi

Tabel 2.3 Hasil klasifikasi *Naïve Bayes*

No	Ya	Tidak	Fakta	Klasifikasi	Prediksi
1	0,002727536	0,13068	Tidak	Tidak	Sesuai
2	0,001075648	0,06534	Tidak	Tidak	Sesuai
3	0,005455072	0	Ya	Ya	Sesuai
4	0,022209936	0	Ya	Ya	Sesuai
5	0,114647463	0	Ya	Ya	Sesuai
6	0,045213084	0	Ya	Ya	Sesuai
7	0,022209936	0	Ya	Ya	Sesuai
8	0,005455072	0,06534	Tidak	Tidak	Sesuai
9	0,028159026	0	Ya	Ya	Sesuai
10	0,056318052	0	Ya	Ya	Sesuai

2.1.4 Tekt Mining

Penambangan teks adalah fase proses analisis data tekstual dari mana sumber data diekstraksi dari sebuah dokumen, misalnya kalimat dari kata-kata data, konsep *text mining* sering digunakan dalam klasifikasi dokumen format teks, dokumen-dokumen ini diklasifikasikan menurut subjek dokumen (Darwis et al., 2021). *Text mining* adalah istilah yang mengacu pada teknik data *mining* yang menganalisis dan Pengolahan data teks terstruktur dan semi-terstruktur. Penambangan teks adalah proses yang seperti data *mining*, tetapi dengan pendapatan yang berbeda. Untuk penambangan teks, pertama diperlukan Setelah pengumpulan data, data harus diproses terlebih dahulu sebelum proses klasifikasi (Pintoko & L., 2018).

2.1.5 Twitter

Media sosial Twitter adalah bagian dari perkembangan media komunikasi yang di buat agar pemakai bisa menuangkan ekspresi, pendapat, aspirasi, kritik, serta bertukar informasi mengenai suatu informasi yang sedang menjadi trending topik, tanpa dibatasi waktu dan tempat. Dimana, pendapat tersebut tersampaikan

langsung, kegiatan mengunggah pendapat di Twitter ini juga sering di bilang dengan tweet. Hal demikian membuat Twitter dijadikan tempat sebagai sumber data teks yang dapat di ambil dan di gunakan untuk berbagai keperluan kebutuhan riset di bidang teknologi informasi (Handayani & Sulistiyawati, 2021).

2.1.6 Term Frequency-Inverse Document Frequency (TF-IDF)

Term weighting ialah proses semua pembobotan kata-kata untuk mengoptimalkan kemampuan analisis *sentimen* dalam proses penambangan teks. Penelitian ini menggunakan *Term Frequency Inverse Document Frequency* (TF-IDF). Istilah frekuensi ($tf(w,d)$) diasumsikan memiliki hubungan kepentingan menurut keseluruhan tampilan teks atau dokumen. *Inverse Documentary Frequency* (IDF) adalah metode pembobotan karakter yang mengontrol tampilan karakter dalam kumpulan teks (Gifari et al., 2022).

Adapun contoh melakukan perhitungan manual TF-IDF adalah sebagai berikut :

TF IDF

Kata kunci : ilmu semantik

Dokumen 1 (D1) : Program logika dan semantik

Dokumen 2 (D2) : Ilmu antar individu

Dokumen 3 (D3) : Dalam program ilmu terdapat transfer ilmu semantik

Jumlah dokumen = 3

Tabel 2.4 Proses TF-IDF

Token	Kk	Tf			Df	D/df	IDF (log D/ df)	D			
		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	0.477	0	0.477	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	3	0.477	0	0	0.477	0
Tfasfer	0	0	0	1	1	3	0.477	0	0	0	0,477

Mencari peringkat. $D1 = 0.176$. $D2 = 0.176$ $D3 = 0.528$

Penjelasan :

- Tabel **token** terdapat beberapa kata yang akan kita bobotkan.
- Tabel **kk** (kata kunci), dimana isi tabel dari kata kunci kita cocokkan dengan tabel **token**, contoh pada tabel token terdapat kata program, nah kita lihat apakah di kata kunci ad kata program apabila tidak maka di labelkan dengan angka 0 dan apabila ada kita labelkan dengan angka 1 begitu seterusnya.
- Tabel **tf** dimana isi tabel tf terbagi menjadi 3 bagian yaitu D1,D2,D3. Untuk pengisian angkanya itu sama seperti kita melabelkan di kata kunci.
- Tabel **df** ialah menghitung dokumen *frequency* jumlah dokumen yang mengandung kata **token**.

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.
 - Dilarang memperbanyak sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sulthan Jambhi

- e. Tabel **D/df** cara mencari nilai tabel D/df ini ialah dengan membagi jumlah dokumen karna dokumennya 3 maka 3 dibagi dengan nilai pada tabel **df**.
- f. Tabel **IDF (log D/df)** untuk mencari nilai pada tabel ini kita menggunakan rumus log dengan angka yang ada dalam tabel **D/df** contoh pada colum pertama di tabel **D/df** adalah 1.5 maka kita mencari log 1.5 yaitu 0.176.
- g. Tabel **W** dimana tahap ini kita mencari bobot dari setiap kata baik dari tabel **kk**, **D1**, **D2**, **D3**. Untuk mencari nilai dari tabelnya kita tinggal memindahkan angka dari tabel **IDF** dimana dengan ketentuan apabila dalam tabel **kk** dan **tf** memiliki nilai 0 maka dalam tabel **W** nilainya 0 juga.
- h. Selanjutnya mencari peringkat yaitu dengan menamahkan kata-kata yang hanya berada dalam kata kunci yaitu ilmu dan semantik pada dokumen 1, dokumen 2, dan dokumen 3 nilai dari masing-masing dokumen tersebut di jumlahkan sehingga dapatla hasil peringkat.

2.1.7 Confusion Matrix

Confusion Matrik adalah metode untuk mengukur suatu kinerja klasifikasi, pada cara kerjanya Confusion Matrik bekerja untuk membandingkan atau untuk melihat hasil klasifikasi yang dibuat oleh sistem dengan hasil klasifikasi yang dibuat. Saat pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Nilai True Negative (TN) merupakan jumlah data 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 sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sutha Jambi

yang terdeteksi dengan benar, sedangkan False Positive (FP) merupakan data negatif namun terdeteksi sebagai data positif (Karsito & Susanti Santi, 2019).

2.2 Analisis *Sentiment* Twitter

Pesan atau teks singkat, yang disebut tweet, merupakan bentuk komunikasi utama di platform media sosial Twitter. Platform Twitter yang penuh dengan komentar dapat memengaruhi pembentukan opini. *Analisis* sentimen dapat dilakukan pada tweet dengan mengidentifikasi dan mengklasifikasikan polaritas teks untuk menentukan apakah suatu dokumen memiliki nilai positif atau negatif sesuai dengan klasifikasi yang ditentukan. Sebagian besar pengguna Twitter bebas mengungkapkan pendapat mereka tentang banyak topik (Watrianthos et al., 2022).

Adapun beberapa tahapan pengolahan data dalam analisis *sentiment* pada Twitter sebagai berikut :

1. Melakukan penarikan data set dari media sosial. Data set yang di tarik merupakan tweet dari masyarakat.
2. *Cleansing* merupakan tahapan pembersihan tanda baca. Contoh kalimat melalui proses pembersihan sebagai berikut :

Tabel 2.5 Proses *Cleansing*

Sebelum	"Vaksin menimbulkan keuntungan dan kerugian di masyarakat http://d1vr.it/Qb83n8 pic.twitter.com/8MhUMO"
Sesudah	"vaksin membawa keuntungan dan kerugian di masyarakat"

Berdasarkan contoh tersebut kita bisa melihat dimana karakter yang dihilangkan berupa URL dan tanda baca.

3. *Case Folding* merupakan tahapan untuk menyamaratakan penggunaan huruf kapital. Contoh teks yang melalui tahapan ini sebagai berikut :

Tabel 2.6 Proses *Case Folding*

Sebelum	" Wajib vaksin COVID untuk mendapatkan sertifikat kepada masyarakat yang sedang melakukan perjalanan jauh "
Sesudah	" wajib vaksin covid untuk mendapatkan sertifikat kepada masyarakat yang sedang melakukan perjalanan jauh "

Berdasarkan contoh tersebut kita dapat melihat dimana kata-kata yang semulanya bercampur antara huruf kecil dan besar setelah melalui proses itu berubah menjadi sama yaitu kecil semua.

4. *Tokenizing* merupakan proses pemecahan kalimat menjadi perkata.

Contoh teks yang telah di *Tokenizing* sebagai berikut :

Tabel 2.7 Proses *Tokenizing*

Sebelum	"pemerintah mewajibkan yang melakukan bepergian harus memiliki sertifikat vaksin covid"
Sesudah	"pemerintah", "mewajibkan", "yang", "melakukan", "bepergian", "harus", "memiliki", "sertifikat", "vaksin", "covid"

Berdasarkan contoh kita dapat melihat dimana kalimat-kalimat tersebut dipotong atau di pecah menjadi bagian-bagian.

5. *Filtering* merupakan langkah menghilangkan kata-kata yang muncul dalam jumlah banyak tetapi tidak dianggap penting (stopwords). Berikut adalah contoh dari kata-kata *stopword* :

ada	di	kalau	pada	yaitu
aku	dia	kami	saja	bila
bapak	ini	lalu	tentu	hari
berbagai	itu	lewat	untuk	masa
cara	jadi	meski	yang	tapi
cuma	juga	oleh	wah	hal

Gambar 2.1 Contoh kata-kata stopword (Yulita et al., 2021).

2.3 Penelitian Yang Relevan

Penelitian yang akan di lakukan terkait dengan penelitian-penelitian terdahulu diantaranya sebagai berikut :

Samsir, et al. (2021) pada penelitian ini di lakukan penambangan teks berbasis dokumen pada Twitter yang di analisis dengan metode algoritma *Naive Bayes*. Penelitian ini menunjukkan bahwa Hasilnya menunjukkan bahwa 30% memandang pembelajaran daring positif, 69% negatif, dan 1% netral. Tingginya opini negatif tersebut disebabkan ketidakpuasan masyarakat terhadap pembelajaran daring.

Asro'i & Februariyanti, (2022) Peneliti menggunakan algoritma *K-Nearest Neighbor* untuk menentukan sentimen pengguna Twitter menggunakan *library scikit-learn* yang populer untuk ilmuwan data. Algoritme ini diterapkan pada 6408 data Tweet yang berisi kata kunci "PPKM" yang dikumpulkan antara 1 Juli 2021 hingga 31 Desember 2021. Hasil *training* menunjukkan tingkat akurasi 69,5%, *recall* 69,5%, dan presisi 68,7%.

Gifari et al., (2022) Tujuan dari penelitian ini adalah untuk mendeskripsikan analisis sentimen dari dokumen film. Metode yang digunakan dalam penelitian ini adalah TF-IDF (*Term Frequency-Inverse Document Frequency*) dan SVM (*Support Vector Machine*). Metode ini dipilih karena kemampuannya untuk menimbang kata dan mengklasifikasikan data berdimensi tinggi. Dari uji skenario yang dilakukan diketahui bahwa algoritma TF-IDF dan SVM dapat digunakan untuk kasus rivew film dengan nilai akurasi 85%, nilai akurasi 100%, nilai pengenalan 70% dan skor F1. Nilai 82%.

Watrianthos et al., (2022) Penelitian ini membuat model klasifikasi menggunakan algoritma *Naive Bayes* pada data twit Twitter dengan mengevaluasi persepsi dan penerimaan audiens terhadap model pembelajaran campuran. Ketika opini diteliti pada model ini, hasilnya diklasifikasikan sebagai tweet positif atau negatif dengan menggunakan metode analisis *sentimen* Twitter. Hasilnya menunjukkan bahwa polarisasi sikap positif dan negatif hampir berimbang: positif 44,51% dan negatif 45,80%.

Vonega et al., (2022) Pada penelitian ini analisis *sentimen* dilakukan dengan menggunakan metode *text mining* dan beberapa *library* seperti *TextBlob*, *VaderSentiment*, serta *SentiWordNet*. Dari data set yang di hasilkan dengan *keyword* “Puan Maharani” memperoleh opini negatif rata-rata hanya 0,1%, rating netral 97,25 dan rating positif 2,55%. Dapat disimpulkan bahwa pengguna Twitter pada umumnya tidak agresif atau defensif saat membahas isu-isu yang mengarah pada pencalonan Puan Maharani di tahun 2024 mendatang. pemilu Presiden.

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 sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suntho Jambi



Pada analisis sentimen terdapat beberapa metode klasifikasi *mechine learning* seperti Naïve Bayes, Support Vector Machine, Logistic Resgion dan Lexion Based, yang digunakan untuk mencari hasil yang terbaik. Namun pada penelitian ini peneliti memilih untuk menggunakan metode *Naïve Bayes Clasifier* (NBC) dikarenakan metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Maka dari pada itu peneliti bertujuan menggunakan metode tersebut untuk menganalisis *sentiment* tentang opini masyarakat terhadap pindahnya ibu kota. Adapun tabel pembanding peneliti terdahulu adalah sebagai berikut :

Tabel 2.8 Tabel pembanding peneliti terdahulu

No	Peneliti	Judul	Metode	Hasil
1.	(Samsir, et al.,2021)	Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes	Metode Naïve Bayes	hasil yang diambil selama penelitian tersebut menunjukkan 30% perasaan positif, 69% perasaan negatif, dan 1% perasaan netral.
2	(Asro'i & Februariyanti, 2022)	Analisis Sentimen Pengguna Twitter terhadap Perpanjangan PPKM Menggunakan Metode K-Nearest Neighbor	K-Nearest Neighbor	Hasil akurasi 69,5%, recall 69,5%, dan presisi 68,7%. Hasilnya kurang memuaskan karena diperlukan ketelitian yang tinggi agar valid ketika digunakan dalam evaluasi kebijakan.
3	(Gifari et al., 2022)	Analisis Sentimen Review Film Menggunakan TF-IDF dan SVM	TF-IDF dan Support Vector Machine	hasil nilai accuracy 0,85, nilai presisi 1,0, nilai recall 0,7 dan nilai skor F1 0,82. Hal ini dikategorikan baik.
4.	(Watrianthos et al., 2022)	Analisis Sentimen Pembelajaran Campuran Pada Twitter Data Menggunakan Algoritma Naïve Bayes	Naïve Bayes	44,51% positif, 45,80% negatif, dan 9,69 % netral. positif menunjukkan bahwa siswa, guru, dan institusi penyelenggara metode pembelajaran campuran.

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

Tabel 2.8 (lanjutan)

5	(Vonega et al., 2022)	Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024	text mining dan beberapa library seperti TextBlob, VaderSentiment, serta SentiWordNet	mendapat rata-rata opini negatif hanya 0,1%, netral 97,25, dan positif 2,55%. Pengguna Twitter pada umumnya tidak agresif saat membahas topik yang mengarah pada pencalonan Puan Maharani di Pilpres 2024 .
---	-----------------------	--	---	---

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

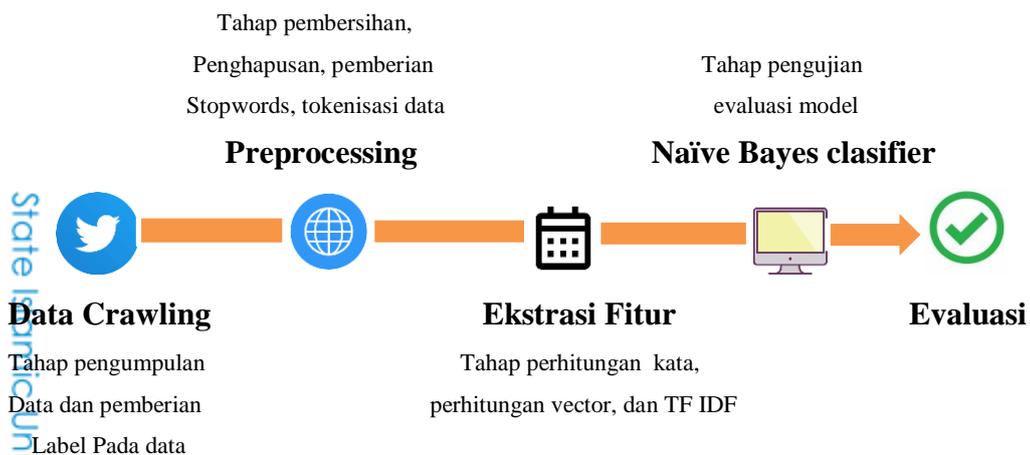
Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
b. Pengutipan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

BAB III

METODE PENELITIAN

3.1 Tahapan Penelitian

Pengumpulan data pada penelitian ini adalah dengan melakukan penarikan data set dari media sosial atau di sebut dengan *Internet Searching*. Data set yang di tarik merupakan tweet dari masyarakat dan tweet itu akan di kelompokkan dengan label positif dan negatif menggunakan metode *Naïve Bayes Clasifier* (NBC).

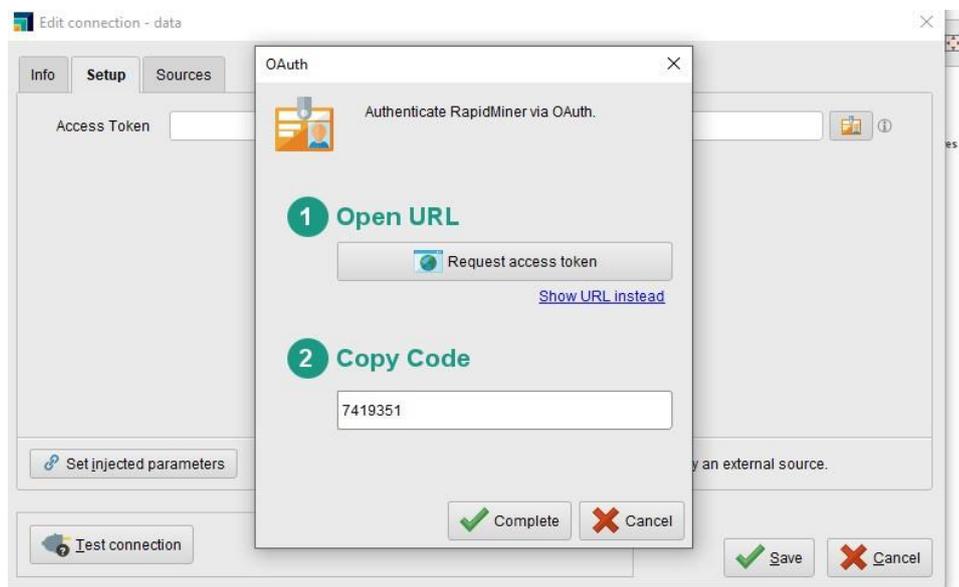


Gambar 3.1 Tahapan Penelitian Menggunakan Klasifikasi *Naïve Bayes*

3.1.1 Data Crawling

Data crawling adalah proses otomatis untuk mengumpulkan dan mengindeks data dari berbagai sumber seperti situs web, database, atau dokumen, pengumpulan data dilakukan dengan crawling data *tweet* menggunakan RapidMiner dengan Keyword *#ibukotapindah*, *#ikn*. Setelah data mentah di tarik lalu data tersebut di simpan ke dalam bentuk CSV (Comma Separated Values)

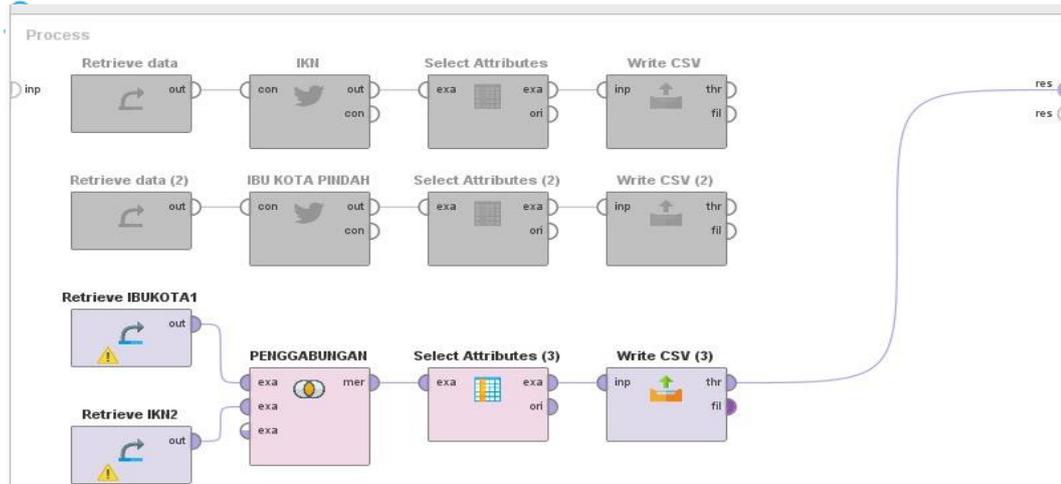
yang nantinya akan digunakan di proses yang lebih lanjut. Rapid Miner adalah aplikasi atau perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data *mining*. Tahapan yang awal dalam penelitian ini adalah *Crawling* Data. Namun sebelum melakukan penarikan data pada Twitter menggunakan aplikasi Rapid Miner kita perlu terlebih dahulu mengkoneksikan antara kedua aplikasi tersebut yaitu Rapid miner dan Twitter, yang mana tahapan pengoneksian dapat kita lihat pada gambar sebagai berikut :



Gambar 3.2 Tahapan pengkoneksian Rapid Miner ke Twitter

Pada gambar di atas yaitu dimana proses menghubungkan antara Rapid Miner dan Twitter dimana nantinya Twitter akan memberikan berupa akses Token yang berupa kode angka yang akan di gunakan dalam proses menghubungkan kedua aplikasi tersebut barulah Rapid Miner bisa di gunakan untuk menarik data.

Adapun tahapan *Crawling* data dapat kita lihat pada gambar di bawah :



Gambar 3.3 Tahapan *Crawling*

Gambar 3.3 adalah tahapan *Crawling* data pada tanggal 6 maret 2023 dengan menggunakan dua kata kunci yaitu IKN dan IBU KOTA PINDAH, dimana data di tarik menggunakan operator *Search Twitter* menggunakan pengaturan data populer atau yang terbaru, kata kunci IKN sebanyak 3866 data mentah dan IBU KOTA PINDAH sebanyak 667 data mentah kemudian data kedua kata kunci tersebut di gabungkan kedalam satu folder dengan menggunakan operator *Append* lalu data disimpan kedalam file CSV yang dimana data tersebut hanya berisi data teks dapat kita lihat pada tabel dibawah :

Tabel 3.1 Tampilan beberapa data mentah hasil *Crawling*

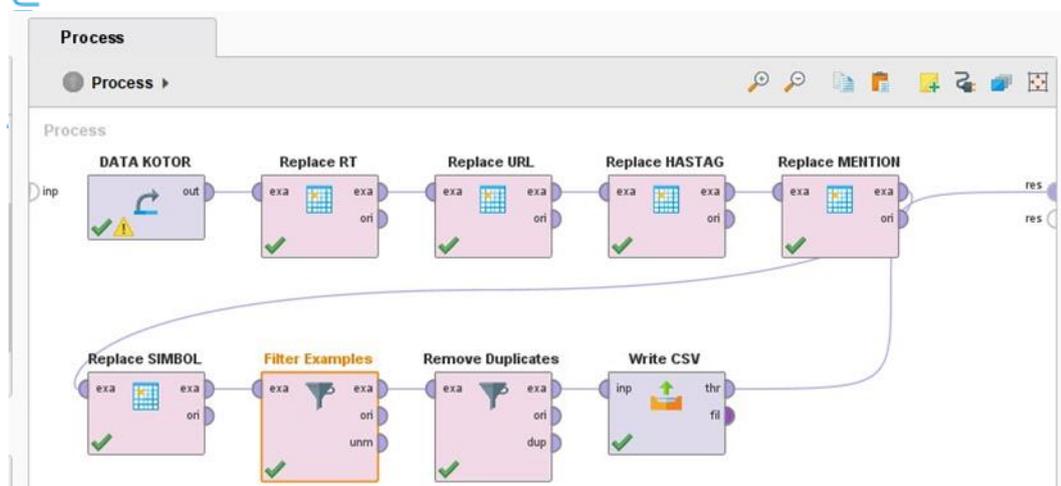
No	Text
1	@rahmawaadityan Kagak warga ibu kota jg,di kota lain yg rumahnya pinggir rel atau yg pakai lahan punya TNI kalau disuruh pindah karena tanahnya atau rumahnya atau lahannya mau di pakai pasti ngamuk-ngamuk nggak terima.yg KAI malah harus kasih kompensasi sesuai bangunan
2	@sociotalker Seharusnya mumpung ada rencana pindah ibu kota ke Kalimantan, sekalian aja reset NPP buat dibangun di Kalimantan yang notabenenya 'aman' dari aktivitas seismik. Toh Jepun yg jelas2 di RoF aja berani bikin NPP...
3	@daffanacom ibu kota mau pindah ke ikn. tp proyek unggulan kayak gini aja bisa mangkrak. yakin gak malu maluin indonesia ?
4	sekarang Ada istilah ibu kota di pindah

Tabel 3.1 (lanjutan)

5	@geloraco Catat: Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta. Penasehat sdh bilang Bos tuch Ibu Kota mau pindah "into the Woods in East Borneo." Jadi hasil kunjunganya Blio menolong Ekonomi Malaysia.
9	@CNNIndonesia Klo warga sulit di pindah, solusinya depo pertamina yg di pindah...ingat pemilik sah Tanah NKRI adalah rakyat..nah tugas pemerintah cari solusi pindahin depo tsb, pindahin ibu kota saja bisa, kenapa pindahin depo rumit amat..hadech
7	RT @VIVAcoid: Menteri Pendayagunaan Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparat sipil...

3.1.2 Preprocessing

Adapun tahap ini adalah mempersiapkan data teks sebelum digunakan pada proses lainnya. Pada tahap ini akan mengubah data teks menjadi bentuk yang lebih baik sehingga menghasilkan informasi teks dengan kualitas yang baik dan siap di gunakan untuk proses selanjutnya. Data yang diperoleh dari hasil penarikan tidak dapat diklasifikasikan secara langsung karena data tersebut mengandung banyak simbol dan kata-kata yang tidak perlu, sehingga diperlukan tahap *Preprocessing* data agar data lebih terstruktur dan bersih sehingga dapat diklasifikasikan ketahap berikutnya. Adapun tahapan pembersihan data dapat kita lihat pada gambar dibawah :



Gambar 3.4 Tahapan pembersihan data

Gambar 3.4 dimana proses pembersihan data kotor dengan menggunakan operator *Replace* RT, URL, HASTAG, MENTION, SIMBOL. Lalu menggunakan operator *Filter Examples* yang gunanya untuk menghapus data yang kosong dan untuk menghapus data yang sama atau dobel digunakan operator *Remove Duplicates*. Adapun tahapannya dapat di lihat sebagai berikut :

Cleansing

Cleansing adalah tahapan dimana karakter dan tanda baca yang tidak diperlukan dihilangkan dari teks, yang mana salah satu fungsi *cleansing* itu sendiri yaitu dapat mengurangi *noise* pada dataset. Contoh karakter yang di hilangkan berupa URL, tag (#), tanda baca seperti titik (.), koma (,) dan karakter lainnya yang tidak digunakan.

a. *Cleansing RT*

Tabel di bawah ini adalah pembersihan kata yang mengandung RT (*retweet*) dimana teks tersebut tidak mempunyai makna dalam analisis *sentiment* sehingga dalam analisis *sentiment* dilakukan *Cleansing RT* dengan menggunakan operator *Replace RT* menggunakan kode **RT @.***: contoh pada tabel dibawah:

Tabel 3.2 *Cleansing RT*

No	Sebelum	Sesudah
1	RT @tedjo_0587: @Rian_Az Dokternya para kadrun, mkanya ngabacotnya ga pake pkiran.. emng klau ibu kota pindah, rakyat juga hrus ikut pinda...	@Rian_Az Dokternya para kadrun, mkanya ngabacotnya ga pake pkiran.. emng klau ibu kota pindah, rakyat juga hrus ikut pinda...
2	RT @VIVAcoid: Menteri Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengumumkan rencana relokasi PNS...	Menteri Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengumumkan rencana relokasi PNS...

Tabel 3.2 (lanjutan)

3	RT @ch_chotimah2: PdhI semua rakyat Indonesia terutama orang Kalimantan seperti saya ini bahagia betul ibu kota negara pindah, hanya dokter...	PdhI semua rakyat Indonesia terutama orang Kalimantan seperti saya ini bahagia betul ibu kota negara pindah, hanya dokter...
4	RT @kumparan: Salah satu anggota ASN merasa resah dan gundah harus pindah tempat kerja ke Ibu Kota Negara baru. Ia menyatakan ketidaksiapan...	Salah satu anggota ASN merasa resah dan gundah harus pindah tempat kerja ke Ibu Kota Negara baru. Ia menyatakan ketidaksiapan...

b. *Cleansing URL*

Tabel di bawah ini adalah contoh pembersihan kata yang mengandung *URL* dimana *URL* tersebut tidak mempunyai makna dalam analisis *sentiment* sehingga dalam analisis *sentiment* dilakukan *Cleansing URL* dengan operator *Replace URL* menggunakan kode **http.*** contoh tahapan ini dapat kita lihat pada tabel dibawah ini :

Tabel 3.3 *Cleansing URL*

No	Sebelum	Sesudah
1	Wakil Ketua DPR Sebut Pemindahan IKN Bukan Sekadar Pindah Tempat https://t.co/5hbyPymET3	Wakil Ketua DPR Sebut Pemindahan IKN Bukan Sekadar Pindah Tempat
2	PNS Pindah ke Ibu Kota Baru Bakal Dapat Segudang Tunjangan, Ini Rinciannya https://t.co/TYgntTito2	PNS Pindah ke Ibu Kota Baru Bakal Dapat Segudang Tunjangan, Ini Rinciannya
3	Untung gue dari bocah udah pindah agak jauh dari ibu kota aowkwkwkwkwk. Kalau di sini mulu mana ngerti gue kalau biji karet di mainin https://t.co/o7TIFX6QDt	Untung gue dari bocah udah pindah agak jauh dari ibu kota aowkwkwkwkwk. Kalau di sini mulu mana ngerti gue kalau biji karet di mainin
4	Apal ibu kota bbrp negara, tp skrg ada yang udah pindah ibu kota, knowledge nya sudah expired https://t.co/7k0uiYKPTO	Apal ibu kota bbrp negara, tp skrg ada yang udah pindah ibu kota, knowledge nya sudah expired
5	Salah satu anggota ASN merasa resah dan gundah harus pindah tempat kerja ke Ibu Kota Negara baru. Ia menyatakan ketidaksiapannya untuk pindah ke IKN. https://t.co/WV0s9wauLk	Salah satu anggota ASN merasa resah dan gundah harus pindah tempat kerja ke Ibu Kota Negara baru. Ia menyatakan ketidaksiapannya untuk pindah ke IKN.

Cleansing HASTAG

Tabel di bawah ini adalah contoh pembersihan kata yang mengandung *HASTAG* dimana *HASTAG* tersebut tidak mempunyai makna dalam analisis *sentiment* sehingga dalam analisis *sentiment* dilakukan *Cleansing HASTAG* dengan operator *Replace HASTAG* menggunakan kode *#.*?* contoh tahapan ini dapat kita lihat pada tabel dibawah ini :

Tabel 3.4 Cleansing HASTAG

No	Sebelum	Sesudah
1	Jika IKN batal, saya janji akan bersujud.#IKN	Jika IKN batal, saya janji akan bersujud.
2	Kenapa bukan Balikpapan yg sdh Kota Besar, naik status jadi Ibukota?#Ibukotapindah	Kenapa bukan Balikpapan yg sdh Kota Besar, naik status jadi Ibukota?
3	Salah satu pertanyaannya yang dijawab Anies ialah apakah dirinya akan melanjutkan pembangunan Ibu Kota Negara.#IKN	Salah satu pertanyaannya yang dijawab Anies ialah apakah dirinya akan melanjutkan pembangunan Ibu Kota Negara.
4	Pada 23 Februari 2023 lalu, Presiden @jokowi meninjau lokasi pembangunan Istana Kepresidenan di Ibu Kota Negara (IKN) Nusantar#IKN	Pada 23 Februari 2023 lalu, Presiden @jokowi meninjau lokasi pembangunan Istana Kepresidenan di Ibu Kota Negara (IKN) Nusantar
5	Presiden optimis kawasan Istana Kepresidenan akan siap digunakan untuk perayaan Hari Ulang Tahun (HUT) Kemerdekaan RI pada bulan Agustus2024 mendatang.#Ibukotapindah	Presiden optimis kawasan Istana Kepresidenan akan siap digunakan untuk perayaan Hari Ulang Tahun (HUT) Kemerdekaan RI pada bulan Agustus 2024 mendatang.

Cleansing MENTION

Tabel di bawah ini adalah contoh pembersihan kata yang mengandung *MENTION* dimana *MENTION* tersebut tidak mempunyai makna dalam analisis *sentiment* sehingga dalam analisis *sentiment* dilakukan *Cleansing MENTION*

dengan operator *Replace MENTION* menggunakan kode `@.*?` contoh tahapan ini dapat kita lihat pada tabel dibawah ini :

Tabel 3.5 *Cleansing MENTION*

No	Sebelum	Sesudah
1	@desliyani_nl iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru, apalagi ikn loh yg mau dijadiin ibu kota. pikir pikir lagi deh ibu kota pindah	iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru, apalagi ikn loh yg mau dijadiin ibu kota. pikir pikir lagi deh ibu kota pindah
2	@daffanacom ibu kota mau pindah ke ikn. tp proyek unggulan kayak gini aja bisa mangkrak. yakin gak malu maluin indonesia ?	ibu kota mau pindah ke ikn. tp proyek unggulan kayak gini aja bisa mangkrak. yakin gak malu maluin indonesia ?
3	@tanyakanrl Nambah macet. Nambah rempong opo opo larangan. Wis pindah ibu kota Arep ngurus apa apa susah. Runyam	Nambah macet. Nambah rempong opo opo larangan. Wis pindah ibu kota Arep ngurus apa apa susah. Runyam
4	@detikcom Hitung untung ruginya dulu..mending biayannya dialihkan untuk pemulihan ekonomi ..blm urgent kayaknya pindah ibu kota.	Hitung untung ruginya dulu..mending biayannya dialihkan untuk pemulihan ekonomi ..blm urgent kayaknya pindah ibu kota.
5	@bliblidotcom Dulu pernah ada yg namanya monas di ibu kota negara, skrg ibu kota negara pindah. Mungkin lendmark negara adalah kwitansi utang wk	Dulu pernah ada yg namanya monas di ibu kota negara, skrg ibu kota negara pindah. Mungkin lendmark negara adalah kwitansi utang wk

Cleansing SIMBOL

Tabel di bawah ini adalah contoh pembersihan kata yang mengandung *SIMBOL* dimana *SIMBOL* tersebut tidak mempunyai makna dalam analisis *sentiment* sehingga dalam analisis *sentiment* dilakukan *Cleansing SIMBOL* dengan operator *Replace SIMBOL* menggunakan kode `[!@#$$%^&*()_+/,.,><;":]` contoh tahapan ini dapat kita lihat pada Tabel 3.6 dibawah ini :

Tabel 3.6 *Cleansing SIMBOL*

No	Sebelum	Sesudah
1	iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru, apalagi ikn loh yg mau dijadiin ibu kota. pikir pikir lagi deh ibu kota pindah	iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru apalagi ikn loh yg mau dijadiin ibu kota pikir pikir lagi deh ibu kota pindah
2	Dokternya para kadrin, mkanya ngabacotnya ga pake pkiran.. emng klau ibu kota pindah, rakyat juga hrus ikut pindah.. haddeeuuhh..	Dokternya para kadrin mkanya ngabacotnya ga pake pkiran emng klau ibu kota pindah rakyat juga hrus ikut pindah haddeeuuhh
3	Pertamina Energy Tower 6 thn berlalu, Erick Thohir sbg Menteri BUMN yg baru & yg berkuasa mengarahkan masa depan dan mimpi Pertamina, termasuk memutuskan apakah gedung itu akan diteruskan atau dihentikan. Ternyata Presiden Jokowi lebih memilih pindah Ibu Kota Baru.	Pertamina Energy Tower thn berlalu Erick Thohir sbg Menteri BUMN yg baru yg berkuasa mengarahkan masa depan dan mimpi Pertamina termasuk memutuskan apakah gedung itu akan diteruskan atau dihentikan Ternyata Presiden Jokowi lebih memilih pindah Ibu Kota Baru
4	Heru membiarkan jakarta dilalap banjir, untuk menguatkan ibu kota mmng harus pindah ke mimpi penghuni istana.	Heru membiarkan jakarta dilalap banjir untuk menguatkan ibu kota mmng harus pindah ke mimpi penghuni istana

f. *Cleansing* data kosong

Tahapan selanjutnya setelah melakukan pembersihan data dengan melakukan *Cleansing* data kosong yaitu tahapan menghapus data yang kosong dari data yang sudah melalui tahapan *Cleansing* dengan menggunakan operator yang bernama *Filter Examples*.

Cleansing data duplikat

Setelah melakukan tahapan penghapusan data yang kosong maka ada satu tahap lagi yaitu menghapus data yang duplikat dengan menggunakan operator yang bernama *Remove Duplicates* data yang mengandung lebih dari satu data yang sama maka akan di hapus supaya mempermudah pada saat melakukan analisis *sentiment*. Setelah tahapan *Cleansing* data dilakukan barulah data tersebut bisa di gunakan ke tahap berikutnya.



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

i. *Case Folding*

Case Folding adalah untuk menyamaratakan penggunaan huruf kapital.

Didalam penulisan tweet pasti terdapat perbedaan bentuk hurup maka perlu dilakukanla yang namanya *Case Folding*.

j. *Tokenizing*

Tokenisasi adalah memecah kalimat menjadi pisah-pisah supaya kata-kata tersebut dapat diproses lebih lanjut dalam analisis teks.

k. *Filtering*

Filtering adalah langkah menghilangkan kata-kata yang muncul dalam jumlah banyak tetapi tidak dianggap penting (*stopwords*). Pada dasarnya stopwords adalah sekumpulan kata yang banyak digunakan dalam berbagai bahasa. Alasan untuk menghapus kata-kata *stopwords* adalah karena terlalu sering muncul, memungkinkan pengguna untuk fokus pada kata-kata yang lebih penting.

3.1.3 Ekstrasi Fitur

merupakan proses penting pada klasifikasi teks untuk mengubah format tekstual yang tidak terstruktur menjadi terstruktur sehingga dapat diproses oleh algoritma *machine learning* untuk mengklasifikasikan ke class yang telah ditentukan sehingga mempeermuda klasifikasi *Naïve Bayes*. Selanjutnya TF-IDF (*Term frequency Inverse Document Frequency*) digunakan untuk menentukan nilai frekuensi sebuah kata dalam sebuah dokumen atau artikel dan juga frekuensi di dalam banyak dokumen. Perhitungan ini menentukan seberapa relevan sebuah

kata dalam sebuah dokumen. Untuk proses dan hasilnya dapat kita lihat pada bab

4.

3.1.4 Pelabelan Manual

Sebelum melanjutkan ketahapan berikutnya adapun data yang telah melalui proses yang telah disebutkan di atas, data tersebut di labelkan secara manual dengan sentiment positif dan negatif, apabila terdapat data yang bernada netral maka data tersebut dilihat dulu jika tidak ada kata-kata seperti mengejek, menghina atau cenderung tidak setuju maka di golongan positif, barulah dilanjutkan ke tahap berikutnya. Berikut adalah beberapa contoh data yang telah dilakukan pelabelan secara manual menggunakan Majority rule, yaitu suara terbanyak yang ditentukan melalui voting untuk mengambil keputusan tanpa mempertimbangkan argumentasi.

O1 = Orang pertama, O2 = Orang kedua, O3 = Orang ketiga, N = Negatif,

P = Positif

Tabel 3.7 Data pelabelan manual

No	Text	O1	O2	O3	Sentimen
1	iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru apalagi ikn loh yg mau dijadiin ibu kota pikir pikir lagi deh ibu kota pindah	N	N	N	Negatif
2	ibu kota mau pindah ke ikn tp proyek unggulan kayak gini aja bisa mangkrak yakin gak malu maluin indonesia ?	N	N	N	Negatif
3	Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparat sipil...	P	P	P	Positif
4	Ga ada urgensinya ibu kota pindah	N	N	N	Negatif

Tabel 3.7 (Lanjutan)

5	Nambah macet Nambah rempong opo opo larangan Wis pindah ibu kota Arep ngurus apa apa susah Runyam	N	N	N	Negatif
6	Bahwa sebenarnya dari jaman Sukarno udah rencana pindah ke Palangkaraya cuma jaman pak Jokowi terealisasi Yg lain cuma bisa wacana doang	P	P	P	Positif
7	tc kayak bakalan tetep jalan pindah ke ikn karna jakarta yg udah overloud banget menurut gua wajar aja kalo pindah ibu kota kalo masalah anggaran besar ya kembali membangun kota baru pasti butuh biaya besar asalkan roadmapnya jelas gaada masalah	P	P	P	Positif
8	Serius tata ruang jakarta udah nggk bisa diselamatin lagi emang paling bener ibu kota pindah ke ikn	P	P	P	Positif
9	Kok lo jadi takut emangnya IKN itu penting? gak lah IKN itu buang ² uang rakyat krn bencinya kalian sama Anies sampe ibu kota negara kalian mau pindah dasar komunis	N	N	N	Negatif
10	Engga sabar si pengen liat ibu kota pindah dan perubahan lalu lintas jakarta	P	P	P	Positif
11	Saya org kalimantan yg ada di bekasi menolak ibu kota pindah ape lu?	N	N	N	Negatif
12	Maksa bangun yg baru	N	N	N	Negatif

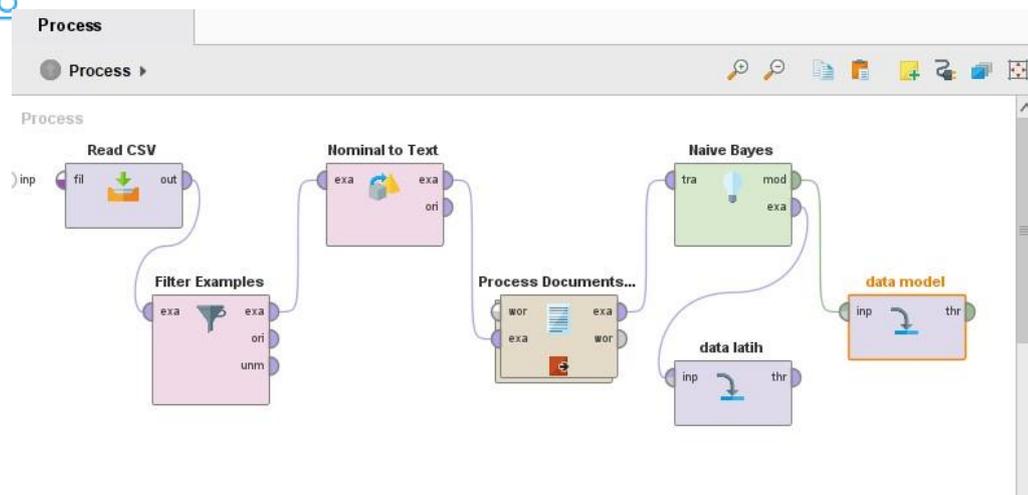
Taebel 3.7 dapat kita lihat di mana beberapa data yang telah dilakukan proses pelabelan secara manual, pada pekabelan manual peneliti melakukan pelabelan dengan menggunakan 3 orang untuk melakukan pelabelan tersebut, jika 2 orang menganggap positif dan 1 orang mengatakan negatif maka *sentimentya* positif begitupun sebaliknya.

3.1.4 Clasificasion Naïve Bayes

Data dalam bentuk tweet di ambil dari Twetter lalu disimpan dalam bentuk file csv. Data dibagikan menjadi dua set data, yaitu data latih dan data uji. Pelabelan akan disediakan untuk membedakan antara tweet negatif dan positif. Metode Naïve Bayes digunakan pada tahap klasifikasi *sentimen* dan interpretasi hasil analisis *sentiment*, selanjutnya data di splitting dengan perbandingan data

training dan data *testing* yang kemudian dikomputasi menggunakan *metode Naïve Bayes* sehingga menghasilkan luaran.

Tahapan awal dalam menggunakan metode *Naïve bayes* pada Rapid Miner adalah membuat model dari metode *Naïve bayes* tersebut yaitu dapat kita lihat pada gambar dibawah ini :



Gambar 3.5 Membuat model *Naïve bayes* pada Rapid Miner

Pada Gambar 3.5 dapat dilihat di mana model *Naïve Bayes* dibentuk yang pertama menggunakan operator yang bernama Read CSV didalam operator tersebut kita masukkan data latih yang sudah kita labeli secara manual selanjutnya menggunakan operator yang bernama *Filter Examples* yang berfungsi untuk menampilkan data yang hanya memiliki *sentimentnya* saja dengan pengaturan filter (*sentiment, is not missing*), kemudian menggunakan operator yang bernama Nominal to Text sebagai pengubah data nominal menjadi data Text, selanjutnya menggunakan operator Process Documents untuk memroses documents yang dalamnya berisi *Tokenize, Transform Cases, Filter Stopwords, Filter Tokens Bylength* selanjutnya kita buat modelnya menggunakan algoritma *Naïve bayes*

Hak Cipta Dilindungi Undang-Undang:

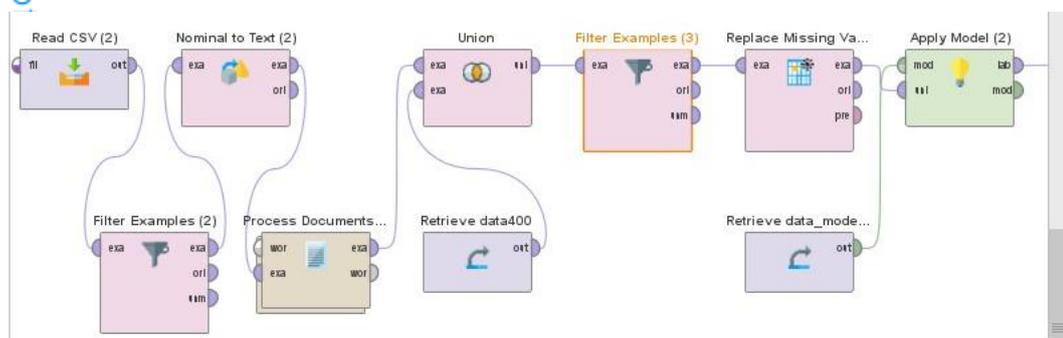
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber asli:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

dengan menggunakan operator *Naïve bayes* setelah model terbentuk barulah kita simpan model dan data latih tersebut menggunakan operator yang bernama Store. Dalam membuat model ini yang di ambil data yang sudah dilabeli sentimennya yang di ambil menggunakan operator *Filter Examples* untuk tampilan datanya yang di proses, tahap ini dapat dilihat pada gambar di bawah :

R...	sentiment	Text
1	negatif	iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru apalagi ikn loh yg mau dijadikan ibu kota pikir pikir lagi deh ibu K...
2	negatif	ibu kota mau pindah ke ikn tp proyek unggulan kayak gini aja bisa mangkrak yakin gak malu malu indonesia ?
3	negatif	Kagak warga ibu kota jgdi kota lain yg rumahnya pinggir rel atau yg pakai lahan punya TNI kalau disuruh pindah karena tanahnya atau...
4	positif	sekarang Ada istilah ibu kota di pindah
5	negatif	Klo warga sulit di pindah solusinya depo pertamina yg di pindahkan pemilik sah Tanah NKRI adalah rakyatnah tugas pemerintah ca...
6	positif	Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparatur si...
7	negatif	Ga ada urgensinya ibu kota pindah
8	positif	Skrg ibu kota Jateng pindah Solo makanya Semarang masuk wilayah Solo
9	negatif	Nambah macet Nambah rempong opo opo larangan Wis pindah ibu kota Arep ngurus apa apa susah Runyam
10	positif	Berantas korupsiTidak korupsi
11	positif	Bansos ada dimanamana
12	positif	Bansos tidak dikorupsi
13	positif	Wakil Ketua DPR Sebut Pemindahan IKN Bukan Sekadar Pindah Tempat
14	positif	Bahwa sebenarnya dari jaman Sukarno udah rencana pindah ke Palangkaraya cuma jaman pak Jokowi terealisasi Yg lain cuma bisa...
15	negatif	kita semua yg salah kok

Gambar 3.6 Data tahap *Filter Examples* dengan pengaturan (*sentiment, is not missing*)

Selanjutnya yaitu tahapan mengisi data yang sentimentnya masi kosong dengan menggunakan model *Naïve bayes* yang kita buat pada gambar 3.5 tahapannya dapat kita lihat pada gambar dibawah :



Gambar 3.7 Proses pelabelan dengan model dan data latih *Naïve Bayes*

Pada Gambar 3.7 ialah proses pelabelan dengan model dan datah latih *Naive Bayes* yang pertama pada tahapan ini adalah memasukan data menggunakan operator *Read CSV* kemudian data di filter dengan menggunakan operator yang bernama *Filter Examples* sama seperti saat membuat model pada tahap sebelumnya tadi namun pada *Filter Examples* ini kebalikan yang tadi yaitu hanya menampilkan data yang sentimennya kosong atau yang belum dilabeli contohnya pada gambar dibawah :

sentiment	Text
?	Tdk ada yg lbh pandir dr yg datang mengemis bk kntr di Jakarta tp akhirnya kantor perwakilan Tesla ASEAN di KL Malaysia Kemu...
?	Tidak mungkin ada kantor pusat Tesla di ASEAN ada di tempat di MalaysiaKL dan di Jakarta Penasehat s...
?	dalam rangka apa?
?	pemutakhiran hidup?
?	Catat Tidak mungkin ada kantor pusat Tesla di ASEAN ada di tempat di MalaysiaKL dan di Jakarta Penasehat sdh bilang Bos tu...
?	Pertamina Energy Tower thn berlalu Erick Thohir sbg Menteri BUMN yg baru yg berkuasa mengarahkan masa depan dan mimp...
?	Trus kalo punya pertamina gak bisa dipindah? Ibu kota aja bisa pindah
?	Rakyat Kalimantan bs sejahtera tanpa ibu kota hrs pindah Kalimantan sangat kaya terbukti wilayah malaysia dan Brunei rakyat s...
?	bhpsemarang
?	Infrastruktur
?	Kurang jauh main lu
?	kalau di jakarta sendiri tanah nya ada tapi yg berada di daerah banjir saja sih kalau mau yg ga banjir paling melipir dikit di daera...
?	Kenapa Soekarno tdk pindahkan Ibukota Jakarta ke Palangkaraya?
?	kita sibuk dgn urusan pribadi kita sampai lupa keselamatan rakyat kita
?	plumpang kita pindah ke ibu kota baru saja

Gambar 3.8 Data tahap *Filter Examples* dengan pengaturan
(*sentiment, no missing*)

Selanjutnya yaitu mengubah data yang mulanya nominal dirubah menjadi teks dengan menggunakan operator yang bernama *Nominal to Text* setelah data dirubah menjadi Teks berikutnya data tersebut diproses menggunakan operator yang bernama *Process Documents* untuk memroses documents yang dalamnya berisi *Tokenize, Transform Cases, Filter Stopwords, Filter Tokens Bylength* sama seperti saat membuat model sebelumnya kemudian operator *Union* diperlukan dalam proses ini untuk memasukkan atau menggabungkan data latih dan data uji

data latih di ambil dari data yang telah memasuki tahap pembentukan model seperti pada gambar 3.5 kemudian menggunakan sekali lagi operator yang bernama *Filter Examples* untuk menampilkan data yang *sentimennya* belum terisi setelah itu menggunakan operator yang bernama *Replace Missing Values* untuk mengubah data yang tidak sama atributnya atau masi tanda tanya dirubah menjadi angka 0, selanjutnya barulah model yang telah di buat di gunakan dengan menggunakan operator yang bernama *Apply Model* lalau di hubungkan model tersebut dan data latih beserta data yang belum mempunyai *sentimennya* barulah proses tersebut di jalankan yang mana Rapid Miner tersebut akan belajar dari data latih dan permodelan data tersebut untuk menampilkan prediksi dari data yang belum ada sentimennya.

3.1.5 Evaluasi

Tahapan ini adalah tahapan terakhir dalam penelitian ini yaitu evaluasi adalah proses memeriksa secara kritis hasil klasifikasi, kegiatan, kebijakan atau sejenisnya. Ini termasuk mengumpulkan informasi tentang kegiatan dan hasil klasifikasi. Tujuannya adalah untuk melakukan penilaian tentang suatu klasifikasi, meningkatkan keefektifannya, dan menimbang keputusan. Dari proses tersebut nantinya kita dapat mengetahui apakah metode yang digunakan sudah sesuai prediksinya.

a. *Split Validation*

Split validation merupakan bentuk klasifikasi *Naïve Bayes* yang berkerja untuk membagi dua data set yaitu data latih dan data tes dengan jumlah data latih



lebih besar dari data tes, data latih digunakan sebagai pelatih data dan data tes digunakan untuk data yang di uji (Karsito & Susanti Santi, 2019).

Gambar 3.9 Ilustrasi *Split Validation*



Cross Validation

Cross Validasion dilakukan untuk pengujian klasifikasi yang menggunakan 5 partisi atau disebut dengan $k = 5$, pengujian ini berkerja 4 sebagai data latih 1 sebagai data uji, yang disebut dengan 5 skema cross Validation.

Tabel 3.8 Ilustrasi skema *5-fold cross validation*

Partisi 1	Partisi 2	Partisi 3	Partisi 4	Partisi 5
Partisi 1	Partisi 2	Partisi 3	Partisi 4	Partisi 5
Partisi 1	Partisi 2	Partisi 3	Partisi 4	Partisi 5
Partisi 1	Partisi 2	Partisi 3	Partisi 4	Partisi 5
Partisi 1	Partisi 2	Partisi 3	Partisi 4	Partisi 5

data uji data latih

3.2 Alat Bantu yang Digunakan

Alat bantu yang digunakan peneliti :

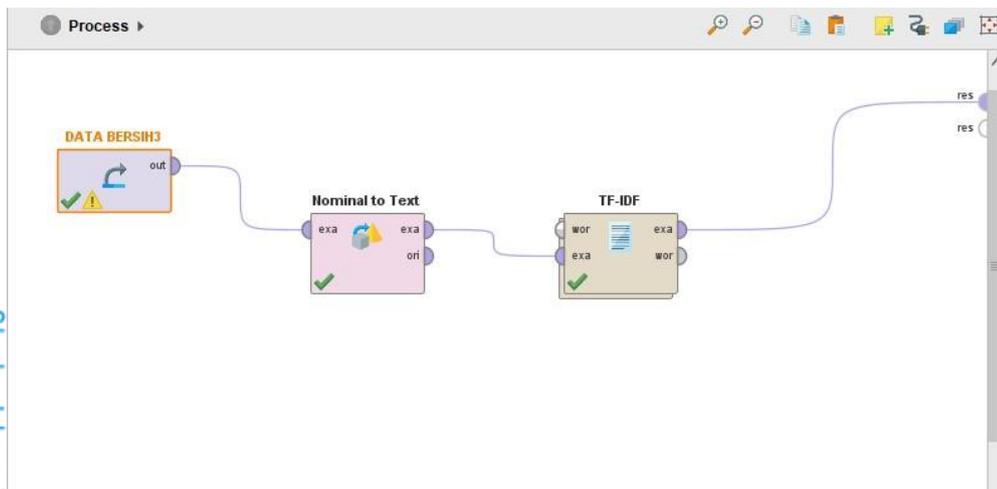
1. PC/Laptop dengan spesifikasi *processor* Intel(R) Core(TM) i5- 8250U CPU @ 1.60GHz 1.80GHz, RAM 8.00 GB, system operasi Windows 10,64-bit.
2. Aplikasi Rapid Miner.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pembahasan dan hasil dari Ekstraksi Fitur

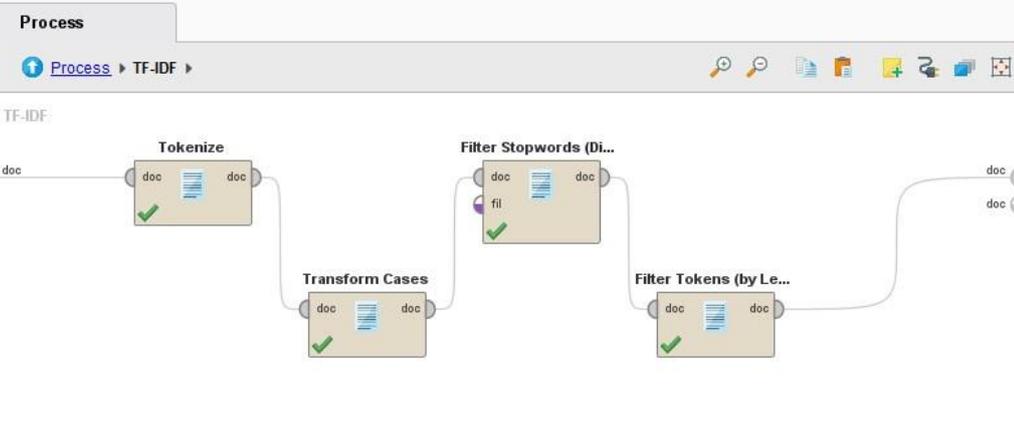
Setelah melalui tahapan *Cleansing* maka data tersebut memasuki tahap Ekstraksi Fitur yang mana di dalam proses tersebut ada istilah yang dinamakan TF-IDF (*Term frequency Inverse Document Frequency*), adapun proses atau tahap TF-IDF dalam Rapid di bawah ini contoh proses tf idf pada Gambar 4.1 berikut :



Gambar 4.1 Proses TF-IDF di Rapid Miner

Sebelum melakukan TF-IDF kita perlu mengubah data set yang sudah di *Cleansing* karena data set yang sudah di *Cleansing* tersebut masih bersifat Nominal sedangkan yang dibutuhkan untuk proses dokumen ini adalah tipe data teks maka data set tersebut harus kita ubah dari Nominal ke Teks dengan menggunakan operator yang bernama *Nominal to Text* seperti yang dapat kita lihat di Gambar 4.1, setelah itu barulah data tersebut bisa di lakukan TF-IDF dengan menggunakan operator yang bernama *Process Documents from Data*, di dalam

operator TF-IDF tersebut kita harus memasukkan beberapa operator lagi seperti yang dapat kita lihat pada gambar di bawah :



Gambar 4.2 Proses didalam operator TF-IDF

Pada Gambar 4.2 adalah gambar proses di dalam operator TF-IDF di mana terdapat beberapa operator lagi yang pertama adalah *Tokenize* yang mana operator ini berfungsi untuk memisahkan menjadi bagian- bagian kalimat yang terpisah atau kata-kata dengan tujuan untuk proses analisis teks selanjutnya adapun gambaran dari proses ini dapat kita lihat pada gambar sebagai berikut :

Open in Turbo Prep Auto Model Filter (1,475 / 1,475 examples): all

Row No.	ABUD	ADA	AI	AIR	AJA	AKAN	AKBP	ALAMI	AMDAL	AMPERA	ANAK	A
1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0

ExampleSet (1,475 examples, 0 special attributes, 5,520 regular attributes)

Gambar 4.3 Gambar hasil dari data yang di *Tokenize*

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

Pada Gambar 4.3 kita bisa melihat di mana kata setiap teks di dalam data tersebut dipecah menjadi kata perkata yang menghasilkan sebanyak 5.520 reguler atribut dari 1.475 data yang mana dapat kita lihat pada pojok kiri bawah gambar 4.3. Pada gambar di atas juga muncul banyak angka 0 angka tersebut merupakan data hasil TF-IDF yang menghitung berapa kali suatu kata dalam data tersebut muncul dalam sebuah dokumen, adapun beberapa contoh kata dari hasil pembobotan kata atau TF-IDF dapat kita lihat pada gambar sebagai berikut :

Word	Attribute Name	Total Occure... ↓	Document Occurences
IKN	IKN	885	824
di	di	477	403
yg	yg	239	183
dan	dan	228	204
kota	kota	181	149
Presiden	Presiden	164	156
yang	yang	160	140

Gambar 4.4 Contoh kata-kata hasil pembobotan TF-IDF

Pada Gambar 4.4 terdapat contoh kata kata yang sudah melalui tahap pembobotan kata atau TF-IDF di mana pada gambar di atas dapat kita lihat kata IKN muncul sebanyak 885 kali dalam 824 data atau dokumen *tweet*, begitu juga dengan kata-kata yang lain yang berhasil di TF-IDF.

Selanjutnya di karnakan kata kata tersebut masi bercampur antara huruf besar dan huruf kecil maka perlu di samaratakan sehingga menjadi huruf kecil semua supaya menjadi kata yang seragam dengan menggunakan operator yang bernama *Transform Cases*.

Setelah menyeragamkan huruf tahapan selanjutnya yaitu menghapus kata-kata yang tidak penting dan tidak mempunyai makna atau bisa disebut dengan

istilah *Stopwords* seperti kata-kata, yang, dan, di, adalah dan masi banyak lagi, untuk menghapus kata-kata tersebut kita membutuhkan operator yang bernama *Filter Stopwords(dictionary)* dimana nantinya didalam operator tersebut kita isikan kumpulan kata kata yang tidak penting tersebut. Yang terkhir dalam proses ini adalah menghilangkan kata kata yang singkat seperti kata-kata yang muncul 2 huruf (yg) kata kata seperti itu di hapus menggunakan operator yang bernama *Filter Tokens (by length)* dan berikut gambar kata kata yang telah dilakukan proses *Stopwords* dan *Filter Tokens (by length)* :

Word	Attribute Name	Total Occurences ↓	Document Occurences
ikn	ikn	977	914
kota	kota	282	234
presiden	presiden	216	203
pembangunan	pembangunan	186	176
pindah	pindah	175	162
nusantara	nusantara	160	157
jokowi	jokowi	158	147
indonesia	indonesia	104	103
negara	negara	99	89
menteri	menteri	95	84
bangun	bangun	77	70

Gambar 4.5 Kata-kata yang sudah di *Stopwors* dan *Filter Tokens*

Pada Gambar 4.5 kita dapat melihat di mana kata-kata di gambar tersebut sudah melalui berbagai proses sehingga hanya menampilkan kata-kata yang berharga dan juka menjadi trending topik dalam kedua kata kunci IKN dan IBU KOTA PINDAH. Supaya tampilannya lebih menarik maka peneliti akan menampilkan sebuah visualisasi yang cukup tren di analisis *sentiment* yaitu berupa *Wordcloud* :



Gambar 4.6 *Wordcloud*

Pada Gambar 4.6 di atas peneliti mengambil 30 contoh kata teratas kecuali kata IKN karena kata IKN adalah kata kunci, kita dapat melihat sebuah kumpulan kata-kata yang banyak muncul atau trending di dalam teks twiiter di mana semakin besar kata-kata pada gambar *Wordcloud* maka semakin besar bobot kata-kata tersebut begitupun sebaliknya. Pada gambar juga terdapat kata-kata lain yang mana kata-kata yang muncul seperti contohnya kata-kata seperti presiden Jokowi ternyata menjadi trending topik atau banyak di sebut dalam Twitter juga.

4.2 Hasil Permodelan Metode *Naiïve Bayes*

Dari data keseluruhan yang tersisa setelah melalui berbagai tahap yang telah dipaparkan sebelumnya tersisalah data sebanyak 1.449 data. Dari data 1.449 tersebut dilakukan 2 kali pengujian yang pertama yang mana 1000 sebagai data latih dan 449 sebagai data uji sedangkan pengujian yang kedua 1.200 sebagai data latih 249 sebagai data uji baru di dapatlah hasil *Accuracy*, presisi dan *Recall* yang dihitung menggunakan rumus :

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

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

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

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

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

Dimana :

- TP adalah True positif, yaitu jumlah data yang terklasifikasi dengan benar oleh sistem.
- TN adalah True negatif, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN adalah false negatif, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
- FP adalah false positif, yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Untuk hasil pengujian manual dapat kita lihat pada tabel dan penjelasan dibawah:

Tabel 4.1 Hasil uji data uji 449 dan data latih 1000

	True negatif	True positif
pred.negatif	67	70
pred.positif	44	268
Total	449	

$$Accuracy = \frac{268+67}{268+67+70+44} * 100\%$$

$$= 74.61\%$$

$$Presisi = \frac{268}{70+268} * 100\%$$

$$= 79.28\%$$

$$Recall = \frac{268}{44+268} * 100\%$$

$$= 85.89\%$$

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

Tabel 4.2 Hasil uji data uji 249 dan data latih 1200

	True negatif	True positif
pred.negatif	41	39
pred.positif	20	149
Total	249	

$$Accuracy = \frac{149+41}{149+41+39+20} * 100\%$$

$$= 76.30\%$$

$$Presisi = \frac{149}{39+149} * 100\%$$

$$= 79.25\%$$

$$Recall = \frac{149}{20+149} * 100\%$$

$$= 79.25\%$$

Berdasarkan pengujian data di atas di mana yang pertama dengan jumlah data latih 1000 dan data uji 449 atau 69% : 31% di dapatkan nilai *Acuraccy* sebesar 74.61%, Presisi 79.28%, *Recall* 85.89% dan pengujian yang kedua dengan data latih 1200 dan data uji 249 atau 83% : 17% didapatkan nilai *Acuraccy* sebesar 76.30%, Presisi 79.25%, *Recall* 79.25%.

4.3 Hasil pengujian *Split Validation* oleh sistem

Untuk hasil pengujian dari sistem dapat kita lihat pada gambar dan penjelasan dibawah :

accuracy: 70.30%

	true negatif	true positif	class precision
pred. negatif	115	142	44.75%
pred. positif	73	394	84.37%
class recall	61.17%	73.51%	

Gambar 4.7 Pengujian data 50% : 50%

Berdasarkan hasil pengujian yang dapat kita lihat pada Gambar 4.7 di mana pengujian di lakukan dengan membagi data uji dan data latih dengan perbandingan 50% :50% di dapatkan nilai *Accuracy* sebesar 70.30%, Presisi 73.51% dan *Recall* 84.37%.

accuracy: 73.79%

	true negatif	true positif	class precision
pred. negatif	78	79	49.68%
pred. positif	35	243	87.41%
class recall	69.03%	75.47%	

Gambar 4.8 Pengujian data 70% : 30%

Hasil pengujian pada Gambar 4.8 di mana pengujian dilakukan dengan membagi data uji dan data latih dengan perbandingan 70% data latih dan 30% data uji, didapatkan nilai *Accuracy* sebesar 73.79%, Presisi 75.47%, *Recall* 87.41%.

Berdasarkan kedua pengujian di atas menunjukkan bahwa semakin besar data latih yang digunakan untuk tahap pengklasifikasian maka semakin besar pula nilai *Accuracy* yang di dapat dengan dibuktikan dua tahapan pengujian *Split Validation* di atas yaitu pengujian secara manual dan pengujian secara otomatis oleh sistem Rapid Miner.

4.4 Hasil pengujian *Cross Validation*

Setelah melakukan pengujian dengan *Split validation* selanjutnya pengujian dengan menggunakan *Cross Validation* di mana pada pengujian ini semua data 1449 di dapatkan *accuracy* sebesar 71.98% yang mana dapat kita lihat pada Gambar 4.9 dibawah ini.

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

accuracy: 71.98% +/- 1.71% (micro average: 71.98%)

	true negatif	true positif	class precision
pred. negatif	248	278	47.15%
pred. positif	128	795	86.13%
class recall	65.96%	74.09%	

Gambar 4.9 pengujian *Cross Validation*

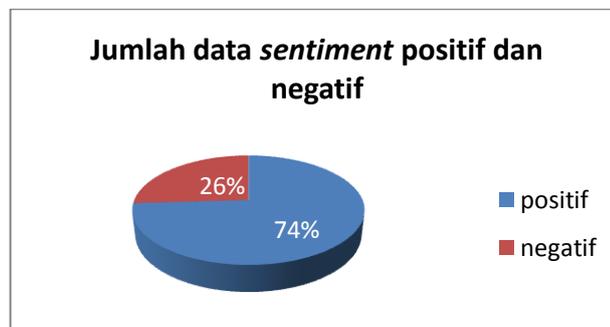
4.5 Hasil percobaan data baru

Selanjutnya yaitu peneliti mencoba memasukkan data baru kedalam klasifikasi ini untuk melihat bagaimana hasil prediksinya adapun data dan prediksinya dapat dilihat pada tabel di bawah :

Tabel 4.3 Percobaan tambah data baru

No	Text	Prediksi	Hasil
1	Indonesia makin maju karena ikn	Positif	Benar
2	Semangat untuk pembangunannya	Positif	Benar
3	Saya bangga jadi warga indonesia	Negatif	Salah
4	Wiih ibu kota baru pasti mantapni	Negatif	Salah
5	Ikn itu gak terlalu penting menghabiskan uang saja	Negatif	Benar
6	Ikn gak penting	Negatif	Benar
7	Indonesia bisa bobrok kalau maksa bangun ikn	Negatif	Benar
8	Menyusahkan rakyat	Negatif	Benar

Berdasarkan data di atas kita dapat melihat dari 8 data di atas yang betul terprediksi dengan benar adalah 6 data dan 2 data salah.



Gambar 4.10 Jumlah *sentiment* positif dan negatif

Berdasarkan gambar diagram di atas menampilkan jumlah data yang bernada positif dan negatif dengan keseluruhan data 1449, *sentiment* yang bernada positif berjumlah 1073 atau 74% dan *sentiment* negatif berjumlah 376 atau 26% ini menunjukkan bahwa *sentiment* yang bernada positif lebih dominan tinggi.

4.6 Evaluasi

Tabel 4.4 Hasil Evaluasi

Perbandingan Pengujian	Acuraccy	Precision	Recall
Manual, data uji 449 dan data latih 1000	74.61%	79.28%	85.89%
Manual, data uji 249 dan data latih 1200	76.30%	79.25%	79.25%
Split Validation, 50% : 50%	70.30%	73.51%	84.37%
Split Validation, 70% : 30%	73.79%	75.47%	87.41%
Cross Validation keseluruhan data	71.98%	74.09%	86.13%

pada Tabel 4.4 telah di lakukan beberapa kali, Pengukuran akurasi menggunakan *confusion matrik* dengan menggunakan algoritma Naïve Bayes Mendapatkan hasil yakni semakin banyak data latih yang digunakan pada tahap pengklasifikasian semakin tinggi akurasi yang di dapat di buktikan dengan pengujian manual data uji 449 mendapatkan akurasi 74.61%, data uji 249 mendapatkan akurasi 76.30% dan pengujian *split validation* oleh sistem data 50% : 50% di dapatkan akurasi 70.30%, data 70% : 30% mendapat akurasi 73.79%. *Cross Validation* menggunakan *fold-5* mendapatkan akurasi 71.98%.

Selanjutnya pengujian dengan memasukkan data baru dari 8 data di yang betul terprediksi dengan benar adalah 6 data dan 2 data salah. Dari 1449 data *sentiment* yang bernada positif berjumlah 1073 atau 74% dan *sentiment* negatif berjumlah 376 atau 26%.

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

Berdasarkan penelitian yang telah dilakukan di mana pada penelitian ini terdapat kekurangan dan kelebihan untuk kekeurangannya sendiri di sini peneliti hanya menggunakan sentiment positif dan negatif saja sedangkan sentiment netral tidak. Untuk kelebihannya pada penelitian ini adalah di mana pada penelitian ini menggunakan *double* pengujian *split validation* di mana data di uji secara manual dan sistem supaya dapat melihat hasil kedua pengujiannya.

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 sebagian dan atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suntho Jambi



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Metode Naïve Bayes dapat digunakan sebagai metode untuk mengklasifikasi *sentiment*. Analisis dengan metode *Naïve Bayes* di mulai dengan *crawling* data dari Twitter pada tanggal 6 maret 2023 sebanyak 4533 data mentah dengan 2 kata kunci IKN dan IBU KOTA PINDAH. Setelah *crawling* selanjutnya pembersihan data atau *preprocessing* yang terdiri dari *cleansing*, *stopword*, dan *casefolding* sehingga menyisakan data yang bersih sebanyak 1449 data. Tahapan terakhir yaitu mengklasifikasikan data dengan metode *Naïve Bayes* dengan menggunakan aplikasi Rapid Miner, pelabelan manual diperlukan sebelum menggunakan metode selanjutnya data di bagi menjadi dua yaitu data latih dan data uji data latih digunakan untuk memberikan pengetahuan kepada rapid miner dalam menggolongkan suatu data sehingga data uji tersebut dapat terprediksi oleh aplikasi Rapid Miner, dari pengujian yang sudah di lakukan dapat disimpulkan bahwa semakin banyak data latih yang digunakan semakin tinggi *Acuraccynya*.

Atas kebijakan pemerintah yang ingin memindahkan Ibu kota Indonesia menimbulkan beragam respon positif dan negatif dari masyarakat, *sentiment* yang paling banyak yaitu positif sebesar 1073 (74%), *sentiment* negatif ditemukan sebanyak 376 (26%). Banyaknya *sentiment* positif dengan jumlah *sentiment* negatif ini menunjukkan bahwa masyarakat Indonesia tidak menutup kemungkinan menyetujui atas kebijakan pemerintah dalam pemindahan Ibu kota Indonesia.

5.2 Saran

Bedasarkan penelitian yang telah dilakukan maka penulis memverikan saran untuk peneliti selanjutnya :

1. Pada penelitian berikutnya diharapkan peneliti menggunakan lebih banyak data supaya lebih bagus, menambahkan sentiment netral dan data emoticon.
2. Menggunakan algoritma klasifikasi yang lain sehingga dapat membandingkan hasil dari klasifikasinya.
3. Pada pelabelan data secara manual diharapkan untuk menggunakan orang yang ahli dalam bahasa indonesia sehingga pelabelan lebih bagus.

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



DAFTAR PUSTAKA

- Agustina, M. P., & Hendry, H. (2021). Sentimen Masyarakat Model Random Forest dan Logistic Regression. *Aiti*, 18(2), 111–124. <https://doi.org/10.24246/aiti.v18i2.111-124>
- Asro'i, A., & Februariyanti, H. (2022). Analisis Sentimen Pengguna Twitter Terhadap Perpanjangan PpkM Menggunakan Metode K-Nearest Neighbor. *Jurnal Khatulistiwa Informatika*, 10(1), 17–24. <https://doi.org/10.31294/jki.v10i1.12624>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Gifari, O. I., Adha, M., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40. <https://doi.org/10.46229/jifotech.v2i1.330>
- Handayani, E. T., & Sulistiyawati, A. (2021). Analisis Sentimen Respon Masyarakat terhadap Kabar Harian Covid-19 pada Twitter Kementerian Kesehatan dengan Metode Klasifikasi Naive Bayes. *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, 2(3), 32–37. <https://doi.org/10.33365/jtsi.v2i3.906>
- Karsito, & Susanti Santi. (2019). Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naive Bayes Di Perumahan Azzura Residence. *Jurnal Teknologi Pelita Bangsa*, 9, 43–48.
- Kevin, V., Que, S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization). 9(2), 162–170.
- Krisdiyanto, T. (2021). Analisis sentimen opini masyarakat Indonesia terhadap kebijakan PPKM pada media sosial Twitter menggunakan Naive bayes classifiers. *Jurnal CoreIT*, 32–37.
- Pintoko, B. M., & L., K. M. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier. *E-Proceeding of Engineering*, 5(3), 8121–8130.
- Prabowo, C. H. (2021). Ibukota DKI Jakarta dalam Menghadapi Pandemi COVID 19. *Jurnal Syntax Admiration*, 2(1), 38–48.
- Rochmanto, H. B., & Hermanto, E. M. P. (2020). Analisis Sentimen Pengguna Twitter Menggunakan Metode NBC dan SVM. *Prosiding Seminar Nasional Statistika IX (SNS) 2020*, 9(2020), 1–11. <http://website.com>
- Samsir¹, Ambiyar², Unung Verawardina³, Firman Edi⁴, R. W., & 1. (2021). Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian 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 sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suttha Jambi

COVID-19 Menggunakan Metode Naïve Bayes. *Jurnal Media Informatika Budidarma*, 5(1), 157–163. <https://doi.org/10.30865/mib.v5i1.2604>

Syahidin, S., & Adnan, A. (2022). Analisis Pengaruh Harga Dan Lokasi Terhadap Kepuasan Pelanggan Pada Bengkel Andika Teknik Kemili Bebesen Takengon. *Gajah Putih Journal of Economics Review*, 4(1), 20–32. <https://doi.org/10.55542/gpjer.v4i1.209>

Vonega, D. A., Fadila, A., & Kurniawan, D. E. (2022). *Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024*. 6(2).

Watrianthos, R., Giatman, M., Simatupang, W., Syafriyati, R., & Daulay, N. K. (2022). Analisis Sentimen Pembelajaran Campuran Menggunakan Twitter Data. *Jurnal Media Informatika Budidarma*, 6(1), 166. <https://doi.org/10.30865/mib.v6i1.3383>

Wiwik, S. (2022). *PERBANDINGAN METODE RANDOM FOREST DAN NAÏVE BAYES CLASSIFIER PADA ANALISIS SENTIMEN ISSUE CLIMATE CHANGE*. Universitas Muhammadiyah Semarang.

Yahya, M. (2018). Pemindahan Ibu Kota Negara Maju dan Sejahtera. *Jurnal Studi Agama Dan Masyarakat*, 14(1), 21. <https://doi.org/10.23971/jsam.v14i1.779>

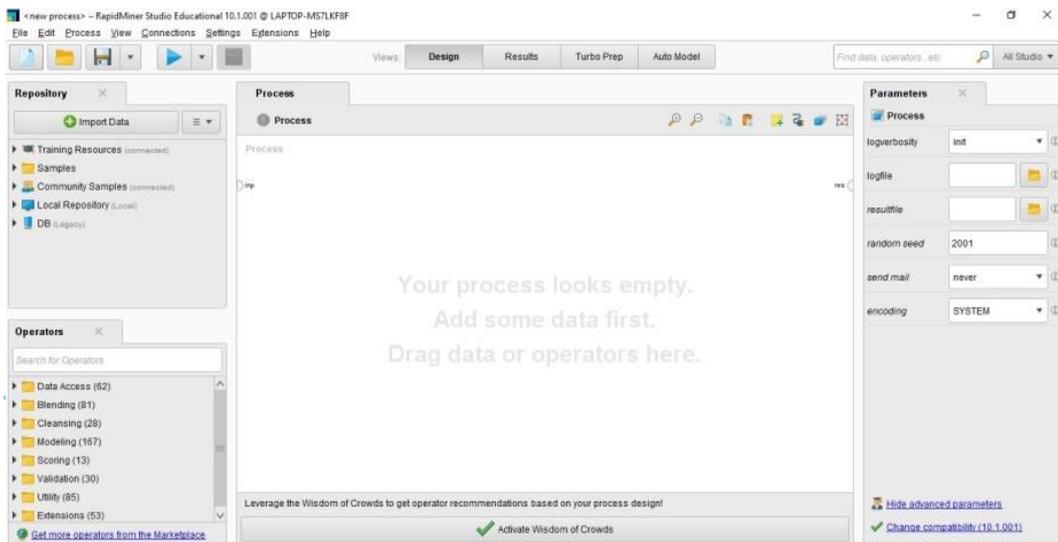
Yulita, W., Dwi Nugroho, E., Habib Algifari, M., Studi Teknik Informatika, P., Teknologi Sumatera, I., Terusan Ryacudu, J., Huwi, W., Agung, J., & Selatan, L. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *Jdmsi*, 2(2), 1–9.

LAMPIRAN

Lampiran 1 Tampilan logo dari aplikasi Rapid Miner

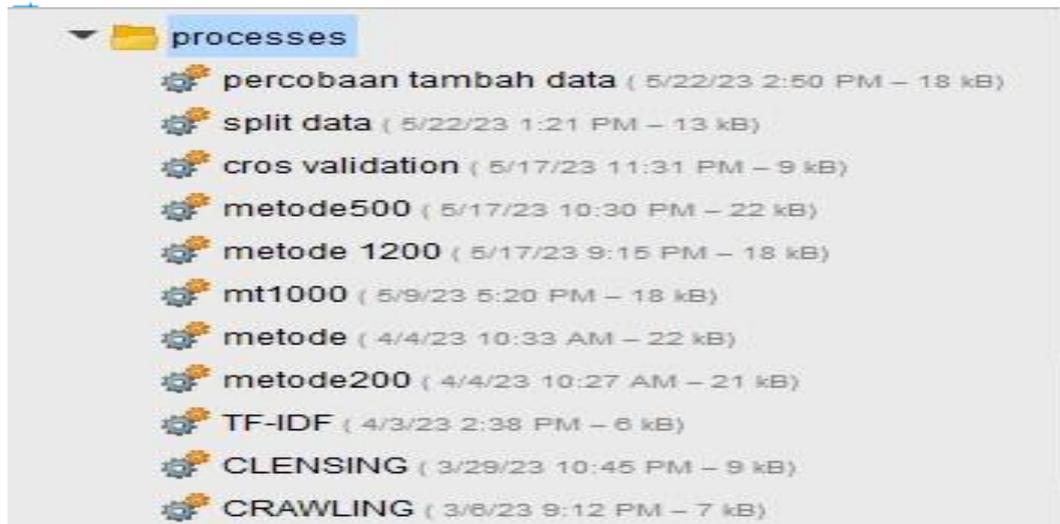


Lampiran 2 Tampilan awal Rapid Miner

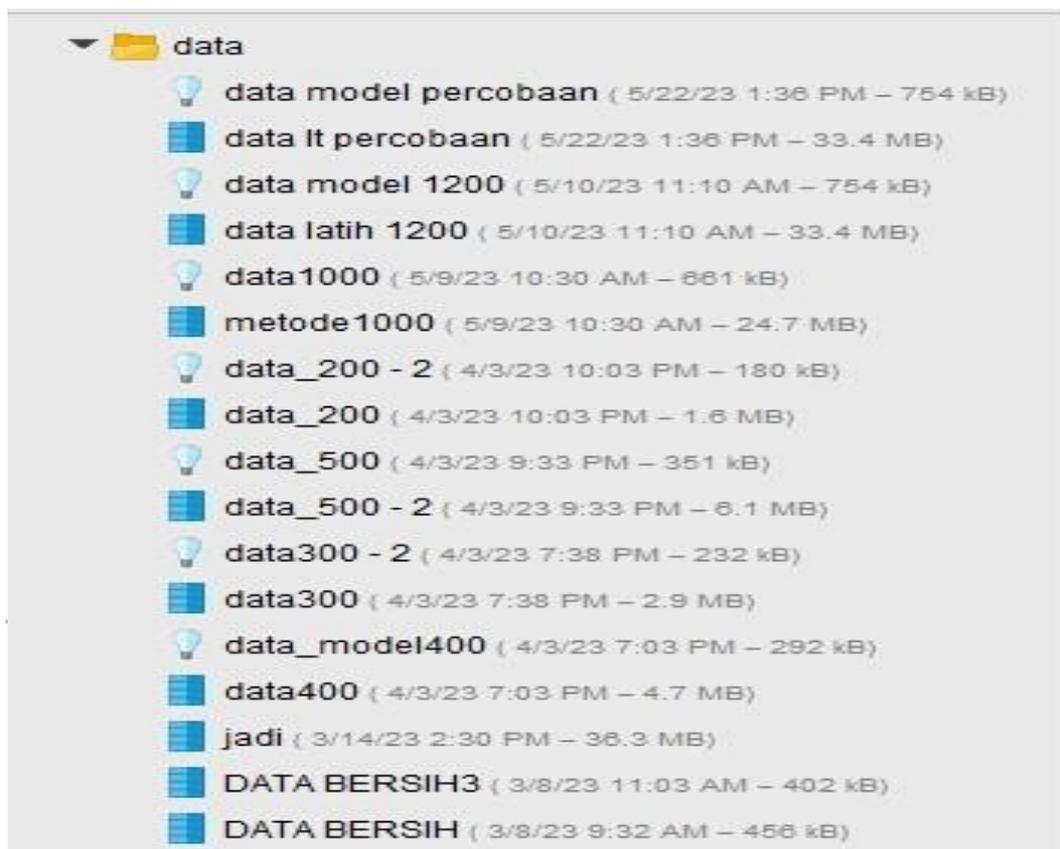


- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber aslinya:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.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

Lampiran 3 Tampilan proses-proses pada Rapid Miner



Lampiran 4 Tampilan data model dan data latih pada Rapid Miner



aha Saifuddin Jambi

@ Hak cipta



UNIVERSITAS ISLAM NEGERI
SULTHAN THAHA SAIFUDDIN
J A M B I

Hak Cipta Dilindungi Undang-Undang:

1. Dilarang mengutip sebagian 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 sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suttha Jambi

Lampiran 5 Tampilan beberapa data kotor yang di tarik dari Twitter

	A
4	RT @Ronnie_Rusli: @geloraco Catat: Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta. Penasehat s...
5	RT @Ronnie_Rusli: @geloraco Catat: Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta. Penasehat s...
6	@desliyani_ni Iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru, apalagi ikn loh yg mau dijadikan ibu kota. pikir pikir lagi deh ibu kota pindah
7	RT @Ronnie_Rusli: @geloraco Catat: Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta. Penasehat s...
8	@daffanacom ibu kota mau pindah ke ikn. tp proyek unggulan kayak gini aja bisa mangkrak. yakin gak malu malu Indonesia ?
9	RT @Ronnie_Rusli: @geloraco Catat: Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta. Penasehat s...
10	RT @Ronnie_Rusli: @geloraco Catat: Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta. Penasehat s...
11	@rahmawaadityan Kakag warga ibu kota jg,di kota lain yg rumahnya pinggir rel atau yg pakai lahan punya TNI kalau disuruh pindah karena tanahnya atau rumahnya atau lahannya mau di pakai pasti ngai
12	@sociotalker Seharusnya mumpung ada rencana pindah ibu kota ke Kalimantan, sekalian aja reset NPP buat dibangun di Kalimantan yang notabenernya 'aman' dari aktivitas seismik. Toh Jepang yg jelas2 c
13	sekarang Ada istilah ibu kota di pindah
14	dalam rangka apa?
15	permutakhiran hidup?
16	@geloraco Catat: Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta. Penasehat sdb bilang Bos tuch Ibu Kota mau pindah "into the Woods in East Borneo." Jad
17	@CNWIndonesia Klo warga sulit di pindah, solusinya depo Pertamina yg di pindah...ingat pemilih sah Tanah NKRI adalah rakyat...nah tugas pemerintah cari solusi pindahin depo tsb, pindahin ibu kota saja
18	.
19	@keraktelor_Pertamina Energy Tower 6 thn berlalu, Erick Thohir sbg Menteri BUMN yg baru & yg berkuasa mengarahkan masa depan dan mimpi Pertamina, termasuk memutuskan apakah gedung itu aka
20	RT @VIVAcoid: Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparat sipil...
21	@nusantara_one Trus kalo punya Pertamina gak bisa dipindah? Ibu kota aja bisa pindah
22	@okezonenews Ga ada urgensinya ibu kota pindah.
23	@irhamnapermana @Muhammad_Saewad Rakyat Kalimantan bs sejahtera tanpa ibu kota hrs pindah, Kalimantan sangat kaya terbukti wilayah Malaysia dan Brunei rakyat sangat sejahtera dan makmur, t
24	RT @VIVAcoid: Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparat sipil...
25	@arifganiesta73 @Heraloebss @Puspem_TNI @glibran_tweet Skrg ibu kota Jateng pindah Solo, makanya Semarang masuk wilayah Solo
26	@tanyakanrl Nambah macet. Nambah repong opo opo larangan. Wis pindah ibu kota Arep ngurus apa apa susah. Runyam
27	RT @VIVAcoid: Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparat sipil...
28	@Kemenkumham_RI @bhp_semarang
29	RT @VIVAcoid: Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparat sipil...

Lampiran 6 Tampilan beberapa data yang sudah melalui tahap preprocessing

2	Tdk ada yg lbh pandir dr yg datang mengemis bk kntr di Jakarta tp akhirnya kantor perwakilan Tesla ASEAN di KL Malaysia Kemungkinan mrk jd sdb dpt info negatif bikin utang atas nama kantor di Ibu kota Jkt tp akhirn
3	Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta Penasehat s...
4	Iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru apalagi ikn loh yg mau dijadikan ibu kota pikir pikir lagi deh ibu kota pindah
5	Ibu kota mau pindah ke ikn tp proyek unggulan kayak gini aja bisa mangkrak yakin gak malu malu Indonesia ?
6	Kagak warga ibu kota jgdi kota lain yg rumahnya pinggir rel atau yg pakai lahan punya TNI kalau disuruh pindah karena tanahnya atau rumahnya atau lahannya mau di pakai pasti ngamukngamuk nggak terimayg KAI m.
7	Seharusnya mumpung ada rencana pindah ibu kota ke Kalimantan sekalian aja reset NPP buat dibangun di Kalimantan yang notabenernya 'aman' dari aktivitas seismik Toh Jepang yg jelas di RoF aja berani bikin NPP
8	sekarang Ada istilah Ibu kota di pindah
9	dalam rangka apa?
10	permutakhiran hidup?
11	Catat Tidak mungkin ada kantor pusat Tesla di ASEAN ada di 2 tempat di Malaysia/KL dan di Jakarta Penasehat sdb bilang Bos tuch Ibu Kota mau pindah "into the Woods in East Borneo" Jadi hasil kunjunganya Blio menol
12	Klo warga sulit di pindah solusinya depo Pertamina yg di pindahingat pemilih sah Tanah NKRI adalah rakyatnah tugas pemerintah cari solusi pindahin depo tsb pindahin ibu kota saja bisa kenapa pindahin depo rumit a
13	Pertamina Energy Tower 6 thn berlalu Erick Thohir sbg Menteri BUMN yg baru yg berkuasa mengarahkan masa depan dan mimpi Pertamina termasuk memutuskan apakah gedung itu akan diteruskan atau dihentikan T
14	Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparat sipil...
15	Trus kalo punya Pertamina gak bisa dipindah? Ibu kota aja bisa pindah
16	Ga ada urgensinya ibu kota pindah
17	Rakyat Kalimantan bs sejahtera tanpa ibu kota hrs pindah Kalimantan sangat kaya terbukti wilayah Malaysia dan Brunei rakyat sangat sejahtera dan makmur bukan iKN yg akan bikin rakyat Kalimantan bs sejahtera tp ;
18	Skrg ibu kota Jateng pindah Solo makanya Semarang masuk wilayah Solo
19	Nambah macet Nambah repong opo opo larangan Wis pindah ibu kota Arep ngurus apa apa susah Runyam
20	bhpsemarang
21	Infrastruktur
22	Berantas korupsiTidak korupsi
23	Bansos ada dimanamana
24	Bansos tidak dikorupsi
25	Wakil Ketua DPR Sebut Pemindahan iKN Bukan Sekadar Pindah Tempat
26	Kurang jauh main lu
27	Bahwa sebenarnya dari jaman Sukarno udah rencana pindah ke Palangkaraya cuma jaman pak Jokowi terealisasi Yg lain cuma bisa wacana doang

Lampiran 7 Data pelabelan 3 orang

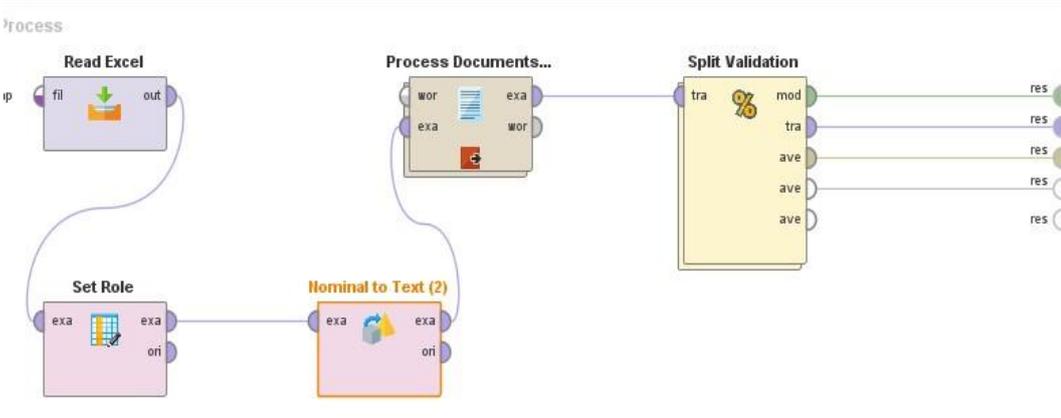
A	B	C	D	E
3 Tidak mungkin ada kantor pusat Tesla di ASEAN ada di tempat di MalaysiaKL dan di Jakarta Penasehat s...	negatif	negatif	negatif	negatif
4 iya lah klo distop otomatis ini banyak ngalamin kendala masalah baru apalagi ikn loh yg mau dijadiin ibu kota pikir pikir lagi deh ibu kc	negatif	negatif	negatif	negatif
5 ibu kota mau pindah ke ikn tp proyek unggulan kayak gini aja bisa mangkrak yakin gak malu malu in Indonesia ?	negatif	negatif	negatif	negatif
6 Kagak warga ibu kota jgdi kota lain yg rumahnya pinggir rel atau yg pakai lahan punya TNI kalau disuruh pindah karena tanahnya atau r	negatif	negatif	negatif	negatif
7 Seharusnya mumpung ada rencana pindah ibu kota ke Kalimantan sekalian aja reset NPP buat dibangun di Kalimantan yang notabene	positif	positif	positif	positif
8 sekarang Ada istilah ibu kota di pindah	positif	negatif	positif	positif
9 dalam rangka apa?	negatif	negatif	positif	negatif
10 pemutakhiran hidup?	positif	positif	positif	positif
11 Catat Tidak mungkin ada kantor pusat Tesla di ASEAN ada di tempat di MalaysiaKL dan di Jakarta Penasehat sdh bilang Bos tuch Ibu Ko	negatif	negatif	negatif	negatif
12 Klo warga sulit di pindah solusinya depo Pertamina yg di pindahingat pemilik sah Tanah NKRI adalah rakyatnah tugas pemerintah cari :	negatif	negatif	positif	negatif
13 Pertamina Energy Tower thn berlalu Erick Thohir sbg Menteri BUMN yg baru yg berkuasa mengarahkan masa depan dan mimpi Pertar	positif	positif	positif	positif
14 Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Abdullah Azwar Anas mengklaim rencana pemindahan aparatur sipi	positif	positif	positif	positif
15 Trus kalo punya Pertamina gak bisa dipindah? Ibu kota aja bisa pindah	positif	positif	positif	positif
16 Ga ada urgensinya ibu kota pindah	negatif	negatif	negatif	negatif
17 Rakyat Kalimantan bs sejahtera tanpa ibu kota hrs pindah Kalimantan sangat kaya terbukti wilayah Malaysia dan Brunei rakyat sangat s	negatif	negatif	negatif	negatif
18 Skrg ibu kota Jateng pindah Solo makanya Semarang masuk wilayah Solo	positif	positif	positif	positif
19 Nambah macet Nambah rempong opo opo larangan Wis pindah ibu kota Arep ngurus apa apa susah Runyam	negatif	negatif	negatif	negatif
20 bhpsemarang	positif	positif	positif	positif
21 Infrastruktur	positif	positif	positif	positif
22 Berantas korupsiTidak korupsi	positif	positif	negatif	positif
23 Bansos ada dimanamana	positif	positif	positif	positif
24 Bansos tidak dikorupsi	positif	positif	positif	positif
25 Wakil Ketua DPR Sebut Pemindahan IKN Bukan Sekadar Pindah Tempat	positif	positif	positif	positif
26 Kurang jauh main lu	negatif	positif	negatif	negatif
27 Bahwa sebenarnya dari jaman Sukarno udah rencana pindah ke Palangkaraya cuma jaman pak Jokowi terealisasi Yg lain cuma bisa wac	positif	negatif	positif	positif
28 kalau di Jakarta sendiri tanah nya ada tapi yg berada di daerah banjir saja sih kalau mau yg ga banjir paling melipir dikit di daerah beka	positif	positif	positif	positif
29 Kenapa Soekarno tdk pindahkan Ibukota Jakarta ke Palangkaraya?	positif	positif	positif	positif
30 kita semua yg salah kok	negatif	negatif	negatif	negatif

- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber aslinya:
 - 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 atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Sunha Jambi

Lampiran 8 Gambaran untuk mencari nilai accuracy manual dalam data

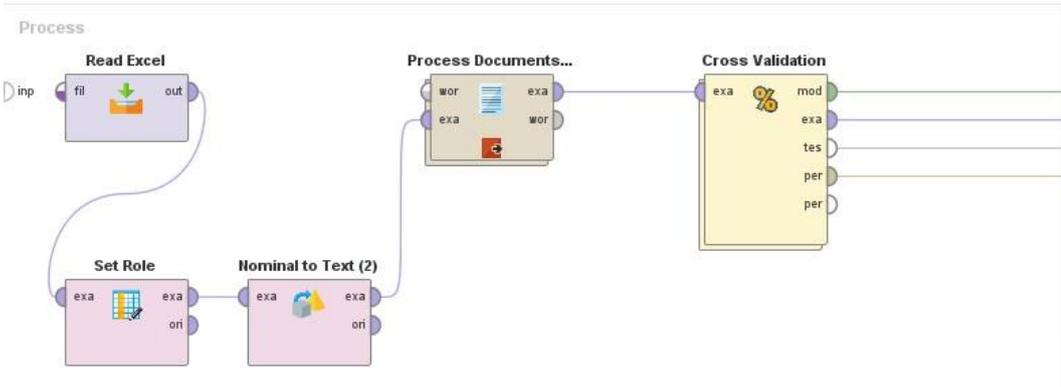
A	B	C	D	E	F
2 bangun ikn butuh mengetahui yg mendalam yg yg mudah termakan isu hoaks mmg tdk senang dgn p Jokowi	negatif	n			
3 upacara bendera Agustus ikn	positif		p		
4 mpr bersidang ibukota negara	negatif				n
5 emang yg menggagas pindah ikn sukar maunya pindah palangkaraya yg disiapkan layoutnya	positif			p	
6 ikn nusantara kemajuan Indonesia	positif		p		
7 presiden Jokowi pemindahan kota Jakarta ikn provinsi Kalimantan timur gag	negatif				n
8 kementerian pupr firlandia bahas kerja pembangunan ikn	positif		p		
9 peninjauan pembangunan kawasan kota nusantara ikn presiden iriana makan siang pekerja proyek pembangunan ikn fasilitas kantin pekerja	positif		p		
10 semoga pembangunan Istana negara ikn lancar cepat selesai	positif		p		
11 optimis Agustus upacara bendera ikn	positif		p		
12 program ikn otak	positif			p	
13 kehadiran ikn nusantara diyakini membawa perubahan Kalimantan	positif		p		
14 kriteria kaltim kota negara dukung penuh pembangunan ikn nusantara	positif		p		
15 presiden negara kemarin menginap kawasan inti ikn foto sejarah Indonesia	positif		p		
16 ikn dibangun lapangan sepak bola Jokowi dibiayai FIFA	positif		p		
17 otomatis dpake persija trus donk klo ikn bgun stadion jg	negatif				n
18 menteri atrkepala bpn hadi tjahjanto menteri investasikepala bkpm bahllil lahadalia menteri bumh erick thohir sekretaris kabinet pramono	positif		p		
19 ikn mangkrak bunga hutang bengkak	negatif	n			
20 bingung proyek ikn digarap kyk gini wkwk	positif		p		
21 pembenci Jokowi ikn gagal alias mangkrak	negatif	n			
22 kepala negara pembangunan lapangan sepak bola dibiayai federasi sepak bola internasional FIFA kepala negara menilai bukti kepercayaan in	positif		p		
23 presiden Jokowi pemindahan kota Jakarta ikn provinsi Kalimantan timur gagasan	positif		p		
24 ikn bagunan berdiridipikirnya mendirikannya pakai jembut	negatif	n			
25 said didu pembangunan ikn sebatas proyek ambisius Jokowi kereta cepat esemka	negatif	n			
26 betapa kelompok ikn galdan bertepuk tangan	positif			p	
27 nunggu indomart soft opening ikn	positif		p		
28 ujungujungnya ikn gagasan	negatif	n			
29 gempa d turki komentar layaknya pakar gempa skrg bahas ikn ahli planologi lucu aneh	positif		p		

Lampiran 9 Tahaan Split Validation pada Rapid Miner



- Hak Cipta Dilindungi Undang-Undang:
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan da menyebutkan sumber aslinya:
 - 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 sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suttha Jambi

Lampiran 10 Tahapan Cross Validation pada Rapid Miner



Lampiran 11 Kamus Stopword

	A	B	C	D	E	F	G	H	I	J	K	L	
1	ada	asal	beginian	berikan	besar	dapat	diinginkan	dipergunakan	dituturkannya	ibaratkan	jika	kecil	kok
2	adalah	asalkan	beginikah	berikut	betul	dari	dijawab	diperkirakan	diucapkan	ibaratnya	jikalau	kedua	kurar
3	adanya	atas	beginilah	berikutnya	betuluh	daripada	dijelaskan	diperlihatkan	diucapkannya	ibu	juga	keduanya	lagi
4	adapun	atau	begitu	berjumlah	biasa	datang	dijelaskannya	diperlukan	diungkapkan	ikut	jumlah	keinginan	lagiai
5	agak	ataukah	begitukah	berkali-kali	biasanya	dekat	dikarenakan	diperlukannya	dong	ingat	jumlahnya	kelamaan	lah
6	agaknya	ataupun	begitulah	berkata	bila	demii	dikatakan	dipersoalkan	dua	ingat-ingat	justru	kelihatan	lain
7	agar	awal	begitupun	berkehendak	bilakah	demikian	dikatakannya	dipertanyakan	dulu	ingin	kala	kelihatannya	lainn
8	akan	awalnya	bekerja	berkeinginan	bisa	demikianlah	dikerjakan	dipunyai	empat	inginkah	kalau	kelima	lalu
9	akankah	bagai	belakang	berkenaan	bisakah	dengan	diketahui	diri	enggak	inginkan	kalaupun	keluar	lama
10	akhir	bagaikan	belakangan	berlainan	boleh	depan	diketahui	dirinya	enggaknya	ini	kalaupun	kembali	lama
11	akhiri	bagaimana	belum	berlalu	bolehkah	di	dikira	disampaikan	entah	inikah	kalian	kemudian	lanju
12	akhirnya	bagaimanakah	belumlah	berlangsung	bolehlah	dia	dilakukan	disebut	entahlah	inilah	kami	kemungkinan	lanju
13	aku	bagaimanapun	benar	berlebihan	buat	diakhiri	dilalui	disebutkan	guna	itu	kamilah	kemungkinannya	lebih
14	akulah	bagi	benarkah	bermacam	bukan	diakhirinya	dilihat	disebutkannya	gunakan	itukah	kamu	kenapa	lewa
15	amat	bagian	benarlah	bermacam-macam	bukanlah	dialah	dimaksud	disini	hal	itulah	kamulah	kepada	lima
16	amatlah	bahkan	berada	bermaksud	bukanlah	diantara	dimaksudkan	disinilah	hampir	jadi	kan	kepadanya	luar
17	anda	bahwa	berakhir	bermula	bukannya	diantaranya	dimaksudkannya	ditambahkan	hanya	jadilah	kan	kesempatan	maca
18	andalah	bahwasanya	berakhirilah	bersama	bulan	diberi	dimaksudnya	ditandaskan	hanyalah	jadinya	kapankah	keseluruhan	maka
19	antar	baik	berakhirnya	bersama-sama	bung	diberikan	diminta	ditanya	hari	jangan	kapanpun	keseluruhannya	maka
20	antara	bakal	berapa	bersiap	cara	diberikannya	dimintai	ditanyai	harus	jangan	karena	keterlalu	maki
21	antaranya	bakalan	berapakah	bersiap-siap	caranya	dibuat	dimisalkan	ditanyakan	haruslah	janganlah	karenanya	ketika	mala
22	apa	balik	berapalah	bertanya	cukup	dibuatnya	dimulai	ditegaskan	harusnya	jauh	kasus	khususnya	mala
23	apaan	banyak	berapapun	bertanya-tanya	cukupkah	didapat	dimulailah	ditunjukkan	hendak	jawab	kata	kini	mam
24	apabila	bapak	berarti	berturut	cukuplah	didatangkan	dimulainya	ditunjuk	hendaklah	jawaban	katakan	kinilah	mam
25	apakah	baru	berawal	berturut-turut	cuma	digunakan	dimungkinkan	ditunjuki	hendaknya	jawabnya	katakanlah	kira	manc
26	apalagi	bawah	berbagai	bertutur	dahulu	diibaratkan	dini	ditunjukkan	hingga	jelaa	katanya	kira-kira	manc
27	apatah	beberapa	berdatangan	berajar	dalam	diibaratkannya	dipastikan	ditunjukkannya	ia	jelaskan	ke	kiranya	manc
28	artinya	begini	berdatangan	berupa	dan	diingat	diperbuat	ditunjuknya	ialah	jelaaah	keadaan	kita	masa
29	se	berapa	berapa	berapa	berapa	berapa	berapa	berapa	berapa	berapa	berapa	berapa	berapa
30	stopword	stopword	stopword	stopword	stopword	stopword	stopword	stopword	stopword	stopword	stopword	stopword	stopword