

**ANALISIS SENTIMEN PADA TWEETS DIVISI
HUMAS POLRI DENGAN METODE
*NAÏVE BAYES CLASSIFIER***

Oleh :

CALDIYASTOVAN MOHI

NIM : T3119095

SKRIPSI

**Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjana**



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2023**

PERSETUJUAN SKRIPSI

**ANALISIS SENTIMEN PADA TWEETS DIVISI
HUMAS POLRI DENGAN METODE
*NAÏVE BAYES CLASSIFIER***

Oleh :

CALDIYASTOVAN MOHI

NIM : T3119095

SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjan
Program Studi Teknik Informatika.
Ini telah disetujui oleh Tim Pembimbing
Gorontalo, Januari 2023

Pembimbing I


Haditsah Annur, M.Kom
NIDN : 0908058403

Pembimbing II


Roys Pakaya, M.Kom
NIDN : 0917098401

PENGESAHAN SKRIPSI

ANALISIS SENTIMEN PADA TWEETS DIVISI HUMAS POLRI DENGAN METODE *NAÏVE BAYES* *CLASSIFIER*

Oleh :

CALDIYASTOVAN MOHI

NIM : T3119095

Diperiksa oleh Panitia Ujian Strata Satu (S1)

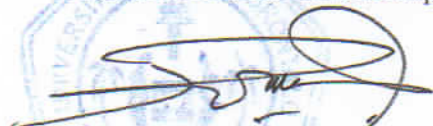
Universitas Ichsan Gorontalo

1. Ketua Penguji
Rezqiwati Ishak, M.Kom
2. Anggota
Apriyanto Alhamad, M.Kom
3. Anggota
Sulistiawaty Ahmad, S.Si, M.Kom
4. Anggota
Haditsah Annur, M.Kom
5. Anggota
Roys Pakaya, M.Kom



Mengetahui

Dekan Fakultas Ilmu Komputer



Irvan Abraham Salihi, M.Kom

NIDN : 0928028101

Ketua Program Studi



Sudirman S. Panna, M.Kom

NIDN : 0924038205

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa :

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, Mei 2023
Yang membuat pernyataan,

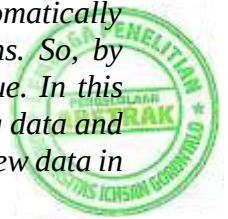


Caldiyastovan Mohi

ABSTRACT

CALDIYASTOVAN MOHI. T3119095. THE SENTIMENT ANALYSIS ON THE TWEETS OF THE PUBLIC RELATIONS DIVISION OF THE INDONESIAN POLICE USING THE NAIVE BAYES CLASSIFIER METHOD

News is a story or information about an event that is hotly discussed. It can be utilized by online media to provide updated news to the public. The use of social media among the public in disseminating information is fast so that distance and time are not an obstacle for social media users to be able to receive and access information developments widely without constraints. Twitter is a social media platform that is widely used by individuals, governments, and organizations, including the police, to publish news about related agencies. By applying one of the functions of text mining, namely the Naïve Bayes Classifier to analyze public sentiment towards tweets from the Indonesian Police Public Relations Division's Twitter account, manually analyzing sentiment on tweets is no longer effective, so a Naïve Bayes Classifier method is needed to automatically analyze sentiment on tweets into positive and negative opinions. So, by using this algorithm, this study gets a fairly high accuracy value. In this study, a high accuracy value of 76% when splitting 10% of testing data and 90% of training data. But when preprocessing and implementing new data in the streamlit framework, it takes up to 1 minute to process the data.



Keywords: Naïve Bayes, Police Public Relations, Twitter, sentiment analysis

ABSTRAK

CALDIYASTOVAN MOHL. T3119095. ANALISIS SENTIMEN PADA TWEETS DIVISI HUMAS POLRI DENGAN METODE NAIVE BAYES CLASSIFIER

Berita merupakan cerita atau keterangan tentang suatu kejadian atau peristiwa yang hangat dibicarakan. Hal ini dapat dimanfaatkan media online untuk memberikan berita terupdate kepada masyarakat. Penggunaan media sosial dikalangan masyarakat dalam menyebarkan informasi dengan cepat sehingga jarak dan waktu tidak menjadi halangan bagi pengguna media sosial untuk dapat menerima dan mengakses perkembangan informasi secara luas tanpa terkendala. Twiter Merupakan salah satu platform media sosial yang banyak digunakan oleh individu, pemerintah dan organisasi, termasuk kepolisian, untuk mempublikasikan berita tentang Instansi terkait. Dengan menerapkan salah satu fungsi dari *Text mining* yaitu *Naïve Bayes Classifier* untuk menganalisis sentimen masyarakat terhadap *tweet* dari akun *twitter* Divisi Humas Polri. menganalisis sentimen pada *tweet* secara manual tidak lagi efektif, maka diperlukan metode *Naïve Bayes Classifier* yang dapat secara otomatis menganalisis sentimen pada *tweet* menjadi opini positif dan negatif. Maka dengan menggunakan algoritma ini penulis mendapatkan nilai akurasi yang cukup tinggi. Pada penelitian ini penulis mendapatkan nilai akurasi yang tinggi sebesar 76% pada saat melakukan *splitting* sebesar 10% data *testing* dan 90% data *training*. Tetapi pada saat melakukan *preprocessing* dan implemetasi data baru pada *framework streamlit* membutuhkan waktu hingga 1 menit untuk dapat memproses data.

Kata kunci : *Naïve Bayes*, Humas Polri, twitter, analisis sentimen

KATA PENGANTAR

Alhamdulillah, penulis dapat menyelesaikan skripsi ini dengan judul: ***“ANALISIS SENTIMEN PADA TWEETS DIVISI HUMAS POLRI DENGAN METODE NAÏVE BAYES CLASSIFIER”***. Skripsi ini disusun sebagai salah satu syarat untuk mengikuti Ujian Akhir guna memperoleh gelar Sarjana Komputer pada Program Studi S1 di Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo.

Penulis sepenuhnya menyadari bahwa skripsi ini tidak mungkin terwujud tanpa ada bantuan dan dorongan dari beberapa pihak, baik bantuan moril maupun materil. Maka dari itu, dengan segala keikhlasan dan kerendahan hati penulis, penulis mengucapkan terima kasih sebesar-besarnya dan penghargaan setinggi-tingginya kepada:

1. Ibu Dr. Hj Juriko Abdussamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan Abraham Salihi S.Kom.,M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Gorontalo;
4. Bapak Sudirman Melangi S.Kom.,M.Kom, selaku WakilDekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Gorontalo;
5. Ibu Irma Surya Kumala Idris, S.Kom.,M.Kom, selaku Wakil Dekan II Bidang Administrasi Umumdan Keuangan Fakultas Ilmu Komputer Universitas Gorontalo;
6. Bapak Sudirman S. Panna, S.Kom.,M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Gorontalo;
7. Ibu Haditsah Annur, S.Kom.,M.Kom, selaku Pembimbing I; Utama Yang telah banyak membantu peneliti dalam meyelesaikan penulisan penyusunan penelitian ini.
8. Bapak Roys Pakaya, S.Kom.,M.Kom, selaku Pembimbing II; Pendamping yang telah banyak membantu penyusunan penelitian ini dari awal hingga akhir.

9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah berusaha mendidik dan mengajarkan berbagai disiplin ilmu kepada kami;
10. Terkhusus Kedua Orang Tua saya tercinta, atas segala kasih sayang, jerih payah, do'a restu serta telah mendukung penulis mencapai cita-cita untuk meraih gelar sarjana
11. Keluarga dan Saudara saya yang tak henti-hentinya memberikan motivasi dan membantu kepada penulis;
12. Rekan-rekan seperjuangan Informatic Enginnering 2019 dan senior-senior saya yang telah banyak membantu dan memberikan dukungan moril kepada penulis;
13. Kepada semua pihak yang ikut membantu dalam penyelesaian skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah SWT melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis sepenuhnya menyadari bahwa apa yang telah dicapai saat ini masih jauh dari kesempurnaan dan kekurangan. Oleh karena itu penulis sangat mengharapkan adanya kritik dan saran yang konstruktif. Penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi semua orang, Aamiin.

Gorontalo, Mei 2023

Caldiyastovan Mohi

DAFTAR ISI

PERSETUJUAN SKRIPSI.....	ii
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN SKRIPSI.....	iv
ABSTRACT	v
ABSTRAK	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
BAB II LANDASAN TEORI	5
2.1 Tinjauan Studi.....	5
2.2 Tinjauan Pustaka.....	7
2.2.1 <i>Twitter</i>	7
2.2.2 Twitter API.....	10
2.2.3 <i>Crawling</i> Data.....	10
2.2.4 <i>Text mining</i>	11
2.2.5 Analisis Sentimen.....	13
2.2.6 <i>Naïve Bayes Classifier</i>	13
2.2.7 Pembobotan Kata	14
2.2.8 <i>Python</i>	15
2.2.9 <i>Confusion Matrix</i>	15
2.3 Kerangka Pikir.....	17
BAB III.....	18
METODE PENELITIAN	18
3.1 Jenis, Metode, Subjek, Waktu, dan Tempat penelitian	18

3.2	Pengumpulan Data.....	18
3.3	Pemodelan/Abstraksi.....	19
3.3.1	Pengembangan Model.....	19
3.3.2	Pra Pengolahan Data	19
3.3.3	Validasi.....	20
3.3.4	Evaluasi Model.....	20
3.3	Pemodelan	20
3.4.1	<i>Preprocessing</i>	22
3.4.2	Pembobotan TF-IDF	22
3.4.3	Klasifikasi <i>Naïve Bayes</i>	23
3.4.4	Hasil Klasifikasi	23
3.4.5	Evaluasi	23
BAB IV		24
HASIL PENELITIAN		24
4.1	Hasil Pengumpulan Data	24
4.2	Hasil Pemodelan.....	27
4.2.1	Hasil Preprocessing.....	28
4.2.2	Hasil Pembobotan TF IDF	47
4.2.3	Klasifikasi Naive Bayes	60
4.2.4	Confusion matrix.....	85
BAB V		93
PEMBAHASAN		93
5.1	Pembahasan Model.....	93
5.1.1	Preprocessing	93
5.1.2	Pembobotan.....	94
5.1.3	Klasifikasi	94
5.1.4	Evaluasi.....	94
5.1.5	Implementasi.....	95
BAB VI		99
PENUTUP		99
6.1	KESIMPULAN	99
6.2	SARAN.....	99
DAFTAR PUSTAKA		101
LAMPIRAN		Error! Bookmark not defined.

DAFTAR TABEL

Tabel 2. 1 : Penelitian Terdahulu	5
Tabel 2. 2 : Case Folding.....	11
Tabel 2. 3 : Tokenization	12
Tabel 2. 4 : Stopword	12
Tabel 2. 5 : Stemming.....	13
Tabel 2. 6 : Confusion Matrix	16
Tabel 4. 1 : Hasil pengumpulan data	24
Tabel 4. 2 : Dataset yang dilabeli	26
Tabel 4. 3 : case folding.....	30
Tabel 4. 4 : Tokenizing.....	32
Tabel 4. 5 : Stopword Removal	37
Tabel 4. 6 : Steamming.....	43
Tabel 4. 7 : Hitung Term Frequency.....	47
Tabel 4. 8 : Normalisasi TF	51
Tabel 4. 9 : Hitung Inverse Document Frequency.....	54
Tabel 4. 10 : Hitung TF-IDF.....	57
Tabel 4. 11 : Hitung Klasifikasi.....	61
Tabel 4. 12 : Data Testing.....	68
Tabel 4. 13 : Hitung Probabilitas.....	69
Tabel 4. 14 : Hitung mean dari atribut.....	69
Tabel 4. 15 : Hitung Standar Deviasi	74
Tabel 4. 16 : Hitung Distribusi Normal	79
Tabel 4. 17 : Hitung Probabilitas dan Hasil Klasifikasi	85
Tabel 5. 1 : Tweet Baru	95

DAFTAR GAMBAR

Gambar 2. 1 : Tweet.....	7
Gambar 2. 2 :Mention.....	8
Gambar 2. 3 : Like	8
Gambar 2. 4 : Direct message.....	8
Gambar 2. 5 : Replay	8
Gambar 2. 6 : Retweet.....	9
Gambar 2. 7 : Hashtag.....	9
Gambar 2. 8 : Trending topic.....	9
Gambar 2. 9 : Source Code Crawling Data Twitter.....	10
Gambar 2. 10 : Kerangka Pikir.....	18
Gambar 3. 1 : pemodelan yang diusulkan	21
Gambar 3. 2 : Preprocessing.....	22
Gambar 4. 1 : Pemodelan	28
Gambar 4. 2 : Preprocessing.....	29
Gambar 4. 3 : Case Folding.....	29
Gambar 4. 4 : visualisasi case folding	31
Gambar 4. 5 : Tokenizing	32
Gambar 4. 6 : Visualisasi Tokenizing.....	36
Gambar 4. 7 : Stopword Removal	37
Gambar 4. 8 : Visualisasi Stopword Removal.....	42
Gambar 4. 9 : Steaming.....	42
Gambar 4. 10 : Visualisasi Steaming.....	46
Gambar 4. 11 : Pembobotan TF-IDF	47
Gambar 4. 12 : Klasifikasi Naive Bayes	60
Gambar 4. 13 : Code Splitting.....	60
Gambar 4. 14 : Code Klasifikasi	61
Gambar 5. 1 : Preprocessing.....	93
Gambar 5. 2 : Pembobotan	94
Gambar 5. 3 : klasifikasi.....	94
Gambar 5. 4 : Evaluasi	95
Gambar 5. 5 : Visualisasi Implementasi Model.....	96
Gambar 5. 6 : Hasil Implementasi 1	97
Gambar 5. 7 : Hasil Implementasi 2	97
Gambar 5. 8 : Hasil Implementasi 3.....	98

DAFTAR LAMPIRAN

Lampiran 1 : Dataset	-
Lampiran 2 : Code Program	-
Lampiran 3 : Surat Izin Penelitian	-
Lampiran 4 : Surat Rekomendasi Peneleitian	-
Lampiran 5 : Surat Bebas Pustaka	-
Lampiran 6 : Surat Bebas Plagiasi	-
Lampiran 7 : Hasil Turnitin	-
Lampiran 8 : Daftar Riwayat Hidup	-

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diera digital sekarang, media online telah menjadi sumber informasi atau berita bagi masyarakat. Berita merupakan cerita atau keterangan tentang suatu kejadian atau peristiwa yang hangat dibicarakan. Berdasarkan hasil riset yang dilakukan oleh lembaga global GFK dan Indonesia Digital Association (IDA) dilaksanakan di lima kota besar di Indonesia sepanjang tahun 2015, presentase konsumsi berita melalui televisi sebesar 91 persen, surat kabar 31 persen dan radio sebesar 15 persen dari 1.521 responden, Angka terendah dibandingkan dengan konsumsi berita melalui online mencapai 96 persen. Hal ini dapat dimanfaatkan media online untuk memberikan berita terupdate kepada masyarakat. [1]

Penggunaan media sosial dikalangan masyarakat dalam menyebarkan informasi dengan cepat sehingga jarak dan waktu tidak menjadi halangan bagi pengguna media sosial untuk dapat menerima dan mengakses perkembangan informasi secara luas tanpa terkendala. Masyarakat biasa maupun perusahaan dan instansi pemerintah kini sudah menggunakan media sosial dalam menyampaikan informasi, aspirasi ataupun mendengar keluhan dari pengguna media sosial.

Kepolisian adalah salah satu instansi pemerintahan yang berperan penting dalam menegakan hukum. Dilihat dari fungsi dan kelembagaan kepolisian, sudah menjadi tugas mereka untuk melindungi masyarakat dari segala bentuk tindakan kriminal dan gangguan yang menimbulkan rasa tidak aman, tidak tertib, dan tidak tenang kepada masyarakat.[2] dengan banyaknya kasus yang terjadi di Indonesia, kepolisian adalah tempat pengaduan masyarakat yang diharapkan bisa menangani kasus-kasus yang terjadi dan dapat diproses sesuai hukum yang ada di Indonesia.

Divisi Humas Polri merupakan unsur pengawas dan pembantu pimpinan dibidang humas di tingkat mabes Polri. Divisi humas polri memiliki akun media sosial salah satunya di *twitter*. Pengguna *twitter* dapat mengetahui informasi terkini terkait kegiatan polri melalui akun Divisi Humas Polri. Divisi Humas Polri

memiliki akun *twitter* yang dimana setiap orang yang memiliki keprihatinan atau saran dapat mempostingnya.

Twitter Merupakan salah satu platform media sosial yang banyak digunakan oleh individu, pemerintah, dan organisasi, termasuk kepolisian, untuk mempublikasikan berita tentang Instansi terkait. Dengan cara ini seseorang dapat mengikuti perkembangan terkini dan menyuarakan pendapat mereka. Sehingga sentimen publik terhadap suatu topik dapat dipelajari dan menjadi sumber informasi dari media sosial.[3]

Analisis sentimen merupakan analisa yang dilakukan pada suatu peristiwa dari pendapat yang berdasarkan pada sikap seseorang tentang suatu objek. Analisis sentimen biasanya melakukan pengumpulan opini untuk mengetahui opini masyarakat dalam postingan di sosial media salah satunya adalah *twitter*. [4] Mengetahui opini publik terhadap suatu objek untuk keperluan tugas akhir maka membutuhkan Analisis sentimen. Ada dua kategori opini publik yaitu negatif dan positif tergantung pada pandangan publik tentang suatu objek tersebut. Sehingga penulis membutuhkan suatu analisis terhadap opini-opini publik dengan tujuan penelitian agar bisa dijadikan tolok ukur baik atau tidaknya persepsi masyarakat tentang polri.

Penilaian positif dan negatif diproses berdasarkan persepsi masyarakat tentang bagaimana aparat polri menangani kasus hukum. Enam dari sepuluh orang yang disurvei mengatakan bahwa berurusan dengan hukum dan polisi biasanya memakan waktu lama dan sulit. Selain itu, lebih dari setengah dari mereka yang disurvei menyatakan bahwa menyuap polisi itu mudah. Bahkan secara resmi 71,4 persen responden menyatakan bahwa diskriminasi dalam penegakan hukum masih dilindungi oleh pihak kepolisian. Menolak menindak pelaku kejahatan yang berasal dari pejabat atau penguasa menunjukkan sikap diskriminatif.

Instansi Indikator Politik Indonesia mengeluarkan hasil survei tingkat kepercayaan masyarakat terhadap Polri menurun tajam pada bulan Agustus 2022, jika dibandingkan pada Bulan Mei 2022. Lembaga tersebut melakukan survei pada 11-17 Agustus 2022. Survei sebelumnya pada bulan Mei, kepercayaan publik untuk polri mencapai 66,7 persen. bahkan pada bulan April 2022 tingkat

kepercayaan masyarakat terhadap Polri mencapai 71,6 persen.[5] dengan hasil survey dari Instansi Politik Indonesia, peneliti melihat perlu adanya analisis sentimen masyarakat terhadap suatu isu yang dipublikasikan lewat *twitter* Divisi Humas Polri.

Meskipun demikian dimungkinkan untuk mengelolah sentimen positif dan negatif yang berkaitan dengan topik *tweet* secara manual, sejumlah besar pengguna juga akan menghasilkan opini yang diungkapkan. Namun, butuh banyak waktu dan tenaga untuk melakukan hal tersebut. Karena menganalisis sentimen pada *tweet* secara manual tidak lagi efektif, maka diperlukan metode *Naïve Bayes Classifier* yang dapat secara otomatis menganalisis sentimen pada *tweet* menjadi opini positif dan negatif. Penulis menggunakan klasifikasi teks, salah satu fungsi dari penambangan teks, untuk menguji algoritma *Naïve Bayes Classifier*, yang dapat menganalisis sentiment pada *tweet*.

Penelitian ini menggunakan algoritma *Naïve Bayes Classifier* untuk teknik pembelajaran dari *Text mining* untuk analisis sentimen. Saat menangani banyak jenis data yang berbeda, algoritma *Naïve Bayes Classifier* dipilih karena kecepatan dan akurasi pemrosesannya yang tinggi. Fakta lain bahwa sejumlah penelitian sebelumnya telah menunjukkan bahwa metode *Naïve Bayes Classifier* memberikan tingkat akurasi dan kinerja yang baik dan telah terbukti, sehingga ini merupakan alasan utama penulis mengapa metode ini dipilih sebagai metode yang diusulkan dalam penelitian ini.[6]

Dari latar belakang di atas, penulis menerapkan salah satu fungsi dari *Text mining* yaitu *Naïve Bayes Classifier* untuk menganalisis sentimen masyarakat terhadap *tweet* dari akun *twitter* Divisi Humas Polri. Sehingga kepolisian dapat mempelajari bagaimana perasaan orang tentang topik tertentu. Informasi ini nantinya akan dijadikan bahan evaluasi untuk membantu Polri meningkatkan pelayanan dan mutu.

Oleh karena itu, penulis mencoba penelitian melakukan analisis sentimen menggunakan *Naïve Bayes Classifier* untuk melihat opini masyarakat terhadap *tweet* Divisi Humas Polri pada media sosial *twitter*. Dengan tujuan utama penulis yaitu untuk melihat berapa banyak masyarakat yang beranggapan positif dan

negatif. Dari hasil penelitian tersebut pihak kepolisian dapat mengevaluasi persepsi masyarakat sehingga Polri dapat meningkatkan pelayanan dan mutu terhadap masyarakat. Selain itu, penulis juga ingin mengetahui tingkat keakuratan klasifikasi sistem menggunakan metode *Naïve Bayes Classifier* yang nantinya akan menjadi bahan referensi peneliti selanjutnya. Obyek yang diklasifikasikan berada pada level kalimat. Berdasarkan latar belakang permasalahan dapat dilakukan penelitian untuk menguji coba Metode *Naïve Bayes Classifier*. Untuk Analisis sentiment pada *tweet* dengan mengangkat judul: ***“Analisis Sentimen Pada Tweets Divisi Humas Polri Dengan Metode Naïve Bayes Classifier”***.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan diatas maka identifikasi masalah pada penelitian ini adalah kurang efektifnya menganalisis sentimen pada *tweet* secara manual.

1.3 Rumusan Masalah

Tahapan identifikasi masalah adalah menentukan permasalahan yang ada dan melakukan riset terhadap permasalahan yang melatar belakangi penelitian. Rumusan masalah pada kasus ini adalah Bagaimana melakukan analisis sentimen pada *tweet* terhadap *twitter* divisi humas Polri dengan metode *Naïve Bayes Classifier*?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah melakukan Analisis Sentimen pada *tweet* terhadap *twitter* divisi humas Polri dengan metode *Naïve Bayes Classifier*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah :

1. Bisa memberikan informasi melalui akun twiiter Divisi Humas Polri kepada Institusi Polri tentang tanggapan publik. Sehingga kepolisian dapat meningkatkan pelayanan masyarakat yang lebih baik.
2. Penelitian ini menjadi bahan referensi analisis sentimen yang menggunakan metode *Naïve Bayes Classifier*.

BAB II

LANDASAN TEORI

2.1 Tinjauan Studi

Pada penelitian ini, metode *Naïve Bayes Classifier* dapat melakukan Analisis Sentimen pada *tweet* yang diupload pada akun *twitter* Divisi Humas Polri ke dalam dua kategori yaitu, kategori Positif dan Negatif di akun *twitter* Divisi Humas Polri. Metode yang akan digunakan adalah *Naïve Bayes* Clasiffier.

Penulis melakukan studi pustaka yang telah dilakukan peneliti-peneliti terdahulu untuk membedakan penelitian yang dilakukan dengan penelitian sebelumnya, adapun tinjauan studi pustaka terdahulu dapat dilihat pada tabel berikut:

Tabel 2. 1 : Penelitian Terdahulu

No	PENELITI	OBJEK	METODE	HASIL
1	Yuna Shophia Dewi Febrianti (2017)	<i>Tweets</i> pada akun <i>twitter</i> STMIK Akakom	<i>Naïve Bayes Classifier</i>	Aplikasi mampu mengklasifikasikan dan menganalisis sentimen yang ada pada akun <i>twitter</i> STMIK Akakom
2	Muliono, Dian Putri Artanti, Abdul Syukur Adi Prihandono, De Rosal I, Moses Setiadi. 2018	Analisa sentimen untuk penilaian pelayanan situs belanja online	<i>Naïve Bayes Classifier</i>	Penelitian ini menggunakan metode <i>Naïve Bayes Classifier</i> untuk melakukan analisis sentimen, dengan jumlah 1200 dataset didapatkan rata-rata akurasi yang cukup bagus dengan nilai 93,33%

No	PENELITI	OBJEK	METODE	HASIL
3	Susanti Fransiska, Yolanda (2019)	<i>Twitter</i> pada review film	Algoritma <i>Naïve Bayes Classifier</i>	Analisis sentimen berfokus pada sentimen, emosi, dan polaritas. Emosi sentimen mengkategorikan data berdasarkan emosi seperti kegembiraan, kesedihan, jijik, dan kejutan
4	Damas Mahardi (2020)	Klasifikasi Pandemi Covid-19 di Indonesia Dengan Algoritma <i>Naïve Bayes</i>	Algoritma <i>Naïve Bayes</i>	Berdasarkan hasil penelitian membuktikan bahwa algoritma <i>Naïve Bayes</i> dapat mengklasifikasikan data Covid-19 dengan probabilitas 0,910 dengan menggunakan 700 data <i>training</i> dan 300 data <i>testing</i>
5	irman Tempola , Miftah Muhammad , Amal Khairan (2018)	Penentuan Status Gunung Berapi	K-Nearest Neighbor (K-NN) dan <i>Naïve Bayes Classifier</i> (NBC)	Berdasarkan pengujian dari dua metode machine learning yang telah diterapkan pada system. Dengan demikian ketika diterapkan dengan <i>Naïve Bayes Classifier</i> akurasi sistem dalam melakukan klasifikasi lebih baik dibandingkan dengan k-nn.

2.2 Tinjauan Pustaka

2.2.1 Twitter

Twitter adalah platform media sosial dengan layanan microblogging yang memungkinkan penggunanya untuk menulis sesuatu, yang dikenal sebagai *tweet*. Ini banyak digunakan oleh masyarakat umum.[7] Penyebaran berita resmi dari berbagai akun kantor berita maupun opini publik melalui *twitter*. Jack dorsey membuat *twitter* pada tahun 2006 dengan alamat <http://www.twitter.com>. Untuk menggunakan *twitter*, anda harus membuat akun terlebih dahulu, dan anda bisa mendaftar di alamat tersebut. *Twitter* memfasilitasi penyebaran informasi yang lebih cepat, yang kemudian akan menjadi bahan diskusi di antara pengguna *twitter*. Keunggulan *twitter* antara lain memiliki jumlah pengikut yang tidak terbatas dan tidak ada timbal balik; misalnya, pengguna A dapat mengikuti akun *twitter* pengguna B meskipun pengguna B tidak mengikuti akun pengguna A. Selain sumber informasi, *Twitter* juga banyak digunakan sebagai sumber data untuk penambangan teks. Ini karena pencarian *twitter* juga bisa berdasarkan pesan status tertentu selain nama akun. Selain itu juga, tidak ada batasan jumlah pengikut yang dapat mengikuti *twitter*, yang dibatasi pada jumlah karakter yang terbatas untuk menyampaikan pesan dengan benar. [8] Berikut ini adalah istilah pada *twitter* :

a. *Tweet*

Tweet adalah status yang ditulis di kotak yang sudah disediakan *twitter* dan mencakup informasi, gambar, trek, opini, dan ulasan. *Tweet* juga disebut pesan. Ada batasan karakter 280 untuk pesan atau informasi yang ingin ditulis.

Terlihat gambar dibawah ini :



Gambar 2. 1 : Tweet

b. *Mention*

Digunakan untuk memanggil pengguna twitter lain pada sebuah *tweet*.

Mention dituliskan dengan '@' diikuti dengan nama pengguna lain.

Terlihat gambar dibawah ini :



Gambar 2. 2 :Mention

c. *Like*

Like merupakan salah satu icon yang digunakan untuk menandai *tweet* yang

di sukai. Terlihat gambar berikut ini :



Gambar 2. 3 : Like

d. *Direct message*

Mengirimkan pesan secara langsung ke orang lain tanpa ada yang mengetahuinya.

Terlihat gambar dibawah ini :



Gambar 2. 4 : Direct message

e. *Replay*

Replay atau balasan bisa digunakan untuk membalas *tweet* pengguna lain.

Terlihat pada gambar berikut :



Gambar 2. 5 : Replay

f. *Retweet*

Retweet adalah penanda apabila kita setuju dengan isi dari *tweet* orang lain kemudian dibagikan ke pengguna *tweet* lain.

Terlihat gambar dibawah ini :



Gambar 2. 6 : Retweet

g. *Hashtag*

Hashtag dapat digunakan untuk mengidentifikasi *tweet* tertentu di *twitter*. Hastag “#” digunakan bersamaan dengan nama topik. Hastag juga meningkatkan visibilitas setiap *tweet* individu.

Terlihat gambar dibawah ini :



Gambar 2. 7 : Hashtag

h. *Trending topic*

Pembicara yang sedang trending karena sedang ramai dibicarakan oleh pengguna *twitter*.

Terlihat gambar dibawah ini :



Gambar 2. 8 : Trending topic

2.2.2 Twitter API

Twitter memperkenalkan API yang tangguh, atau antarmuka pemrograman aplikasi, yang memungkinkan banyak pengguna yang mengakses data *twitter*. Data *twitter* dapat digunakan untuk berbagai keperluan, termasuk mencari informasi populer, mendapatkan *mention*, dan mendapatkan *Retweet* dari akun di *twitter* untuk mendongkrak pengembangan aplikasi.[4] API (*Application Programming Interface*) adalah sekumpulan perintah, fungsi, dan protokol yang dapat digunakan untuk membangun perangkat lunak pada sistem operasi tertentu. Ini juga berfungsi sebagai dokumentasi pembuatan perangkat lunak dengan antarmuka, fungsi, kelas, dan struktur. Pemrograman dapat menghubungkan satu perangkat lunak ke perangkat lunak lainnya menggunakan API. Kunci konsumen dan rahasia konsumen diperlukan untuk mengakses API *twitter*, yang dapat diakses pada halaman <http://dev.twitter.com>, kedua kunci tersebut berfungsi untuk memberi tahu *twitter* tentang aplikasi yang dibuat.

Terdapat berbagai jenis sistem API yang dapat digunakan, termasuk sistem operasi, library, dan web.

2.2.3 Crawling Data

Crawling Data pada *twitter* merupakan suatu langkah pengambilan data atau mengunduh data dari server *Twitter* dengan bantuan *Application Programming Interface (API) Twitter* baik berupa *tweet* maupun data user name pengguna *twitter*. [4] Dalam proses pengambilan data atau *Crawling* data dari *twitter* dimana data tersebut untuk keperluan tugas akhir peneliti. *Crawling* data ini bisa dilakukan dengan cara membuat program dengan memasukkan kata kunci yang kita inginkan untuk mencari *tweet*. Misalnya, “@divHumas_Polri” pada saat memasukan kata kunci berupa *mention* user name pada program yang kemudian akan di proses untuk mengambil *tweet* yang *mention* ke akun Divisi Humas Polri. Kumpulan data *tweet* tersebut adalah data yang akan digunakan untuk tugas akhir.

```
for i,tweet in enumerate(sntwitter.TwitterSearchScrapper('@divHumas_Polri since:2022-01-01 until:2022-10-28').get_items()):
    if i > 1000:
        break
    tweets_list.append([tweet.date, tweet.id, tweet.content, tweet.user.username, analyze_sentiment(tweet.content)])
```

Gambar 2. 9 : Source Code Crawling Data Twitter

2.2.4 Text mining

Text mining atau Penambangan teks adalah proses yang menggunakan teknologi penambangan data untuk menemukan pola dalam teks yang dapat digunakan dalam menemukan informasi yang relevan dengan tujuan tertentu. Karena tidak ada struktur yang melekat pada teks, proses penambangan teks menggunakan tahap awal untuk mempersiapkan teks agar dapat diubah menjadi lebih terstruktur. Proses *Text mining* dibagi menjadi tiga tahap utama, yaitu proses awal (*text Preprocessing*), transformasi text (*text transformation*), dan penemuan pola (*pattern discovery*).[9] Penambangan teks dapat digunakan pada pelacakan topik, kategorisasi, dan analisis sentimen.

Didalam penambangan teks dilakukan proses pengubahan data menjadi data yang teratur sesuai dengan kebutuhan, proses tersebut merupakan Teks *Preprocessing*. *Text Preprocessing* merupakan tahap pembersihan, dan mengubah teks menjadi sederhana agar selanjutnya dapat diproses.[8] Berikut ini tahap dalam proses tersebut.

- a. *Case folding* merupakan pengubahan huruf besar menjadi huruf kecil pada semua ulasan yang ada pada dokumen data *training* dan data *testing*. Hanya huruf “a” sampai dengan huruf “z” yang diterima dan karakter selain huruf akan dihilangkan.[8]

Tabel 2. 2 : Case Folding

Sebelum <i>case folding</i>	Hasil <i>case folding</i>
PAK PRESIDEN @jokowi TOLONG USAT TUNTAS KASUS SUPIR DI PAPUA YANG DIBUNUH GEROMBOLAN ORANG BERSIFAT SETAN! USUT JUGA KENAPA APARAT @DivHumas_Polri HANYA MENONTON TANPA MELINDUNGI KORBAN! PADAHAL DISITU APARAT BERSENJATA LENGKAP!	pak presiden jokowi tolong usat tuntas kasus supir di papua yang dibunuh gerombolan orang bersifat setan usut juga kenapa aparat divhumas polri hanya menonton tanpa melindungi korban padahal disitu aparat bersenjata lengkap jangan pandang bulu penindakan hukum takut gt

JANGAN PANDANG BULU PENINDAKAN HUKUM! TAKUT - >	
---	--

- b. *Tokenization* adalah pemisahan kalimat (*string*) yang setiap kalimatnya disusun menjadi potongan tunggal.[8]

Tabel 2. 3 : Tokenization

Sebelum <i>Tokenization</i>	Hasil <i>Tokenization</i>
polisi harus jujur dan santun karna kebanyakan boong dan arogan	Polisi Harus Jujur dan santun karna kebanyakan boong dan arogan

- c. *Stopword* adalah proses penghilang kata yang tidak berkaitan dengan topik dokumen, karena kata tersebut tidak akan mempengaruhi keakuratan dalam klasifikasi sentimen.[8]

Tabel 2. 4 : Stopword

Sebelum Stopword	Hasil Stopword
polisi harus jujur dan santun karena kebanyakan bohong dan arogan	polisi harus jujur dan santun kebanyakan bohong dan arogan

- d. *Stemming* adalah proses untuk mengubah kata yang terdapat dalam suatu dokumen ke kata akarnya (*root word*) yang terdapat dalam sistem IR.[8]

Tabel 2. 5 : Stemming

Sebelum Stemming					Hasil Stemming				
polisi	harus	jujur	dan	Santun	polisi	harus	jujur	dan	Santun
karena	kebanyakan	bohong	dan	arogan	karena	banyak	bohong	dan	arogan

2.2.5 Analisis Sentimen

Studi komputasi tentang penilaian, pendapat, dan perasaan orang tentang suatu entitas atau peristiwa dikenal sebagai analisis sentiment. Analisis sentimen adalah metode untuk memproses data teks secara otomatis untuk mengekstrak data dan memahami opini. Ada dua jenis opini dalam analisis sentiment: positif dan negatif. Instansi terkait dapat mempelajari tentang berbagai reaksi masyarakat terhadap suatu layanan, peristiwa, atau produk menggunakan analisis sentimen. Umpan balik komunitas dan pakar berfungsi sebagai dasar untuk analisis sentimen.[10]

2.2.6 Naïve Bayes Classifier

Naïve Bayes Classifier adalah teknik klasifikasi statistik yang memperhitungkan probabilitas kemungkinan kelas, termasuk probabilitas signifikansi kelas. Saat menggunakan metode ini untuk mengkategorikan data *twitter*, data tersebut akan dikategorikan menurut skor positif dan negatifnya. Pengklasifikasian *Naïve Bayes* adalah metode klasifikasi berdasarkan teorema bayes. Teorema bayes adalah teori statistik yang dapat diterapkan untuk mempelajari hipotesis tertentu. Bayes adalah teknik prediksi berbasis probabilistik berdasarkan teorema bayes dengan asumsi independensi yang kuat [11]. *Naïve Bayes* memiliki ciri utama seperti asumsi dari masing-masing kondisi atau kejadian yang sangat kuat. Metode tersebut digunakan untuk pengklasifikasian sentimen pada Tugas Akhir, karena mencakup sejumlah keunggulan yang berbeda satu sama lain, berakurasi tinggi, cepat, dan sederhana.

Yang perlu dilakukan adalah menghitung nilai probabilitas masing-masing dengan persamaan (1).

$$P(W_j | V_j) = \frac{n_j + 1}{n + [kosakata]} \dots \dots \dots (1)$$

$P(W_j | V_j)$ = Peluang Kategori j Ketikas Munculnya Kata i
 $n_j + 1$ = Frekuensi Munculnya Kata Pada Sebuah Kategori
 n = Jumlah Semua Kata Pada Suatu Kategori Dalam Dokumen
 $[kosakata]$ = Jumlah Seluruh Kata (Distinc) Pada Seluruh Data Latih

Setelah itu, menghitung probabilitas kategori dengan menggunakan persamaan (2).

$$P(V_j) = \frac{[docsj]}{[contoh]} \dots \dots \dots (2)$$

$P(V_j)$ = Probabilitas Dokumen Kategori
 $[docsj]$ = Jumlah Semua Dokumen Pada Sebuah Kategori
 $[contoh]$ = Jumlah Total Data Yang Dilatih

Untuk proses pengklasifikasian, data yang dimasukan belum diketahui kategorinya. Peroses ini dilakukan pencarian kata pada data yang dimasukan sesuai dengan data latih. Untuk klasifikasi data yang dimasukan menggunakan persamaan (3).

$$V_{MAP} = \operatorname{argmax}_{v_j \in} \sum_i P(V_j) \Pi_i P(a_i | V_j) \dots \dots \dots (3)$$

Fase pengujian dan pelatihan implementasi *Naïve Bayes* terdiri dari dua proses, yaitu proses *testing* dan *training*. Proses latihan digunakan untuk membangun model analisis yang dapat digunakan untuk menganalisis dalam jumlah besar sebagai pengklasifikasian data mentah yang baru.

2.2.7 Pembobotan Kata

Pembobotan kata adalah suatu fitur yang dibutuhkan untuk transformasi data sehingga teks digital bisa diubah menjadi model yang efisien agar proses analisis siap untuk dijalankan. Bobot *term* dapat dianalogikan sebagai data biner

(true and false), frekuensi, dan frekuensi invers dokumennya (TF-IDF).[6] TF-IDF adalah bobot perhitungan untuk setiap kata pada dokumen. Metode TF-IDF digunakan untuk mengetahui seberapa sering kata yang muncul didalam dokumen.[12] TF-IDF dapat dilakukan dengan beberapa cara sebagai berikut.

$$IDF = \log \frac{D}{DF} \dots\dots\dots(4)$$

$$TF - IDF = tf * idf \dots\dots\dots(5)$$

Keterangan :

- D : Jumlah Dokumen Dalam Dataset
- DF : Jumlah Dokumen Yang Mengandung Kata Tersebut
- tf : *Term Frequency*
- idf : *Inversi Document Frequency*

TF-IDF adalah metode yang sering digunakan karena mudah untuk dipahami, efisien dan hasil yang akurat.

2.2.8 Python

Python merupakan bahasa pemrograman yang banyak digunakan perusahaan besar maupun developer untuk mengembangkan berbagai macam aplikasi berbasis desktop, web, dan mobile. peneliti menggunakan *Python* yang mempunyai sintaksis kode yang sangat jelas dan juga dilengkapi fungsional yang tinggi sehingga mempermudah peneliti dalam menyelesaikan tugas akhir yang terkait dengan pemrograman.[13]

python adalah bahasa pemrograman yang terbilang mudah dan efisien akan tetapi bahasa perograman ini termasuk dalam high level. Bahasa perograman *python* dirancang dengan memperhatikan kemudahan developer agar bisa dengan lebih efisien dalam mengerjakan pekerjaan.

2.2.9 Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi sistem klasifikasi yang berdasarkan kemampuan parameter prediktif suatu model. parameter prediktif adalah parameter yang mengukur kualitas set uji statistik yang diberikan berdasarkan atribut yang ada dalam sampel.

Akurasi dinyatakan dalam presentase, sehingga aturan dalam akurasi 100% yang artinya seluruh kasus yang terdapat dalam aturan klasifikasi, diklasifikasikan dengan benar pada kelas yang akan diprediksi.[11] pengguna indikator prediksi memerlukan pengukuran jumlah probabilitas yang diprediksi lebih besar atau sama. *Confusion Matrix* dapat dilihat pada tabel berikut

Tabel 2. 6 : Confusion Matrix

Actual Class	Predicted Class		
		Kelas=Positif	Kelas=Negatif
	Kelas = Positif	TP(True Positive)	FN(False Negatif)
	Kelas = Negatif	FP(False Positive)	TN(True Negatif)

$$akurasi = \frac{(TP+TN)}{TP+TN+FP+FN} \times 100 \dots \dots \dots (6)$$

$$precision = \frac{TP}{TP+FP} \times 100 \dots \dots \dots (7)$$

$$recall = \frac{TP}{TP+FN} \times 100 \dots \dots \dots (8)$$

$$f - measure = \frac{(2 \times recall \times presisi)}{(recall + presisi)} \dots \dots \dots (9)$$

Keterangan :

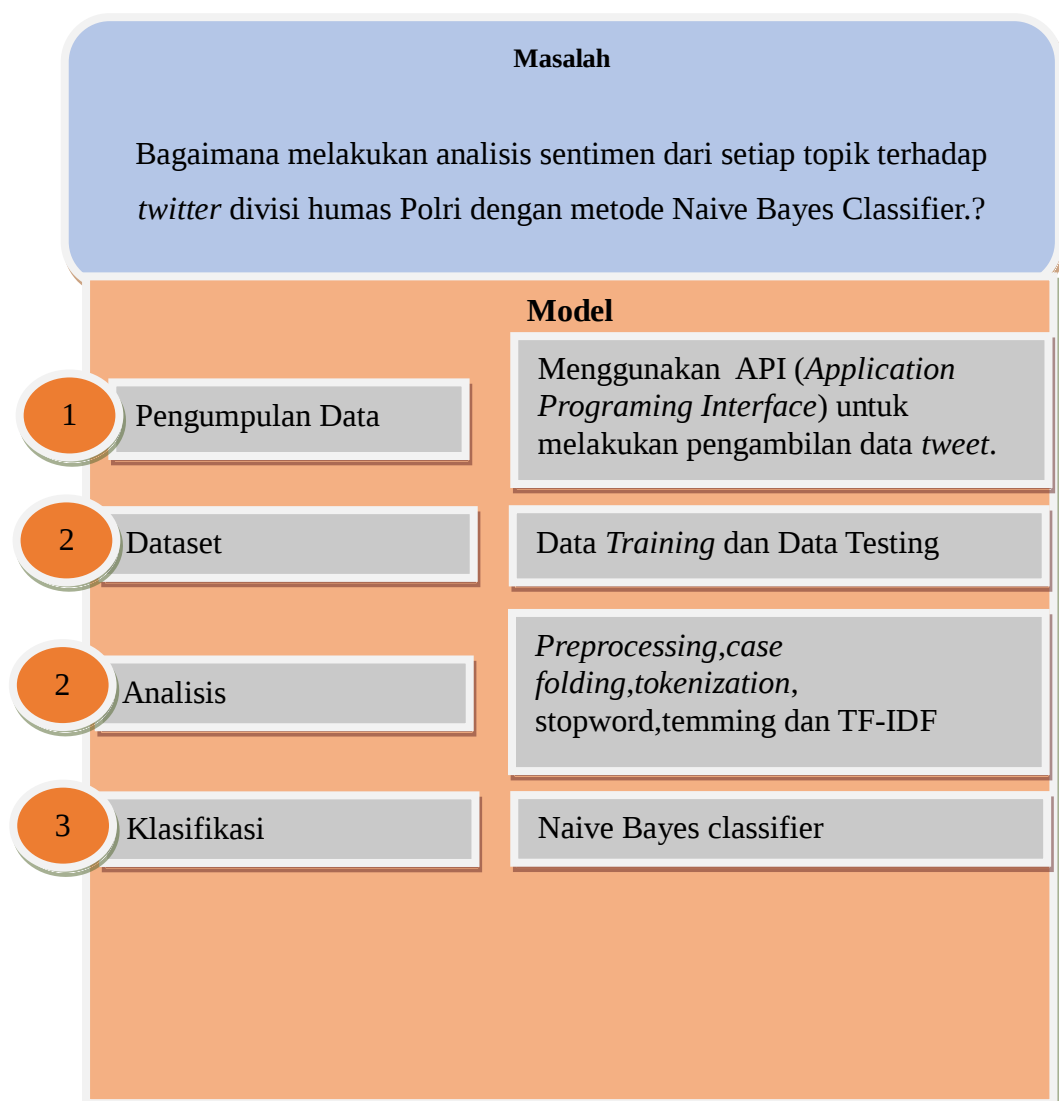
TP(*True Positive*) : jumlah data yang memiliki kategori positif diklasifikasikan positif.

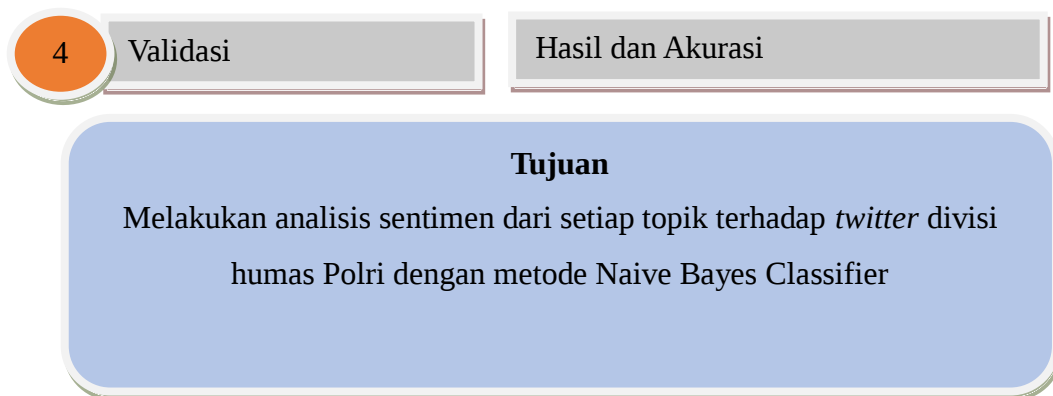
TN(*True Negatif*) : jumlah data yang memiliki kategori negatif diklasifikasikan negatif.

FN(*False Negatif*) : jumlah data yang memiliki kategori positif diklasifikasikan negatif.

TN(*True Negatif*) : jumlah data yang memiliki kategori negatif diklasifikasikan positif.

2.3 Kerangka Pikir





Gambar 2. 10 : Kerangka Pikir

BAB III

METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Waktu, dan Tempat penelitian

Dipandang dari tingkat penerapannya, maka penelitian ini adalah penelitian terapan. Penelitian ini merupakan penelitian kuantitatif Jika dipandang dari suatu informasi yang diolah. Jika berdasarkan dari perilaku terhadap data, maka penelitian ini adalah penelitian konfirmatori. Penelitian ini memanfaatkan studi kasus di akun *twitter* Divisi Humas Polri. Dengan demikian jenis penelitian ini adalah deskriptif. Seubjek penelitian ini Analisis Sentimen Pada *Tweets* Divisi Humas Polri menggunakan Metode *Naïve Bayes Classifier*.

3.2 Pengumpulan Data

Penelitian ini menggunakan data sekunder yang di ambil dari akun *twitter* resmi Divisi Humas Polri. Data yang diambil untuk membuat kamus opini yang akan menjadi data *training*. Kamus data opini dibuat dengan mengumpulkan data opini melalui media data online. Kamus data opini yang dikumpulkan diklasifikasi menjadi kamus opini negatif dan positif. Untuk melakukan klasifikasi peneliti menggunakan bantuan secara online ataupun offline untuk melakukan klasifikasi agar data kamus sesuai dan meminimalkan subjektifitas.

Aplikasi yang dikembangkan menggunakan data internal dari *twitter* sebagai bagian dari proses penentuan kalimat, baik positif maupun negatif. kategori tersebut termasuk sebagai proses pengklasifikasian. Klasifikasi pada suatu dokumen termasuk kelas positif dan kelas negatif. Pada pengumpulan data ini juga menggunakan study literatur dengan data sekunder yang didapatkan dari *twitter*.

3.3 Pemodelan/Abstraksi

Pengembangan Model

Metode Pengklasifikasi *Naïve Bayes Classifier* digunakan dalam prosedur atau langkah utama penelitian ini. Metode klasifikasi dirancang untuk didistribusikan dan didasarkan pada probabilitas sederhana, membuat asumsi bahwa tidak ada kelas teknik klasifikasi termasuk metode ini, yang cepat prosesnya dan mudah digunakan.

Pra Pengolahan Data

Menggunakan library pada *python*, ini adalah langkah pertama dalam memproses data teks menjadi analisis sentimen. Peneliti melakukan tahapan analisa data yang akan diambil dan apa saja yang akan digunakan pada data tersebut. Pengumpulan data dilakukan dengan menggunakan data scraper dari *tweets* pada media sosial *twitter* yang telah tersedia *twitter API* (application programming Interface) yang diambil langsung dari akun media sosial *twitter*. Pengambilan data *tweets* menggunakan tools scrapping atau *Crawling* data dengan penelitian menggunakan kata kunci yaitu Divisi Humas Polri. Pada pengolahan data ada beberapa metode yang digunakan yaitu *Case folding*, mengubah huruf besar menjadi huruf kecil dan menghilangkan tanda baca. *Tokenization*, proses memisahkan sekumpulan karakter pada teks menjadi kata. *Filtering stopword*, menghapus kata atau karakter yang tidak digunakan. *Steamming*, mengubah kata berimbuhan menjadi kata dasar.

Validasi

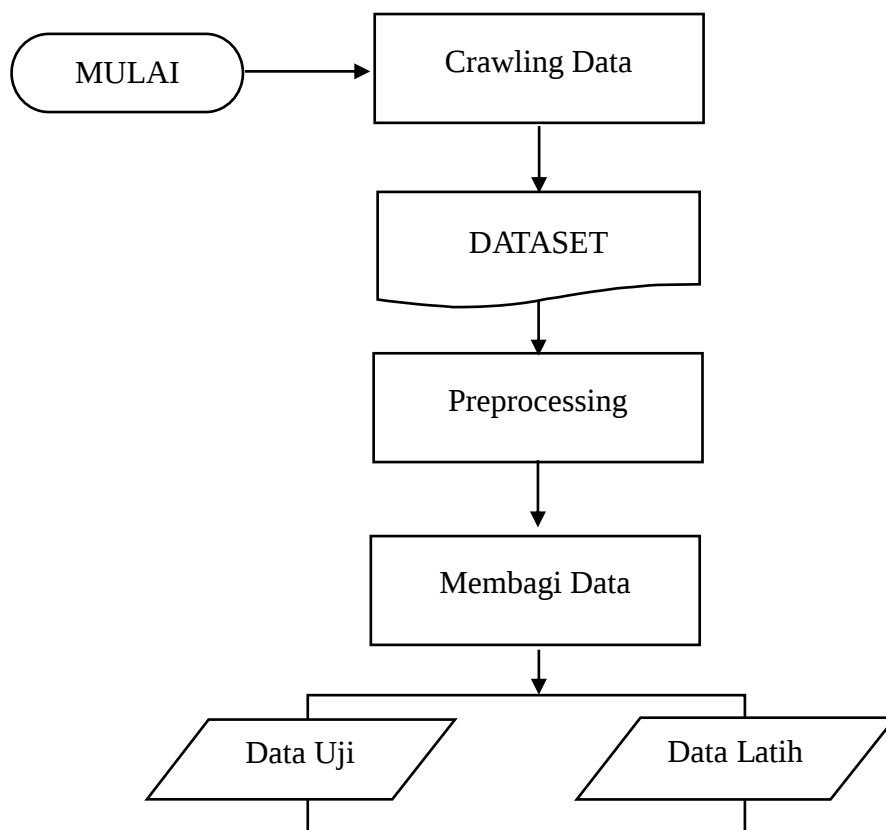
Data latih dan data uji adalah dua komponen validasi. Kinerja model yang merupakan pengklasifikasian *Naïve Bayes Classifier* digunakan untuk menentukan akurasi model.

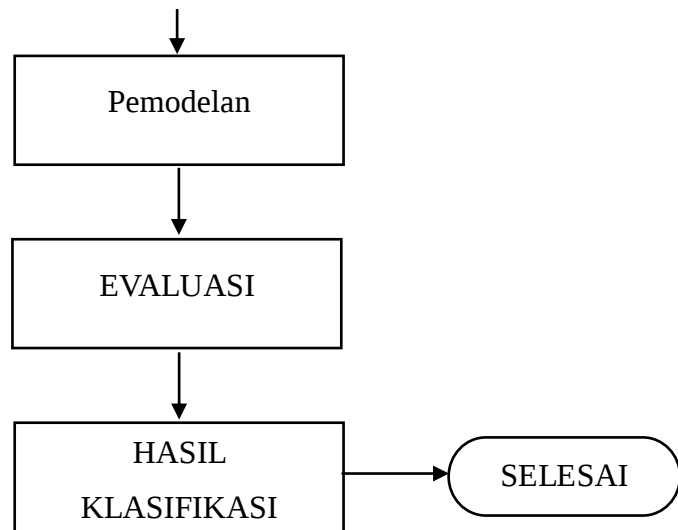
Evaluasi Model

Evaluasi dilakukan dengan menganalisis hasil klasifikasi. Kemudian proses evaluasi menggunakan *Confusion Matrix*.

3.3 Pemodelan

Model yang diusulkan dapat digambarkan menggunakan flowchart berikut ini :

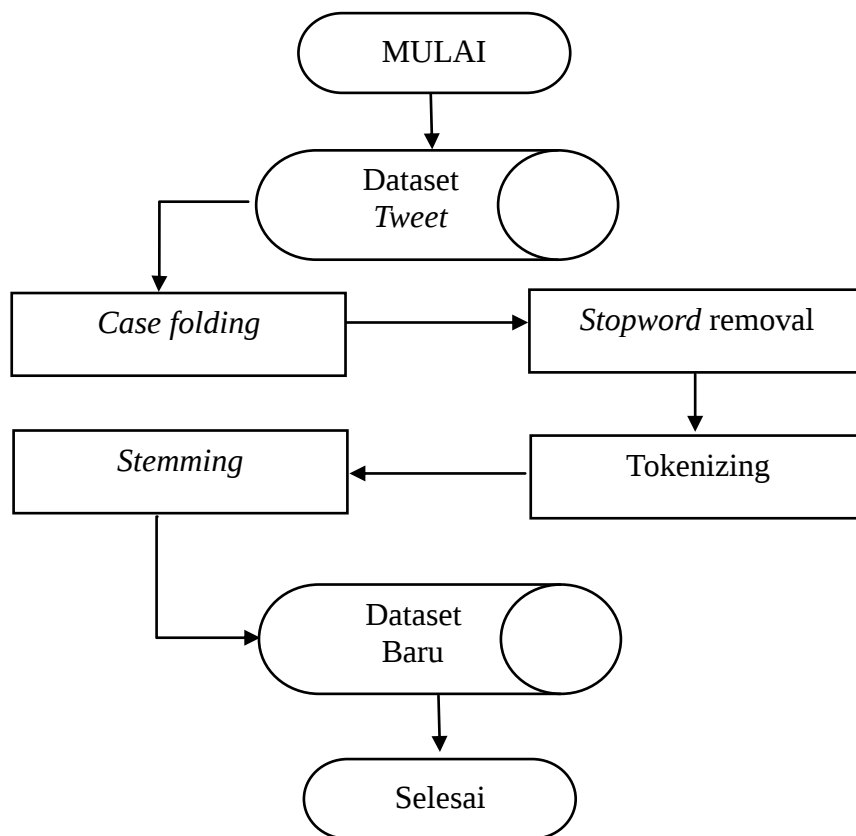




Gambar 3. 1 : pemodelan yang diusulkan

Preprocessing

Tahapan setelah dataset yaitu *Preprocessing* data. *Preprocessing* data adalah tahap pengolahan data mentah menjadi data yang akan diproses pada tahapan selanjutnya. Ada beberapa tahapan *Preprocessing* yang akan dilakukan yaitu merubah huruf besar menjadi huruf kecil dan menghilangkan seluruh tanda baca (*case folding*), memisahkan sekumpulan karakter dalam teks menjadi suatu kata (*tokenization*), menghilangkan karakter-karakter atau kata yang tidak digunakan (*filtering stopwords*), dan merubah kata berimbuhan menjadi kata dasar (*stemming*).



Gambar 3. 2 : Preprocessing

Pembobotan TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) tahapan yang dilakukan untuk perhitungan pembobotan untuk mengetahui seberapa sering suatu *term* kata yang muncul dalam setiap dokumen dengan menggunakan rumus

perhitungan.[10] semua kata akan dilakukan pembobotan sehingga setiap dokumen akan mempunyai bobot masing-masing. Kemudian dilakukan proses *training* untuk dokumen yang siap dan dilakukan proses klasifikasi.

Klasifikasi Naïve Bayes

Pada tahap ini dilakukan klasifikasi sentimen dengan menggunakan metode *Naïve Bayes Classifier*. Menyiapkan data yang telah diberi label sebelumnya untuk digunakan dalam proses klasifikasi (data *training*), merupakan langkah awal dalam proses metode klasifikasi *Naïve Bayes*. Penggunaan library *python* akan memproses penetapan sentimen ke jumlah data yang disiapkan sebelumnya. Peneliti juga akan secara manual memberi label data dengan membagi tanggapan menjadi kategori positif dan negatif. Kemudian keakuratan proses klasifikasi yang dilakukan akan diukur setelah semua data telah dilabeli.

Hasil Klasifikasi

Output data yang dihasilkan merupakan hasil akurasi yang didapatkan dari metode yang digunakan. Penyajian masing-masing sentimen berdasarkan hasil klasifikasi merupakan hasil lain yang ditunjukkan pada penelitian ini. Ada dua kategori sentimen pada penelitian ini yaitu kategori positif dan kategori negatif.

Evaluasi

Evaluasi dilakukan untuk memverifikasi hasil dari klasifikasi dengan melakukan pengukuran keakuratan pada metode klasifikasi. Tahap evaluasi ini menggunakan metode confusion matrix untuk pembandingan hasil dari klasifikasi yang dilakukan pada sistem sesuai dengan hasil yang sebenarnya.

BAB IV

HASIL PENELITIAN

4.1 Hasil Pengumpulan Data

Pada penelitian ini menggunakan dataset *tweet* berbahasa indonesia yang diambil dari akun *twitter* Divisi Humas Polri. Kemudian peneliti melakukan pemrosesan dan pengklasifikasian pada dataset berdasarkan pemodelan yang diusulkan. Jika terjadi outlier atau *tweet* yang tidak berhubungan dengan topik yang akan diteliti dan tidak memenuhi kriteria untuk dijadikan dataset pada proses pengumpulan dataset, maka peneliti akan menyeleksi *tweet* tersebut dari dataset agar tidak mengganggu pada saat proses klasifikasi.

Pengumpulan data pada penelitian ini menggunakan *library* pada *python* yang dilakukan dengan menggunakan data *scraper* dari *tweets* pada media sosial *twitter* yang telah tersedia pada *twitter API (application programming Interface)* dan diambil langsung dari akun media sosial *twitter*. Pengambilan data *tweets* menggunakan *tools scrapping* atau *crawling* data dengan ini peneliti menggunakan kata kunci yaitu Divisi Humas Polri. Dari hasil *crawling* data diperoleh 1001 *tweet*. Kemudian dilakukan seleksi pada *tweet* yang memenuhi kriteria untuk diproses sehingga peneliti menyeleksi data *tweet* 1001 menjadi 122 data *tweet* yang siap untuk diproses. Dari 1001 data yang di seleksi, banyak *tweet* yang mengandung iklan dan menggunakan hastag @divhumaspolri sehingga jumlah data *tweet* menurun menjadi 122 data *tweet*. Selain itu juga, dari data 1001 *tweet* yang diperoleh pada proses *crawling* terdapat banyak *retweet*. Setelah dilakukan seleksi data, penghapusan *tweet* yang tidak berkaitan dengan topik, penghapusan *retweet* dan sentimen yang dianggap netral, maka peneliti memperoleh 122 data *tweet* yang kemudian siap untuk diproses. Berikut adalah tabel hasil pengumpulan data.

Tabel 4. 1 : Hasil pengumpulan data

No	Waktu Tweet	Text	Username	label
1	2022-10-27 23:59:51+00:00	Lestarkan Semangat Persatuan & Kesatuan	The4_9Institute	Positif

No	Waktu Tweet	Text	Username	label
2	2020-02-28 14:06:09+00:00	Terima kasih atas responnya, Ndan poldajateng Kami tunggu perkembangan kasus tersebut! ☐	Nasikin_MA	Positif
3	2022-10-27 23:59:35+00:00	Dibikin ga masuk akal biar susah dan milih buat nembak aja, mayan dapet cuan tuhh☐	yesimsoquiet	Negatif
4	2022-10-27 23:57:20+00:00	Mantap Pak ListyoSigitP	The4_9Institute	Positif
5	2022-10-27 23:56:54+00:00	Tilang saja sesuai aturan ...!! jgn buat2 aturan dewe☐	Ojoklalenya	Positif
6	2022-10-27 23:56:25+00:00	Delik pidana pencemaran nama baik RI1 sudah terbukti nie,. harusnya tetap bisa proses hukum makhluk satu itu.	ApriliaLin	Positif
7	2022-10-27 23:55:48+00:00	Tidak laku klo gini	SSobrah	Negatif
8	2022-10-27 23:55:28+00:00	dan jajaran kepolisian lainnya bisa segera ditangkap pelakunya, sebelum kepercayaan masyarakat hilang, sehingga akan dihakimi massal.	bariekly_	Negatif
9	2022-10-27 23:54:58+00:00	Saya juga termasuk salah satu yang lolos, kemudian saya pernah ditanya apa keluhannya saat melakukan tes lapangan tsb, dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb, namun jawabn yg diberikn sngt tdk memuaskan.	srohimah109	Negatif

No	Waktu Tweet	Text	Username	label
10	2022-10-27 23:54:17+00:00	Setuju. Ayo donk pak tunjukkan kalau Polri tidak berpihak, gerakan para buzzer ini masif dan organized lho. Tindak mereka pak.	akuboekankamu	Positif
.....				
122	2022-10-27 22:02:55+00:00	Ini anak sudah kurang ajar Marwah Presiden jokowi seenaknya dia samakan dengan Tukan Obat pinggir jalan... Tindak...☐	anggorosabdo	Positif

Setelah mendapatkan data dari twitter langkah selanjutnya adalah Peneliti meminta bantuan untuk pelabelan sentimen negatif dan positif pada data tweet di Polda Gorontalo, bagian Humas Polda Gorontalo. Kemudian data yang telah dilabeli siap untuk dilakukan proses selanjutnya. Berikut adalah tabel data yang telah dilabeli.

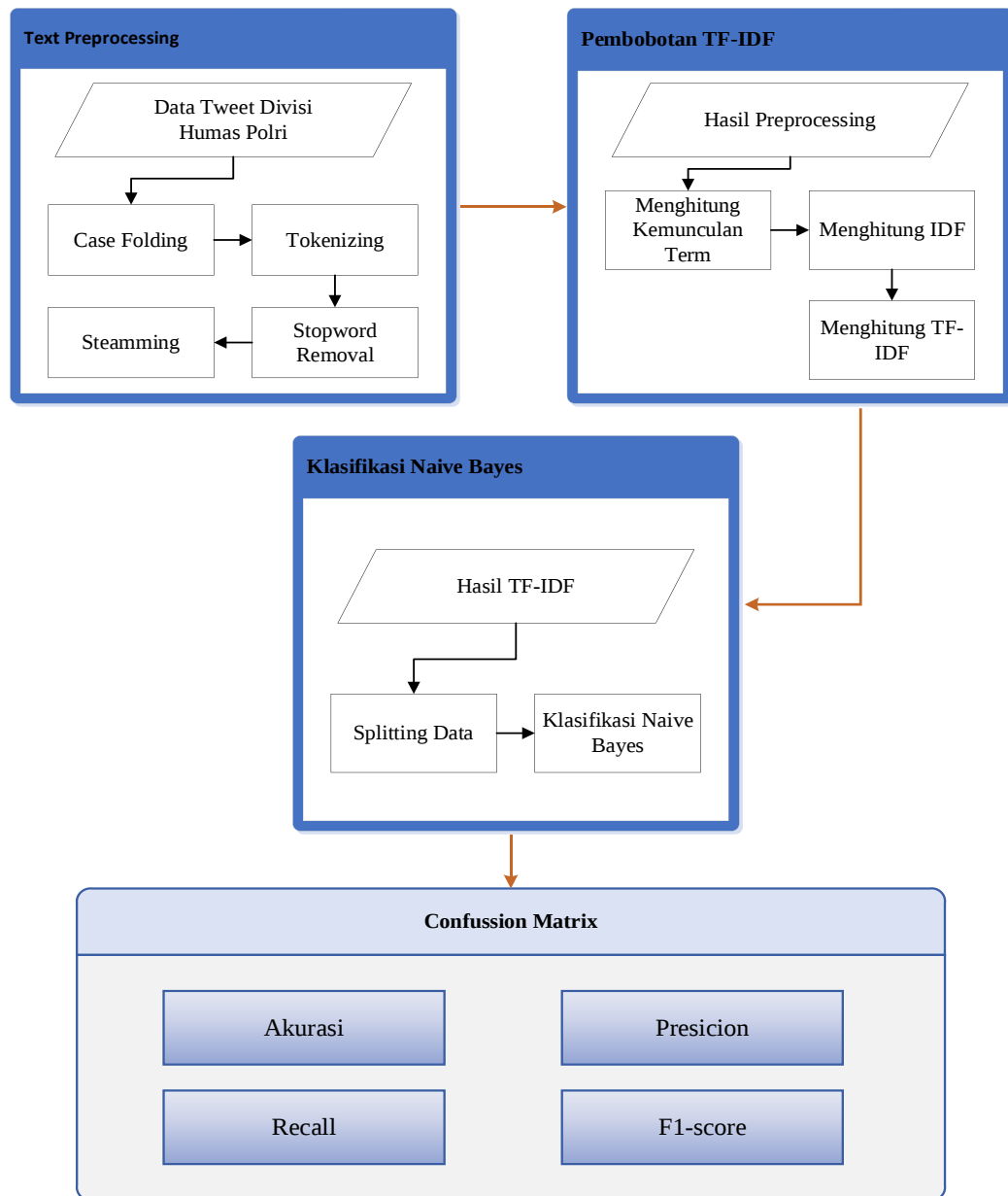
Tabel 4. 2 : Dataset yang dilabeli

No	text	label
1	Lestarkan Semangat Persatuan & Kesatuan	Positif
2	Terima kasih atas responnya, Ndan poldajateng Kami tunggu perkembangan kasus tersebut! ☐	Positif
3	Dibikin ga masuk akal biar susah dan milih buat nembak aja, mayan dapet cuan tuh☐	Negatif
4	Mantap Pak ListyoSigitP	Positif
5	Tilang saja sesuai aturan ...!! jgn buat2 aturan dewe☐	Positif
6	Delik pidana pencemaran nama baik RI1 sudah terbukti nie,. harusnya tetap bisa proses hukum makhluk satu itu.	Positif

No	text	label
7	Tidak laku klo gini	Negatif
8	dan jajaran kepolisian lainnya bisa segera ditangkap pelakunya, sebelum kepercayaan masyarakat hilang, sehingga akan dihakimi massal.	Negatif
9	Saya juga termasuk salah satu yang lolos, kemudian saya pernah ditanya apa keluhannya saat melakukan tes lapangan tsb, dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb, namun jawabn yg diberikn sngt tdk memuaskan.	Negatif
10	Setuju. Ayo donk pak tunjukkan kalau Polri tidak berpihak, gerakan para buzzer ini masif dan organized lho. Tindak mereka pak.	Positif
.....		
122	Ini anak sudah kurang ajar Marwah Presiden jokowi seenaknya dia samakan dengan Tukan Obat pinggir jalan... Tindak...□	Positif

4.2 Hasil Pemodelan

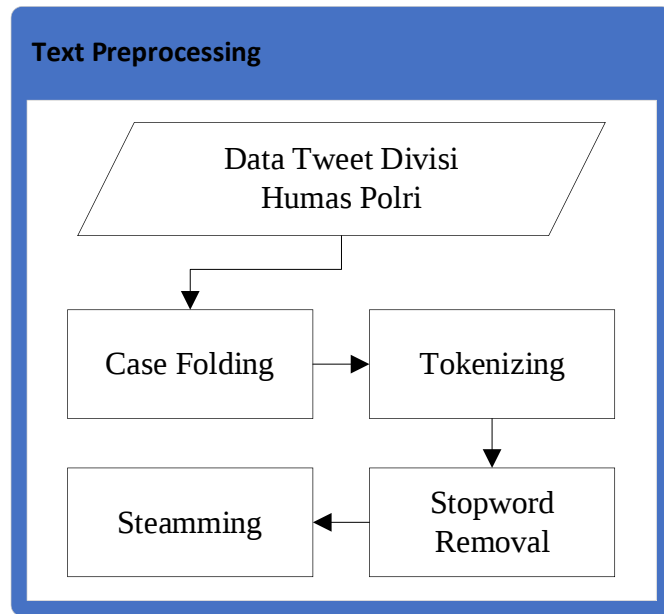
Model yang dibangun pada penelitian ini akan dilakukan secara terstruktur mulai dari *preprocessing*, pembobotan TF.IDF, klasifikasi *Naïve Bayes*, dan kemudian di evaluasi menggunakan confusion matrix. Penulis Membuat *flowchart* untuk model yang dibangun:



Gambar 4. 1 : Pemodelan

4.2.1 Hasil Preprocessing

Preprocessing dilakukan setelah dataset siap untuk diolah. Pada tahapan *preprocessing* terdapat beberapa proses yang akan dilakukan sehingga data menjadi terstruktur dengan tahapan *Case folding*, *Tokenizing*, *Stopword removal*, dan *Stemming*. Peneliti membuat *flowchart* mengenai *preprocessing*.



Gambar 4. 2 : Preprocessing

Pada tahapan *preprocessing* peneliti menggunakan framework streamlit berbasis *python*, sehingga hasil yang didapatkan dari proses *Case folding* adalah sebagai berikut :

1) *Case folding*

Case folding merupakan pengubahan huruf besar menjadi huruf kecil pada semua *tweet* yang ada pada dokumen data trening dan data *testing*. Berikut ini adalah *sourcecode python* pada *framework streamlit* untuk proses *case folding*.

```

103     # Case Folding
104     for k in dataset.keys():
105         hasil_case_folding = dataset[k].lower()
106         dataset[k] = hasil_case_folding
107     st.write("Case Folding:", dataset)

```

Gambar 4. 3 : Case Folding

Selanjutnya penulis membuat tabel untuk *case folding* dari 1-10 dokumen. Berikut adalah tabel hasil dari *case folding* dari dokumen 1-10

Tabel 4. 3 : case folding

No	text	Case Folding
1	Lestarikan Semangat Persatuan & Kesatuan	lestarikan semangat persatuan kesatuan
2	Terima kasih atas responnya, Ndan poldajateng Kami tunggu perkembangan kasus tersebut! □	terima kasih atas responnya, ndan poldajateng kami tunggu perkembangan kasus tersebut
3	Dibikin ga masuk akal biar susah dan milih buat nembak aja, mayan dapet cuan tuhh□	dibikin ga masuk akal biar susah dan milih buat nembak aja mayan dapet cuan tuhh
4	Mantap Pak ListyoSigitP	mantap pak listyosigitp
5	Tilang saja sesuai aturan ...!! jgn buat2 aturan dewe□	tilang saja sesuai aturan jgn buat2 aturan dewe
6	Delik pidana pencemaran nama baik RI1 sudah kebukti nie,. harusnya tetap bisa proses hukum makhluk satu itu.	delik pidana pencemaran nama baik ri1 sudah kebukti nie harusnya tetap bisa proses hukum makhluk satu itu
7	Tidak laku klo gini	tidak laku klo gini
8	dan jajaran kepolisian lainnya bisa segera ditangkap pelakunya, sebelum kepercayaan masyarakat hilang, sehingga akan dihakimi massal.	dan jajaran kepolisian lainnya bisa segera ditangkap pelakunya sebelum kepercayaan masyarakat hilang sehingga akan dihakimi massal
9	Saya juga termasuk salah satu yang lolos, kemudian saya pernah ditanya apa keluhannya saat melakukan tes lapangan tsb, dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb, namun jawabn yg diberikn sngt tdk memuaskan.	saya juga termasuk salah satu yang lolos kemudian saya pernah ditanya apa keluhannya saat melakukan tes lapangan tsb dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb namun jawabn yg diberikn sngt tdk memuaskan


```

109 # Tokenizing
110 tokenizer = RegexpTokenizer(r"\w+")
111 for k in dataset.keys():
112     tokens = tokenizer.tokenize(dataset[k])
113     dataset[k] = tokens
114 st.write("Tokenisasi", dataset)

```

Gambar 4. 5 : Tokenizing

selanjutnya penulis membuat tabel untuk *Tokenizing* dari 1-10 dokumen. Berikut adalah tabel hasil dari *Tokenizing* dari dokumen 1-10.

Tabel 4. 4 : Tokenizing

No	Case Folding	Tokenisasi
1	lestarikan semangat persatuan kesatuan	lestarikan
		semangat
		persatuan
		kesatuan
2	terima kasih atas responnya ndan poldajateng kami tunggu perkembangan kasus tersebut	terima
		kasih
		atas
		responnya
		ndan
		poldajateng
		kami
		tunggu
		perkembangan
		kasus
		tersebut
3	dibikin ga masuk akal biar susah dan milih buat nembak aja mayan dapet cuan tuh	dibikin
		ga
		masuk
		akal
		biar
		susah
		dan
		milih
		buat
		nembak
		aja

No	Case Folding	Tokenisasi
		mayan
		dapet
		cuan
		tuhh
4	mantap pak listyosigitp	mantap
		pak
		lisetyosigitp
5	tilang saja sesuai aturan jgn buat2 aturan dewe	tilang
		saja
		sesuai
		aturan
		jgn
		Buat2
		aturan
		dewe
6	delik pidana pencemaran nama baik ri1 sudah bukti nie harusnya tetap bisa proses hukum makhluk satu itu	delik
		pidana
		pencemaran
		nama
		baik
		ri1
		sudah
		bukti
		nie
		harusnya
		tetap
		bisa
		proses
		hukum
		makhluk
		satu
		itu
7	tidak laku klo gini	tidak
		laku

No	Case Folding	Tokenisasi
		klo
		gini
8	dan jajaran kepolisian lainnya bisa segera ditangkap pelakunya sebelum epercayaan masyarakat hilang sehingga akan dihakimi massal	dan
		jajaran
		kepolisian
		lainnya
		bisa
		segera
		ditangkap
		pelakunya
		sebelum
		kepercayaan
		masyarakat
		hilang
		sehingga
		akan
9	saya juga termasuk salah satu yang lolos kemudian saya pernah ditanya apa keluhannya saat melakukan tes lapangan tsb dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb namun jawabn yg diberikn sngt tdk memuaskan	dihakimi
		massal
		saya
		juga
		termasuk
		salah
		satu
		yang
		lolos
		kemudian
		saya
		pernah
		ditanya
		apa
		keluhannya
		saat
		melakukan
		tes
		lapangan
		tsb

No	Case Folding	Tokenisasi
		dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb namun jawaban yg diberikn sngt tdk memuaskan
10	setuju ayo donk pak tunjukkan kalau polri tidak berpihak gerakan para buzzer ini masif dan organized lho tindak mereka pak	setuju ayo donk pak tunjukkan kalau polri tidak berpihak gerakan para buzzer ini masif dan organized lho tindak

No	Case Folding	Tokenisasi
		mereka
		pak
....		
122	foto satu peristiwa baik, namun lebih baik tunjukkan dgn peraturan yg melayani dan memudahkan seluruh rakyat indonesia	foto
		satu
		peristiwa
		baik
		namun
		lebih
		baik
		tunjukkan
		dgn
		peraturan
		yg
		melayani
		dan
		memudahkan
		seluruh
		rakyat
		indonesia

Pada framework stremlit berbasis *python* didapatkan hasil *Tokenizing* pada gambar berikut.

```

Tokenisasi

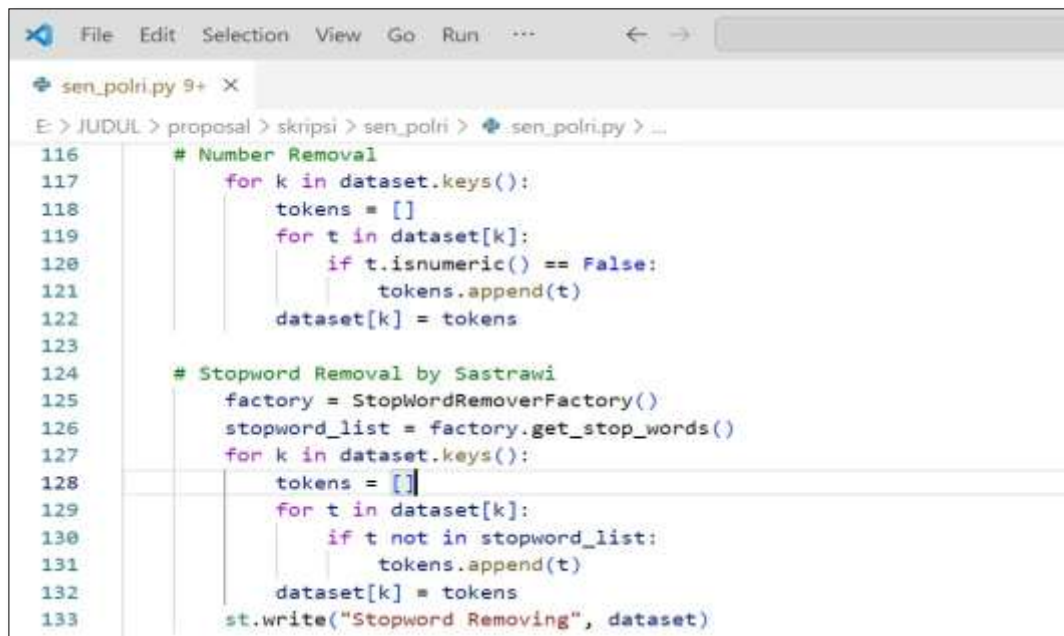
"1":[
0:"lestarikan"
1:"semangat"
2:"persatuan"
3:"kesatuan"]
"2":[
0:"terima"
1:"kasih"
2:"atas"
3:"responnya"
4:"ndan"
5:"poldajateng"

```

Gambar 4. 6 : Visualisasi Tokenizing

3) Stopword removal

Stopword adalah proses penghilang kata yang tidak berkaitan dengan topik dokumen. Berikut ini adalah sourcecode *python* pada framework *streamlit* untuk proses *Stopword removal*.



```
116 # Number Removal
117 for k in dataset.keys():
118     tokens = []
119     for t in dataset[k]:
120         if t.isnumeric() == False:
121             tokens.append(t)
122     dataset[k] = tokens
123
124 # Stopword Removal by Sastrawi
125 factory = StopWordRemoverFactory()
126 stopword_list = factory.get_stop_words()
127 for k in dataset.keys():
128     tokens = []
129     for t in dataset[k]:
130         if t not in stopword_list:
131             tokens.append(t)
132     dataset[k] = tokens
133 st.write("Stopword Removing", dataset)
```

Gambar 4. 7 : Stopword Removal

selanjutnya penulis membuat tabel untuk *Stopword removal* dari 1-10 dokumen. Berikut adalah tabel hasil dari *Stopword removal* dari dokumen 1-10.

Tabel 4. 5 : Stopword Removal

No	Tokenisasi	Stopword Removing
1	lestarikan	lestarikan
	semangat	semangat
	persatuan	persatuan
	kesatuan	kesatuan
2	terima	terima
	kasih	kasih
	atas	atas
	responnya	responnya
	ndan	ndan

No	Tokenisasi	Stopword Removing
	poldajateng	poldajateng
	kami	
	tunggu	tunggu
	perkembangan	perkembangan
	kasus	kasus
	tersebut	tersebut
3	dibikin	dibikin
	ga	ga
	masuk	masuk
	akal	akal
	biar	biar
	susah	susah
	dan	
	milih	milih
	buat	buat
	nembak	nembak
	aja	aja
	mayan	mayan
	dapet	dapet
	cuan	cuan
	tuhh	tuhh
4	mantap	mantap
	pak	pak
	lisetyosigitp	lisetyosigitp
5	tilang	tilang
	saja	
	sesuai	sesuai
	aturan	aturan
	jgn	jgn
	buat2	buat
	aturan	aturan
	dewe	dewe
6	delik	delik
	pidana	pidana
	pencemaran	pencemaran
	nama	nama

No	Tokenisasi	Stopword Removing
	baik	baik
	ri1	ri
	sudah	
	kebukti	kebukti
	nie	nie
	harusnya	harusnya
	tetap	tetap
	bisa	
	proses	proses
	hukum	hukum
	makhluk	makhluk
	satu	satu
	itu	
7	tidak	
	laku	laku
	klo	klo
	gini	gini
8	dan	
	jajaran	jajaran
	kepolisian	kepolisian
	lainnya	lainnya
	bisa	
	segera	segera
	ditangkap	ditangkap
	pelakunya	pelakunya
	sebelum	
	kepercayaan	kepercayaan
	masyarakat	masyarakat
	hilang	hilang
	sehingga	
	akan	
	dihakimi	dihakimi
	massal	massal
9	saya	
	juga	
	termasuk	termasuk
	salah	salah
	satu	satu

No	Tokenisasi	Stopword Removing
	yang	
	lolos	lolos
	kemudian	kemudian
	saya	
	pernah	pernah
	ditanya	ditanya
	apa	apa
	keluhannya	keluhannya
	saat	
	melakukan	melakukan
	tes	tes
	lapangan	lapangan
	tsb	tsb
	dan	
	kemudian	kemudian
	saya	
	menyatakan	menyatakan
	hal	
	sama	sama
	seperti	
	ini	
	kepada	
	polisi	polisi
	tsb	tsb
	namun	
	jawaban	jawaban
	yg	yg
	diberikn	diberikn
	sngt	sngt
	tdk	tdk
	memuaskan	memuaskan
10	setuju	setuju
	ayo	ayo
	donk	donk
	pak	pak
	tunjukkan	tunjukkan
	kalau	kalau
	polri	polri

No	Tokenisasi	Stopword Removing
	tidak	
	berpihak	berpihak
	gerakan	gerakan
	para	
	buzzer	buzzer
	ini	
	masif	masif
	dan	
	organized	organized
	lho	lho
	tindak	tindak
	mereka	
	pak	pak
....		
122	foto	foto
	satu	satu
	peristiwa	peristiwa
	baik	baik
	namun	
	lebih	lebih
	baik	baik
	tunjukkan	tunjukkan
	dgn	dgn
	peraturan	peraturan
	yg	yg
	melayani	melayani
	dan	
	memudahkan	memudahkan
	seluruh	seluruh
	rakyat	rakyat
	indonesia	indonesia

Pada framework stremlit berbasis *python* didapatkan hasil *Stopword removal* pada gambar berikut.

```
Stopword Removing :
"1": [0: "lestarikan"
1: "semangat"
2: "persatuan"
3: "kesatuan"]
"2": [0: "terima"
1: "kasih"
2: "atas"
3: "responnya"
4: "ndan"
5: "poldajateng"
6: "tunggu"
7: "perkembangan"
8: "kasus"
9: "tersebut"]
"3": [0: "dibikin"
1: "ga"
2: "masuk"
3: "akal"
4:
.]
```

Gambar 4. 8 : Visualisasi Stopword Removal

4) *Steaming*

Proses ini dilakukan penghapusan imbuhan dari setiap kata. Sehingga mendapatkan kata dasarnya. Berikut ini adalah sourcecode *python* pada framework *streamlit* untuk proses *Steaming*.

```
135 # Stemming by Sastrawi
136     factory = StemmerFactory()
137     stemmer = factory.create_stemmer()
138     for k in dataset.keys():
139         tokens = []
140         for t in dataset[k]:
141             tokens.append(stemmer.stem(t))
142         dataset[k] = tokens
143
144 # Mengembalikan Format Dataset Awal
145     for k in dataset.keys():
146         dataset[k] = " ".join(dataset[k])
147     st.write("steaming", dataset)
```

Gambar 4. 9 : Steaming

selanjutnya penulis membuat tabel untuk *Stopword removal* dari 1-10 dokumen. Berikut adalah tabel hasil dari *Stopword removal* dari dokumen 1-10.

Tabel 4. 6 : Steaming

No	Stopword Removing	Steaming
1	lestarikan	lestari
	semangat	semangat
	persatuan	satu
	kesatuan	satu
2	terima	terima
	kasih	kasih
	atas	atas
	responnya	responnya
	ndan	ndan
	poldajateng	poldajateng
	tunggu	tunggu
	perkembangan	kembang
	kasus	kasus
	tersebut	sebut
3	dibikin	bikin
	ga	ga
	masuk	masuk
	akal	akal
	biar	biar
	susah	susah
	milih	milih
	buat	buat
	nembak	nembak
	aja	aja
	mayan	mayan
	dapet	dapat
	cuan	caun
	tuhh	tuhh
4	mantap	mantap
	pak	pak
	lisetyosigitp	lisetyosigitp
5	tilang	tilang
	sesuai	sesuai
	aturan	atur
	jgn	jgn
	buat2	buat

No	Stopword Removing	Steaming
	aturan	atur
	dewe	dewe
6	delik	delik
	pidana	pidana
	pencemaran	cemar
	nama	nama
	baik	baik
	ri1	ri
	kebukti	bukti
	nie	nie
	harusnya	harus
	tetap	tetap
	proses	proses
	hukum	hukum
	makhluk	makhluk
	satu	satu
7	laku	laku
	klo	klo
	gini	gin
8	jajaran	jajaran
	kepolisian	polisi
	lainnya	lain
	segera	segera
	ditangkap	tangkap
	pelakunya	laku
	kepercayaan	percaya
	masyarakat	masyarakat
	hilang	hilang
	dihakimi	hakim
	massal	massal
9	termasuk	masuk
	salah	salah
	satu	satu
	lolos	lolos
	kemudian	kemudian
	pernah	pernah
	ditanya	tanya

No	Stopword Removing	Steaming
	apa	apa
	keluhannya	keluh
	melakukan	laku
	tes	tes
	lapangan	lapang
	tsb	tsb
	kemudian	kemudian
	menyatakan	nyata
	sama	sama
	polisi	polisi
	tsb	tsb
	jawaban	jawabn
	yg	yg
	diberikn	diberikn
	sngt	sngt
	tdk	tdk
	memuaskan	muas
10	setuju	tuju
	ayo	ayo
	donk	donk
	pak	pak
	tunjukkan	tunjuk
	kalau	kalau
	polri	polri
	berpihak	pihak
	gerakan	gera
	buzzer	buzzer
	masif	masif
	organized	organized
	lho	lho
	tindak	tindak
	pak	pak
....		
122	foto	foto
	satu	satu
	peristiwa	peristiwa
	baik	baik

No	Stopword Removing	Steaming
	lebih	lebih
	baik	baik
	tunjukkan	tunjuk
	dgn	dgn
	peraturan	atur
	yg	yg
	melayani	layan
	memudahkan	mudah
	seluruh	seluruh
	rakyat	rakyat
	indonesia	indonesia

Pada framework stremlit berbasis *python* didapatkan hasil *Steaming* pada gambar berikut.

Steaming :

"1":
"lestari semangat satu satu"

"2":
"terima kasih atas responnya ndan poldajateng tunggu kembang kasus sebut"

"3":
"bikin ga masuk akal biar susah milih buat nembak aja mayan dapet cuan tuhh"

"4":
"mantap pak listyosigitp"

"5":
"tilang sesuai atur jgn buat atur dewe"

"6":
"delik pidana cemar nama baik ri bukti nie harus tetap proses hukum makhluk satu"

"7":
"laku klo gin"

"8":
"jajar polisi lain segera tangkap laku percaya masyarakat hilang hakim massal"

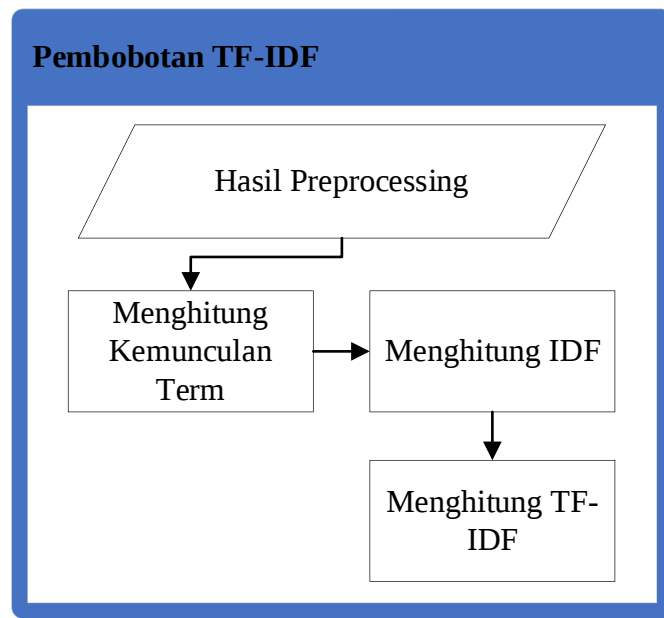
"9":
"masuk salah satu lolos kemudian pernah tanya apa keluhan laku tes lapang tsb kemudian nyata sama polisi tsb jawabn yg diberikn sngt tdk muas"

"10":
"tahu apa dengk pak tunjuk kalau naki nihak gara buzzer massif organized lha"

Gambar 4. 10 : Visualisasi Steaming

4.2.2 Hasil Pembobotan TF IDF

Tahapan selanjutnya setelah *preprocessing* dilakukan, kemudian peneliti memberikan bobot pada setiap kalimat dengan menggunakan pembobotan TF-IDF. Pada penggabungan dua konsep perhitungan yang digunakan dalam pembobotan TF-IDF yaitu inverse dan frekuensi kemunculan kata. Peneliti membuat *flowchart* untuk pembobotan TF-IDF berikut ini.



Gambar 4. 11 : Pembobotan TF-IDF

1) Term Frequency

Term Frequency atau TF merupakan proses yang dilakukan untuk perhitungan frekuensi kemunculan kata berdasarkan term yang didapatkan. Untuk Dokumen pada tabel diberi keterangan D1-D122. Berikut adalah Tabel hasil dari perhitungan term frequency.

Tabel 4. 7 : Hitung Term Frequency

No	Term	Term Frekuensi											
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10		D122
1	lestari	1	0	0	0	0	0	0	0	0	0	...	0
2	semangat	1	0	0	0	0	0	0	0	0	0	...	0
3	satu	2	0	0	0	0	1	0	0	1	0	...	1
4	terima	0	1	0	0	0	0	0	0	0	0	...	0

No	Term	Term Frekuensi											
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10		D122
5	kasih	0	1	0	0	0	0	0	0	0	0	...	0
6	atas	0	1	0	0	0	0	0	0	0	0	...	0
7	responnya	0	1	0	0	0	0	0	0	0	0	...	0
8	ndan	0	1	0	0	0	0	0	0	0	0	...	0
9	poldajateng	0	1	0	0	0	0	0	0	0	0	...	0
10	tunggu	0	1	0	0	0	0	0	0	0	0	...	0
11	kembang	0	1	0	0	0	0	0	0	0	0	...	0
12	kasus	0	1	0	0	0	0	0	0	0	0	...	0
13	sebut	0	1	0	0	0	0	0	0	0	0	...	0
14	bikin	0	0	1	0	0	0	0	0	0	0	...	0
15	ga	0	0	1	0	0	0	0	0	0	0	...	0
16	masuk	0	0	1	0	0	0	0	0	1	0	...	0
17	akal	0	0	1	0	0	0	0	0	0	0	...	0
18	biar	0	0	1	0	0	0	0	0	0	0	...	0
19	susah	0	0	1	0	0	0	0	0	0	0	...	0
20	milih	0	0	1	0	0	0	0	0	0	0	...	0
21	buat	0	0	1	0	0	0	0	0	0	0	...	0
22	nembak	0	0	1	0	0	0	0	0	0	0	...	0
23	aja	0	0	1	0	0	0	0	0	0	0	...	0
24	mayan	0	0	1	0	0	0	0	0	0	0	...	0
25	dapat	0	0	1	0	0	0	0	0	0	0	...	0
26	caun	0	0	1	0	0	0	0	0	0	0	...	0
27	tuhh	0	0	1	0	0	0	0	0	0	0	...	0
28	mantap	0	0	0	1	0	0	0	0	0	0	...	0
29	pak	0	0	0	1	0	0	0	0	0	2	...	0
30	lisetyosigitp	0	0	0	1	0	0	0	0	0	0	...	0
31	tilang	0	0	0	0	1	0	0	0	0	0	...	0
32	sesuai	0	0	0	0	1	0	0	0	0	0	...	0
33	jgn	0	0	0	0	1	0	0	0	0	0	...	0
34	buat2	0	0	0	0	1	0	0	0	0	0	...	0
35	atur	0	0	0	0	2	0	0	0	0	0	...	1
36	dewe	0	0	0	0	1	0	0	0	0	0	...	0
37	delik	0	0	0	0	0	1	0	0	0	0	...	0
38	pidana	0	0	0	0	0	1	0	0	0	0	...	0
39	cemar	0	0	0	0	0	1	0	0	0	0	...	0
40	nama	0	0	0	0	0	1	0	0	0	0	...	0
41	baik	0	0	0	0	0	1	0	0	0	0	...	2

No	Term	Term Frekuensi											
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10		D122
42	ri1	0	0	0	0	0	1	0	0	0	0	...	0
43	bukti	0	0	0	0	0	1	0	0	0	0	...	0
44	nie	0	0	0	0	0	1	0	0	0	0	...	0
45	harus	0	0	0	0	0	1	0	0	0	0	...	0
46	tetap	0	0	0	0	0	1	0	0	0	0	...	0
47	proses	0	0	0	0	0	1	0	0	0	0	...	0
48	hukum	0	0	0	0	0	1	0	0	0	0	...	0
49	makhluk	0	0	0	0	0	1	0	0	0	0	...	0
50	laku	0	0	0	0	0	0	1	1	1	0	...	0
51	klo	0	0	0	0	0	0	1	0	0	0	...	0
52	gin	0	0	0	0	0	0	1	0	0	0	...	0
53	jajaran	0	0	0	0	0	0	0	1	0	0	...	0
54	polisi	0	0	0	0	0	0	0	1	1	0	...	0
55	lain	0	0	0	0	0	0	0	1	0	0	...	0
56	segera	0	0	0	0	0	0	0	1	0	0	...	0
57	tangkap	0	0	0	0	0	0	0	1	0	0	...	0
58	percaya	0	0	0	0	0	0	0	1	0	0	...	0
59	masyarakat	0	0	0	0	0	0	0	1	0	0	...	0
60	hilang	0	0	0	0	0	0	0	1	0	0	...	0
61	hakim	0	0	0	0	0	0	0	1	0	0	...	0
62	massal	0	0	0	0	0	0	0	1	0	0	...	0
63	salah	0	0	0	0	0	0	0	0	1	0	...	0
64	lolos	0	0	0	0	0	0	0	0	1	0	...	0
65	kemudian	0	0	0	0	0	0	0	0	2	0	...	0
66	pernah	0	0	0	0	0	0	0	0	1	0	...	0
67	tanya	0	0	0	0	0	0	0	0	1	0	...	0
68	apa	0	0	0	0	0	0	0	0	1	0	...	0
69	keluh	0	0	0	0	0	0	0	0	1	0	...	0
70	tes	0	0	0	0	0	0	0	0	1	0	...	0
71	lapang	0	0	0	0	0	0	0	0	1	0	...	0
72	tsb	0	0	0	0	0	0	0	0	2	0	...	0
73	nyata	0	0	0	0	0	0	0	0	1	0	...	0
74	sama	0	0	0	0	0	0	0	0	1	0	...	0
75	jawabn	0	0	0	0	0	0	0	0	1	0	...	0
76	yg	0	0	0	0	0	0	0	0	1	0	...	1
77	diberikn	0	0	0	0	0	0	0	0	1	0	...	0
78	sngt	0	0	0	0	0	0	0	0	1	0	...	0

No	Term	Term Frekuensi											
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10		D122
79	tdk	0	0	0	0	0	0	0	0	1	0	...	0
80	puas	0	0	0	0	0	0	0	0	1	0	...	0
81	tuju	0	0	0	0	0	0	0	0	0	1	...	0
82	ayo	0	0	0	0	0	0	0	0	0	1	...	0
83	donk	0	0	0	0	0	0	0	0	0	1	...	0
84	tunjuk	0	0	0	0	0	0	0	0	0	1	...	1
85	kalau	0	0	0	0	0	0	0	0	0	1	...	0
86	polri	0	0	0	0	0	0	0	0	0	1	...	0
87	pihak	0	0	0	0	0	0	0	0	0	1	...	0
88	gerak	0	0	0	0	0	0	0	0	0	1	...	0
89	buzzer	0	0	0	0	0	0	0	0	0	1	...	0
90	masif	0	0	0	0	0	0	0	0	0	1	...	0
91	organized	0	0	0	0	0	0	0	0	0	1	...	0
92	lho	0	0	0	0	0	0	0	0	0	1	...	0
93	tindak	0	0	0	0	0	0	0	0	0	1	...	0
....													
94	foto	0	0	0	0	0	0	0	0	0	0	...	1
95	peristiwa	0	0	0	0	0	0	0	0	0	0	...	1
96	lebih	0	0	0	0	0	0	0	0	0	0	...	1
97	dgn	0	0	0	0	0	0	0	0	0	0	...	1
98	layan	0	0	0	0	0	0	0	0	0	0	...	1
99	mudah	0	0	0	0	0	0	0	0	0	0	...	1
100	seluruh	0	0	0	0	0	0	0	0	0	0	...	1
101	rakyat	0	0	0	0	0	0	0	0	0	0	...	1
102	indonesia	0	0	0	0	0	0	0	0	0	0	...	1
Panjang Dokumen :		4	10	14	3	7	14	3	11	24	15	0	15

Untuk mencari Term frequency di atas dengan cara menghitung jumlah kata yang sama dan muncul pada setiap dokumen, kemudian menjumlahkan tiap kata yang muncul dari tiap-tiap dokumen yang ada.

Dari hasil term frequency tersebut langkah selanjutnya melakukan proses normalisasi pada Term Frequency dengan cara membagi masing-masing dokumen dengan panjang dokumen. Untuk Dokumen pada tabel diberi keterangan D1-D122. Berikut ini tabel dari TF Normalisasi.

Tabel 4. 8 : Normalisasi TF

Hasil TF Normalisasi											
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10		D122
0,25	0	0	0	0	0	0	0	0	0		0
0,25	0	0	0	0	0	0	0	0	0		0
0,5	0	0	0	0	0,071	0	0	0,042	0		0,067
0	0,1	0	0	0	0	0	0	0	0		0
0	0,1	0	0	0	0	0	0	0	0		0
0	0,1	0	0	0	0	0	0	0	0		0
0	0,1	0	0	0	0	0	0	0	0		0
0	0,1	0	0	0	0	0	0	0	0		0
0	0,1	0	0	0	0	0	0	0	0		0
0	0,1	0	0	0	0	0	0	0	0		0
0	0,1	0	0	0	0	0	0	0	0		0
0	0,1	0	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0,042	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0,071	0	0	0	0	0	0	0		0
0	0	0	0,333	0	0	0	0	0	0		0
0	0	0	0,333	0	0	0	0	0	0,133		0
0	0	0	0,333	0	0	0	0	0	0		0
0	0	0	0	0,143	0	0	0	0	0		0
0	0	0	0	0,143	0	0	0	0	0		0
0	0	0	0	0,143	0	0	0	0	0		0
0	0	0	0	0,143	0	0	0	0	0		0
0	0	0	0	0,286	0	0	0	0	0		0,067
0	0	0	0	0,143	0	0	0	0	0		0

Hasil TF Normalisasi											
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10		D122
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0,133
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0,071	0	0	0	0		0
0	0	0	0	0	0	0,333	0,091	0,042	0		0
0	0	0	0	0	0	0,333	0	0	0		0
0	0	0	0	0	0	0,333	0	0	0		0
0	0	0	0	0	0	0	0,091	0	0		0
0	0	0	0	0	0	0	0,091	0,042	0		0
0	0	0	0	0	0	0	0,091	0	0		0
0	0	0	0	0	0	0	0,091	0	0		0
0	0	0	0	0	0	0	0,091	0	0		0
0	0	0	0	0	0	0	0,091	0	0		0
0	0	0	0	0	0	0	0,091	0	0		0
0	0	0	0	0	0	0	0,091	0	0		0
0	0	0	0	0	0	0	0,091	0	0		0
0	0	0	0	0	0	0	0	0,042	0		0
0	0	0	0	0	0	0	0	0,042	0		0
0	0	0	0	0	0	0	0	0,083	0		0
0	0	0	0	0	0	0	0	0,042	0		0
0	0	0	0	0	0	0	0	0,042	0		0
0	0	0	0	0	0	0	0	0,042	0		0
0	0	0	0	0	0	0	0	0,042	0		0
0	0	0	0	0	0	0	0	0,042	0		0
0	0	0	0	0	0	0	0	0,042	0		0
0	0	0	0	0	0	0	0	0,083	0		0
0	0	0	0	0	0	0	0	0,042	0		0

[illegible]

2) IDF (Inverse Document Frequency)

Sebelum melakukan perhitungan IDF, langkah selanjutnya adalah menghitung jumlah DF. Setelah dilakukan perhitungan DF kemudian melakukan perhitungan IDF. Dengan cara logaritma jumlah dokumen di

bagi Dokumen Frekuensi. Berikut ini merupakan tabel dari perhitungan DF dan IDF.

Tabel 4. 9 : Hitung Inverse Document Frequency

No	Term	Dokumen Frekuensi	IDF
1	lestari	1	1,041
2	semangat	1	1,041
3	satu	4	0,439
4	terima	1	1,041
5	kasih	1	1,041
6	atas	1	1,041
7	responnya	1	1,041
8	ndan	1	1,041
9	poldajateng	1	1,041
10	tunggu	1	1,041
11	kembang	1	1,041
12	kasus	1	1,041
13	sebut	1	1,041
14	bikin	1	1,041
15	ga	1	1,041
16	masuk	2	0,74
17	akal	1	1,041
18	biar	1	1,041
19	susah	1	1,041
20	milih	1	1,041
21	buat	1	1,041
22	nembak	1	1,041
23	aja	1	1,041
24	mayan	1	1,041
25	dapat	1	1,041
26	caun	1	1,041
27	tuhh	1	1,041
28	mantap	1	1,041
29	pak	2	0,74
30	lisetyosigitp	1	1,041
31	tilang	1	1,041
32	sesuai	1	1,041

No	Term	Dokumen Frekuensi	IDF
33	jgn	1	1,041
34	buat2	1	1,041
35	atur	2	0,74
36	dewe	1	1,041
37	delik	1	1,041
38	pidana	1	1,041
39	cemar	1	1,041
40	nama	1	1,041
41	baik	2	0,74
42	ri1	1	1,041
43	bukti	1	1,041
44	nie	1	1,041
45	harus	1	1,041
46	tetap	1	1,041
47	proses	1	1,041
48	hukum	1	1,041
49	makhluk	1	1,041
50	laku	3	0,564
51	klo	1	1,041
52	gin	1	1,041
53	jajaran	1	1,041
54	polisi	2	0,74
55	lain	1	1,041
56	segera	1	1,041
57	tangkap	1	1,041
58	percaya	1	1,041
59	masyarakat	1	1,041
60	hilang	1	1,041
61	hakim	1	1,041
62	massal	1	1,041
63	salah	1	1,041
64	lolos	1	1,041
65	kemudian	1	1,041
66	pernah	1	1,041
67	tanya	1	1,041
68	apa	1	1,041
69	keluh	1	1,041
70	tes	1	1,041

No	Term	Dokumen Frekuensi	IDF
71	lapang	1	1,041
72	tsb	1	1,041
73	nyata	1	1,041
74	sama	1	1,041
75	jawabn	1	1,041
76	yg	2	0,74
77	diberikn	1	1,041
78	sngt	1	1,041
79	tdk	1	1,041
80	puas	1	1,041
81	tuju	1	1,041
82	ayo	1	1,041
83	donk	1	1,041
84	tunjuk	2	0,74
85	kalau	1	1,041
86	polri	1	1,041
87	pihak	1	1,041
88	gerak	1	1,041
89	buzzer	1	1,041
90	masif	1	1,041
91	organized	1	1,041
92	lho	1	1,041
93	tindak	1	1,041
....		...	...
94	foto	1	1,041
95	peristiwa	1	1,041
96	lebih	1	1,041
97	dgn	1	1,041
98	layan	1	1,041
99	mudah	1	1,041
100	seluruh	1	1,041
101	rakyat	1	1,041
102	indonesia	1	1,041

Proses perhitungan DF yaitu dengan menjumlahkan kata yang muncul dari setiap dokumen kemudian dilakukan logaritma jumlah dokumen di bagi dengan jumlah kata yang muncul pada tiap-tiap dokumen.

3) Perhitungan TF-IDF

Tahapan terakhir pembobotan TF-IDF yaitu melakukan perhitungan TF-IDF itu sendiri. Perhitungan TF-IDF adalah dengan melakukan perkalian dari nilai TF dan IDF. Untuk Dokumen pada tabel diberi keterangan D1-D122. Dibawah ini adalah tabel dari perhitungan manual TF-IDF.

Tabel 4. 10 : Hitung TF-IDF

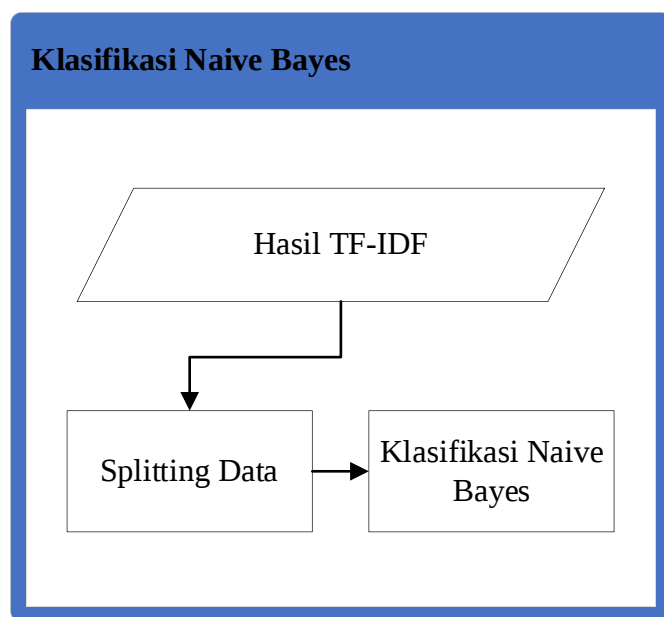
TF * IDF											
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	..	D122
0,26	0	0	0	0	0	0	0	0	0	..	0
0,26	0	0	0	0	0	0	0	0	0	..	0
0,22	0	0	0	0	0,031	0	0	0,018	0	..	0,029
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0,104	0	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,053	0	0	0	0	0	0,031	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0,074	0	0	0	0	0	0	0	..	0
0	0	0	0,347	0	0	0	0	0	0	..	0
0	0	0	0,246	0	0	0	0	0	0,098	..	0

TF * IDF											
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	..	D122
0	0	0	0,347	0	0	0	0	0	0	..	0
0	0	0	0	0,149	0	0	0	0	0	..	0
0	0	0	0	0,149	0	0	0	0	0	..	0
0	0	0	0	0,149	0	0	0	0	0	..	0
0	0	0	0	0,149	0	0	0	0	0	..	0
0	0	0	0	0,212	0	0	0	0	0	..	0,05
0	0	0	0	0,149	0	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,053	0	0	0	0	..	0,098
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0,074	0	0	0	0	..	0
0	0	0	0	0	0	0,188	0,051	0,024	0	..	0
0	0	0	0	0	0	0,347	0	0	0	..	0
0	0	0	0	0	0	0,347	0	0	0	..	0
0	0	0	0	0	0	0	0,095	0	0	..	0
0	0	0	0	0	0	0	0,067	0,031	0	..	0
0	0	0	0	0	0	0	0,095	0	0	..	0
0	0	0	0	0	0	0	0,095	0	0	..	0
0	0	0	0	0	0	0	0,095	0	0	..	0
0	0	0	0	0	0	0	0,095	0	0	..	0
0	0	0	0	0	0	0	0,095	0	0	..	0
0	0	0	0	0	0	0	0,095	0	0	..	0
0	0	0	0	0	0	0	0	0,044	0	..	0
0	0	0	0	0	0	0	0	0,044	0	..	0
0	0	0	0	0	0	0	0	0,086	0	..	0
0	0	0	0	0	0	0	0	0,044	0	..	0

Untuk mendapatkan nilai TF-IDF maka dilakukan dengan mengalikan nilai TF dengan IDF yang sudah didapatkan sebelumnya.

4.2.3 Klasifikasi Naive Bayes

Proses Klasifikasi *Naïve Bayes* dilakukan Setelah hasil Dokumen-dokumen tersebut diberikan bobot menggunakan TF-IDF. Pada tahapan ini peneliti melakukan klasifikasi menggunakan algoritma *Naïve Bayes classifier* dengan *library Scikit-Learn*. Dibawah ini adalah *flowchart* untuk proses klasifikasi *Naïve Bayes*.



Gambar 4. 12 : Klasifikasi Naive Bayes

Splitting data adalah tahapan yang dilakukan untuk membagi dataset menjadi data *testing* dan data *training*, proses *splitting* dilakukan sebelum proses klasifikasi. Berikut adalah source code dari proses *splitting* data.

```
163 t_size = st.radio(
164     "Pilih pembagian splitting data",
165     (0.1,0.2,0.3))
166 st.write("size splitting :", t_size)
167 if st.button('Mulai klasifikasi'):
168     st.success("3. naive Bayes")
169     st.write("Kernel GaussianNB")
170     #vectorizer = TfidfVectorizer(max_features=2500)
171     from sklearn.naive_bayes import GaussianNB
172     model_g = GaussianNB()
173     from sklearn.model_selection import train_test_split
174     v_data = vectorizer.fit_transform(dataset.values()).toarray()
175     X_train, X_test, y_train, y_test = train_test_split(v_data, dnb['label'], test_size=t_size, random_state=0)
```

Gambar 4. 13 : Code Splitting

Selanjutnya proses klasifikasi, peneliti menggunakan *library Scikit-Learn* pada streamlit untuk proses klasifikasi *Naïve Bayes*.

```

167     if st.button('Mulai klasifikasi'):
168         st.success("3. Naive Bayes")
169         st.write("Kernel GaussianNB")
170         #vectorizer = TfidfVectorizer(max_features=2500)
171         from sklearn.naive_bayes import GaussianNB
172         model_g = GaussianNB()
173         from sklearn.model_selection import train_test_split
174         v_data = vectorizer.fit_transform(dataset.values()).toarray()
175         X_train, X_test, y_train, y_test = train_test_split(v_data, dnb['label'], test_size=t_size, random_state=0)
176         nbmodel = model_g.fit(X_train, y_train)
177         nbmodel.class_count_
178         from sklearn.metrics import accuracy_score
179         y_preds = model_g.predict(X_test)
180         predict = pd.array(y_preds)
181         true_label = pd.array(y_test)
182         st.write("Data Testing Sebenarnya")
183         st.write(true_label)
184         st.write("Data Testing Vg Diprediksi")
185         st.write(predict)
186         proba = nbmodel.predict_proba(X_test)
187         st.write("Nilai Probabilitas")
188         st.write(proba)
189         st.write('nilai akurasi adalah ',
190                 accuracy_score(y_test, y_preds)*100)
191

```

Gambar 4. 14 : Code Klasifikasi

kemudian peneliti melakukan perhitungan untuk mendapatkan hasil klasifikasi berdasarkan model yang di bangun. Pada perhitungan ini peneliti menggunakan *microsoft excel*. Untuk total data *training* menggunakan 10 dokumen dari hasil perhitungan TF-IDF dengan jumlah 6 dokumen sentimen positif dan 4 dokumen sentimen negatif. Untuk Dokumen pada tabel diberi keterangan D1-D122. Berikut adalah tabel dari data *training*.

Tabel 4. 11 : Hitung Klasifikasi

Dokumen	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
lestari	0,26	0	0	0	0	0	0	0	0	0
semangat	0,26	0	0	0	0	0	0	0	0	0
satu	0,22	0	0	0	0	0,03	0	0	0,02	0

masuk	ga	bikin	sebut	kasus	kembang	tunggu	poldajateng	ndan	responnya	atas	kasih	terima
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0,104	0,104	0,104	0,104	0,104	0,1	0,104	0,1	0,104	0,104
0,053	0,07	0,074	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0,031	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

lisetyosigitp	pak	mantap	tuhh	caun	dapat	mayan	aja	nembak	buat	milih	susah	biar	akal
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0,07	0,07	0,074	0,074	0,07 7	0,074	0,07	0,074	0,074	0,07	0,07
0,347	0,25	0,347	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0,1	0	0	0	0	0	0	0	0	0	0	0	0

harus	nie	bukti	ri1	baik	nama	cemar	pidana	delik	dewe	atur	buat2	jgn	sesuai	tilang
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,15	0,21	0,149	0,15	0,149	0,149
0,074	0,07	0,074	0,07	0,05	0,074	0,074	0,074	0,07	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

percaya	tangkap	segera	lain	polisi	jajaran	gin	klo	laku	makhluk	hukum	proses	tetap
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,074	0,074	0,074	0,07
0	0	0	0	0	0	0,35	0,35	0,19	0	0	0	0
0,095	0,095	0,095	0,1	0,067	0,095	0	0	0,05	0	0	0	0
0	0	0	0	0,031	0	0	0	0,02	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

lapang	tes	keluh	apa	tanya	pernah	kemudian	lolos	salah	massal	hakim	hilang	masyarakat
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,095	0,095	0,095	0,095
0,044	0,04	0,044	0,04	0,044	0,044	0,086	0,04	0,044	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0

polri	kalau	tunjuk	donk	ayo	tuju	puas	tdk	sngt	diberikn	yg	jawabn	sama	nyata	tsb
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0,04	0,04	0,04	0,044	0,03	0,044	0,04	0,044	0,09
0,07	0,07	0,05	0,07	0,1	0,07	0	0	0	0	0	0	0	0	0

Sentimen	tindak	lho	organized	masif	buzzer	gerak	pihak
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
1	0,07	0,1	0,07	0,07	0,07	0,07	0,07

Kemudian untuk data *testing* penulis mengambil satu data dari data baru yang sudah di hitung nilai TF-IDF untuk pengklasifikasian, dengan harapan naive bayes mampu mengklasifikasikan data *testing* yang di pilih.

Tabel 4. 12 : Data Testing

Dokumen	foto	peristiwa	lebih	dgn	layan	mudah	seluruh	rakyat	indonesia
D11	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07	0,07

Selanjutnya menentukan nilai probabilitas masing masing sentimen dengan menjumlahkan banyaknya sentimen negatif dan sentimen positif.

Kemudian masing-masing sentimen di bagi jumlah keseluruhan sentimen. Berikut adalah tabel menghitung nilai probabilitas.

Tabel 4. 13 : Hitung Probabilitas

Hitung Probabilitas Tiap Sentimen		
Sentimen	Jumlah Training	Probabilitas
positif	6,00	0,6
negatif	4,00	0,4
Total	10,00	

Setelah menentukan nilai probabilitas proses selanjutnya melakukan perhitungan mean dari atribut tiap-tiap sentimen, dengan cara menjumlahkan tiap-tiap nilai sentimen positif dan melakukan cara yang sama pada sentimen negatif dari masing-masing atribut. Kemudian hasil dari penjumlahan tiap-tiap sentimen pada masing-masing atribut dibagi dengan jumlah masing-masing sentimen negatif dan positif. Berikut adalah tabel dari perhitungan mean dari atribut tiap-tiap sentimen.

Tabel 4. 14 : Hitung mean dari a tribut

Kelas	positif	negatif
lestari	0,04	0,00
semangat	0,04	0,00
satu	0,04	0,00
terima	0,02	0,00
kasih	0,02	0,00
atas	0,02	0,00

caun	dapat	mayan	aja	nembak	buat	milih	susah	biar	akal	masuk	ga	bikin	sebut	kasus	kembang	tunggu	poldajaten g	ndan	responnya
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02	0,02	0,02	0,02	0,02	0,02	0,02
0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,00	0,00	0,00	0,00	0,00	0,00	0,00

harus	nie	bukti	ri1	baik	nama	cemar	pidana	delik	dewe	atur	buat2	ign	sesuai	tilang	lisetyosigitp	pak	mantap	tuhh
0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,01	0,02	0,04	0,02	0,02	0,02	0,02	0,06	0,06	0,06	0,00
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,02

salah	massal	hakim	hilang	masyarakat	percaya	tangkap	segera	lain	polisi	jajaran	gin	klo	laku	mahluk	hukum	proses	tetap
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,01
0,01	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,02	0,09	0,09	0,07	0,00	0,00	0,00	0,00

tindak	0,01	0,00
lho	0,01	0,00
organized	0,01	0,00
masif	0,01	0,00
buzzer	0,01	0,00
gera k	0,01	0,00
pihak	0,01	0,00
polri	0,01	0,00
kalau	0,01	0,00

Selanjutnya menghitung standart deviasi pada masing-masing sentimen di tiap-tiap atribut. Peneliti menggunakan rumus di *microsoft excel* “=STDEV.P(nilai atribut di masing masing sentimen)” . berikut adalah tabel perhitungan dari standart deviasi.

Tabel 4. 15 : Hitung Standar Deviasi

Kelas	positif	negatif
lestari	0,10	0,00
semangat	0,10	0,00
satu	0,08	0,01
terima	0,04	0,00

harus	nie	bukti	ri1	baik	nama	cemar	pidana	delik	dewe	atur	buat2	jgn	sesuai	tilang	lisetyosigitp	pak	mantap	tuhh	caun	dapat
0,03	0,03	0,03	0,03	0,02	0,03	0,03	0,03	0,03	0,06	0,08	0,06	0,06	0,06	0,06	0,13	0,09	0,13	0,00	0,00	0,00
0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,03	0,03	0,03

tindak	0,03	0,00
lho	0,03	0,00
organized	0,03	0,00
masif	0,03	0,00
buzzer	0,03	0,00
gerak	0,03	0,00
pihak	0,03	0,00

Menghitung distribusi normal dari atribut masing-masing sentimen. Peneliti menggunakan rumus di *microsoft excel* =NORM.DIST(nilai atribut data *testing* dan nilai masing-masing atribut pada data *training*)” . berikut adalah tabel perhitungan dari distribusi normal.

Tabel 4. 16 : Hitung Distribusi Normal

Sentimen	positif	negatif
probabilitas	0,6	0,4
lestari	0,61	1,00
semangat	0,61	1,00
satu	0,64	1,00
terima	0,91	1,00

segera	lain	polisi	jajaran	gin	klo	laku	makhluk	hukum	proses	tetap	harus	nie	bukti	ri1	baik	nama	cemar
1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98	1,00	0,98	0,98
0,87	0,87	0,95	0,87	0,46	0,46	0,52	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

nyata	tsb	lapang	tes	keluh	apa	tanya	pernah	kemudian	lolos	salah	massal	hakim	hilang	masyarakat	percaya	tangkap
1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
1,00	0,90	1,00	1,00	1,00	1,00	1,00	1,00	0,90	1,00	1,00	0,87	0,87	0,87	0,87	0,87	0,87

tindak	lho	organized	masif	buzzer	gerak	pihak	polri	kalau	tunjuk	donk	ayo	tuju	puas	tdk	sngt	diberikn	yg	jawabn	sama
0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	1,00	0,99	0,99	0,99	1,00	1,00	1,00	1,00	1,00	1,00	1,00
1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Langkah terakhir adalah menghitung probabilitas tiap sentimen dan klasifikasi hasil akhir. Berikut ini adalah tabel hasil dari klasifikasi.

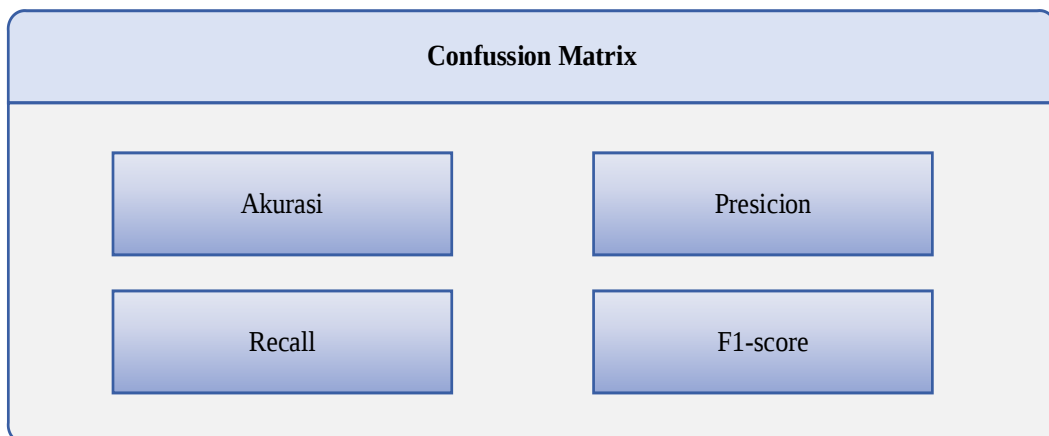
Tabel 4. 17 : Hitung Probabilitas dan Hasil Klasifikasi

PROBABILITAS NAÏVE BAYES	
PROBABILITAS	Sentimen
0,5923958044	POSITIF
0,4000000000	NEGATIF

HASIL AKHIR KLASIFIKASI	
POSITIF	0,59

4.2.4 Confusion matrix

Dalam evaluasi model hal yang perlu dilakukan adalah dengan melakukan pengujian terhadap model yang sudah dibangun, peneliti melakukan pengujian model menggunakan *confusion matrix*. berikut adalah flowchart confusion matrix



Gambar 4. 15 : Confussion Matrix

Confusion matrix digunakan untuk mengevaluasi sistem klasifikasi yang berdasarkan kemampuan parameter prediktif suatu model. parameter prediktif adalah parameter yang mengukur kualitas set uji statistik yang diberikan berdasarkan atribut yang ada dalam sampel.

Dibawah ini adalah dataset yang digunakan untuk melakukan pengujian terhadap model yang telah dibangun.

Tabel 4. 18 : Dataset

No	Text	sentimen aktual
84	ada kasus di Palopo thn 2021, bahkan WARTAWAN yg seharusnya dilindungi oleh UU Pers nyatanya dihukum menurut UU ITE, KARENA MELAPORKAN TINDAK KEJAHATAN KORUPSI entah di mana rasa keadilan Hakim dan ketaatan pada UU Pers. Terus koruptornya??? https://t.co/IPcZtB1DoI	Negatif
97	Tolong dong pak, pengawalan pejabat masih banyak brisik pakai sirine...	Positif
10	Saya juga termasuk salah satu yang lolos, kemudian saya pernah ditanya apa keluhannya saat melakukan tes lapangan tsb, dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb, namun jawabn yg diberikn sngt tdk memuaskan.	Negatif
71	ujian sim kek ujian jd pemain sirkus. mohon dipertimbangkan ulang ujian praktek tersebut. Saya gak puas pak.	Positif
8	dan jajaran kepolisian lainnya bisa segera ditangkap pelakunya, sebelum kepercayaan masyarakat hilang, sehingga akan dihakimi massal.	Negatif
45	Faktanya yang banyak membunuh rakyat dan korup itu aparat ,kok	Positif
22	Hoax dan Ujaran Kebencian dapat mengancam persatuan \n\nLaporkan dan Janban Ikut Menyebarkan\	Positif
7	Delik pidana pencemaran nama baik RI1 sudah terbukti nie,. harusnya tetap bisa proses hukum makhluk satu itu.	Positif

No	Text	sentimen aktual
61	Langsung tinas saja pak ..sudah banyak warga masyarakat yg tertinas karena hal tersebut	Positif
90	Tangkap pa polisi orang yg seperti ini	Positif
86	Liat aja bnyk kasus,setelh memviralkn jln atau apa,ujung2nya rakyat minta maaf,aku curiga rakyat itu di takut2i uu pencemarn nama baik	negatif
109	Berbenah dan evaluasi harus dilakukan . Ini untuk kebaikan polri jg kedepannya. Jgn sampai kejadian yg udah2 terjadi lagi.Terima Kasih	Positif
33	ujian sim menurut saya harus di lihat kembali, cek di lapangan berapa yg pakai ujian + berapa yg hanya langsung foto jadi sim nya	Negatif

Selanjutnya melakukan proses pengujian terhadap model yang telah dibangun sesuai dengan dataset, kemudian data *testing* akan dibandingkan dengan sentimen aktual yang telah didapatkan dari implementasi model yang dibangun. pada tabel confusion matrix diberi keterangan FP(False Positif), FN(False Negatif), TP(True Positif) dan TN(True Negatif). Berikut adalah tabel perbandingan sentimen yg didapatkan.

Tabel 4. 19 : Perbandingan sentimen

No	Text	sentimen aktual	prediksi	confusion matrix
84	ada kasus di Palopo thn 2021, bahkan WARTAWAN yg seharusnya dilindungi oleh UU Pers nyatanya dihukum menurut UU ITE, KARENA MELAPORKAN TINDAK KEJAHATAN KORUPSI entah di mana rasa keadilan Hakim dan ketaatan pada UU Pers. Terus koruptornya??? https://t.co/IPcZtB1DoI	Negatif	Positif	FP
97	Tolong dong pak, pengawalan pejabat masih banyak brisik pakai sirine...	Positif	Positif	TP
10	Saya juga termasuk salah satu yang lolos, kemudian saya pernah ditanya apa keluhannya saat melakukan tes lapangan tsb, dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb, namun jawabn yg diberikn sngt tdk memuaskan.	Negatif	Negatif	TN
71	ujian sim kek ujian jd pemain sirkus. mohon dipertimbangkan ulang ujian praktek tersebut. Saya gak puas pak.	Positif	Positif	TP
8	dan jajaran kepolisian lainnya bisa segera ditangkap pelakunya, sebelum kepercayaan masyarakat hilang, sehingga akan dihakimi massal.	Negatif	Negatif	TN
45	Faktanya yang banyak membunuh rakyat dan korup itu aparat ,kok	Positif	Positif	TP

No	Text	sentimen aktual	prediksi	confusion matrix
22	Hoax dan Ujaran Kebencian dapat mengancam persatuan \n\nLaporkan dan Janban Ikut Menyebarkan\	Positif	Positif	TP
7	Delik pidana pencemaran nama baik RI1 sudah terbukti, harusnya tetap bisa proses hukum makhluk satu itu.	Positif	Positif	TP
61	Langsung tinas saja pak ..sudah banyak warga masyarakat yg tertinas karena hal tersebut	Positif	Positif	TP
90	Tangkap pa polisi orang yg seperti ini	Positif	Positif	TP
86	Liat aja bnyk kasus,setelh memviralkn jln atau apa,ujung2nya rakyat minta maaf,aku curiga rakyat itu di takut2i uu pencemarn nama baik	negatif	Positif	FP
109	Berbenah dan evaluasi harus dilakukan . Ini untuk kebaikan polri jg kedepannya. Jgn sampai kejadian yg udah2 terjadi lagi.Terima Kasih	Positif	Positif	TP
33	ujian sim menurut saya harus di lihat kembali, cek di lapangan berapa yg pakai ujian + berapa yg hanya langsung foto jadi sim nya	Negatif	Positif	FP

Tahapan terakhir evalusai model yaitu melakukan perhitungan manual *confusion matrix* untuk mengetahui nilai akurasi, *presicion*, *recall* dan *f1-score* dari klasifikasi tersebut. pada tabel confussion matrix diberi keterangan FP(False

Positif), FN(False Negatif), TP(True Positif) dan TN(True Negatif). Berikut adalah tabel perhitungan manual *confusion matrix*.

Tabel 4. 20 : Hitung Confusion Matrix

Predicted class Actual Class	Positif	Negatif
Positif	TP = 8	FN = 0
Negatif	FP = 3	TN = 2

Selanjutnya melakukan proses perhitungan manual dengan menggunakan rumus berikut ini.

1. Akurasi

Untuk menghitung akurasi ini dilakukan dengan menggunakan persamaan (10) berikut

$$\begin{aligned}
 \text{akurasi} &= \frac{(TP+TN)}{TP+TN+FP+FN} \times 100 & (10) \\
 &= \frac{(8 + 2)}{8 + 2 + 3 + 0} \\
 &= \frac{10}{13} = 0,76 \times 100 = 76\%
 \end{aligned}$$

2. Precission

Untuk menghitung Precission ini dilakukan dengan menggunakan persamaan (11) berikut.

$$\begin{aligned}
 \text{Presisi} &= \frac{(TP)}{FP+TP} & (11) \\
 &= \frac{8}{3+8} = \frac{8}{11} = 0,73
 \end{aligned}$$

3. Recall

Untuk menghitung Precision ini dilakukan dengan menggunakan persamaan (12) berikut.

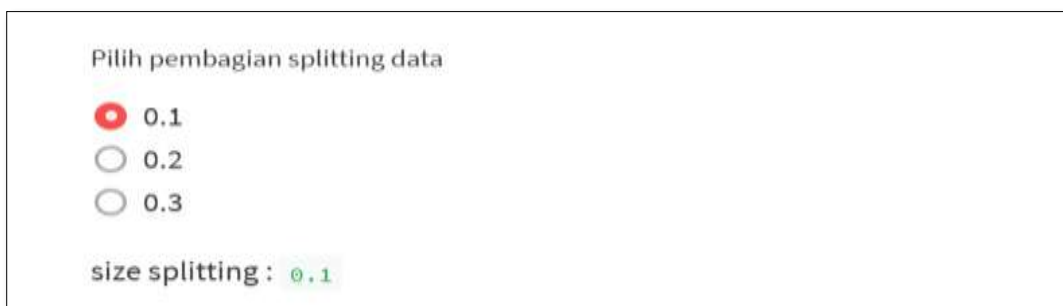
$$\begin{aligned}
 Recall &= \frac{TP}{FN+TP} \\
 &= \frac{8}{0+8} = \frac{8}{8} = 1
 \end{aligned}
 \tag{12}$$

4. F-Measure

Untuk menghitung F-Measure ini dilakukan dengan menggunakan persamaan (12) berikut.

$$\begin{aligned}
 f - measure &= \frac{(2 \times recall \times presisi)}{(recall + presisi)} \\
 &= \frac{2 \times 1 \times 0,73}{1 + 0,73} \\
 &= \frac{1,46}{1,73} = 0,84
 \end{aligned}
 \tag{13}$$

Dari pengujian menggunakan perhitungan manual *confusion matrix* dapat dibandingkan dengan hasil visualisasi dari *confusion matrix* menggunakan *framework streamlit* berbasis *python*. Berikut adalah gambar dari visualisasi *confusion matrix* dengan *splitting* 0,1 dari 122 dataset.



Gambar 4. 16 : Visualisasi Splitting



Gambar 4. 17 : Visualisasi Confusion Matrix

BAB V

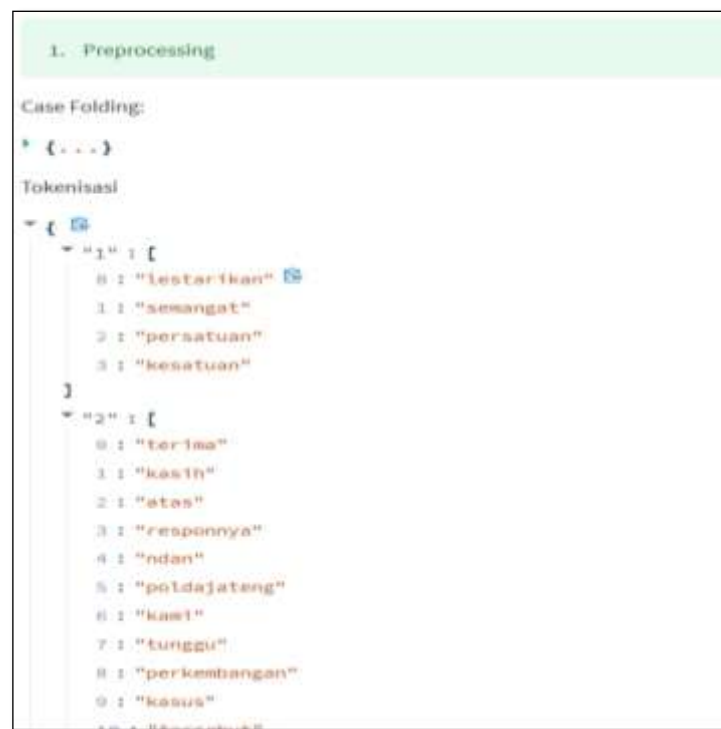
PEMBAHASAN

5.1 Pembahasan Model

pada pembangunan model peneliti memanfaatkan framework streamlit berbasis python untuk melakukan klasifikasi dengan algoritma Naive bayes Classifier. Penggunaan streamlit ini bersifat open source yang dibuat untuk membangun aplikasi web dibidang sains data dan machine learning yang interaktif.

5.1.1 Preprocessing

Dari tahapan preprocessing pada model yang dibangun masih terdapat permasalahan yang membuat penelitian ini bekerja tidak maksimal, contohnya saja menjalankan preprocessing pada pemodelan membutuhkan waktu kurang lebih 1 menit untuk dapat memproses 122 dataset. Permasalahan lain terdapat pada proses Tokenizing yaitu melakukan pemisahan kata per kata tetapi tidak bekerja pada 2 kata yang tidak menggunakan space. Berikut adalah preprocessing pada pemodelan.



Gambar 5. 1 : Preprocessing

5.1.2 Pembobotan

Pada tahapan pembobotan kata, peneliti menggunakan metode TF-IDF. ini bertujuan untuk memberikan bobot pada setiap kata. sehingga hasil yang di dapatakn pada pemodelan adalah sebagai berikut.



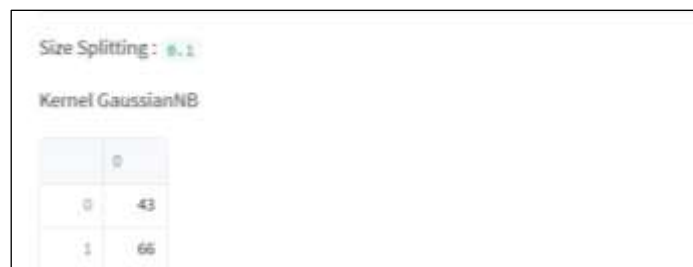
2. Pembobotan Dokumen (TF-IDF)

	bagus	bagusanya	bagus	bagus	is	bagus	bagus	bagus	bagus	bagus	bagus	bagus	bagus	bagus	bagus	bagus	bagus	bagus	bagus	bagus
4	0	0	0	0	0	0	0	0	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	0	0	0	0	0	0	0	0
6	0	0	1	0	0	0	0	0	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	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	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	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	1	0	0	0	1	0	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	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Gambar 5. 2 : Pembobotan

5.1.3 Klasifikasi

Peneliti menggunakan metode Naive Bayes Classifier untuk melakukan klasifikasi dengan menggunakan *library Scikit-learn* pada framework streamlit. Hasil yang didapat adalah 43 nilai negatif dan 66 nilai positif. Berikut adalah gambar dilakukannya klasifikasi.



Size Splitting: 0.1

Kernel GaussianNB

	0	
0		43
1		66

Gambar 5. 3 : klasifikasi

5.1.4 Evaluasi

Metode penelitian yang dimulai dari pengumpulan data sampai dengan melakukan perhitungan nilai probabilitas menggunakan Naive Bayes Classifier peneliti mendapat nilai akurasi 76% dari 122 dataset, dimana dari 122 data tersebut dilakukan splitting yaitu 10% data *testing* dan 90% data *training*,

sehingga didapatkan nilai akurasi 76%, sedangkan pada *splitting* 20% data *testing* dan 80% data *training* hanya mendapatkan nilai akurasi 72% dan *splitting* 30% data *testing* dan 70% data *training* hanya mendapatkan nilai akurasi sebesar 64%. Berikut adalah visualisasi *splitting* pada framework *streamlit* dengan nilai akurasi tertinggi.



Gambar 5. 4 : Evaluasi

5.1.5 Implementasi

pada tahapan implementasi model dilakukan proses uji coba terhadap model yang dibangun. Implementasi model bekerja untuk menguji *tweet-tweet* yang baru sehingga bisa diklasifikasikan menjadi positif atau negatif.

Tweet-tweet yang baru dan akan diuji pada implementasi model yang telah dibangun menggunakan *framework streamlit* berbasis *python* dapat dilihat pada tabel berikut.

Tabel 5. 1 : Tweet Baru

No	text	Label
1	@LH_265 @DewantoroHelen Jujur, saya mulai meragukan tingkat kepuasan masyarakat terhadap institusi @DivHumas_Polri bisa mencapai 80%.	

No	text	Label
2	@Irwan2yah @DivHumas_Polri @kemendagri @tjahjo_kumolo Terakhir paling klarifikasi dan khilaf. Materai 10k. Kapan mari ne Indonesia kita bisa bebas dari orang-orang seperti mereka. Tolong polri segera brantas semua golongan meraka. ☐	
3	Anggota Bhabinkamtibmas Polsek Senori Polres Tuban Melaksanakan Pengamanan Pemungutan Suara Pilkades Tahun 2022 di Desa Rayung dan Desa Katerban Kec, Senori @DivHumas_Polri @HumasPoldaJatim @humas_restuban @PolsekSenori #polisi #polisijatim #polri	

Berikut adalah visualisasi dari implementasi model yang telah dibangun untuk menguji *tweet-tweet* baru.



Gambar 5. 5 : Visualisasi Implementasi Model

Kemudian *tweet* yang baru dimasukan kedalam kolom text yang tersedia untuk melakukan pengujian terhadap *tweet* apakah dapat diklasifikasikan menjadi sentimen positif atau sentimen negatif.

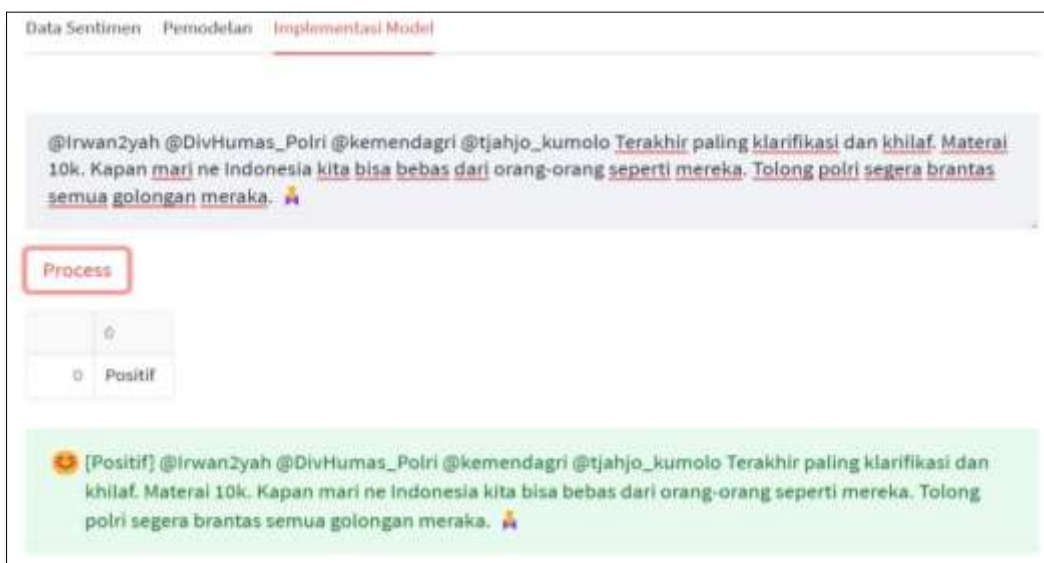
Berikut adalah hasil visualisasi dari implementasi model pada *streamlit* dengan memasukkan *tweet* yang pertama



Gambar 5. 6 : Hasil Implementasi 1

Dari *tweet* yang telah diuji pada implementasi model, dapat dilihat bahwa *tweet* tersebut diklasifikasikan sebagai sentimen “Negatif”.

Selanjutnya adalah hasil visualisasi dari implementasi model pada *streamlit* dengan memasukkan *tweet* yang kedua.



Gambar 5. 7 : Hasil Implementasi 2

Dari *tweet* yang telah diuji pada implementasi model, dapat dilihat bahwa *tweet* tersebut diklasifikasikan sebagai sentimen “Positif”.

selanjutnya adalah hasil visualisasi dari implementasi model pada *streamlit* dengan memasukkan *tweet* yang ketiga.



Gambar 5. 8 : Hasil Implementasi 3

Dari *tweet* yang telah diuji pada implementasi model, dapat dilihat bahwa *tweet* tersebut diklasifikasikan sebagai sentimen “Positif”.

BAB VI

PENUTUP

6.1 KESIMPULAN

Pada penelitian ini penulis telah melakukan pemodelan menggunakan metode *Naïve Bayes classifier*, maka berdasarkan hasil yang diketahui bahwa algoritma *Naïve Bayes* mampu mengklasifikasikan sentimen pada *tweet* dari Divisi Humas Polri. Dengan menggunakan algoritma ini penulis mendapatkan nilai akurasi yang cukup tinggi. Dengan memanfaatkan 122 data *tweet* penulis mendapatkan hasil 61% data positif dan 39% data negatif dengan akurasi sebesar 76%. Hasil menunjukkan tingkat nilai positif lebih besar dari nilai negatif. Dengan demikian banyak masyarakat yang merasa puas dengan pelayanan polisi di Indonesia berdasarkan data dari twitter. Kemudian peneliti melakukan pengujian terhadap 5 data baru opini masyarakat tentang pelayanan kepolisian dengan memanfaatkan implementasi pada pemodelan, penulis mendapatkan hasil dengan 2 data negatif dan 3 data positif yang kemudian peneliti berharap ini akan menjadi bahan evaluasi bagi pihak kepolisian untuk meningkatkan pelayanan terhadap masyarakat.

6.2 SARAN

Berdasarkan kesimpulan yang telah dituangkan penulis, maka ada beberapa saran yang perlu untuk disampaikan, ini merupakan saran yang diberikan peneliti untuk tindak lanjut atau pengembangan yang lebih mendalam bagi peneliti-peneliti selanjutnya, berikut ini adalah saran untuk mencapai tujuan yaitu :

- 1) Pada penelitian selanjutnya, untuk dapat mempertimbangkan sampel data *tweet* atau komentar pada media sosial untuk dilakukan analisis.
- 2) Pada penelitian selanjutnya, menambahkan model *deployment* atau membangun *website* agar hasil prediksi dapat diakses secara *online*.
- 3) Pada penelitian selanjutnya, disarankan untuk menggunakan lebih dari satu metode untuk melakukan klasifikasi.
- 4) Pada penelitian selanjutnya, proses implementasi dapat memproses data dengan jumlah yang banyak.

- 5) Pada penelitian selanjutnya, menambahkan fitur untuk tweet dari twiter secara *real time* pada model yang akan dibangun.
- 6) Pada Peneliti Selanjutnya, Menambahkan Preprocessing Tokenizing untuk memisahkan kata yang tidak menggunakan *space*.

DAFTAR PUSTAKA

- [1] G. Tika, A. Adiwijaya, and N. D. W. Cahyani, "Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan Multilayer Perceptron," *eProceedings Eng.*, vol. 6, no. 1, pp. 3631–3640, 2019, [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/8503>
- [2] Consejería de Educación. Junta de Castilla y León, "濟無No Title No Title No Title," *Protoc. prevención y Organ. del regreso a la Act. lectiva en los centros Educ. Castilla y León para el curso Esc. 2020/2021*, no. 2, pp. 1–19, 2020.
- [3] S. Fransiska, "ANALISIS SENTIMEN TWITTER UNTUK REVIEW FILM MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER (NBC) PADA SENTIMEN R PROGRAMMING," *J. Siliwangi*, vol. 5, no. 2, 2019.
- [4] B. M. Pintoko and K. M. L., "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier," *e-Proceeding Eng.*, vol. 5, no. 3, pp. 8121–8130, 2018.
- [5] A. Zaki, M. Faiz, "Survei Indikator: Tingkat Kepercayaan Publik terhadap Polri Menurun," *nasional.tempo.co*, 2022. <https://nasional.tempo.co/read/1626759/survei-indikator-tingkat-kepercayaan-publik-terhadap-polri-menurun> (accessed Mar. 28, 2023).
- [6] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes," *J. Sains dan Inform.*, vol. 7, no. 2, pp. 155–164, Dec. 2021, doi: 10.34128/jsi.v7i2.331.
- [7] O. Penelitian, D. Metode Penelitian, and O. Penelitan, "BAB III."
- [8] S. Nurul, J. Fitriyyah, N. Safriadi, E. Esyudha, and P. #3, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes", [Online]. Available: <http://dev.twitter.com>.

- [9] K. Sihotang and R. Ghaniy, “Penerapan Metode Naïve Bayes Classifier Untuk Penentuan Topik Tugas Akhir Pada Website Perpustakaan STIKOM Binaniaga,” *J. Ilm. Teknol. dan Sains*, vol. 9, pp. 63–72, 2019.
- [10] R. Yasmin, “Covid-19 Menggunakan Metode Naive Bayes Classifier Pada Media Sosial Twitter Covid-19 Menggunakan Metode Naive Bayes,” 2021.
- [11] M. Imron, “Penerapan Data Mining Algoritma Naives Bayes Dan Part Untuk Mengetahui Minat Baca Mahasiswa Di Perpustakaan Stmik Amikom Purwokerto,” *J. Telemat.*, vol. 10, no. 2, pp. 121–135, 2017.
- [12] V. I. Santoso, G. Virginia, and Y. Lukito, “Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine,” *J. Transform.*, vol. 14, no. 2, p. 72, 2017, doi: 10.26623/transformatika.v14i2.439.

LAMPIRAN

DATASET

No	Dateline	text	Username	label
1	2022-10-27 23:59:51+00:00	Lestarikan Semangat Persatuan & Kesatuan	The4_9Institute	Positif
2	2020-02-28 14:06:09+00:00	Terima kasih atas responnya, Ndan poldajateng Kami tunggu perkembangan kasus tersebut! ☐	Nasikin_MA	Positif
3	2022-10-27 23:59:35+00:00	Dibikin ga masuk akal biar susah dan milih buat nembak aja, mayan dapet cuan tuhh☐	yesimsoquiet	Negatif
4	2022-10-27 23:57:20+00:00	Mantap Pak ListyoSigitP	The4_9Institute	Positif
5	2022-10-27 23:56:54+00:00	Tilang saja sesuai aturan ...!! jgn buat2 aturan dewe☐	Ojoklalenya	Positif
6	2022-10-27 23:56:25+00:00	Delik pidana pencemaran nama baik RI1 sudah kebukti nie,. harusnya tetap bisa proses hukum makhluk satu itu.	ApriliaLin	Positif
7	2022-10-27 23:55:48+00:00	Tidak laku klo gini	SSobrah	Negatif
8	2022-10-27 23:55:28+00:00	dan jajaran kepolisian lainnya bisa segera ditangkap pelakunya, sebelum kepercayaan masyarakat hilang, sehingga akan dihakimi massal.	bariekly_	Negatif

No	Dateline	text	Username	label
9	2022-10-27 23:54:58+00:00	Saya juga termasuk salah satu yang lolos, kemudian saya pernah ditanya apa keluhannya saat melakukan tes lapangan tsb, dan kemudian saya menyatakan hal sama seperti ini kepada polisi tsb, namun jawabn yg diberikn sngt tdk memuaskan.	srohimah109	Negatif
10	2022-10-27 23:54:17+00:00	Setuju. Ayo donk pak tunjukkan kalau Polri tidak berpihak, gerakan para buzzer ini masif dan organized lho. Tindak mereka pak.	akuboekankamu	Positif
11	2022-10-27 23:53:29+00:00	jenis sangsi seperti ini ada di pasal brp ayat brp UU lalu lintas ya. Katanya gak boleh bawa2 Identitas tertentu, lha ini malah aparat penegak Hukum sendiri yg kasih contoh.	kumur69	Negatif
12	2022-10-27 23:52:10+00:00	Bpom kenapa tidak di periksa dan di audit selaku pengawas peredaran obat. Kok bahan berbahaya bisa lolos uji klinis ? ?	kecenk_spike	Negatif
13	2022-10-27 23:50:31+00:00	Ayo pak segera ditindak, jangan sampe nanti di katakan B***I	rizieqdivist	Positif
14	2022-10-27 23:48:45+00:00	Lanjutkan saja pak polisi, biar yg beginian jadi efek jera.masa seenaknya minta maaf	Sasorishared	Positif
15	2022-10-27 23:48:34+00:00	Apa benar semua polisi gangguan jiwa saat ini	BangTAJIR8	Negatif
16	2022-10-27 23:47:45+00:00	tolong dibantu cek apakah info dari orang ini benar.. kalo tidak terbukti.. mohon agar dilakukan tindakan tegas. Tmksh.	HrBmntyo21	Positif

No	Dateline	text	Username	label
17	2022-10-27 23:45:13+00:00	Semoga polri terus berbenah dan benar2 menjadi pengayom di masyarakat...aamiiin	arifsugiri	Positif
18	2022-10-27 23:44:39+00:00	Waduh Penerapan Hukum yang tidak pada tempatnya.Jangan diteruskan deh pakk, pls.Jangan dicampur aduk deh□□	Msimsim2ND	Negatif
19	2022-10-27 23:42:29+00:00	maaf pak polisi sebenarnya hal yang seperti ini tidak berguna kita mendapatkan SIM itu bukan untuk balapan tapi untuk keperluan sehari2 makanya ketika test dilapangan itu bukan seperti ini tapi yang bersipat keseharian bgmn cara patuh ketika melintas jalan KA	PrianganSelatan	Positif
20	2022-10-27 23:41:16+00:00	Sblum tambah merepotkan aparat,lebih baik kepolisian melarang/membatasi kegiatan itu!!	S2MByM1	Positif
21	2022-10-27 23:40:10+00:00	Hoax dan Ujaran Kebencian dapat mengancam persatuan \n\nLaporkan dan Janban Ikut Menyebarkan\	jakulasi	Positif
22	2022-10-27 23:36:57+00:00	apakah anda akan membiarkan perempuan stres Kadrin ini menyebar kebohongan seperti ini.	NusantaraBerstu	Negatif
23	2022-10-27 23:36:31+00:00	dr semenjak pendidikan mereka semua sudah penuh kepalsuan, maka menjabat dan kerja DILAPANGAN ilmu menipu memanipulasi dan mencari jalan haram di terapkan. Sudah sangat parah ListyoSigitP	mouchethung	Negatif

No	Dateline	text	Username	label
24	2022-10-27 23:36:02+00:00	aparap kok gitu ya..aneh aja	tulalitwidos	Negatif
25	2022-10-27 23:35:29+00:00	bapak ListyoSigitP kyknya kerja sendiri.. dikangkangi anak buahnya..	Seribudapat3	Negatif
26	2022-10-27 23:33:30+00:00	Katanya di DAMKAR DKI ada korupsi senilai Rp. 6 triliyun lebih yg kayaknya Abas gak mungkin tidak terlibat !, siap2 para kadrun akan bsrjuang mati2an membela boss Yaman !!!	MamanSu74843985	Negatif
27	2022-10-27 23:32:41+00:00	polisi harus jujur dan santun karna kebanyakan boong dan arogan	loyaloy29070088	Negatif
28	2022-10-27 23:31:32+00:00	Makin rusak ni Polri ListyoSigitP	hoax_bukan	Negatif
29	2022-10-27 23:30:53+00:00	salut buat Polri yg dg seriusnya ikut menangani berbagai kasus,sampai kasus virus aja gk segan ² Polri turun tangan□	sanipar7	Positif
30	2022-10-27 23:29:32+00:00	DivHumas_Polri ini harusnya bikin tim khusus buat berantas premanisme. Ungakapannya gini, "Utk bisa menangkap anjing, polisi harus pake anjing yg lbh liar dan gila".	rizieqdivist	Positif
31	2022-10-27 23:29:29+00:00	Apakah proses pendaftaran kendaraan roda tiga listrik tidak sama untuk setiap polda?	Priyatno	Positif
32	2022-10-27 23:29:19+00:00	ujian sim menurut saya harus di lihat kembali, cek di lapangan berapa yg pakai ujian + berapa yg hanya langsung foto jadi sim nya	rudiwaluyo	Negatif

No	Dateline	text	Username	label
33	2022-10-27 23:28:46+00:00	Kalau pengendara mobil banyak yang kesal pada pengendara motor yg melaju meliuk-liuk di jalanan, coba datang dan lihat ujian praktek di Samsat utk SIM C. Coba tanya ke , utk apa ujian zigzag dan putaran angka 8?	indrabustami	Negatif
34	2022-10-27 23:28:15+00:00	Menuntut orang lain berubah dan bertanggung jawab dgn lantang! Padahal segelintir #aremania yg memicu kejadian dgn menerobos masuk lapangan! Perbaiki diri sendiri dulu sam baru menuntut! Negara sebagian dr kalian masuk lapangan banyak orang mati!	KOeyanz	Positif
35	2022-10-27 23:26:38+00:00	Tolong pakk... Hentikan kelakuan polisi yg begini ini!! Konyol....☹	Ojoklalenya	Negatif
36	2022-10-27 23:26:23+00:00	As usual ..DivHumas_Polri bekerja setelah heboh dan viral dulu di jagat medsos seperti yg pernah disampaikan Jendral ListyoSigitP ...	Great22630501	Negatif
37	2022-10-27 23:25:40+00:00	Ya itu masih berlaku .tajam kebawah tumpul ke atas. Pejabat kita kadang klo ngomong seakan akan tegas .tegas cuma ngomong doang. Seperti pak ListyoSigitP dll . Prakteknya cuma angin lalu	_Goentoer_	Negatif

No	Dateline	text	Username	label
38	2022-10-27 23:24:52+00:00	BRAVO POLRI & TNI,...terdepan pelaksana kegiatan VAKSINASI...RAKYAT SEHAT...EKONOMI TUMBUH PESAT..□□	Dargomb33306384	Positif
39	2022-10-27 23:24:08+00:00	Tajam ke bawah Tumpul ke atas. Hukum seperti dagelan....	_riverheaven	Negatif
40	2022-10-27 23:23:59+00:00	pa tolong di perbaiki sistem uji test pembuatan sim nya karena sistem yang sekarang sangat menyusahkan masyarakat	diyan04	Negatif
41	2022-10-27 23:23:18+00:00	Itulah kelemahan penegakan hukum di negeri kita ini. Aparat tidak mendukung kerja keras Presiden untuk menegakkan hukum yg adil. Hukum masih terbang pilih. Hukum masih pilih kasih. Hukum masih mengutamakan duit. Penegak hukum kita tegas hanya pada yg lemah.	_Goentoer_	Negatif
42	2022-10-27 23:23:00+00:00	Berharap kepada aparat, tapi aparat tak dapat diharapkan□	kingpulangpagi	Negatif
43	2022-10-27 23:21:51+00:00	Ditangkap kemudian dilepas setelah bersepakat yaaaa percuma....	_riverheaven	Negatif
44	2022-10-27 23:21:48+00:00	Faktanya yang banyak membunuh rakyat dan korup itu aparat ,kok	RomziTrilaksana	Negatif
45	2022-10-27 23:21:44+00:00	Harus yg bikin fitnah dan hoax di usut dan di proses. Tidak di biarkan di lepas. Karena fitnah lebih kejam dari pembunuhan maka polisi harus menangkap siapapun yg menyebarkan fitnah.	_Goentoer_	Positif

No	Dateline	text	Username	label
46	2022-10-27 23:21:22+00:00	Maaf kami Takut dibalikin pencemaran nama baik. Kami cuma bercanda saja kok.	PaimoMontok	Positif
47	2022-10-27 23:20:50+00:00	Pak Polisi, tolong mereka harus dihukum seberat2nya...utk teman2 Bang Napi waktu dan tempat agar dipersiapkan...	suryoirawan	Positif
48	2022-10-27 23:20:28+00:00	Sungguh aneh ini kasus pidana tapi enggak kesentuh pidananya hingga sdh sampai THN:2022....pd kemana nih penegak hukum dinegri ini???	DofieeeeDofi	Negatif
49	2022-10-27 23:19:35+00:00	Sbenarnya hukum rajam sampe mati bs brlaku sih kyknya buat pelaku kek gini.. Kalo hukum ala d negara kita, sbnrnya kasian korban.. Mo pelaku pd tobat jungkir balik juga traumanya bakalan susah hilang.. Mnding klo beneran pd berubah.. Ato penjara slamanya tnpa remisi..	Ahmad74Syahid	Positif
50	2022-10-27 23:19:33+00:00	Mantap Pak Polisi	The4_9Institute	Positif
51	2022-10-27 23:19:17+00:00	Semakin terdepan kepercayaan Rakyat kepada Polri, selamat bekerja Pak Polisi yg kami cintai. KPK mana ya ?	NaysaPutri14	Positif
52	2022-10-27 23:18:23+00:00	Hahaha banyak bicara kaya jagoan.. Diciduk polisi juga dijamin meweeekkk□□□□□	pongoh_andre	Positif
53	2022-10-27 23:18:03+00:00	berhutang apa bangsa ini pd KOMISARIS PELNI ini hingga kalian biarkan ybs seenaknya menista agama mayoritas di sini?	otomoyotoM	Negatif

No	Dateline	text	Username	label
54	2022-10-27 23:17:53+00:00	Semoga segera ditangkap	SautSitinjak4	Positif
55	2022-10-27 23:17:12+00:00	Dikasih apem, entar 2minggu udah bebas. Liat aja	aba_faqih	Negatif
56	2022-10-27 23:16:10+00:00	Bejad..!!! Entah bgm perasaan ayahnya melihat anaknya menjadi korban Mohon pendampingan trhdp korban dan para pelaku dijerat hukuman yg seberat2nya.	BangMaestro	Positif
57	2022-10-27 23:11:16+00:00	Semoga #BuzzeRpLaknat seperti dia bisa mendapatkan haknya untuk diadili setelah rezim ini tumbang. Cc	imbangimedia	Positif
58	2022-10-27 23:09:59+00:00	masa gini sih Pantesan pada protes masyarakat pak. Pantesan ada tagar #percumalaporpolisi	bobcatid	Negatif
59	2022-10-27 23:08:10+00:00	Kira2 Pasca lengser ini anus di pemprov DKI ada praktek kelebihan Bayar (korupsi) pd di tangkapin gak ya ??? untung hukum korupsi gak berlaku di jakarta, berlaku hanya daerah lain.	Jebret2020	Positif
60	2022-10-27 23:05:52+00:00	Langsung tindas saja pak ..sudah banyak warga masyarakat yg tertindas karena hal tersebut	kulit_bundar_	Positif

No	Dateline	text	Username	label
61	2022-10-27 23:04:25+00:00	polisi hrs serius jika ada laporan kejahatan seksual! apalagi yg pake modus pesta tahun baru dsj biasanya korban dikasih miras/narkoba diperkosa saat tak sadar diri kmudian diancam utk penuhi keinginan pelaku!	AikHoki	Positif
62	2022-10-27 23:04:03+00:00	Setahu saya...cadar nya biasanya di ikat di luar jilbab.. Sangat jarang ada mengikat cadar di dalamkrn pasti menyulitkan saat memakai dan menyulitkan saat ingin membuka ato makan minum.	Ahmad41400773	Positif
63	2022-10-27 23:03:36+00:00	Saya tidak percaya lapor Polisi lagi. Banyak boongnya.proses kasus harus nunggu lama	Chengho_maulana	Negatif
64	2022-10-27 23:03:16+00:00	Seharusnya orang sprti ini hrs ditindak tegas. Saya mohon sama pak jokowi sbg Panglima tertinggi di Republik ini agar menginstruksikan kpd seluruh aparat polri maupun tni, utk brtindak tegas kpd orang2 yg mrusak tatanan Demokrasi di Republik ini. Tangkap mereka.	ToLumampa89	Positif
65	2022-10-27 23:03:06+00:00	Nnti tertangkap minta maaf ujung ujungnya wkwkwkwkwk ...!!! Galak di medsos doang coba di hadapan pak presiden Jokowi Lo .. kek .ayam sayur nyalinya	MiftaArifrahma6	Positif

No	Dateline	text	Username	label
66	2022-10-27 23:02:05+00:00	Kabareskrim Sebut Penolakan Pembangunan Gereja di Karimun Terkait Pilkada. Kami dukung Polri utk menindak provokator Intoleransi, krn bibit ini sgt berbahaya memecah belah sesama anak bangsa indonesia. Belajar dari Pilkada DKI yg brutal.	yvonetandiarak	Positif
67	2022-10-27 23:01:49+00:00	Yah polisinya salah..tapi radikalisme dan terorisme ada kan? Mau nutup nutupi?	dandanillalimbo	Negatif
68	2022-10-27 23:01:08+00:00	Pak dari pada kami cari tempat ngopi bareng dengan beliau ini. Saran saya temani dulu pak. Bahaya juga kalau seluruh BATAK menemani dia'kan	petruspasaribuu	Positif
69	2022-10-27 23:00:36+00:00	segera menampakan wajah pelaku. Selamat bekerja #BravoPolri	MasOjie23	Positif
70	2022-10-27 23:00:14+00:00	ujian sim kek ujian jd pemain sirkus. mohon dipertimbangkan ulang ujian praktek tersebut. Saya gak puas pak.	bayugirisuta	Negatif
71	2022-10-27 22:59:42+00:00	Memang jika anggota yang memiliki integritas tinggi pasti akan melakukan tugasnya dengan baik sesuai dengan perintah dari institusi polri	putrinabila220	Positif
72	2022-10-27 22:57:44+00:00	Lebih bahaya golongan elu dgn selalu menjual agama dan SARA! Mereka mau makar terhadp pemerintahan yg sah! Harusnya golongan elu orang yg ditangkap!!	JamesChimo	Positif

No	Dateline	text	Username	label
73	2022-10-27 22:57:44+00:00	jangan tambah goblok dong wahai bapak2 polisi . ini masuknya dah terlalu privasi. jangan idiot deh, otaknya belum malfungsi efek rokoan ama duit pungli kan? Hadeeeeh	winseoblue	Negatif
74	2022-10-27 22:57:44+00:00	Mungkin masalahnya kepolisian belum bisa bertindak jika tidak ada laporan dari masyarakat yg diculik. Pasalnya dalam mengusut sebuah peristiwa pelanggaran pidana harus ada pihak yang merasa dirugikan atas peristiwa itu...□□□	1254Abp	Positif
75	2022-10-27 22:57:22+00:00	Sebuah terobosan yang sangat tepat yang sudah diambil polri hal ini akan meminimalisir pungli oknum polisi pada saat tilang manual karena sistemnya sudah diganti dengan cara e-tilang.	putrinabila220	Positif
76	2022-10-27 22:57:08+00:00	Sebuah kebijakan yang sangat tepat yang sudah diambil polri dengan adanya e-tilang akan mengatasi adanya pungli oknum polisi pada saat tilang manual.	AlyaSalimah	Negatif
77	2020-02-28 07:02:31+00:00	Ini terjadi masih di bawah negara berlandaskan Pancasila lho ya. Kebayang ga kalau negara beneran di ubah jadi khilafah seperti yg mereka-mereka inginkan??	BlitzSee	Negatif
78	2022-10-27 22:53:17+00:00	coba sekali sekali denger keluhan tentang test yang absurd	memosntn	Negatif

No	Dateline	text	Username	label
79	2022-10-27 22:51:42+00:00	Sepakat mba puanmaharani_ri apa yang sudah katakan oleh pak jokowi ini harus diingat dan dijalankan dengan baik oleh polri.	putrinabila220	Positif
80	2022-10-27 22:50:16+00:00	Klo pelanggar lainnya non muslim gimana? Lalu apa hubungannya baca quran sama pelanggaran lalin?	OlaAdonara	Positif
81	2022-10-27 22:48:17+00:00	LOVE_AG4EVER___ Terlalu sadis makhluk ini mghina presiden NKRI	Annessautma	Negatif
82	2022-10-27 22:47:27+00:00	KELANJUTANNYA, HANYA 1 DARI 3 KASUS YG DITINDAKLANJUTI dan sepi dari pemberitaan. Juga tidak diketahui seperti apa kelanjutannya	dawiguna	Negatif
83	2022-10-27 22:46:40+00:00	ada kasus di Palopo thn 2021, bahkan WARTAWAN yg seharusnya dilindungi oleh UU Pers nyatanya dihukum menurut UU ITE, KARENA MELAPORKAN TINDAK KEJAHATAN KORUPSI entah di mana rasa keadilan Hakim dan ketaatan pada UU Pers. Terus koruptornya???	dawiguna	Negatif
84	2022-10-27 22:46:22+00:00	mohon maaf, moral kami, biar kami yg urus sendiri. Tidak usah buat peraturan yg merugikan orang yg usaha dari penyewaan tempat tinggal sementara.	DeasRamaRambo	Negatif

No	Dateline	text	Username	label
85	2022-10-27 22:44:55+00:00	Liat aja bnyk kasus,setelh memviralkn jln atau apa,ujung2nya rakyat minta maaf,aku curiga rakyat itu di takut2i uu pencemarn nama baik	Verraagustian_	Negatif
86	2022-10-27 22:44:20+00:00	UU ITE secara efektif MENCEGAH RAKYAT MENYIARKAN TINDAK KEJAHATAN yg dilakukan oleh pejabat, oleh anak pejabat, oleh cucu pejabat DAN OLEH DEMIKIAN, KEJAHATAN DAPAT TERUS BERLANGSUNG tapi UU kayak begini, nampaknya, tdk akan diubah oleh DPR sebaliknya, bye2 UU Perampasan aset	dawiguna	Positif
87	2022-10-27 22:41:46+00:00	Disinilah dikit2 pencemaran nama baik,pasl ini berbhya skli,kita uplod pejabt bis kena pencemarn nama baik	Verraagustian_	Positif
88	2022-10-27 22:41:35+00:00	Penghinaan yang luar biasa terhadap Kepala Negara. Segera tangkap dan diproses hukum	Bayak_vanLingga	Positif
89	2022-10-27 22:40:52+00:00	ADA YG DIDUGA MELAKUKAN TINDAK PENYIKSAAN itu yg dipublikasi, lantas atas UU ITE, Nikita ditangkap walaupun ada perintah Kapolri dlm kasus UU ITE ok lah Yg diduga lakukan tindak penyiksaan, apakah BEBAS??	dawiguna	Positif
90	2022-10-27 22:36:35+00:00	Tangkap pa polisi orang yg seperti ini	BuddzzzB	Positif

No	Dateline	text	Username	label
91	2022-10-27 22:36:12+00:00	Gimna drun ko di cabut,stelh bikin heboh di cabut,shrnnya orng2 yg bikin gaduh jga di serok pa	Verraagustian_	Positif
92	2022-10-27 22:34:26+00:00	Terima kasih ata responnya pak. Ditunggu proses selanjutnya	SupriyanY	Positif
93	2022-10-27 22:34:15+00:00	Bareskrim Polri Lakukan Pemeriksaan terhadap Korban Penipuan Investasi PT Rifan Financindo #investasi #investasibodong #rifanfinancindoberjangka #penipuan #berita #beritaviral	parboaboa	Positif
94	2022-10-27 22:33:32+00:00	Bener banget. Kalau ada e-tilang itu bakal nemberantas pungli. Aku setuju dg terobosan yg dilakukan Polri ini.	wiiwiie11	Positif
95	2022-10-27 22:32:59+00:00	Mantap bpak Kapolri, tegas dan jgn main2 lgi dgn cara2 lama... BRAVO POLRI KU... 🇮🇩🇮🇩	sonyvanambonese	Positif
96	2022-10-27 22:31:41+00:00	Tolong dong pak, pengawalan pejabat masih banyak brisik pakai sirine...	NovaArdiann86	Negatif
97	2022-10-27 22:31:27+00:00	ga percaya masih ada satu dua aparat hukum yg punya nurani. Ibarat mutiara ditumpukan lumpur. Barakallah.	summa_isme	Positif
98	2022-10-27 22:31:23+00:00	Tuduhan seriusss ini, harus dimintai keterangannya...tidak ada hubungan orang2 yg disebutkan di atas dengan teroris dan radikalisme. Mereka tidk niat ganti ideologi Negara, tidak teror bom bunuh diri.	Konrad46253015	Positif
99	2020-02-28 17:10:41+00:00	mohon atensinya pak bersihkan organisasi dari paham khilafah anti	ronindrawan	Positif

No	Dateline	text	Username	label
		pancasila		
100	2022-10-27 22:29:06+00:00	..nangis nggak ya kira2 kalau dikerangkeng ??? ...yg begini nih pantasnya mmng tidak dibebas liarkan..	HerryYuwono6	Positif
101	2022-10-27 22:29:02+00:00	Terrorist atau tololis?! Rakyat dah pada Pinter.Tuh Polisi pegang Pistol nggak pake sarung tangan,pistol mainan anak ² dpt beli dimana tuh?! Sekalian senjata Tajam berkarat yg udh dipamerin di KM50 juga keluarin semua□□□	DjuremiReinke	Negatif
102	2022-10-27 22:27:45+00:00	Sarana Cukuran Tersedia, Kasi Propam : Terima Kasih Kami Kepada pak Kapolres Tana Toraja	MardjuniRamlah	Positif
103	2022-10-27 22:27:03+00:00	tolong agar tes uji sim nya di evaluasi ulang karena telalu menyulitkan bagi masyarakat	diyan04	Negatif
104	2022-10-27 22:25:03+00:00	harusnya polri tegas lah dengan berita begini.	jokower_sejati	Positif
105	2022-10-27 22:24:16+00:00	Tuhkan. Polisi itu selalu mengayomi masyarakat kok, turun ke masyarakat dan bantu2 di masyarakat. Polisi bukan lembaga yg menakutkan. Ya ngga ?	mawar12z	Positif
106	2022-10-27 22:23:26+00:00	Itu lintasan apa jalur belok MotoGP sih munyer" belok g karuan gitu. Mana lulus 5x kalo aku yg Ndak omesh jalan sempit □□	lig_1066	Negatif

No	Dateline	text	Username	label
107	2022-10-27 22:22:41+00:00	Coba Buzzer GOBLOK ini tangkap sudah membuat Gaduh Video lama di muat lagi ini sengaja buat adu domba antar umat beragama.	kaconk_bar	Positif
108	2022-10-27 22:22:39+00:00	Berbenah dan evaluasi harus dilakukan . Ini untuk kebaikan polri jg kedepannya. Jgn sampai kejadian yg udah2 terjadi lagi.Terima Kasih	mawar12z	Positif
109	2022-10-27 22:21:00+00:00	Bahas apa? POLRI ITU SUDH PARAAH DRI DULU DNA DIBIARKAN, DRI POLSEK,POLRES,POLDA MABES ISINYA ORG2 SALAH DIDIK SEMUA	mouchethung	Negatif
110	2022-10-27 22:20:22+00:00	Pak hoax kaya gini jgn dibiarkan,memecah belah kerukunan masyarakat.	iroel_san	Positif
111	2022-10-27 22:20:05+00:00	Sandiwara? coretan itu bukti nyata emang begitu adanya dr Polsek polres Polda sampai pusat	mouchethung	Positif
112	2022-10-27 22:18:11+00:00	Mesti ada klarifikasi dari DPR, pemerintah pusat dan dari orang ini juga, ttg pernyataannya dlm video ini, agar rakyat tidak bingung dan saling mencurigai. Ditunggu klarifikasi terbuka dari pihak2 terkait. Terimakasih ☺. Salam #IndonesiaDamai.	Bentir16	Positif
113	2022-10-27 22:18:10+00:00	Masih ada aja yang bilang polri jelek, padahal yang ngelakuin kejahatan hanya beberapa. Penting banget si untuk tidak menyamaratakan orang dengan yang lainnya.	ayunurayu3	Positif

No	Dateline	text	Username	label
114	2022-10-27 22:16:55+00:00	Iptu Witrionaldi Memimpin Ops Kepolisian Cipta Kondusif dan ops Kejahatan Jalanan Malam Hari guna Mengantisipasi pelaku tawuran, Begal dan 3C di jalan raya pasar Babelan Desa Babelan Kota, Kamis (27/10/2022)	PolsekBabelan	Positif
115	2022-10-27 22:14:53+00:00	berbenah dan evaluasi adalah upaya yg harus dilakukan polri saat ini	kstmlm	Positif
116	2022-10-27 22:11:31+00:00	Terima kasih atas responnya pak. semoga kasusnya cepat selesai	Sugy55722869	Positif
117	2022-10-27 22:06:59+00:00	Bikin haru kontennya, semua aparat aku harap bisa kyak gini setelah reformasi kultural yg dicanangkan pak dan pak jokowi bisa direalisasikan #SiapKawalPolri	PujiAst61751000	Positif
118	2022-10-27 22:05:47+00:00	@ekagumilars @permadiaktivis @jokowi @DivHumas_Polri Laporan sampe mulut dower dan kaki gempor pun gak bakalan di proses polisi.. Anggap itu rombongan topeng monyet..	noerzam92	Negatif
119	2022-10-27 22:04:29+00:00	Hal ini bermula ketika adanya kerusakan mobil patwal. Silakan usut tuntas Bismillah rapatkan barisan OngisnadeNet tetap tegar dulur-dulur malang #MafiaWasit	Cococin1	Positif
120	2022-10-27 22:04:20+00:00	siapkawalpolri ListyoSigitP integritas sebuah institusi sangat penting, sebab ia menjadi faktor dipercaya atau tidaknya oleh masyarakat ListyoSigitP	kstmlm	Positif

No	Dateline	text	Username	label
121	2022-10-27 22:02:55+00:00	Ini anak sudah kurang ajar Marwah Presiden jokowi seenaknya dia samakan dengan Tukan Obat pinggir jalan... Tindak...□	anggorosabdo	Positif
122	2020-12-28 23:24:12+00:00	Foto satu peristiwa baik, namun lebih baik tunjukkan dgn peraturan yg melayani dan memudahkan seluruh rakyat Indonesia	gunardi_saja	Positif

CODE PROGRAM

```
import streamlit as st
from PIL import Image
import pandas as pd
import streamlit as st
from PIL import Image
import pandas as pd
import streamlit as st
from PIL import Image
import pandas as pd
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory
import StopWordRemoverFactory
from nltk.tokenize import RegexpTokenizer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
vectorizer = TfidfVectorizer(max_features=2500)
model_g = GaussianNB()
st.set_page_config(
    page_title="ANALISIS SENTIMEN",
    page_icon="🐼",
    layout="wide",
    initial_sidebar_state="expanded",
    menu_items={
        'Get Help': 'https://www.extremelycoolapp.com/help',
        'Report a bug': "https://www.extremelycoolapp.com/bug",
        'About': "# This is a header. This is an *extremely* cool app!"
    }
)
```

```

)
st.header('ANALISI SENTIMEN PADA TWEETS DIVISI HUMAS POLRI')
st.image ('image/logomhp.png', width=500)
st.sidebar.image ('image/analisis.png', width=280, )
excel_file = 'datapolri.xlsx'
sheet_name = 'Sheet1'
df = pd.read_excel(excel_file,
                    sheet_name=sheet_name,
                    header=0)
jumlahdata = df.No.count()
dataset = {}
sentimen = {}
text = []
label = []
for i in range(0, jumlahdata):
    dataset[df.No[i]] = df.text[i]
    sentimen[df.No[i]] = df.label[i]
    label.append(df.label[i])
    text.append(df.text[i])
for i in range(0, jumlahdata):
    if (df.label[i] == 0): #Negatif
        st.sidebar.error(df.text[i], icon='❌',)
    if (df.label[i] == 1): #Positif
        st.sidebar.info(df.text[i], icon="ℹ️")
    # else:
    #     st.info(df.Tweet[i],icon="ℹ️")
excel_file = 'dataset_polri.xlsx'
sheet_name = 'datapolri'
df = pd.read_excel(excel_file,
                    sheet_name=sheet_name,

```

```

        header=0)
jumlahdata = df.No.count()
dataset = {}
sentimen = {}
text = []
label = []
No = []
for i in range(0, jumlahdata):
    dataset[df.No[i]] = df.text[i]
    sentimen[df.No[i]] = df.label[i]
    label.append(df.label[i])
    text.append(df.text[i])
    No.append(df.No[i])
dnb = pd.DataFrame (columns=['text','label'])
dnb ['label'] = label
dnb ['text'] = text

tab1, tab2, tab3 = st.tabs(["Data Sentimen", "Pemodelan", "Implementasi
Model"])
with tab1:
    st.header("Data Sentimen")
    dnb = pd.DataFrame(columns=['text', 'label'])
    dnb['label'] = label
    dnb['text'] = text
    st.write(dnb)

with tab2:
    st.header("Pemodelan")
    options_form = st.form("options")
    preprocessing = options_form.form_submit_button('Preprocessing')
    pembobotan = options_form.form_submit_button('Pembobotan')

```

```

t_size = options_form.radio("Pilih pembagian splitting data",
                             (0.1,0.2,0.3))
klasifikasi = options_form.form_submit_button('Klasifikasi')
evaluasi = options_form.form_submit_button('Evaluasi')
if preprocessing:
    st.success("1. Preprocessing")
# Case Folding
    for k in dataset.keys():
        hasil_case_folding = dataset[k].lower()
        dataset[k] = hasil_case_folding
    st.write("Case Folding:", dataset)
# Tokenizing
    tokenizer = RegexpTokenizer(r"\w+")
    for k in dataset.keys():
        tokens = tokenizer.tokenize(dataset[k])
        dataset[k] = tokens
    st.write("Tokenisasi", dataset)
# Number Removal
    for k in dataset.keys():
        tokens = []
        for t in dataset[k]:
            if t.isnumeric() == False:
                tokens.append(t)
        dataset[k] = tokens
# Stopword Removal by Sastrawi
    factory = StopWordRemoverFactory()
    stopword_list = factory.get_stop_words()
    for k in dataset.keys():
        tokens = []
        for t in dataset[k]:
            if t not in stopword_list:

```

```

        tokens.append(t)
    dataset[k] = tokens
    st.write("Stopword Removing", dataset)
# Stemming by Sastrawi
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    for k in dataset.keys():
        tokens = []
        for t in dataset[k]:
            tokens.append(stemmer.stem(t))
        dataset[k] = tokens
# Mengembalikan Format Dataset Awal
    for k in dataset.keys():
        dataset[k] = " ".join(dataset[k])
    st.write("stemming", dataset)
if pembobotan:
    st.success("2. Pembobotan Dokumen (TF-IDF)")
    #Frekuensi Kemunculan Kata
    tf = CountVectorizer()
    term_doc_matrix = tf.fit_transform(dataset.values())
    TAMPIL=pd.DataFrame(term_doc_matrix.toarray(), index=dataset.keys(),
columns=tf.get_feature_names_out())
    st.write(TAMPIL)
    # Pembobotan TF-IDF
    vectorizer = TfidfVectorizer(max_features=2500)
    model_g = GaussianNB()
    v_data = vectorizer.fit_transform(dataset.values()).toarray()
    #st.write(v_data)
if klasifikasi:
    st.write("Size Splitting :", t_size)
    st.write("Kernel GaussianNB")

```

```

#vectorizer = TfidfVectorizer(max_features=2500)
from sklearn.naive_bayes import GaussianNB
model_g = GaussianNB()
from sklearn.model_selection import train_test_split
v_data = vectorizer.fit_transform(dataset.values()).toarray()
    X_train, X_test, y_train, y_test = train_test_split(v_data, dnb['label'],
test_size=t_size, random_state=0)
    nbmodel = model_g.fit(X_train, y_train)
    nbmodel.class_count_
        from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report
    y_preds = nbmodel.predict(X_test)
    predict = pd.Series(y_preds)
    y_true = pd.Series(y_test)
    st.write("Data Testing Aktual")
    st.write(y_true)
    st.write("Data Testing Yg Diprediksi")
    st.write(predict)
    proba = nbmodel.predict_proba(X_test)
    st.write("Nilai Probabilitas")
    st.write(proba)
    st.write('nilai akurasi adalah ',
accuracy_score(y_test, y_preds)*100)
if evaluasi:
    st.success("3. Evaluasi")
    v_data = vectorizer.fit_transform(dataset.values()).toarray()
        X_train, X_test, y_train, y_test = train_test_split(v_data, dnb['label'],
test_size=t_size, random_state=0)
        model_g.fit(X_train, y_train)
        y_preds = model_g.predict(X_test)
        from sklearn.metrics import classification_report, confusion_matrix

```

```

        st.write(confusion_matrix(y_test, y_preds))
        st.write(classification_report(y_test, y_preds))
with tab3:
    st.header("Masukan data Uji baru")
    options_form = st.form("options_form")
    txt = options_form.text_area("", height=5, placeholder='Masukkan Tweet Baru..!')
    proses = options_form.form_submit_button()
    if proses:
        for k in dataset.keys():
            hasil_case_folding = dataset[k].lower()
            dataset[k] = hasil_case_folding
        # Tokenizing
        tokenizer = RegexpTokenizer(r"\w+")
        for k in dataset.keys():
            tokens = tokenizer.tokenize(dataset[k])
            dataset[k] = tokens
        # Number Removal
        for k in dataset.keys():
            tokens = []
            for t in dataset[k]:
                if t.isnumeric() == False:
                    tokens.append(t)
            dataset[k] = tokens
        # Stopword Removal by Sastrawi
        factory = StopWordRemoverFactory()
        stopword_list = factory.get_stop_words()
        for k in dataset.keys():
            tokens = []
            for t in dataset[k]:
                if t not in stopword_list:

```

```

        tokens.append(t)
    dataset[k] = tokens
# Stemming by Sastrawi
factory = StemmerFactory()
stemmer = factory.create_stemmer()
for k in dataset.keys():
    tokens = []
    for t in dataset[k]:
        tokens.append(stemmer.stem(t))
    dataset[k] = tokens
# Mengembalikan Format Dataset Awal
for k in dataset.keys():
    dataset[k] = " ".join(dataset[k])
# Frekuensi Kemunculan Kata
tf = CountVectorizer()
term_doc_matrix = tf.fit_transform(dataset.values())
pd.DataFrame(term_doc_matrix.toarray(), index=dataset.keys(),
columns=tf.get_feature_names_out())
# Pembobotan TF.IDF
vectorizer = TfidfVectorizer(max_features=2500)
model_g = GaussianNB()
v_data = vectorizer.fit_transform(dataset.values()).toarray()
X_train, X_test, y_train, y_test = train_test_split(v_data, dnb['label'],
test_size=t_size, random_state=0)
model_g.fit(X_train, y_train)
y_preds = model_g.predict(X_test)
tweet = txt
v_data = vectorizer.transform([tweet]).toarray()
y_predss = model_g.predict(v_data)
st.write(y_predss)
dnb = pd.DataFrame(columns=['No', 'text', 'label'])

```

```

dnb['label'] = label
dnb['text'] = text
dnb['No'] = No
for i in range (0, jumlahdata):
    if y_predss == "Positif":
        coba = "Positif"
        new_tweet = {"No": str(i+2), "text": txt, "label": coba }
        df = dnb.append(new_tweet, ignore_index=True)
    else:
        coba = "Negatif"
        new_tweet = {"No": str(i+2), "text": txt, "label": coba }
        df = dnb.append(new_tweet, ignore_index=True)
df.to_excel('dataset_polri.xlsx',
            sheet_name='datapolri', index=False)
# dengan asumsi bahwa 1 merupakan label positif
if y_predss == "Positif":
    st.success('[Positif] '+txt, icon='✅')
else:
    st.error('[Negatif] '+txt, icon='❌')

```



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
LEMBAGA PENELITIAN

Kampus Unisan Gorontalo Lt.3 - Jln. Achmad Nadjamuddin No. 17 Kota Gorontalo
Telp: (0435) 8724466, 829975 E-Mail: lembagapenelitian@unisan.ac.id

Nomor : 4476/PIP/LEMLIT-UNISAN/GTO/XII/2022

Lampiran : -

Hal : Permohonan Izin Penelitian

Kepada Yth,

Kapolda Gorontalo

di,-

Tempat

Yang bertanda tangan di bawah ini :

Nama : Dr. Rahmisyari, ST.,SE.,MM

NIDN : 0929117202

Jabatan : Ketua Lembaga Penelitian

Meminta kesediannya untuk memberikan izin pengambilan data dalam rangka penyusunan **Proposal / Skripsi**, kepada :

Nama Mahasiswa : Caldiyastopan Mohi

NIM : T3119095

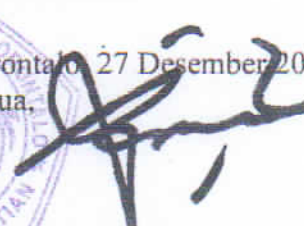
Fakultas : Fakultas Ilmu Komputer

Program Studi : Teknik Informatika

Lokasi Penelitian : POLDA GORONTALO

Judul Penelitian : KLASIFIKASI TOPIK BERITA DAN ANALISIS SENTIMEN
PADA TWEETS DIVISI HUMAS POLRI DENGAN METODE
NAIVE BAYES CLASSIFIER

Atas kebijakan dan kerja samanya diucapkan banyak terima kasih.

Gorontalo, 27 Desember 2022
Ketua,

Dr. Rahmisyari, ST.,SE.,MM
NIDN 0929117202

+



SURAT KETERANGAN
Nomor : SKET / 02 / V / 2023

Yang bertanda tangan di bawah ini :

Nama : DESMONT HARJENDRO A.P., S.I.K., M.T.
Pangkat / NRP : AKBP / 78120909
Jabatan : KABID HUMAS POLDA GORONTALO
Kesatuan : POLDA GORONTALO

Dengan ini menerangkan bahwa :

Nama Mahasiswa : CALDIYASTOVAN MOHI
NIM : T3119095
Program Studi : TEKNIK INFORMATIKA
Fakultas : FAKULTAS ILMU KOMPUTER
Universitas : UNIVERSITAS ICHSAN GORONTALO

Telah selesai melaksanakan Penelitian Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo dengan Judul **"ANALISIS SENTIMEN PADA TWEETS DIVISI HUMAS POLRI DENGAN METODE NAIVE BAYES CLASSIFIER"** di Polda Gorontalo bertempat di Bidhumas Polda Gorontalo.

Demikian Surat Keterangan ini dibuat dengan sebenar-benarnya untuk dapat dipergunakan seperlunya.

Gorontalo, 08 Mei 2023



a.n. KEPALA KEPOLISIAN DAERAH GORONTALO
KABID HUMAS

DESMONT HARJENDRO A.P., S.I.K., M.T.
AJUN KOMISARIS BESAR POLISI NRP 78120909



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO**

FAKULTAS ILMU KOMPUTER

UPT. PERPUSTAKAAN FAKULTAS

SK. MENDIKNAS RI NO. 84/D/0/2001

Jl. Achmad Nadjamuddin No.17 Telp(0435) 829975 Fax. (0435) 829976 Gorontalo

SURAT KETERANGAN BEBAS PUSTAKA

No : 010/Perpustakaan-Fikom/V/2023

Perpustakaan Fakultas Ilmu Komputer (FIKOM) Universitas Ichsan Gorontalo dengan ini menerangkan bahwa :

Nama Anggota : Caldiyastovan Mohi

No. Induk : T3119095

No. Anggota : M202338

Terhitung mulai hari, tanggal : Selasa, 09 Mei 2023, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di pergunakan sebagaimana mestinya.



Gorontalo, 09 Mei 2023

**Mengetahui,
Kepala Perpustakaan**

Apriyanto Alhamad, M.Kom

NIDN : 0924048601



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

UNIVERSITAS ICHSAN GORONTALO

FAKULTAS ILMU KOMPUTER

SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001

Jl. Achmad Najamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT REKOMENDASI BEBAS PLAGIASI

No. 145/FIKOM-UIG/R/V/2023

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom
NIDN : 0928028101
Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

Nama Mahasiswa : Caldiyastovan Mohi
NIM : T3119095
Program Studi : Teknik Informatika (S1)
Fakultas : Fakultas Ilmu Komputer
Judul Skripsi : Analisis Sentimen Pada Tweets Divisi Humas Polri
Dengan Metode Naive Bayes Classifier

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi **Turnitin** untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar **12%**, berdasarkan Peraturan Rektor No. 32 Tahun 2019 tentang Pendeteksian Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ichsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan **BEBAS PLAGIASI** dan layak untuk diujikan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.



Irvan Abraham Salihi, M.Kom
NIDN. 0928028101

Gorontalo, 10 Mei 2023
Tim Verifikasi,

Zulfrianto Y. Lamasigi, M.Kom
NIDN. 0914089101

Terlampir :
Hasil Pengecekan Turnitin

PAPER NAME

Skripsi_T3119095_Caldiyastovan Mohi.d
ocx

AUTHOR

T3119095-CALDIYASTOVAN MOHI caldi
mohi5@gmail.com

WORD COUNT

15357 Words

CHARACTER COUNT

70127 Characters

PAGE COUNT

102 Pages

FILE SIZE

3.7MB

SUBMISSION DATE

May 9, 2023 7:34 PM GMT+8

REPORT DATE

May 9, 2023 7:35 PM GMT+8

● 12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 12% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 0% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less than 10 words)

12% Overall Similarity

Top sources found in the following databases:

- 12% Internet database
- 3% Publications database
- Crossref database
- Crossref Posted Content database
- 0% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	repository.ub.ac.id	2%
	Internet	
2	repository.uin-suska.ac.id	1%
	Internet	
3	andi.ddns.net	1%
	Internet	
4	eprints.utdi.ac.id	1%
	Internet	
5	core.ac.uk	1%
	Internet	
6	jurnal.untan.ac.id	<1%
	Internet	
7	ejurnal.teknokrat.ac.id	<1%
	Internet	
8	ejournal.amikompurwokerto.ac.id	<1%
	Internet	

9	ecampus.pelitabangsa.ac.id	Internet	<1%
10	media.neliti.com	Internet	<1%
11	jtiik.ub.ac.id	Internet	<1%
12	123dok.com	Internet	<1%
13	mwrsf.unl.edu	Internet	<1%
14	vdocuments.mx	Internet	<1%
15	eprints.radenfatah.ac.id	Internet	<1%
16	researchgate.net	Internet	<1%
17	ejurnal.seminar-id.com	Internet	<1%
18	docplayer.info	Internet	<1%
19	ejournal.um-sorong.ac.id	Internet	<1%
20	jurnal.iaii.or.id	Internet	<1%

21	eprints.umm.ac.id Internet	<1%
22	coursehero.com Internet	<1%
23	video.kompas.com Internet	<1%
24	dspace.uui.ac.id Internet	<1%
25	journal.admi.or.id Internet	<1%
26	repositori.uma.ac.id Internet	<1%
27	senafti.budiluhur.ac.id Internet	<1%
28	repository.upnjatim.ac.id Internet	<1%
29	Frandy Ristianto, Nurmalasari Nurmalasari, Ani Yoraeni. "Impementasi ... Crossref	<1%
30	repository.unimus.ac.id Internet	<1%
31	eprints.unisbank.ac.id Internet	<1%
32	etheses.uin-malang.ac.id Internet	<1%

33

id.123dok.com

Internet

<1%

34

repo.stikesicme-jbg.ac.id

Internet

<1%

Daftar Riwayat Hidup



Nama	: Caldiyastovan Mohi
Nim	: T3119095
Tempat, Tanggal Lahir	: Tolotio, 13 Mei 2000
Alamat	: Desa Lembah Hijau
Agama	: Islam
Kewarganegaraan	: WNI
Email	: caldimohi5@gmail.com

Riwayat pendidikan dan pekerjaan

1. Tahun 2013, Menyelesaikan Pendidikan di Sekolah Dasar Negeri 5 Bonepantai
2. Tahun 2016, Menyelesaikan Pendidikan di Sekolah Menengah Pertama Negeri 3 Bonepantai
3. Tahun 2019, Menyelesaikan Pendidikan di Sekolah Menengah Kejuruan Negeri 3 Gorontalo
4. Tahun 2019, Diterima menjadi mahasiswa di Perguruan Tinggi Swasta Universitas Ichsan Gorontalo