ANALISIS SENTIMEN TERKAIT *HASHTAG* KENAIKAN BBM PADA SOSIAL MEDIA TWITTER MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

PROPOSAL TUGAS AKHIR

Diajukan Oleh :

Zaneta Hanes Armairan 8020190230

Untuk Persyaratan Penelitian Dan Penulisan Tugas Akhir

Sebagai Akhir Proses Studi Strata 1

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DINAMIKA BANGSA JAMBI

2022

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal : ANALISIS SENTIMEN TERKAIT *HASHTAG* BBM NAIK PADA SOSIAL MEDIA TWITTER MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

a. Nama Lengkap : Zaneta Hanes Armairan
b. NIM : 8020190230
c. Jenis Kelamin : Laki-laki
d. Tempat/Tanggal Lahir : Jambi, 10 Februari 2002
e. Alamat : Jln. Lingkar Selatan, Perumahan Griya Paal Merah Asri, RT. 23, Kecamatan Paal Merah, Kelurahan Paal Merah, Jambi Selatan, Kota Jambi
f. No. Telepon : 081218172330
g. Email : zanetahanesarmairan.9e@gmail.com

PERNYATAAN HASIL EVALUASI

Nama/NIM : Zaneta Hanes Armairan / 8020190230

Prodi : ~~SI~~/TI/~~SK~~

Judul : ANALISIS SENTIMEN TERKAIT *HASHTAG* BBM NAIK PADA SOSIAL MEDIA TWITTER MENGGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*

Hasil Evaluasi : Disetujui / Disetujui dengan perbaikan / Ditolak *)

1. Catatan
   - ☐ Alasan penolakan proposal tugas akhir
   - ☐ Proposal tugas akhir tidak relevan dengan program studi
   - ☐ Pernah ada topik sejenis
   - ☐ Metode yang dipakai tidak jelas
   - ☐ Masalah terlalu sempit
   - ☐ ……...

Mengetahui,

Ketua Program Studi Teknik Informatika

BENY, S.KOM, M.SC

YDB.07.84.055

*) Coret yang tidak perlu

BAB I
PENDAHULUAN

1.1 LATAR BELAKANG

Salah satu permasalahan yang dialami masyarakat Indonesia saat ini, adalah tentang Bahan Bakar Minyak (BBM) yang sedang naik. Hal ini, membuat masyarakat Indonesia untuk hidup berhemat dan ada juga yang memutuskan untuk beralih ke jenis bahan bakar minyak yang lain. Misalnya, dari yang sebelumnya memakai jenis BBM Pertamax kemudian beralih ke Pertalite. Hal ini tentu menimbulkan pro dan kontra dikalangan masyarakat. Hampir seluruh masyarakat Indonesia memberikan respon negatif terhadap kebijakan kenaikan BBM, dikarenakan banyak masyarakat dari kalangan menengah ke bawah yang ekonominya terkadang sulit. Sehingga, mereka merasa berat untuk membayar biaya BBM yang begitu besar, sedangkan hidup mereka masih merasa kekurangan. Selain itu, ada juga dampak positif yang ditimbulkan dari adanya kebijakan kenaikan BBM, seperti masyarakat lebih memilih naik transporatasi umum atau *nebeng* ketimbang membawa kendaraan pribadinya[1]. Sejak ada wacana kebijakan kenaikan BBM, masyarakat banyak memberikan opini di media sosial.

Perkembangan media yang begitu pesat, memunculkan banyak media *online* dari media berita bahkan sampai media sosial. Media sosial saja sudah begitu banyak, antara lain Facebook, Twitter, Instagram, Line dan masih banyak lagi. Media sosial ini tidak hanya digunakan sebagai sarana pertemanan, tetapi sudah banyak digunakan untuk kegiatan lain, seperti media promosi, media edukasi dan media informasi bagi masyarakat untuk mengetahui dan menanggapi setiap kebijakan yang dilakukan oleh pemerintah. Salah satu media sosial, khususnya Twitter merupakan aplikasi yang marak digunakan pada kalangan usia baik tua maupun muda. Tercatat jumlah pengguna aktif harian di Twitter selama tahun 2022, yaitu sebanyak 237,8 juta[2]. Dilansir dari We Are Social (Sebuah situs web yang

menyajikan data beserta tren yang dibutuhkan dalam memahami internet, media sosial dan perilaku *e-commerce* setiap tahun secara berkala), jumlah pengguna Twitter di Indonesia mencapai 18,45 juta pada tahun 2022[2]. Pengguna Twitter dapat menyampaikan opini melalui *tweet* yang mereka *post* dan kemudian dapat dianalisis untuk melihat bagaimana sentimen yang dimunculkan terhadap setiap kebijakan yang diwacanakan pemerintah. Dari berbagai macam *tweet* yang diposting tersebutlah yang menjadi alasan untuk dipakainya analisis sentimen.

Pada penelitian ini, analisis sentimen dilakukan untuk melihat opini masyarakat yang ditujukan untuk setiap *tweet* yang terkait dengan *hashtag* kenaikan BBM, opini tersebut bisa dimasukkan kategori opini positif maupun negatif. Besarnya sentimen yang ditujukan pada *tweet* yang mengandung *hashtag* kenaikan BBM bisa dijadikan sebuah parameter bagi pemerintah untuk melihat seberapa setujunya masyarakat Indonesia terhadap kebijakan kenaikan BBM tersebut.

Penentuan polaritas positif atau negatifnya suatu opini dapat dilakukan secara manual dengan memanfaatkan disiplin ilmu bahasa atau komunikasi, tetapi seiring bertambahnya sumber opini maka semakin banyak waktu yang dibutuhkan untuk mengklasifikasinya. Besarnya polaritas *tweet* yang ditunjukkan oleh masyarakat bisa dijadikan parameter untuk melihat setuju atau tidaknya masyarakat terhadap kebijakan kenaikan BBM. Oleh karena itu, penelitian ini menggunakan fungsi *text mining* dan pembelajaran mesin untuk mengklasifikasi polaritas opini dari data *tweet* dengan menggunakan algoritma *Naïve Bayes Classifier* untuk mendapatkan hasil klasifikasi *tweet* positif maupun negatif.

1.2 PERUMUSAN MASALAH

Berdasarkan latar belakang permasalahan yang telah diuraikan diatas, maka dapat diambil rumusan yang akan menjadi pembahasan penelitian ini, yaitu : “Bagaimana cara menganalisis sentimen pada *tweet* yang terkait dengan *hashtag*

kebijakan kenaikan BBM pada media sosial Twitter dengan menggunakan algoritma *Naïve Bayes Classifier*".

1.3 BATASAN MASALAH

Agar penelitian ini terfokus dan tidak keluar dari topik pembahasan, adapun yang menjadi batasan masalah pada penelitian ini, adalah :

1. Data yang di *crawling* hanya berupa *tweet* berbahasa Indonesia dengan data sebanyak 1000 *tweet*.
2. Data yang digunakan sebagai data *training* sebanyak 250 *tweet* dan data *testing* 100 *tweet*.
3. *Keyword* yang mengandung *hashtag* yang terkait dengan kebijakan kenaikan BBM.
4. Algoritma yang digunakan dalam mengklasifikasi data adalah *Naïve Bayes Classifier* dan tidak ada perbandingan dengan algoritma lain.
5. Pengujian hasil analisis menggunakan *tools* RapidMiner.
6. *Output* yang dihasilkan dari data *tweet* diklasifikasikan menjadi opini positif dan negatif.

1.4 TUJUAN PENELITIAN

Adapun tujuan dari penelitian dan penyusunan proposal tugas akhir ini, adalah :

1. Melihat sejauh mana tingkat akurasi algoritma *Naïve Bayes Classifier* pada tahapan analisis sentimen.
2. Mengklasifikasikan nilai sentimen pada *hashtag* yang terkait dengan kebijakan kenaikan BBM secara keseluruhan dari pengguna Twitter di Indonesia berdasarkan kategori sentimen yang telah ditentukan, yaitu positif atau negatif.

1.5 MANFAAT PENELITIAN

Adapun manfaat yang didapat dari penelitian dan penyusunan proposal tugas akhir ini, adalah :

1. Dapat mengetahui informasi tentang bagaimana tanggapan masyarakat tentang kebijakan kenaikan BBM dan mengetahui kecenderungan opini masyarakat pada hasil sentimen untuk pengguna Twitter di Indonesia.
2. Bagi penulis, dapat mengembangkan pengetahuan dan wawasan yang baru bahwa terdapat data *tweet* yang dapat diolah dan dilihat nilai sentimennya.
3. Bagi pembaca, dengan adanya penelitian ini dapat digunakan sebagai referensi atau acuan untuk penelitian selanjutnya.

1.6 SISTEMATIKA PENULISAN

Adapun sistematika penulisan yang digunakan dalam menyusun penelitian ini, adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini, penulis membahas tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini, penulis membahas tentang teori-teori dan pendapat para ahli yang berhubungan langsung dengan permasalahan yang dianalisis, antara lain : pengertian *data mining*, *text mining*, analisis sentimen, *Naïve Bayes Classifier* dan *hashtag*.

BAB III : METODOLOGI PENELITIAN

Pada bab ini, menjelaskan tentang kerangka kerja penelitian, metode pengumpulan data, metode pengklasifikasian dan alat bantu yang digunakan selama mengerjakan penelitian.

BAB IV : ANALISIS

Pada bab ini akan dilakukan tahap analisis menggunakan algoritma *Naïve Bayes Classifier* dengan memanfaatkan data *tweet* yang telah dikumpulkan.

BAB V : HASIL ANALISIS DAN REKOMENDASI

Pada bab ini, menjelaskan hasil analisis dalam bentuk visualisasi dan rekomendasi berdasarkan penelitian yang telah dilakukan.

BAB VI : PENUTUP

Bab ini merupakan penutup dari penelitian ini, yang terdiri dari kesimpulan dari bab-bab sebelumnya dan saran-saran yang sekiranya dapat bermanfaat bagi pihak yang bersangkutan pada penelitian ini.

BAB II

LANDASAN TEORI

2.1 ANALISIS SENTIMEN

Analisis sentimen atau *sentiment analysis* adalah merupakan perpaduan dari *data mining* dan *text mining*, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi. *Sentiment analysis* merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentimen yang terkandung dalam sebuah opini. Pada *sentiment analysis* terdiri dari 3 jenis opini, yaitu opini positif, opini negatif dan opini netral, sehingga dengan *sentiment analysis* perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui *feedback* masyarakat atau pun para ahli.

Sentimen mengacu pada fokus topik tertentu, pernyataan pada suatu topik mungkin akan berbeda makna dengan pernyataan yang sama pada subjek yang berbeda, oleh karena itu pada beberapa penelitian, terutama pada *review* produk, perkerjaan didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses *sentiment analysis*[3].

2.2 APLIKASI

Menurut (Hasan Abdurahman dan Asep Ririh Riswaya, 2014), aplikasi adalah program siap pakai yang dapat digunakan untuk menjalankan perintah-perintah dari pengguna aplikasi tersebut dengan tujuan mendapatkan hasil yang lebih akurat sesuai dengan tujuan pembuatan aplikasi tersebut, aplikasi mempunyai arti yaitu pemecahan masalah yang menggunakan salah satu teknik pemrosesan data aplikasi

yang biasanya berpacu pada sebuah komputansi yang diinginkan atau diharapkan maupun pemrosesan data yang diharapkan[4].

Secara umum, aplikasi merupakan alat terapan yang difungsikan secara khusus dan terpadu sesuai dengan kemampuan yang dimilikinya. Aplikasi merupakan suatu perangkat komputer yang siap pakai bagi penggunanya (*user*).

2.3 MEDIA SOSIAL

Menurut Van Dijk (dalam Nasrullah, 2015) Media sosial adalah platform media yang memfokuskan pada eksistensi pengguna yang memfasilitasi mereka dalam beraktifitas maupun berkolaborasi.

Sedangkan menurut Meike dan Young (dalam Nasrullah, 2015) mengartikan kata media sosial sebagai konvergensi antara komunikasi personal dalam arti saling berbagi diantara individu (*to be share one-to-one*) dan media publik untuk berbagi kepada siapa saja tanpa ada kekhususan individu.

Kemudian menurut Boyd (dalam Nasrullah, 2015) menjelaskan bahwa media sosial sebagai kumpulan perangkat lunak yang memungkinkan individu maupun komunitas untuk berkumpul, berbagi, berkomunikasi dan dalam kasus tertentu saling berkolaborasi atau bermain[5].

Dari beberapa definisi tentang media sosial diatas, dapat disimpulkan bahwa media sosial adalah suatu bentuk interaksi atau komunikasi yang dilakukan antar seseorang (dua orang atau lebih) yang menggunakan suatu perangkat lunak melalui suatu perangkat komputer.

2.4 TWITTER

Twitter merupakan sebuah perangkat lunak media sosial yang dikembangkan untuk semua orang. Dimana pada perangkat lunak ini, seseorang bisa membuat

suatu kiriman, baik berupa foto atau video, mencari berita terhangat yang sedang dibicarakan saat ini, mencari kiriman atau *tweet* orang lain dengan menggunakan tanda pagar (*hashtag*) dan lain-lain.

Bagi pengguna, Twitter tidak hanya sekedar untuk mengirim suatu foto atau video saja, tetapi juga dapat menemukan banyak komunitas yang bisa *sharing* informasi secara *real time*. Mulai dari yang membicarakan teknologi, olahraga, kesehatan, cinta dan masih banyak lagi dan juga untuk membagikan (*sharing*) foto, link ke *website* atau blog, serta ruangan untuk menelurkan ide-ide penggunanya. Fasilitas ini setidaknya bisa menjadi awal mula bagi pengguna untuk menciptakan *personal brand* sendiri di dunia digital[6].

2.5 *HASHTAG*

Hashtag adalah suatu frasa kata kunci yang dieja tanpa menggunakan spasi, dengan tanda tagar di bagian depannya, *Hashtag* ini nantinya akan mengikat percakapan publik dari seluruh pengguna yang berbeda ke dalam suatu aliran. Anda bisa menemukannya dengan mencari kata kunci ataupun memilih hashtag, atau dengan menggunakan suatu *tools* tertentu. Agar postingan yang dibalut dengan *hashtag* ini bisa muncul di penelusuran semua orang, maka postingan dengan menggunakan *hashtag* ini harus bersifat lebih publik[7].

The Tweeting Typographic Conventions mengatakan *hashtag* atau tagar adalah serangkaian karakter yang didahului dengan simbol # (*hash*, atau *pound*), seperti pada #barcamp. Tagar tidak mengizinkan adanya spasi. Dalam tagar bila terdapat dua atau lebih kata, huruf awal dari setiap kata dapat ditulis dengan huruf besar, seperti dalam #SouthAfrica, bukan #southafrica, atau bahkan #South Afrika". Tagar dapat berisi angka tetapi tidak bisa seluruhnya terdiri dari angka. Tagar biasanya digunakan untuk mengelompokkan suatu pesan pada media sosial. Sehingga pesan yang mengandung tagar dapat dengan mudah dicari.Tanda pagar (tagar), atau dalam bahasa Inggris disebut *hashtag*, pada awalnya adalah sebuah tanda untuk menunjukkan nomor (misalnya “#1″, yang merupakan singkatan dari

“nomor satu”) atau disebut dengan tanda nomor (*number sign*). Istilah tanda nomor, saat itu hanya populer di Kanada. Di sebagian besar Amerika Serikat dikenal sebagai tanda pon (*pound sign*). Sementara di luar Amerika Utara, tanda ini disebut sebagai tanda pagar (*hash key*) yang biasa digunakan pada pesawat telepon[8].

2.6 *DATA MINING*

Menurut (Larose, 2006), definisi *data mining* adalah proses menemukan sesuatu yang bermakna oleh suatu korelasi baru, pola dan juga tren yang terdapat dengan cara memilah-milah data yang berukuran besar, dimana data tersebut disimpan dalam *repository*, menggunakan teknologi sosialisasi pola serta statistik dan teknik matematika[9].

Data mining adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting pada data. Proses pengumpulan dan ekstraksi informasi tersebut dapat dilakukan menggunakan perangkat lunak dengan bantuan perhitungan statistika, matematika, ataupun teknologi *Artificial Intelligence* (AI). Data mining sering disebut juga *Knowledge Discovery in Database* (KDD)[10].

2.7 *TEXT MINING*

Menurut (Ronen Feldman dan James Sanger dalam buku *The Text Mining Handbook*), *text mining* dapat didefinisikan secara luas sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen-komponen dalam *data mining* yang salah satunya adalah kategorisasi. *Text mining* bisa dianggap subjek riset yang tergolong baru. *Text mining* dapat memberikan solusi 14 dari permasalahan seperti pemrosesan, pengorganisasian/pengelompokkan dan menganalisa *unstructured text* dalam jumlah besar.

Dalam memberikan solusi, *text mining* mengadopsi dan mengembangkan banyak teknik dari bidang lain, seperti *Data mining*, *Information Retrieval*, Statistik dan Matematik, *Machine Learning*, *Linguistic*, *Natural Language Processing*, dan *Visualization*. Kegiatan riset untuk *text mining* antara lain ekstraksi dan

penyimpanan *text*, preprocessing akan konten *text*, pengumpulan data statistik dan *indexing* dan analisa konten[11].

2.8 *NAÏVE BAYES CLASSIFIER*

Menurut (Olson Delen, 2008) menjelaskan *Naïve Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dr "master" tabel keputusan.

Naïve Bayes Classifier bekerja sangat baik dibanding dengan *model classifier* lainnya. Hal ini dibuktikan oleh Xhemali , Hinde Stone dalam jurnalnya "*Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*" mengatakan bahwa "*Naïve Bayes Classifier* memiliki tingkat akurasi yg lebih baik dibanding model *classifier* lainnya".

Keuntungan penggunaan *Naïve Bayes Classifier* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai *variable independent*, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians[12].

2.9 RAPIDMINER

RapidMiner merupakan perangakat lunak yang bersifat terbuka (*open source*). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik.

RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk *input*, *output*, *data preprocessing* dan visualisasi. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin *data mining* yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan munggunakan bahasa java sehingga dapat bekerja di semua sistem operasi.

RapidMiner sebelumnya bernama **YALE** (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund[13].

2.10 WEKA

WEKA merupakan *tools machine learning* yang praktis. "*Waikato Environment for Knowledge Analysis*" atau yang dikenal dengan WEKA dibuat di Universitas Waikato, New Zealand, yang dikhususkan untuk menunjang bidang penelitian, pendidikan dan berbagai pengaplikasiannya dalam *data mining*. *Software* dibangun menggunakan *class Java* dengan metode berorientasi objek dan dapat dijalankan hampir pada semua *platform*, dalam penggunaanya WEKA sangat mudah diterapkan pada beberapa tingkatan yang berbeda. WEKA menyediakan implementasi algoritma pembelajaran *state of the art* yang dapat diterapkan pada *dataset* dari *command line*.

Dalam WEKA terdapat *tools* yang berguna untuk *preprocessing* data, klasifikasi, regresi, *clustering*, aturan asosiasi, dan visualisasi. Penggunaannya dapat dilakukan *preprocess* pada data, memasukkan dalam sebuah skema pembelajaran, dan menganalisis *classifier* yang dihasilkan oleh performanya, semua dikerjakan tanpa menulis kode program. Salah satu contoh penggunaan WEKA dengan menerapkan sebuah metode pembelajaran kedalam dataset serta menganalisis hasilnya untuk memperoleh informasi tentang data, atau menerapkan beberapa metode dan membandingkan performanya untuk dipilih.

Definisi lain dari WEKA adalah alat yang dapat digunakan untuk *preprocessing dataset* dan membuat *user* dapat terfokus pada algoritma yang akan digunakan tanpa terlalu memperhatikan detail seperti pembacaan data dari file, implementasi algoritma *filtering* dan penyediaan kode untuk evaluasi hasil (Purnamasari, Henharta, Sasmita, Ihsani dan Wicaksana, 2013)[14].

2.11 PENELITIAN SEJENIS

Penelitian sejenis adalah sebagai modal acuan penulis untuk melakukan penelitian yang sebelumnya pernah dilakukan, diambil dari karya tulisan orang lain yang dapat dilihat dari tabel 2.1 berikut :

Tabel 2.1 Penelitian Sejenis

No	Penelitian dan Tahun	Judul Penelitian	Algoritma	Kesimpulan
1	(Mahardika dan Zuliarso, 2018)	ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA *NAIVES BAYES CLASSIFIER*	*Naïve Bayes Classifier*	Dari hasil pengujian, dapat disimpulkan bahwa metode *Naïve Bayes Classifier* dalam melakukan klasifikasi *tweet* sentimen negatif dan positif dengan 300 data latih dan 100 data uji mendapat hasil akurasi sebesar 97%. Dengan hasil akurasi yang cukup tinggi yaitu 97% maka metode *Naive Bayes Classifier*

				dapat digunakan untuk melakukan klasifikasi *tweet* dengan sentimen negatif dan positif secara otomatis.
2	(Rahayu, 2017)	ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL : *TWEET* TWITTER MENGGUNAKAN *NAÏVE BAYES*	*Naïve Bayes Classifier*	Penelitian ini menyimpulkan bahwa, aplikasi ini dapat mendowload data dari *Twitter* melalui API *streaming* dengan pengaturan data Bahasa Indonesia. Proses klasifikasi semakin akurat jika data latih yang digunakan dalam pembelajaran berjumlah banyak, akan tetapi dapat juga mengurangi keakuratan jika kata-kata yang terdapat pada *tweet* tersebut mengalami bias atau bermakna ganda,

				sehingga implementasi *Naïve Bayes* pada penelitian ini menghasilkan nilai akurasi 93%, serta presisi sentimen positif 90%, presisi sentimen negatif 90% dan presisi sentimen netral sebesar 100%.
3	(Amelia, Darmansah, Prastiwi dan Purbaya, 2022)	IMPLEMENTASI ALGORITMA *NAIVE BAYES* TERHADAP ANALISIS SENTIMEN OPINI MASYARAKAT INDONESIA MENGENAI DRAMA KOREA PADA TWITTER	*Naïve Bayes Classifier*	Penelitian ini menyimpulkan bahwa, analisis sentimen opini masyarakat terhadap drama Korea melalui Twitter terdiri dari pengambilan 100 data. *Tools* yang digunakan dalam proses analisis yaitu Orange versi 3.30.2 dengan menggunakan metode algoritma *Naive Bayes*. Hasil penelitian yang sudah dilakukan menunjukkan

				persentase performa diatas 69% dengan perhitungan algoritma *Naive Bayes* terhadap opini masyarakat melalui Twitter mengenai drama Korea, dengan persentase tingkat *accuracy* sebesar 69%, *precision* sebesar 73%, *recall* sebesar 69% dan F1 *Score* sebesar 69%.

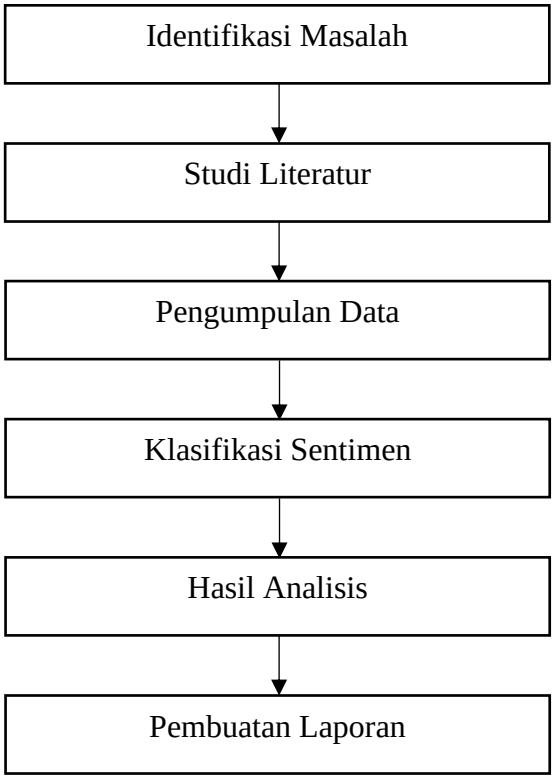
Dari beberapa studi literatur diatas, maka dapat disimpulkan bahwa kelebihan dari algoritma *Naïve Bayes Classifier* mampu menghasilkan klasifikasi teks yang cukup akurat. Pada penelitian ini, penulis menggunakan *tweet* yang mengandung *hashtag* #BBMNaik untuk melihat nilai sentimennya. Penulis memanfaatkan studi literatur diatas untuk dijadikan sebagai referensi saat melakukan penelitian.

BAB III

METODOLOGI PENELITIAN

3.1 KERANGKA KERJA PENELITIAN

Kerangka kerja penelitian merupakan tahap yang dilakukan selama mengerjakan penelitian. Kerangka kerja penelitian ini dibuat untuk mempermudah proses pencapaian hasil penelitian dan kerangka kerja penelitian ini bertujuan sebagai pedoman alur pikir penelitian yang harus diikuti agar tidak keluar dari ruang lingkup penelitian. Adapun kerangka kerja penelitian dapat dilihat pada gambar 3.1



Gambar 3.1 Kerangka Kerja Penelitian

Berdasarkan pada gambar kerangka kerja penelitian pada halaman sebelumnya, maka dapat diuraikan pembahasan dari setiap tahapannya sebagai berikut :

1. Identifikasi Masalah

   Pada tahapan ini, penulis mengidentifikasi apa saja yang menjadi masalah tentang klasifikasi dalam menganalisis sentimen pengguna Twitter mengenai *hashtag* terkait kebijakan kenaikan BBM dengan menggunakan algoritma *Naïve Bayes Classifier*, sehingga penulis dapat menentukan langkah kerja dan menentukan apa saja yang dibutuhkan dalam penelitian ini.

2. Studi Literatur

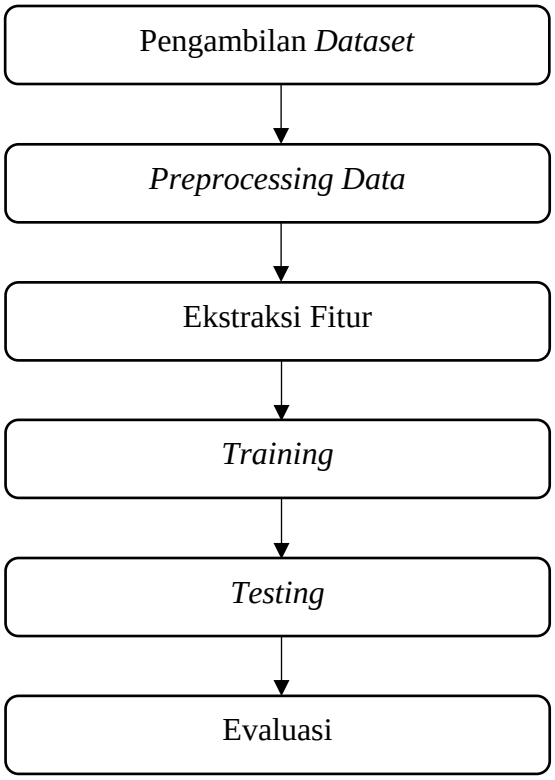
   Pada tahap ini, penulis melakukan pencarian tentang landasan teori yang dapat diperoleh dari berbagai sumber, seperti : buku, jurnal ilmiah, paper dan juga artikel ilmiah dari akademisi di internet yang mencakup tentang algoritma *Naïve Bayes Classifier*, *data mining*, analisis sentimen, RapidMiner, Twitter dan *hashtag* yang bertujuan untuk melengkapi penelitian, baik berupa konsep penelitian dan teori pada setiap pokok pembahasan penelitian.

3. Pengumpulan Data

   Pada tahap ini, pengumpulan data menggunakan bahasa pemrograman Python dengan Library Twitterscrapper, *tweet* akan diperoleh dalam bentuk file CSV, *keyword* yang dimasukkan dalam pengambilan *tweet* adalah *hashtag* “#BBMNaik”, “#KenaikanBBM” dan “#HargaBBMNaik” dengan periode pada tahun 2022 bulan Januari sampai dengan bulan September 2022, sehingga data yang akan dikumpulkan hanya berupa *tweet* yang terkait dengan *hashtag* kenaikan BBM pada jangka waktu di tahun tersebut. Setelah mendapatkan *tweet*, data akan diolah secara manual untuk mendapatkan data latih yang berkualitas. *Tweet* yang tidak berkaitan dengan *hashtag* kenaikan BBM akan dihilangkan.

4. Klasifikasi Sentimen

Pada tahap ini, pengklasifikasian sentimen akan dilakukan dengan beberapa tahapan. Adapun alur tahapan klasifikasinya dapat dilihat pada gambar 3.2



Gambar 3.2 Klasifikasi Sentimen (Nurhuda et al, 2016 : 39)

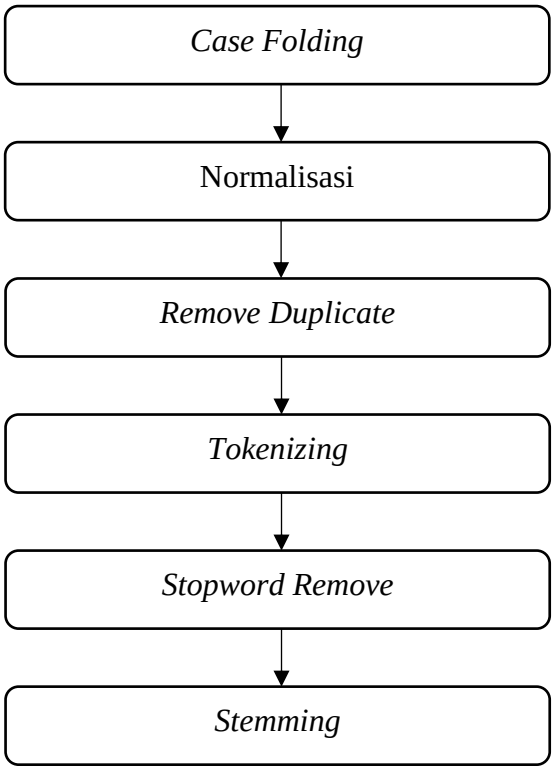
a. Pengumpulan *Dataset*

Dilakukan dengan cara memanfaatkan API dari pihak Twitter melalui teknik *streaming* data. Hasil dari *streaming* data, kemudian disimpan dalam bentuk file CSV. Kemudian, dilakukan *preprocessing* untuk menjadikan sederhana dimensi dari *dataset*. Selanjutnya, *dataset* tersebut kembali dipilah melalui ekstraksi fitur, dengan cara memanfaatkan *part-of-speech tagging* untuk mendapatkan kelas kata. Setelah mendapatkan kelas kata dari *dataset*, dilakukan *filtering* opini untuk menentukan suatu *tweet* termasuk opini atau bukan. Hasil dari *tweet* yang sudah masuk kategori opini, akan dipisahkan menjadi 2 kategori sentimen, yaitu sentimen yang mempunyai

polaritas positif dan negatif. Hasil *tweet* yang dipisahkan tadi, nantinya akan digunakan sebagai *dataset training* dimana proses pemisahannya dilakukan secara manual. Setelah *dataset training* terbentuk, barulah klasifikasi *tweet* menggunakan metode *Naïve Bayes Classifier* dimulai.

b. *Preprocessing Data*

Sebelum *dataset tweet* siap digunakan, maka akan terlebih dahulu dilakukan *preprocessing data*, sehingga *dataset* telah bersih dan siap digunakan dalam proses selanjutnya. Berikut adalah diagram alir mengenai *preprocessing data*.



Gambar 3.3 *Preprocessing Data*

Seperti yang dilihat pada gambar 3.3 diatas, bahwa tahap pertama dari *preprocessing data* adalah *case folding*, yaitu mengubah semua huruf kapital menjadi huruf kecil atau *lowercase*. Kemudian, dilakukan normalisasi berupa

penghapusan semua karakter selain *string*, serta penghapusan beberapa karakteristik dari data Twitter, misalnya *@username*, *#hashtag*, http://URL dan "RT" atau kata yang menandakan kalau itu perulangan *tweet*. Proses *remove duplicate* dilakukan untuk menghapus data yang ganda dalam proses pengambilan *tweet*. Pada tahapan *tokenizing*, akan dilakukan pemecahan kalimat berdasarkan kata yang telah ditentukan. Pada proses *stopwords*, dilakukan untuk menghapus setiap kata yang tidak ada hubungannya dengan analisis sentimen, seperti kata ganti, kata depan dan kata sambung. Pada tahap terakhir, yaitu *stemming* akan dilakukannya pembuangan kata yang memiliki imbuhan dan menjadikannya sebagai kata dasar.

c. Ekstraksi Opini

   Pada tahap ini, akan dilakukan penyaringan terhadap *tweet* yang telah dilakukan *preprocessing* data sebelumnya untuk mendapatkan *tweet* yang hanya mengandung opini atau sentimen, selanjutnya *dataset* di *filter* menggunakan *rule based* yang menunjukkan kalimat opini.

5. Hasil Analisis

   Pada tahapan ini, penulis menganalisis hasil dari pengklasifikasian sentimen dengan menggunakan algoritma *Naïve Bayes Classifier*, dengan memanfaatkan beberapa data *training*. Hasil yang didapatkan bisa dilihat apakah hasil tersebut bernilai *tweet* positif atau *tweet* negatif.

6. Pembuatan Laporan

   Setelah semua tahapan penelitian dilakukan, maka akan dibuat laporan sebagai dokumentasi penelitian agar dapat dimanfaatkan pada waktu yang akan datang untuk menjadi acuan bagi peneliti sendiri maupun orang lain.

3.2 ALAT BANTU PENELITIAN

Dalam melakukan penelitian, penulis juga memerlukan alat bantu. Alat bantu yang digunakan dalam melakukan analisis terbagi menjadi dua bagian, baik *hardware* maupun *software*. Adapun alat bantu yang digunakan, yaitu :

1. Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan dalam penelitian ini, yaitu :

a. Laptop dengan *processor* Intel(R) Core(TM) i3-10110U CPU @ 2.10GHz 2.59 GHz
b. RAM 4 GB

2. Perangkat Lunak (*Software*)

Perangkat lunak yang penulis gunakan dalam penelitian ini, yaitu :

a. Sistem operasi Windows 10
b. Microsoft Office 2021
c. RapidMiner
d. Google Chrome

3.3 JADWAL PENELITIAN

No	Kegiatan	Minggu Pelaksanaan															
		September				Oktober				November				Desember			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Identifikasi Masalah																
2	Studi Literatur																
3	Pengumpulan Data																
4	Klasifikasi Sentimen																
5	Hasil Analisis																
6	Pembuatan Laporan																

DAFTAR PUSTAKA

[1] N. B. Putri, "Kenaikan Harga BBM, Pro atau Kontra?," *www.kumparan.com*, 2022. https://kumparan.com/novanda-bemika/kenaikan-harga-bbm-pro-atau-kontra-1xrYpRAhqHD (diakses 14 September 2022).

[2] M. A. Rizaty, "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022," *www.dataindonesia.id*, 2022. https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022 (diakses 14 September 2022).

[3] E. M. Sipayung, H. Maharani, dan I. Zefanya, "**済無**No Title No Title," *Peranc. Sist. Anal. Sentimen Komentar Pelangg. Menggunakan Metod. Naive Bayes Classif.*, vol. 8, no. 9, hal. 958–965, 2016.

[4] S. R. Adi Widarma, "PERANCANGAN APLIKASI GAJI KARYAWAN PADA PT. PP LONDON SUMATRA INDONESIA Tbk. GUNUNG MALAYU ESTATE - KABUPATEN ASAHAN," *J. Teknol. Inf.*, vol. 1, 2019.

[5] B. A. B. Ii *et al.*, "Pemanfaatan Medsos Untuk Efektifitas Komunikasi," *Info Singk. Bid. Polit. Dalam Negeri Pus. Penelit. Badan Keahlian DPR RI*, vol. 12, no. 15, hal. 25–30, 2020.

[6] www.detik.com, "10 Alasan Kenapa Anda Harus Main Twitter," *www.detik.com*, 2012. https://inet.detik.com/tips-dan-trik/d-1927026/10-alasan-kenapa-anda-harus-main-twitter (diakses 14 September 2022).

[7] Ibnu, "Hashtag Adalah: Pengertian, Manfaat dan Strategi Tepat Dalam Menggunakannya," *www.accurate.id*, 2021. https://accurate.id/marketing-manajemen/hashtag-adalah/ (diakses 14 September 2022).

[8] "Jurnalhastag219gantipres (1)."

[9] R. Mulyawan, “Mengenal Pengertian Data Mining: Apa itu Penambangan Data? Menurut Ahli, Sejarah, Fungsi, Jenis, Proses dan Tahapannya!,” *www.rifqimulyawan.com*, 2022. https://rifqimulyawan.com/blog/pengertian-data-mining/ (diakses 14 September 2022).

[10] R. Setiawan, “Apa itu Data Mining dan Bagaimana Metodenya?,” *www.dicoding.com*, 2021. https://www.dicoding.com/blog/apa-itu-data-mining/ (diakses 14 September 2022).

[11] I. Hamil, “Bab 2 tinjauan pustaka 2.1.,” vol. d, no. 1953, hal. 5–12, 1991.

[12] M. H. Widianto, “Algoritma Naive Bayes,” *www.binus.ac.id*. https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/ (diakses 14 September 2022).

[13] A. Umar, “RapidMiner, Definisi dan Fitur-Fiturnya,” *www.abdumar.com*. https://www.abdumar.com/2021/03/rapidminer-definisi-dan-fitur-fiturnya.html?m=1 (diakses 30 Juli 2022).

[14] www.jakhi.live, “Mesin yang Bertugas Menata Data Supaya Lebih Mudah dan Cepat ditampilkan ke Pelanggan Adalah,” *www.jakhi.live*. https://jakhi.live/mesin-yang-bertugas-menata-data-supaya-lebih-mudah-dan-cepat-ditempilkan-ke-pelanggan-adalah (diakses 14 September 2022).

LAMPIRAN

YAYASAN DINAMIKA BANGSA
BIRO ADMINISTRASI AKADEMIK DAN KEMAHASISWAAN
UNIVERSITAS DINAMIKA BANGSA

KARTU RENCANA STUDI (KRS)

Nama Mahasiswa	: Zaneta Hanes Armairan	Pemb. Akademik	: Desi Kisbianty, ST, M.S.I
Nomor Mahasiswa	: 8020190230	Semester	: Ganjil 2022/2023
Program Studi	: Teknik Informatika	Angkatan	: 2019
Jenjang Studi	: Strata 1	IP Sebelumnya	: 3.79

No	Kode MK	Nama Matakuliah	SKS	B/U	Ruang
1	MPTI16P313	Networking Advanced	3	B	Kampus Thehok, Labor Cisco 4.3 Se
2	MPTI16P304	Aplikasi Enterprise	3	B	Kampus Kobar, R. 2.15 Kamis,14:45
3	PRTI168623	Skripsi	6	B	R. Dosen Minggu,08:00
#	TISISK7	TOEFL	-	B	Kampus Kobar, R. 2.17 Selasa 15:30 Kampus Kobar, R. 2.17 Rabu 15:30 -
#	-	MTA Microsoft Certified: Azure AI Fundamentals	-	-	-
Total Kredit Yang Diambil semester ini			12		

Jambi, 14 September 2022
Pembimbing Akademik
ttd
Desi Kisbianty, ST, M.S.I
NIK : YDB.09.82.068

Mahasiswa yang mengontrak **Tugas Akhir / Skripsi Wajib mendaftar di sintap.unama.ac.id*
Mahasiswa yang mengontrak **Proyek Penelitian / Kerja Praktek Wajib mendaftar di sintap.unama.ac.id*
**Dokumen ini sah tanpa tanda tangan Pembimbing Akademik*

Daftar *Stopword*

ada
adalah
adanya
adapun
agak
agaknya
agar
akan
akankah
akhir
akhiri
akhirnya
aku
akulah
amat
amatlah
anda
andalah
antar
antara
antaranya
apa
apaan
apabila
apakah

apalagi

apatah

artinya

asal

asalkan

atas

atau

ataukah

ataupun

awal

awalnya

bagai

bagaikan

bagaimana

bagaimanakah

bagaimanapun

bagi

bagian

bahkan

bahwa

bahwasanya

baik

bakal

bakalan

balik

banyak

bapak

baru

bawah

beberapa

begini

beginian

beginikah

beginilah

begitu

begitukah

begitulah

begitupun

bekerja

belakang

belakangan

belum

belumlah

benar

benarkah

benarlah

berada

berakhir

berakhirlah

berakhirnya

berapa

berapakah

berapalah

berapapun

berarti

berawal

berbagai

berdatangan

beri

berikan

berikut

berikutnya

berjumlah

berkali-kali

berkata

berkehendak

berkeinginan

berkenaan

berlainan

berlalu

berlangsung

berlebihan

bermacam

bermacam-macam

bermaksud

bermula

bersama

bersama-sama

bersiap

bersiap-siap

bertanya

bertanya-tanya

berturut

berturut-turut

bertutur

berujar

berupa

besar

betul

betulkah

biasa

biasanya

bila

bilakah

bisa

bisakah

boleh

bolehkah

bolehlah

buat

bukan

bukankah

bukanlah

bukannya

bulan

bung

cara

caranya

cukup

cukupkah

cukuplah

cuma

dahulu

dalam

dan

dapat

dari

daripada

datang

dekat

demi

demikian

demikianlah

dengan

depan

di

dia

diakhiri

diakhirinya

dialah

diantara

diantaranya

diberi

diberikan

diberikannya

dibuat

dibuatnya

didapat

didatangkan

digunakan

diibaratkan

diibaratkannya

diingat

diingatkan

diinginkan

dijawab

dijelaskan

dijelaskannya

dikarenakan

dikatakan

dikatakannya

dikerjakan

diketahui

diketahuinya

dikira

dilakukan

dilalui

dilihat

dimaksud

dimaksudkan

dimaksudkannya

dimaksudnya

diminta

dimintai

dimisalkan

dimulai

dimulailah

dimulainya

dimungkinkan
dini
dipastikan
diperbuat
diperbuatnya
dipergunakan
diperkirakan
diperlihatkan
diperlukan
diperlukannya
dipersoalkan
dipertanyakan
dipunyai
diri
dirinya
disampaikan
disebut
disebutkan
disebutkannya
disini
disinilah
ditambahkan
ditandaskan
ditanya
ditanyai
ditanyakan
ditegaskan
ditujukan

ditunjuk

ditunjuki

ditunjukkan

ditunjukkannya

ditunjuknya

dituturkan

dituturkannya

diucapkan

diucapkannya

diungkapkan

dong

dua

dulu

empat

enggak

enggaknya

entah

entahlah

guna

gunakan

hal

hampir

hanya

hanyalah

hari

harus

haruslah

harusnya

hendak
hendaklah
hendaknya
hingga
ia
ialah
ibarat
ibaratkan
ibaratnya
ibu
ikut
ingat
ingat-ingat
ingin
inginkah
inginkan
ini
inikah
inilah
itu
itukah
itulah
jadi
jadilah
jadinya
jangan
jangankan
janganlah

jauh

jawab

jawaban

jawabnya

jelas

jelaskan

jelaslah

jelasnya

jika

jikalau

juga

jumlah

jumlahnya

justru

kala

kalau

kalaulah

kalaupun

kalian

kami

kamilah

kamu

kamulah

kan

kapan

kapankah

kapanpun

karena

karenanya

kasus

kata

katakan

katakanlah

katanya

ke

keadaan

kebetulan

kecil

kedua

keduanya

keinginan

kelamaan

kelihatan

kelihatannya

kelima

keluar

kembali

kemudian

kemungkinan

kemungkinannya

kenapa

kepada

kepadanya

kesampaian

keseluruhan

keseluruhannya

keterlaluan

ketika

khususnya

kini

kinilah

kira

kira-kira

kiranya

kita

kitalah

kok

kurang

lagi

lagian

lah

lain

lainnya

lalu

lama

lamanya

lanjut

lanjutnya

lebih

lewat

lima

luar

macam

maka

makanya

makin

malah

malahan

mampu

mampukah

mana

manakala

manalagi

masa

masalah

masalahnya

masih

masihkah

masing

masing-masing

mau

maupun

melainkan

melakukan

melalui

melihat

melihatnya

memang

memastikan

memberi

memberikan

membuat

memerlukan
memihak
meminta
memintakan
memisalkan
memperbuat
mempergunakan
memperkirakan
memperlihatkan
mempersiapkan
mempersoalkan
mempertanyakan
mempunyai
memulai
memungkinkan
menaiki
menambahkan
menandaskan
menanti
menanti-nanti
menantikan
menanya
menanyai
menanyakan
mendapat
mendapatkan
mendatang
mendatangi

mendatangkan

menegaskan

mengakhiri

mengapa

mengatakan

mengatakannya

mengenai

mengerjakan

mengetahui

menggunakan

menghendaki

mengibaratkan

mengibaratkannya

mengingat

mengingatkan

menginginkan

mengira

mengucapkan

mengucapkannya

mengungkapkan

menjadi

menjawab

menjelaskan

menuju

menunjuk

menunjuki

menunjukkan

menunjuknya

menurut
menuturkan
menyampaikan
menyangkut
menyatakan
menyebutkan
menyeluruh
menyiapkan
merasa
mereka
merekalah
merupakan
meski
meskipun
meyakini
meyakinkan
minta
mirip
misal
misalkan
misalnya
mula
mulai
mulailah
mulanya
mungkin
mungkinkah
nah

naik

namun

nanti

nantinya

nyaris

nyatanya

oleh

olehnya

pada

padahal

padanya

pak

paling

panjang

pantas

para

pasti

pastilah

penting

pentingnya

per

percuma

perlu

perlukah

perlunya

pernah

persoalan

pertama

pertama-tama

pertanyaan

pertanyakan

pihak

pihaknya

pukul

pula

pun

punya

rasa

rasanya

rata

rupanya

saat

saatnya

saja

sajalah

saling

sama

sama-sama

sambil

sampai

sampai-sampai

sampaikan

sana

sangat

sangatlah

satu

saya

sayalah

se

sebab

sebabnya

sebagai

sebagaimana

sebagainya

sebagian

sebaik

sebaik-baiknya

sebaiknya

sebaliknya

sebanyak

sebegini

sebegitu

sebelum

sebelumnya

sebenarnya

seberapa

sebesar

sebetulnya

sebisanya

sebuah

sebut

sebutlah

sebutnya

secara

secukupnya
sedang
sedangkan
sedemikian
sedikit
sedikitnya
seenaknya
segala
segalanya
segera
seharusnya
sehingga
seingat
sejak
sejauh
sejenak
sejumlah
sekadar
sekadarnya
sekali
sekali-kali
sekalian
sekaligus
sekalipun
sekarang
sekarang
sekecil
seketika

sekiranya

sekitar

sekitarnya

sekurang-kurangnya

sekurangnya

sela

selain

selaku

selalu

selama

selama-lamanya

selamanya

selanjutnya

seluruh

seluruhnya

semacam

semakin

semampu

semampunya

semasa

semasih

semata

semata-mata

semaunya

sementara

semisal

semisalnya

sempat

semua

semuanya

semula

sendiri

sendirian

sendirinya

seolah

seolah-olah

seorang

sepanjang

sepantasnya

sepantasnyalah

seperlunya

seperti

sepertinya

sepihak

sering

seringnya

serta

serupa

sesaat

sesama

sesampai

sesegera

sesekali

seseorang

sesuatu

sesuatunya

sesudah

sesudahnya

setelah

setempat

setengah

seterusnya

setiap

setiba

setibanya

setidak-tidaknya

setidaknya

setinggi

seusai

sewaktu

siap

siapa

siapakah

siapapun

sini

sinilah

soal

soalnya

suatu

sudah

sudahkah

sudahlah

supaya

tadi

tadinya

tahu

tahun

tak

tambah

tambahnya

tampak

tampaknya

tandas

tandasnya

tanpa

tanya

tanyakan

tanyanya

tapi

tegas

tegasnya

telah

tempat

tengah

tentang

tentu

tentulah

tentunya

tepat

terakhir

terasa

terbanyak

terdahulu

terdapat

terdiri

terhadap

terhadapnya

teringat

teringat-ingat

terjadi

terjadilah

terjadinya

terkira

terlalu

terlebih

terlihat

termasuk

ternyata

tersampaikan

tersebut

tersebutlah

tertentu

tertuju

terus

terutama

tetap

tetapi

tiap

tiba

tiba-tiba

tidak

tidakkah

tidaklah

tiga

tinggi

toh

tunjuk

turut

tutur

tuturnya

ucap

ucapnya

ujar

ujarnya

umum

umumnya

ungkap

ungkapnya

untuk

usah

usai

waduh

wah

wahai

waktu

waktunya

walau

walaupun

wong

yaitu

yakin

yakni

yang