BAB I

PENDAHULUAN

1.1 Latar Belakang

Presiden merupakan seseorang yang bertindak sebagai pemimpin dalam suatu negara dengan tugasnya yaitu sebagai kepala negara dan kepala pemerintahan. Selama 77 tahun Negara Indonesia telah berdiri, negara ini dipimpin oleh beberapa orang berbeda seperti Dr. Ir. H. Soekarno (1945 – 1967), Jendral TNI H.M. Soeharto (1967 – 1998), Prof. Dr. Ing. B. J. Habibie (1998 – 1999), K. H. Abdurrahman Wahid (Gusdur) (1999 – 2001), Megawati Soekarnoputri (2001 – 2004), Susilo Bambang Yudhoyono (2004 – 2009, 2009 – 2014), dan yang menjabat saat ini Ir. H. Joko Widodo (2014 – 2019, 2019 – Sekarang) [1].

Presiden di Indonesia dipilih melalui masyarakat dengan melalui proses demokrasi yaitu pemilihan presiden (pilpres) yang dilaksanakan setiap 5 tahun sekali. Menjadi seorang presiden memiliki beberapa persyaratan yang dimana persyaratan tersebut adalah seseorang tidak diperbolehkan menjadi presiden apabila orang tersebut sebelumnya telah menjadi presiden selama 2 periode secara berturut – turut [2], yang dalam hal ini presiden Indonesia saat ini sudah tidak bisa mencalonkan kembali menjadi Presiden pada pilpres selanjutnya yang akan terlaksana pada tahun 2024. Berdasarkan tersebut banyak bermunculan survei elektabilitas terhadap beberapa tokoh publik yang memiliki elektabilitas baik yang menjadikan tokoh ini bisa dijadikan bakal calon presiden Indonesia di pilpres pada tahun 2024. Pada hasil dari beberapa survei tersebut memiliki hasil yang cukup

beragam dimana peneliti mengambil 3 hasil survei dari lembaga survei berbeda yang ditemukan pada suatu lini berita yang memiliki judul “Zulhas Didorong Nyapres 2024, Survei Membuktikan...”[3]–[5]. Dari hasil ketiga survei tersebut peneliti melakukan pengambilan nilai rata – rata dari ketiga survei tersebut dengan hasilnya empat tokoh publik dengan nilai rata – rata elektabilitas terbesar adalah Ganjar Pranowo 19,2%, Anies Baswedan 14,2%, Prabowo Subianto 14%. Dari hasil survei tersebut dan juga adanya lini berita yang sudah mempublikasi berita tersebut memiliki adanya kemungkinan penyampaian pendapat masyarakat atas hasil survei tersebut dengan media sosial.

Secara umum pemilihan umum lahir dari konsep atau gagasan besar demokrasi, keadilan dam kesetaraan bagi seluruh individu disuatu negara dalam segala aspek kehidupan. Dalam pelaksanaan pemilu diperlukan peran aktif warga masyarakat untuk menyukseskan pemilihan umum ini. Masyarakat diberikan akses dan ruang untuk berperan aktif menjadi bagian dari sebuah pemilihan umum. Meskipun terkadang peran masyarakat masih cenderung hanya mengikuti prosedur dan momentum saja [6].

Pengguna media sosial yang ada pada Indonesia saat ini sudah mencapai 170 juta pengguna di awal tahun 2021, dengan media sosial yang cukup favorit adalah twitter dengan pembuktiannya dengan hasil dari survei bahwa Indonesia menempati sebagai posisi ke-6 dengan peringkat dunia penggunaan twitter, selain dari itu di Indonesia pengguna dengan jenjang umur 16 – 64 tahun mencapai tingkat 63,6% yang bersesuaian dengan sudah layak menjadi pemilih dalam pilpres [7].

Dalam memasuki tahun-tahun politik saat ini akan banyak bertebaran informasi dan komentar dari masyarakat terhadap isu bakal calon Presiden pada pemilu 2024 ini, dalam hal itu banyak lembaga survei yang telah mengeluarkan hasil survei terhadap beberapa tokoh yang memiliki elektabilitas untuk bisa menjadi calon presiden, berdasarkan hal tersebut juga banyak warganet yang menyampaikan pendapat, dari pendapat tersebut bisa dibuat kesimpulan mengenai sentimen warga masyarakat terhadap suatu tokoh bakal calon presiden tersebut, komentar atau pendapat yang disampaikan akan sangat beragam dimulai dari dukungan terhadap pelaksanaan pemilu, penggiringan opini publik, ujaran kebencian dan komentar-komentar lainnya. kemajuan teknologi saat ini mengakibatkan penyampaian pendapat dapat dengan mudah dipublikasikan melalui media sosial. Salah satunya adalah melalui media twitter, twitter menjadi salah satu media sosial yang paling sering digunakan masyarakat dalam mengemukakan pendapatnya karena dianggap bebas. Hingga saat ini media twitter menjadi media yang sangat bebas dan luas dalam menyampaikan informasi sehingga dalam memasuki tahun politik ini akan ada banyak asumsi dari masyarakat terhadap bakal calon Presiden pada pemilu 2024, komentar-komentar yang diberikan sangatlah beragam sehingga menimbulkan fenomena politik, dikarenakan komentar yang diberikan setiap individu unik dan beragam maka diperlukan sebuah analisa publik terkait pendapat isu terhadap bakal calon Presiden pada pemilu 2024, dengan melakukan analisis opini masyarakat di media sosial twitter terhadap isu bakal calon Presiden pada pemilu 2024 nanti kita dapat mengetahui respon masyarakat Indonesia terhadap gelaran tersebut seperti apa.

Sesuai dengan keadaan dan isu terhadap bakal calon Presiden pada pemilu 2024, banyaknya pendapat-pendapat yang dikemukakan tentang isu tersebut, baik itu pendapat positif, negatif dan netral, dari pendapat tersebut diketahui berapa berapa banyak yang mendukung, tidak mendukung ataupun bersifat netral. Untuk mengetahui hal tersebut dibutuhkan data yang valid dan mengelompokkan nya sesuai polaritas dari tanggapan tersebut dan dengan harapan dari penelitian ini bisa memberikan bahan referensi kepada masyarakat dalam memilih presiden di pilpres yang akan datang. Selain itu dapat dijadikan acuan untuk terhindar dari berita hoax, opini negatif, dan ujaran kebencian kepada satu pihak.

Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral. Saat ini, peneliti secara luas menggunakan analisis sentimen sebagai salah satu cabang penelitian dalam ilmu komputer. Jejaring sosial, seperti *Twitter*, umumnya digunakan dalam analisis sentimen untuk menentukan persepsi. [8]

Text mining adalah algortima komputer yang bertujuan untuk menggali suatu informasi atau mengekstrak informasi baru yang sebelumnya belum diketahui ketika belum dilakukan proses *text mining* yang berasal dari teks yang tidak terstruktur dalam jumlah besar. Atau dalam pengertian lain *text mining* adalah kegiatan menambang dari suatu data berupa text untuk menemukan sebuah inti dari suatu dokumen atau kata-kata tertentu apa yang bisa mewakili dokumen dan menganalisa keterhubungan antar dokumen yang diperiksa oleh komputer. [9]

Metode *Naive Bayes* adalah metode klasifikasi dalam penambangan teks yang digunakan dalam analisis sentimen. Metode ini berpotensi baik dalam klasifikasi dalam hal presisi dan komputasi data. *Naive Bayes* banyak digunakan dalam teknik klasifikasi terutama *Twitter* menggunakan beberapa metode seperti Unigram *Naive Bayes*, *Multinomial Naive Bayes*, dan *Maximum Entropy Classification*. Fitur utama klasifikasi *Naive Bayes* untuk mendapatkan hipotesis yang kuat dari setiap kondisi atau peristiwa. Perhitungan kategori probabilitas dalam *Naive Bayes* menggunakan pendekatan algoritma *Bayes* menggunakan persamaan. [10]

Teorema Bayes menerangkan hubungan antara probabilitas terjadinya peristiwa A dengan syarat peristiwa B telah terjadi dan probabilitas terjadinya peristiwa B dengan syarat peristiwa A telah terjadi. Teorema ini didasarkan pada prinsip bahwa tambahan informasi dapat memperbaiki probabilitas. Teorema Bayes ini bermanfaat untuk mengubah atau memutakhirkan (mengupdate) probabilitas yang dihitung dengan tersedianya data dan informasi tambahan. [11]

Berdasarkan uraian latar belakang diatas, maka penulis membuat aplikasi yang berjudul **"Analisis Sentimen *Public Twitter* Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Metode *Naive Bayes*".**

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, pokok permasalahannya adalah bagaimana merancang dan membangun Aplikasi Analisis Sentimen *Public Twitter* Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Metode *Naive Bayes*?.

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah merancang dan membangun Aplikasi Analisis Sentimen *Public Twitter* Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Metode *Naive Bayes*.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

1. Data diambil dari media sosial *twitter.*
2. Aplikasi ini berbasis *web.*
3. Aplikasi ini hanya bisa di akses oleh *admin*
4. Data cuitan dari setiap tokoh bakal calon Presiden Republik Indonesia 2024 - 2029.
5. Data komentar yang di kumpulkan sebelum tahun 2024.
6. Jumlah data sebanyak 1000 data komentar.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat menambah pemahaman lebih lanjut mengenai ilmu yang dipelajari selama kuliah serta bisa mengetahui pendapat – pendapat pengguna *twitter* terhadap calon Presiden pada pemilu 2024.

1.6 Sistematika Penulisan

Sistematika penulisan dari skripsi ini terdiri dari lima bagian utama sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah tujuan penelitian, batasan masalah, manfaat penelitian dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini berisi teori-teori yang digunakan pada penelitian ini. Teori-teori yang berhubungan dengan landasan dalam pembuatan aplikasi.

BAB 3 METODOLOGI PENELITIAN

Bab ini berisi tahapan–tahapan dalam pengumpulan data, perancangan sistem perumusan masalah dan analisa.

BAB 4 ANALISA DAN PERANCANGAN

Bab ini berisi analisa dan perancangan Media pembelajaran yang akan dibangun atau dikembangkan.

BAB 5 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi dari analisa dan perancangan dan pengujian pada aplikasi yang berhasil dibangun.

BAB 6 PENUTUP

Bab ini berisi rangkuman dari hasil penelitian yang telah dilakukan dan saran–saran untuk pengembangan aplikasi atau penelitian selanjutnya.

BAB II

LANDASAN TEORI

2.1 Analisis

Analisis merupakan suatu upaya penyelidikan untuk melihat, mengamati,mengetahui, menemukan, memahami, menelaah, mengklasifikasi, dan mendalami serta menginterpretasikan fenomena yang ada [8].

Analisis Sistem sebagai penguraian dari suatu sistem informasi yang utuh ke dalam bagian - bagian komponennya dengan maksud untuk mengidentifikasikan dan mengevaluasi permasalahan dan kesempatan [9].

Analisis merupakan suatu kegiatan dalam rangka mempelajari suatu bentuk benda,kegiatan,fakta maupun fenomena dalam mempelajari hubungan antar bagian yang terkait dari berbagai fungsi yang menjadi satu. [10].

Analisis dapat didefinisikan sebagai penguraian dari suatu sistem informasi yang utuh ke dalam bagian komponennya dengan maksud untuk mengevaluasi permasalahan, kesempatan, hambatan yang terjadi dan kebutuhan yang diharapkan sehingga dapat diusulkan perbaikannya [11].

Analisis adalah kegiatan berpikir untuk menguraikan suatu keseluruhan menjadi komponen sehingga dapat mengenal tanda-tanda komponen, hubungannya satu sama lain dan fungsi masing-masing dalam satu keseluruhan yang terpadu [12].

2.2 Analisis Sentimen

Sentiment analysis atau analisis sentimen dalam Bahasa Indonesia adalah sebuah teknik atau cara yang digunakan untuk mengidentifikasi bagaimana sebuah

sentimen diekspresikan menggunakan teks dan bagaimana sentimen tersebut bisa dikategorikan sebagai sentimen positif maupun sentimen negatif [13].

Analisis sentimen adalah bidang *interdisipliner*, sebuah bidang dimana digabung dengan cara *probabilistic weighting model* untuk memberi skor pada pada setiap fitur [13].

Analisis sentimen adalah proses yang digunakan untuk menentukan opini, emosi dan sikap yang dicerminkan melalui teks, dan biasanya diklasifikasikan menjadi opini negatif dan positif [13].

Analisis sentimen adalah tugas menemukan opini dari penulis tentang suatu identitas tertentu. Untuk melakukan analisis sentimen ada beberapa algoritma yang dapat digunakan salah satunya adalah algoritma *Naive Bayes* [14].

Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral. Saat ini, peneliti secara luas menggunakan analisis sentimen sebagai salah satu cabang penelitian dalam ilmu komputer. Jejaring sosial, seperti *Twitter*, umumnya digunakan dalam analisis sentimen untuk menentukan persepsi public [15].

2.3 *Public*

Publik adalah sejumlah manusia yang memiliki kebersamaan berfikir, perasaan, harapan, sikap atau tindakan yang benar dan baik berdasarkan nilai-nilai norma yang merasa memiliki [16].

Menurut Gruth dan Marsh dalam Estawara (2010), publik adalah setiap kelompok orang yang memiliki minat atau nilai-nilai bersama dalam situasi

tertentu, terutama kepentingan atau nilai-nilai mereka mungkin bertindak atas kesediaan [17].

Publik adalah sekelompok orang/individu yang memiliki kepentingan dan minat yang sama akan sesuatu hal [18].

2.4 *Twitter*

Twitter sebagai salah satu media baru dalam kategori media sosial yang terhubung dengan internet, menjadikannya sebuah media untuk menyebarkan pesan melalui komunikasi massa [19].

Twitter adalah sebuah media sosial yang berbentuk *microblogging* yang memberikan fasilitas kepada penggunya untuk untuk mengirim dan membagikan informasi atau pendapatnya yang tidak lebih dari 140 karakter, atau yang disebut sebagai *tweet* atau *twit* [20].

Twitter adalah sebuah situs *web* yang dimiliki dan dioperasikan oleh *Twitter Inc*, dimana *twitter* menawarkan jejaring sosial berupa *mikroblog* sehingga memungkinkan penggunanya untuk mengirim dan membaca pesan yang disebut kicauan (*tweets*). *Tweet* bisa dilihat secara umum namun pengirim dapat membatasi penerima pesan ke daftar orang terdekat saja. Pengguna dapat melihat kicauan penulis lain yang dikenal dengan sebutan pengikut (*follower*). Karena jumlah karakter huruf pada *Twitter* yang dibatasi inilah *Twitter* banyak digunakan sebagai media untuk menulis pesan singkat dan cepat [21].

Twitter adalah suatu jaringan sosial yang dilambangkan serupa dengan burung sebagai kicauan atau pesan untuk menerima atau memberikan informasi secara up to date dengan detail yakni: tanggal, waktu dan zona tercantumkan

dikronologi twitter sesuai pesan yang disajikan pada saat itu, yang dimaksud dengan kicauan adalah teks tulisan hingga 140 karakter yang ditampilkan pada halaman profil pengguna [21].

Twitter adalah situs micro blogging yang dioperasikan oleh Twitter, Inc. Disebut micro blogging karena situs ini memungkinkan penggunanya mengirim dan membaca pesan seperti blog pada umumnya. Pesan tersebut dinamakan tweet, yaitu teks tulisan sebanyak 140 karakter yang ditampilkan pada halaman profil pengguna [21].

2.5 Calon Presiden

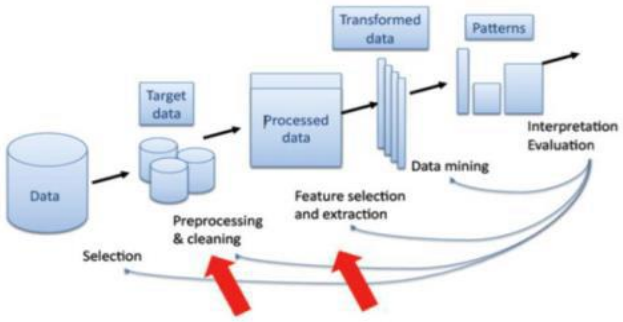
Pengertian calon presiden yang dikutip dari glosarium.org yaitu : orang-orang yang memenuhi syarat sebagai calon presiden dan namanya terdaftar di Komisi Pemilihan Umum sebagai peserta pemilihan presiden. Seseorang bisa mencalonkan diri menjadi presiden/wakil presiden di Indonesia harus memenuhi syarat presiden dan wakil presiden Republik Indonesia [26].

2.6 *Text Mining*

Text mining adalah algortima komputer yang bertujuan untuk menggali suatu informasi atau mengekstrak informasi baru yang sebelumnya belum diketahui ketika belum dilakukan proses *text mining* yang berasal dari *teks* yang tidak terstruktur dalam jumlah besar. Atau dalam pengertian lain *text mining* adalah kegiatan menambang dari suatu data berupa *text* untuk menemukan sebuah inti dari suatu dokumen atau kata-kata tertentu apa yang bisa mewakili dokumen dan menganalisa keterhubungan antar dokumen yang diperiksa oleh komputer [23].

Text mining adalah satu langkah dari *analisis teks* yang dilakukan secara otomatis oleh komputer untuk menggali informasi yang berkualitas dari suatu rangkaian *teks* yang terangkum dalam sebuah dokumen. Prosedur utama dalam metode ini terkait dengan menemukan kata-kata yang dapat mewakili isi dari dokumen untuk selanjutnya dilakukan analisis keterhubungan antar dokumen dengan menggunakan metode statistik tertentu seperti *analisis* kelompok, *klasifikasi* dan *asosiasi* [24].

Tahapan dalam *text mining* secara umum dapat dilihat pada gambar dibawah ini**.**



Gambar 2.1 Tahapan *Text Mining*

2.7 Data *Training*

Data *training* adalah data yang digunakan sebagai acuan untuk membangun model *klasifikasi*. Pada skenario pengujian ini data *training* sudah dipersiapkan sebelumnya dan disimpan ke dalam dataset. Hasil data *training* yang didapat akan diuji dengan data baru yang masih belum diketahui *classtype* kategorinya [25].

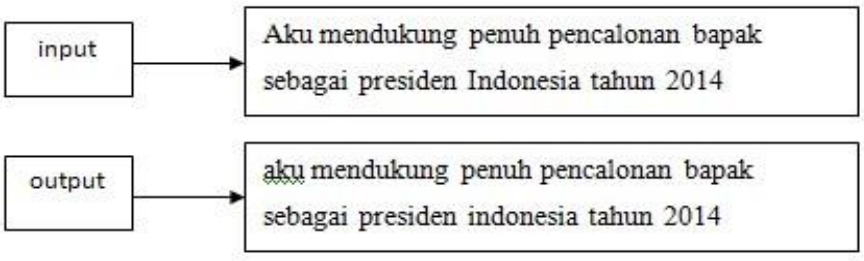
2.8 *Preprocessing*

Preprocessing untuk data yang akan di gunakan dalam proses analisa sentimen, dimana data yang kita proses akan kita ambil informasi yang terkandung didalamnya dalam hal sentimen penulisnya yaitu negaitf, positif atau netral. Guna memudahkan dalam mengelola data maka data perlu kita berikan analisa sentimen secara manual dengan membaca maksud dari kalimat yang ada dalam sentimen tersebut, sehingga dapat diberikan penilaian bahwa sentimen tersebut merupakan setimen negatif, positif atau netral [26].

2.9 *Case Folding*

Case folding adalah mengubah semua huruf besar atau kapital dalam Twitter menjadi huruf kecil. Hanya huruf ‘a’ sampai ‘z’ yang diterima. Karakter selain huruf dihilangkan dan dianggap *delimiter*. *Delimiter* adalah urutan satu karakter atau lebih yang dipakai untuk membatasi atau memisahkan data yang disajikan dalam *plain text*[25].

Contoh dari tahap ini seperti yang ada dalam Gambar 2.2.



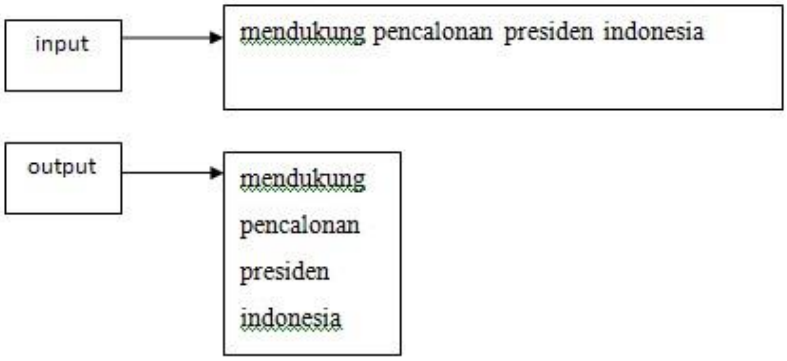
Gambar 2.2. Proses *Case Folding*

2.10 *Tokenizing*

Tokenizing atau tokenisasi adalah proses yang paling awal dalam melakukan *text mining*. Dalam proses ini, input stream yang didapat dari file *text* akan dipecah-

pecah menjadi bagian bagian yang lebih kecil. Sebagai contoh pemecahan kalimat menjadi kata-kata (*tokens*) [26].

Contoh dari tahap ini seperti yang ada dalam Gambar 2.3.



Gambar 2.3. Proses *Tokenizing*

2.11 *Stop Word*

Langkah yang selanjutnya adalah *Stop Word Removal*. *Stop Word* sendiri didefinisikan sebagai sekumpulan kata yang tidak berhubungan (*irrelevant*) dengan subyek utama yang dimaksud, meskipun kata tersebut sering muncul didalam data yang digunakan. Kata-kata yang dimaksud biasanya adalah jenis kata sambung, imbuhan, dan lain sebagainya [27].

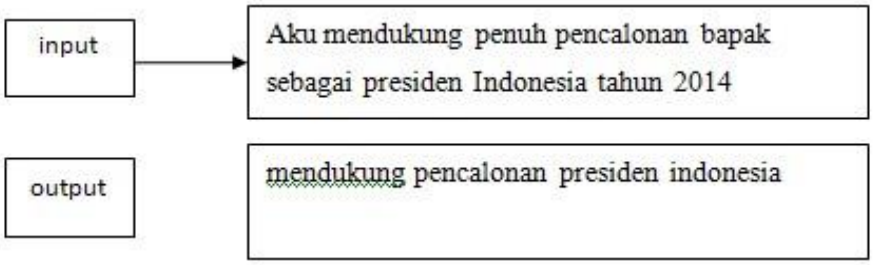
2.12 Normalisasi

Normalisasi adalah memastikan bahwa tidak terjadi Partial Dependency, yaitu kondisi ketika salah satu atribut dalam satu entitas bergantung pada sebagian kunci primer yang ada di entitas tersebut. Dalam penelitian ini, bentuk kedua dari normalisasi ini dapat dilihat pada penjabaran [28].

2.13 *Filtering*

Filtering adalah tahap mengambil kata penting dari hasil proses token. Bisa menggunakan algoritma *stoplist* atau *wordlist*. *Filtering* dapat juga diartikan sebagai proses mengambil kata – kata penting dari hasil proses token atau penghapusan stopwords [29].

Contoh dari tahap ini seperti yang ada dalam Gambar 2.4.



Gambar 2.4. Proses *Filtering*

2.14 *Stemming*

Stemming digunakan untuk memeriksa relevannya beberapa dokumen dibutuhkan proses *stemming*. *Stemming* adalah di mana proses pemetaan kata pada kalimat berimbuhan menjadi kata asli (tanpa kata imbuhan awalan, akhiran, sisipan, kombinasi) yang di jalankan algoritma tertentu [30].

2.15 *Naive Bayes*

Secara umum algoritma *naive bayes* untuk membangun sebuah hasil kesimpulan adalah sebagai berikut "*Naive Bayes Classifer* merupakan salah satu algoritma dalam teknik data mining yang menerapkan teori bayes dalam klasifikasi. Teorema keputusan *bayes* adalah pendekatan statistik yang fundamental dalam pengenalan pola (*Pattern Recognition*), *Naive bayes* didasarkan pada asumsi

penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai *output* [31].

Naive Bayes merupakan suatu bentuk *klasifikasi* data dengan menggunakan metode probabilitas dan statistik. Algoritma *Naive Bayes* dapat diartikan sebagai sebuah metode yang tidak memiliki aturan, *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan *klasifikasi* dengan cara melihat *frekuensi* tiap *klasifikasi* pada data *training* [32].

Naive bayes merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan aturan *bayes* dengan asumsi ketidaktergantungan yang kuat. Selain itu naive bayes juga dapat menganalisis variabel-variabel yang paling mempengaruhinya dalam bentuk peluang [33].

a) Hitung probabilitas *prior* setiap kategori, yang menjadi kategori ada 3 kategori yanitu positif, netral, negatif.

$$\text{p}\frac{\frac{postif}{netral}}{negatif} = \frac{x(\frac{\frac{postif}{netral}}{negatif})}{|c|} \quad (1)$$

b) Menghitung probabilitas *likelihood* setiap term dari semua dokumen, Banyaknya term tergantung pada hasil praproses data.

$$\text{P(w|positif/netral/negatif)} = \frac{(nk\left(\frac{\frac{positif}{netral}}{negatif}\right)+1}{(n.\frac{\frac{positif}{netral}}{negatif}+|r|} \quad (2)$$

c) Proses klasifikasi data uji yang termasuk ke dalam data training dengan memilih kata data uji. Nilai yang lebih tinggi merupakan kelas baru dari data tersebut.

$$p\left(\frac{postif}{netral}{negatif}|d\right) = P\left(\frac{positif}{netral}{negatif}\right) \text{ x np(w|positif/netral/negatif)} \quad (3)$$

2.16 *Term Frequency - Inverse Document Frequency (Tf-Idf)*

Algoritma Term Frequency – Inverse Document Frequency (TF-IDF) merupakan salah satu algoritma yang sangat populer untuk melakukan klasifikasi teks yang dapat digunakan untuk menganalisis hubungan antara sebuah kalimat dengan sekumpulan dokumen. Metode *TF-IDF* merupakan metode untuk menentukan frekuensi relatif dari setiap kata-kata atau token-token yang mana setiap kata tersebut akan diberikan pembobotan berupa nilai berdasarkan penting atau tidaknya suatu kata dalam dokumen berdasarkan jumlah kemunculan kata dalam suatu dokumen dan mengukur kata-kata tersebut terhadap keseluruhan dokumen yang ada. Metode *TF-IDF* ini merupakan gabungan dari dua dua konsep untuk penghitungan bobot yaitu *Tf (term frequency)* dan *idf (inverse document frequency)* [34].

$$Idf = \log\left(\frac{N}{df}\right) \quad (4)$$

2.17 Aplikasi

Mengemukakan bahwa Aplikasi adalah *software* yang dibuat oleh suatu perusahaan komputer untuk mengerjakan tugas-tugas tertentu, misalnya *Ms.World, Ms.Excel* [35].

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. Pengertian aplikasi secara umum adalah alat terapan yang difungsikan secara khusus dan terpadu sesuai kemampuan yang dimilikinya, aplikasi merupakan suatu perangkat komputer yang siap pakai bagi *user* [36].

Aplikasi dapat diartikan sebagai suatu program berbentuk perangkat lunak yang berjalan pada suatu sistem tertentu yang berguna untuk membantu berbagai kegiatan yang dilakukan oleh manusia [37].

2.18 *UML(Unified Modeling Language)*

Unified Modeling Language merupakan salah satu metode pemodelan visual yang digunakan dalam perancangan dan pembuatan sebuah *software* yang berorientasikan pada objek [38].

Pada perkembangan teknik pemograman berorientasi objek, muncullah sebuah standarisasi bahasa pemodelan untuk pembangunan perangkat lunak yang dibangun dengan menggunakan teknik pemograman berorientasi objek yaitu *Unified Modeling Language (UML)* [39].

UML merupakan bahasa visual untuk pemodelan dan komunikasi mengenai sebuah sistem dengan menggunakan diagram dan teks-teks pendukung [40].

2.18.1 *Use Case Diagram*

Use Case Diagram merupakan diagram yang harus dibuat pertama kali saat pemodelan perangkat lunak berorientasi objek dilakukan [18].

Use case atau *diagram use case* merupakan pemodelan untuk kelakuan (*behavior*) sistem informasi yang akan dibuat [21].

Use Case merupakan langkah pertama dalam memodelkan sebuah sistem. *Use Case* merupakan pemodelan untuk kebutuhan sebuah sistem fungsional, setiap *Use Case* digambarkan sebagai kunci dari suatu skenario yang dilakukan oleh aktor dan diringkas dalam sebuah batas sistem, setiap *Use Case* dihubungkan dengan sebuah garis notasi [21].

2.18.2 *Class Diagram*

Class Diagram dibuat setelah diagram *use case* dibuat terlebih dahulu. Pada diagram ini harus menjelaskan hubungan apa saja yang terjadi antara suatu objek dengan objek lainnya sehingga terbentuklah suatu sistem aplikasi [22].

Diagram kelas atau *class diagram* menggambarkan struktur sistem dari segini pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem [19].

Class diagram adalah salah satu pemodelan yang cukup penting dalam *UML*, fungsinya adalah untuk membuat sebuat *logical models* dari sebuah sistem [21].

2.18.3 *Activitiy Diagram*

Activity Diagram menggambarkan*workflow* (aliran kerja) atau aktifitas dari sebuah sistem atau proses bisnis [22].

Diagram aktivitas atau *activity diagram* menggambarkan *workflow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis atau menu yang ada pada perangkat lunak [19].

Activity diagram merupakan pemodelan yang menggambarkan sebuah sistem kerja dari sebuah objek atau sebuah sistem, sebuah *activity diagram* digambarkan dengan sebuah alur secara terstruktur proses kerja dari *use case* yang sedang diproses dari titik awal sampai titik akhir, setiap aktivitas digambarkan dengan notasi-notasi sesuai fungsinya [21].

2.18.4 *Sequence Diagram*

Sequence diagram menggambarkan kelakuan objek pada *use case* dengan mendeskripsikan waktu hidup objekdan *massage* yang dikirimkan dan diterima antar objek. Oleh karena itu untuk menggambar diagram sekuen maka harus diketahui objek-objek yang terlibat dalam sebuah *use case* beserta metode-metode yang dimiliki kelas yang diinstansiasi menjadiobjek itu". Membuat diagram sekuen juga dibutuhkan untuk melihatskenario yang ada pada *use case* [23].

Sequence diagram merupakan *UML* yang menggambarkan interaksi antar objek di dalam dan disekitar sistem, termasuk pengguna, *display*, dan sebagainya berupa *message* yang digambarkan terhadap waktu [19].

Sequence diagram merupakan interaksi object dalam sistem, dapat dilihat pada gambar dibawah ini object-object yang ada di sistem [24].

2.19 *PHP (Hypertext Preprocessor)*

Hypertext Markup Language (*HTML*) adalah bahasa yang digunakan untuk menulis halaman *web*. *HTML* merupakan pengembangan dari standar. Pemformatan dokumen *teks* yaitu *Standard Generalized Markup Language* (*SGML*). *HTML* sebenaranya adalah dokumen *ASCII* atau *teks* biasa, yang dirancang untuk tidak tergantung pada stuatu sistem operasi tertentu [25].

PHP merupakan singkatan dari *Hypertext Preprocessor*. Ia merupakan bahasa berbentuk skrip yang ditempatkan dalam *server* dan diproses di *server*. Hasilnyalah yang dikirimkan ke klien, tempat pemakai menggunakan *browser* [26].

PHP adalah bahasa pemrograman untuk dijalankan melalui halaman *web*, umumnya digunakan untuk mengolah informasi di *internet* [27].

2.20 *MySQL(My Structured Query Language)*

MySQL adalah sebuah perangkat lunak sistem manajemen basis data *SQL (database management system)* atau *DBMS* yang *multithread, multi-user*, dengan sekitar 6 juta instalasi di seluruh dunia. *MySQL AB* membuat *MySQL* tersedia sebagai perangkat lunak gratis dibawah lisensi *General Public License (GPL)*, tetapi mereka juga menjual dibawah lisensi komersial untuk kasus-kasus dimana penggunaannya tidak cocok dengan penggunaan *GPL* [28].

MySQL merupakan singkatan dari *My Structured Query Language* adalah sebuah program *database server* yang mampu menerima dan mengirimkan data dengan sangat cepat serta multi *user*, sedangkan *SQL* adalah bahasa perintah dasar yang ada didalamnya [29].

2.21 *XAMPP*

XAMPP merupakan paket web server berbasis *open source* yang dapat dipasang pada beberapa sistem operasi yang ada *(Windows, Linux, dan Mac OS)* (Riyanto dan Slamet, 2020). Menurut Kurniawan (2019), *XAMPP* merupakan *tool* yang menyediakan paket perangkat lunak ke dalam satu buah paket. Dengan menginstal *XAMPP* maka tidak perlu lagi melakukan instalasi dan konfigurasi *web*

server apache, PHP, dan *MySQL* secara manual. *XAMPP* akan menginstalasi dan mengkonfigurasikan secara otomatis.[29]

2.22 *Confusion Matrix*

Pengukuran terhadap kinerja suatu sistem klasifikasi merupakan hal yang penting. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. *Confusion Matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

Berdasarkan jumlah keluaran kelasnya, sistem klasifikasi dapat dibagi menjadi 4 jenis yaitu klasifikasi binary, multiclass, multi-label dan hierarchical. Pada klasifikasi binary, data masukan dikelompokkan ke dalam salah satu dari dua kelas. Jenis klasifikasi ini merupakan bentuk klasifikasi yang paling sederhana dan banyak digunakan. Contoh penggunaannya antara lain dalam sistem yang melakukan deteksi orang atau bukan, sistem deteksi kendaraan atau bukan, dan sistem deteksi pergerakan atau bukan.

Sementara itu, pada bentuk klasifikasi multi-class, data masukan diklasifikasikan menjadi beberapa kelas. Sebagai contoh sistem yang dapat mengklasifikasikan jenis kendaraan seperti sepeda, sepeda motor, mobil, bus, truk, dan sebagainya. Bentuk klasifikasi multi-label pada dasarnya sama dengan multi-class dimana data dikelompokkan menjadi beberapa kelas, namun pada klasifikasi multi-label, data dapat dimasukkan dalam beberapa kelas sekaligus. Bentuk klasifikasi yang terakhir adalah hierarchical. Data masukan dikelompokkan

menjadi beberapa kelas, namun kelas tersebut dapat dikelompokkan kembali menjadi kelas-kelas yang lebih sederhana secara hirarkis. Contohnya dalam penelitian ini, arah pergerakan dikelompokkan menjadi 12 arah pergerakan yang tentunya dapat disederhanakan menjadi 4 arah.

Pada pengukuran kinerja menggunakan *confusion matrix*, terdapat 4 istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive (TP), True Negative(TN), False Positive (FP)* dan *False Negative (FN). Nilai True Negative (TN)* merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive (TP)* merupakan data positif yang terdeteksi benar. *False Negative (FN)* merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif.

Pada jenis klasifikasi *binary* yang hanya memiliki 2 keluaran kelas, *confusion matrix* dapat disajikan seperti pada Tabel 2.1.

Tabel 2.1 Tabel *Confusion Matrix*

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (True Positive)	FN (False Negative)
Negatif	FP (False Positive)	TN (True Negative)

Berdasarkan nilai *True Negative (TN), False Positive (FP), False Negative (FN),* dan *True Positive (TP)* dapat diperoleh nilai akurasi, presisi dan *recall*. Nilai

akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan 2.1. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Nilai presisi dapat diperoleh dengan Persamaan (6), nilai *recall* diperoleh dengan persamaan (7). Nilai akurasi diperoleh dengan Persamaan (5), Nilai *eror rate* diperoleh dengan Persamaan (8). [46]

$$Akurasi = \frac{TP+\mathrm{TN}}{TP+TN+FP+FN} \times 100\% \quad (5)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

$$Eror\ Rate = \frac{FP+FN}{TP+TN+FP+FN} \times 100\% \quad (8)$$

dimana:

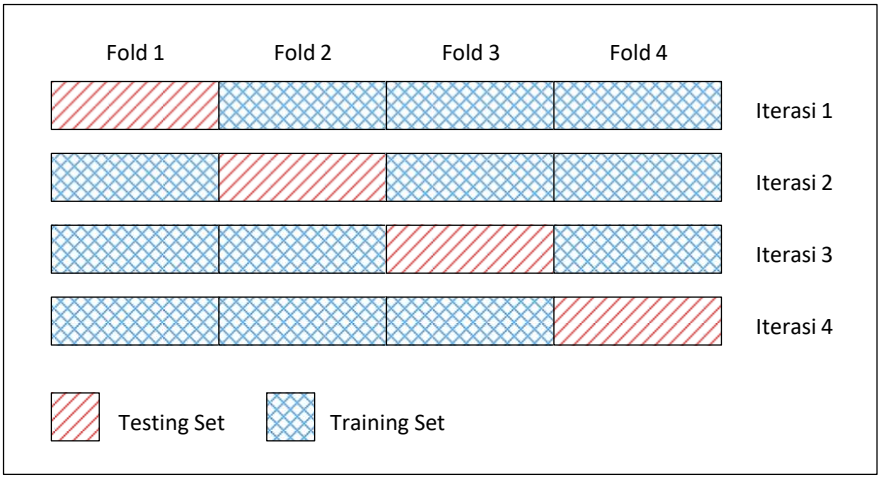
True Positive (TP) : Merupakan data positif yang diprediksi benar
True Negative (TN) : Merupakan data negatif yang diprediksi benar.
False Postive (FP) : Merupakan data negatif namun diprediksi sebagai data positif.

False Negative (FN) : Merupakan data positif namun diprediksi sebagai data negative.

2.23 *K-Fold Cross Validation*

K-fold cross validation adalah teknik yang dapat digunakan apabila memiliki jumlah data yang terbatas (jumlah *instance* tidak banyak). *K-fold cross validation* merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak. *K-fold cross validation* diawali dengan membagi data sejumlah *n-fold* yang diinginkan. Dalam proses *cross validation* data akan dibagi dalam n buah partisi dengan ukuran yang sama D1,D2,D3..Dn selanjutnya proses *testing* dan *training* dilakukan sebanyak n kali. Dalam iterasi ke-I partisi akan menjadi data *testing* dan sisanya akan menjadi data *training*. Untuk penggunaan jumlah *fold* terbaik untuk uji validitas, dianjurkan menggunakan 10-*fold cross validation* dalam model [47].

K-fold cross validation yang sering digunakan adalah 10- *fold cross validation.* Manfaat menggunakan *k-fold cross-validation* adalah setiap *record* muncul dalam set tes tepat satu kali, kekurangannya adalah bahwa tugas validasi yang diperlukan dibuat lebih sulit [48]. Berikut adalah skema dari *K-Fold cross validation* :



Gambar 2.5. Pembagian data pada metode 10-*Fold Cross Validation*

2.24 *Flowchart*

Flowchart merupakan gambaran secara grafik dari langkah-langkah dan urutan-urutan prosedur dari suatu program yang dibuat. *Flowchart* menggambarkan urutan logika dari suatu prosedur pemecahan masalah, sehingga *flowchart* merupakan langkah-langkah penyelesaian masalah yang dituliskan dalam simbol-simbol tertentu [34].

Flowchart adalah cara untuk menjelaskan tahap-tahap pemecahan masalah dengan merepresentasikan simbol-simbol tertentu yang mudah dipahamii, mudah digunakan dan standar [35].

Tujuan penggunaan *flowchart* adalah untuk menggambarkan suatu tahapan penyelesaian masalah secara sederhana, terurai dan rapi dengan menggunakan simbol-simbol yang standar yang dapat di mengerti oleh programmer [35].

2.25 Penelitian Terkait

Tabel 2.3. Penelitian Terkait

No	Penulisan dan Tahun	Judul	Metode	Hasil
1.	(Syarifuddinn, M. 2020).	Analisis Sentimen Opini Publik Mengenai Covid-19 Pada *Twitter* Menggunakan Metode *Naive Bayes* Dan Knn	*Naïve Bayes*	Berdasarkan hasil penelitian, dapat ditarik beberapa kesimpulan diantaranya adalah COVID19 merupakan penyakit yang tidak sedikit meresahkan masyarakat umum, apalagi ditambah dengan kicauan di media social tentang COVID-19 yang semakin merabak, salah satunya adalah *twitter* yang menambah masyarakat katakutan.
2.	(Samsir, S.,Ambiyar, A., Verawardina, U.,Edi, F., & Watriantho s, R. 2021).	Analisis Sentimen Pembelajaran Daring Pada *Twitter* di Masa Pandemi COVID-19 Menggunakan Metode *Naive Bayes*.	*Naïve Bayes*	Penelitian ini menerapkan algoritma Naïve Bayes untuk klasifikasi sentimen. Analisis sentimen dalam study ini menggunakan data *Twitter* dengan kata kunci ‘pembelajaran daring’, ‘kuliah’, ‘belajar’, ‘online’, ‘daring’, dan tagar #BelajarDariRumah yang difilter dengan kata kunci ‘online’ dan ‘rumah’ pada twit-twit dalam bahasa Indonesia pada minggu pertama November 2020. Penelitian pada periode tersebut menunjukkan 30% sentimen positif, 69% sentimen negatif, dan 1% netral.
3.	(Nugroho, A. 2018).	Analisis Sentimen Pada Media Sosial *Twitter* Menggunakan *Naive Bayes Classifier* Dengan Ekstrasi Fitur *N-Gram*.	*Naive Bayes*	Berdasarkan dari hasil penelitian terhadap model algoritma *naïve bayes* yang di kombinasikan dengan ekstraksi fitur n-gram terhadap analisa perseberan tweet terhadap isu kenaikan tarif dasar listrik, maka dapat ditarik kesimpulan bahwa algoritma naïve bayes mampu mencapai tingkat akurasi 89,67%, dan pengaruh ekstraksi fitur n-gram yang diterapkan dapat meningkatkan nilai akurasi pada algoritma *naïve bayes* sekitar 2,33%, yaitu menjadi

				92,00% dalam pengklasifikasian data tweet.
4.	Mahardhika, Y. S., & Zuliarso, E. (2018).	Analisis Sentimen Terhadap Pemerintahan Joko Widodo Pada Media Sosial *Twitter* Menggunakan Algoritma *Naives Bayes Classifier*	*Naive Bayes*	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.
5.	Pintoko, B. M., & Lhaksmana, K. M. (2018).	Analisis Sentimen Jasa Transportasi Online pada *Twitter* Menggunakan Metode *Naive Bayes Classifier*.	*Naive Bayes*	Berdasarkan hasil analisis yang dilakukan terkait sentimen terhadap jasa transportasi online dengan menggunakan metode *Naive Bayes* untuk menentukan kelas prediksi, didapatkan hasil sentimen positif sebesar 88.60% dan sentimen negatif sebesar 11.40% dengan akurasi sebesar 86.80%. Hasil menunjukkan tingkat sentimen positif lebih besar dibandingkan dengan tingkat sentimen negatif. Dengan hasil penelitian ini bisa dilihat bahwa banyak masyarakat yang merasa puas dengan layanan jasa transportasi online di Indonesia berdasarkan data dari *Twitter*.
6.	Sudiantoro, A. V., & Zuliarso, E. (2018).	Analisis Sentimen *Twitter* Menggunakan Text Mining Dengan Algoritma *Naive Bayes*	*Naive Bayes*	Algoritma *Naive Bayes Classifier* sangat efektif untuk digunakan sebagai proses klasifikasi tweet yang dibutuhkan dalam sistem analisis sentiment ini dimana nilai yang di dapatkan dalam pengujian sampai 84%.
7.	Syarifuddinn, M. (2020).	Analisis Sentimen Opini Publik Terhadap Efek Psbb Pada *Twitter*	*Naive Bayes*	Dalam studi ini, upaya dilakukan untuk mengklasifikasikan sentimen analisis dari tweet di twitter tentang efek PSBB bagi masyarakat di Indonesia. Untuk merangkum pandangan masyarakat ini teknik

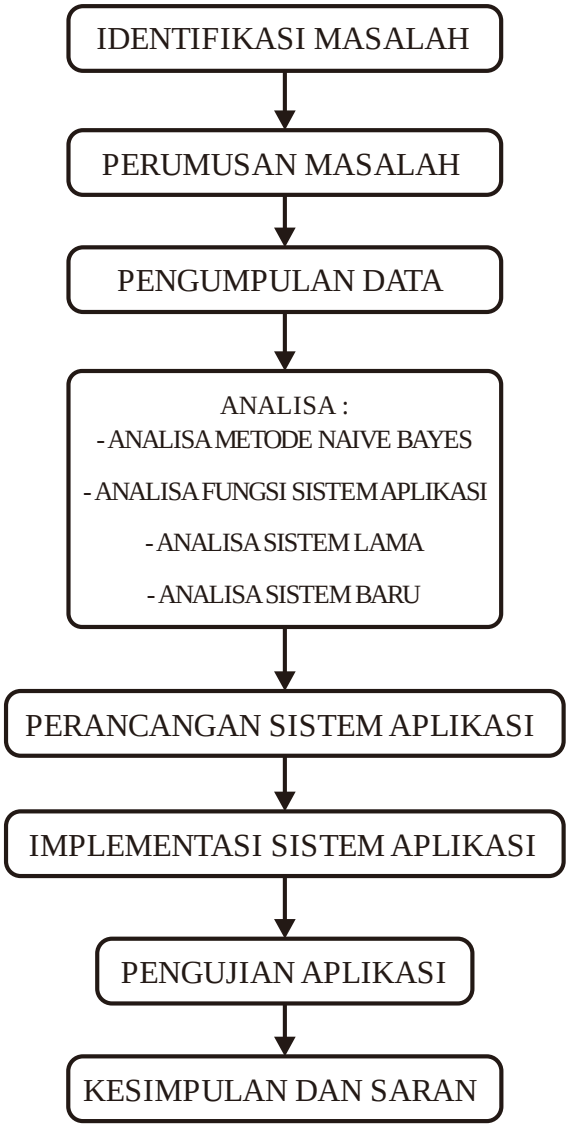
		Dengan Algoritma *Decision Tree-Knn-Naive Bayes*.		penambangan teks digunakan, dan penggalian data menggunakan tiga algoritma yang berbeda yaitu Decision Tree, K-NN, dan *Naive Bayes*. Tiga algoritma memprediksi label pada dataset. Dari hasil pengujian secara keseluruhan algoritma *Decision Tree* memiliki nilai tertinggi dari nilai *accuracy* dan *recall*.
8.	Naraswati, N. P. G., Nooraeni, R., Rosmilda, D. C., Desinta, D., Khairi, F., & Damaiyanti, R. (2021).	Analisis Sentimen Publik dari *Twitter* Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan *Naive Bayes Classification*	*Naive Bayes*	Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa masyarakat lebih banyak memberikan sentimen negatif terhadap kebijakan penanganan COVID-19 khususnya PSBB, wajib masker, dan jam malam yaitu sebesar 72,72 persen. Pada sentimen positif, tiga kata dengan frekuensi kemunculan terbanyak yaitu demo, jakarta, dan kerja.
9.	Ruhyana, N. (2019).	Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil/Genap Pada *Twitter* Dengan Metode Klasifikasi *Naive Bayes*.	*Naive Bayes*	Dari hasil penelitian untuk mencari sentiment dalam hal penerapan sistem ganjil genap untuk tanggapan pada pengguna twitter dan dilakukan analisa sentiment dengan label postif dan negatif dengan metode klasifikasi *naive bayes* pengujian dilakukan dengan data training dan data testing yang di ambil dari *twitter* menggunakan software *rapid miner* sehingga memudahkan memberikan hasil dari penelitian tersebut dengan hasil *accuracy* 86,67%, presisi 71,43% dan recall 80,00%.
10.	Ramadhan, D. A., & Setiawan, E. B. (2019).	Analisis Sentimen Program Acara Di Sctv Pada *Twitter* Menggunakan	*Naive Bayes*	Berdasarkan hasil penelitian untuk Analisis Sentimen Program Acara di SCTV pada *Twitter* menggunakan Metode *Naive Bayes* dan *Support Vector* Machine yang telah dilakukan mendapatkan Labelling

		Metode *Naive Bayes* Dan *Support Vector Machine*		terbaik didapatkan dengan cara *Labelling* Otomatis sedangkan akurasi terbaik dari Metode *Naive Bayes* dan *Support Vector Machine* adalah Metode *Support Vector* Machine dengan Seluruh Program Acara didapatkan hasil akurasi 88,57% berikut perkategori yang diperoleh Berita didapatkan hasil akurasi 79,81%, ISSN : 2355-9365 e-*Proceeding of Engineering* : Vol.6, No.2 Agustus 2019 | Page 9742 *Entertaiment* didapatkan hasil akurasi 89,80%, Sinetron didapatkan hasil akurasi 73,68% dan FTV didapatkan hasil akurasi 87,74%.

BAB III

METODOLOGI PENELITIAN

Penelitian ini dilakukan dengan melaksanakan tahapan demi tahapan yang berhubungan. Tahapan-tahapan tersebut dijabarkan dalam metode penelitian. Metode penelitian diuraikan kedalam bentuk skema yang jelas, teratur, dan sistematis. Berikut tahapan-tahapan penelitian dapat dilihat pada gambar 3.1



Gambar 3.1 Metodologi Penelitian

Penjelasan dari tahapan-tahapan penelitian pada gambar 3.1 dapat dilihat pada penjelasan dibawah ini:

3.1 Identifikasi Masalah

Identifikasi masalah merupakan tahapan awal yang dilakukan dalam penelitian ini bertujuan untuk mengamati masalah yang terjadi yaitu mengklasifikasikan komentar pengguna *twitter* terhadap bakal calon presiden Indonesia di pilpres pada tahun 2024. Langkah untuk dapat memahami masalah yang telah ditentukan ruang lingkup masalah atau batasannya. Dengan menganalisa masalah yang telah ditemukan tersebut, maka diharapkan masalahnya dapat dipahami dengan baik. Maka dari itu diperlukan sebuah sistem baru, yaitu dengan membuat suatu aplikasi yang mampu mengklasifikasikan komentar pengguna *twitter* pada masalah terhadap bakal calon presiden Indonesia di pilpres pada tahun 2024.

3.2 Perumusan Masalah

Berdasarkan hasil dari tahapan identifikasi masalah sebelumnya, maka tahapan selanjutnya adalah perumusan masalah. Pada tahapan perumusan masalah akan dirumuskan masalah yang dianggap sebagai penelitian dalam tugas akhir ini. Permasalahan-permasalahan yang menjadi rumusan masalah dalam penelitian ini didapatkan dari penelitian, terkait data pengamatan pendahuluan sebelumnya. Solusi yang didapatkan pada tahapan perumusan masalah yang akan menjadi judul penelitian tugas akhir ini yaitu " Aplikasi Analisis Sentimen *Public Twitter* Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Metode *Naive Bayes* Berbasis web*".* Penggunaan metode *Naïve Bayes* dalam Aplikasi ini didasari

oleh keunggulan metode *Naïve Bayes* yaitu *Naive Bayes* merupakan suatu bentuk klasifikasi data dengan menggunakan metode probabilitas dan statistik.

3.3 Pengumpulan Data

Pengumpulan data adalah tahapan-tahapan yang bertujuan dalam memperoleh data-data informasi yang berhubungan dengan penelitian tugas akhir ini. Pada tahapan pengumpulan data ini juga berguna untuk mengumpulkan semua kebutuhan data yang akan diproses nantinya menggunakan metode *Naïve Bayes*. Dalam pengumpulan data ini penulis mendapatkannya dari komentar pengguna *twitter* terhadap bakal calon presiden Indonesia di pilpres pada tahun 2024.

3.4 Analisa

Tahapan selanjutnya adalah melakukan analisa metode sistem dari penelitian Skripsi ini. Adapun tahapan analisa dalam penelitian Skripsi ini adalah sebagai berikut:

3.4.1 Analisa Metode *Naïve Bayes*

Tahapan ini yaitu menganalisa cara kerja dari metode *Naïve Bayes*, dengan menganalisa metode ini, nantinya bisa di terapkan pada aplikasi yang akan di bangun, karena permasalahan yang ditemukan oleh penulis dalam penelitian ini yaitu dapat mengetahui bagaimana pendapat public terhadap bakal calon presiden Indonesia di pilpres pada tahun 2024.

3.4.2 Analisa Fungsi Sistem Aplikasi

Setelah melakukan tahapan analisis terhadap metode *Naïve Bayes*, maka selanjutnya adalah analisa fungsional sistem yang akan dibangun. Adapun tahapan-tahapan analisis fungsional yaitu dalam pembuatan *flowchart.*

3.4.3 Analisa Sistem Lama

Analisa sistem lama diperlukan untuk mengetahui prosedur-prosedur awal dalam kasus yang sedang diteliti, agar dapat dibuatkan sistem baru yang diharapkan akan menyempurnakan sistem lama.

3.4.4 Analisa Sistem Baru

Setelah menganalisa sistem lama, maka tahapan selanjutnya dengan menganalisa sistem yang baru. Analisa dalam pembuatan sistem ini menggunakan metode *Naïve Bayes* serta penggunaan U*nified Modelling Language* untuk menganalisa kebutuhan sistem.

3.5 Perancangan Sistem Aplikasi

Setelah tahapan analisa selesai dilakukan, maka tahapan selanjutnya adalah perancangan sistem. Tahapan perancangan sistem terdiri dari :

3.5.1 Perancangan Basis Data

Setelah dilakukannya analisa sistem yang akan dibuat, maka tahap berikutnya ialah analisa dan perancangan basis data yang kita lakukan untuk melengkapi komponen dalam pembuatan sistem.

3.5.2 Perancangan Struktur Menu

Rancangan struktur menu ini kita perlu untuk memberikan gambaran terhadap menu-menu atau fitur pada sistem yang akan dibuat.

3.5.3 Perancangan Antar Muka (*Interface)*

Dalam mempermudah komunikasi antara sistem dengan pengguna, maka perlu di rancang antar muka (*interface).* Dalam perancangan *interface* hal

terpenting yang harus dilakukan ialah bagaimana menciptakan tampilan yang baik dan mudah dimengerti oleh pengguna.

3.6 Implementasi Sistem Aplikasi

Beberapa komponen pendukung yang memiliki peran yang sangat penting dalam implementasi sistem diantaranya adalah perangkat keras (*hardware*) dan perangkat lunak (*software*). Adapun spesifikasi dari perangkat keras (*hardware*) dan perangkat lunak (*software*) yang di gunakan sebagai berikut :

1. Perangkat keras (*hardware*), antara lain :

   Prosesor : *Intel Core i5-8250U*

   Memory (RAM) : 8192 MB

   System type : 64-bit *Operating System*

   Harddisk : HDD 1 TB

2. Perangkat Lunak (*Software*), antara lain :

   Sistem operasi : *Windows* 10

3.7 Pengujian Aplikasi

Pengujian sistem ini dilakukan dengan cara menggunakan *Black Box* berikut pengertiannya :

3.7.1 *Black Box Testing*

Pengujian ini memakai teknik *black box*, dimana yang diuji adalah fungsi-fungsi yang digunakan untuk membuat sebuah *website analisis sentiment*. Pengujian ini dilakukan dengan membuka menu-menu yang mengiakan fungsi-fungsi dari sentimen analisis. Pengujian ini juga memastikan *website analisis sentiment* berjalan dengan baik [43].

Salah satu metode pengujian yang berfokus pada spesifikasi fungsionalitas dari perangkat lunak disebut *Black Box Testing*. Pengujian ini memberikan gambaran atas sekumpulan kondisi masukan dan melakukan pengujian pada uraian fungsional program. Untuk menguji kesalahan yang tidak dapat dicakup oleh *White Box Testing*, maka solusi lainnya dapat menggunakan *Black Box Testing. Black Box Testing* digunakan untuk mendeteksi permasalahan berikut [44] :

a. Fungsi yang salah atau hilang.
b. Kesalahan pada *interface.*
c. Kesalahan struktur data dan basis data.
d. Kesalahan fungsi.
e. Kesalahan deklarasi dan terminasi.

Apabila terjadi *error* atau tidak sesuai dengan tujuan yang diharapkan maka dilakukan penganalisa sistem kembali hingga tidak ditemukan *error.*

3.8 Kesimpulan dan Saran

Tahapan terakhir adalah menarik kesimpulan dari hasil penelitian terkait penggunaan metode *Naive Bayes* pada aplikasi ini, Pada tahapan ini juga berisikan saran peneliti bagi pembaca untuk melakukan pengembangan terhadap penelitian ini kedepannya.