**SENTIMEN ANALISIS PILPRES 2024 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN *NAÏVE BAYES CLASSIFIER***

**SKRIPSI**

**Karya tulis sebagai salah satu syarat**

**untuk memperoleh gelar Tingkat Sarjana**

**Oleh**

**MUCHAMMAD FAHD ISHAMUDDIN**

**NPM : 411550050180048**



**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS LANGLANGBUANA**

**2023**

LEMBAR PENGESAHAN

**SENTIMEN ANALISIS PILPRES 2024 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER**

**Oleh**

**MUCHAMMAD FAHD ISHAMUDDIN**

**NPM : 411550050180048**

**untuk memperoleh gelar Tingkat Sarjana dari**

**Program Studi Teknik Informatika**

**Universitas Langlangbuana**

Menyetujui

Pembimbing 1 Pembimbing 2

(Prof. Dr. Pembimbing 1, M.T.) (Dr. Pembimbing 2, S.T., M.T.)

NIDN: <NIDN> NIDN: <NIDN>

Mengetahui

Dekan Ketua Program Studi

Fakultas Teknik Teknik Informatika

(Dr. Dekan Fakultas Teknik, M.T.) (Dr. Kaprodi TIF, M.Si)

NIDN: <NIDN> NIDN: <NIDN>

LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR

Saya yang bertanda tangan dibawah ini :

Nama : MUCHAMMAD FAHD ISHAMUDDIN

NPM : 411550050180048

Judul Tugas Akhir : SENTIMEN ANALISIS PILPRES 2024 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER

Menyatakan dengan sebenarnya bahwa penulisan Tugas Akhir ini berdasarkan hasil penelitian, pemikiran dan pemaparan asli dari saya sendiri, baik untuk naskah laporan maupun kegiatan pembangunan aplikasi yang tercantum sebagai bagian dari Tugas Akhir ini. Jika terdapat karya orang lain, saya akan mencantumkan sumber yang jelas.

Demikian 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 dan sanksi lain sesuai dengan peraturan yang berlaku di Universitas Langlangbuana. Demikian pernyataan ini saya buat dalam  keadaan sadar tanpa paksaan dari pihak manapun.

Bandung, 14 April 2023

Yang membuat pernyataan,

\_\_\_MUCHAMMAD FAHD ISHAMUDDIN\_\_\_

NPM : 411550050180048

*Dipersembahkan kepada kedua orang tuaku dan guru-guruku, \*

*semoga Allah selalu memberkati mereka.*

KATA PENGANTAR

Penulis sangat berterima kasih pada Prof. Dr. Marsilam Hutabarat, M.T. Dam Dr. Jaya Sumarna, S.T., M.T. sebagai Pembimbing, atas segala saran, bimbingan dan nasehatnya selama penelitian berlangsung dan selama penulisan Skripsi ini.

Terima kasih disampaikan kepada Kementerian Riset Teknologi dan Pendidikan Tinggi atas bantuan Beasiswa Pendidikan Pascasarjana (BPPs) yang diterima selama pendidikan program Tingkat Sarjana ini.

(dan seterusnya)

ABSTRAK

Pengiriman barang dalam sebuah ekspedisi idealnya dilakukan oleh para kurir dengan membawa sejumlah barang ke sejumlah tujuan, dengan tujuan-tujuan yang dipilih memiliki jalur yang kurang lebih searah. Namun dalam kenyataannya, tidak selalu terpenuhi, sehingga kadang seorang kurir mendapatkan suatu rute pengiriman dengan jalur yang terpencar-pencar. Tujuan-tujuan dengan jalur yang terpencar ini perlu dicarikan alternatif rute yang paling efektif sehingga bisa menghemat biaya bahan bakar serta mempercepat waktu pengiriman. Untuk keperluan tersebut, berbagai kemungkinan jalur tersebut dimasukkan ke dalam struktrur data graf, dan dicari rute-rute yang paling efektif untuk ditempuh. Rute-rute yang muncul biasanya bersifat acak, karena itu berbagai kemungkinan rute akan dimaksukkan ke dalam sistem komputasi setiap saat pencarian rute ini dibutuhkan. Permodelan dan simulasi pencarian rute ini dibuat dengan memanfaatkan kepustakaan API grafis OpenGL, dengan bahasa pemrograman C++, serta dimodelkan dengan pendekatan berorientasi objek menggunakan notasi UML.

Kata Kunci: *Machine Learning, Naïve bayes classifier, Pilpres 2024, Prabowo, Ganjar Pranowo, Anies Baswedan.*

ABSTRACT

Delivery of goods on an expedition is ideally carried out by couriers carrying a number of goods to several destinations, with the selected destinations having a likely similar route. But in reality, this ideal condition is not always fulfilled, so sometimes a courier gets a delivery route with scattered routes. Within the destinations of these scattered routes, it’s need to be find the most effective alternative route so that it can save fuel costs and speed up delivery times. For this purpose, the various possible paths are organized in the graph data structure, and the most effective routes are searched for. The routes are usually random, therefore various possible routes will be input into the computation system whenever this route search is needed. This route search modeling and simulation application was created using the OpenGL graphical API library, using the C ++ programming language, and modeled with an object-oriented approach using UML notation.

Keywords: *Machine Learning, Naïve bayes classifier, Pilpres 2024, Prabowo, Ganjar Pranowo, Anies Baswedan.*

DAFTAR ISI

[LEMBAR PENGESAHAN ii](#_Toc134729774)

[LEMBAR PERNYATAAN KEASLIAN TUGAS AKHIR iii](#_Toc134729775)

[KATA PENGANTAR v](#_Toc134729776)

[ABSTRAK vi](#_Toc134729777)

[ABSTRACT vii](#_Toc134729778)

[DAFTAR ISI viii](#_Toc134729779)

[DAFTAR GAMBAR xi](#_Toc134729780)

[DAFTAR TABEL xii](#_Toc134729781)

[DAFTAR SINGKATAN DAN ISTILAH xiii](#_Toc134729782)

[DAFTAR SIMBOL xiv](#_Toc134729783)

[DAFTAR LAMPIRAN xv](#_Toc134729784)

[BAB I PENDAHULUAN I-1](#_Toc134729785)

[I.1 Latar Belakang I-1](#_Toc134729786)

[I.2 Rumusan Masalah I-3](#_Toc134729787)

[I.3 Batasan Masalah I-3](#_Toc134729788)

[I.4 Tujuan Penelitian I-3](#_Toc134729789)

[I.5 Keluaran Penelitian I-3](#_Toc134729790)

[I.6 Sistematika Penulisan I-4](#_Toc134729791)

[BAB II LANDASAN TEORI II-1](#_Toc134729792)

[II.1 Teori Terkait Permasalahan II-1](#_Toc134729793)

[II.1.1 *Sentiment Analysis* II-1](#_Toc134729794)

[II.1.2 *Naïve Bayes Classifier* II-1](#_Toc134729795)

[II.1.3 Machine Learning II-2](#_Toc134729796)

[II.2 Teori Pendukung II-3](#_Toc134729797)

[II.2.1 Python II-3](#_Toc134729798)

[II.2.2 Text Mining II-3](#_Toc134729799)

[II.2.3 Jupyter Notebook II-4](#_Toc134729800)

[II.2.4 Preprocessing II-4](#_Toc134729801)

[II.2.5 Monte Carlo Cross Validation II-7](#_Toc134729802)

[II.2.6 Cross Industry Standard Process for Data Mining (CRISP-DM) II-7](#_Toc134729803)

[II.3 Penelitian-penelitian Terdahulu II-8](#_Toc134729804)

[II.3.1 Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit II-8](#_Toc134729805)

[II.3.2 Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u II-8](#_Toc134729806)

[II.3.3 Sentiment Analysis Menggunakan Naïve Bayes Classifier pada Tweet Tentang Zakat II-9](#_Toc134729807)

[BAB III METODOLOGI PENELITIAN III-1](#_Toc134729808)

[III.1 Metode Penelitian III-1](#_Toc134729809)

[III.2 Metodologi Pengembangan Sistem III-1](#_Toc134729810)

[III.2.1 Cross Industry Standard Process for Data Mining III-1](#_Toc134729811)

[III.2.2 Scrapping III-1](#_Toc134729812)

[III.2.3 Pendekatan Supervised learning III-1](#_Toc134729813)

[III.2.4 Modeling III-1](#_Toc134729814)

[III.2.5 Monte Carlo Cross Validation III-1](#_Toc134729815)

[III.3 Tahapan Penelitian III-1](#_Toc134729816)

[BAB IV HASIL DAN PEMBAHASAN IV-1](#_Toc134729817)

[IV.1 Inception Phase IV-1](#_Toc134729818)

[IV.1.1 Analisis Proses Bisnis IV-1](#_Toc134729819)

[IV.1.2 Requirements IV-1](#_Toc134729820)

[IV.2 Elaboration Phase IV-1](#_Toc134729821)

[IV.2.1 Analisis IV-1](#_Toc134729822)

[IV.2.2 Design IV-1](#_Toc134729823)

[IV.3 Construction Phase IV-2](#_Toc134729824)

[IV.3.1 Implementasi IV-2](#_Toc134729825)

[IV.3.2 Pengujian IV-2](#_Toc134729826)

[IV.3.3 Deployment IV-2](#_Toc134729827)

[BAB V SIMPULAN DAN SARAN V-1](#_Toc134729828)

[V.1 Simpulan V-1](#_Toc134729829)

[V.2 Saran V-1](#_Toc134729830)

[DAFTAR PUSTAKA xv](#_Toc134729831)

DAFTAR GAMBAR

Gambar 2.1 Diagram Konektivitas Modul-modul GPS 3

Gambar 2.2 Contoh Diagram Usecase 5

Gambar 3.1 Business Usecase Sistem Yang Sedang Berjalan 25

Gambar 3.2 Business Usecase Sistem Yang Akan Dikembangkan 30

Gambar 4.1 Diagram Sekuens Use Case Authentication 60

DAFTAR TABEL

Tabel 3.1 Tabel Kebutuhan Fungsional Sistem 26

Tabel 3.2 Tabel Skenario Usecase Authencitacion 29

Tabel 4.1 Daftar Kebutuhan Tabel-tabel Database 62

Tabel 4.2 Tabel Rencana Pengujian Perangkat Lunak 70

DAFTAR SINGKATAN DAN ISTILAH

|  |  |
| --- | --- |
| **Singkatan** | **Kepanjangan** |
| RUP | *Rational Unified Process* |
| UML | *Unified Modelling Language* |
| GPS | *Global Positioning System* |
| OOSE | *Object Oriented Software Engineering* |

DAFTAR SIMBOL

1. **SImbol-simbol Dalam Diagram Usecase**

|  |  |
| --- | --- |
| **Simbol** | **Arti** |
|  | Aktor, pelaku dari sebuah usecase |
|  | Usecase, sekumpulan aksi yang membentuk suatu kesatuan lojik untuk menyelesaikan suatu kebutuhan fungsional |
|  | Asosiasi, menunjukkan siapa aktor yang bertanggung jawab atas suatu usecase |

**Sumber: Booch, 2005**

1. **Simbol-simbol Dalam Diagram Sekuens**

|  |  |
| --- | --- |
| **­** | **Arti** |
| IMG_256 | Kelas/Objek Interface/Boundary, kelas/objek yang merupakan penghubung antara aktor dengan sistem, atau menghubungkan sistem dengan sistem lain yang terkait (API) |
| IMG_256 | Kelas/Objek Controller, kelas//objek yang mengandung lojik dari sistem/usecase, library, atau kumpulan fungsi-fungsi dan prosedur yang digunakan dalam sistem. |
| IMG_256 | Kelas/Objek Etitas, kelas yang merepresentasikan data yang diolah di dalam sistem |

**Sumber: Booch, 2005**

DAFTAR LAMPIRAN

[LAMPIRAN A: STRUKTUR ORGANISASI PT CEPAT KIRIM xvi](#_Toc65825259)

[LAMPIRAN B: HASIL SURVEY PENGGUNA xvii](#_Toc65825260)

[LAMPIRAN C: KODE PROGRAM xviii](#_Toc65825261)

# PENDAHULUAN

## Latar Belakang

Indonesia merupakan negara yang memiliki bentuk pemerintahan presidensial dan demokrasi. Pemerintahan presidensial berarti kepemimpinan pada negara tersebut dipimpin oleh seorang presiden, Demokrasi berarti kekuasaan tertinggi ada di tangan rakyat sehingga yang dapat memilih siapa pemimpin pada negara tersebut. Seperti yang kita ketahui jika pemilihan presiden diadakan dengan Pemilu (Pemilihan Umum) yang bertujuan untuk menentukan eksekutif dan legislatif serta diselenggerakan oleh KPU (Komisi Pemilihan Umum).

Pemilu pertama kali dilaksanakan pada 29 September 1955 untuk memilih anggota DPR dilanjutkan pada 15 Desember 1955 untuk memilih anggota Dewan Konstituante, pada saat ini Indonesia masih dipimpin oleh Ir. Soekarno. Pada tahun 1967 ada SUPERSEMAR (Surat Perjanjian Sebelas Maret) yang menyatakan penyerahan kepemimpinan dari Ir. Soekarno kepada Soeharto. Pemilu pada tahun 1971, 1977, 1997 ketiga pemilu tersebut hanya digunakan untuk memilih DPR hingga akhirnya pada tahun 1999, setelah masa kepemimpinan Ir. B. J. Habibie, FREng. Diadakan pemilihan presiden (pilpres) tetapi melalui sidang paripurna MPR,dengan mencatatkan Abdurrahman Wahid (gusdur) menjadi presiden ke-4 Indonesia dan didampingi Megawati Soekarnoputri sebagai wakil presiden, hingga pada tahun 2004 merupakan pertama kalinya pemilihan presiden dilakukan secara luberjurdil (langsung, umum, bebas, rahasia, jujur dan adil) dengan memenangkan Susilo Bambang Yudhoyono sebagai presiden dan Jusuf Kalla sebagai wakil presiden. Terakhir dilaksanakn pada 2019 yang dimenangkan oleh petahan yakni Ir. H. Joko Widodo.

Hingga sekarang pemilihan presiden masih dilakukan untuk menentukan pemimpin negara Indonesia, tetapi ada yang berbeda antara zaman terdahulu dengan zaman sekarang, pada zaman sekarang kita bisa mengetahui reaksi masyarakat terhadap pemilu dan juga calon presiden yang diusung, apalagi melalui media sosial yang merupakan komponen primer manusia era modern.



**Gambar 1. 1** Grafik Pengguna Aktif Media Sosial di Indonesia (dataindonesia.id, 2022)

Pengguna media sosial di Indonesia per tahun 2022 dapat dilihat pada gambar 1.1 di atas dan menunjukkan sudah mencapai 191 juta jiwa dan total masyarakat Indonesia ialah 275 juta jiwa, berarti sudah mencapai 69% jiwa di Indonesia menggunakan sosial media hal ini bisa menjadi alasan media sosial adalah media yang bisa menjadi pertimbangan dalam melihat elektabilitas maupun perspektif masyarakat terhadap Pemilu, hal tersebut merupakan tugas dari juru kampanye atau tim pemenangan dari suatu partai atau calon yang ingin manyalonkan dirinya di Pemilu 2024 untuk menaikkan elektabilitas partai atau calon yang ia dukung. Media sosial yang massif dan menjadi persebaran opini dari masyarakat Indonesia ialah twitter, karena banyaknya pendapat yang disampaikan dari warganet Indonesia akan menghasilkan berbagai macam reaksi, maka dapat dilakukan sentiment analisis untuk mendapatkan nada emosional tweet warganet Indonesia, algoritma yang banyak digunakan untuk mendapatkan sentiment ialah naïve bayes, dengan menghitung probabilitas dengan dasar bayes theorem.

Setelah mendalami permasalahan tersebut, maka penulis tertarik untuk menganalisa data dari twitter tentang Pilpres 2024 dan menuangkannya pada penelitian Skripsi yang berjudul ”**SENTIMEN ANALISIS PILPRES 2024 PADA MEDIA SOSIAL TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER**”

## Rumusan Masalah

Berdasarkan dari latar belakang yang sudah ditulis, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

Bagaimana seorang juru kampanye dapat menguji tweet yang sudah diunggah dan yang sudah disediakan juru kampanye suatu calon/partai ini positif, negative ataupun netral.

## Batasan Masalah

Adapun Batasan masalah pada penelitian ini adalah sebagai berikut:

Data yang digunakan lebih dari seratus ribu (100000) data

Tweet yang didapat ialah tweet sejak januari 2022 hingga desember 2022

Output ialah data sentiment terhadap pilpres 2024 dan web sederhana pengujian sentimen

## Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka yang menjadi tujuan penelitian yang akan dilakukan antara lain:

1. Menerapkan metode naïve bayes classifier guna melakukan sentiment analisis dari social media twitter agar menjadi pertimbangan kontestan politik.
2. Mengetahui akurasi terbaik dan melakukan sentiment analysis tentang pilpres 2024 guna mengetahui bentuk opini positif, negative dan netral.

## Keluaran Penelitian

Dalam penelitian ini luaran yang dihasilkan adalah sebagai berikut:

1. Laporan Penelitian, Python Notebook dan visualisasi data
2. Web sederhana penerapan model sentiment analisis

## Sistematika Penulisan

Untuk memberikan gambaran mengenai penelitian ini, maka disusun sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN : membahas secara singkat mengenai latar belakang,

rumusan masalah, batasan masalah, tujuan penelitian, keluaran penelitian dan

sistematika penulisan.

BAB II LANDASAN TEORI : membahas tentang artikel-artikel jurnal dari karya

para peneliti sebelumnya yang berguna dalam proses pembuatan sistem.

BAB III METODOLOGI PENELITIAN : menjelaskan bagaimana urutan langkah

penyelesaian masalah berdasarkan rumusan masalah, serta penjelasan metode

penelitian yang akan digunakan.

BAB IV HASIL DAN PEMBAHASAN : tentang hasil penelitian dan hasil analisis

dari sistem.

BAB V SIMPULAN DAN SARAN : simpulan yang diperoleh dari analisis sistem

dan saran yang bermanfaat.

# LANDASAN TEORI

## Teori Terkait Permasalahan

### *Sentiment Analysis*

Analisis sentimen adalah pengumpulan pandangan orang tentang setiap peristiwa yang terjadi dalam kehidupan nyata. Dalam situasi seperti itu di mana dunia sedang melalui, memahami emosi dari orang-orang berdiri sangat penting. Skenario kubur dimana orang tidak bisa keluar dari rumah mereka menuntut eksplorasi-ing apa orang-orang benar-benar berpikir tentang keseluruhan skenario. Oleh karena itu, penulis telah merencanakan pekerjaan ini di bawah menghadapi situasi yang menuntut terutama di media social (Chakraborty,K. 2020).

Analisis sentimen adalah proses untuk mengidentifikasi dan mengenali atau mengkategorikan emosi pengguna atau pendapat untuk layanan apa pun seperti film, masalah produk, acara, atau setiap atribut adalah positif, negatif atau netral. Sumber untuk analisis ini adalah saluran komunikasi sosial yaitu situs Web yang meliputi *review*, forum diskusi, *blog*, *micro-blog*, *Twitter* dll. Bidang penelitian ini sangat populer saat ini karena data pendapatnya di mana pengguna dapat menemukan ulasannya layanan apa pun yang berguna untuk kehidupan sehari-hari mereka. Besar jumlah data opini disimpan dalam bentuk digital. Untuk topik tertentu atau pendapat analisis sentimen yang menghubungkan penambangan data bekerja dan memberikan output. (Mehta, P. and Pandya, S., 2020)

### *Naïve Bayes Classifier*

*Naïve Bayes Classifier* adalah metode klasifikasi berdasarkan teorema Bayes. Pengklasifikasi *Naïve Bayes* dikenal lebih baik daripada beberapa metode klasifikasi lainnya. Karena pertama, ciri utama dari *Naïve Bayes* adalah asumsi independensi (naif) yang sangat kuat dari setiap kondisi atau peristiwa. Kedua, modelnya simple dan mudah dibuat. Ketiga, model dapat diimplementasikan untuk set data yang besar. Dasar salah satu teorema *Naïve Bayes* yang digunakan adalah rumus Bayes sebagai berikut: (Han, Kamber, & Pei, 2012).

Metode Naïve bayes classifier berasal dari bayes theorem yang ditemukan oleh Thomas bayes pada tahun 1770. Teorema bayes adalah sebuah teorema dengan dua penafsiran berbeda. Teorema ini menyatakan seberapa jauh derajad kepercayaan subjektif harus berubah secara rasional ketika diberikan petunjuk baru. Teori ini juga berasal dari penerapan teori probabilitas.

Berikut persamaan 1 adalah teori naïve bayes:

|  |  |  |
| --- | --- | --- |
|  |  | Persamaan ( 1 ) |

**P(H|e)** = peluang kejadian **H** apabila **e** terjadi

**P(e|H)** = peluang kejadian **e** apabila **H** terjadi

**P(H)** = probabilitas kejadian (H)

**P(e)** = probabilitas (e) atau disebut prior probability. Berlaku jika (e) ≠ 0

### Machine Learning

*Machine learning* atau dalam Bahasa Indonesia dikenal dengan pembelajaran mesin adalah aplikasi dari disiplin ilmu kecerdasan buatan (*Artificial Intelligence*). Konsep dari machine learning adalah memberikan kemampuan kepada computer untuk belajar secara mandiri dari sekumpulan data yang sudah diberikan sebelumnya, dengan menggunakan algoritma dan model untuk membuat prediksi. Fokus utama dari machine learning adalah untuk menemukan sebuah pola yang tepat dari sekumpulan data, sehingga dapat menghasilkan suatu model untuk melakukan proses *input-output* tanpa menggunakan kode program secara eksplisit (A. K. Tiwari, 2017). *Machine learning* dibagi dalam 3 bentuk, yakni *supervised learning*, *unsupervised learning* dan *generative learning*. Sentiment analisis dengan algoritma naïve bayes menggunakan metode *supervised Learning*.

1. **Supervised Learning**

*Supervised learning* adalah bidang pengenalan pola dan statistik dalam ilmu komputer. Ini adalah studi ilmiah tentang algoritma dan model statistik, yang digunakan untuk melakukan tugas tertentu secara efisien, tanpa menggunakan instruksi eksplisit, tetapi mengandalkan model. Algoritme pembelajaran yang diawasi membangun model matematika dari data sampel untuk membuat prediksi tanpa memerlukan pemrograman eksplisit untuk melakukan tugas. (Yin, Q. 2020).

*Supervised Learning* adalah suatu metode untuk menciptakan *artificial intelligence* (AI), untuk mengidentifikasi pola dalam kumpulan data yang tidak di klasifikasikan atau tidak di beri label. Algoritma yang bertujuan untuk memperkirakan fungsi pemetaan sehingga ketika ada variabel *input* ( X ) kita dapat memprediksi variabel *output* ( Y ). Algoritma supervised learning dapat digunakan untuk memproses berbagai jenis data, mulai data yang terstruktur hingga yang tidak terstruktur. (Altamevia, F. 2023)

## Teori Pendukung

### Python

Python adalah bahasa pemrograman komputer open source untuk tujuan umum. Ini dioptimalkan untuk kualitas perangkat lunak, produktivitas pengembang, portabilitas program, dan integrasi komponen. Python digunakan oleh setidaknya ratusan ribu pengembang dunia di berbagai bidang seperti skrip Internet, pemrograman sistem, antarmuka pengguna, kustomisasi produk, pemrograman numerik, dan banyak lagi. Secara umum dianggap menjadi salah satu dari empat atau lima bahasa pemrograman yang paling banyak digunakan di dunia hari ini.(Mark Lutz, 2011)

Python adalah bahasa pemrograman interpretative yang dianggap mudah dipelajari serta berfokus pada keterbacaan kode, dengan kata lain python diklaim sebagai Bahasa pemrograman yang memiliki kode kode pemrograman yang sangat jelas, lengkap dan mudah untuk dipahami. (JUD, 2019)

### Text Mining

*Text mining* adalah salah satu bidang yang sampai saat ini masih berkembang dengan pesat, dengan tugasnya dalam mengekstraksi atau mengumpulkan informasi yang bermakna dari teks alami suatu bahasa. Ini dapat diartikan sebagai proses menganalisis suatu teks untuk kemudian diekstrak informasi-informasi yang berguna dari teks tersebut untuk tujuan terentu. Dalam budaya modern, teks adalah salah satu media dalam pertukaran informasi, dibanding dengan database, teks tidak terstuktur, memiliki bermacam-macam bentuk, dan lebih sulit ditangani menggunakan algoritma tertentu.(Witten : 2004)

Pada kasus ini, yaitu text mining sumber data yang berupa teks tidak memiliki struktur yang jelas dan memiliki bermacam bentuk, sehingga disebut sebagai unstructured data. Maka dari itu, butuh proses untuk membuat data menjadi lebih terstuktur sehingga ekstraksi informasi dari teks akan lebih mudah, tepat, dan sangat penting dalam proses text mining. Sumber data yang digunakan, yaitu novel berbahasa Indonesia merupakan unstructured data, sehingga butuh proses untuk membuat data menjadi lebih terstuktur. Salah satunya adalah dengan diawali oleh preprocessing, yang mana nanti akan menghasilkan fitur yang lebih represenatif dibanding sumber data novel berbahasa Indonesia yang belum dipersiapkan dan masih tidak berstruktur.

### Jupyter Notebook

Jupyter adalah organisasi non-profit untuk mengembangkan software interaktif dalam berbagai bahasa pemrograman. Notebook adalah satu software buatan Jupyter, adalah aplikasi web open-source yang memungkinkan Anda membuat dan berbagi dokumen interaktif yang berisi kode liv e, persamaan, visualisasi, dan teks naratif yang kaya.(B. Prijono : 2019)

Pada penelitian ini, Jupyter Notebook digunakan sebagai salah satu text editor untuk menuliskan kode-kode program Python serta memvisualisasikan data hasil olah pada sistem ini.

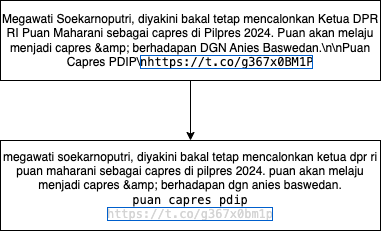
### Preprocessing

Proses preprocessing merupakan tahap dimana sumber data input diolah Kembali sebelum kemudian diproses lebih lanjut dan dijadikan bahan data pada machine leardning. Pada kasus teks tweet, text mining masih memliliki struktur yang bercampur dimana pada tweet masih ada mention dan link, serta data yang NaN pada data hasil scrapping, sehingga dibutuhkan proses yang merubah bentuknya menjadi data yang terstruktur. Proses ini akan melakukan penyeragaman case tweet yang merubah semua tweet menjadi lowercase, menghilangkan tanda mention serta username termention, menghilangkan tautan pada tweet, kemudian membuat token dari data input, sehingga data lebih bersih, terstruktur dan dapat diolah lebih lanjut.

Pada penelitian ini, tahap preprocessing yang diterapkan adalah *Case Folding, Lemmatization, Stopword Removal dan Tokenizing.*

#### Case Folding

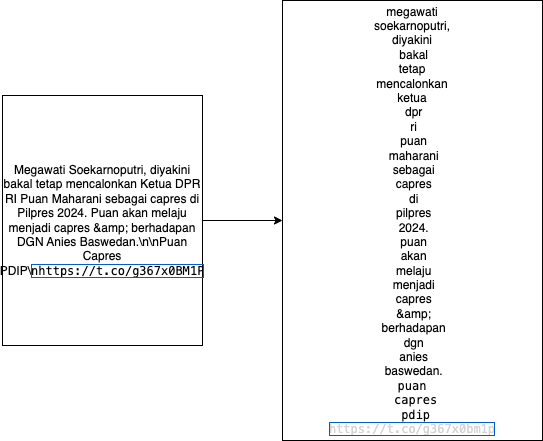
Case folding digunakan untuk menyeragamkan seluruh tesx dalam case yang seragam, abik menjadi huruf kecil (lowercase) ataupun huruf kapital (uppercase). Case folding digunakan pada penelitian ini adalah penyeragaman menjadi lowercase. Contoh dari penerapan case folding bisa dilihat pada gambar dibawah ini:



**Gambar 2. 1** contoh case folding

#### Tokenizing

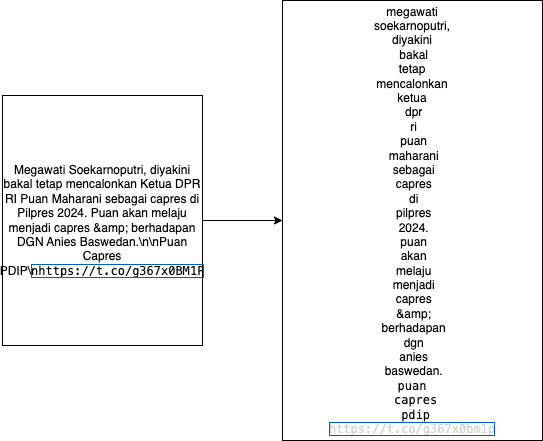
Proses Tokenizing adalah proses dimana string dipotong menjadi beberapa bagian dengan melihat delimiternya, seperti tipe kapitalisasi, keberadaan digit, tanda baca, karakter special dan sebagainya. Pemecahan dokumen menjadi kata – kata tunggal silakukan dengan caara men-scan dokumen dan setiap kata akan teridentifikasi atau terpisahkan dengan kalimatnya oleh delimiter. Tokenizing adalah proses diamana data input dibagi menjadi beberapa token sesuai dengan jumlah kalimat menggunakan delimiter ‘.’ Pada teks input contoh tokenizing ada pada gambar dibawah.



**Gambar 2. 2** contoh tokenizing

#### Stemming

Stemming adalah teknik pada natural language processing yang digunakan untuk mengembalikan kata kepada kata dasarnya yang disesuaikan dengan kamus Bahasa Indonesia, proses stemming dilakukan dengan menggunakan library sastrawi. Stemming digunakan pada kebutuhan yang berhubungan dengan text mining seperti information retrieval yang dilakukan pada tahap preprocessing.



**Gambar 2. 3** contoh stemming

#### Stopword Removal

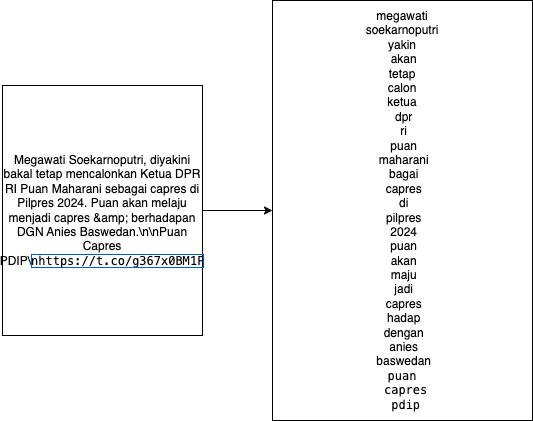
Stopword Removal adalah tahap pemilihan kata-kata yang dianggap penting. Terdapat dua metode yang dapat digunakan dalam tahap stopword removal, yakni:

* **Stoplist**

Pada metode ini, kita menyuaokan kumpulan kata yang tidak deskriptif/tidak pendting yang disebut stoplist. Kata yang termasuk ke dalam stoplist akan dibuang dan tidak digunakan pada proses selanjutnya.

* **Wordlist**

Wordlist merupakan kebalikan dari stoplist, pada metode ini kita menyiapkan kumpulan kata yang deskriptif yang disebut wordlist. Hanya kata yang termasuk ke dalam wordlist yang akan digunakan pada proses selanjutnya, sementara kata lainnya akan dibuang



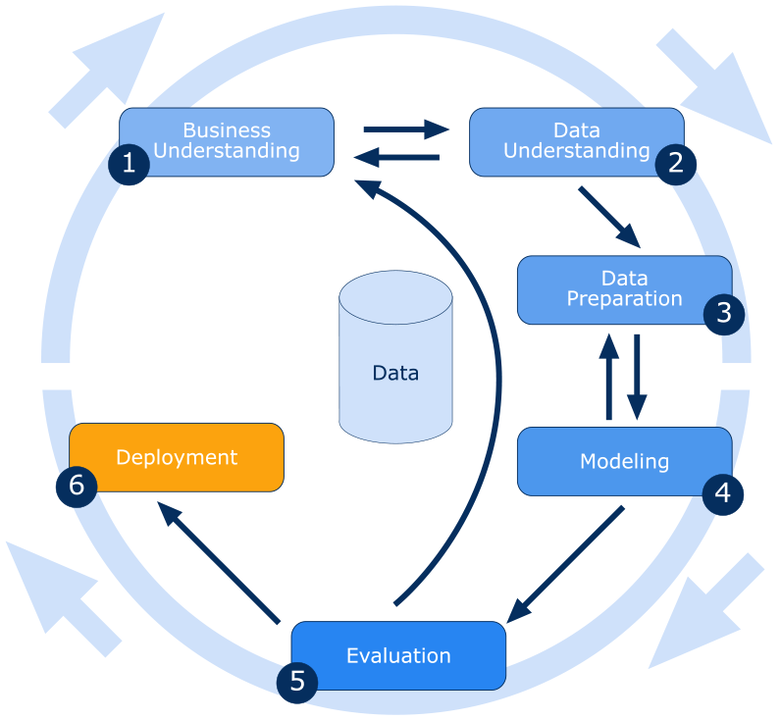
**Gambar 2. 4** contoh stopword removal

### Monte Carlo Cross Validation

blabalablabla

### Cross Industry Standard Process for Data Mining (CRISP-DM)

nanan



## Penelitian-penelitian Terdahulu

State of the art adalah hal yang cukup penting bagi penelitian

### Implementasi Metode Naïve Bayes untuk Analisis Sentimen Warga Jakarta Terhadap Kehadiran Mass Rapid Transit

Sarika A, Helena N, Noor F, Ika N.(2019), melakukan penelitian menggunakan data dari sosial media yaitu Twitter dengan keyword “MRTJakarta” yang dilakukan selama masa uji coba public MRT yaitu dari tanggal 5 – 23 maret 2019. Tweet yang diambil sebanyak 1000 tweet (800 tweet untuk training dan 200 tweet untuk testing). Dalam peneliltian ini naive bayes dapat memprediksi sentimen dari tweet yang sudah dikumpulkan terkait animo masyarakat terhadap MRTJakarta dengan akurasi sebesar 75%.

### Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u

Ike V, Bagas S.(2022), melakukan penelitian ini dapat diambil kesimpulan bahwa Algoritma Naïve Bayes Classifier dapat melakukan analisis sentimen dengan benar dan melakukan klasifikasi secara otomatis setelah melalui tahapan- tahapan proses, yaitu Preprocessing data, pembobotan kata, membuat model untuk klasifikasi otomatis dan dibuatnya data training untuk melatih klasifikasi pada data testing. Tahapan proses tersebut dapat berjalan dengan baik dan mengklasifikasikan data dengan parameter positif dan negatif. Setelah dilakukan 3 kali pengujian didapatkan hasil akurasi 80%, 80%, dan 85%. Didapatkan hasil akurasi paling tinggi pada pengujian terakhir yakni sebesar 85%. Dengan pengujian menggunakan 3 dataset yang memiliki jumlah data yang berbeda, dan setelah mendapatkan hasil tingkat akurasi dari proses analisis sentimen dapat disimpulkan bahwa jumlah dataset dalam pengujian sangat berpengaruh terhadap tingkat akurasi Algoritma Naïve Bayes Classifier. Hal ini ditunjukan oleh hasil tingkat akurasi pada pengujian ketiga dengan 3000 dataset mendapatkan nilai akurasi 85%, lebih besar daripada pengujian pertama dengan 1000 dataset yang hanya memiliki akurasi sebesar 80%.

### Sentiment Analysis Menggunakan Naïve Bayes Classifier pada Tweet Tentang Zakat

Adhyaksa H (2020), hasil klasifikasi sentiment dari 50 tweet data uji menggunakan algoritma naïve bayes dengan seleksi fitur Term-Frequency serta metode lexicon Based, didapatkan jumlah sentiment positif yang lebih dominan dibandingkan sentiment negative maupun netral dikarenakan pada pengujian dengan metode lexicon based terdapat lebih banyak tweet yang mengandung kata dalam kamus lexicon positif disbanding kata dalam kamus lexicon negative. Selanjutnya, pada pengujian dengan masing-masing seleksi fitur, sentiment positif lebih dominan dikarenakan tidak keseimbangan jumlah sentiment positif, negative dan netral dalam klasifikasi data latih menggunakan metode lexicon based dimana sentiment positif lebih besar sehingga system lebih condong dalam mengklasifikasi sentiment positif.

# METODOLOGI PENELITIAN

## Metode Penelitian

Metode penelitian yang digunakan ialah pendekatan supervised learning. metode ini memiliki tahapan tahapan yaitu pengumpulan data, preprocessing data, pelabelan data, pemilihan model dan prediksi.

## Metodologi Pengembangan Sistem

### Cross Industry Standard Process for Data Mining

bbbb

### Scrapping

<Alasan pemilihan model proses yang digunakan>

### Pendekatan Supervised learning

<Alasan pemilihan permodelan yang digunakan >

### Modeling

<Alasan pemilihan pendekatan yang digunakan >

### Monte Carlo Cross Validation

<Alasan pemilihan tools permodelan yang digunakan >

## Tahapan Penelitian

<Tahapan penelitian ini menggabungkan langkah-langkah Metode Penelitian dan Metode Pengembangan Sistem yang dilakukan dalam kegiatan Skripsi. Ini adalah variasi dari Rencana Kegiatan pada Proposal namun dalam bentuk rincian langkah demi langkah penelitian yang dilakukan.   
Contoh berikut adalah Tahapan Penelitian Rekayasa dengan model proses Metode Waterfall>

1. Mendefinisikan Masalah <dilaksanakan di Bab I>
2. Mengkaji penelitian terdahulu <dilaksanakan di Bab II dan III>
3. Menganalisis proses bisnis <dilaksanakan di Bab VI.I>
4. Menganalisis kebutuhan fungsional sistem <dilaksanakan di Bab IV.2>
5. Membuat desain sistem <dilaksanakan di Bab IV.3>
6. Membuat implementasi sistem <<dilaksanakan di Bab IV.4>
7. Melakukan pengujian sistem <<dilaksanakan di Bab IV.5>

# HASIL DAN PEMBAHASAN

<CONTOH di sini untuk Pengembangan Perangkat Lunak menggunakan *Rational Unified Process* (RUP), jadi Proses Bisnis masuk ke dalam *Inception Phase*. **Untuk Model Proses yang lain**, biasanya sistematikanya adalah Bab IV.1 Proses Bisnis, dengan rincian, Bab IV.1.1 Proses Bisnis Sistem Berjalan dam IV.1.2 Proses Bisnis Sistem Usulan>

## Inception Phase

### Analisis Proses Bisnis

#### Proses Bisnis Sistem Berjalan

<Gambar Proses Bisnis Sistem yang sedang berjalan disertai narasi>

#### Proses Bisnis Usulan (Bila ada)

<Gambar Proses Bisnis Usulan yang sedang berjalan disertai narasi>

#### Business Usecase Diagram

#### Business Object Model

### Requirements

#### Daftar Kebutuhan Fungsional

#### Daftar Kebutuhan Non Fungsional

## Elaboration Phase

### Analisis

#### System Usecase Diagram

#### Usecase Scenario

### Design

#### Sequence Diagram

#### State Chart Diagram

#### Class Diagram

#### Desaim Kelas-kelas Interface/View

#### Desain Kelas-kelas Controller

#### Desain Kelas-kelas Model

## Construction Phase

### Implementasi

#### Realisasi Usecase

#### Spesifikasi Kelas

#### *Screenshoot/Pseudocode* Kelas-kelas Inteface/View

#### Implementasi/*Pseudocode* Kelas-kelas Controller

#### Desain Tabel/*Pseudocode* Kelas-kelas Model

### Pengujian

#### Rencana Pengujian

#### Deskripsi Pengujian

#### Hasil Pengujian

### Deployment

#### Package Diagram

#### Deployment Diagram

# SIMPULAN DAN SARAN

## Simpulan

<Sampaikan simpulan sesuai dengan point-point tujuan penelitian>

## Saran

<Sampaikan saran-saran yang bisa dilakukan oleh pengguna atau peneliti selanjutnya untuk menyempurnakan penelitian ini>

DAFTAR PUSTAKA

Albarda (2010): Portal Kantor Digital, *Proceeding Seminar TIK Nasional ke-6*., STEI Institut Teknologi Bandung.

**LINK**

**JURNAL**

**PROSIDING**

Ali, N., Babar, M.A. (2009):Modeling Service Oriented Architectures of Mobile Applications by Extending SoaML with Ambients, *IEEE Computer Society vol. 978-0-7695-3784-9/09*

Budi, Setiawan, “*Panduan Penggunaan Notasi UML dalam Pengembangan Aplikasi GIS*”, [*http://setiawanbudi.blogspot.com/artikel?17*,](http://setiawanbudi.blogspot.com/artikel?17,) diambil pada tanggal 17 Agustus 2018 pukul 19.00 WIB

Creswell, J.W. (2003): *Research Design Qualitative, Quantitive, and Mixed Methods Approaches*, Sage Publications, Second Edition, London

**BUKU**

Dörner, D., Draxler,S., Pipek, V., and Wulf, V.(2009): End Users at the Bazaar: Designing Next-Generation Enterprise Resource Planning Systems, *IEEE Software vol. 0740-7459/09*

**JURNAL**

Erl, Thomas. (2005): *Service Orienterd Architecture,* Prentice Hall, New Jersey.

**BUKU**

Fei, H.R., Xiang, W.H., Qian, X., Pei, J.X., Hui, L. (2009): Service oriented decentralized access control for military systems in Net-Centric Environment, *IEEE Computer Society vol. 978-0-7695-3643-9*

**JURNAL**

Sugiman, A., Djubiantono, T., Aziz, F., Subandi, J.S., Wagiman, R. A., Aceng, D.C., Hardi, W.E., Elan, J.M., Darsono, A.C, dan Agus (2015) : Pengukuran Kadar Gangguang Sinyal GPS Menggunakan Model Bernoulli, *Jurnal Telematika,* 48, 661-667.

**BUKU + EDITOR**

Wijaya, R. (1996) : *Diagnosis Penyakit Tipus dengan Metode PCR,* Disertasi Program Doktor, Universitas Langlangbuana, 25 – 29.

**DISERTASI/  
TESIS/  
SKRIPSI**

Zuckerberg, Mark, “How to Obtain Accurate GPS Location With Android OS”, *IEEE Online Journal, Edition IX/18, 2018*, [*http://ejournal.ieee.org/artickel?1123*,](http://ejournal.ieee.org/artickel?1123,) diambil pada tanggal 20 Semptember 2018 pukul 19.00 WIB

**LINK**

Galih, Sandika, “*Belajar NodeJS | 1. Apa Itu NodeJS*?”, Tutorial Web Programming UNPAS, https://youtu.be/sSLJx5t4OJ4?t=687, diambil pada tanggal 20 Februari 2021 pukul 19.00 WIB

**Sumber YouTube**

gcuomo (24 October 2013). "*JavaScript Everywhere and the Three Amigos* (Into the wild BLUE yonder!)". [www.ibm.com](http://www.ibm.com), https://en.wikipedia.org/wiki/Node.js#cite\_note-6, diambil pada tanggal 20 Februari 2021 pukul 19.00 WIB

**Sumber Wikipedia**

LAMPIRAN A: Python Notebook Scrapping

LAMPIRAN B: Python notebook cleaning data

LAMPIRAN C: Python notebook SENTIMEN ANALISIS