

# ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP RENCANA KENAIKAN PPN MENJADI 12% DI MEDIA SOSIAL X DENGAN METODE NAÏVE BAYES

Jesica Kristovani Siagian<sup>1</sup>, Painem<sup>2\*</sup>

<sup>1,2</sup>Teknik Informatika, Fakultas Teknik Informasi, Universitas Budi Luhur, Jakarta, Indonesia

Email: <sup>1</sup>1911510483@student.budiluhur.ac.id, <sup>2\*</sup>painem@budiluhur.ac.id

(\* : corresponding author)

**Abstrak-** Pajak adalah salah satu sumber pendapatan atau penerimaan bagi suatu negara termasuk Indonesia. Pajak juga merupakan sumber penerimaan terbesar bagi suatu negara dan tidak dapat dipisahkan sejak dahulu. Salah satu sumber dana pembangunan terbesar didapatkan dari departemen pajak. Pajak yang ditetapkan oleh pemerintah salah satunya ialah Pajak Pertambahan Nilai (PPN). PPN itu sendiri merupakan pajak atas konsumsi atas pembelian suatu barang atau pengguna jasa yang sifatnya objektif sehingga dapat dikatakan PPN merupakan pajak berdasarkan objek yang digunakan oleh pembeli. Kenaikan PPN dapat menimbulkan beberapa masalah, termasuk potensi berbelanja di luar negeri seiring dengan kenaikan harga barang dan jasa di Indonesia. Dengan kenaikan PPN ini, maka harga barang akan mengalami kenaikan dan penjualan akan mengalami penurunan sehingga berdampak pada turunnya sektor barang dan jasa. Rencana kenaikan PPN ini kemudian viral di perbincangkan di media sosial salah satunya yaitu twitter. Banyak Masyarakat memperdebatkan rencana kenaikan sehingga terjadinya pro dan kontra terhadap rencana kenaikan PPN ini. Oleh sebab itu penelitian ini bertujuan untuk melakukan klasifikasi tweet dari twitter mengenai PPN 12% dengan menggunakan metode naive bayes classifier. Beberapa tahap yang dilakukan yaitu *crawling* data, pelabelan data, *preprocessing*, ekstraksi fitur dengan *bag of words*, klasifikasi dengan naive bayes classifier dan pengujian dengan *confusion matrix*. *Dataset* merupakan hasil *crawling* twitter mulai tanggal 1 Maret hingga 15 Mei 2024. Dari 468 *dataset*, hasil pengujian mendapatkan akurasi 83% sebesar presisi 68,8% dan *recall* 78,6%.

**Kata kunci:** PPN 12%, analisis sentimen, *naive bayes classifier*, *bag of words*

## ANALYSIS OF INDONESIAN PUBLIC SENTIMENT TOWARDS THE PLAN TO INCREASE VAT TO 12% ON X SOCIAL MEDIA USING THE NAÏVE BAYES METHOD

**Abstract-** Taxes are one of the sources of revenue or income for a country, including Indonesia. Taxes are also the largest source of revenue for a country and have been integral for a long time. One of the largest sources of development funds comes from the tax department. One of the taxes established by the government is Value Added Tax (VAT). VAT itself is a consumption tax on the purchase of goods or services, and it is objective in nature, meaning VAT is a tax based on the object used by the buyer. An increase in VAT can lead to several issues, including the potential for shopping abroad due to rising prices of goods and services in Indonesia. With this VAT increase, the prices of goods will rise, and sales will decrease, affecting the goods and services sector. The plan to increase VAT then went viral and was discussed on social media, including Twitter. Many people debated the plan, leading to both support and opposition regarding the VAT increase. Therefore, this study aims to classify tweets on Twitter about the 12% VAT increase using the Naive Bayes classifier method. The steps undertaken include data crawling, data labelling, pre-processing, feature extraction with bag of words, classification using the Naive Bayes classifier, and testing with a confusion matrix. The dataset consists of Twitter data crawled from March 1 to May 15, 2024. From 468 datasets, the testing results achieved an accuracy of 83%, with a precision of 68.8% and recall of 78.6%.

**Keywords:** 12% VAT, Sentiment Analysis, Naive Bayes Classifier, Bag of Words

### 1. PENDAHULUAN

Seiring dengan perkembangan ekonomi di dunia, maka aktivitas suatu perdagangan semakin meningkat di setiap sektor. Aktivitas perdagangan selalu berhubungan dengan pemenuhan kebutuhan yang dibutuhkan oleh suatu negara kepada negara lainnya. Maka dari itu, untuk melancarkan aktivitas tersebut dan juga membantu memberikan pemasukan negara. Suatu negara harus menetapkan pajak untuk setiap barang yang diperjual belikan

Pajak adalah sumber pemasukan negara yang memiliki kemampuan besar dalam menunjang program kerja pemerintah guna mendukung tujuan pemerintah [1] Pajak yang ditetapkan oleh pemerintah salah satunya ialah Pajak Pertambahan Nilai (PPN). PPN itu sendiri Pajak yang dikenakan pada setiap peningkatan nilai barang atau

jasa yang dihasilkan akibat proses produksi di berbagai jalur perusahaan dan dikonsumsi di dalam negeri sehingga dapat dikatakan PPN merupakan pajak berdasarkan objek yang digunakan oleh pembeli.[2]

Berdasarkan Undang-Undang Nomor 7 Tahun 2021 mengenai Harmonisasi Peraturan Perpajakan, pemerintah berencana untuk menaikkan tarif Pajak Pertambahan Nilai (PPN) menjadi 12% pada tahun depan. Pasal 7 ayat (2) dalam Bab 4 UU HPP secara khusus menyebutkan bahwa tarif PPN akan berubah dari 11% menjadi 12% mulai 1 Januari 2025. Dari perspektif pemerintah, peningkatan tarif PPN ini bertujuan untuk mengoptimalkan penerimaan negara, yang diharapkan dapat mendukung pemulihan ekonomi dan pembangunan.[3]

Kenaikan PPN dapat menimbulkan beberapa masalah, termasuk potensi berbelanja di luar negeri seiring dengan kenaikan harga barang dan jasa di Indonesia. Dengan kenaikan PPN ini, maka harga barang akan mengalami kenaikan dan penjualan akan mengalami penurunan. Kenaikan PPN juga akan meningkatkan biaya produksi dan konsumsi. Sehingga berdampak pada turunnya sektor barang dan jasa. Produktivitas yang rendah ini dapat mempengaruhi bertambahnya pengangguran dan mengurangi pendapatan serta konsumsi masyarakat.

## 2. METODE PENELITIAN

### 2.1. Data Penelitian

*Dataset* atau data yang digunakan dalam penelitian ini dikumpulkan dari Twitter mulai tanggal 1 Maret 2024 sampai dengan tanggal 14 Mei 2024. Dari jangka waktu tersebut didapatkan sebanyak 468 *record* data. Proses pengambilan data atau *crawling data* menggunakan *library python* yang bernama *tweet-harvest*. Hasil dari *crawling data* berupa *.csv*. Data yang diambil merupakan data berbahasa Indonesia dengan kata kunci “*ppn 12%*”. Berikut merupakan contoh data *tweet* hasil dari *crawling* yang dapat dilihat pada Gambar 1 berikut:

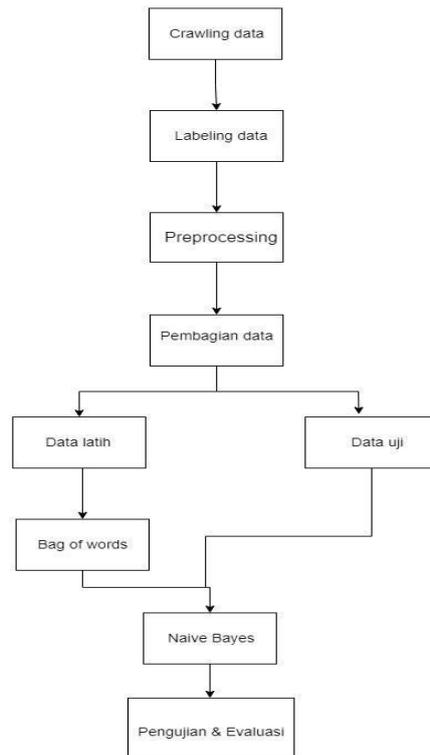
created_at	full_text	Username
Tue May 14 23:24:4	@DarneNyoman @Shopeeholic @DoankWarto Pajak ppn naik 12% ucap salah satu mentri alasannya pada mlh kelanjutan ada juga yg doa maksa jatah kursi mentri... negara buat panggung standup komedi rakyat mlongo sok iyes	fmaulana179
Tue May 14 16:15:2	Catatan artikel ini justru memberikan dukungan ke Prabowo. Coba lihat mana ada elit Prabowo yang menulis mereka sedang sibuk rebutan kursi kekuasaan untuk menjabat menjadi menteri? PPN naik sebesar 12% aja bisa menambah 50-70 triliun pendapatan negara. Jika naik jadi 25% kayak	semut_daun01
Tue May 14 14:16:4	Rencana PPN Naik 12 Persen PKS : Masyarakat Kian Terhimpit Industri Semakin Terpukul. #PKSpembelaRakyat #PKSuntukIndonesia <a href="https://t.co/aaWYThi12K">https://t.co/aaWYThi12K</a>	pksbogor
Tue May 14 11:59:4	Kenaikan PPN disetujui DPR sejak thn 2021. Jadi 11% thn 2022 jadi 12% thn 2024. Tapi kata SMI PPN yg naik jd 12% tergantung presiden baru. Kl Pres Prabowo batalkan akan bilang lembek ditekan rakyat. Sahkan akan disebut diktator. Jd penonton yg baik ahh.. <a href="https://t.co/u3BbTLGyyj">https://t.co/u3BbTLGyyj</a>	Widyarenee
Tue May 14 11:40:5	PPN naik jadi 12% ges..	kangmasai
Tue May 14 07:48:1	Teknologi&amp	FPKSDPRRI
Tue May 14 06:18:1	Rencana PPN Naik 12 Persen Aleg PKS: Masyarakat Kian Terhimpit Industri Semakin Terpukul <a href="https://t.co/RKcJ8RV3YE">https://t.co/RKcJ8RV3YE</a> via @FPKSDPRRI   @eck	PKSejahtera
Tue May 14 04:09:5	Rencana kenaikan PPN sangat menghimpit masyarakat. Ini memukul mundur daya beli masyarakat yang saat ini dihadapkan berbagai tekanan perekonomian ~@EckyAwal <a href="https://t.co/N9p6TJ43vb">https://t.co/N9p6TJ43vb</a>	Dwiagus51141596

Gambar 1. Hasil *Crawling Data*

### 2.2. Penerapan Metode

Dalam melakukan penelitian ini, terdapat beberapa proses yang perlu dilakukan. Tujuannya dibuat proses tersebut agar penelitian ini dapat berjalan sesuai dengan rencana. Berikut beberapa prosesnya pada Gambar 2:

- Crawling data* Sebuah proses dalam pengumpulan data dengan menerapkan Search API pada media twitter.[4]
- Labeling* data oleh pakar yang merupakan seorang pegawai pajak. Tujuan dari pelabelan data untuk mengklasifikasikan data berdasarkan karakteristik dari isi tweet atau kalimat. Pelabelan ini dibagi menjadi dua klasifikasi, yaitu positif dan negatif.
- Preprocessing* yaitu proses pembersihan huruf, karakter dan kata yang tidak penting. Tujuannya dilakukan proses tersebut untuk menghasilkan sebuah *dataset* menjadi lebih terstruktur dan bersih
- Pembagian data, menjadi Data Latih dan Data Uji.
- Bag of Words* adalah representasi sederhana yang digunakan dalam pemrosesan bahasa alami (NLP) dan pengambilan informasi “IR”, juga dikenal sebagai model ruang vector. [5]
- Naïve Bayes Classifier* untuk menghitung nilai probabilitasnya. Algoritma ini termasuk kedalam kategori klasifikasi data *mining* yang menerapkan teknik statistik dan probability. Fungsi algoritma ini yaitu meramalkan kejadian di masa yang akan datang. [6]



**Gambar 2.** Metode Penelitian

### 2.3. Rancangan Pengujian

Pengujian dilakukan dengan menggunakan tahap klasifikasi Naive Bayes Menguji melalui *Confusion matrix* membandingkan label actual dengan label prediksi. Tujuan dari *Confusion matrix* adalah untuk mengukur kinerja model klasifikasi yang sudah dibuat.

### 2.4. Pre-processing

Merupakan tahapan pembersihan data agar memiliki pola yang sama sebelum dilakukan proses selanjutnya. Tahap *pre-processing* yang digunakan antara lain:

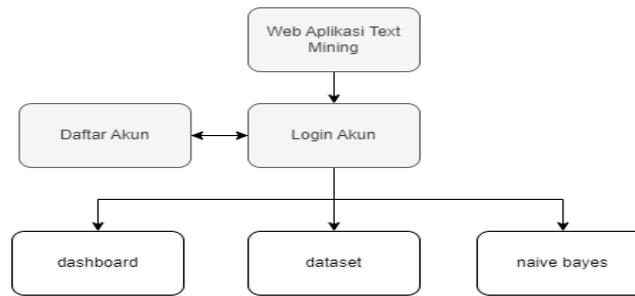
- Case Folding* digunakan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Sehingga huruf 'a' sampai 'z' yang diambil, karakter selain huruf akan dihilangkan, begitu juga dengan huruf-huruf kapital [7].
- Cleansing* merupakan tahapan dimana karakter dan tanda baca yang tidak diperlukan dihilangkan dari teks, yang dimana salah satu fungsi *cleansing* itu sendiri yaitu dapat mengurangi noise pada *dataset* [8].
- Slang Words* bertujuan untuk memperbaiki kata yang tidak baku menjadi baku dan memperbaiki kata yang disingkat. [9]
- Stop Words* Merupakan (*Common Words*) atau kata-kata umum yang sering muncul tidak memberikan informasi penting, contohnya 'yang', 'di', 'ke', dan lain-lain [10].
- Stemming* merupakan pemrosesan teks yang mengubah kata imbuhan menjadi kata dasar, contohnya me, ber, an [7].

### 2.5. Rancangan Basis Data

Untuk melakukan penelitian ini, kita perlu merancang sistem database untuk menyimpan data yang dibutuhkan saat menjalankan aplikasi dan untuk memudahkan pembuatan aplikasi.

### 2.6. Rancangan Menu

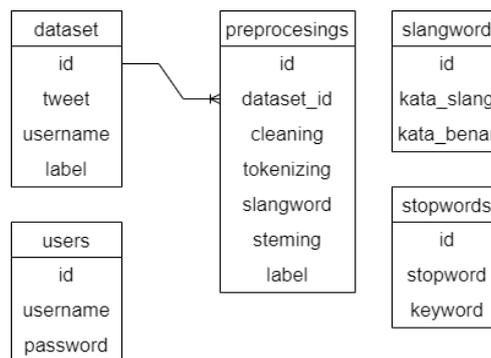
Rancangan menu Gambar 3 digunakan untuk merancang menu pada aplikasi untuk memudahkan pengguna agar tidak tersesat saat menggunakan aplikasi. Rancangan menu pada sistem ini terdiri dari menu *dashboard*, *dataset* dan *naive bayes*.



Gambar 3. Rancangan Menu

## 2.7. Rancangan LRS

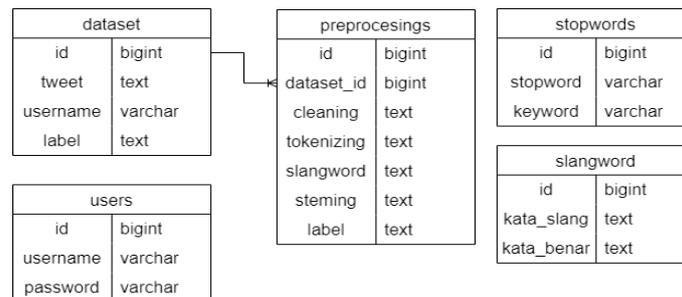
Sistem Gambar 4 digunakan dalam penelitian ini dirancang menggunakan *Logical Relational Structure* (LRS).



Gambar 4. Rancangan LRS

## 2.8. Class Diagram

Berikut Gambar 5 adalah rancangan *Class Diagram* yang digunakan untuk sistem pada penelitian ini.

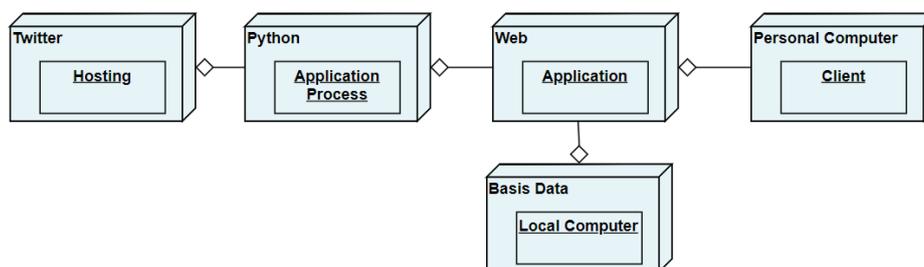


Gambar 5. Class Diagram

## 3. HASIL DAN PEMBAHASAN

### 3.1. Deployment Diagram

*Deployment Diagram* Gambar 6 adalah gambaran tentang proses- proses dalam sistem yang sedang berjalan dan menjelaskan bagaimana sistem tersebut saling terkait.



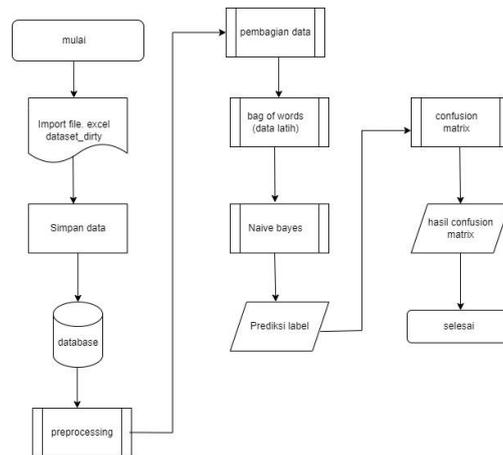
Gambar 6. Deployment Diagram

### 3.2. Implementasi Metode

Implementasi pada penelitian dilakukan dengan enam langkah yang diproses harus berurutan. Tahapannya yaitu pengumpulan data atau *crawling* data, proses *labeling*, *preprocessing*, pembagian data menjadi data latih dan uji, *bag of words*, dan klasifikasi menggunakan Naive Bayes.

### 3.3. Flowchart

*Flowchart* adalah *diagram* yang menunjukkan langkah-langkah dan keputusan untuk menyelesaikan suatu proses pemrograman. Setiap langkah yang ada pada *flowchart* disajikan dalam bentuk *diagram* dan dihubungkan dengan garis atau panah. *Flowchart* Gambar 7 ini menggambarkan langkah-langkah atau proses yang dilakukan oleh sistem.



Gambar 7. Flowchart Keseluruhan Sistem

### 3.4. Algoritma

Algoritma merupakan suatu langkah atau metode yang dirancang berurutan dan secara teratur, untuk menyelesaikan masalah yang berkaitan dengan suatu kegiatan atau instruksi. Berikut adalah algoritma yang dibuat berdasarkan *flowchart* yang telah dibuat sebelumnya.

#### a. Algoritma Keseluruhan Proses Sistem

Algoritma 1 ini menggambarkan langkah-langkah pada keseluruhan system.

Algoritma 1. Algoritma Keseluruhan Proses Sistem

1	Start
2	Import file excel <i>dataset_dirty</i>
3	Simpan <i>dataset_dirty</i> ke basis data
4	Lakukan sub proses <i>preprocessing</i>
5	Lakukan sub proses pembagian data
6	Lakukan sub proses <i>bag of words</i> menggunakan data latih
7	Algoritma Naive Bayes
8	Keluaran label prediksi data uji
9	Lakukan sub proses <i>confusion matrix</i>
10	Keluaran hasil <i>confusion matrix</i>

#### b. Algoritma *Import Dataset*

Algoritme ini menggambarkan atau proses import *dataset* menggunakan file excel bertipe .xls dan .xlsx ke basis data.

Algoritma 2. Algoritma *Import dataset*

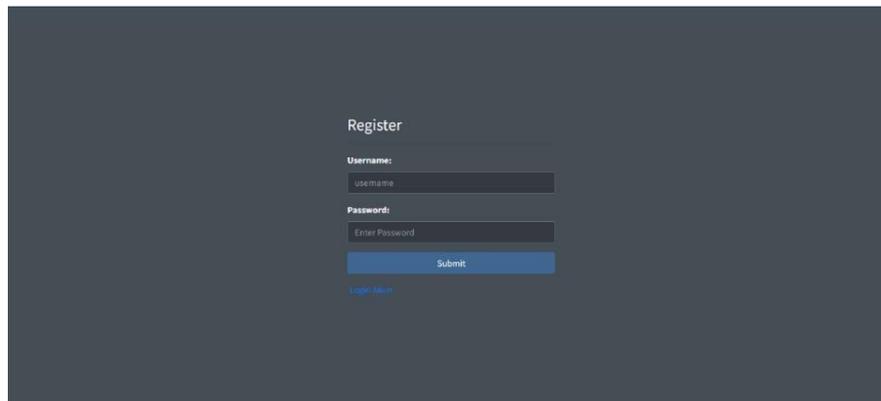
1	Start
2	Import file excel bertipe .xls atau .xlsxIf (tipe_file == .xls .xlsx)
3	Masukkan data ke dalam basis dataElse
4	Kembali ke import <i>dataset</i>
5	End

### 3.5. Pengujian

Pengujian merupakan salah satu kegiatan yang harus dilakukan pada saat proses pengembangan sistem, keakuratan atau kesamaan hasil yang diperoleh dengan sistem yang dirancang akan dievaluasi, dianalisis, dan ditentukan. Dalam penelitian ini, akurasi, presisi, dan *recall* diuji dengan menerapkan algoritma *naive bayes classifier*.

#### a. Tampilan Layar Register

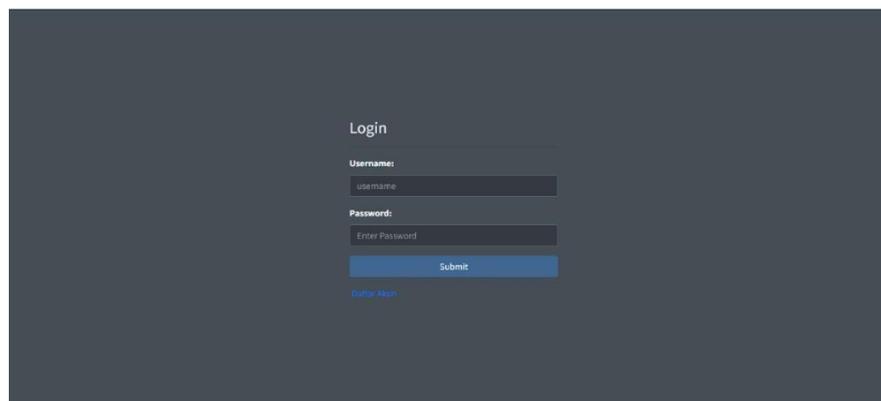
Pada halaman Gambar 8 ini pengguna diharapkan mendaftarkan *username* dan *password* agar bisa menggunakan aplikasi *text mining*.



Gambar 8. Tampilan Layar Register

#### b. Tampilan Layar Login

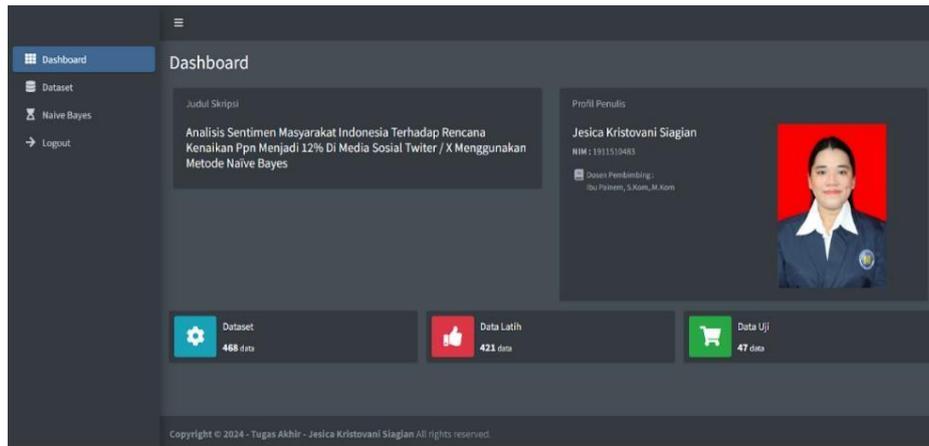
Pada halaman Gambar 9 ini pengguna diharapkan mengisi *username* dan *password* yang sebelumnya sudah didaftarkan, agar dapat menggunakan aplikasi *text mining*.



Gambar 9. Tampilan Layar Login

#### c. Tampilan Layar Dashboard

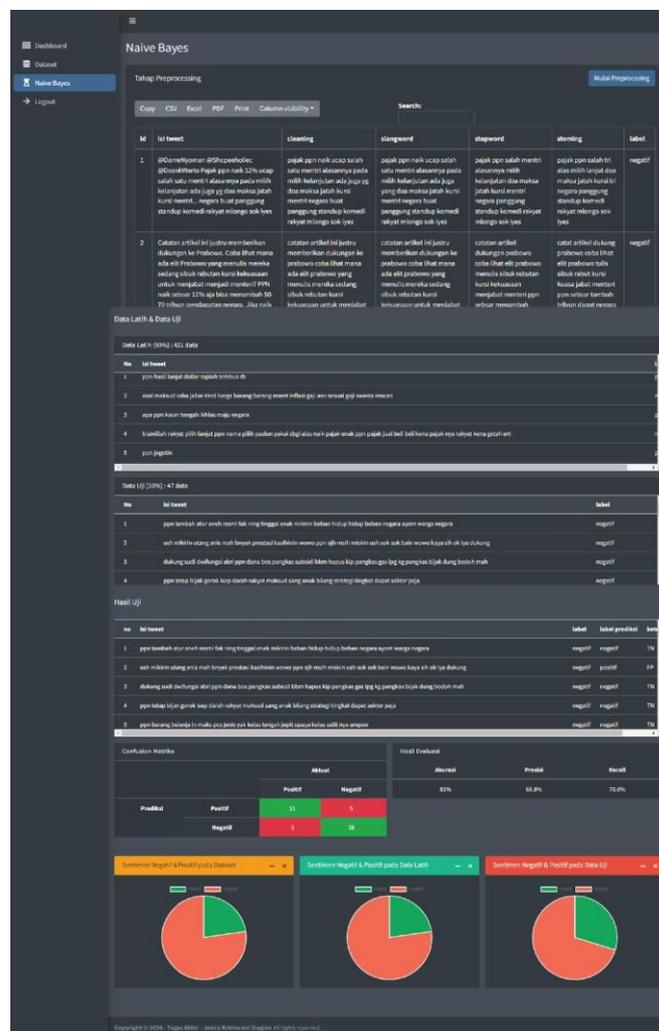
Halaman Gambar 10 ini merupakan tampilan pertama saat pengguna masuk ke aplikasi. Pada halaman ini terdapat informasi mengenai judul penelitian penulis, biodata singkat penulis, serta statistik jumlah *dataset*, data latih dan data uji yang digunakan pada penelitian ini.



Gambar 10. Tampilan Layar Dashboard

#### d. Tampilan Layar Naïve Bayes

Pada halaman Gambar 11 ini terdapat beberapa proses dari aplikasi *text mining* ini. Mulai dari proses *preprocessing*, pembagian data latih dan data uji, pengujian menggunakan *confusion* matriks, hasil nilai akurasi, presisi dan *recall*, serta terdapat *chart* untuk membandingkan sentimen positif dan negatif pada *dataset*, data uji dan data latih.



Gambar 11. Tampilan Layar Naïve Bayes

#### 4. KESIMPULAN

Berdasarkan Berdasarkan 468 *dataset* dari hasil *crawling* data di twitter, mengenai sentimen Masyarakat Indonesia terhadap rencana kenaikan PPN menjadi 12%. 362 data (77,3%) bersentimen negatif dan 106 data (22,7%) bersentimen positif sejak 1 Maret 2024 – 15 Mei 2024. Pada penelitian ini dilakukan dengan melewati 6 proses penting, yaitu: proses *crawling* data, *labeling*, *preprocessing*, pembagian data, ekstraksi fitur dengan *Bag of Words*, klasifikasi dengan *Naive Bayes Classifier* dan pengujian dengan *confusion matrix*. Hasil pengujian dan evaluasi mendapatkan nilai akurasi sebesar 83%, presisi 68,8% dan *recall* 78,6%.

#### DAFTAR PUSTAKA

- [1] I. Hidayat and S. Gunawan, "Pengaruh Kesadaran Wajib Pajak, Sanksi Perpajakan, Dan Kualitas Pelayanan Perpajakan Terhadap Kepatuhan Wajib Pajak Dalam Membayar Pajak Bumi Dan Bangunan," *Jurnal Manajemen dan Ilmu Pendidikan (MANAZHIM)*. vol. 4, no. 1, pp. 110-132, 2022.
- [2] S. Wijaya and K. R. Arsini, "Fasilitas Ppn Tidak Dipungut Atau Dibebaskan: Perbedaan Dan Permasalahan," *Jurnal Manajemen Sumber Daya Manusia, Administrasi dan Pelayanan Publik*. vol. 8, no. 1, pp, 91-104, 2021.
- [3] I. M. Putri, "Kenaikan Ppn 12% Dan Dampaknya Terhadap Ekonomi," *Jurnal Ilmiah MEA (Manajemen, Ekonomi, dan Akuntansi)*, vol. 8, no. 2, pp. 934-944, 2024.
- [4] N. T. Luchia, S. N. Auliani, H. Handayani, N. W. Azani, and R. Adha, " Analysis of Twitter User Sentiments for the Aplikasi TikTok Application Using Naïve Bayes Clasifier Algorithm," *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, pp. 100-104, 2023.
- [5] R. F. R. Pohan, D. E. Ratnawati and I. Arwani, " Implementasi Algoritma Support Vector Machine dan Model Bag-of-Words dalam Analisis Sentimen mengenai PILKADA 2020 pada Pengguna Twitter, " *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 10, pp 4924-4931, 2022.
- [6] A. Aditya and A. Wibowo, "Analisis Sentimen Menggunakan Metode Naïve Bayes Berdasarkan Opini Masyarakat Dari Twitter Terhadap Perang Rusia dan Ukraina," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, vol. 1, no. 1, September 2022, pp. 551-558.
- [7] N. Ula, C. Setianingsih, and R. A. Nugrahaeni, "Sistem Rekomendasi Lagu Dengan Metode Content-Based Filtering Berbasis Website," *e-Proceeding of Engineering*, vol. 8, no. 6, pp. 12193-12199, 2021.
- [8] K. A. Lubis, M. T. A. Bangsa and A. Yudertha, "Analisis Sentimen Opini Masyarakat Terhadap Pindahnya Ibu Kota Indonesia Dengan Menggunakan Klasifikasi Naïve Bayes," *Jurnal Teknoinfo*, vol. 18, no. 1, pp. 226-238, 2024.
- [9] T. M. Iryana, Indriati and P. P. Adikara, "Analisis Sentimen Masyarakat Terhadap Mass Rapid Transit Jakarta Menggunakan Metode Naïve Bayes Dengan Normalisasi Kata," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 6, pp. 2753-2760, 2021.
- [10] M. S. Anwar, I. M. I. Subroto and S. Mulyono, "Sistem Pencarian E-Journal Menggunakan Metode Stopword Removal dan Stemming Berbasis Android," *Prosiding Konferensi Ilmiah Mahasiswa Unissula (KIMU) 2 Klaster Engineering, Universitas Islan Sultas Agung*, 18 Oktober 2019, pp. 58-70.