

PENERAPAN NAIVE BAYES CLASSIFIER UNTUK ANALISIS SENTIMEN TERHADAP LAYANAN BIMBINGAN BELAJAR DI LKP ELEKTRON SAMARINDA

Fikriawan Mufti Haq ¹⁾, Pitrasacha Adytia ²⁾, Wahyuni ³⁾

Teknik Informatika, STMIK Widya Cipta Dharma
Jl. M. Yamin No. 25, Samarinda, Kalimantan Timur, 75123
Email : fikriawanjunior@gmail.com ¹⁾, pitra@wicida.ac.id ²⁾, wahyuni@wicida.ac.id ³⁾

ABSTRAK

Lembaga Pendidikan Non-Formal saat ini dihadapkan pada persaingan yang semakin ketat. Untuk dapat berkembang dan bertahan, lembaga tidak hanya menawarkan produk atau jasa yang berkualitas, tetapi juga memberikan layanan yang baik kepada customer nya. Dengan banyaknya LKP yang membuka jasa bimbingan belajar merupakan suatu ancaman nyata bagi LKP ELEKTRON. Agar bisa terus bersaing dengan LKP lainnya, dengan itu diperlukan perbaikan dan peningkatan layanan terhadap siswa maupun calon siswa Bimbingan Belajar di LKP ELEKTRON untuk tetap menjadikan LKP ELEKTRON sebagai salah satu pilihan LKP terfavorit yang ada di Samarinda.

Pada penelitian ini menggunakan metode dari Keputusan Menteri Ketenagakerjaan Republik Indonesia Nomor 299 Tahun 2020 Tentang Penetapan Standar Kompetensi Kerja Nasional Indonesia (SKKNI) dengan pengklasifikasian ulasan dari siswa/i aktif bimbingan belajar LKP Elektron ke beberapa sentimen diantaranya positif, negatif dan netral menggunakan algoritma Naive Bayes Classifier (NBC).

Pada pembuatan model NBC menggunakan 80% Data latih dan 20% Data uji dari 291 Data Ulasan. Dengan tingkat akurasi pada model yang sudah dibuat adalah 84.7 %. Selanjutnya model akan diterapkan kedalam sebuah sistem berbasis website menggunakan Streamlit untuk menganalisis sentimen terhadap layanan bimbingan belajar di LKP ELEKTRON Samarinda berdasarkan ulasan-ulasan pada siswa/i yang aktif mengikuti bimbingan belajar.

Kata Kunci : Analisis Sentimen, Klasifikasi Naive Bayes, Layanan, SKKNI

1. PENDAHULUAN

Lembaga Pendidikan Non-Formal saat ini dihadapkan pada persaingan yang semakin ketat. Untuk dapat berkembang dan bertahan, lembaga tidak hanya menawarkan produk atau jasa yang berkualitas, tetapi juga memberikan layanan yang baik kepada customer nya. Layanan merupakan faktor penting yang dapat mempengaruhi keputusan customer dalam memilih produk atau jasa. Customer yang mendapatkan layanan yang baik akan lebih cenderung untuk kembali dan merekomendasikan kepada kerabatnya ataupun orang lain.

Pada tahun 2021, LKP ELEKTRON pernah melakukan upaya peningkatan layanan bimbingan belajarnya, pihak LKP ELEKTRON membuat survey kelayakan pelayanan menggunakan Google Forms untuk teknik pengumpulan datanya. Akan tetapi untuk melakukan penilaian kelayakan pelayanan di LKP ELEKTRON masih menggunakan teknik observasi dengan mengumpulkan dan melihat ulasan siswa secara satu per satu yang membutuhkan banyak waktu, sehingga sulit dalam mengkategorikan dan menganalisis ulasannya. Untuk mengetahui kecenderungan komentar serta informasi yang terdapat dalam ulasan bukan hal yang mudah.

Oleh karena itu, dibutuhkan suatu metode yang dapat mengolah data tersebut dengan cara cepat dan otomatis untuk mengkategorikan ulasan-ulasan tersebut baik ulasan positif, negatif, maupun netral pada tiap tahunnya. Metode yang paling sederhana dari pengklasifikasi probabilitas dan memiliki tingkat kesalahan yang sangat minimum dibanding dengan algoritma lainnya adalah algoritma Naive Bayes Classifier. Metode tersebut dapat digunakan untuk mengetahui kecenderungan komentar dari ulasan siswa/i yang

aktif belajar di Bimbingan Belajar pada LKP ELEKTRON. Dari kecenderungan komentar ini dapat memberitahu pemilik LKP ELEKTRON tentang bagaimana sentimen terhadap layanan bimbingan belajar dan dari hasil analisis sentimen tersebut dapat mengambil sebuah tindakan yang sesuai untuk meningkatkan layanan bimbingan belajarnya. Selain itu hasil analisis sentimen ini dapat dijadikan sebagai bahan evaluasi untuk tahun berikutnya.

2. BATASAN MASALAH

Adapun batasan-batasan masalah dalam penelitian ini adalah :

1. Data yang digunakan adalah data ulasan dari siswa aktif Bimbingan Belajar pada LKP ELEKTRON Samarinda menggunakan Google forms.
2. *Output* dari analisis sentimen disajikan pada aplikasi berbasis *website*.
3. Algoritma yang digunakan adalah Naive Bayes Classifier.
4. Bahasa pemrograman pada penelitian ini menggunakan Python.
5. Metode Penelitian yang digunakan adalah Metode SKKNI (Standar Kompetensi Kerja Nasional Indonesia) pada bidang keahlian *Artificial Intelligence* yang berdasarkan Pemetaan Standar Kompetensi dari Keputusan Menteri Ketenagakerjaan RI No. 299 Tahun 2020.
6. *Library/Tools* yang dipakai dalam melakukan *pre-processing* menggunakan *Natural Language Toolkit* (NLTK).
7. *Library/Tools* yang dipakai untuk analisis dan manipulasi data menggunakan Pandas.

8. *Library/Tools* yang dipakai untuk memproses pembobotan kata pada metode *Term Frequency – Inverse Document Frequency* (TF-IDF) menggunakan *Scikit-Learn*.
9. *Library/Tools* yang dipakai untuk menciptakan visualisasi data menggunakan *Matplotlib*.
10. Metode yang dipakai untuk melakukan evaluasi terhadap performa model yang dibuat yaitu menggunakan metode *Confusion Matrix*.
11. *Library/Tools* yang dipakai untuk membangun sebuah *output* dari hasil analisis sentimen berupa *website* menggunakan *web service* dari Python yaitu *Streamlit*.
12. Metode Pengujian yang dipakai untuk melakukan evaluasi terhadap *website* yang dibuat yaitu menggunakan *Black Box Testing* dan *White Box Testing*.

3. DASAR TEORI

Suatu penelitian yang tidak dilandasi dengan teori yang lengkap menyebabkan pemecahan masalah tidak akan mencapai sasaran. Pada Penelitian ini ada beberapa teori-teori yang di-lakukan oleh peneliti untuk membahas suatu permasalahan yang dihadapi.

3.1 Naive Bayes Classifier

Naive Bayes Classifier merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Metode paling sederhana dari pengklasifikasian probabilitas adalah Naive Bayes Classifier. Naive Bayes adalah salah satu algoritma pem-belajaran induktif yang paling efektif dan efisien untuk machine learning dan data mining. Performa Naive Bayes yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keidependenan atribut (tidak ada kaitan antar atribut). Asumsi keidependenan atribut ini pada data sebenarnya jarang terjadi, namun walaupun asumsi keidependenan atribut tersebut dilanggar performa pengklasifikasian Naive Bayes cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris. Teorema pada NBC dikombinasikan dengan "naive" dimana diasumsikan kondisi antar atribut saling bebas. [1]

Cara kerja Naive Bayes Classifier diatur oleh teorema Bayes yang bekerja sesuai prinsip probabilitas bersyarat (conditional probability). Probabilitas bersyarat merepresentasikan suatu probabilitas atau peluang dari suatu peristiwa, yang mengacu pada peristiwa terkait yang telah terjadi. Berikut rumus probabilitas bersyarat :

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

Keterangan rumus (1) :

$P(A|B)$ = Probabilitas bersyarat A yang diberikan oleh B

$P(B|A)$ = Probabilitas bersyarat B yang diberikan oleh A

$P(A)$ = Probabilitas kejadian A

$P(B)$ = Probabilitas kejadian B

Untuk keperluan eksperimen, percobaan klasifikasi dilakukan dengan menggunakan judul berita sebanyak 5 buah. Dokumen tersebut dimanfaatkan sebagai dokumen latih (training). Satu artikel yang berbeda lainnya digunakan sebagai data uji (testing). [2]

Tabel 1. Data latih & uji untuk percobaan klasifikasi

Dokumen Judul	Judul Berita	Kategori
Doc Latih 1	Drama 6 Gol: Barcelona Petik Kemenangan, Messi Bikin Rekor Lagi	Olahraga
Doc Latih 2	YLKI Sebut Tarif MRT Rp 8.500 Cukup Fair dan Diterima Masyarakat	Ekonomi
Doc Latih 3	Zinedine Zidane Resmi Melatih Real Madrid Lagi	Olahraga
Doc Latih 4	Anggaran Naik, Kuota Mudik Gratis Bertambah Jadi 54 Ribu Orang	Ekonomi
Doc Latih 5	Bocoran Jersey Baru Barcelona Dikritik Netizen	Olahraga
Doc Uji 1	Prediksi Real Madrid vs Barcelona: Tuan Rumah Sedang Dertawakan	Belum diketahui

Selanjutnya melakukan tahap *Preprocessing* yang merupakan tahap awal penelitian yang bertujuan untuk menyamakan bentuk kata, mengurangi frekuensi kosa kata yang ada dalam dokumen agar menjadi data yang dapat diolah pada proses selanjutnya. Proses ini memiliki beberapa tahapan yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*.

Tabel 2. Hasil sampel dokumen setelah preprocessing

Dokumen Judul	Kata Dasar	Kategori
Doc Latih 1	drama gol barcelona petik menang messi bikin rekor	Olahraga
Doc Latih 2	ylki tarif mrt rp fair terima masyarakat	Ekonomi
Doc Latih 3	zinedine zidane resmi latih real madrid	Olahraga
Doc Latih 4	anggaran kuota mudik gratis tambah ribu orang	Ekonomi
Doc Latih 5	bocor jersey barcelona kritik netizen	Olahraga
Doc Uji 1	prediksi real madrid vs barcelona tuan rumah tawa	Belum diketahui

Kemudian, pada tahap *training* menggunakan 5 sampel dokumen latih seperti pada Tabel 2. Terdapat 32 kosa kata berbeda yang terdapat pada kelima dokumen training. Nantinya setiap kosa kata tersebut dihitung probabilitasnya terhadap dokumen yang sudah diketahui kategorinya.

Tabel 3. Kosa kata berbeda dari 5 dokumen training

Kosa Kata Hasil Training			
drama	ylki	zidane	garis
gol	tarif	resmi	tambah
barcelona	mrt	latih	ribu
petik	rp	real	orang
menang	fair	madrid	bocor
messi	terima	anggar	jersey
bikin	masyarakat	kuota	kritik
rekor	zinedine	mudik	netizen

Tabel 3 merupakan 32 kosa kata berbeda yang dihitung probabilitasnya, berdasarkan kategori olahraga dan ekonomi menggunakan perhitungan persamaan. Berikut perhitungan persamaan yang digunakan :

$$P(C_i) = \frac{f_d(C_i)}{|D|} \quad (2)$$

Keterangan rumus (2) :

$P(C_i)$ = Probabilitas kelas C_i

$f_d(C_i)$ = Frekuensi dokumen yang termasuk dalam kelas C_i

$|D|$ = Jumlah total dokumen dalam keseluruhan dataset

Dan juga, menggunakan perhitungan persamaan dari rumus $P(A|B)$ yang sudah mengalami penambahan proses *Laplacian smoothing* yang digunakan untuk proses klasifikasi dokumen. *Laplace smoothing* merupakan metode yang banyak digunakan, sekaligus *smoothing* yang disebut sebagai default *smoothing* dan *smoothing* tertua yang pernah di implementasikan pada Naive Bayes Classifier. *Laplace smoothing* juga disebut dengan *Add-one smoothing* karena metode *smoothing* ini menambahkan angka 1 pada setiap frekuensi token [3] . Rumus naive bayes dengan *laplace smoothing*, dapat dihitung menggunakan formula sebagai berikut :

$$P(W_k|C_i) = \frac{n_k+1}{n+|vocabulary|} \quad (3)$$

Keterangan rumus (3) :

$P(W_k|C_i)$ = Probabilitas bersyarat W_k yang diberikan oleh C_i

W_k = Kata ke-k dalam kosakata

C_i = Kelas ke-i. Kelas ini mewakili kategori sentimen

n_k = Jumlah kemunculan kata W_k dalam kelas C_i

n = Jumlah total kata dalam kelas C_i

$|vocabulary|$ = Ukuran kosakata (jumlah kata unik) dalam seluruh dataset

Pengerjaan menggunakan rumus (2) dan (3) :

P(Olahraga)	$= \frac{3}{ 5 } = 0,6$
P(drama Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(gol Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(barcelona Olahraga)	$= \frac{2+1}{19+ 32 } = 0,0588$
P(petik Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(menang Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(messi Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(bikin Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(rekor Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(zinedine Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(zidane Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(resmi Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(latih Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(real Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(madrid Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$

P(bocor Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(jersey Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(kritik Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$
P(netizen Olahraga)	$= \frac{1+1}{19+ 32 } = 0,03921$

Dengan mengacu pada persamaan yang sama dengan kategori olahraga, perhitungan dokumen berdasarkan kategori ekonomi menghasilkan angka :

P(Ekonomi)	$= \frac{2}{ 5 } = 0,4$
P(ylki Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(tarif Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(mrt Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(rp Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(fair Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(terima Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(masyarakat Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(anggar Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(kuota Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(mudik Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(gratis Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(tambah Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(ribu Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$
P(orang Ekonomi)	$= \frac{1+1}{14+ 32 } = 0,04347$

Kemudian, pada tahapan testing menggunakan 1 dokumen uji dan setiap kata yang ada pada dokumen uji dihitung probabilitasnya berdasarkan kategori yang sudah diketahui yaitu kategori olahraga dan kategori ekonomi. Dokumen testing yang akan digunakan seperti pada Tabel 4.

Tabel 4. Dokumen testing

Dokumen Judul	Input	Output
Doc Uji	prediksi real madrid vs barcelona tuan rumah tawa	Belum Diketahui

Selanjutnya dihitung probabilitas dokumen uji berdasarkan kategori olahraga dengan mengacu pada rumus persamaan (3) sebagai berikut :

P(prediksi Olahraga)	$= \frac{0+1}{19+ 32 } = 0,0196$
P(real Olahraga)	$= \frac{1+1}{19+ 32 } = 0,0392$
P(madrid Olahraga)	$= \frac{1+1}{19+ 32 } = 0,0392$
P(vs Olahraga)	$= \frac{0+1}{19+ 32 } = 0,0196$
P(barcelona Olahraga)	$= \frac{2+1}{19+ 32 } = 0,0588$
P(tuan Olahraga)	$= \frac{0+1}{19+ 32 } = 0,0196$
P(rumah Olahraga)	$= \frac{0+1}{19+ 32 } = 0,0196$
P(tawa Olahraga)	$= \frac{0+1}{19+ 32 } = 0,0196$

Adapun perhitungan yang dilakukan terhadap dokumen uji berdasarkan kategori ekonomi menghasilkan angka yang semua sama, yaitu 0,0219. Hasil probabilitas dokumen testing pada kategori olahraga adalah :

$$\begin{aligned}
 P(\text{Olahraga}) &= P(\text{Olahraga}) + P(\text{prediksi} \mid \text{Olahraga}) + \\
 &\quad P(\text{real} \mid \text{Olahraga}) + P(\text{madrid} \mid \text{Olahraga}) + \\
 &\quad P(\text{vs} \mid \text{Olahraga}) + P(\text{barcelona} \mid \text{Olahraga}) + \\
 &\quad P(\text{tuan} \mid \text{Olahraga}) + P(\text{rumah} \mid \text{Olahraga}) + \\
 &\quad P(\text{tawa} \mid \text{Olahraga}) \\
 &= 0,6 + 0,0196 + 0,0392 + 0,0392 + 0,0196 + \\
 &\quad 0,0588 + 0,0196 + 0,0196 + 0,0196 \\
 &= \mathbf{0,8352}
 \end{aligned}$$

Hasil probabilitas dokumen testing pada kategori ekonomi adalah:

$$\begin{aligned}
 P(\text{Ekonomi}) &= P(\text{Ekonomi}) + P(\text{prediksi} \mid \text{Ekonomi}) + \\
 &\quad P(\text{real} \mid \text{Ekonomi}) + P(\text{madrid} \mid \text{Ekonomi}) + \\
 &\quad P(\text{vs} \mid \text{Ekonomi}) + P(\text{barcelona} \mid \text{Ekonomi}) + \\
 &\quad P(\text{tuan} \mid \text{Ekonomi}) + P(\text{rumah} \mid \text{Ekonomi}) + \\
 &\quad P(\text{tawa} \mid \text{Ekonomi}) \\
 &= 0,4 + 0,0219 + 0,0219 + 0,0219 + 0,0219 + \\
 &\quad 0,0219 + 0,0219 + 0,0219 + 0,0219 \\
 &= \mathbf{0,5752}
 \end{aligned}$$

Dari hasil perhitungan probabilitas diatas yaitu $P(\text{Olahraga}) = 0,8352 > P(\text{Ekonomi}) = 0,5752$ maka dapat disimpulkan bahwa dokumen uji tersebut dikategorikan sebagai **Kategori Olahraga**.

3.2 Streamlit

Streamlit adalah *framework open source* dari Python yang memungkinkan untuk membuat aplikasi *website* menggunakan bahasa Python dalam mengaplikasi model dari *machine learning* atau *data science* [4]. Streamlit dapat diartikan yaitu kerangka kerja *website* yang ditujukan untuk menyebarkan model dan visualisasi dengan mudah menggunakan bahasa Python, yang cepat dan minimalis tetapi juga memiliki tampilan yang cukup baik serta ramah pengguna [5]. Tersedia *widget* bawaan untuk masukan pengguna, seperti pengunggahan gambar, penggeser, masukan teks, dan elemen *hypertext markup language* (HTML) lain yang sudah dikenal, seperti *checkboxes* dan *radio buttons*. Streamlit merupakan aplikasi yang tidak berbayar dan pengguna tidak perlu memiliki pengetahuan pengembangan *front-end* yang mahir untuk mengoperasikannya. Streamlit dapat dijalankan pada editor Anaconda serta bahasa Python seri 3.7 ke atas, tetapi tidak mendukung pada editor Jupyter notebook, sehingga harus dikonversi ke editor Pycharm atau Visual Code. Tampilan beranda pada aplikasi Streamlit dapat dipisahkan menjadi dua bagian, yaitu *buttons*, untuk pemilihan menu, serta tampilan *visual chart*. Hal ini menyebabkan dibutuhkan *library* NumPy serta Pandas untuk menampilkan grafik.

3.3 Standar Kompetensi Kerja Nasional Indonesia

Standar Kompetensi Kerja Nasional Indonesia (SKKNI) adalah rumusan kemampuan kerja yang mencakup aspek pengetahuan, keterampilan, dan/atau keahlian serta sikap kerja yang relevan dengan pelaksanaan tugas dan syarat jabatan yang ditetapkan [6]. SKKNI dikembangkan melalui konsultasi dengan industri terkait, untuk memastikan kesesuaian kebutuhan di tempat kerja.

Peraturan Menteri tenaga Kerja Republik Indonesia nomor 05 tahun 2012 menjelaskan bahwa SKKNI adalah rumusan kemampuan kerja yang mencakup aspek pengetahuan, keterampilan dan/atau keahlian, serta sikap kerja relevan dengan pelaksanaan tugas dan syarat jabatan yang ditetapkan sesuai dengan ketentuan peraturan perundang-undangan. SKKNI disusun untuk menetapkan unit-unit kompetensi yang akan dijadikan sebagai alat ukur kemampuan kerja seseorang yang meliputi aspek pengetahuan, keterampilan dan sikap kerja sebagaimana yang disyaratkan oleh pengguna jasa sumber daya manusia, dalam hal ini industri.

Adapun proses tahapan pengembangan dari metodologi berdasarkan keputusan menteri ketanagakerjaan RI nomor 299 tahun 2020 tentang penetapan SKKNI kategori informasi dan komunikasi golongan pokok aktivitas pemrograman, konsultasi computer dan kegiatan YBDI bidang keahlian artificial intelligence subbidang *data science*, bisa dilihat pada tabel 5.

Tabel 5. Pemetaan SKKNI Nomor 299 Tahun 2020

Tujuan Utama	Fungsi Kunci	Fungsi Utama	Fungsi Dasar
Menemukan pengetahuan insight atau pola yang bermanfaat dari data untuk berbagai keperluan	Meng-analisis kebutuhan organisasi	<i>Business understanding</i>	Menentukan objektif bisnis
			Menentukan tujuan teknis <i>data science</i>
			Membuat rencana proyek <i>data science</i>
		<i>Data understanding</i>	Mengumpulkan data
			Menelaah data
			Memvalidasi data
	Meng-embang-kan model	<i>Data preparation</i>	Memilah data
			Membersihkan data
			Mengkonstruksi data
			Menentukan label
		Mengintegrasikan data	
		<i>Modeling</i>	Membangun skenario pengujian
	Membangun model		
	<i>Model evaluation</i>	Mengevaluasi hasil pemodelan	
		Melakukan review proses pemodelan	
	Meng-gunakan model yang di-hasilkan	<i>Deployment</i>	Membuat rencana deployment model
			Melakukan deployment model
			Membuat rencana pemeliharaan
Melakukan pemeliharaan model			
<i>Evaluation</i>		Melakukan review proyek <i>data science</i>	
		Membuat laporan akhir proyek <i>data science</i>	

4. METODOLOGI PENELITIAN

4.1 Waktu dan Tempat Penelitian

Penelitian ini mulai dilaksanakan pada bulan Maret 2024 sampai dengan bulan Agustus 2024 dan dilaksanakan di LKP ELEKTRON yang berlokasi di Jl. Raden Wolter Monginsidi, Samarinda, Kalimantan Timur.

4.2 Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian :

a. Studi Pustaka

Data dan informasi didapat dari buku, artikel ilmiah, berita, maupun sumber kredibel lainnya yang reliabel dan juga sesuai dengan topik penelitian yang dilakukan.

b. Kuisisioner

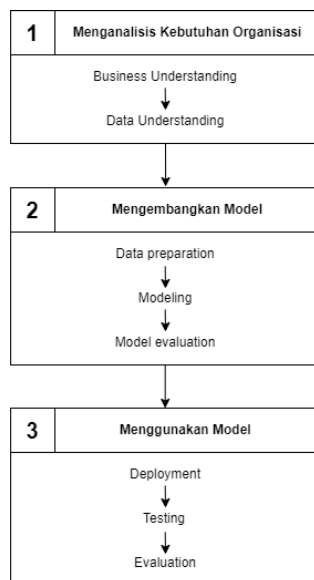
Memberikan pertanyaan kepada siswa/i aktif bimbingan belajar di LKP ELEKTRON Samarinda. Pertanyaan yang diajukan akan disebarakan secara *online* menggunakan Google forms.

c. Observasi

Menyebarkan Kuisisioner secara *online* untuk menjangkau semua siswa aktif bimbingan belajar dan datang langsung ke tempat bimbingan belajar di LKP ELEKTRON Samarinda untuk melakukan wawancara kepada salah satu stafnya.

4.3 Metode Pengembangan Sistem

Pada penelitian ini menggunakan metode dari Keputusan Menteri Ketenagakerjaan Republik Indonesia Nomor 299 Tahun 2020 Tentang Penetapan Standar Kompetensi Kerja Nasional Indonesia (SKKNI) dan Penambahan Tahapan *Testing* untuk menguji skrip dan program yang dibuat. Berikut kerangka kerja pada penelitian ini :



Gambar 1. Kerangka kerja pengembangan sistem

a. Business Understanding

Pada tahap ini peneliti melakukan sesuai dengan yang sudah tertulis di ketentuan tersebut seperti, Menentukan Objektif Bisnis, dalam penelitian ini objek dari penelitian adalah Siswa/i LKP ELEKTRON Samarinda. Setelah itu Menentukan Tujuan Teknis, dalam penelitian ini memiliki tujuan yaitu mengetahui sentimen terhadap ulasan dari siswa/i yang aktif pada Layanan Bimbingan Belajar di LKP ELEKTRON Samarinda. Kemudian, peneliti melakukan

pembuatan rencana proyek kedepannya terkait penelitian yang akan dilakukan.

b. Data Understanding

Tahap *Data Understanding* ini merupakan tahapan mengumpulkan data untuk diteliti, dalam penelitian ini dilakukan dengan cara menyebarkan Kuisisioner kepada siswa/i yang aktif bimbingan belajar LKP ELEKTRON menggunakan Platform dari microsoft yaitu Google forms. Kemudian menelaah dan mengeksplorasi terhadap data yang sudah dikumpulkan. Dan juga, memvalidasi data yang sudah dikumpulkan.

c. Data Preparation

Pada *Data Preparation*, Peneliti melakukan labeling data secara manual untuk menentukan setiap ulasan apakah mengandung nilai positif, negatif, ataupun netral. Setelah itu peneliti akan melakukan beberapa tahapan *Preprocessing* yang menggunakan metode *Natural Language Processing* dengan *library/tools* dari NLTK untuk melakukan tugas-tugas seperti:

1. *Case Folding*, mengubah semua huruf menjadi huruf kecil agar tidak ada kata-kata yang memiliki arti atau makna yang berbeda. Serta membersihkan emoji, karakter spesial dan angka.
2. *Tokenizing*, mengubah suatu kalimat menjadi kata-kata yang terpisah agar pada data tersebut setiap katanya memiliki nilai.
3. Normalisasi, melakukan pemuatan data berupa kamus singkatan kata dalam Bahasa Indonesia yang kemudian dilakukan proses normalisasi terhadap kata-kata yang berupa singkatan tertentu pada kolom Ulasan/ Komentar.
4. *Filtering (Stopword Removal)*, mengambil kata-kata penting dari tahap normalisasi dan menghilangkan kata-kata umum yang tidak memiliki makna.
5. *Stemming*, menghilangkan imbuhan yang berada di awal ataupun di akhir, sehingga terbentuknya kata dasar.

Setelah *Preprocessing* dilakukan, tahap selanjutnya yang dilakukan dalam *Data Preparation* adalah tahap pembobotan terhadap kata menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* dengan *library/tools* dari Scikit-Learn.

d. Modelling

Pada tahap *modeling* ini, data yang sudah diproses pada tahap data *preparation* akan dipisah menjadi 2 bagian data yaitu data latih dan data uji. Pada data latih, data terlebih dahulu dibuat seimbang (*balance*) menggunakan teknik *oversampling*. Setelah itu data latih akan dijadikan model *Machine Learning* menggunakan algoritma Naive Bayes Classifier. Pada data Uji, data akan diuji menggunakan hasil dari model yang telah dibuat untuk memprediksi data uji.

e. Model Evaluation

Pada tahap ini peneliti melakukan evaluasi dan mengukur kinerja dari model yang sudah diuji di tahap modeling menggunakan metode *confusion matrix*. Tahap ini akan menghasilkan *Precision* yang berfungsi untuk kecocokan antara bagian data yang diambil dengan informasi yang dibutuhkan. *Recall* yang berfungsi untuk mengukur tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *Accuracy* yang berfungsi untuk mengukur tingkat kedekatan antara nilai yang didapat terhadap nilai

sebenarnya. Dan juga, *F-Measure* yang merupakan perhitungan evaluasi dalam information retrieval yang menggabungkan *precision* dan *recall*.

f. Deployment

Pada tahap *deployment* ini, peneliti mengimplementasikan *model machine learning* dengan algoritma *naive bayes classifier* dan hasil analisis sentimen yang sudah dilakukan menggunakan *library/tools* dengan *web service* dari Python yaitu *streamlit* agar analisis sentimen yang sudah dibangun dapat ditampilkan secara *visual* dalam sebuah *website* yang interaktif dan memudahkan pemilik LKP ELEKTRON dalam membaca hasil dari analisa data ulasan siswa/i.

g. Testing

Pada tahap *testing*, peneliti menguji dan memeriksa kode atau skrip dari program dari analisis sentimen yang sudah dilakukan dengan menggunakan *white box testing*. Karena pada penelitian ini dilakukan hingga tahap *deployment*, maka peneliti juga menguji hasil dari *web service* yang sudah diimplementasikan dengan menggunakan metode *black box testing*.

h. Evaluation

Setelah dilakukan *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Model Evaluation*, *Deployment*, dan *Testing*. Maka pada tahap ini peneliti membuat laporan akhir guna mencatat atau menulis mengenai analisis sentimen yang sudah dibangun, dalam kasus ini membuat laporan akhir dalam bentuk Skripsi.

5. IMPLEMENTASI

Berikut adalah penjabaran mengenai tahapan – tahapan dalam membangun sistem analisis sentimen menggunakan Metode SKNI dan algoritma Naive Bayes Classifier berbasis *website*.

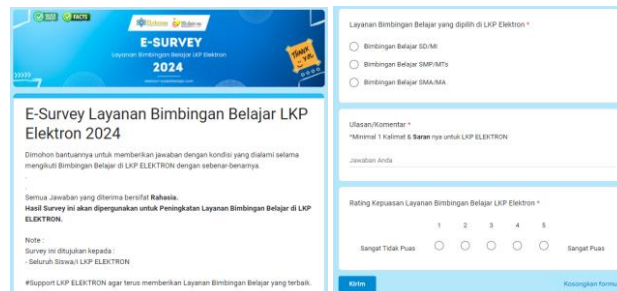
a. Business Understanding

Pada tahapan *Business Understanding*, peneliti melakukan penentuan objektif bisnis dan tujuan teknis *data science*. Pada objektif bisnis, LKP ELEKTRON dapat meningkatkan layanan bimbingan belajarnya pada tiap tahunnya untuk menjadikan sebagai salah satu LKP terfavorit khususnya di Kota Samarinda. Selain itu juga, tujuan *data science* dari penelitian ini adalah mengelompokkan ulasan siswa/i ke beberapa sentimen menggunakan model dari algoritma Naive Bayes Classifier. Selanjutnya menjalankan rencana proyek *data science* yaitu dengan cara melakukan tahapan dari *Data Understanding* sampai tahapan *Evaluation* secara berurutan, sehingga menghasilkan sebuah Laporan Akhir.

b. Data Understanding

Pada tahapan *Data Understanding*, Peneliti mengumpulkan data dengan cara melakukan Survey terhadap siswa/i aktif pada LKP ELEKTRON secara *online*. Survey dilakukan menggunakan *platform* dari Microsoft yaitu Google Forms. Kuisisioner yang digunakan dalam survey bersifat Kuisisioner terbuka, yang berisikan pertanyaan terbuka untuk siswa dapat menjawab sesuai pendapatnya pada Ulasan/Komentar terhadap Layanan Bimbingan Belajar di LKP ELEKTRON 2024.

Hasil Kuisisioner yang telah didapatkan pada Google forms akan diunduh dengan format CSV. Dan menggunakan *library* pandas untuk membaca file yang telah di download.



Gambar 2. Survey dari Google Forms

Untuk mempermudah dalam pemanggilan fitur pada kolom di *Dataframe*, maka dilakukanlah pengurangan kata, dengan mengambil kata pertama pada tiap kolom. Pada kolom "Layanan Bimbingan Belajar yang dipilih di LKP Elektron" diubah menjadi "Layanan", Pada kolom "Ulasan/Komentar" diubah menjadi "Ulasan", dan Pada Kolom "Rating Kepuasan Layanan Bimbingan Belajar LKP Elektron" diubah menjadi "Rating". Berikut hasil dari gambar 2 dengan menampilkan 5 data dari baris pertama.

Timestamp	Layanan	Ulasan	Rating
25/06/2024 15:10:32	Bimbingan Belajar SMA/MA	Terus Dikembangkan	4
25/06/2024 15:10:50	Bimbingan Belajar SMP/MTs	keren	4
25/06/2024 15:11:30	Bimbingan Belajar SMP/MTs	bagus	5
25/06/2024 15:13:59	Bimbingan Belajar SMA/MA	nyaman	5
25/06/2024 15:14:10	Bimbingan Belajar SMA/MA	Tidak ada saran	5

Gambar 3. Mengubah nama kolom

Pada Survey berhasil mengumpulkan 307 data dari Siswa/i, Pada Layanan Bimbingan Belajar SMA/MA memberikan Ulasan terbanyak dengan total 227 Ulasan. Dan siswa/i memberikan Rating 5 dengan total 203 data yang berarti sangat puas terhadap layanan bimbingan belajar.

	Timestamp	Layanan	Ulasan	Rating
count	307	307	307	307
unique	307	3	293	5
top	02/07/2024 21:49:10	Bimbingan Belajar SMA/MA	bagus	5
freq	1	227	4	203

Gambar 4. Ringkasan statistik

c. Data Preparation

Pada tahapan *Data Preparation*, Peneliti melakukan tahapan-tahapan seperti *Labeling*, *Case Folding*, *Tokenizing*, *Normalisasi*, *Filtering* (*Stopword Removal*), *Stemming* dan TF-IDF untuk pembersihan kata dan pembobotan kata dari ulasan siswa/i aktif bimbingan belajar di LKP ELEKTRON.

1. Labeling

Pada tahapan *Labeling*, Peneliti melakukan *Labeling* secara manual dari 307 data ulasan untuk dikelompokkan berdasarkan sentimen yang terdiri dari sentimen positif, sentimen negatif dan sentimen netral. Sentimen positif yang berarti sebuah ulasan tersebut memiliki perasaan yang baik tentang suatu hal. Sentimen negatif berarti sebuah ulasan tersebut memiliki perasaan yang kurang baik tentang suatu

hal. Dan sentimen netral berarti sebuah ulasan tersebut tidak memiliki perasaan yang baik ataupun kurang baik. Peneliti menggunakan 2 kolom yaitu kolom Ulasan dan Kolom Sentimen untuk melakukan ke tahapan ke selanjutnya.

	Ulasan	Sentimen
0	Terus Dikembangkan	Positif
1	keren	Positif
2	bagus	Positif
3	nyaman	Positif
4	Tidak ada saran	Netral
5	lebih bersih	Negatif

Gambar 5. Hasil labeling

2. Case Folding

Pada tahapan *Case Folding*, Peneliti melakukan pembersihan kata dengan menjadikan semua huruf ke huruf kecil agar tidak ada kata-kata yang memiliki arti atau makna yang berbeda dan mengubah emoji menjadi teks. Serta menghapus karakter spesial dan angka yang terdapat pada data ulasan siswa/i.

	Ulasan	Sentimen	Clean
0	Terus Dikembangkan	Positif	terus dikembangkan
1	keren	Positif	keren
2	bagus	Positif	bagus
3	nyaman	Positif	nyaman
4	Tidak ada saran	Netral	tidak ada saran
5	lebih bersih	Negatif	lebih bersih
6	perbaiki meja meja yang rusak	Negatif	perbaiki meja meja yang rusak

Gambar 6. Hasil case folding

3. Tokenizing

Pada tahapan *Tokenizing*, peneliti melakukan pemisahan kata dari data ulasan siswa/i sehingga data yang dihasilkan akan berbentuk List dengan Tipe data String. Tahapan ini menggunakan *library* NLTK (*Natural Language Tool Kit*) untuk pemrosesan datanya.

	Ulasan	Sentimen	Clean	Token
0	Terus Dikembangkan	Positif	terus dikembangkan	[terus, dikembangkan]
1	keren	Positif	keren	[keren]
2	bagus	Positif	bagus	[bagus]
3	nyaman	Positif	nyaman	[nyaman]
4	Tidak ada saran	Netral	tidak ada saran	[tidak, ada, saran]
5	lebih bersih	Negatif	lebih bersih	[lebih, bersih]
6	perbaiki meja meja yang rusak	Negatif	perbaiki meja meja yang rusak	[perbaiki, meja, meja, yang, rusak]

Gambar 7. Hasil tokenizing

4. Normalisasi

Pada tahapan *Normalisasi*, peneliti melakukan pemuan data berupa kamus singkatan kata dalam bahasa indonesia yang kemudian dilakukan proses normalisasi terhadap data ulasan siswa/i yang mengandung kata singkatan tertentu menjadi kata utuh.

	Ulasan	Sentimen	Clean	Token	Normalisasi
300	bagus bgtttt fav mudah di mengerti dan cepett	Positif	bagus bgtttt fav mudah di mengerti dan cepett	[bagus, bgtttt, fav, mudah, di, mengerti, dan, ...]	[bagus, banget, favorit, mudah, di, mengerti, ...]
301	Nyaman, bersih, pengajar yang asik juga ramah	Positif	nyaman bersih pengajar yang asik juga ramah	[nyaman, bersih, pengajar, yang, asik, juga, r, ...]	[nyaman, bersih, pengajar, yang, asik, juga, ...]
302	senang bgt bisa ngerasin bimmel disini, selai...	Positif	senang bgt bisa ngerasin bimmel disini selai...	[senang, bgt, bisa, ngerasin, bimmel, disini, ...]	[senang, banget, bisa, ngerasin, bimmel, disi, ...]
303	Bagus	Positif	bagus	[bagus]	[bagus]
304	Sudah sangat baik	Netral	sudah sangat baik	[sudah, sangat, baik]	[sudah, sangat, baik]
305	Bimbingan elektron sdh sangat baik, proses be...	Positif	bimbingan elektron sdh sangat baik proses be...	[bimbingan, elektron, sdh, sangat, baik, prose, ...]	[bimbingan, elektron, sudah, sangat, baik, pro, ...]
306	sukses selalu elektronn hati hati hati	Netral	sukses selalu elektronn hati hati hati	[sukses, selalu, elektronn, hati, hati, hati, ...]	[sukses, selalu, elektron, hati, hati, hati, ...]

5. Filtering (Stopword Removal)

Pada tahapan *filtering (Stopword removal)*, peneliti mengambil kata-kata penting dari tahap normalisasi dan menghilangkan kata-kata umum yang tidak memiliki makna menggunakan kamus stopwords indonesia. Dan penambahan data pada kamus stopwords oleh peneliti untuk menghilangkan kata-kata yang tidak memiliki makna seperti "ehehehhe", "xixixi" dan sejenisnya.

	Ulasan	Sentimen	Clean	Token	Normalisasi	Stopwords
300	bagus bgtttt fav mudah di mengerti dan cepett	Positif	bagus bgtttt fav mudah di mengerti dan cepett	[bagus, bgtttt, fav, mudah, di, mengerti, dan, ...]	[bagus, banget, favorit, mudah, di, mengerti, ...]	[bagus, banget, favorit, mudah, di, mengerti, cepett]
301	Nyaman, bersih, pengajar yang asik juga ramah	Positif	nyaman bersih pengajar yang asik juga ramah	[nyaman, bersih, pengajar, yang, asik, juga, r, ...]	[nyaman, bersih, pengajar, yang, asik, juga, ...]	[nyaman, bersih, pengajar, asik, ramah]
302	senang bgt bisa ngerasin bimmel disini, selai...	Positif	senang bgt bisa ngerasin bimmel disini selai...	[senang, bgt, bisa, ngerasin, bimmel, disini, ...]	[senang, banget, bisa, ngerasin, bimmel, disi, ...]	[senang, banget, ngerasin, bimmel, teman, asy, ...]
303	Bagus	Positif	bagus	[bagus]	[bagus]	[bagus]
304	Sudah sangat baik	Netral	sudah sangat baik	[sudah, sangat, baik]	[sudah, sangat, baik]	[]
305	Bimbingan elektron sdh sangat baik, proses be...	Positif	bimbingan elektron sdh sangat baik proses be...	[bimbingan, elektron, sdh, sangat, baik, prose, ...]	[bimbingan, elektron, sudah, sangat, baik, pro, ...]	[bimbingan, elektron, proses, belajar, mengaja, ...]
306	sukses selalu elektronn hati hati hati	Netral	sukses selalu elektronn	[sukses, selalu, elektronn]	[sukses, selalu, elektron]	[sukses, elektron]

Gambar 8. Hasil Filtering (Stopword Removal)

6. Stemming

Pada tahapan *stemming*, peneliti menghilangkan imbuhan yang berada di awal ataupun di akhir kata pada data ulasan siswa/i, sehingga membentuk kata dasar. Pada proses *stemming*, peneliti menggunakan *library* dari NLTK yaitu *StemmerFactory* untuk meningkatkan akurasi tugas pemrosesan teks seperti pencarian, klasifikasi, dan pengelompokan kata. Contoh imbuhan awalan (*prefiks*) seperti ber-, di-, dan per-. Dan contoh imbuhan akhiran seperti -kan, -an, dan -nya.

	Ulasan	Sentimen	Clean	Token	Normalisasi	Stopwords	Stemmer
300	bagus bgtttt fav mudah di mengerti dan cepett	Positif	bagus bgtttt fav mudah di mengerti dan cepett	[bagus, bgtttt, fav, mudah, di, mengerti, dan, ...]	[bagus, banget, favorit, mudah, di, mengerti, ...]	[bagus, banget, favorit, mudah, di, mengerti, cepett]	[bagus, banget, favorit, mudah, di, mengerti]
301	Nyaman, bersih, pengajar yang asik juga ramah	Positif	nyaman bersih pengajar yang asik juga ramah	[nyaman, bersih, pengajar, yang, asik, juga, r, ...]	[nyaman, bersih, pengajar, yang, asik, juga, ...]	[nyaman, bersih, pengajar, asik, ramah]	[nyaman, bersih, ajar, asik, ramah]
302	senang bgt bisa ngerasin bimmel disini, selai...	Positif	senang bgt bisa ngerasin bimmel disini selai...	[senang, bgt, bisa, ngerasin, bimmel, disini, ...]	[senang, banget, bisa, ngerasin, bimmel, teman, asy, ...]	[senang, banget, ngerasin, bimmel, teman, asy, ...]	[senang, banget, ngerasin, bimmel, teman, asy, ...]
303	Bagus	Positif	bagus	[bagus]	[bagus]	[bagus]	[bagus]
304	Sudah sangat baik	Netral	sudah sangat baik	[sudah, sangat, baik]	[sudah, sangat, baik]	[]	[]
305	Bimbingan elektron sdh sangat baik, proses be...	Positif	bimbingan elektron sdh sangat baik proses be...	[bimbingan, elektron, sdh, sangat, baik, prose, ...]	[bimbingan, elektron, sudah, sangat, baik, pro, ...]	[bimbingan, elektron, proses, belajar, mengaja, ...]	[bimbing, elekt, proses, ajar, ajar, santai, ...]
306	sukses selalu elektronn hati hati hati	Netral	sukses selalu elektronn	[sukses, selalu, elektronn]	[sukses, selalu, elektron, hati, hati, hati, ...]	[sukses, elektron, hati, hati, hati, ...]	[sukses, elektron, hati, hati, hati, ...]

Gambar 9. Hasil stemming

Hasil Data *stemming* disatukan kembali menjadi sebuah kalimat yang memiliki makna pada tiap kata dari data ulasan siswa/i yang akan digunakan sebagai bahan untuk melakukan pelatihan dan pengujian pada model Naive Bayes Classifier.

	Ulasan	Sentimen	Clean	Token	Normalisasi	Stopwords	Stemmer	Stemmer_text
300	bagus bgtttt fav mudah di mengerti dan cepett	Positif	bagus bgtttt fav mudah di mengerti dan cepett	[bagus, bgtttt, fav, mudah, di, mengerti, dan, ...]	[bagus, banget, favorit, mudah, di, mengerti, ...]	[bagus, banget, favorit, mudah, di, mengerti, cepett]	[bagus, banget, favorit, mudah, di, mengerti]	bagus banget favorit mudah di mengerti cepett
301	Nyaman, bersih, pengajar yang asik juga ramah	Positif	nyaman bersih pengajar yang asik juga ramah	[nyaman, bersih, pengajar, yang, asik, juga, r, ...]	[nyaman, bersih, pengajar, yang, asik, juga, ...]	[nyaman, bersih, pengajar, asik, ramah]	[nyaman, bersih, ajar, asik, ramah]	nyaman bersih ajar asik ramah
302	senang bgt bisa ngerasin bimmel disini, selai...	Positif	senang bgt bisa ngerasin bimmel disini selai...	[senang, bgt, bisa, ngerasin, bimmel, disini, ...]	[senang, banget, bisa, ngerasin, bimmel, teman, asy, ...]	[senang, banget, ngerasin, bimmel, teman, asy, ...]	[senang, banget, ngerasin, bimmel, teman, asy, ...]	senang banget ngerasin bimmel teman asy gur...
303	Bagus	Positif	bagus	[bagus]	[bagus]	[bagus]	[bagus]	bagus
304	Sudah sangat baik	Netral	sudah sangat baik	[sudah, sangat, baik]	[sudah, sangat, baik]	[]	[]	
305	Bimbingan elektron sdh sangat baik, proses be...	Positif	bimbingan elektron sdh sangat baik proses be...	[bimbingan, elektron, sdh, sangat, baik, prose, ...]	[bimbingan, elektron, sudah, sangat, baik, pro, ...]	[bimbingan, elektron, proses, belajar, mengaja, ...]	[bimbing, elekt, proses, ajar, ajar, santai, ...]	bimbing elektron proses ajar ajar santai mudah...
306	sukses selalu elektronn hati hati hati	Netral	sukses selalu elektronn	[sukses, selalu, elektronn]	[sukses, selalu, elektron, hati, hati, hati, ...]	[sukses, elektron, hati, hati, hati, ...]	[sukses, elektron, hati, hati, hati, ...]	sukses elektron hati hati hati

Gambar 10. Hasil gabungan stemming

Setelah melakukan menggabungkan hasil stemming, peneliti menghapus data kosong pada fitur "Stemmer_text". Terdapat 16 data kosong pada fitur "Stemmer_text" dan Tersisa 291 data ulasan siswa/i dari LKP ELEKTRON yang nantinya akan dipergunakan untuk tahapan selanjutnya.

	Timestamp	Layanan	Ulasan	Rating	Sentimen	Clean	Token	Normalisasi	Stopwords	Stemmer	Stemmer_text
count	291	291	291	291	291	291	291	291	291	291	291
unique	235	3	282	5	3	279	277	274	269	267	267
top	23-06-24 15:14	Bimbingan Belajar SMA/MA	bagus	5	Positif	bagus	[keren]	[bagus]	[bagus]	[bagus]	bagus
freq	6	217	4	193	170	5	5	7	8	8	8

Gambar 11. Ringkasan statistik akhir

7. TF-IDF

Pada tahapan TF-IDF, peneliti menggunakan TF-IDF untuk melihat kata yang sering muncul di dalam ulasan siswa/i di LKP ELEKTRON. Disini peneliti ingin melihat 10 data yang sering muncul. Berikut ranking kata dari 10 kata yang memiliki frekuensi tertinggi berdasarkan dari 291 ulasan.

term	rank
0	ajar 58.034845
3	elektron 48.911437
4	guru 43.793022
6	moga 34.473071
2	bagus 31.812567
9	saran 28.647657
8	paham 26.679010
1	asyik 25.165996
5	kelas 23.224087
7	nyaman 22.440047

Gambar 12. Ranking Kata

d. Modelling

Pada tahapan *Modelling*, Peneliti menggunakan *library sklearn* untuk pembuatan model Naive Bayes Classifier. Untuk mempermudah dalam pembuatan model, Sentimen yang awalnya huruf di konversi menjadi angka. Dengan mengubah nama Sentimen Positif bernilai 1, Sentimen Netral bernilai 0 dan Sentimen Negatif bernilai -1.

Selanjutnya peneliti, membagi data latih dan data uji untuk pembuatan model Naive Bayes Classifier. Pada data dari hasil *Pre-paration* dibagi menjadi 80:20, dimana 80% adalah data latih dan 20% adalah data Uji. Data latih sebanyak 232 data dan Data uji sebanyak 59 Data.

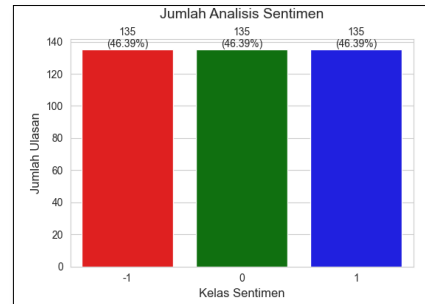
```
y_train.value_counts()
```

Sentimen	count
1	135
0	52
-1	45

Name: count, dtype: int64

Gambar 13. Sentimen pada data latih

Data sentimen pada data latih harus Balance/Seimbang agar meningkatkan akurasi dari akurasi Model yang ingin dibuat. Untuk melakukan keseimbangan Data latih, diperlukannya *oversampling* agar semua kategori sentimen mempunyai nilai yang sama. Pada penelitian ini menggunakan Random Over Sampler sebagai teknik *oversampling*. Berikut adalah hasil visualisasi setelah melakukan *oversampling* pada data latih.



Gambar 14. Visualisasi hasil oversampling

Selanjutnya pembuatan model dari data latih yang sudah di *oversampling* menggunakan *library* dari naive bayes yaitu *MultinomialNB*. Model yang sudah dibuat, harus dievaluasi terlebih dahulu sebelum diterapkan untuk melakukan analisa data E-survey yang baru.

e. Model Evaluation

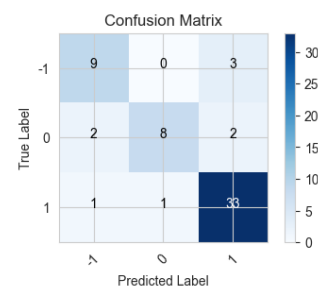
Pada tahapan *Model Evaluation*, Peneliti melakukan penghitungan akurasi dan laporan klasifikasi beserta *mapheat* dari *confusion matrix*. Peneliti menggunakan *library* *itertools* dan *sklearn*. Berikut hasil akurasi dan juga laporan klasifikasi dari sebuah model yang telah dibuat menggunakan metode *oversampling* dari Random Over Sampler (ROS) yang memiliki akurasi sebesar 0.847 atau 84.7%.

Akurasi Model Naive Bayes : 0.847457627118644

Laporan Klasifikasi :				
	precision	recall	f1-score	support
Negatif	0.75	0.75	0.75	12
Netral	0.89	0.67	0.76	12
Positif	0.87	0.94	0.90	35
accuracy			0.85	59
macro avg	0.84	0.79	0.81	59
weighted avg	0.85	0.85	0.84	59

Gambar 15. Akurasi dan Laporan Klasifikasi

Selanjutnya pembuatan visualisasi data menggunakan *heatmap* dari Evaluasi model dengan metode *Confusion matrix* untuk menampilkan jumlah data uji yang di-klasifikasikan sehingga memudahkan dalam mengevaluasi akurasi suatu sistem klasifikasi. Berikut *heatmap* dari metode *Confusion Matrix* yang memiliki total 59 Data Uji.

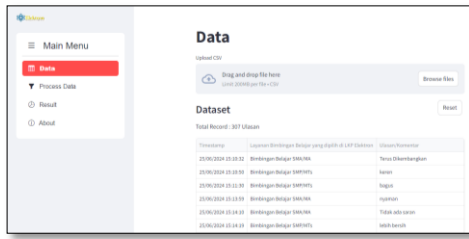


Gambar 16. Confusion Matrix

f. Deployment

Pada tahapan *Deployment*, peneliti menggunakan *web service* dari Python yaitu *streamlit* untuk mendeployment model yang sudah dibangun dapat ditampilkan secara *visual* dalam sebuah *website* yang interaktif.

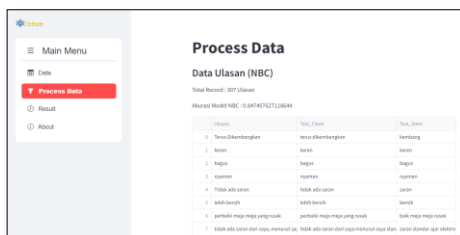
1. Halaman Menu Data



Gambar 17. Halaman menu data

Pada halaman menu data, menampilkan *input data file* dan *Dataset* dari hasil inputan data yang telah diupload pada menu data. Pada menu ini juga terdapat info *total record* dari *Dataset* dan *button reset* untuk membersihkan semua data.

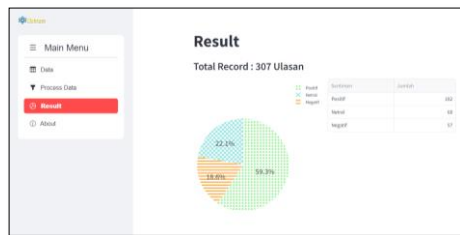
2. Halaman Menu Process Data



Gambar 18. Halaman menu process data

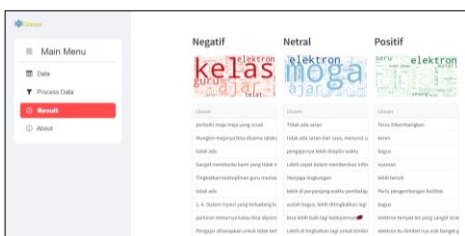
Pada halaman menu *Process Data*, menampilkan kolom *Ulasan*, *Text_Clean*, *Text_Stem* dan *Sentimen* dari hasil pemrosesan data menggunakan model *Naive Bayes Classifier*. Pada menu ini juga terdapat info *total record* dan tingkat akurasi pada model yang dipakai.

3. Halaman Menu Result



Gambar 19. Halaman menu result

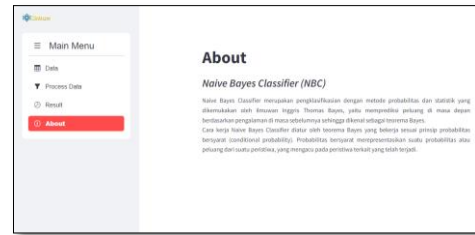
Pada halaman menu *Result*, menampilkan *Grafik* dalam bentuk *Pie Chart* dan pengklasifikasian data berdasarkan sentimen positif, sentimen netral dan sentimen negatif. Pada menu ini juga terdapat info *total record* dan jumlah data pada setiap sentimennya. Dan terdapat halaman lanjutan pada menu *result* yang bisa dilihat pada gambar 20.



Gambar 20. Halaman lanjutan menu result

Pada halaman lanjutan menu *Result*, menampilkan kumpulan kata kata yang berbentuk *wordcloud* dan menampilkan list data ulasan dari sentimen positif, sentimen netral dan sentimen negatif untuk memudahkan dalam mengambil kesimpulan pada tiap sentimennya.

4. Halaman Menu About



Gambar 21. Halaman menu about

Pada halaman *about*, menampilkan tentang dari model *Naive Bayes Classifier* yang digunakan untuk pemrosesan data dari *Ulasan Siswa/i aktif di LKP ELEKTRON*.

g. Testing

Pada tahapan *Testing*, Peneliti akan melakukan 2 jenis *Testing*, yaitu *Black box Testing* dan *White box Testing* sebagai alat pengujian agar sistem dapat bekerja semestinya.

Tabel 6. *Black box testing*

No	Kegiatan Pengujian	Hasil yang diharapkan	Jumlah Pengujian		
			1	2	3
1	Input Data E-survey	Pada menu ini, file akan disimpan pada <i>database</i> dengan format <i>CSV</i> .	✓	✓	✓
2	Reset Data E-Survey	Pada menu ini, semua file yang ada di <i>database</i> akan di hapus dan aplikasi siap untuk digunakan ulang.	✓	✓	✓
3	Proses Data E-survey	Pada menu ini, data akan diproses hingga menghasilkan sentimen pada tiap ulasan menggunakan model dari <i>NBC</i> . Data yang sudah diproses akan disimpan pada <i>Database</i> .	✓	✓	✓
4	Hasil Akhir Data E-Survey	Pada menu ini, data akan disajikan dalam bentuk <i>pie chart</i> dan <i>wordcloud</i> sesuai dengan sentimen.	✓	✓	✓

Keterangan :

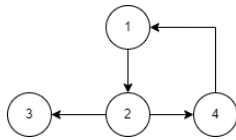
Penguji 1 : Mahasiswa/Peneliti

Penguji 2 : Admin LKP Elektron

Penguji 3 : Pemilik LKP Elektron

Pada tahapan *White box testing*, peneliti perlu mengetahui secara dalam *source code* yang akan diuji dengan menggunakan teknik *basis path* untuk mengukur kompleksitas dari suatu program dengan cara mengidentifikasi semua jalur yang mungkin dilalui oleh program. Dan juga perhitungan *Cyclomatix Complexity V(G)* untuk menentukan jumlah path yang selanjutnya digunakan untuk *test case* dari *source code* yang akan diuji.

1. *White box testing* input data e-survey

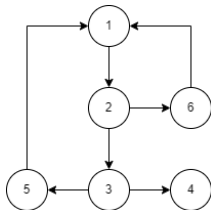


Gambar 22. Flowgraph input data e-survey

Tabel 7. Test case input data e-survey

No	Alir Program	Keterangan	Hasil Pengujian
1	1-2-4-1	Jika tidak terdapat file upload	Berhasil
2	1-2-3	Jika terdapat file upload dalam bentuk CSV	Berhasil

2. *White box testing* reset data e-survey

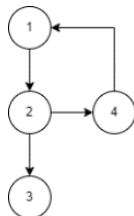


Gambar 23. Flowgraph reset data e-survey

Tabel 8. Test case reset data e-survey

No	Alir Program	Keterangan	Hasil Pengujian
1	1-2-3-4	Berhasil Menghapus seluruh file	Berhasil
2	1-2-3-5-1	Gagal menghapus file, jika salah satu file tidak ada didalam database.	Berhasil
3	1-2-6-1	Jika tidak terdapat file didalam database	Berhasil

3. *White box testing* proses data e-survey



Gambar 24. Flowgraph proses data e-survey

Tabel 9. Test case proses data e-survey

No	Alir Program	Keterangan	Hasil Pengujian
1	1-2-3	Jika data E-survey berhasil di prediksi menggunakan NBC	Berhasil
2	1-2-4-1	Tidak adanya Data E-Survey yang diupload	Berhasil

h. *Evaluation*

Pada tahapan Evaluation, peneliti melakukan review dari seluruh tahapan pengembangan sistem yang sudah dikerjakan. Dan peneliti membuat laporan akhir berupa skripsi yang diharapkan menjadi bahan referensi untuk peneliti lainnya.

6. **KESIMPULAN**

Berdasarkan hasil pembahasan dari Penerapan Naive Bayes Classifier untuk Analisis Sentimen Terhadap Layanan Bimbingan Belajar di LKP ELEKTRON, maka dapat ditarik kesimpulan sebagai berikut :

1. Sistem untuk Penerapan Naive Bayes Classifier untuk Analisis Sentimen Terhadap Layanan Bimbingan Belajar di LKP ELEKTRON dibuat berbasis *website* menggunakan Streamlit.
2. Proses membangun model pada sistem menggunakan metode SKKNI yang memiliki 7 Tahapan dan penambahan tahapan Testing pada Sistem yang dibuat, tahapannya terdiri : *Business Understanding, Data Understanding, Data Preparation, Modelling, Model Evaluation, Deployment, Testing, Evaluation.*
3. Tersisa 291 Data Bersih dari 307 Data ulasan Siswa/i aktif di LKP ELEKTRON yang didapat dari hasil tahapan pada *Data Preparation.*
4. Pada model Naive Bayes Classifier, dibangun menggunakan 80% Data latih dan 20% Data uji dari 291 Data Ulasan.
5. Penggunaan *Random Over Sampler* untuk melakukan *oversampling* pada data latih agar data tiap sentimen bernilai sama (*Balance*)
6. Tingkat Akurasi yang didapat pada Evaluasi model yaitu 0.847 atau dipersentasekan menjadi 84.7 %.

DAFTAR PUSTAKA

- [1] Syarli, & Asrul Ashari Muin. (2016). Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi). *Jurnal Ilmiah Ilmu Komputer*, 2(1).
- [2] Hidayat, E. Y., & Rizqi, M. A. (2020). Klasifikasi Dokumen Berita Menggunakan Algoritma Enhanced Confix Stripping Stemmer dan Naïve Bayes Classifier. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2).
- [3] Listiowarni, I., & Rahayu Setyaningsih, E. (2018). Analisis Kinerja Smoothing pada Naive Bayes untuk Pengkategorian Soal Ujian. *Jurnal Teknologi Dan Manajemen Informatika*, 4(2).
- [4] Ferdyanandi, M., Setiawan, N. Y., & Abdurrachman Bachtiar, F. (2022). Prediksi Potensi Penjualan Makanan Beku Berdasarkan Ulasan Pengguna Shopee Menggunakan Metode Decision Tree Algoritma C4.5 Dan Random Forest (Studi Kasus Dapur Lilis). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(2).
- [5] Widi Hastomo, Nur Aini, Adhitio Satyo Bayangkari Karno, & L.M. Rasdi Rere. (2022). Metode Pembelajaran Mesin untuk Memprediksi Emisi Manure Management. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 11(2).
- [6] Falakh, F. (2022). Identifikasi Standar Kompetensi Kerja Nasional Indonesia (SKKNI) pada Pengembangan Kurikulum Program Studi Teknik Lingkungan. *Journal on Education*, 4(4).