ANALISIS SENTIMEN TERHADAP GENERASI Z DALAM ETIKA KERJA PADA MEDIA SOCIAL MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Firman Frezy Pradana, Karina Auliasari, Mira Orisa
Teknik Informatika, Institut Teknologi Nasional Malang
Jalan Raya Karanglo km 2 Malang, Indonesia
2118112@scholar.itn.ac.id

ABSTRAK

Etika kerja merupakan norma yang menjadi pedoman dalam bertindak dan bersikap dalam lingkungan kerja. Generasi z kini sedang memasuki masa transisi menuju kedewasaan dan mulai memasuki dunia kerja. Karena tumbuh dan berkembang di era teknologi , generasi ini memiliki karakter yang berbeda dari generasi sebelumnya seperti fleksibel dan adaptis, peduli terhadap kesehatan mental dan kesejahteraan, serta rasa keingintahuan yang tinggi. Namun, karena perbedaan tersebut menimbulkan beragam respon dari masyarakat, baik mendukung maupun mengkritik. Tujuan penelitian untuk menganalisis sentimen masyarakat terhadap isu etika kerja generasi z, dengan mengimplementasikan metode *support vector machine*, untuk mengklasifikan sudut pandang masyarakat ke dalam kategori positif atau negatif. Hasil evaluasi pengujian mendapatkan nilai *recall* sebesar 60% ,*precision* sebesar 68%, dan *accuracy* sebesar 69%. Hasil klasifikasi sentimen menunjukkan sentimen negatif mendominasi, menandakan kecenderungan opini masyarakat terhadap etika kerja generasi z mengarah pada pandangan negatif.

Kata kunci: *Etika Kerja, Generasi z, Support Vector Machine, Analisis Sentimen, z, Media Sosial*

1. PENDAHULUAN

Etika kerja merupakan prinsip norma yang menjadi pedoman dalam bertindak dan bersikap di dalam lingkungan kerja. Etika kerja berperan sebagai faktor utama dalam mendorong peningkatan kinerja dan perkembangan suatu instansi, karena etika yang baik akan memberikan kontribusi besar dalam mencapai tujuan perusahaan dengan hasil yang optimal. Selain itu, etika kerja yang diterapkan dengan konsisten terbukti memiliki pengaruh besar terhadap peningkatan kualitas individu maupun instansi[1].

Generasi Z atau Gen Z merupakan kelompok individu yang lahir pada rentan waktu antara tahun 1997 hingga 2012. Kelompok usia ini, berada pada fase transisi menuju kedewasaan, di mana sebagian besar telah menyelesaikan pendidikan menengah atau tinggi dan mulai memasuki dunia kerja. Karena tumbuh dan berkembang di tengah pesatnya teknologi digital, gen z memiliki *prespektif* yang unik terhadap etika kerja, yang pada gilirannya mendorong perubahan dalam dunia kerja. Gen z memiliki pendekatan kerja yang berbeda, seperti lingkungan kerja fleksibel, aturan yang longgar, serta kepedulian terhadap kesejahteraan dan kesehatan mental sementara semangat belajar , keterbukaan terhadap umpan balik, dan kemampuan *multitasking*, menjadikan generasi z tenaga kerja yang cekatan dan inovatif [2].

Pendekatan kerja yang ditunjukkan oleh gen z tentunya menimbulkan beragam respon dari masyarakat, baik dalam bentuk dukungan maupun kritik. Sejumlah karyawan rekan kerja gen z mulai menyuarakan opini pada media sosial. Banyak anggapan kurangnya motivasi dalam bekerja serta keterampilan yang dinilai buruk pada gen z [3].

Guna mendapatkan pemahaman yang lebih terhadap isu etika kerja, pendekatan yang dapat digunakan adalah analisis sentimen. analisis sentimen adalah proses pendekatan terstruktur untuk menilai kecenderungan emosional atau sikap dalam suatu teks terhadap object tertentu. Tujuan analisis sentimen adalah mengungkapkan dan mengklasifikasikan opini, sentimen, dan emosi dari teks tidak terstruktur seperti media sosial [4].

Metode yaang digunakan dalam penelitian menggunakan algoritma *support vector machine(SVM)* untuk mengklasifikasikan opini masyarakat menjadi 2 kategori yaitu positif dan negatif. Algoritma SVM didasarkan pada pencarian *hyperplane* optimal yang berfungsi sebagai pembatas pemisah antara dua kelas. Algoritma SVM unggul dalam hal akurasi yang tinggi, efisiensi penggunaan memori, serta kemampuan dalam menangani data yang tidak terdistribusi secara normal [5]

Hasil penelitian ini diharapkan dapat membantu dalam pengembangan studi kecerdasan *komputasi*, khususnya dalam penerapan dan pengembangan pembelajaran mesin untuk analisis sentimen. Selain itu , hasil analisis sentimen yang diperoleh dapat dijadikan evaluasi berbagai pihak dalam merumuskan kebijakan ataupun strategi yang lebih baik untuk generasi saat ini dan ke depannya.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian yang dilakukan Herlinda berjudul "Analisis *Sentimen* Masyarakat Terhadap Generasi Z Dalam Dunia Kerja Pada Media Sosial Twitter Menggunakan Metode *Naïve Bayes*" bertujuan menganalisis opini gen z terhadap dunia kerja dengan menggunakan metode *naïve bayes* dan

membandingkan akurasi pelabelan *lexicon* dan manual. Hasil evaluasi menunjukkan pelabelan manual memiliki akurasi lebih tinggi dibandingkan pelabelan *lexicon* saat *stopword* digunakan, yaitu akurasi 87% berbanding 76%, namun setelah *stopword* dihilangkan, akurasi pelabelan *lexicon* meningkat signifikan hingga 96%, melampaui akurasi pelabelan manual yang mencapai 90% [6]

Yahya dalam penelitian berjudul "Analisis Sentimen Untuk Deteksi Ujaran Kebencian Pada Media Sosial Terkait Pemilu 2024 Menggunakan Metode *Support Vector Machine*" berfokus pada merancang sistem berbasis *web* dengan menerapkan metode *SVM* untuk mendeteksi ujaran kebencian yang berkaitan dengan pemilu dalam kalimat bahasa Indonesia. Hasil dari pengujian 100 data *tweet* menunjukkan algoritma berhasil dalam mendeteksi ujaran kebencian dan non-kebencian dengan nilai *reccal* 76% , presicion 96% dan tingkat akurasi 81% [7].

Dalam penelitiannya yang berjudul "Pemetaan Opini Publik Terhadap Perubahan Kebijakan BPJS Kesehatan Dengan Pendekatan Support Vector Machine(SVM) Dalam Analisis Sentimen", Damayanti menganalisis respon masyarakat terhadap perubahan kebijakan iuran BPJS kesehatan melalui penerapan metode *SVM*. Berdasarkan hasil pengujian, sistem berhasil mengidentifikasi sentimen terhadap perubahan kebijakan BPJS dengan tingkat akurasi sebesar 94,28% [8].

Dalam penelitian yang berjudul "Analisis Sentimen Terhadap Pengguna Gojek Dan Grab Pada Media Sosial Twitter Menggunakan *Random Forest* " Melia menyatakan bahwa tujuan penelitian untuk mengklasifikasikan opini pengguna terhadap layanan transportasi online dengan menggunakan algoritma *Random Forest*. Dari penelitian yang dilakukan hasil menunjukkan komentar positif lebih mendominasi dari pada komentar negatif, hal tersebut menunjukkan algoritma *Random Forest* cukup baik dalam melakukan klasifikasi dengan nilai akurasi sebesar 76,25% [9]

Penelitian yang dilakukan Ramadani dengan judul "Perbandingan Algoritma *Support Vector Machine, Decision Tree,* dan Logistic Regresion Pada Analisis Sentimen Ulasan Aplikasi *Netflix*". Penelitian ini bertujuan menganalisis sentimen terhadap penggunaan *netflik*. Hasil pengujian antara 3 metode menunjukkan algoritma *SVM* konsisten memberikan klasifikasi yang lebih baik dengan tingkat akurasi rata-rata 88.18% dan nilai tertinggi pada *K-Fold Cross Validation* mencapai 92.69% [10].

2.2. Text Mining

Text mining merupakan proses menganalisis serta mengekstrak informasi yang relevan dari suatu dokumen. Tujuannya untuk memperoleh informasi yang bernilai dari sejumlah dokumen , terutama dalam kumpulan dokumen yang berukuran besar. Dalam beberapa kasus, *text mining* sering digunakan dalam menyelesaikan permasalahan klasifikasi, *clustering*, dan ekstraksi informasi. Secara umum, *text mining* mengadopsi metode data *mining*, namun perbedaan terletak pada sumber data. *Text mining* mengambil pola dari teks yang tidak terstruktur, sementara data *mining* berasal dari data yang terstruktur [11].

2.3. Analiis Sentimen

Analisis sentimen merupakan cabang dari Pemrosesan Bahasa Alami (NLP) yang mengembangkan sistem untuk mengidentifikasi dan mengekstrak opini dari teks. Analisis sentimen kerap digunakan untuk menggambarkan pandangan masyarakat tentang produk, merek, layanan, politik, maupun isu yang sedang berkembang. Analisis sentimen bertujuan untuk mengklasifikasikan teks, kalimat, atau pendapat ke dalam kelas opini positif, negatif, atau netral [8] yang kemudian dijadikan pengambilan keputusan.

2.4. Preprocessing

Dalam analisis sentimen, *preprocessing* langkah awal dalam pengolahan data yang bertujuan untuk menghilangkan elemen elemen yang tidak dibutuhkan dalam analisis data. Tahapan dalam preproessing melibatkan proses menyeragamkan huruf menjadi huruf kecil(*casefolding*), penghilangan karakter yang tidak relevan (*cleansing*), mengubah teks menjadi daftar kata yang terpisah atau token(*tokenization*), mengubah kata sesuai kaidah bahasa baku(*normalisasi*), mengidentifikasi kata negasi(*convert negasi*), penghapusan kata umum(*stopword*), mengubah menjadi kata dasar(*stemming*). Proses ini diperlukan agar kalimat lebih terstruktur dengan baik dan lebih mudah diproses oleh sistem [12].

2.5. Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency (*TF-IDF*) merupakan Teknik yang digunakan untuk mereprsentasikan kata dalam bentuk numerik berdasarkan frekuensi kemunculan dan tingkat relevansinya dalam korpus atau kalimat. Untuk mencari TF-IDF menggunakan persamaan (1), namun sebelum mencari TF-IDF terlebih dahulu di cari nilai TF(*Term Frekuensi*) persamaan (2) dan IDF(*Inverse dokumen Frekuensi*) persamaan (3) [13].

$$w_{t,d} = tf_{t,d} \times idf_t \quad (1)$$

$$tf_{t,d} = 1 + \log_{10}(f_{t,d}), jika\ f_{t,d} > 0 \quad (2)$$

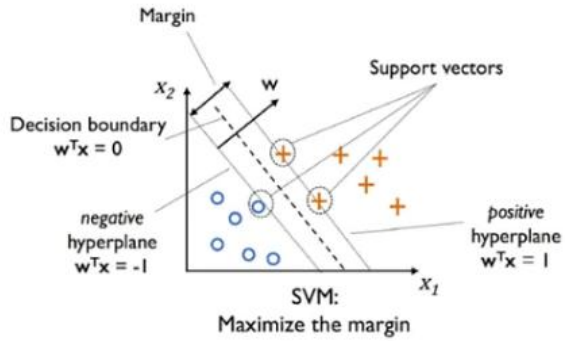
$$idf_t = \log_{10}\left(1 + {}^{N}/_{df_t}\right) \quad (3)$$

Keterangan :

$w_{t,d}$ = Bobot kata(*term)* terhadap dokumen *d*
tf_{td} = Nilai TF *term* dalam dokumen *d*
idf_{td} = Nilai IDF pada *term*
N = Banyak dokume yang ada
df_t = Banyaknya dokumen yang mengandung *term*
f_{td} = Banyak *term* yang muncul dalam dokumen *d*

2.6. Support Vector Machine

SVM atau *Support Vector Machine* didasarkan pada prinsip dasar linear *classifier* yaitu klasifikasi yang dapat dipisahkan secara linear [14]. merupakan algoritma dalam *machine learning* yang diterapkan untuk tugas klasifikasi diantara dua kelas data dan dipisahkan dengan garis solid di tengah disebut *hypreplane*. Hyperplane tersebut digunakan untuk memisahkan kelas -1 dan +1. *Svm* mampu memisahkan data secara linear dengan menggunakan dua variabel yaitu x sebagai data fitur teks dan y sebagai label kelas. *Support vector* merupakan data yang berada paling dekat dengan hyperplane dan akan dihitung oleh *svm* untuk menentukan hyperplane paling optimal. Hyperplane terbaik diperoleh dengan memilih margin terbesar antara dua kelas. Margin merupakan jarak antara hyperplane dengan data terdekat dari masing-masing kelas yang menentukan batas pemisah optimal [5]



Gambar 1. Margin Maksimal

Persamaan (4) merupakan model matematis yang digunakan untuk mencari hyperplane.

$$w \times x_i + b = 0 \quad (4)$$

Data x_i dibagi menjadi dua kelas -1 dan 1. Di dalam data x_i, yang termasuk dalam kelas -1 dijelaskan dengan rumus persamaan(5). Dan pada kelas 1 di jelaskan pada persamaan(6)

$$w \times x_i + b < 0\,, y_i = -1 \quad (5)$$

$$w \times x_i + b > 0\,, y_i = 1 \quad (6)$$

Keterangan :
x_i = Data input
y_i = Label yang di berikan
w = weigh
b = bias

Selain menangani data linear, SVM dapat menangani data non-linear dengan menerapkan fungsi *kernel*. Tujuan penggunaan *kernel* adalah untuk merepresentaikan data ke ruang dimensi yang lebih tinggi, dengan memisahkan data non-linear secara linear. Berikut *kernel* yang digunakan pada metode SVM [15]:

Kenel linear

$$k(x_i, x) = (x_i, x) \quad (7)$$

Kernel *Radial Basis Funcion* (RBF)

$$k(x_i, x) = \exp \frac{-||x_i x||^2}{2\sigma^2} \quad (8)$$

Kernel Polynomial

$$k(x_i, x) = (x_i, x)^d \quad (9)$$

Kernel Sigmoid

$$k(x_i, x) = tahn(\sigma(x_i x) + C) \quad (10)$$

2.7. Evaluasi Klasifikasi

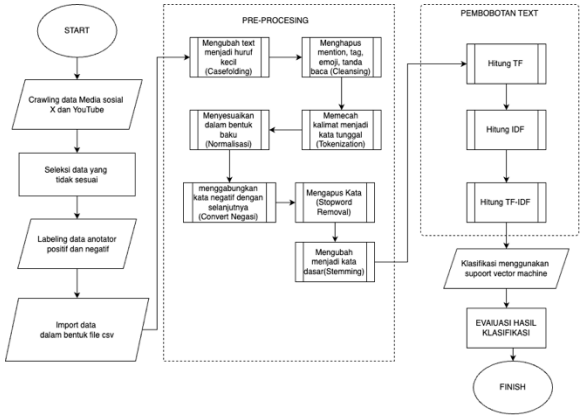
Evaluasi klasifikasi bertujuan untuk mengukur sejauh mana performa model dalam melakukan prediksi yang akurat. *Confusion matrix* digunakan untuk mengevaluasi model klasifikasi karena menyajikan hasil prediksi data uji, termasuk jumlah prediksi benar(*true*) dan salah(*false*). Tingkat keakuratan hasil dievaluasi menggunakan 3 nilai yaitu recall, precision, dan akurasi. *Recall*(*True Positive Rate*) membandingkan data positif yang teridentifikasi benar dengan seluruh data positif, *precision*(*Positive Predictive Value*) membandingkan prediksi positif yang benar dengan total prediksi positif, *accuracy* mengukur rasio prediksi benar terhadap seluruh data [16].

Berikut formula untuk mencari ke 3 nilai:

1. $Recall = \frac{TP}{TP+FP} \times 100\%$ (11)
2. $Precision = \frac{TP}{TP+FN} \times 100\%$ (12)
3. $Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \times 100\%$ (13)

3. METODE PENELITIAN

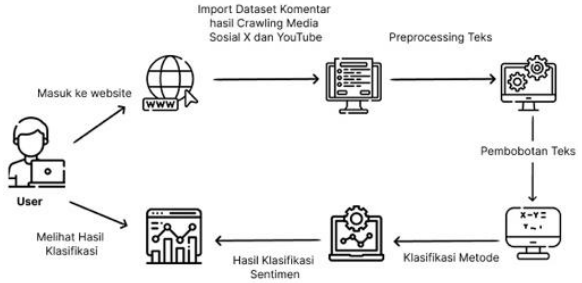
3.1. Flowchart Sistem



Gambar 2. Flowchart Sistem

Pada gambar 2 dimulai dari proses crawling data media social dari media social *X* dan *YouTube.* Data yang tidak relevan di seleksi, lalu annotator memberi label positif dan negative, kemudian data tersebut di simpan dalam format csv. Saat melakukan import data ke dalam program melalui file berformat csv. Selanjutnya data memasuki proses preprocessing, pada tahapan ini data akan di bersihkan agar lebih terstruktur. Setelah itu , dilakukan proses pembobotan dengan menggunakan algoritma pembobotan tf-idf. Hasil tf-idf akan diubah ke dalam format matrix untuk klasifikasi *svm.* Hasil dari metode svm kemudian dievaluasi untuk mengukur akurasi model.

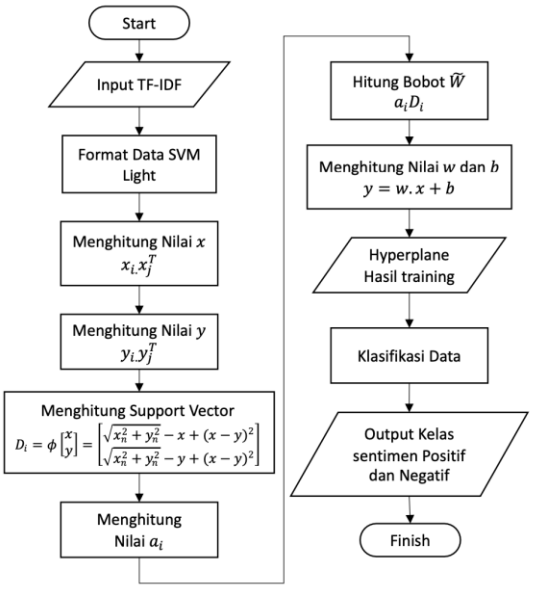
3.2. Block Diagram Sistem



Gambar 3. Diagram Blok Sistem

Stuktur Gambar 3 , user akan masuk ke dalam sistem melalui website, kemudian mengimpor data opini atau komentar media sosial *YouTube* dan *X* yang telah dilabeli sentimen positif dan negatif. Data ini selanjutnya diproses melalui tahapan preprocessing teks, seperti pembersihan dan normalisasi agar menjadi lebih terstruktur. Setelah itu dilanjutkan pada tahapan pembobotan teks untuk menentukan Tingkat kepentingan setiap kata di dalam dokumen. Nilai bobot yang diperoleh digunakan sebagai input proses klasifikasi metode. Hasil dari klasifikasi kemudian dievaluasi dan ditampilkan kepada pengguna sebagai output data sentimen.

3.3. Flowchart Metode



Gambar 4. flowchart metode svm

Gambar 4 menggambarkan tahapan klasifikasi data menggunakan algoritma *svm* dengan pendekatan *kernel linear*. Dataset yang digunakan merupakan data latih yang diperoleh setelah dilakukan pembagian dataset, antar data latih(training) dan data uji(testing). Data latih tersebut diubah menjadi format *svm*. Setelah itu dilakukan perhitungan terhadap nilai fitur x(teks) dan y(label). Sistem menghitung *support vector* dengan *kernel trik phi* , hasil *kernel* digunakan untuk menghitung nilai *support vector* tiap kalimat. Hitung nilai a_i, w(*bobot*) dan b(*bias*) atau *hyperplane*. *Hyperplane* ini terbentuk berdasarkan data *training* yang telah diklasifikasikan oleh algoritma SVM. Dari hasil *training*, terbentuklah sebuah model klasifikasi. Model tersebut kemudian diuji menggunakan data *testing*, dan hasilnya adalah prediksi kelas berupa sentimen positif maupun negatif, sesuai dengan data *input* yang diberikan.

3.4. Arsitektur Sistem

Tabel 1. Arsitektur Sistem

Input	Proses	Output
1. komentar 2. Label	1. Pengumpulan data *2.* Pre-*processing* data 3. Pembobotan TF-IDF 4. Modelling menggunakan *svm* 5. Evaluasi menggunakan *confusion matrix*	1. Klasifikasi positif atau negatif tiap komentar. 2. Distribusi positif dan negatif dalam bentuk grafis 3. Frekuensi kata pada sentimen positif dan negatif (*worldclooud*)

Tabel 1. Arsitektur sistem terdiri atas 3 proses yaitu input, proses, output. Input menjelaskan jenis data yang akan digunakan, dalam kasus analisis sentimen yaitu komentar yang diberikan label. Dalam Proses, terdapat proses pengumpulan data yaitu hasil crawling data dari media sosial, pre-processing mencangkup pembersihan data, kemudian pembobotan menggunakan *tf-idf, modelling* klasifikasi menggunakan *Support vector machine, dan evaluasi* menggunakan *confusion matrix* . Bagian output menampilkan hasil klasifikasi *svm*, distribusi positif dan negative, dan *wordcloud*. Berikut tahapan preprocessing:

3.5. *Casefolding*

Tabel 2. Proses Casefolding

Input	Output
@hrdbacot Keluhan HRD soal genZ yaitu mereka belum siap bekerja dalam sistem dan organisasi perusahaan. Jadinya baru sebulan seminggu bahkan beberapa hari kerja udah out. Katanya burnout underpaid gak work life balance. Kayaknya genZ butuh lebih digembleng.	@hrdbacot keluhan hrd soal genz yaitu mereka belum siap bekerja dalam sistem dan organisasi perusahaan. jadinya baru sebulan seminggu bahkan beberapa hari kerja udah out. katanya burnout underpaid gak work life balance. kayaknya genz butuh lebih digembleng.

Pada Tabel 2 merupakan tahapan *casefolding*, dimana mengubah semua kata menjadi huruf kecil. Perubahan pada kalimat terlihat pada kata “HRD”, “genZ” yang dikonversi dalam huruf kecil.

3.6. *Cleaning*

Tabel 3. Proses Cleaning

Input	Output
@hrdbacot keluhan hrd soal genz yaitu mereka belum siap bekerja dalam sistem dan organisasi perusahaan. jadinya baru sebulan seminggu bahkan beberapa hari kerja udah out. katanya burnout underpaid gak work life balance. kayaknya genz butuh lebih digembleng.	keluhan hrd soal genz yaitu mereka belum siap bekerja dalam sistem dan organisasi perusahaan jadinya baru sebulan seminggu bahkan beberapa hari kerja udah out katanya burnout underpaid gak work life balance kayaknya genz butuh lebih digembleng

Tabel 3 merupakan tahapan *cleaning*. Dalam *cleansing* komponen yang tidak di perlukan di hapus, seperti URL, *mention*, dll. Perubahan pada kalimat di atas yaitu menghilangkan *mention* @hrdbacot dan pungtuasi.

3.7. Tokenisasi

Tabel 4. Proses Tokenisasi

Input	Output
keluhan hrd soal genz yaitu mereka belum siap bekerja dalam sistem dan organisasi perusahaan jadinya baru sebulan seminggu bahkan beberapa hari kerja udah out katanya burnout underpaid gak work life balance kayaknya genz butuh lebih digembleng	'keluhan','hrd','soal', 'genz','yaitu','mereka', 'belum','siap','bekerja', 'dalam','sistem','dan', 'organisasi','perusahaan', 'jadinya','baru','sebulan', 'seminggu','bahkan', 'beberapa', 'hari', 'kerja', 'udah', 'out', 'katanya', 'burnout', 'underpaid', 'gak', 'work', 'life', 'balance', 'kayaknya', 'genz', 'butuh', 'lebih', 'digembleng'

Tabel 4 merupakan tahapan *tokenization*. Dimana teks akan diubah menjadi bagian-bagian kecil kata. Hasil dari *tokenization* yaitu kata atau sub-kata yang dipisah dengan koma dan spasi untuk membantu mesin memahami dan memproses teks untuk kemudian dianalisis lebih lanjut.

3.8. Normalisasi

Tabel 5. Proses Normalisasi

Input	Output
keluhan, hrd, soal, genz, yaitu, mereka, belum, siap, bekerja, dalam, sistem, dan, organisasi, perusahaan, jadinya, baru, sebulan, seminggu, bahkan, beberapa, hari, kerja, udah, out, katanya, burnout, underpaid, gak, work, life, balance, kayaknya, genz, butuh, lebih, digembleng	keluhan, hrd, soal, genz, yaitu, mereka, belum, siap, bekerja, dalam, sistem, dan, organisasi, perusahaan, sehingga, baru, sebulan, seminggu, bahkan, beberapa, hari, kerja, sudah, keluar, menurutnya, burnout, underpaid, tidak, kerja, hidup, seimbang, sepertinya, genz, perlu, lebih, digembleng

Tabel 5 merupakan tahapan normalisasi. Tahapan ini kata-kata slang maupun istilah gaul yang tidak baku dalam teks diubah menjadi bentuk baku. Terlihat perubahan beberapa kata “jadinya” menjadi “sehingga”, “out” menjadi “keluar” dan “gak” menjadi “tidak”. Tahapan ini dilakukan untuk menyamakan representasi kata agar konsisten dengan kosakata yang digunakan model.

1. Convert Negasi

Tabel 6. Proses Convert Negasi

Input	Output
keluhan, hrd, soal, genz, yaitu, mereka, belum, siap, bekerja, dalam, sistem, dan, organisasi, perusahaan, jadinya, baru, sebulan, seminggu, bahkan, beberapa, hari, kerja, udah, out, katanya, burnout, underpaid, gak, work, life, balance, kayaknya, genz, butuh, lebih, digembleng	keluhan, hrd, soal, genz, yaitu, mereka, belum_siap, bekerja, dalam, sistem, dan, organisasi, perusahaan, sehingga, baru, sebulan, seminggu, bahkan, beberapa, hari, kerja, sudah, keluar, menurutnya, burnout, underpaid, tidak_kerja, hidup, seimbang, sepertinya, genz, perlu, lebih, digembleng

Tabel 6 tahapan convert negasi, kata negasi digabungkan dengan kata selanjutnya dengan menggunakan simbol *underscore*(“_”). Ini berfungsi untuk mempertahankan kata-kata negatif yang ada pada kalimat. Pada hasil diatas kata “belum siap” dinegasiakan menjadi“belum_siap” dan kata “ tidak kerja” menjadi “tidak_kerja”

3.9. Stopword

Tabel 7. Proses *Stopword*

Input	Output
keluhan, hrd, soal, genz, yaitu, mereka, belum_siap, bekerja, dalam, sistem, dan, organisasi, perusahaan, sehingga, baru, sebulan, seminggu, bahkan, beberapa, hari, kerja, sudah, keluar, menurutnya, burnout, underpaid, tidak_kerja, hidup, seimbang, sepertinya, genz, perlu, lebih, digembleng	keluhan soal genz belum_siap bekerja sistem organisasi perusahaan sebulan seminggu hari kerja keluar burnout underpaid tidak_kerja hidup seimbang genz perlu digembleng

Tabel 7 merupakan tahapan *stopword removal*. Langkah ini penting karena kata yang umum yang lazim secara berulang dalam dokumen teks, namun tidak memiliki kontribusi signifikan akan dihilangkan. Pada hasil di atas kata “hrd”,”yaitu”, ”mereka” dan lainnya dihapus karena tidak memberikan informasi.

3.10. Stemming

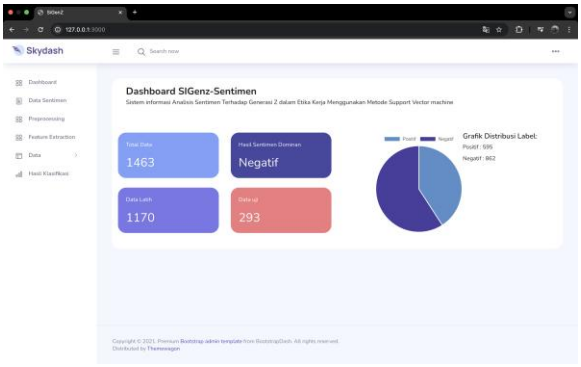
Tabel 8. Proses Stemming

Input Proses	Output Proses
keluhan soal genz belum_siap bekerja sistem organisasi perusahaan sebulan seminggu hari kerja keluar burnout underpaid tidak_kerja hidup seimbang genz perlu digembleng	keluh soal genz belum siap kerja sistem organisasi perusahaan bulan minggu hari kerja keluar burnout underpaid tidak kerja hidup imbang genz perlu gembleng

Tabel 8 merupakan tahapan *stemming*, dimana setiap kata akan dikembalikan ke bentuk dasarnya dengan cara menghapus imbuhan, baik berupa awalan, akhiran, maupun kombinasi keduanya. Perubahan kata pada kalimat di atas "keluhan" menjadi "keluh", "bekerja" menjadi "kerja", "digembleng" menjadi "gembleng". Simbol underscore("_") pada kata sambung negasi akan ikut dihilangkan.

4. HASIL DAN PEMBAHASAN

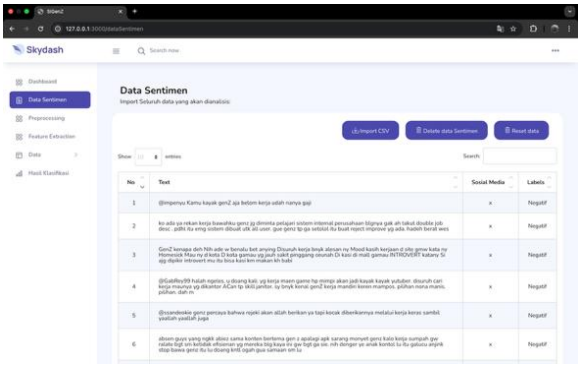
4.1. Halaman Dashboard



Gambar 5. Tampilan Dashboard

Ditampilkan Gambar 5. yaitu hasil implementasi dari dashboard. Halaman dashboard menyajikan informasi terkait total data, jumlah data latih dan uji, serta distribusi label sentimen positif dan negatif dalam bentuk grafik. Selain itu ditampilkan sentimen dominan sebagai representasi opini berdasarkan hasil data uji.

4.2. Halaman Data Sentimen

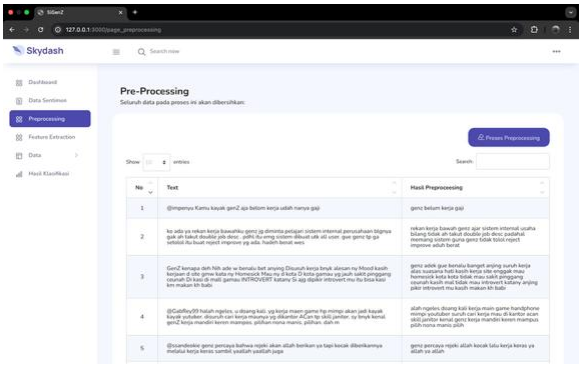


Gambar 6. Tampilan Data Sentimen

Visualisasi yang ditampilkan pada Gambar 6. merupakan hasil implementasi data sentimen. Menu ini digunakan untuk mengimpor dan mengelola dataset. Tersedi tombol import untuk mengunggah file csv, serta tombol delete dan reset untuk menghapus data yang tersimpan di basis data.

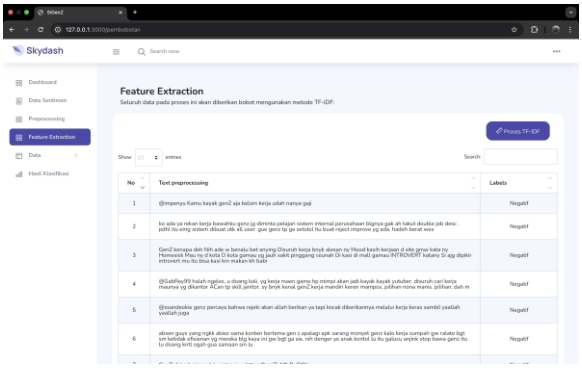
4.3. Halaman Pre-procesing

Gambar 7. Merupakan hasil implementasi preprocessing data, di mana data mentah hasil import sentimen akan diproses melalui tahapan seperti cleaning ,normalisasi, dan tokenisasi untuk menghasilkan data yang bersih dan siap digunakan dalam analisis sentimen.



Gambar 7. Tampilan Preprocessing

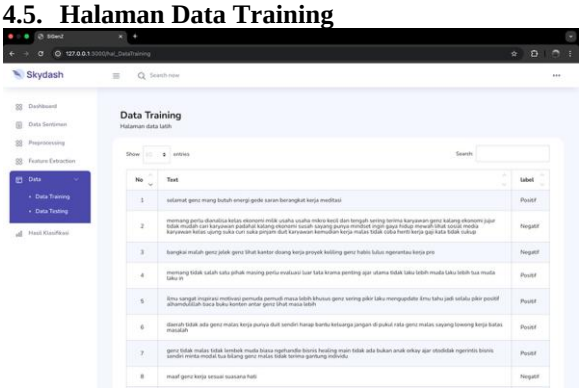
4.4. Halaman Feature Extraction



Gambar 8. Tampilan Feature Extraction

Antarmuka yang ditampilkan pada Gambar 8. merupakan implementasi feature extraction. Menu ini akan mengubah teks menjadi data numeric dengan mengunakan metode TF-IDF. Data kemudian dibagi 80% untuk training dan 20% untuk testing.

4.5. Halaman Data Training



Gambar 9. Tampilan Data Training

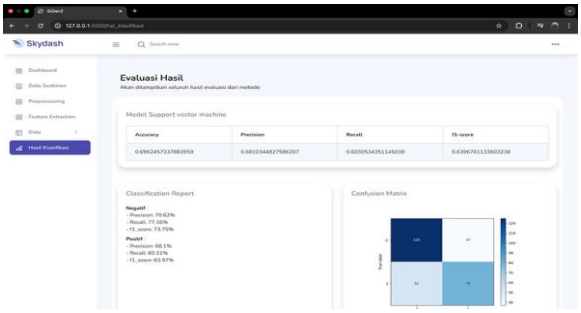
Gambar 9. merupakan implementasi data training. Halaman ini menampilkan 80% data hasil split dari *feature extraction* yang disiapkan untuk pelatihan model.

4.6. Halaman Data Training

Gambar 10. Tampilan Data Testing

Pada Gambar 10. merupakan implementasi data testing. Halaman ini menampilkan 20% data hasil split dari *feature extraction* yang disiapkan untuk pengujian model.

4.7. Halaman Hasil Klasifikasi



Gambar 11. Tampilan Hasil Analisis

Gambar 11. menampilkan halaman hasil analisis, yang menyajikan evaluasi kinerja model SVM menggunakan *confusion matrix*, serta matrix seperti accuaracy,*precision*,*recall* dalam *classification report.* Selain itu halaman ini menampilkan hasil prediksi terhadap data uji untuk memberikan gambaran performa model secara menyeluruh.

4.8. Evaluasi Pengujian Model

Pada tahapan evaluasi model dilakukan pengujian untuk melihat performa model. Evaluasi kinerja sistem dilakukan melalui perhitung nilai berdasarkan tabel nilai *confusion matrix*. Data yang digunakan untuk penhujian berasal dari data testing yang diperoleh melalui pembagian *dataset* dengan rasio 80:20. Selanjutnya , label hasil annotator dibandingkan dengan label yang dihasilkan oleh sistem untuk mengevaluasi kinerja model.

Tabel 9. Hasil Nilai Confusion Matrix

Predicted Classs	**Actual Class**		**Total**
	Positif	**Negatif**	
Positif	79	37	116
Negatif	52	125	177
Total	131	162	293

Setelah didapatkan nilai *confusion matrix*, selanjutnya dijadikan dasar untuk menghitung nilai *recall* dengan persamaan (11) ,*precision* dengan persaman(12) dan *accuracy* dengan persamaan (13). Berikut formula untuk menghitung masing-masing evaluasi.

$$Precision = \frac{TP}{TP + FP} = \frac{79}{79 + 37} \times 100\% = 68\%$$

$$Recall = \frac{TP}{TP + FN} = \frac{79}{79 + 52} \times 100\% = 60\%$$

$$Accuracy = \frac{TP + TN}{Total\ Test} = \frac{204}{293} \times 100\% = 69\%$$

Berdasarkan tabel 9. sebanyak 79 opini yang diprediksi sebagai positif, dan memang faktanya positif. Sementara itu 37 opini diprediksi positif, tetapi faktanya negatif. Sebanyak 125 opini yang di prediksi negatif, dan faktannya negatif, sedangkan 52 kalimat diprediksi negatif, namun faktanya positif. Hasil evaluasi menunjukkan rata rata akurasi dengan jumlah data uji 293 opini, menghasilkan nilai precision sebesar 68%, recall sebesar 60%, dan Ttingkat akurasi 69%.

4.9. Evaluasi K-FOLD Cross Validation

Metode *K-Fold cross validation* merupakan metode evaluasi dalam pembelajaran mesin untuk mengevaluasi performa model, namun tetap mempertahankan sebagian data evaluasi. Tujuan untuk mengurangi terjadinya *overfitting* dan melatih model pada data yang belum pernah dilatih sebelumnya pada tiap iterasi. Nilai k(*folds*) yang digunakan dalam k-fold cross validation adalah 5 dan 10.

Tabel 10. Cros validation k=5

C_val	**Accuracy**
K1	0.6709
K2	0.6752
K3	0.7607
K4	0.7094
K5	0.6752
Rata-rata *C_val*	0.6983

Berdasarkan hasil tabel 10. Diperoleh nilai akurasi rata-rata pada proses evaluasi menggunakna k-crosss validation dengan k=5 sebesar 0.6983. berikutnya hasil evaluasi dengan k=10 :

Tabel 11. Cross validation k=10

C_val	**Akurasi**
K1	0.6581
K2	0.7179
K3	0.6239
K4	0.6838
K5	0.7350
K6	0.7607
K7	0.7350
K8	0.7009
K9	0.6581
K10	0.6752
Rata-rata *C_val*	0.6949

Pada tabel 11. diperoleh nilai akurasi rata-rata pada proses evaluasi menggunakna k-fold cross validation dengan k=10 sebesar 0,6949

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan , sistem mampu melakukan klasifikasi pada opini positif dan negatif terkait etika kerja generasi z. Hasil penelitian menunjukkan bahwa sentimen negatif mendominasi, menandakan kecenderungan opini masyarakat terhadap etika kerja generasi z mengarah pada pandangan negatif. Keberhasilan sistem dalam mengklasifikasikan data sebanyak 293 data, menghasilkan nilai recall 60%, precision 68%, dan accuracy 69%, sedangkan untuk pengujian *k-fold cross validation* dengan pembagian data ke dalam 5 dan 10 *fold* menghasilkan nilai rata-rata sebesar 0,69. Untuk penelitian berikutnya disarankan untuk meningkatkan proses preprocessing, dan menambahkan jumlah dataset, serta menggunakan pendekatan algoritma lain, seperti deep learning guna mengevaluasi lebih mendalam terkait isu etika kerja generasi z.

DAFTAR PUSTAKA

[1] M. M. P. Putra, E. Robyardi, and Heryati, "Pengaruh Etika Kerja Dan Lingkungan Kerja Terhadap Kualitas Kerja Pegawai Pada Direktorat Kementrian Kelautan Distrik Navigasi Kelas 1 Palembang," *Jurnal Ilmiah Wahana Pendidikan*, vol. 10, pp. 641–648, Jan. 2024.

[2] Wijiati and U. Chasanah, "PENGARUH GAYA HIDUP GENERASI Z (FLEKSIBILITAS KERJA DAN KOMPENSASI) TERHADAP PEMILIHAN PEKERJAAN," Yogyakarta, Jun. 2024.

[3] M. G. Syahdi, R. Sugiarti, and F. Suhariadi, "Motivasi Kerja Generasi Z : Kajian Literatur," *Action Research Literate*, vol. 8, no. 2, pp. 398–412, Mar. 2024.

[4] D. S. Prasetyo, K. Auliasari, and Y. A. Pranoto, "Analisis Sentimen Pada Komentar Media Sosial Terkait Isu Joki Dengan Menggunakan Metode LSTM Sentiment Analysis on Social Media Comments Regarding Jockey Issues Using the LSTM Method," Malang, 2025. [Online]. Available: https://jurnal.unimed.ac.id/2012/index.php/cess

[5] I. S. Aisah, B. Irawan, and T. Suprapti, "ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS SENTIMEN ULASAN APLIKASI AL QUR'AN DIGITAL," Dec. 2023.

[6] V. A. Herlinda, C. S. K. Aditya, and D. R. Chandranegara, "Analisis Sentimen Masyarakat Terhadap Generasi Z Dalam Dunia Kerja Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes," *REPOSITOR*, vol. 6, no. 4, pp. 405–414, Nov. 2024.

[7] R. D. Yahya, S. A. Wibowo, and N. Vendyansyah, "ANALISIS SENTIMEN UNTUK DETEKSI UJARAN KEBENCIAN PADA MEDIA SOSIAL TERKAIT PEMILU 2024 MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," 2024.

[8] Damayanti, D. I. Efendi, D. Solihudin, C. L. Rohmat, and S. E. Permana, "PEMETAAN OPINI PUBLIK TERHADAP PERUBAHAN KEBIJAKAN BPJS KESEHATAN DENGAN PENDEKATAN SUPPORT VECTOR MACHINE(SVM) DALAM ANALISIS SENTIMEN," Feb. 2024.

[9] Melia, B. Irawan, and O. Nurdiawan, "ANALISIS SENTIMEN TERHADAP PENGGUNA GOJEK DAN GRAB PADA MEDIA SOSIAL TWITTER MENGGUNAKAN RANDOM FOREST," Cirebon, Oct. 2023.

[10] N. C. Ramadani, I. Tahyudin, and A. S. Barkah, "Perbandingan Algoritma Support Vector Machine, Decision Tree, dan Logistic Regresion Pada Analisis Sentimen Ulasan Aplikasi Netflix," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 2, pp. 110–117, Aug. 2024, doi: 10.25077/TEKNOSI.v10i2.2024.110-117.

[11] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *Media Online*, vol. 2, no. 5, pp. 178–185, Apr. 2022, [Online]. Available: https://djournals.com/klik

[12] D. Atmajaya, A. Febrianti, and H. Darwis, "Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 4, pp. 2173–2181, 2023.

[13] R. A. Pranata, Rudiman, and N. A. Verdikha, "METODE PEMBOBOTAN TF-IDF UNTUK KLASIFIKASI TEKS QUICK COUNT PEMILIHAN WAKIL PRESIDEN INDONESIA 2024 PADA X TWITTER DENGAN METODE SVM," vol. 18, no. 2, pp. 126–138, Aug. 2024, doi: 10.47111/JTI.

[14] Muh. F. Rizki, K. Auliasari, and R. P. Prasetya, "ANALISIS SENTIMENT CYBERBULLYING PADA SOSIAL MEDIA TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," Sep. 2021.

[15] E. R. M. Sholihah, I. G. S. M. Diyasa, and E. Y. Puspaningrum, "PERBANDINGAN KINERJA KERNEL LINEAR DAN RBF SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN ULASAN PENGGUNA KAI ACCESS PADA GOOGLE PLAY STORE," *Jurnal Mahasiswa Teknik Informatika*, vol. 8, no. 1, pp. 728–733, Feb. 2024.

[16] G. Rininda, I. H. Santi, and S. Kirom, "PENERAPAN SVM DALAM ANALISIS SENTIMEN PADA EDLINK MENGGUNAKAN PENGUJIAN CONFUSION MATRIX," Oct. 2023