

ANALISIS SENTIMEN TWITTER TERHADAP UU CIPTA KERJA DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

TWITTER SENTIMENT ANALYSIS OF THE CIPTA KERJA LAW USING THE NAÏVE BAYES ALGORITHM AND SUPPORT VECTOR MACHINE

Frencis Matheos Sarimole¹, Andrian Nur Ihsan²

^{1,2}Program Studi Teknik Informatika Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika
andriannurihsan.sdjcell@gmail.com

ABSTRACT

The purpose of the Cipta Kerja Law is to encourage investment and create jobs through simplification of the bureaucratic and licensing systems, making it easier for business people, especially UMKM. Many people reject the Cipta Kerja Law because they are seen as not taking sides with the community, especially workers and employees. Several articles of the Cipta Kerja Law are considered detrimental to workers. One of them concerns dismissal and reduced severance pay. Many people disagree with the ratification of the law. The purpose of this research is to understand and find out public opinion on the Cipta Kerja Law, whether this opinion belongs to the positive sentiment class or the negative sentiment class and to find out the accuracy of the percentage results for each class. Tests were carried out using the Naïve Bayes algorithm and Support Vector Machine (SVM) Based on the test results of the Naïve Bayes algorithm model, it produces an accuracy value of 75.43% and a Support Vector Machine of 81.31%. From the sentiment analysis testing process using 1442 data, it produces predictions of 970 Positive Sentiments and 472 Negative Sentiments.

Keywords: Analysis, Sentiment, Twitter, Naïve Bayes, Support Vector Machine

ABSTRAK

Undang-Undang Cipta Kerja tujuannya adalah untuk mendorong investasi dan penciptaan lapangan kerja melalui penyederhanaan sistem birokrasi dan perizinan, kemudahan bagi pelaku usaha, khususnya UKM. Banyak masyarakat yang menolak UU Cipta Kerja karena dinilai tidak memihak kepada masyarakat, khususnya kepada buruh dan karyawan. Beberapa pasal UU Cipta Kerja dianggap merugikan pekerja. Salah satunya menyangkut pemecatan dan nilai pesangon yang dikurangi. Banyak masyarakat yang tidak setuju dengan pengesahan UU tersebut. Tujuan dari penelitian ini adalah untuk memahami dan mengetahui opini masyarakat atas UU Cipta Kerja, apakah opini tersebut masuk dalam kelas sentimen positif atau kelas sentimen negatif serta mengetahui hasil akurasi persentasi dari setiap kelas. Pengujian dilakukan menggunakan algoritma Naïve Bayes dan Support Vector Machine (SVM) Berdasarkan hasil pengujian model algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 75,43% dan Support Vector Machine sebesar 81,31%, Dari proses pengujian sentimen analisis menggunakan 1442 data, menghasilkan prediksi sebanyak 970 Sentimen Positif dan 472 Sentimen Negatif.

Kata Kunci: Analisis, Sentimen, Twitter, Naïve Bayes, Support Vector Machine.

PENDAHULUAN

Menurut laporan terbaru *We Are Social* ditahun 2023, dikatakan bahwa jumlah pengguna internet di Indonesia ada sekitar 212,9 juta orang. Dari jumlah itu, ada 167 orang yang menggunakan internet untuk mengakses media sosial. Salah satu media sosial yang digunakan adalah twitter, untuk pengguna twitter saat ini mencapai 24 juta orang. Tidak heran banyak *tweet* dari para penggunanya yang membahas tentang berbagai hal seperti: hiburan,

pendidikan, pekerjaan, politik, termasuk tentang agama. Salah satu isu yang *trending topic* dan menjadi perbincangan banyak orang di twitter adalah tentang Undang-Undang Cipta Kerja.

Disebutkan Undang-Undang Cipta Kerja tujuannya adalah untuk mendorong investasi dan penciptaan lapangan kerja melalui penyederhanaan sistem birokrasi dan perizinan, kemudahan bagi pelaku usaha, khususnya UKM, ekosistem investasi yang kondusif, dan penciptaan

lapangan kerja untuk memenuhi kebutuhan tenaga kerja yang terus bertambah yang menghambat pertumbuhan.

Banyak masyarakat yang menolak UU Cipta Kerja karena dinilai tidak memihak kepada masyarakat, khususnya kepada buruh dan karyawan. (Kartikasari and Fauzi 2021) Beberapa pasal UU Cipta Kerja dianggap merugikan pekerja. Salah satunya menyangkut pemecatan dan nilai pesangon yang dikurangi. Banyak masyarakat yang tidak setuju dengan pengesahan UU tersebut. Bahkan dari 9 fraksi yang ada di DPR hanya ada 2 fraksi yang tidak setuju dengan disahkannya UU Cipta Kerja yaitu Fraksi Demokrat dan fraksi PKS. Banyak dari elemen buruh, karyawan dan mahasiswa yang melakukan aksi demo untuk menolak UU Cipta Kerja.

Hal ini yang menjadi trending topic di twitter, diantara banyaknya penolakan terhadap UU Cipta Kerja, tidak sedikit juga yang setuju terhadap disahkannya UU Cipta Kerja. Dari fenomena tersebut dapat digunakan sebagai bahan analisis sentimen untuk mengetahui presentase positif dan presentase negatif pendapat masyarakat Indonesia tentang masalah ini.

Analisis sentimen atau opinion mining merupakan metode analisis berbasis komputasi mengenai pendapat, sentimen, dan emosi. (Liu and Tsai 2010) Analisis sentimen digunakan untuk melihat kecenderungan suatu sentimen atau pendapat, apakah pendapat tersebut cenderung beropini positif atau negatif. (Nurzahputra and Muslim 2016) Sebelum dilakukan analisis sentimen, perlu dilakukan preprocess dengan menggunakan metode *text mining* untuk mengolah data teks. Praproses teks tersebut meliputi case folding, *tokenizing*, stop- words, dan *stemming*. Penulis pada penelitian ini akan melakukan analisis sentimen kepada para pengguna twitter terhadap UU cipta kerja dengan menggunakan metode algoritma Naive Bayes dan Support Vector Machine

METODE

Metode yang digunakan pada penelitian dalam analisis sentimen ini

menggunakan pendekatan *Natural Language Processing (NLP)* dengan penerapan algoritma naïve bayes dan algoritma support vector machine.

Data Penelitian

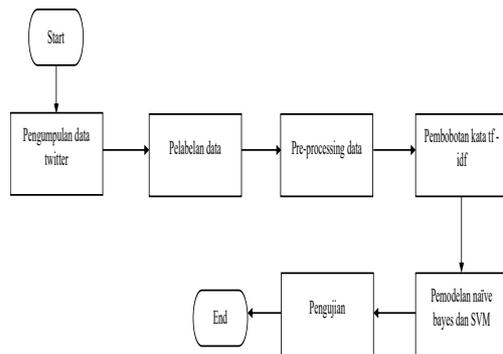
Data yang digunakan dalam penelitian ini adalah kumpulan tweet mengenai UU cipta kerja pada tanggal 21 oktober 2021 hingga 27 maret 2023. Data yang didapat sebanyak 1442 tweet. Data yang telah dikumpulkan ialah sebagai input sistem agar dapat diproses lebih lanjut. Adapun contoh data hasil *crawling* yakni sebagai berikut:

Tabel 1. Contoh Tweet

Tweet	Label	Keterangan
Dengan adanya Perpu Cipta kerja bikin Izin usaha menjadi super cepat. Perizinan terintegrasi dengan sistem online (OSS)Perppu Ciptaker Sah	1	Positif
@BossTemlen ini orang ngomong base on data apa berasal dari angan angan dia sih... dia lupa abis teken UU Cipta kerja yg sangat merugikan lingkuan hidup..	-1	Negatif

Penerapan Metodologi

Dalam penelitian ini penulis menggunakan beberapa metode untuk memproses data, berikut adalah metode yang digunakan:



Gambar 1. Tahapan Metodologi

Pengumpulan Data

Pengumpulan data twitter ini menggunakan teknik *crawling* data yaitu proses pengambilan data dari twitter yang diproses dengan Jupyter Notebook, berfungsi untuk mengumpulkan *dataset* yang akan diolah serta merupakan tahap awal dalam penelitian ini. Data yang dikumpulkan merupakan data tweet dengan keyword “Cipta Kerja”. Dari hasil *crawling* tersebut berhasil mendapatkan data berjumlah 1442 data *tweet* yang selanjutnya akan dilakukan pengolahan agar menjadi data yang berkualitas dalam proses analisis sentimen.

Pelabelan Data

Data yang telah di ambil dari twitter akan dilakukan pelabelan secara otomatis dengan *library textblob* berdasarkan *polarity*. Pada proses ini data tweet hanya membagi kedalam kelas positif (polaritas=1) dan kelas negatif (polaritas=-1). Penilaian parameter pada label positif mengandung kalimat – kalimat yang baik, seperti pujian, semangat, dukungan dan lain sebagainya. Dan kalimat negatif adalah kalimat yang mengandung kalimat kurang baik seperti hinaan, putus asa, cacian, ejekan dan lain sebagainya. (Sisfotenika 2023)

Data Preparation (Preprocessing)

Data yang diambil merupakan data yang tidak terstruktur sehingga perlu dilakukan proses *preprocessing*. Dalam proses ini, kumpulan data *tweet* melalui proses *preprocessing* untuk melakukan transformasi data. (Muttaqin and Bachtiar 2016) Pada tahap ini dilakukan penyiapan

data teks yang memiliki beberapa permasalahan seperti *noise*, tidak lengkap. Proses tersebut meliputi beberapa tahap yaitu:

a. Casefolding

Tahap *Casefolding* merupakan proses dalam text *preprocessing* yang dilakukan untuk menyeragamkan karakter pada data. Proses case folding adalah proses merubah semua huruf menjadi huruf kecil.

b. Tokenizing

Tokenizing adalah tugas memisahkan deretan kata di dalam kalimat, paragraph atau halaman menjadi token atau potongan kata. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca, *mention*, *hashtag* dan *link*.

c. Stopwords

Filtering (stopwords removal). *Stop word* didefinisikan sebagai term yang tidak berhubungan (irrelevant) dengan subyek utama dari database meskipun katater sebut sering kali hadir di dalam dokumen. Contoh *stop words* adalah ada, adalah, adanya, adapun, agak, dll.

d. Stemming

Stemming merupakan proses tahap pencarian kata dasar dari hasil proses *filtering* dengan membuang kata imbuhan. Sebagai contoh, kata terdaftar, pendaftaran berbagi-pakai term stem umum daftar, dan dapat diperlakukan sebagai bentuk lain dari kata ini.

Pembobotan Kata

Metode TF-IDF merupakan metode penghitungan bobot setiap kata yang paling umum digunakan dalam temu kembali informasi. Cara ini juga dikenal efektif, mudah, dan memberikan hasil yang akurat. (Azis Maarif 2015) Metode ini akan menghitung nilai *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* untuk setiap token (*word*) pada setiap dokumen dalam korpus. Cara ini akan menghitung bobot setiap token pada dokumen dengan rumus:

$$tf_{ij} = \frac{f_a(i)}{\max_{j \in d} f_a(j)}$$

Term Frequency istilah menentukan frekuensi (tingkat frekuensi) istilah dalam dokumen. Sedangkan frekuensi dokumen adalah banyaknya dokumen yang terdapat istilah. Seperti yang anda lihat pada rumus di atas, yang digunakan dalam perhitungan adalah nilai *term frequency* dari *document frequency*. N disebut dengan jumlah seluruh dokumen dan df menyatakan nilai *document frequency* dari *term* yang akan dicari *inversnya*. Nilai *invers document frequency* (idf) diperoleh dari perhitungan berikut:

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

N adalah jumlah total *document* dalam *corpus*, $N = |D|$. $|\{d \in D : t \in d\}| = df(t)$, adalah jumlah dokumen yang mengandung *term* t. IDF juga dapat dituliskan dalam bentuk:

$$idf(t, D) = \log \left(\frac{N}{df(t)+1} \right)$$

Penambahan 1 untuk menghindari pembagian terhadap 0 jika df(t) tidak ditemukan pada corpus. Setelah itu *apply* L1 *normalization* dengan fungsi *normalize* () pada *library scikitlearn reprocessing* sehingga formulanya tepat sama dengan definisi TF diawal tulisan. L1 *Normalization* ditulis dalam bentuk:

$$\|x\|_1 = \sum_i |x_i|^1$$

Selanjutnya lakukan perkalian antara TF dan IDF dengan fungsi *multiply* (), sehingga didapatkan TFIDF vector. Penggunaan *smooth_idf = False* seperti dibahas sebelumnya agar fungsi IDF berbentuk, $idf(t) = \log [n / df(t)] + 1$.

Pemodelan Naïve Bayes

Setelah melalui *praprocessing* data dan *vectorizer* selanjutnya dibuat model yang akan digunakan untuk melakukan klasifikasi pada data uji. Proses ini dilakukan dengan menggunakan bantuan *library* pada bahasa pemograman Python3 yang bernama *scikit-learn* untuk proses klasifikasi. Pada proses klasifikasi digunakan data uji sebanyak 20% dari data keseluruhan.

Pemodelan Support Vector Machine

Seperti halnya metode Naïve Bayes, tahap ini membagi data set menjadi dua bagian yaitu data latih dan data uji dengan rasio *test* dan *train* yaitu 20% dari keseluruhan data.

Pengujian Model

Metode yang digunakan untuk evaluasi adalah *confusion matrix*. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi.(Ainurrohmah 2021)

Tabel 2. Confusion Matrix

Nilai prediksi	Nilai aktual	
	Positive	Negative
Postive	TruePositive	FalseNegativ e
Negativ e	FalsePositiv e	TrueNegative

Tabel di atas merupakan tabel *confusion matrix* dengan keterangan sebagai berikut:

- a. TP (*True Positive*) = jumlah data nilai aktual kelas positif dan nilai prediksi kelas positif
- b. TN (*True Negative*) = jumlah data nilai aktual negatif dan nilai prediksi negatif
- c. FP (*False Positive*) = jumlah data nilai aktual positif dan nilai prediksi negatif
- d. FN (*False Negative*) = jumlah data nilai aktual negatif dan nilai prediksi positif

Akurasi adalah nilai rasio data *tweet* yang sudah terdeteksi dalam pengujian. Nilai akurasi dapat menunjukkan kedekatan antara nilai prediksi sistem dan prediksi manusia(Azhari, Situmorang, and Rosnelly 2021). Berikut rumusnya:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Presisi adalah nilai ketepatan sistem mengenai informasi sistem untuk menunjukan data positif dan data negatif yang benar. Nilai presisi ini dihasilkan dari nilai prediksi positif berbanding dengan jumlah nilai yang positif(Azhari, Situmorang, and Rosnelly 2021), dengan rumus sebagai berikut:

$$Presisi = \frac{TP}{TP + FP}$$

TP + FP

Recall adalah nilai yang menunjukkan tingkat keberhasilan untuk mengetahui kembali informasi mengenai data positif dan negatif yang benar. *Recall* dihasilkan dari jumlah nilai true positif dibanding nilai aktual positif (Azhari, Situmorang, and Rosnelly 2021), sebagai berikut:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Spesifikasi adalah kebenaran memprediksi negatif dibandingkan dengan jumlah data negatif (Irwansyah Saputra 2022), dengan rumus sebagai berikut:

$$\text{Spesifikasi} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

F1 Score adalah perbandingan rata-rata presisi dan *recall* yang dibobotkan (Irwansyah Saputra 2022), dengan rumus sebagai berikut:

$$\text{F1 Score} = \frac{2 \times \text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}}$$

HASIL DAN PEMBAHASAN

Pengumpulan Data

Hasil Pengumpulan data diambil dari twitter dengan yang kata kunci “Cipta Kerja”. *Crawling* data yang diambil dalam bentuk teks yang merupakan tweet dari pengguna twitter. Data yang diambil pada tanggal 21 oktober 2021 hingga 27 maret 2023 data yang didapat sebanyak 1442 tweet.

Tabel 3. Contoh Hasil Crawling

Data Hasil Crawling
Gw sempet nanya, sih, apakah mbak CS familiar atau pernah akses @AHU_Online dan @OSS_id. Jawabnya tentu saja belum. Terus gw jelasin, deh, kalo PT Perorangan ini implementasi UU Cipta Kerja, ngurusnya serba online. Ada nomor pendirian dari Kemenkumham.
Besok diminta ngejelasin perppu cipta kerja yang baru, padahal sampe skrg w sendiri masih bingung knp ini perppu tahun 2022 kok bisa aturan pelaksanaanya mengacu ke pp tahun 2021. Retroaktif?
Lmao

Perpu cipta kerja ada apa dengan nya kenapa malah jadi bikin ribet sih. Yang kerja 5 hari dalam seminggu selama ini sebenarnya siapa, orang kantornya kalo yang dilapangan mana ada justru hari kerjanya full termasuk tanggal merah tetap kerja biarpun di BUMN.

Preprocessing Data

Tahapan *praprocessing* data perlu dilakukan karena beberapa kalimat *tweet* yang didapatkan tidak sepenuhnya menggunakan kata baku dan menggunakan bahasa indonesia yang baik. Preprocessing dilakukan menggunakan bantuan library pada bahasa pemrograman Python. *Praprocessing* data dilakukan dengan tahap *Case Folding*, *Tokenizing*, *Stopwords*, *Stemming* sehingga menghasilkan data bersih dan siap untuk lanjut pada proses berikutnya.

a. Casefolding

Tahap *Casefolding* merupakan proses yang dilakukan untuk menyeragamkan karakter pada data. Proses *casefolding* adalah proses merubah semua huruf menjadi huruf kecil. (Aditia Rakhmat Sentiaji et al. 2014)

Tabel 4. Hasil Casefolding

Sebelum Casefolding	Hasil Casefolding
Gw sempet nanya, sih, apakah mbak CS familiar atau pernah akses @AHU_Online dan @OSS_id. Jawabnya tentu saja belum. Terus gw jelasin, deh, kalo PT Perorangan ini implementasi UU Cipta Kerja, ngurusnya serba online. Ada nomor pendirian dari Kemenkumham.	gw sempet nanya, sih, apakah mbak cs familiar atau pernah akses @ahu_online dan @oss_id. jawabnya tentu saja belum. terus gw jelasin, deh, kalo pt perorangan ini implementasi uu cipta kerja, ngurusnya serba online. ada nomor pendirian dari kemenkumham.

b. Tokenizing

Tokenizing adalah tugas memisahkan deretan kata di dalam kalimat, *paragraph* atau halaman menjadi token atau potongan kata. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca, mention, hastag dan link. (Aditia Rakhmat Sentiaji et al. 2014)

Tabel 5. Hasil *Tokenizing*

Sebelum <i>Tokenizing</i>	Hasil <i>Tokenizing</i>
Gw sempet nanya, sih, apakah mbak CS familiar atau pernah akses @AHU_Online dan @OSS_id. Jawabnya tentu saja belum. Terus gw jelasin, deh, kalo PT Perorangan ini implementasi UU Cipta Kerja, ngurusnya serba online. Ada nomor pendirian dari Kemenkumham..	['gw', 'sempet', 'nanya', 'sih', 'apakah', 'mbak', 'cs', 'familiar', 'atau', 'pernah', 'akses', 'online', 'dan', 'id', 'jawabnya', 'tentu', 'saja', 'belum', 'terus', 'gw', 'jelasin', 'deh', 'kalo', 'pt', 'perorangan', 'ini', 'implementasi', 'uu', 'cipta', 'kerja', 'ngurusnya', 'serba', 'online', 'ada', 'nomor', 'pendirian', 'dari', 'kemenkumham']

c. *Stopwords*

Stop word didefinisikan sebagai term yang tidak berhubungan (irrelevant) dengan subyek utama dari database meskipun katater sebut sering kali hadir di dalam dokumen. (Aditia Rakhmat Sentiaji et al. 2014)

Tabel 6. Hasil *Stopwords*

Sebelum <i>Stopwords</i>	Hasil <i>Stopwords</i>
Gw sempet nanya, sih, apakah mbak CS familiar atau pernah akses @AHU_Online dan @OSS_id. Jawabnya tentu saja belum. Terus gw jelasin, deh, kalo PT Perorangan ini implementasi UU Cipta Kerja, ngurusnya serba online. Ada nomor pendirian dari Kemenkumham..	['gw', 'sempet', 'nanya', 'mbak', 'cs', 'familiar', 'akses', 'online', 'id', 'gw', 'jelasin', 'deh', 'pt', 'perorangan', 'implementasi', 'uu', 'cipta', 'kerja', 'ngurusnya', 'serba', 'online', 'nomor', 'pendirian', 'kemenkumham']

d. *Stemming*

Stemming merupakan proses tahap pencarian kata dasar dari hasil proses filtering dengan membuang kata imbuhan. (Aditia Rakhmat Sentiaji et al. 2014)

Tabel 7. Hasil *Stemming*

Sebelum <i>Stemming</i>	Hasil <i>Stemming</i>
Gw sempet nanya, sih, apakah mbak CS familiar atau pernah akses @AHU_Online dan @OSS_id. Jawabnya tentu saja belum. Terus gw jelasin, deh, kalo PT Perorangan ini implementasi UU Cipta Kerja, ngurusnya serba online. Ada nomor pendirian dari Kemenkumham..	['gw', 'sempet', 'tanya', 'mbak', 'cs', 'familiar', 'akses', 'online', 'id', 'gw', 'jelas', 'deh', 'pt', 'orang', 'implementasi', 'uu', 'cipta', 'kerja', 'ngurusnya', 'serba', 'online', 'nomor', 'diri', 'kemenkumham']

Pembobotan Kata (TF-IDF)

Metode TF-IDF merupakan metode penghitungan bobot setiap kata yang paling umum digunakan dalam temu kembali informasi. Cara ini juga dikenal efektif, mudah, dan memberikan hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)* untuk setiap token (*word*). Berikut beberapa kata yang sudah memiliki nilai:

Tabel 8. Hasil Pembobotan Kata

Term	TF	TF-IDF
aksi	0.0625	0.26142149153041117
download	0.0625	0.41128994608030933
hp	0.0625	0.36796824729531274
masingmasi ng	0.0625	0.41128994608030933
jagajaga	0.0625	0.41128994608030933

tangkap	0.0625	0.342626678038 55246
ditanyain	0.0625	0.385948376823 54904
demo	0.0625	0.212661581683 56272
kasih	0.0625	0.294302310023 95985
muka	0.0625	0.367968247295 31274
lho	0.0625	0.317285108781 79217
perpu	0.0625	0.176214938849 46166

Pemodelan dan Pengujian

a. Pemodelan

```

from sklearn.model_selection import train_test_split
x = THEET_DATA[["text"]]
y = THEET_DATA[["label"]]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=10)

from sklearn.metrics import classification_report, f1_score, accuracy_score, precision_recall_fscore_support as score,
confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

def predict(model, X):
    y = model.predict(X)
    return y

def metrics(y_true, y_pred):
    precision, recall, fscore, support = score(y_true, y_pred, average="macro")
    accuracy = accuracy_score(y_true, y_pred)
    print("Accuracy : {}".format(accuracy))
    print("Precision : {}".format(precision))
    print("Recall : {}".format(recall))
    print("F1-Score : {}".format(fscore))

    print(classification_report(y_true, y_pred, digits=4))
    mat = confusion_matrix(y_true, y_pred, labels=[-1,1])
    sns.heatmap(mat, square=True, annot=True, cbar=True, xticklabels=[-1,1], yticklabels=[-1,1])
    plt.xlabel("Predicted Label")
    plt.ylabel("True Label")
    plt.show()
    
```

Gambar 2. Code Pemodelan

Pemodelan diatas diawali dengan membagi data latih dan data uji, dimana data latih sebanyak 80 % dan data uji 20%, pemodelan diatas juga membuat model untuk pengujian *confusion matrix* dengan menggunakan performance matrix seperti *Accuracy, precision, recall, dan fscore*.

b. Pengujian Algoritma Naïve Bayes

```

from sklearn.naive_bayes import MultinomialNB #naive bayes classifier
from sklearn.pipeline import make_pipeline
from sklearn.feature_extraction.text import TfidfVectorizer

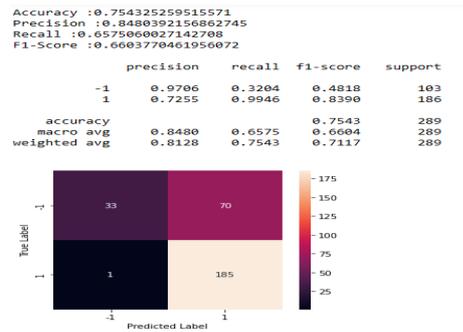
nbmodel = make_pipeline(TfidfVectorizer(), MultinomialNB())

nbmodel.fit(X_train, y_train)

y_test_pred = predict(nbmodel, X_test)
metrics(y_test, y_test_pred)
    
```

Gambar 3. Code Pengujian Naïve Bayes

Berikut adalah output dari pengujian algoritma naïve bayes menggunakan *confusion matrix*,



Gambar 4. Output Pengujian

Dari hasil output diatas akurasi yang didapat adalah 75.43 %, Dari *confusion matrix* dapat dijelaskan bahwa model mengklasifikasikan secara benar sebesar 185 data sebagai positif, dan 33 sebagai data negatif.

c. Pengujian Algoritma Support Vector Machine

Setelah menentukan uji klasifikasi dengan algoritma naïve bayes, maka selanjutnya menentukan parameter model klasifikasi menggunakan algoritma support vector machine (SVM) dengan *code* python sebagai berikut,

```

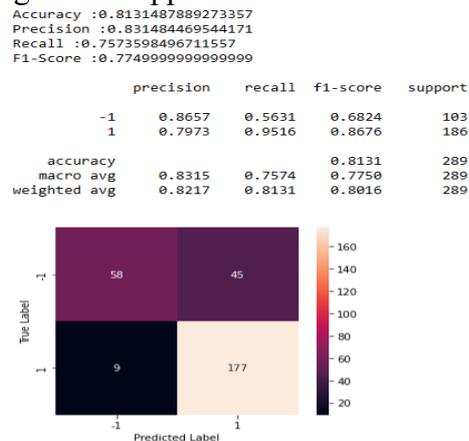
from sklearn.svm import LinearSVC #svm classifier
svmmodel = make_pipeline(TfidfVectorizer(), LinearSVC())

svmmodel.fit(X_train, y_train)

y_test_pred = predict(svmmodel, X_test)
metrics(y_test, y_test_pred)
    
```

Gambar 5. Code Pengujian SVM

Berikut adalah output dari pengujian algoritma support vector machine di atas,



Gambar 6. Output Pengujian

Dari hasil output diatas akurasi yang didapat adalah 81.31 %, Dari *confusion matrix* dapat dijelaskan bahwa model mengklasifikasikan secara benar sebesar

177 data sebagai positif, dan 58 sebagai data negatif.

SIMPULAN

Berdasarkan hasil pengujian algoritma Naive Bayes Classifier yang telah dilakukan ada beberapa hal yang dihasilkan, antara lain:

- a. Pada penelitian ini, algoritma Naive Bayes menghasilkan nilai akurasi sebesar 75,43%. dan memiliki nilai Precision sebesar 84,80%, nilai Recall sebesar 65,75% dan nilai F1-Score sebesar 66,03%.
- b. algoritma Support Vector Machine menghasilkan nilai akurasi sebesar 81,31%. memiliki nilai Precision sebesar 83,14%, nilai Recall sebesar 75,73% dan nilai F1-Score sebesar 77,49%.
- c. Berdasarkan hasil pengujian dari kedua metode, metode Support Vector Machine (SVM) mempunyai nilai yang lebih tinggi dibandingkan dengan metode Naive Bayes sehingga yang kita ambil adalah metode Support Vector Machine (SVM).

DAFTAR PUSTAKA

- Aditia Rakhmat Sentiaji, Adam Mukaharil Bachtiar et al. (2014). Analisis Sentimen Terhadap Acara Televisi Berdasarkan Opini Publik. *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*.
- Ainurrohman. (2021). Akurasi Algoritma Klasifikasi Pada Software Rapidminer Dan Weka. *Prosiding Seminar Nasional Matematika* 4: 493–99.
<https://journal.unnes.ac.id/sju/index.php/prisma/>.
- Azhari, Mulkan, Zakaria Situmorang, and Rika Rosnelly. (2021). “Perbandingan Akurasi, Recall, Dan Presisi Klasifikasi Pada Algoritma C4.5, Random Forest, SVM Dan Naive Bayes. *Jurnal Media Informatika Budidarma* 5(2): 640.
- Azis Maarif, Abdul. (2015). Penerapan Algoritma Fuzzy.” *Universitas Dian Nusantara*.
repository.unair.ac.id/29371/3/15
BAB II.pdf.
- Irwansyah Saputra, Dinar Ajeng Kristiyanti. (2022). *Machine Learning Untuk Pemula*. Informatika Bandung.
- Kartikasari, Hesty, and Agus Machfud Fauzi. (2021). “Penolakan Masyarakat Terhadap Pengesahan Omnibus Law Cipta Kerja Dalam Perspektif Sosiologi Hukum.” *Doktrina: Journal of Law* 4(1): 39–52.
- Liu, Chin-Hung, and Wei-Shih Tsai. (2010). “The Effects of Service Quality and Lifestyle on Consumer Choice of Channel Types: The Health Food Industry as an Example.” *African Journal of Business Management* 4(6): 1023–39.
<http://www.academicjournals.org/AJBM>.
- Muttaqin, Firdaus Akhmad, and Adam Mukaharil Bachtiar. (2016). Implementasi Teks Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak ‘Dodo Kids Browser. *Jurnal Ilmiah Komputer dan Informatika*: 1–8.
- Nurzahputra, Aldi, and Much Aziz Muslim. (2016). Analisis Sentimen Pada Opini Mahasiswa Menggunakan Natural Language Processing. *Seminar Nasional Ilmu Komputer (Snik)*: 114–18.
- Sisfotenika, Jurnal. (2023). Analisis Sentimen Dengan Algoritma SVM Dalam Tanggapan Netizen Terhadap Berita Resesi 2023 Analysis Sentiment Using the SVM Algorithm in Netizen Responses to News of the 2023 Recession.13(1): 53–64.