

ANALISIS SENTIMEN RANCANGAN UNDANG-UNDANG TINDAK PIDANA KEKERASAN SEKSUAL/UNDANG-UNDANG TINDAK PIDANA KEKERASAN SEKSUAL PENGGUNA *TWITTER* MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER*

Adam Panca Putra Pinaria¹, Widodo², Murien Nugraheni³

¹ Mahasiswa Prodi Pendidikan Teknik Informatika dan Komputer, Teknik Elektro, FT – UNJ

² Dosen Prodi Pendidikan Teknik Informatika dan Komputer, Teknik Elektro, FT – UNJ

³ Dosen Prodi Sistem Teknologi Informasi, Teknik Elektro, FT – UNJ

¹ adampancaputrapinaria_1512618050@mhs.unj.ac.id, ² widodo@unj.ac.id, ³ muriennugraheni@unj.ac.id

Abstrak

Penelitian ini bertujuan untuk melakukan analisis sentimen Rancangan Undang-Undang Tindak Pidana Kekerasan Seksual/Undang-Undang Tindak Pidana Kekerasan Seksual pengguna *Twitter* menggunakan metode *Naïve Bayes Classifier* dan mengevaluasi kinerja klasifikasinya. Analisis sentimen pada penelitian ini menggunakan metode *Naïve Bayes Classifier* dengan pembobotan *Term Frequency-Inverse Document Frequency*. Proses pengumpulan data dilakukan menggunakan pustaka *Snsrape Python*. Data pakai yang digunakan untuk dibersihkan hanya yang bersentimen positif dan negatif. Proses pembersihan data melalui tahapan *preproses teks* yaitu *Case Folding*, *Cleansing*, *Tokenizing*, *Stopword Removal*, *Stemming* dan *Normalization*. Analisis sentimen yang dihasilkan berupa 12.807 data, dengan sentimen negatif sebanyak 10.263 data (80,1%) dan sentimen positif sebanyak 2.544 data (19,9%). Sentimen pengguna *twitter* Indonesia mengarah pada sentimen negatif. Pengujian kinerja model klasifikasi metode analisis sentimen penelitian ini menggunakan teknik *Confusion Matrix* pada hasil klasifikasi data uji dengan nilai *Accuracy* klasifikasi algoritma sebesar 93%, nilai *Precision* klasifikasi sebesar 99%, nilai *Recall* klasifikasi sebesar 87% dan *F- Measure* sebesar 93%. Dari nilai pengujian yang dihasilkan dapat disimpulkan model klasifikasi dengan metode algoritma *Naïve Bayes Classifier* memiliki kinerja yang sangat baik.

Kata kunci : *Analisis Sentiment, Naïve Bayes Classifier, Confusion Matrix, UUTPKS, Twitter.*

1. Pendahuluan

Teknologi informasi yang semakin pesat berkembang memberikan manusia kemudahan dalam setiap keperluannya. Salah satu hasil perkembangan teknologi tersebut adalah *platform* digital komunikasi yang disebut media sosial. Media sosial memberikan kemudahan komunikasi instan serta menjadi *platform* berbagi informasi secara publik. Kemudahan tersebut dimanfaatkan untuk melakukan periklanan, berbagi opini dll (Brogan, 2010). Penggunaan media sosial dalam publikasi kebijakan dan pernyataan resmi oleh pemerintah atau organisasi memberikan kesempatan masyarakat dalam merespons secara langsung. Komunikasi langsung ini juga menjadi keunggulan media sosial daripada media cetak. Media sosial yang digunakan untuk menyampaikan opini salah satunya adalah *Twitter*. Banyak pengguna menggunakan *Twitter* sebagai bentuk kampanye atau demonstrasi digital. hal ini dapat dilihat dari demonstrasi September 2019 yang menjadikan beberapa media sosial salah satunya *Twitter* sebagai lokasi demonstrasi digitalnya.

Demonstrasi September 2019 memperlihatkan keterlibatan media sosial dengan *tweet* yang diikuti tanda pagar (#) mengenai topik atau tuntutan demonstrasi tertentu. Salah satu topik atau tuntutan demonstrasi September 2019 adalah "... pengesahan/disahkannya Rancangan Undang-undang Penghapusan Kekerasan Seksual (RUU PKS)" dengan tanda pagar #SahkanRUUPKS (Nasional.kompas.com, 2019). Selama tuntutan diberikan terjadinya proses keluar masuknya RUU tersebut dari daftar Program Legislasi Nasional (Prolegnas). RUU ini pun berubah nama pada akhir Agustus 2021 menjadi Rancangan Undang-undang Tindak Pidana Kekerasan Seksual (RUU TPKS). Perubahan tersebut dilakukan oleh Badan Legislasi (Baleg) Dewan Perwakilan Rakyat (DPR). Perubahan nama ini juga mengubah tanda pagar tuntutan menjadi #SahkanRUUTPKS. Pada akhir tahun 2021 RUU TPKS

Baleg menetapkan RUU TPKS untuk masuk dalam Prioritas 2022. Pengesahan RUU TPKS menjadi usul inisiatif DPR dilaksanakan pada awal tahun 2022. RUU TPKS kemudian resmi disahkan menjadi UU No 12 tahun 2022 tentang Tindak Pidana Kekerasan Seksual (UU TPKS) pada tanggal 12 April 2022 (Indonesia, 2022). Pembahasan hingga pengesahan undang-undang ini menjadi polemik di masyarakat karena adanya penolakan mulai dari fraksi partai politik, pelajar hingga masyarakat umum. Opini penolakan tersebut juga hadir melawan opini yang mendukung pengesahan di media sosial *Twitter* dengan tanda pagar #TolakRUUTPKS dan #GagalkanRUUTPKS.

Pendapat masyarakat Indonesia yang terbagi dua yaitu mendukung atau menolak RUU TPKS/UU TPKS sangat banyak di *Twitter* sehingga tidak diketahui secara pasti apakah masyarakat mendukung atau menolaknya. Pendapat tersebut menjadi data yang dapat digunakan dalam proses analisis sentimen. Analisis sentimen sendiri digunakan untuk melihat kecenderungan opini atau pendapat dari sebuah masalah atau topik tertentu.

Analisis sentimen memiliki beberapa metode atau algoritma yang dapat digunakan. Beberapa metode tersebut adalah *Naïve Bayes Classifier* (NBC), *Support Vector Machine* (SVM), dan *k-Nearest Neighbour* (k-NN) dan lain-lain. Setiap analisis sentimen suatu topik dengan suatu metode perlu dievaluasi kinerjanya dengan Teknik analisis data yang sesuai untuk melihat keakuratan metode yang digunakan. Analisis sentimen pada topik RUU TPKS/UU TPKS di *Twitter* sendiri belum dilakukan. Sehingga belum adanya model klasifikasi yang bisa digunakan untuk analisis sentimen topik RUU TPKS/UU TPKS di *Twitter* dan belum diketahuinya kinerja dan akurasi dari suatu metode yang digunakan. Salah satu metode yang digunakan adalah metode *Naïve Bayes Classifier* (NBC). Menurut Ratnawati (Ratnawati, 2018), NBC merupakan sebuah metode klasifikasi yang berakar pada Teorema *Bayes*. Metode pengklasifikasian ini menggunakan metode probabilitas dan statistik yaitu memprediksi peluang berdasarkan pengalaman di masa sebelumnya (Teorema *Bayes*) dengan ciri utamanya adalah asumsi yang sangat kuat (naif) akan ketergantungan dari masing-masing kondisi/kejadian.

NBC sendiri memiliki tingkat akurasi yang tinggi dan juga sering digunakan dalam pemodelan model klasifikasi. Hal ini dapat dilihat dari penelitian Ratnawati (Ratnawati, 2018) dan Naraswati, Dkk (2021). Dalam penelitian Ratnawati (2018) dengan judul “Implementasi Algoritma *Naive Bayes* Terhadap Analisis Sentimen Opini Film Pada *Twitter*” menggunakan *5 fold cross validation* sedangkan penelitian Naraswati, Dkk (Naraswati, et al., 2021) “Analisis Sentimen Publik dari *Twitter* Tentang Kebijakan Penanganan COVID-19 di Indonesia Dengan *Naïve Bayes Classification*”. Pada penelitian Ratnawati (Ratnawati, 2018) analisis sentimen yang dilakukan memiliki akurasi sebesar 90 % dengan rincian nilai *precision* 92%, *recall* 90% dan *f-measure* 90% dengan komposisi 80% data latih dan 20% data uji. Pada penelitian (Naraswati, et al., 2021) dengan data valid uji sebanyak 2.330 *tweets* diperoleh tingkat akurasi klasifikasi sebesar 87,34%, sensitivitas sebesar 93,43%, dan spesifisitas 71,76%. NBC juga memiliki beberapa keunggulan yaitu dapat dipakai untuk data kuantitatif maupun kualitatif serta tidak memerlukan jumlah data yang banyak. Pada proses analisis NBC juga tidak perlu melakukan *training* dengan data yang banyak dan memiliki perhitungan yang cepat dan efisien. Pengklasifikasian dokumen dengan NBC dapat disesuaikan dengan kebutuhan setiap orang, serta dapat digunakan untuk klasifikasi masalah biner maupun *multi*. Berdasarkan latar belakang tersebut penulis mengadopsi algoritma *Naïve Bayes Classifier* (NBC) sebagai metode dalam penelitian ini dengan mengangkat judul “Analisis Sentimen Rancangan Undang-Undang Tindak Pidana Kekerasan Seksual/Undang-Undang Tindak Pidana Kekerasan Seksual Pengguna *Twitter* Menggunakan Algoritma *Naïve Bayes Classifier*”.

2. Dasar Teori

2.1. Text Mining

Text mining merupakan analisis teks yang dapat mengekstrak sumber data dari dokumen dengan tujuan menemukan kata-kata yang dapat menggambarkan isi dokumen sehingga dapat dilakukan analisis hubungan, keterkaitan dan kelas antar dokumen (Lesmeister, 2015).

2.2. Analisis Sentimen

Analisis sentimen adalah metode yang digunakan untuk mengekstrak data opini untuk memahami dan memproses data teks untuk memastikan sentimen yang terkandung dalam opini (Fransiska & Yolanda, 2019).

2.3. RUUTPKS / UUTPKS

“Rancangan Undang-Undang tentang Tindak Pidana Kekerasan Seksual (RUUTPKS) adalah instrumen penting untuk membangun moralitas masyarakat dan bangsa yang berkemanusiaan yang adil dan beradab berdasarkan ketuhanan Yang Maha Esa, serta langkah penting untuk membangun ketahanan keluarga, yang merupakan pilar penting dalam pencegahan kekerasan seksual” (Republik Indonesia, 2021). Undang-Undang tentang Tindak Pidana Kekerasan Seksual (UU TPKS) adalah Undang-Undang No. 12 Tahun 2022. UU TPKS

dalam kasus kekerasan seksual menjadi landasan hukum yang memberikan layanan, pendampingan, perlindungan, pemulihan, restitusi dan rehabilitasi pada pihak yang terkait dan membutuhkan (Indonesia, 2022).

2.4. *Twitter*

Twitter adalah layanan media sosial dan *microblogging online* yang memungkinkan pengguna untuk mengirim dan membaca pesan berbasis teks hingga 280 karakter, yang disebut *Tweet* (Ratnawati, 2018).

2.5. *Google Colaboratory*

Google Colaboratory atau *Google Colab* adalah lingkungan *Jupyter notebook* gratis yang berbasis *cloud*. *Google Colab* mampu digunakan untuk menulis dan menjalankan kode, menyimpan dan membagikan hasil analisis, serta mengakses sumber daya komputasi yang canggih, semuanya dari *browser* secara gratis (Gunawan, et al., 2020).

2.6. *Python*

Python merupakan bahasa pemrograman interaktif dan berjalan di berbagai platform atau aplikasi yang berbeda. Ada banyak dukungan *library* untuk menulis program *python* untuk pemrosesan data dan *machine learning* (Bujang, et al., 2021).

2.6.1. *Snsrape*

Snsrape adalah pustaka *Python* yang dapat mengatur dan mengelola informasi di *Twitter*, memungkinkan pengguna untuk dengan mudah mengambil data *tweet* tanpa batas secara langsung (Mulyani, 2022).

2.7. *Naive Bayes Classifier*

Naïve Bayes Classifier merupakan teknik klasifikasi berdasarkan teorema *Bayes* dengan asumsi independen antara prediksi. *Naïve Bayes Classifier* memprediksi peluang masa depan berdasarkan pengalaman masa lalu dikenal sebagai teorema *Bayes* (Ali, et al., 2019). Dengan *Naive Bayes*, data berskala kecil dapat diolah untuk memberikan hasil prediksi real-time, menurut Wongkar & Angdresey (2019). Rumus *Bayes* dalam (Muslehatin, Ibnu, & Mustakim, 2017) secara umum dapat diberikan sebagai berikut:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Keterangan :

X = Data dengan *class* yang belum diketahui

H = Hipotesis data X merupakan suatu *class* spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (*posteriori prob.*)

P(H) = Probabilitas hipotesis H (*prior prob.*)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

2.8. *Preprocessing*

Preprocessing diperlukan untuk menghapus data yang tidak diperlukan dan mengubah data teks tidak terstruktur menjadi data teks terstruktur atau semi terstruktur (Sivi, 2018).

2.8.1. *Case Folding*

Case folding adalah proses mengubah karakter dalam data menjadi huruf kecil (*lowercase*) dan menghilangkan non-karakter (Indraloka & Santosa, 2017).

2.8.2. *Cleansing*

Cleansing adalah proses menghilangkan kata-kata yang tidak perlu untuk mengurangi *noise* (Syarif, Herdiani, & Astuti, 2019).

2.8.3. *Tokenizing*

Tokenizing merupakan proses pemenggalan kalimat menjadi kata-kata yang dikenal dengan sebutan token. Token yang dimaksud adalah kata yang dipisahkan oleh spasi (Harijiatno, 2019).

2.8.4. *Stopword Removal*

Tujuan dari *stopword removal* adalah untuk menghilangkan kata-kata yang bersifat umum dan sering muncul serta memiliki sedikit relevansi dengan teks (Syarif, Herdiani, & Astuti, 2019).

2.8.5. Stemming

Hasil dari proses *stemming* ini adalah pembentukan akar kata. Akar kata adalah bagian kata yang tersisa setelah dihilangkan imbuhan (Indraloka & Santosa, 2017).

2.8.6. Normalization

Pada tahap ini, bentuk tidak baku diubah menjadi bentuk baku untuk menghindari duplikasi kata dengan makna yang sama (Harijatio, 2019).

2.9. Term Frequency – Inverse Document Frequency (TF-IDF)

Term Frequency (TF) adalah frekuensi atau jumlah kemunculan kata dalam dokumen tertentu, yang menunjukkan pentingnya kata tersebut dalam setiap dokumen. *Document Frequency* (DF) adalah frekuensi dokumen yang berisi kata tersebut menunjukkan seberapa umum kata tersebut, Sedangkan *Inverse Document Frequency* (IDF) adalah fungsi kebalikan dari nilai *Document Frequency* (DF). Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan proses yang dilakukan untuk mengubah data dari data teks menjadi data numerik untuk memberi bobot kata demi kata dengan menghitung TF dikalikan IDF. Jika kata tersebut sering muncul dalam sebuah dokumen maka bobot kata semakin besar. Bobot kata semakin kecil jika kata muncul dalam beberapa dokumen (Septian, Fachrudin, & Nugroho, 2019).

Rumus persamaan dalam pembobotan TF-IDF adalah :

$$TF \cdot IDF = TF \times IDF \quad (2)$$

$$IDF = \log (D/DF) \quad (3)$$

Keterangan:

TF = frekuensi term dalam dokumen

IDF = kemunculan sebuah term dalam banyak dokumen

DF = frekuensi dokumen yang memuat term

D = dokumen

2.10. Confusion Matrix

Confusion matrix adalah alat yang digunakan untuk mengukur seberapa baik *classifier* digunakan untuk mengenali rangkap dari kelas yang berbeda (Han, Kamber, & Pei, 2012). Pengukuran yang diterapkan pada Confusion Matrix dilakukan dengan menghitung *Accuracy*, *Precision*, *Recall*, dan *f-measures* yang terkait dengan operasi nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang merupakan nilai pada *confusion matrix*.

Tabel 2.1 Confusion Matrix

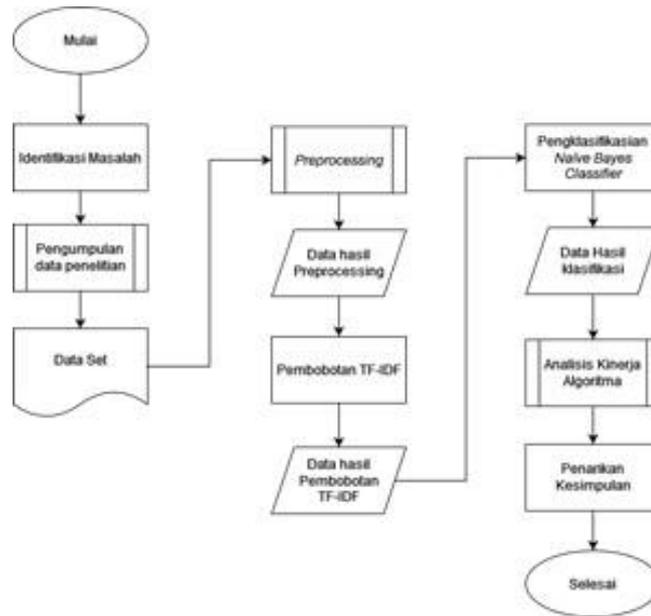
		Kelas Prediksi	
		<i>True Positive</i> (TP)	<i>False Positive</i> (FP)
Kelas Sebenarnya		<i>False Negative</i> (FN)	<i>True Negative</i> (TN)

Sumber: (Han, Kamber, & Pei, 2012)

3. Metodologi

3.1. Diagram Alir Penelitian

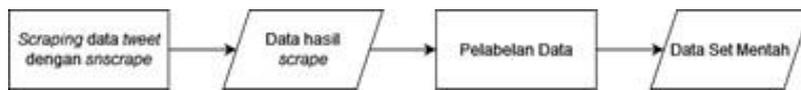
Penelitian ini dilakukan dengan melakukan eksperimen melalui simulasi di komputer dengan menggunakan bahasa pemrograman *python*. Diagram alir penelitian ini dapat dilihat pada gambar 3.1 Berikut diagram alir penelitian :



Gambar 3.1 Diagram Alir Penelitian

3.1.1. Pembuatan *Dataset* Penelitian

Langkah awal pada penelitian ini adalah dengan melakukan pembuatan *dataset* penelitian. Adapun tahapan pembuatan *dataset* dapat dilihat pada gambar 3.2 berikut.

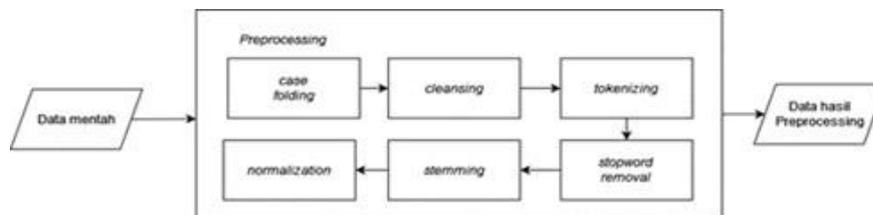


Gambar 3.2 Tahapan Pembuatan *Dataset* Penelitian

Data yang telah terkumpul kemudian diberi label positif, negatif dan netral. Hasil dari pelabelan data kemudian diambil data yang memiliki label positif dan negatif untuk kemudian digunakan. Data dengan label netral dilepas karena tidak akan digunakan dalam proses klasifikasi.

3.1.2. Pengolahan Data Penelitian

Data mentah yang masih kotor dimasukkan ke tahap ini untuk diproses. Hasil dari proses ini adalah dokumen yang berkualitas yang mempermudah proses klasifikasi. Tahapan pengolahan data dapat dilihat pada gambar 3.3.



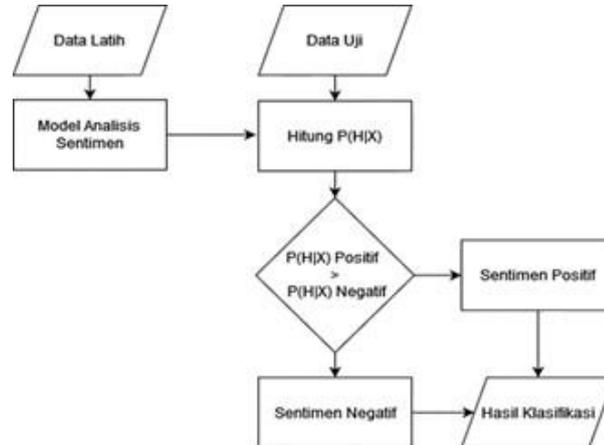
Gambar 3.3 Tahapan *Preprocessing*

Tabel 3.1 Contoh Hasil Proses *Preprocessing*

Tweet	Setelah Proses Pre-Processing
@__AnakKolong Hidup dari uang Negara, tp merongrong persatuan dan kesatuan bangsa 😏😏😏 Tindakan nyata nyaris tak terdengar UUTPKS aja ditentang, emangnya cewe ga layak dilindungi.. 😏😏😏	['hidup', 'uang', 'negara', 'rongrong', 'satu', 'satu', 'bangsa', 'tindak', 'nyata', 'nyaris', 'dengar', 'cewek', 'layak', 'lindung']

3.1.3. Pengklasifikasian dengan *Naïve Bayes Classifier*

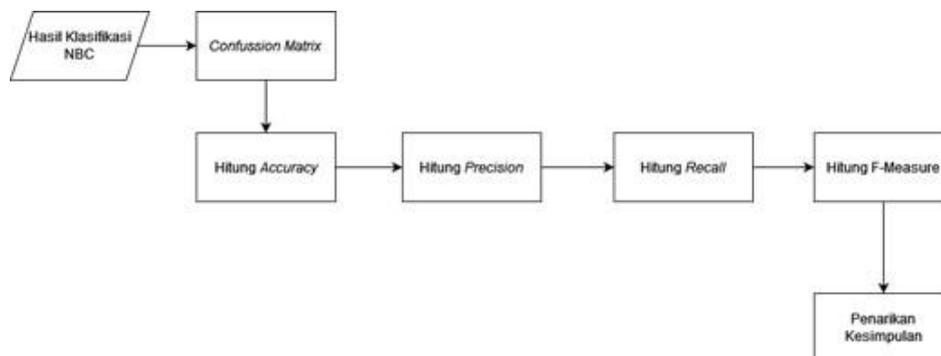
Pada tahap ini proses klasifikasi menggunakan metode *Naïve Bayes Classifier* yang dibagi menjadi dua proses, yaitu proses latih dan proses uji. Pada tahap ini dilakukan proses latih terlebih dahulu untuk pelatihan, selanjutnya dilakukan proses uji dengan mengacu probabilitas dari *dataset* latih. Diagram alir tahapan klasifikasi *Naïve Bayes Classifier* ini ditunjukkan pada gambar 3.4 berikut.



Gambar 3.4 Tahapan Klasifikasi *Naïve Bayes Classifier*

3.1.4. Analisis Kinerja Algoritma

Pada tahap ini performa dari *classifier* kemudian akan dievaluasi menggunakan tabel *Confusion Matrix* dengan membandingkan hasil klasifikasi dengan label data uji sebenarnya untuk menghitung nilai *Accuracy*, *Precision*, *Recall*, dan *F-Measure*. Tahapan analisis dapat dilihat pada gambar 3.5.



Gambar 3.5 Tahapan Analisis Kinerja Algoritma

3.1. Teknik dan Prosedur Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data *tweet* yang didapatkan dari media sosial *Twitter*. Data yang dikumpulkan berupa data teks *tweet* pada *Twitter* yang diambil dengan menggunakan pustaka *snsrape python* di *google Colab*. Data yang diambil pada penelitian ini menggunakan kata kunci ‘SahkanRUUTPKS’, ‘TolakRUUTPKS’, ‘GagalkanRUUTPKS’, ‘KekerasanSeksual’ dan ‘UUTPKS’. Pada proses pelabelan *tweet* dibedakan menjadi 3 kelas, yaitu nilai positif, netral dan nilai negatif. Penelitian ini hanya mengambil data dengan label nilai positif dan negatif kemudian akan di *preprocessing*. Data tersebut akan dibagi menjadi data latih dan data uji dengan perbandingan 80% data latih dan 20% data uji.

3.2. Teknik Analisis Data

/Persamaan untuk menemukan nilai *Accuracy*, *Precision*, *Recall*, *F-Measure*:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (6)$$

$$F - \text{Measure} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (7)$$

Nilai pada *Accuracy*, *Precision*, *Recall*, dan *F-Measure* dapat memberikan perbandingan klasifikasi model dengan klasifikasi yang sebenarnya. Perbandingan tersebut membantu penarikan kesimpulan mengenai kinerja algoritma yang digunakan dalam nilai persen (%). Nilai persen tersebut digunakan untuk menilai kinerja algoritma.

4. Hasil dan Analisis

4.1. Pengumpulan Data

Data yang sudah diambil menggunakan pustaka *snsrscrape* kemudian digabung dan dilabeli yang selanjutnya disebut data utama mentah berjumlah 14.476 *tweets* dengan sentimen positif sebanyak 2.711 *tweets*, sentimen negatif sebanyak 11.685 *tweets* dan sentimen netral sebanyak 80 *tweets*. Dengan hanya mengambil data *tweets* sentimen positif dan negatif, maka didapatkan data utama mentah sebanyak 14.396 *tweets*.

4.2. Preproses Teks Data

Case Folding

Tabel 4.1 Sampel Hasil Case Folding

Teks Awal	Teks Case Folding
Dengan adanya "penumpang gelap" yang akan menguntungkan golongan LGBT, sexual consent, matriofocal dan pemulihan korban kekerasan seksual dengan perspektif feminis, maka sudah tepat jika menyuarakan #gagalkanruutpks https://t.co/WKGV2Iuzlp	dengan adanya "penumpang gelap" yang akan menguntungkan golongan lgbt, sexual consent, matriofocal dan pemulihan korban kekerasan seksual dengan perspektif feminis, maka sudah tepat jika menyuarakan #gagalkanruutpks https://t.co/wkgv2iuzlp

Cleansing

Tabel 4.2 Sampel Hasil Cleansing

Teks Case Folding	Teks Cleansing
dengan adanya "penumpang gelap" yang akan menguntungkan golongan lgbt, sexual consent, matriofocal dan pemulihan korban kekerasan seksual dengan perspektif feminis, maka sudah tepat jika menyuarakan #gagalkanruutpks https://t.co/wkgv2iuzlp	dengan adanya penumpang gelap yang akan menguntungkan golongan lgbt sexual consent matriofocal dan pemulihan korban kekerasan seksual dengan perspektif feminis maka sudah tepat jika menyuarakan

Tokenizing

Tabel 4.3 Sampel Hasil Tokenizing

Teks Cleansing	Teks Tokenizing
dengan adanya penumpang gelap yang akan menguntungkan golongan lgbt sexual consent matriofocal dan pemulihan korban kekerasan seksual dengan perspektif feminis maka sudah tepat jika menyuarakan	['dengan', 'adanya', 'penumpang', 'gelap', 'yang', 'akan', 'menguntungkan', 'golongan', 'lgbt', 'sexual', 'consent', 'matriofocal', 'dan', 'pemulihan', 'korban', 'kekerasan', 'seksual', 'dengan', 'perspektif', 'feminis', 'maka', 'sudah', 'tepat', 'jika', 'menyuarakan']

Stopword Removal

Tabel 4.4 Sampel Hasil Stopword Removal

Teks Tokenizing	Teks Stopword Removal
-----------------	-----------------------

['dengan', 'adanya', 'penumpang', 'gelap', 'yang', 'akan', 'menguntungkan', 'golongan', 'lgbt', 'sexual', 'consent', 'matriofocal', 'dan', 'pemulihan', 'korban', 'kekerasan', 'seksual', 'dengan', 'perspektif', 'feminis', 'maka', 'sudah', 'tepat', 'jika', 'menyuarakan']	['penumpang', 'gelap', 'menguntungkan', 'golongan', 'lgbt', 'sexual', 'consent', 'matriofocal', 'pemulihan', 'korban', 'kekerasan', 'seksual', 'perspektif', 'feminis', 'menyuarakan']
---	--

Stemming

Tabel 4.5 Sampel Hasil Stemming

Teks Stopword Removal	Teks Stemming
['penumpang', 'gelap', 'menguntungkan', 'golongan', 'lgbt', 'sexual', 'consent', 'matriofocal', 'pemulihan', 'korban', 'kekerasan', 'seksual', 'perspektif', 'feminis', 'menyuarakan']	['tumpang', 'gelap', 'untung', 'golong', 'lgbt', 'sexual', 'consent', 'matriofocal', 'pulih', 'korban', 'keras', 'seksual', 'perspektif', 'finis', 'suara']

Normalization

Tabel 4.6 Sampel Hasil Normalization

Teks Stemming	Teks Normalization
['tumpang', 'gelap', 'untung', 'golong', 'lgbt', 'sexual', 'consent', 'matriofocal', 'pulih', 'korban', 'keras', 'seksual', 'perspektif', 'finis', 'suara']	['tumpang', 'gelap', 'untung', 'golong', 'lgbt', 'seksual', 'konsen', 'matriofocal', 'pulih', 'korban', 'keras', 'seksual', 'perspektif', 'feminis', 'suara']

Dari 14.396 data *tweets* yang di preproses didapatkan 12.807 data *tweets* bersih dengan komposisi sentimen 2.544 data *tweets* positif dan 10.263 data *tweets* negatif. Data hasil *preprocessing text* kemudian dibagi menjadi data latih sebanyak 80% dan data uji sebanyak 20%.

4.3. Klasifikasi Dengan Naïve Bayes Classifier (NBC)

Berdasarkan data hasil prediksi didapat sebanyak 1127 data *tweets* positif dan 1435 data *tweets* negatif. Dengan komposisi *true positive* (TP) sebanyak 1114 data, *true negative* (TN) 1268 data, *false positive* (FP) 13 data dan *false negative* (FN) 167 data.

Tabel 4.7 Confusion Matrix Hasil Klasifikasi

Confusion Matrix		Kelas Prediksi	
		Positif (1)	Negatif (0)
Kelas Sebenarnya	Positif (1)	1114	167
	Negatif (0)	13	1268

4.4. Pembahasan

Berikut hasil perhitungan confusion matrix berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{1114+1268}{1114+1268+13+167} = 0,93 \quad (4)$$

$$Precision = \frac{TP}{TP+FP} = \frac{1114}{1114+13} = 0,99 \quad (5)$$

$$Recall = \frac{TP}{TP+FN} = \frac{1114}{1114+167} = 0,87 \quad (6)$$

$$F - Measure = 2 \times \frac{Recall \times Precision}{Recall+Precision} = 2 \times \frac{0,87 \times 0,99}{0,87+0,99} = 0,93 \quad (7)$$

Dari hasil perhitungan diatas diketahui nilai *Accuracy* sebesar 93%, nilai *Precision* sebesar 99%, nilai *Recall* sebesar 87% dan *F- Measure* sebesar 93%.

4.5. Aplikasi Hasil Penelitian

Hasil Penelitian ini berupa model klasifikasi analisis sentimen rancangan undang-undang tindak pidana kekerasan seksual/undang-undang tindak pidana kekerasan seksual di *twitter* dengan menggunakan algoritma *Naïve Bayes Classifier* (NBC) dan *Term Frequency – Inverse Document Frequency* (TF – IDF) untuk dilihat performa atau akurasi dari algoritma tersebut. Performa suatu model atau algoritma klasifikasi didapatkan dengan

pengujian performa menggunakan metode *confusion matrix*. Model ini diharapkan dapat membantu untuk mengklasifikasikan atau menganalisis sentimen terhadap topik yang berkaitan dengan rancangan undang-undang tindak pidana kekerasan seksual/undang-undang tindak pidana kekerasan seksual.

5. Kesimpulan dan Saran

5.1. Kesimpulan

Analisis sentimen pada opini mengenai Rancangan Undang-Undang Tindak Pidana Kekerasan Seksual/Undang-Undang Tindak Pidana Kekerasan Seksual pengguna *twitter* dengan kata kunci “SahkanRUUTPKS”, “TolakRUUTPKS”, “GagalkanRUUTPKS”, “KekerasanSeksual” dan “UUTPKS” didapatkan sebanyak 12.807 data, dengan sentimen positif sebanyak 10.263 data (80,1%) dan sentimen negatif sebanyak 2.544 data (19,9%). Sentimen masyarakat Indonesia terhadap Rancangan Undang-Undang Tindak Pidana Kekerasan Seksual/Undang-Undang Tindak Pidana Kekerasan Seksual di *twitter* mengarah kepada sentimen negatif. Hasil evaluasi model klasifikasi menggunakan algoritma *Naive Bayes Classifier* (NBC) berdasarkan pengujian *confusion matrix* diketahui nilai *accuracy* sebesar 93%, nilai *precision* sebesar 99%, nilai *recall* sebesar 87% dan *F-Measure* sebesar 93%. Dari nilai pengujian yang dihasilkan dapat disimpulkan model klasifikasi dengan algoritma *Naive Bayes Classifier* (NBC) memiliki kinerja yang sangat baik.

5.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, berikut ini adalah saran yang dapat dijadikan bahan pertimbangan untuk penelitian-penelitian selanjutnya:

- Melakukan penambahan kata pada kamus normalisasi agar data semakin bersih mengingat gaya penulisan *tweet* pengguna Indonesia banyak menggunakan kata tidak baku/gaul.
- Menggunakan algoritma berbeda untuk penelitian selanjutnya sebagai metode pembandingan performa atau akurasi algoritma untuk menemukan algoritma yang sesuai dengan topik analisis.

Daftar Pustaka:

- Ali, L., Khan, S. U., Golilarz, N. A., Yakubu, I., Qasim, q., Noor, A., & Nour, R. (2019). A Feature-Driven Decision Support System for Heart Failure Prediction Based on X2 Statistical Model and Gaussian Naive Bayes. *Hindawi Computational and Mathematical Methods in Medicine*, 1-8.
- Brogan, C. (2010). *Social Media 101: Tactics and Tips to Develop Your Business Online*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Bujang, S. D., Selamat, A., Ibrahim, R., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. *IEEE Access*, vol. 9, 95608-95621.
- Gunawan, T. S., Ashraf, A., Riza, B. S., Haryanto, E. V., Rosnelly, R., Kartiwi, M., & Janin, Z. (2020). Development of video-based emotion recognition using deep learning with Google Colab. *TELKOMNIKA Telecommunication, Computing, Electronics and Control Vol. 18 No. 5*, 2463-2471.
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: concepts and Techniques (3rd ed)*. Massachusetts: Morgan Kauffman.
- Harijianto, S. D. (2019). Analisis Sentimen Pada Twitter Menggunakan Multinomial Naive Bayes. *J. Phys. A Math. Theor Vol. 8 No. 5*, 1-104.
- Indonesia, R. (2022). *Undang-Undang Tindak Pidana kekerasan Seksual*. Jakarta: Indonesia.
- Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *Jurnal Sains dan Seni ITS Vol. 6, No. 2*, A51-A56.
- Lesmeister, C. (2015). *Mastering Machine Learning with R*. Birmingham: Packt Publishing.
- Mulyani, S. (2022). *Penggunaan Alter Account (Akun Anonim) Sebagai Pemicu Kekerasan Berbasis Gender Online (KBGO) di Media Sosial Twitter (Studi Deskriptif Kuantitatif Pada Korban Kekerasan Berbasis Gender Online (KBGO) Di Kota Bandung)*. Bandung: Universitas Pendidikan Indonesia.
- Muslehatin, W., Ibnu, M., & Mustakim. (2017). Penerapan Naive Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9* (hal. 250-256). Pekanbaru: Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau.
- Naraswati, N. P., Rosmilda, D. C., Desinta, D., Khairi, F., Damaiyanti, R., & Nooraeni, R. (2021). Analisis Sentimen Publik dari Twitter Tentang Kebijakan Penanganan Covid-19 di Indonesia dengan Naives Bayes Classification Vol. 10, No. 1. *Sistemasi: Jurnal Sistem Informasi*, 222-238.
- Nasional.kompas.com. (2019, September 24). *Ramai-ramai Turun ke Jalan, Apa yang Dituntut Mahasiswa?* Diambil kembali dari nasional.kompas.com:

- <https://nasional.kompas.com/read/2019/09/24/15440851/ramai-ramai-turun-ke-jalan-apa-yang-dituntut-mahasiswa>
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika Vol. 03, No. 01*, 50-59.
- Republik Indonesia, B. L. (2021). *Naskah Akademik Rancangan Undang-Undang Republik Indonesia Nomor... Tahun... Tentang Tindak Pidana Kekerasan Seksual*. Jakarta: Indonesia.
- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *INSYST: Journal of Intelligent System and Computation Vol. 1 No. 1*, 43-49.
- Sivi, N. A. (2018). *Analisis Sentimen Pada Media Sosial Menggunakan Metode Naive Bayes Classifier*. Bandar Lampung: Informatics & Business Institute Darmajaya.
- Syarif, R. D., Herdiani, A., & Astuti, W. (2019). Identifikasi Cyberbullying pada Komentar Instagram menggunakan Metode Lexicon-Based dan Naïve Bayes Classifier (Studi kasus: Pemilihan Presiden Indonesia Tahun 2019). *e-Proceeding of Engineering : Vol.6, No.2* (hal. 8838-8851). Bandung: Telkom University.
- Wongkar, M., & Angdresey, A. (2019, October). *Sentiment analysis using Naive bayes Algorithm of the data crawler: Twitter*. In 2019 Fourth International Conference on Informatics and Computing (ICIC) (pp. 1-5).