Teknik Informatika

HATE SPEECH DETECTION PADA VIDEO MENGGUNAKAN METODE *KNN* DAN *NAIVE BAYES*

Christopher Kelvin Pintoro Kwan*, Vincentius Riandaru Prasetyo, Fitri Dwi Kartikasari

Fakultas Teknik Universitas Surabaya, Raya Kalirungkut, Surabaya 60293

*Corresponding author: s160420021@student.ubaya.ac.id

Abstract—*Hate speech has had many negative impacts in Indonesia, such as riots, physical and verbal altercations, divisions in society, and many more. Social media is the place to spread hate speech most quickly. Not only through text posts, It is quite common to find hate speech in the form of videos. In this research, researchers will create a model that applies machine learning models to detect hate speech in videos, where currently most machine learning models are used to detect hate speech in text form only. In its application, the model will convert the input video into text using Google API. Then classification will be carried out using KNN to classify whether the video is hate speech or not, and Naive Bayes to classify the context of the video. In an unbalanced dataset, the classification results obtained for hate speech classification were 74% and for video context classification the accuracy was 45%. In a balanced dataset but overfitting occurs, the accuracy obtained in hate speech classification is 93% and in video context classification the accuracy is 55%. Based on the test results, it was found that the model used can have good accuracy if the dataset used is balanced between labels and there is no overfitting on the labels.*

Keywords: *hate speech, machine learning, knn, naive bayes*

Abstrak—*Hate speech* atau ujaran kebencian sudah memberikan banyak dampak yang negatif di Indonesia seperti kerusuhan, pertengkaran fisik maupun verbal, perpecahan di masyarakat, dan masih banyak lagi. Sosial media menjadi tempat untuk menyebarkan *hate speech* paling cepat. Tidak hanya melalui postingan teks, cukup sering juga ditemukan *hate speech* berbentuk video. Dalam penelitian ini, peneliti akan membuat model yang menerapkan model *machine learning* untuk mendeteksi adanya *hate speech* dalam video dimana saat ini kebanyakan model *machine learning* digunakan untuk mendeteksi *hate speech* dalam bentuk teks saja. Dalam penerapannya, model akan mengubah video yang diinput menjadi teks menggunakan *Google API*. Kemudian klasifikasi akan dilakukan menggunakan KNN untuk mengklasifikasikan apakah video *hate speech* atau bukan, dan *naive bayes* untuk mengklasifikasikan konteks dari video. Pada *dataset* yang tidak seimbang hasil klasifikasi yang didapatkan pada klasifikasi *hate speech* adalah 74% dan klasifikasi konteks video didapatkan akurasi sebesar 45%. Pada *dataset* yang seimbang namun terjadi *overfitting* akurasi yang didapatkan pada klasifikasi *hate speech* adalah 93% dan pada klasifikasi konteks video didapatkan akurasi 55%. Berdasarkan hasil uji coba didapatkan bahwa model yang digunakan dapat memiliki akurasi yang baik apabila *dataset* yang digunakan seimbang antar label dan tidak ada *overfitting* pada label.

Kata kunci*: hate speech, machine learning, knn, naive bayes*

Pendahuluan

Hate speech atau ujaran kebencian merupakan sebuah dampak negatif yang muncul dari berkembangnya teknologi terutama sosial media, dimana pengguna menjadi bebas dalam mengekspresikan diri mereka baik melalui unggahan video ataupun teks. Ujaran kebencian adalah tindakan komunikasi yang dilakukan oleh suatu individu atau kelompok dalam bentuk provokasi, hasutan, hinaan, kepada individu atau kelompok lain dalam hal berbagai aspek seperti ras, warna kulit, gender, cacat, orientasi seksual, kewarganegaraan, agama dan lain-lain (Zulkarnain, 2020). Penyalahgunaan teknologi informasi untuk menyebarkan ujaran kebencian sering dilakukan untuk tujuan pribadi, seperti menciptakan rasa permusuhan terhadap individu atau kelompok tertentu dalam bentuk SARA serta mengurangi tingkat keterpilihan seseorang dalam menduduki jabatan tertentu (Sepima et al., 2021). Akses internet atau media sosial mendorong berkembangnya homogenitas masyarakat yang membuat beberapa masyarakat yang tidak terbiasa dengan lingkungan seperti itu menyebabkan pertentangan dengan orang lain di media sosial (Kusuma, 2019).

Berdasarkan Kemp (2023), Jumlah pengguna internet di Indonesia per Januari 2023 tercatat mencapai 212,9 juta. Penetrasi internet di Indonesia saat ini mencapai 77 persen (212,9 juta jiwa)

sementara sisanya, yaitu sekitar 23 persen (63,51 juta jiwa), belum terhubung dengan jaringan internet. Menurut Kusuma (2019), penyebab tingginya tingkat penyebaran ujaran kebencian di media sosial adalah perkembangan teknologi yang sangat pesat tidak berbanding lurus dengan tingkat literasi digital masyarakat. Rendahnya literasi digital membuat masyarakat kesusahan dalam menyaring konten-konten positif karena terjadi banjir informasi yang menghasilkan interaksi negatif antar pengguna, sehingga mendorong perilaku tidak bertanggung jawab.

Pemerintah sendiri sudah melakukan upaya pencegahan untuk menanggulangi terjadinya ujaran kebencian dengan membuat peraturan yaitu Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik. Upaya pencegahan terjadinya ujaran kebencian dengan memberikan penyuluhan ataupun sosialisasi kepada masyarakat luas mengenai informasi dampak media elektronik jika tidak digunakan dengan bijak, etika menggunakan media sosial dengan memberikan pengetahuan hukum mengenai UU ITE (G. Ambar Wulan, 2021). Penyuluhan dapat membuat masyarakat mengerti akan dampak dari penyebaran ujaran kebencian. Akan tetapi, meski penyuluhan tentang UU ITE sering dilakukan hal tersebut tidak dapat mencegah masyarakat untuk menyebarkan ujaran kebencian sepenuhnya, mengingat masih rendahnya SDM di Indonesia yang membuat penyuluhan tidak efektif dan juga tidak merata ke semua masyarakat. Oleh karena itu, sistem dari teknologi itu sendiri yang dapat mencegah tersebarnya ujaran kebencian, dengan membuat *Hate Speech Detection System* yang akan mengidentifikasi apakah postingan yang dibuat oleh pengguna merupakan ujaran kebencian atau bukan sebelum pengguna mengunggahnya ke media sosial.

Hate Speech Detection pada *video* akan menggunakan analisis sentimen dalam mengidentifikasi *input video* yang akan di *input* oleh pengguna. Metode klasifikasi yang akan penulis gunakan dalam penelitian ini adalah metode klasifikasi *KNN* dan *Naive Bayes*. *KNN* akan digunakan untuk mengklasifikasikan konteks pembicaraan dalam video seperti individu, sekelompok orang, suku, ras, agama, dan antar-golongan (SARA), dan *Naive Bayes* akan digunakan untuk mengklasifikasikan apakah *video* tersebut adalah *hate speech* atau tidak. KNN adalah metode klasifikasi yang fleksibel baik untuk *binary classification* atau *multi-class classification.* Dalam klasifikasi ini KNN akan digunakan untuk melakukan klasifikasi *multi-class* dimana KNN mengklasifikasikan konteks dari video yang terdapat 6 *class*. Menurut Vincentius dan Hendrik (2021), KNN merupakan metode klasifikasi yang cocok untuk digunakan dalam masalah yang memiliki ukuran kelas 3 atau lebih dan ketika jumlah *dataset* yang digunakan meningkat, KNN dapat mengurangi *error* pada klasifikasi yang akan dilakukan. *Naive Bayes* akan digunakan sebagai metode klasifikasi *hate speech* atau bukan *hate speech* karena *Naive Bayes* merupakan metode yang cocok digunakan dalam klasifikasi data yang berbentuk teks khususnya *multinomial naive bayes* yang akan digunakan dalam penelitian ini. Utamanya, *Multinomial Naive Bayes* digunakan untuk mengkategorikan teks, contohnya seperti klasifikasi dokumen, filtrasi *spam,* dan analisis sentimen dimana dengan menggunakan *naive bayes* sebuah model klasifikasi dapat dengan cepat diciptakan dan melakukan prediksi dengan menggunakan model tersebut dengan cepat (Abbas et al, 2019). *Naive Bayes* adalah metode klasifikasi yang berbasis probabilitas sehingga memiliki model yang simpel oleh karena itu, *Naive Bayes* tidak membutuhkan sumber daya komputasi yang tinggi dan kecepatan klasifikasinya cenderung cepat. Dalam penelitian ini, penulis menggunakan *naive bayes* untuk melakukan klasifikasi *hate speech* dan *bukan hate speech* yang merupakan *binary classification* karena dalam penelitian yang dilakukan oleh Wu dan Bhandari, *multinomial naive bayes* memiliki akurasi yang lebih tinggi dalam *binary classification* daripada *multiclass classification.*

Metode Penelitian

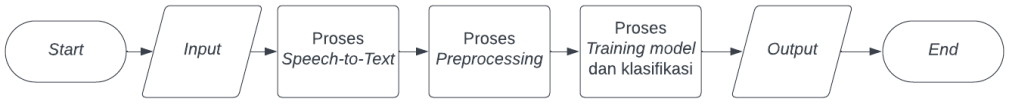
A. Dataset

Dataset yang akan digunakan dalam penelitian ini memiliki 3 fitur. Fitur yang pertama berisikan teks-teks dari video-video yang sudah pernah beredar di sosial media seperti *youtube*,

twitter, *instagram*, *facebook*, dan *tiktok.* Teks akan didapatkan secara manual dari *youtube*, *twitter*, *instagram*, *facebook*, dan *tiktok.* Video akan diunduh kemudian diekstrak menjadi teks menggunakan *speech-to-text.*

B. Proses

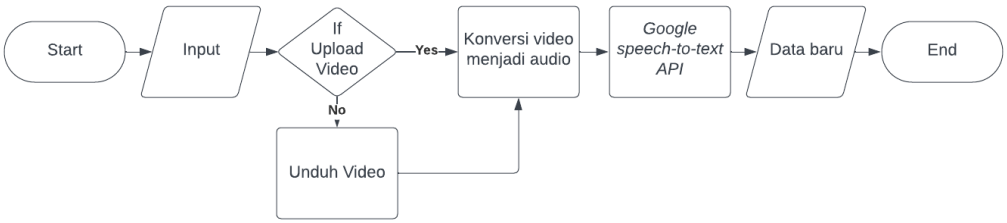
Proses yang akan terjadi dalam penelitian ini dapat terbagi menjadi 3 bagian yaitu proses *speech-to-text,* proses *preprocessing,* dan proses *training model* dan klasifikasi. Sistem akan melakukan proses setelah menerima input dari *user interface*. Setelah sistem menerima input, *user interface* akan mengirimkan input menuju proses. Proses yang akan terjadi adalah proses *speech-to-text*, kemudian proses *preprocessing,* dan kemudian proses *training model*. Setelah proses selesai sistem akan melakukan klasifikasi dan *output* akan dikirimkan ke *user interface* untuk ditampilkan. Alur dari sistem dapat terlihat pada gambar 1.1.



Gambar 1. *Flowchart* alur proses.

a) Proses *Speech-To-Text*

Sistem akan melakukan klasifikasi setelah pengguna meng*upload* video atau memasukkan *link youtube* dari video yang akan diklasifikasikan. Apabila *input* berupa *link youtube* maka sistem akan mengunduh video tersebut terlebih dahulu. Setelah sistem mendapatkan video tersebut maka, sistem akan melakukan konversi video yang telah diupload menjadi *file* audio dengan format *".wav".* Setelah sistem mengkonversi file video menjadi audio sistem akan melakukan *speech-to-text* menggunakan *Google speech-to-text API* dan dialog dalam file audio akan menjadi data baru. Alur proses *speech-to-text* seperti pada gambar 1.2.

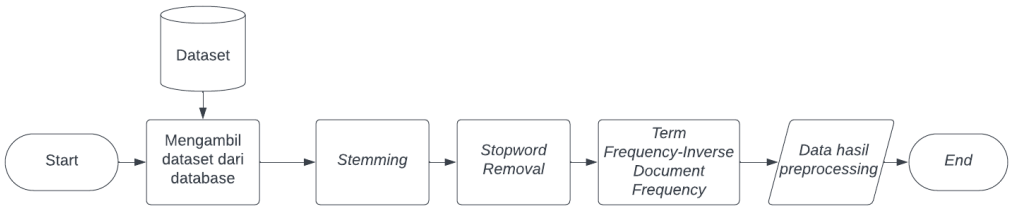


Gambar 2. Flowchart alur proses *speecht-to-text.*

b) Proses *Preprocessing*

Kemudian, sistem akan mengirimkan *query* ke *database* untuk mendapatkan *dataset*. *Query* akan dijalankan untuk setiap *array* untuk mengambil setiap kolom pada *dataset. Dataset* akan dibagi menjadi 3 variabel *array* yaitu *text*, *hate speech label*, dan konteks. Dimana *text* akan menjadi variabel bebas dan *hate speech label* dan konteks akan menjadi variabel terikat.

Setelah semua *dataset* didapatkan sistem akan melakukan data *pre-processing* pada dataset *text* dan pada data baru dengan melakukan *stemming* dan *stopword removal*. Kemudian sistem akan melakukan pembobotan pada dataset *text* dan juga data baru menggunakan *Term Frequency-Inverse Document Frequency (TF-IDF).* Pembobotan akan digunakan untuk mengubah kalimat menjadi angka berdasarkan frekuensi munculnya kata tersebut dalam dataset. Pembobotan ini akan digunakan oleh sistem untuk melakukan training pada model menggunakan dataset dan juga testing pada data baru. Alur proses *preprocessing* dapat terlihat pada gambar 1.3.



Gambar 3. Flowchart alur proses preprocessing.

c) Proses *Training Model dan Klasifikasi*

Setelah pembobotan pada *dataset* dan data baru, menggunakan hasil pembobotan *dataset text* dan juga *dataset hate speech label* sistem akan melakukan *train* pada model menggunakan metode *naive bayes* untuk mengklasifikasikan *hate speech* atau bukan. Sistem kemudian akan menentukan nilai K yang akan digunakan dalam metode *KNN* menggunakan rumus $k = \sqrt{jumlah\ dataset}$ apabila hasil yang didapatkan desimal maka sistem akan membulatkan hasil tersebut. Setelah itu sistem akan melakukan klasifikasi konteks menggunakan hasil pembobotan *dataset text* dan *dataset* konteks yang di *train* pada model menggunakan metode *KNN* dengan perhitungan distance *euclidean*. Sistem juga akan melakukan klasifikasi konteks hanya pada video yang diklasifikasikan sebagai *hate speech* dengan hanya menggunakan data yang memiliki label *hate speech* untuk menjadi percobaan apakah hasil yang dihasilkan lebih akurat atau tidak. Setelah sistem mendapatkan semua hasil klasifikasi maka *output* akan ditampilkan pada *user interface*. Alur proses *training model* terlihat pada gambar 1.4.

Gambar 4. Flowchart alur proses *training model* dan klasifikasi.

C. Evaluasi

Model akan membagi *dataset* menjadi 80% untuk *data training* dan 20% untuk *data testing.* Evaluasi model akan dilakukan dengan menggunakan *confusion matrix.* Dengan menggunakan *confusion matrix* evaluasi dapat dilakukan dengan menghitung akurasi, *precision*, dan *recall* dari model. Bentuk umum dari *confusion matrix* terdapat pada tabel 1.

Tabel 1
Bentuk Umum Confusion Matrix

		Nilai Sebenarnya	
		Positif	**Negatif**
Nilai Prediksi	Positif	*True Positive* (TP)	*False Positive* (FP)
	Negatif	*False Negative* (FN)	*True Negative* (TN)

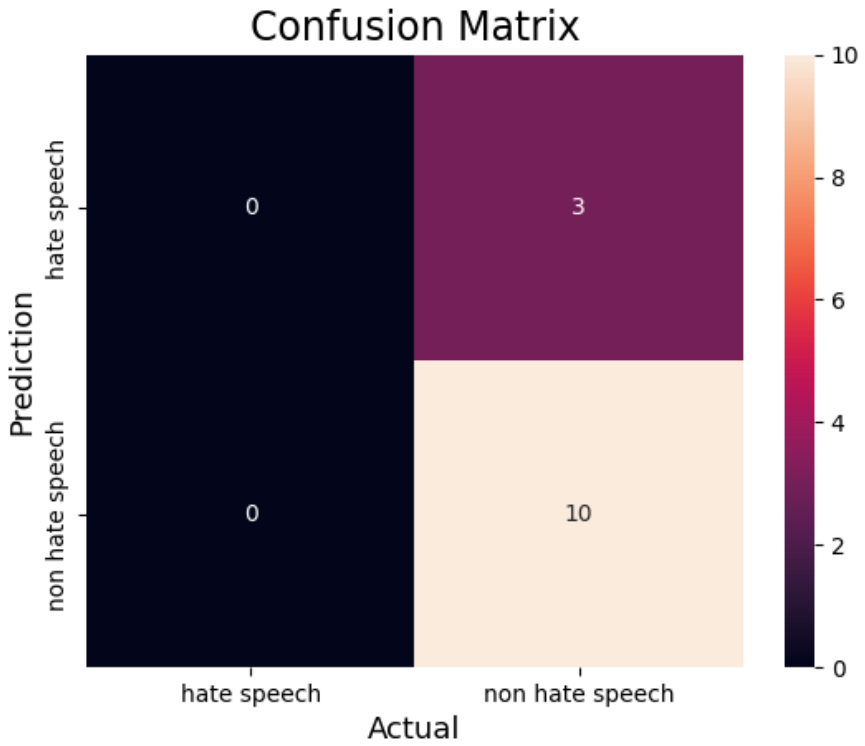
Hasil

Hasil penelitian akan diambil dari hasil evaluasi yang telah dilakukan. *Dataset* yang digunakan dalam uji coba berjumlah 62 data. Pada klasifikasi *hate speech,* perbandingan jumlah data berdasarkan label dapat terlihat pada tabel 2.

Tabel 2
Jumlah Data Berdasarkan Label Hate Speech

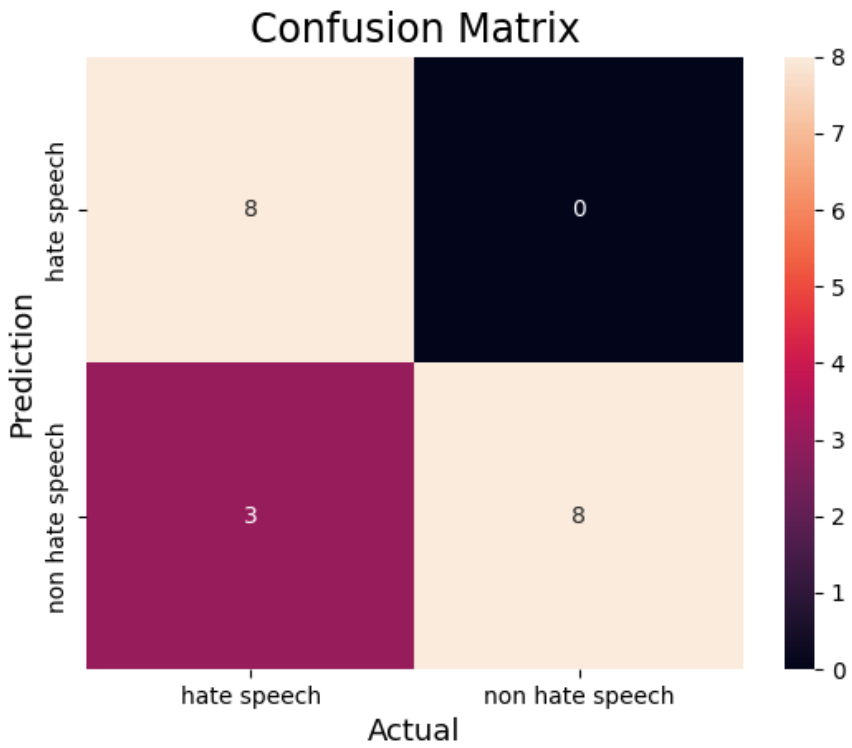
Label	Jumlah Data
Hate Speech	16
Non Hate Speech	46

Dapat terlihat perbedaan jumlah data yang berlabel *hate speech* dan *non hate speech.* Data *testing* yang digunakan berjumlah 13 data. Tingkat akurasi yang didapatkan dengan menggunakan klasifikasi *naive bayes* adalah 76%.



Gambar 5. Confusion matrix klasifikasi *hate speech.*

Dilakukan juga klasifikasi untuk label *hate speech* menggunakan *dataset* yang telah di *oversampling* sehingga data menjadi seimbang dan didapatkan peningkatan akurasi menjadi 86% dengan *confusion matrix* seperti pada gambar 6.



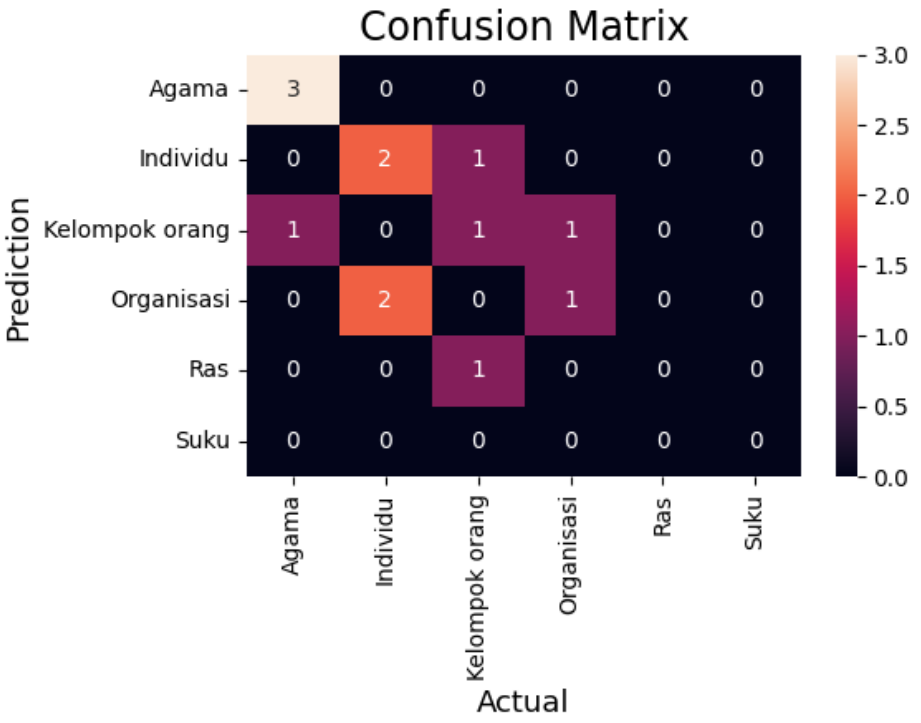
Gambar 6. Confusion matrix klasifikasi *hate speech* dengan *dataset oversampling.*

Pada klasifikasi konteks data pada *dataset* yang digunakan juga tidak seimbang. Perbandingan data pada *dataset* dapat terlihat pada tabel 3.

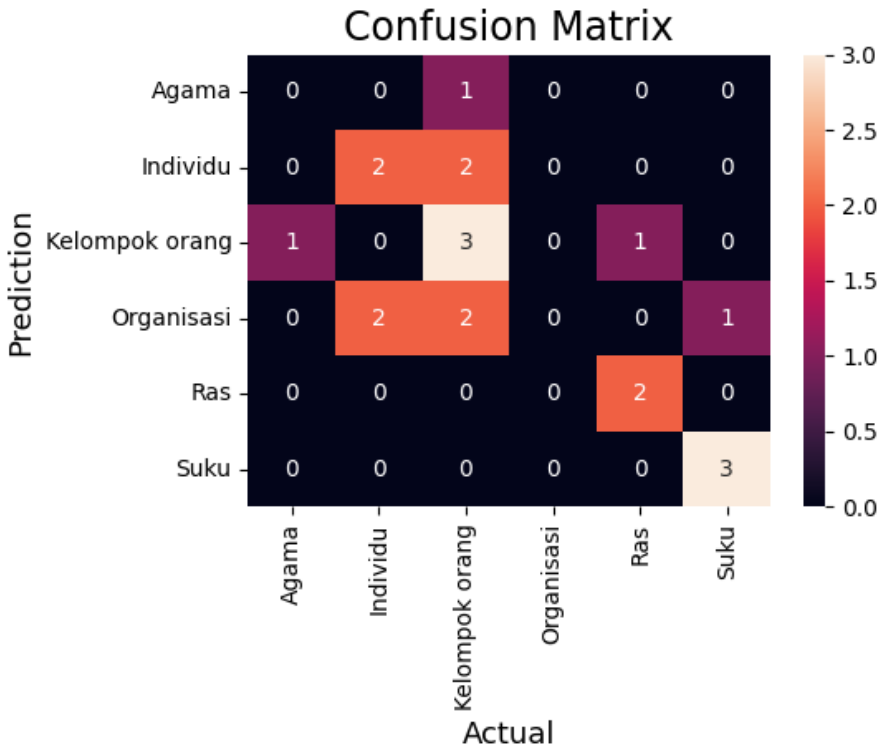
Tabel 3
Jumlah Data Berdasarkan Label Hate Speech

Label	Jumlah Data
Agama	13
Individu	15
Kelompok Orang	13
Organisasi	16
Ras	4
Suku	1

Hasil akurasi klasifikasi yang didapatkan adalah 52% pada *dataset* dengan keadaan normal. Ketika dilakukan *ovesampling* pada *dataset* didapatkan penurunan sebesar 41%. Hasil *confusion matrix* dari *dataset* normal dapat terlihat pada gambar 7 dan gambar 8.



Gambar 7. Confusion matrix klasifikasi konteks.



Gambar 8. Confusion matrix klasifikasi konteks dengan *dataset oversampling.*

Diskusi

Berdasarkan hasil penelitian yang dilakukan dapat terlihat bahwa pelabelan data pada *dataset* sangat mempengaruhi akurasi dari klasifikasi. Pada *dataset* yang digunakan oleh sistem menunjukkan adanya tanda-tanda *overfitting* pada *dataset* hal ini disebabkan oleh ketidakseimbangan data. Tingkat akurasi yang didapatkan dengan menggunakan klasifikasi *naive bayes* adalah 76%, meskipun cukup tinggi hal ini terjadi karena adanya *overfitting* data pada label *non hate speech*. Hal ini terbukti pada gambar 5, terdapat *confusion matrix* dari pengujian ini dimana data dengan label *hate speech* yang digunakan hanya 3 data dengan akurasi 0% yang menandakan bahwa sistem tidak dapat menghasilkan klasifikasi *hate speech* sedangkan *non hate speech* berjumlah 10 data dengan akurasi 100%.

Ketika dilakukan *oversampling dataset hate speech* hasil akurasi yang didapatkan menggunakan metode klasifikasi *naive bayes* meningkat menjadi 86%. Dibandingkan dengan sebelum sistem melakukan *oversampling,* dengan dilakukannya *oversampling* sistem dapat menghasilkan klasifikasi *hate speech*. Pada Gambar 6.10, terdapat *confusion matrix* dari percobaan ini dan dapat terlihat data yang digunakan untuk *testing* lebih seimbang pada 19 data *testing* dengan 11 data berlabel *hate speech* dan 8 data berlabel *non hate speech.* Akurasi yang didapatkan juga cukup seimbang apabila dibandingkan dengan *dataset* sebelum *oversampling,* yaitu 72% akurasi untuk data berlabel *hate speech* dan 100% akurasi untuk data berlabel *non hate speech.*

Pada klasifikasi konteks juga terlihat adanya *overfitting* pada *dataset,* dimana Perbedaan pada label agama, individu, kelompok orang, dan organisasi tidak terlalu banyak. Namun, apabila dibandingkan dengan jumlah data ras dan suku dapat terlihat perbedaan yang sangat banyak. Hal ini menyebabkan terjadinya *overfitting* pada 4 label mayoritas tersebut yang dapat terlihat pada *confusion matrix* pada gambar 7. Pada *confusion matrix* terlihat bahwa tidak ada ras dan suku yang masuk kedalam data *testing,* dan hasil klasifikasi dominan pada 4 label mayoritas tersebut. Akurasi yang didapatkan dari klasifikasi ini cukup tinggi yaitu 52% dari 13 data.

Akurasi yang didapatkan setelah dilakukan *oversampling* adalah berkurang menjadi 41%. Namun, terlihat perbedaan pada *confusion matrix* pada gambar 8, dimana lebih banyak hasil klasifikasi dengan label ras dan suku. Namun akurasi yang berkurang menandakan bahwa terjadinya *overfitting* yang membesar pada *dataset*. Berdasarkan hal tersebut didapatkan bahwa pada label *multiclass* apabila dilakukan *oversampling dataset, overfitting* pada data minoritas cukup tinggi yang membuat akurasi berkurang.

Penelitian juga dilakukan menggunakan *stratified k-fold cross validation* dimana akurasi rata-rata yang didapatkan oleh *naive bayes* pada *dataset* normal adalah 74% dan klasifikasi konteks yang dilakukan oleh *KNN* adalah 45%. Ketika dilakukan *oversampling* pada *dataset* didapatkan peningkatan drastis dari *naive bayes* dimana akurasi *naive bayes* meningkat menjadi 94%. Namun, *KNN* hanya meningkat menjadi 55%. Hal ini menunjukkan bahwa *naive bayes* hanya membutuhkan data yang seimbang agar mendapatkan hasil yang baik, dan metode ini tidak sensitif pada *overfitting* data. Sedangkan *KNN* sangat sensitif pada *dataset* yang digunakan. Dimana *KNN* membutuhkan data yang seimbang dan juga minim *overfitting* sehingga dapat memberika akurasi yang baik

Berbeda dengan penelitian-penelitian yang dilakukan sebelumnya seperti pada penelitian *Bangla Hate Speech Detection in Videos Using Machine Learning* yang berfokus pada perbandingan metode-metode *machine learning, dataset* yang digunakan cukup baik hingga mendapatkan akurasi lebih dari 90% pada semua metode. Terdapat juga Indonesia Hate Speech Detection Using Deep Learning dimana hasil dari klasifikasi yang tidak jauh berbeda. Penelitian ini juga membuktikan hasil dari penelitian *Detection of Hate Speech in Videos Using Machine Learning.* Dimana penelitian ini berfokus pada perbandingan akurasi pada *dataset binary class* dan *multi class.* Hasil penelitian ini menunjukkan bahwa klasifikasi *binary class* memiliki akurasi yang lebih baik daripada *multiclass.* Pada penelitian yang dilakukan penulis, hasil yang didapatkan juga sama dimana klasifikasi label *hate speech* memiliki akurasi yang lebih baik dari pada klasifikasi konteks. Dengan menggunakan metode *naive bayes* dilakukan klasifikasi *hate speech* dan konteks menggunakan *dataset*

oversampling, didapatkan akurasi dari klasifikasi *hate speech* sebesar 94% dan akurasi dari konteks sebesar 71%.

Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan penulis memiliki beberapa kesimpulan. Kesimpulan pertama adalah bahwa sistem yang dibuat dapat membantu pengguna mengidentifikasi video *hate speech* atau bukan. Kesimpulan kedua hasil klasifikasi *hate speech* sistem menggunakan *naive bayes* akan memiliki akurasi yang lebih baik apabila *dataset* yang digunakan memiliki data yang seimbang. Berdasarkan validasi klasifikasi sistem, akurasi dari *naive bayes* menggunakan *dataset* yang tidak seimbang adalah 74% dan apabila *dataset* seimbang menjadi 93%. Kesimpulan terakhir adalah hasil klasifikasi konteks sistem menggunakan *KNN* akan memiliki akurasi yang lebih baik apabila data yang dimiliki seimbang dan pelabelan yang baik sehingga terhindar dari *overfitting* yang menyebabkan metode *KNN* menjadi tidak akurat. Dimana berdasarkan hasil validasi klasifikasi sistem, didapatkan akurasi dari *KNN* ketika menggunakan *dataset* tidak seimbang dan terjadi *overfitting* adalah 45% dan ketika *dataset* seimbang dan terjadi sedikit *overfitting* didapatkan kenaikan akurasi menjadi 55%.

Daftar Referensi

Kemp,S. (2021,Februari 9). *Digital 2023: INDONESIA.* https://datareportal.com/reports/digital-2023-indonesia

G. Ambar Wulan, R. G. M. P. M. (2021). Pencegahan Kejahatan Ujaran Kebencian di Indonesia. *Jurnal Ilmu Kepolisian*, *14*(3), 19. https://doi.org/10.35879/jik.v14i3.278

Kusuma, R. A. (2019). Dampak Perkembangan Teknologi Informasi dan Komunikasi terhadap Perilaku Intoleransi dan Antisosial di Indonesia. *MAWA'IZH: JURNAL DAKWAH DAN PENGEMBANGAN SOSIAL KEMANUSIAAN*, *10*(2), 273–290. https://doi.org/10.32923/maw.v10i2.932

Lopez-Bernal, D., Balderas, D., Ponce, P., & Molina, A. (2021). Education 4.0: Teaching the basics of knn, lda and simple perceptron algorithms for binary classification problems. *Future Internet*, *13*(8). https://doi.org/10.3390/fi13080193

Ray, S. (2019). Introduction to Machine Learning and Different types of Machine Learning Algorithms. In *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*.

Sepima, A., Siregar, G., & Siregar, S. A. (2021). Penegakan Hukum Ujaran Kebencian di Republik Indonesia. In *Jurnal Retentum: Vol. Vol 2* (Issue 1 Februari).

Zulkarnain, Z. (2020). UJARAN KEBENCIAN (HATE SPEECH) DI MASYARAKAT DALAM KAJIAN TEOLOGI. *Studia Sosia Religia*, *3*(1). https://doi.org/10.51900/ssr.v3i1.7672

Hossain Junaid, M. I., Hossain, F., & Rahman, R. M. (2021). Bangla Hate Speech Detection in Videos Using Machine Learning. *2021 IEEE 12th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2021*, 347–351. https://doi.org/10.1109/UEMCON53757.2021.9666550

Sutejo, T. L., & Lestari, D. P. (2019). Indonesia Hate Speech Detection Using Deep Learning. *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*, 39–43. https://doi.org/10.1109/IALP.2018.8629154

Wu, C. S., & Bhandary, U. (2020). Detection of Hate Speech in Videos Using Machine Learning. *Proceedings - 2020 International Conference on Computational Science and Computational Intelligence, CSCI 2020*, 585–590. https://doi.org/10.1109/CSCI51800.2020.00104