DOI: 10.47002/seminastika.v3i1.232
3rd SEMINASTIKA 2021 *E-ISSN: 2655-0881*

KLASIFIKASI DAN ANALISIS DATASET KOMENTAR VIDEO YOUTUBE MENGGUNAKAN SUPERVISED LEARNING

CLASSIFICATION AND ANALYSIS OF YOUTUBE VIDEO COMENT DATASET USING SUPERVISED LEARNING

Finki Dona Marleny*[1], Ihdalhubbi Maulida[2], Bayu Nugraha [3]
[1,2]Informatika Fakultas Teknik Universitas Muhammadyah Banjarmasin
[3]Sistem Informasi Fakultas Sains & Teknologi Universitas Sari Mulia

*E-mail: *finkidona@gmail.com*

ABSTRAK

Interaksi di dalam sosial media dapat di lihat dari komentar-komentar sebagai umpan balik dari setiap kegiatan yang ada di media sosial, mulai dari status yang berupa teks, gambar maupun video. Dari berbagai respon pada kolom komentar tersebut diperoleh sebuah informasi dari data yang tidak terstruktur sehingga perlu adanya suatu teknik untuk mendefinisikan nilai informasi Fokus dalam penelitian ini adalah untuk memverifikasi kebenaran dan menggali nilai informasi yang terstruktur sehingga dapat menggambarkan kejadian dan topik yang terhubung dari komentar-komentar yang ada di dalam video youtube yang menjadi objek penelitian ini. Dari hasil pengujian di atas dapat dilihat nilai performa dari hasil pengujian menggunakan metode Naïve Bayes mendapatkan akurasi sebesar 57,50%, sedangkan dengan menggunakan metode KNN mendapatkan akurasi 88.06%..

***Kata kunci**:Youtube, Komentar, Klasifikasi, Video*

ABSTRACT

Interaction in social media can be seen from comments as feedback from every activity on social media, starting from status in the form of text, images and videos. From the various responses in the comments column, information is obtained from unstructured data so that there is a need for a technique to define the value of information. contained in the youtube video which is the object of this research. From the test results above, it can be seen that the performance value of the test results using the Naïve Bayes method gets an accuracy of 57.50%, while using the KNN method gets an accuracy of 88.06%.

***Keywords:** Youtube, Coment, Classification, Video*

PENDAHULUAN

Media sosial dan kebutuhan manusia modern saat ini tidak terlepas untuk saling berinteraksi, saling berbagi, sebagai fasilitas pembelajaran maupun hiburan. Interaksi di dalam sosial media dapat di lihat dari komentar-komentar yang tersedia di berbagai platform sebagai umpan balik dari setiap kegiatan yang ada di media sosial, mulai dari status yang berupa teks, gambar maupun video[1]. Dari komentar-komentar tersebut banyak terdapat opini yang terkadang positif maupun negatif, opini tersebut dapat dikaitkan dengan respon yang beragam, semakin banyak komentar maka semakin banyak pengguna yang berinteraksi, keadaan ini kadang menimbulkan interaksi yang memiliki respon tidak terduga. Dari berbagai respon pada kolom komentar tersebut diperoleh sebuah informasi dari data yang tidak terstruktur sehingga perlu adanya suatu teknik untuk mendefinisikan nilai informasi sehingga dapat lebih terstruktur. Dari data yang telah diestrak didapatkan suatu informasi tentang opini atau pendapat dari pengguna media sosial terhadap entitas tertentu[2].

Menganalisis pemantauan opini publik di Internet atau media sosial untuk pengumpulan informasi opini publik berupa

analisis pencarian, pengelompokan informasi, analisis ucapan, dan analisis prediksi tren yang selanjutnya memverifikasi kebenaran dari nilai suatu informasi[3]. Informasi yang telah diproses dapat di kelompokan dan menggunakan berbagai Teknik pemrosesan teks salah satunya dapat dengan fitur ekstrasi teknik cross validation terhadap model Naïve bayes[4].

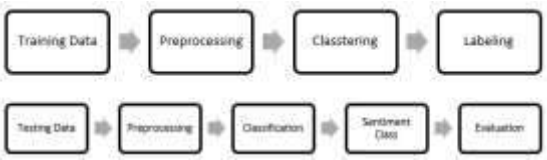
Dalam proses penggalian nilai suatu informasi yang terstruktur dan diproses dengan teknik tertentu agar pendapat yang memiliki nilai informasi tersebut bisa diolah dan dianalisa sehingga menghasilkan informasi yang bernilai dan dapat membantu banyak pihak dalam mengambil keputusan yang tepat yaitu dengan sentiment analisis yang merupakan studi komputasi mengenai pendapat, perilaku dan emosi seseorang terhadap entitas yang menggambarkan individu, kejadian atau topik[5].

Sedangkan untuk mengklasifikasikan kumpulan data komentar di video Youtube untuk analisis sentimen bisa menggunakan salah satu metode klasifikasi yaitu naive bayes[6]. Naive bayes merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat [7]. Metode sederhana ini juga merupakan salah satu metode kerja tercepat untuk analisis sentiment publik [8].

Fokus dalam penelitian ini adalah untuk memverifikasi kebenaran dan menggali nilai informasi yang terstruktur sehingga dapat menggambarkan kejadian dan topik yang terhubung dari komentar-komentar yang ada di dalam video youtube yang menjadi objek penelitian ini.

METODOLOGI

Klasifikasi dan analisis dataset komentar video youtube ini menggunakan tahapan Supervised Learning. Tahapan dalam metode yang digunakan adalah sebagai berikut:



Gambar 1. Tahapan Metode

Sumber Data

Data yang di gunakan adalah data dari akun Youtube:
https://www.youtube.com/channel/UC9lS7c1X0H-PILsaC0gem1g

Pada table 1 terdapat ID Video yang merupakan komentar video yang akan di analisis, kemudian ada jenis kelamin pengguna yang berkomentar menunjukan persentase komentar berdasarkan jeniskelamin ditiap video, selanjutnya usia pengguna yang mengomentari video tersebut.

Tabel 1. Sumber Dataset

ID Video	Pria	Wanita	Usia
7DhgTS4gWdc	40%	60%	24-34
VBJ2c4ptThw	48,2%	51,8%	24-34
ehJldQV3bfs	40%	60%	24-34
			35-44
BAzWeuO9fiE	48,1%	51,9 %	24-34

Preprocessing

Dalam Tahap Preprocessing Data dilakukan beberapa tahapan proses, yaitu:

1. Case folding

Berfungsi untuk menyeragamkan bentuk huruf menjadi huruf kecil. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital.

2. Tokenizing

Pada proses tokenization ini, semua kata yang ada di dalam tiap dokumen dipisahkan dan dihilangkan tanda bacanya, serta dihilangkan jika terdapat simbol atau apapun yang bukan huruf.

3. Stopword removal

Pada tahap ini, kata-kata yang tidak relevan akan dihapus, kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen.

4. Stemming

Stemming adalah proses pencarian kata dasar dengan menghilangkan imbuhan.

Naïve Bayes

Metode Naïve bayes dirumuskan sebagai berikut:

$$\boldsymbol{P(A|B) = \frac{P(B|A).(A)}{P(B)}} \qquad (1)$$

Data komentar yang terdapat pada metode ini akan direpresentasikan oleh pasangan atribut x1,x2,x3, ...xn. Dimana x1 adalah kata pertama dari dokumen, x2 adalah kata kedua dan seterusnya. Metode Naive Bayes akan mencari

probabilitas tertinggi untuk melakukan proses klasifikasi pada tweet yang akan diuji.

$$V_{MAP} = \underset{V_j \epsilon V}{\operatorname{argmax}} \prod_{n=1}^{n} P(x_i|V_j)\, P(V_j) \quad (2)$$

$$P(x_i|V_j) = \frac{n_k + 1}{n + |kosakata|} \quad (3)$$

K-Nearest Neighbor

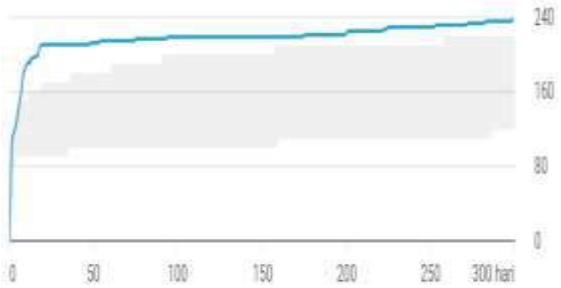
Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut[10]. Kedekatan didefinisikan dalam jarak metrik, seperti jarak Euclidean. Jarak Euclidean dapat dicari dengan menggunakan persamaan berikut ini:

$$Dxy = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (4)$$

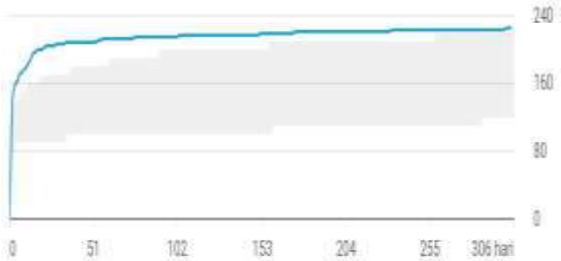
HASIL DAN PEMBAHASAN

Hasil Analisis komentar video dari peneltian ini dapat dibagi menjadi beberapa bagian pembahasan. Yaitu bagian proses dan hasil pengujian akurasi metode naïve bayes dan KNN.

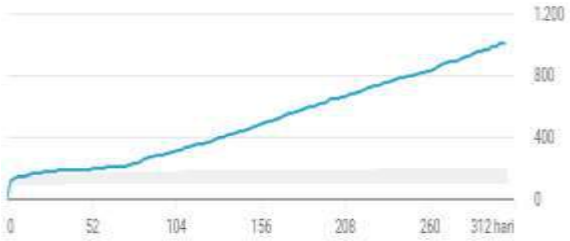
Berikut ini adalah gambaran dari retensi penonton video yang akan diteliti.



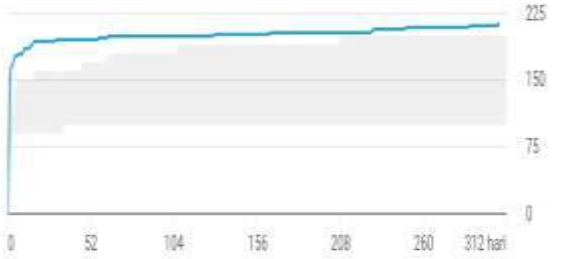
Gambar 2. Analisis Video ID : 7DhgTS4gWdc



Gambar 3. Analisis Video ID:VBJ2c4ptThw

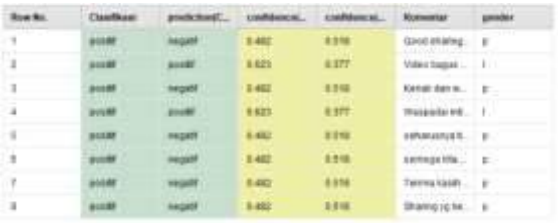


Gambar 4. Analisis Video ID : ehJldQV3bfs

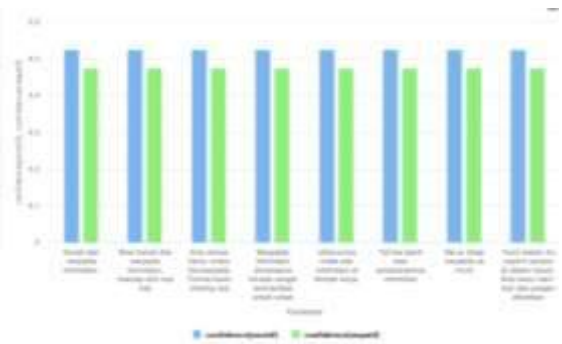


Gambar 5. Analisis VideoID: BAzWeuO9fiE

Grafik diatas menunjukan ringkasan jumlah penonton pada rentang waktu 312 hari dari awal unggahan video, video dengan ID: ehJldQV3bfs lebih banyak diminati.

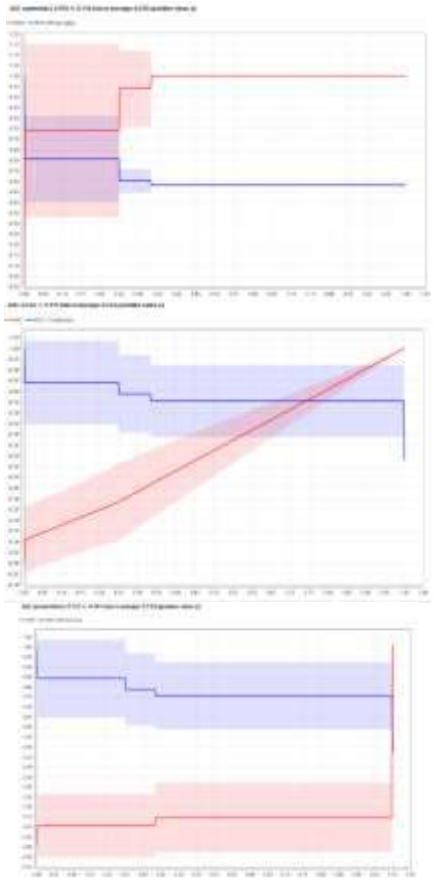


Gambar 6. Gambar klasifikasi prediksi

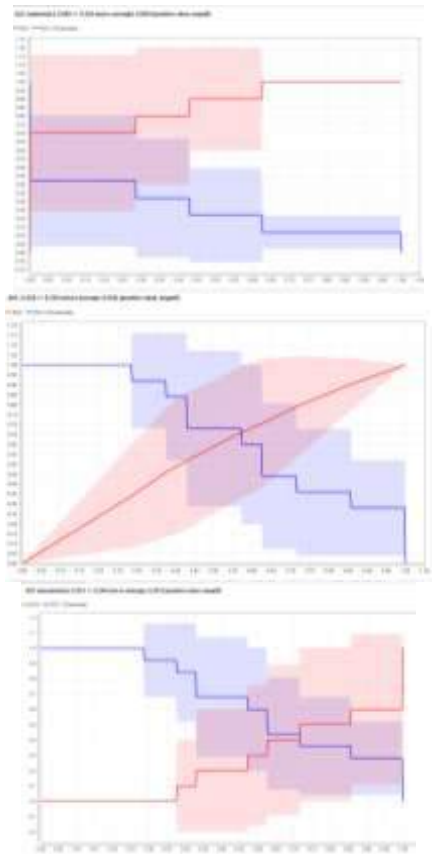


Gambar 7. Confidence Positif dan negative

grafik di atas menunjukan analisis sentiment positif dan negatif yang mana sudah mulai terlihat klasifikasi prediksi dari pengelompokan berdasarkan atribut komentar. Dari grafik diatas menunjukan confidence positif dan negatif, dari gambaran hasil diagram tersebut lebih dominan komentar positif.



Gambar 8. AUC OPTIMISTIC naïve bayes



Gambar 9. AUC OPTIMISTIC KNN

KESIMPULAN

Pengujian dilakukan dengan membagi dataset dengan ratio 80%:20%. Proses akan diuji menggunakan metode Naïve Bayes dan KNN. Hasil pengujian akan ditunjukan dengan akurasi, confusion matrix, recall, dan presisi. Akurasi yang didapatkan menghasilkan seberapa dekat nilai prediksi dengan nilai sebenarnya. Berdasarkan pengujian di atas dapat dilihat nilai performa atau akurasi dari pengujian menggunakan metode Naïve Bayes mendapatkan akurasi sebesar 57,50%, sedangkan dengan menggunakan metode KNN mendapatkan akurasi 88.06%.

Tabel 2. Hasil Perbandingan

Metode	Akurasi
Naïve Bayes	57.50%
KNN	88.06%

SARAN

Dalam pengujian yang dilakukan terdapat kesenjangan akurasi antara penggunaan metode KNN dan Naïve bayes, kesenjangan ini terjadi karena data latih yang di ujikan terlalu sedikit, sehingga hasil yang di proses kurang optimal, untuk mengoptimalkan hasil pengujian dapat menggunakan data komentar yang lebih banyak dan beragam sehingga pola data terlihat lebih terstruktur dan lebioh baik lagi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

[1] Marleny, F. (2020). ANALISIS PENGGUNA WHATSAPP TERHADAP KESALAHAN MENGIRIM PESAN TEKS MENGGUNAKAN METODE KLASIFIKASI. *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, *5*(1), 19-24.

[2] Zy, A. T., & Nugroho, A. 2018. Comparison Algorithm Classification *Naive Bayes*, Decision Tree, and Neural Network for Analysis Sentiment. International Conference on

Economic, Business, and Accounting, 1(c), 115–115.
X. Hu and C. Lu, "Research on Key Technologies of Internet Public Opinion

[3] Monitoring and Analysis System," *2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 2018, pp. 165-169, doi: 10.1109/ITOEC.2018.8740536
Yasah, K. S. (2020). Analisa Sentimen Tweet Indonesia Menggunakan Fitur

[4] Ekstrasi Dan Teknik Cross Validation Terhadap Model Naïve Bayes. *Jurnal SIGMA*, *10*(4), 189-194.
Medhat, Walaa, Hassan, Ahmed, & Korashy, Hoda, 2014, Sentiment Analysis

[5] Algorithms And Applications: A Survey, Ain Shams Engineering Journal (2014) 5, 1093–1113

Feldman, R and Sanger, J. 2007. The Text Mining Handbook: Advanced

[6] Feldman, R and Sanger, J. 2007. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press:NewYork

[7] Byna, A., & Basit, M. (2020). Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, *9*(3), 407-411.

[8] Susanto, A., Maula, M. A. I., Mulyono, I. U. W., & Sarker, M. K. (2021). Sentiment Analysis on Indonesia Twitter Data Using Naïve Bayes and K-Means Method. *Journal of Applied Intelligent System*, *6*(1), 40-45.

[9] Wu, J. (2012). *Advances in K-means clustering: a data mining thinking*. Springer Science & Business Media.

[10] Shi, K., Li, L., Liu, H., He, J., Zhang, N., & Song, W. (2011, September). An improved KNN text classification algorithm based on density. In *2011 IEEE International Conference on Cloud Computing and Intelligence Systems* (pp. 113-117). IEEE.