

## PERBANDINGAN ALGORITMA DALAM ANALISA SENTIMEN KRISIS EVERGRANDE PADA KANAL BERITA YOUTUBE

Henny Leidiyana<sup>1</sup>

Fakultas Teknik Informatika, Prodi Sistem Informasi, Universitas Bina Sarana Informatika

Email: <sup>1</sup>henny.hnl@bsi.ac.id

### Abstrak

Selama beberapa tahun terakhir, YouTube telah menjadi video online besar dengan jutaan pengguna. Banyak dari video ini berisi komentar pengguna dalam jumlah yang signifikan dan menarik untuk dilakukan kegiatan mengekstrak tren dari komentar tersebut. Dalam penelitian ini dilakukan analisa sentimen dari komentar pada berita melalui kanal YouTube terkait topik perusahaan swasta terbesar di Cina dan salah satu pengembang real estate yang mengalami krisis keuangan sehingga gagal membayar hutang. Dengan menggunakan teknik pembelajaran mesin, analisa sentimen yang telah dilakukan meliputi tahap persiapan data, ekstraksi fitur menggunakan teknik Count Vectorizer kemudian pemodelan dengan algoritma SVM, Multinomial Naive Bayes, Decision Tree, dan Logistic regression. Hasil terbaik diperoleh pada penerapan algoritma SVM dan Multinomial Naive Bayes dengan akurasi 84%, algoritma Logistic Regression 81% dan terendah pada algoritma Decision Tree dengan akurasi 73%.

**Kata Kunci:** *Evergrande, Count Vectorizer, YouTube*

### PENDAHULUAN

Evergrande adalah salah satu perusahaan swasta terbesar di Cina dan salah satu pengembang real estate terkemuka. Evergrande mengambil pinjaman untuk memenuhi permintaan real estate di kota-kota besar seperti Beijing, Shanghai, Shenzhen, dan Guangzhou. Namun kemudian terjadi penurunan harga properti di kota-kota kecil ditambah pada bulan Agustus 2021 pemerintah Beijing menerapkan aturan ketat atas industri real estate yang bertujuan untuk mengekang utang dan membuat sektor real estate lebih terjangkau bagi warga China, akibatnya Evergrande dikabarkan tidak mampu lagi melakukan pembayaran bunga pinjamannya. (Iswara, 2021)

Hal ini menjadi topik hangat dalam pemberitaan pada youtube dan media sosial lainnya. Netizen ramai memberikan komentar tentang peristiwa ini. Berdasarkan fenomena ini penulis tertarik untuk melakukan pengamatan terhadap data dari media

sosial youtube untuk analisa sentiment yang saat ini tengah menjadi tema penelitian yang populer.

Berbagai teknik analisa sentimen telah banyak dilakukan, dan penerapan pada bermacam-macam domain yang berbeda sudah banyak dilakukan seperti dalam bidang bisnis untuk review produk (Rasool et al., 2019)(Lutfi et al., 2018), review penjualan (Lutfi et al., 2018), analisa pelayanan (Chinnam et al., 2020), bisnis perhotelan (Farisi et al., 2019), pendidikan (Sigmawaty & Adriani, 2019), kesehatan (Trisnawati et al., 2021), kebijakan pemerintah (Laurenz & Eko Sedyono, 2021), hiburan (Abbas et al., 2019) menggunakan data yang bersumber dari media sosial seperti youtube (Baravkar et al., 2021), twitter (Suppala & Rao, 2019), facebook (Rianto & Pratama, 2021), Instagram (Naf'an et al., 2019). Teknik analisa sentiment yang cukup banyak digunakan seperti Naive Bayes SVM, Decision Tree, Random Forest (Thomas et al., 2021). Penelitian-

penelitian tersebut banyak yang melakukan komparasi. Misalkan pada penelitian yang dilakukan oleh Aravind Veluchamy yaitu perbandingan model klasifikasi sentimen teks menggunakan vektorisasi TF-IDF dengan hasil tertinggi pada model logistic regression dengan akurasi 90%.

Penelitian ini bertujuan untuk melakukan perbandingan metode mesin pembelajaran untuk analisa sentiment terhadap berita krisis Evergrande menggunakan data komentar pada video youtube kanal berita luar negeri berbahasa Inggris. Dataset komentar diambil menggunakan pemrograman python kemudian dilabeli dengan VADER. Setelah itu dilakukan preprocessing terhadap dataset berlabel barulah pembuatan model. Algoritma yang digunakan dalam penelitian ini yaitu SVM, Linear regression, Decsion tree, dan Multinomial Naïve Bayes. Terakhir yaitu evaluasi terhadap model, dilakukan dengan mengukur kinerja menggunakan akurasi.

#### Algoritma Klasifikasi

1. Support Vector Machine.  
SVM adalah metode klasifikasi yang sangat akurat. SVM sebuah metode untuk klasifikasi data linier dan nonlinier SVM menggunakan pemetaan nonlinier untuk mengubah data training ke dimensi yang lebih tinggi. Dalam dimensi baru inilah dilakukan pencarian hyperplane pemisah yang optimal linier (batas yang yang memisahkan komentar positif, negative, dan netral). SVM menemukan hyperplane menggunakan support vector (Han, 2011).
2. Linear Regression  
LR digunakan untuk menghubungkan satu variabel dependen kategoris dengan satu atau lebih variabel independen. LR mencoba untuk menemukan hyper-plane yang memaksimalkan kesenjangan pemisahan antara kelas.(Ankit & Saleena, 2018)
3. Decision Tree  
Decision tree adalah pembelajaran pohon keputusan dari pelatihan berlabel kelas. Pohon keputusan adalah struktur pohon seperti diagram alur, di mana setiap simpul internal (nonleaf node) menunjukkan tes pada atribut, setiap cabang mewakili hasil dari tes, dan setiap

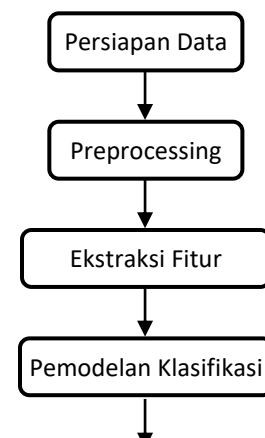
simpul daun (atau simpul terminal) memegang label kelas.(Han, 2011)

4. Multinomial Naive Bayes  
Multinomial NB adalah pengklasifikasi teks yang digunakan dalam masalah pemrosesan bahasa alami (NLP). Naive Bayes memprediksi tag teks dengan menghitung probabilitas setiap tag untuk teks tertentu dan kemudian menampilkan tag yang tertinggi. Pengklasifikasi Multinomial NB adalah jenis pengklasifikasi NB dan sering digunakan sebagai dasar untuk klasifikasi teks (Abbas et al., 2019)

#### METODE

Salah satu tugas pemrosesan bahasa alami yang banyak digunakan dalam berbagai masalah bisnis adalah Klasifikasi Teks. Tujuan klasifikasi teks adalah untuk secara otomatis mengklasifikasikan dokumen teks ke dalam satu atau lebih kategori yang ditentukan. Klasifikasi Teks adalah contoh tugas pembelajaran mesin yang diawasi (*supervised learning*) karena kumpulan data berlabel yang berisi dokumen teks dan labelnya digunakan untuk melatih pengklasifikasi. Salah satu contoh klasifikasi teks adalah analisa sentimen dari media sosial. Langkah yang dilakukan terdiri dari tahapan:

1. Persiapan data: Data komentar yang diperoleh dilabeli dengan metode VADER
2. Preprocessing: Pada langkah ini dilakukan penghapusan karakter khusus, lemmatisasi, menghapus stopword, dan tokenisasi
3. Ekstraksi Fitur: Langkah selanjutnya adalah ekstraksi Fitur di mana dataset komentar yang telah dilabeli diubah menjadi fitur datar agar dapat dibuat model pembelajaran mesin.
4. Pembuatan Model: Pada langkah ini model pembelajaran mesin dilatih pada kumpulan data berlabel.
5. Evaluasi: Pada langkah ini mengukur kinerja model klasifikasi menggunakan akurasi



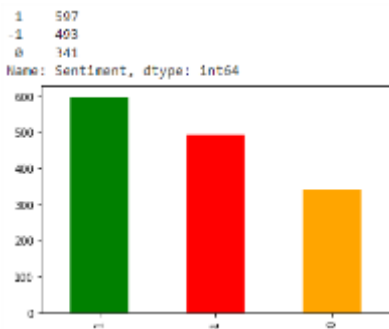
**HASIL DAN PEMBAHASAN**

Dalam analisa sentimen di sini menggunakan pemrograman python. Dataset yang digunakan sebanyak 1431 komentar video youtube kanal berita berbahasa inggris dengan topik krisis evergrande. Komentar yang diperoleh dilabeli dengan metode VADER lalu disimpan dengan format csv.

Komentar	Sentiment Scores	Sentiment
people doole right almost good government mea...	0.6926	1
market stabilizes here bulls will likely come ...	-0.6486	-1
china could easily call international debts of...	-0.0772	-1
this would phone buying puts whilst saying not...	0.4266	1
china oceanwide development francisco seized t...	-0.3612	-1

Gambar 1. Sampel Dataset Berlabel

Komentar diberi label '1' untuk menunjukkan komentar positif, '-1' untuk menunjukkan komentar negatif dan label '0' menunjukkan komentar yang netral. Gambar 1 menunjukkan sampel dataset yang sudah diberi label. Gambar 2 adalah grafik yang menunjukkan proporsi kelas setelah pelabelan dengan VADER.



Gambar 2. Proporsi Kelas setelah Pelabelan dengan VADER

Dataset dimuat untuk dilakukan proses ekstraksi fitur. Dataset tersebut kemudian dipecah menjadi

data latihan dan data uji. Pembuatan model nantinya bertujuan memprediksi label pada dataset pengujian. Berikutnya adalah melakukan preprocessing terhadap data berlabel. Tahapan yang dilakukan yaitu:

1. Menghapus karakter khusus dan karakter beraksen. Karakter yang dikenal sebagai tanda aksen dapat menyebabkan masalah dalam langkah-langkah preprocessing seperti lemmatisasi. Oleh karena itu, penghapusan karakter aksen ini sebagai upaya pencegahan. Pada langkah ini juga dilakukan penghapusan karakter khusus dan emoji.
2. Lemmatisasi. Lemmatisasi adalah proses mendapatkan akar morfologis dari kata. Dalam banyak kasus, lemmatisasi memungkinkan mesin untuk mengenali tense yang berbeda dari kata yang sama.
3. Menghapus stopword. Stopwords adalah kata-kata yang memiliki sedikit atau tidak ada artinya, seperti kata penunjuk orang.
4. Tokenisasi. Tokenisasi adalah proses memisahkan kata-kata dari kalimat menjadi unit individu yang akan digunakan untuk ekstraksi fitur.

Langkah selanjutnya adalah ekstraksi fitur di mana dataset komentar yang telah diberi label diubah menjadi fitur datar (*flat feature*) agar dapat dibuat model pembelajaran mesin. Pada langkah ini, data diubah menjadi vektor fitur dan fitur baru akan dibuat menggunakan dataset yang ada. Teknik yang digunakan yaitu count vectorizer. Count vectorizer adalah notasi matriks dari dataset di mana setiap baris mewakili dokumen dari korpus, setiap kolom mewakili istilah dari korpus, dan setiap sel mewakili jumlah frekuensi istilah tertentu dalam dokumen tertentu.

Setelah dilakukan ekstraksi fitur, dataset siap untuk dilatih dengan algoritma machine learning. Data dibagi secara random ke dalam data latih dan data uji untuk mendapatkan akurasi. Pemodelan dengan SVM, logistic regression, multinomial naïve bayes, dan decision tree. Penerapan SVM menggunakan fitur yang diekstrak dengan count vector menghasilkan akurasi 84%, logistic regression 81%, decision tree 73%, dan multinomial NB 84%. SVM

memiliki akurasi yang sama dengan multinomial NB. Decision tree menghasilkan performansi terendah. Ringkasan kinerja dari setiap pengklasifikasi dapat dilihat pada Tabel 1.

Tabel 1. Pebandingan Kinerja Pengklasifikasi

	Accuracy (%)
SVM	84
Linear classifier	81
Decision Tree	73
Multinomial NB	84

**PENUTUP**

**Simpulan**

Analisa sentimen yang telah dilakukan meliputi tahap persiapan data yang diperoleh melalui video berita kanal YouTube yang kemudian dilakukan pelabelan dengan teknik VADER. Langkah selanjutnya adalah ekstraksi fitur terhadap data berlabel menggunakan teknik Count Vectorizer yang hasilnya digunakan sebagai dataset untuk pembuatan model. Algoritma yang digunakan dalam pembuatan model yaitu SVM, Logistic Regression, Decision Tree, dan Multinomial Naive Bayes. Akurasi model SVM dan Multinomial NB sama yaitu 84%, LR 81%, dan yang terkecil decision tree 73%.

**Saran**

Penelitian lebih lanjut untuk peningkatkan dapat dilakukan beberapa cara antara lain menggunakan ekstraksi fitur lainnya seperti *word embedding* atau *bag of words*. Kemudian dalam pembuatan model dapat dilakukan penerapan *deep learning*, dan bisa juga dilakukan *hyperparameter tuning* dalam pemodelan. Penelitian juga dapat dikembangkan dengan menghubungkan antara sentiment krisis evergrande dengan harga saham.

**DAFTAR PUSTAKA**

Abbas, M., Ali, K., Jamali, A., Ali Memon, K., & Aleem Jamali, A. (2019). Multinomial Naive Bayes Classification Model for Sentiment Analysis. *IJCSNS International Journal of Computer Science and Network Security*, 19(3), 62. <https://www.researchgate.net/publication/334451164>

Ankit, & Saleena, N. (2018). An Ensemble Classification System for Twitter Sentiment Analysis. *Procedia Computer Science*, 132(Iccids), 937–946. <https://doi.org/10.1016/j.procs.2018.05.109>

Baravkar, A., Jaiswal, R., Chhoriya, J., & Tekwani, P. B. (2021). Sentimental Analysis of YouTube Videos. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 554–562. <https://doi.org/10.32628/cseit2172112>

Chinnam, S. K., Reshmi Khadherbhi, S., Sandhya Krishna, P., & Anveshini, D. (2020). Sentiment analysis in services provided by telecommunications. *International Journal of Advanced Science and Technology*, 29(3), 9167–9176.

Farisi, A. A., Sibaroni, Y., & Faraby, S. Al. (2019). Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier. *Journal of Physics: Conference Series*, 1192(1). <https://doi.org/10.1088/1742-6596/1192/1/012024>

Han, J. W. (2011). *Data Mini Concepts and Techniques*. Morgan Kaufman.

Iswara, A. J. (2021). *Krisis Evergrande: Awal Mula Petaka, Utang Rp 4 Kuadriliun, dan Ruginya Ribuan Orangle*. Kompas.Com. <https://www.kompas.com/global/read/2021/10/02/203602470/krisis-evergrande-awal-mula-petaka-utang-rp-4-kuadriliun-dan-ruginya?page=all>

Laurensz, B., & Eko Sedyono. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 10(2), 118–123. <https://doi.org/10.22146/jnteti.v10i2.1421>

Lutfi, A. A., Permanasari, A. E., & Fauziati, S. (2018). Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine. *Journal of Information Systems Engineering and Business Intelligence*, 4(1), 57. <https://doi.org/10.20473/jisebi.4.1.57-64>

Naf'an, M. Z., Bimantara, A. A., Larasati, A., Risondang, E. M., & Nugraha, N. A. S. (2019).

- Sentiment Analysis of Cyberbullying on Instagram User Comments. *Journal of Data Science and Its Applications*, 2(1), 88–98.  
<https://doi.org/10.21108/jdsa.2019.2.20>
- Rasool, A., Tao, R., Marjan, K., & Naveed, T. (2019). Twitter Sentiment Analysis: A Case Study for Apparel Brands. *Journal of Physics: Conference Series*, 1176(2).  
<https://doi.org/10.1088/1742-6596/1176/2/022015>
- Rianto, A., & Pratama, A. R. (2021). Sentiment Analysis of Covid-19 Vaccination Posts on Facebook in Indonesia With Crowdtangle. *Jurnal Riset Informatika*, 3(4), 353–362.  
<https://doi.org/10.34288/jri.v3i4.282>
- Sigmawaty, D., & Adriani, M. (2019). Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information) 12/2 (2019), 91-102. DOI: <http://dx.doi.org/10.21609/jiki.v12i2.745>. *Jurnal Ilmu Komputer Dan Informasi (Journal of ...)*, 2, 75–84.  
<https://core.ac.uk/download/pdf/296598902.pdf>
- Suppala, K., & Rao, N. (2019). Sentiment analysis using naïve bayes classifier. *International Journal of Innovative Technology and Exploring Engineering*, 8(8), 264–269.  
<https://doi.org/10.14445/22312803/ijett-v68i4p141>
- Thomas, S., Yuliana, & Noviyanti, P. (2021). Study Analisis Metode Analisis Sentimen pada YouTube. *Journal of Information Technology*, 1(1), 1–7.  
<https://doi.org/10.46229/jifotech.v1i1.201>
- Trisnawati, S., Lampung, B., Hermanto, R., Lampung, B., Priantoro, A. U., Lampung, B., Sabita, H., Lampung, B., Fauzi, C., & Lampung, B. (2021). *Twitter Sentiment Analysis on The use of Sinovac Vaccine in Indonesia*. 77–80.