



Penerapan Data Mining Untuk Analisis Data Sosial Media Menggunakan Metode K-Nearest Neighbour (K-NN)

Della Natasya¹, Zaehol Fatah²

¹Program Studi Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy Situbondo

²Program Studi Sistem Informasi, Sains dan Teknologi, Universitas Ibrahimy Situbondo

ntsyahtdran14@gmail.com, zaeholfatah@gmail.com

Abstrak

Penerapan data mining pada analisis data sosial media telah menjadi alat penting dalam mengidentifikasi pola, tren, dan wawasan yang relevan dari data yang sangat besar dan beragam. Studi ini berfokus pada penggunaan teknik data mining untuk menganalisis data sosial media dengan tujuan memahami perilaku pengguna, sentimen publik, dan pengaruh tren sosial terhadap berbagai sektor industri. Metodologi yang digunakan meliputi pengumpulan data, pra-pemrosesan data, analisis setimen, dan clustering. Hasil penelitian ini menunjukkan bagaimana data mining dapat memberikan wawasan mendalam yang bermanfaat bagi perusahaan untuk strategi pemasaran, manajemen reputasi, dan pengambilan keputusan bisnis yang lebih baik. Penelitian ini juga menyoroti tantangan dan solusi dalam mengelola data sosial media yang besar dan beragam, serta pentingnya etika dalam pengolahan data.

Kata Kunci: Data Mining, Sosial Media, Metode K-Nearest Neighbour

PENDAHULUAN

Pada analisis penelitian ini untuk penerapan data mining dengan metode K-Nearest Neighbour (K-NN) akan mengupas secara rinci pembahasan tentang penerapan data mining untuk analisis data sosial media menggunakan metode K-Nearest Neighbour (KNN). Data mining adalah proses ekstraksi informasi dan pola yang berguna dari data besar. Ini juga disebut sebagai proses penemuan pengetahuan, penggalian dari beberapa pengetahuan yang didapat dari data, ekstraksi pengetahuan atau analisis data atau pola. Jadi dengan semikian itu bahwa Data Mining didefinisikan sebagai mengekstraksi informasi dari kumpulan data yang sangat besar. Dengan makna lain kita dapat menggali ilmu pengetahuan melalui data yang ada. Data Mining adalah proses ekstraksi informasi dan pola yang berguna dari data besar. Ini juga disebut dengan proses penemuan suatu ilmu pengetahuan melalui analisis data. (Lugianti et al., 2022)

Penulis memilih metode K-Nearest Neighbour (KNN) sebagai metode penelitian dalam studi kasus penerapan data mining pada analisis data sosial media yang sudah diambil. Algoritma KNN digunakan untuk mengklasifikasikan data dengan menganalisis preferensi pengguna dari peringkat yang diberikan. (*Final-MGFRTS+Mutiara+1027-1033*, n.d.) Dengan beberapa proses yang dilakukan dari pengumpulan data, data diproses, dan dibagi menjadi set pelatihan dan pengujian untuk mengoptimalkan model. Kelebihan lain dari metode KNN adalah terbukti efektif efektif dalam klasifikasi data, tetapi kualitas hasil sangat bergantung pada kualitas data yang dianalisis. (Arrohman & Fatah, 2024)

Teknologi sudah berkembang dan semakin memberi banyak pengaruh dan juga menularkan manfaat untuk kehidupan manusia. Pengaruh besar sosial media menumbuhkan popularitas media sosial yang meningkat di kalangan masyarakat luas. Media sosial lahir sebagai perpaduan antara arus komunikasi dan perkembangan teknologi, sehingga memberikan kemudahan bagi orang yang ingin menggunakan jejaring sosial atau hubungan sosial dengan seluruh komponen di seluruh dunia karena sosial media juga merupakan platform online yang memiliki bermacam kegunaan, minat, seperti aktivitas kelompok ataupun pribadi. (Nugraha et al., 2019)

Sosial media atau biasa disebut juga media sosial merupakan sebuah media untuk bersosialisasi satu dengan yang lain yang dilakukan secara online dengan memungkinkan manusia untuk saling berkomunikasi tanpa batas ruang dan waktu. Tidak sembarangan orang yang menggunakan sosial media, karenanya pun sosial media dilegalkan atau diizinkan oleh pemerintah sehingga menghasilkan Undang-undang Informasi dan Transaksi Elektronik (UU ITE) untuk membatasi public untuk memberikan kritik. (Muhamad Ayub & Sofia Farzanah Sulaeman, 2022)

Salah satu contoh dari media sosial yang banyak digunakan oleh masyarakat adalah telepon cerdas atau yang disebut *smartphone*. Benda ini sendiri dilengkapi berbagai fitur yang banyak diminati seperti *chatting*, *email*, *sms*, *mms*, *browsing* serta fasilitas sosial media lainnya. Dalam media sosial, ada tiga bentuk yang merujuk kepada makna bersosial yaitu pengenalan (*cognition*), komunikasi (*communicate*), dan kerjasama (*cooperation*). Jejaring sosial merupakan situs dimana setiap orang bisa membuat web page pribadi, kemudian terhubung dengan teman-teman untuk berbagi komunikasi dan informasi. Jejaring sosial terbesar antara lain, Facebook, Whatsapp, Instagram, Tik-tok, Twitter dan lainnya. (Mulyono, 2021)

Di dalam buku yang berjudul Fenomena Medsos oleh penulis Muthi' Ahmad, S.H dicontohkan macam-macam media sosial diantaranya adalah Facebook, Youtube, Twitter, Skype, Instagram, Line, dan masih banyak lagi. Adapun

beberapa dampak Media Sosial diantaranya adalah sebagai sarana untuk menjalin hubungan sesama manusia, memudahkan dalam menambah dan bertukar ide dan masih banyak lagi. (Muthi' Ahmad, n.d.)

Dampak positif dalam perkembangan media sosial antara lain yakni memudahkan masyarakat dalam hal komunikasi, mencari dan mengakses berbagai informasi dan data dengan mudah, memperluas relasi, menambah wawasan, dan masih banyak hal lainnya lagi. Adapun dampak negatif dari media sosial ini biasanya dilakukan oleh opnum yang tidak bisa memanfaatkan media sosial dengan baik atau menyalah gunakan media sosial seperti, membuat orang kecanduan atau ketergantungan sehingga orang menjadi malas dan kurang memperhatikan lingkungan sekitar. Contoh yang bisa berdampak pada psikis juga seperti *buyling*, *tantrum*, dan lainnya. (Mulyono, 2021)

METODE

Tahapan Pengambilan Data

Data yang diambil untuk dianalisa pada jurnal ini bersumber dari kaggle mengenai data penggunaan media sosial yang pada data tersebut berisi jenis kelamin, pekerjaan, jenis platform, dan waktu penggunaan. Dan data tersebut akan diteliti untuk diklasifikasi dalam analisis kuantitas perempuan dan laki-laki dalam penggunaan media sosial dengan menggunakan aplikasi rapid miner melalui penerapan metode K-Nearest Neighbor (K-NN)

Klasifikasi

Klasifikasi adalah teknik dasar dalam data mining yang melibatkan pengkategorian *instance* atau sampel data ke dalam kelas atau kategori yang telah ditentukan berdasarkan fiturnya. Definisi lain dari klasifikasi adalah metode supervised learning yang mengategorikan populasi data yang besar. (Setiawan et al., 2023). “ Menurut penelitian yang dilakukan oleh Sidik et al (2020), ‘Klasifikasi ini menjadi lebih populer dikarenakan dapat menangani rentang data yang lebih luas dibandingkan dengan regresi’ (dikutip oleh Ahmad Yogiarto, Ahmad Homaidi, Zaehol Fatah, 2024).” (Yogiarto et al., 2024)

Media Sosial

Menurut KBBI, media adalah alat komunikasi, dalam konteks pembahasan ini yang dimaksud adalah media sebagai alat komunikasi. Sedangkan aspek sosial adalah permasalahan yang berkaitan dengan masyarakat atau kepentingan umum. Media sosial sendiri merupakan bagian penting dari perkembangan internet. Jejaring sosial dapat dipahami sebagai *platform* digital yang menyediakan fungsi untuk melakukan aktifitas sosial bagi setiap penggunanya. (Dr. Wira Yudha Alam S. IP. et al., 2023)

Rapid miner

Rapid Miner adalah suatu Analisis teks yang bekerja untuk memproses data mining, dengan melibatkan penggalian pola dari kumpulan data besar dan menggabungkannya dengan Teknik statistik, kecerdasan buatan, dan basis data. Tujuan dari analisis teks ini adalah untuk menghasilkan informasi yang akurat dan bermakna dari teks yang sangat besar atau kompleks. (Jundanuddin et al., 2024)

Data Mining

Data Mining adalah metode yang memungkinkan para penggunanya untuk mengakses data yang besar dalam waktu yang relatif cepat. Atau dengan kata lain data mining merupakan suatu alat atau aplikasi menggunakan analisis statistik pada data melalui suatu proses ekstraksi atau penggalian data dan informasi yang belum diketahui sebelumnya. (Ardilla et al., 2021)

Data mining adalah metode yang memungkinkan para penggunanya untuk mengakses data yang besar dalam waktu yang relative cepat. Atau dengan kata lain data mining merupakan suatu alat dan aplikasi menggunakan analisis statistic pada data melalui suatu alat dan aplikasi menggunakan analisis statistic pada data melalui suatu proses ekstraksi atau penggalian data dan informasi yang belum diketahui sebelumnya. Secara sederhana data mining merupakan proses penggalian suatu data yang berujung pada penemuan informasi terbaru dengan cara mencari pola atau aturan tertentu dari sejumlah data yang sangat besar, sehingga cara kerja dari data mining sebenarnya adalah untuk memeriksa database yang berukuran besar guna menemukan pola atau bentuk yang baru sehingga berguna dalam proses pengambilan Keputusan. (Ardilla et al., 2021)

Data mining memiliki serangkaian proses yang perlu dilakukan sebelum mendapatkan informasi baru. Yaitu, pembersihan data, pembersihan proses penghilangan noise dan data yang tidak konsisten, proses integrasi data yang menggabungkan data dari sumber data yang berbeda, pemilihan data, dan proses seleksi. Data yang telah diambil untuk digunakan dalam proses data mining, meninggalkan data yang tidak digunakan. Konversi data adalah proses mengubah data ke dalam format yang dapat digunakan dalam perhitungan algoritmik dan evaluasi pola. Merupakan Proses analisis hasil proses penambangan menggunakan satuan pengukuran, dan penyajian pengetahuan, adalah proses menampilkan hasil proses penambangan. (Nurhidayah, 2021)

Gambar 1. Contoh gambar

K-Nearest Neighbor (K-NN)

Metode K-NN merupakan suatu metode untuk klasifikasi, metode ini melakukan klasifikasi terhadap suatu objek yang berdasarkan kepada data training yang mempunyai jarak yang paling dekat dari obyek tersebut. Algoritma K-Nearest Neighbor atau yang biasa disebut KNN merupakan metode pengklasifikasian data yang bekerja relatif dengan cara yang lebih sederhana dibandingkan metode pengklasifikasian data lainnya. (Aprilian et al., 2020)

KNN adalah metode jenis supervised machine learning yang dapat digunakan untuk klasifikasi dan regresi. Cara kerjanya menggunakan data latih dan mengklasifikasi data tes berdasarkan jarak ke data latih tersebut.

Langkah-langkah klasifikasi sebagai berikut:

1. Menentukan nilai k.
2. Menghitung jarak antara data tes yang baru dan tetangga terdekat sejumlah k pada data latih
3. Memeriksa dengan *voting*, di kelas manakah yang terbanyak jumlah tetangganya. Yang memiliki jumlah tetangga terdekat terbanyak akan dipilih sebagai hasil kelas klasifikasi. (Budhi Widodo, 2022)

HASIL DAN PEMBAHASAN**Penerapan Data Mining Menggunakan Rapid Miner**

Tahapan awal dalam pengaplikasian data mining pada rapid miner seperti gambar dibawah ini:

Gambar 2. Contoh gambar

Evaluasi

Tahapan selanjutnya ini adalah tahapan evaluasi untuk mengavaluasi hasil dari algoritma K-Nearest untuk mengetahui kesimpulan dari proses analisis data mining.

Gambar 3. Contoh gambar

Pada gambar di atas merupakan hasil evaluasi yang dihasilkan oleh Performance Vector. Dalam konteks data mining, performance vector adalah kunci untuk mengidentifikasi area yang perlu perbaikan dan perbandingan. Pada gambar tersebut Performance vector mengacu pada metrik atau vektor yang digunakan untuk mengevaluasi kinerja modal

Klasifikasi

Tahapan selanjutnya ini adalah tahapan evaluasi untuk mengavaluasi hasil dari algoritma K-Nearest untuk mengetahui kesimpulan dari proses analisis data mining.

Gambar 4. Contoh gambar

Dari gambar tersebut dapat disimpulkan menurut data klasifikasi tersebut menghasilkan nilai akurasi 54,29% dari proses pengerjaan yang dihasilkan.

KESIMPULAN

Berdasarkan proses panjang dari pengerjaan analisis data mining media sosial dengan metode KNN dapat disimpulkan, bahwa efektifitas KNN terbukti efektif untuk klasifikasi data teks dari media sosial. Dan berdasarkan evaluasi kerja penggunaan rapid miner memudahkan proses data mining dengan antarmuka yang user-friendly dan berbagai alat yang tersedia untuk preprocessing, pemodelan, dan evaluasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penulisan jurnal ini. Terimakasih kepada Universitas Ibrahimy terlebih Fakultas Sains dan Teknologi terutama Dosen Pembimbing Bapak Zaehol Fatah, M. Kom. Dan seluruh teman-teman yang sudah mendukung dan banyak memberi arahan dan terimakasih juga kepada kedua orang tua yang selalu memberi dukungan baik finansial maupun doa.

DAFTAR PUSTAKA

- Aprilian, R., Habibi, R., Setyawan, M. Y. H., & Awangga, R. M. (2020). *Algoritma KNN dalam memprediksi cuaca untuk menentukan tanaman yang cocok sesuai musim*. Kreatif. <https://books.google.co.id/books?id=69f9DwAAQBAJ>
- Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Manuhutu, M. A., Ridwan, M., & Wardhani, A. K. (2021). *DATA MINING DAN APLIKASINYA*. Penerbit Widina. <https://books.google.co.id/books?id=53FXEAAAQBAJ>
- Arrohman, S., & Fatah, Z. (2024). *Gudang Jurnal Multidisiplin Ilmu Prediksi Diabetes Menggunakan Algoritma Klasifikasi K-Nearest Neighbors (K-NN) pada Perempuan Indian Pima*. 2, 220–226.
- Dr. Wira Yudha Alam S. IP., S. E. M. S. M. M. I. P., Pandian, A. P., Cindy, M. R., Billa, E. J. S., Firmansyah, M. R., Anggraini, D. S., & Sofatunisa, A. (2023). *Sosial Media dan Strategi Pemasaran*. MEGA PRESS NUSANTARA. <https://books.google.co.id/books?id=fYj9EAAAQBAJ>
- Final-MGHFRS+Mutiara+1027-1033*. (n.d.).
- Jundanuddin, M., Fatah, Z., Munazilin, A., Informasi, S., Sains, F., Sukorejo, U. I., Komputer, I., Sains, F., & Sukorejo, U. I. (2024). *Implementasi Data Mining Pada Penilaian Kinerja Guru Menggunakan Metode K-Nearest Neighbor (K-NN) Studi Kasus Pada Mi Alfabari Silo-Jember Implementation Of Data Mining On Teacher Performance Appraisal Using K-Nearest Neighbor (K-NN) Method Case Study*. 13(105), 869–880.
- Lugianti, D., Indramawan Nugroho, B., & Surorejo, S. (2022). Penerapan Sistem Interaksi Manusia dan Komputer Pada Website e-Learning di STMIK Tegal. *Jurnal Minfo Polgan*, 11(2), 36–44. <https://doi.org/10.33395/jmp.v11i2.11639>
- Muhamad Ayub, & Sofia Farzanah Sulaeman. (2022). Dampak Sosial Media Terhadap Interaksi Sosial Pada Remaja: Kajian Sistematis. *Jurnal Penelitian Bimbingan Dan Konseling*, 7(1), 21–32.
- Mulyono, F. (2021). Dampak Media Sosial bagi Remaja. *Jurnal Simki Economic*, 4(1), 57–65. <https://doi.org/10.29407/jse.v4i1.66>
- Muthi' Ahmad, S. H. (n.d.). *Fenomena Medsos (Studi Fenomena Dampak Negatif Media Sosial Terhadap Keharmonisan Rumah Tangga)*. GUEPEDIA. <https://books.google.co.id/books?id=udkWEAAAQBAJ>
- Nugraha, A., Chrisnanto, Y. H., & Yuniarti, R. (2019). Prediksi Sentimen Pada Sosial Media Twitter Mengenai Produk Smartphone Menggunakan Algoritma K-NN Classification. *Sensasi*, 251–258.
- Nurhidayah. (2021). *Peguruang: Conference Series*. 3(November), 10–14.
- Setiawan, Z., Fajar, M., Priyatno, A. M., Putri, A. Y. P., Aryuni, M., Yuliyanti, S., Widiputra, H., Meilani, B. D., Ibrahim, R. N., & Azdy, R. A. (2023). *BUKU AJAR DATA MINING*. PT. Sonpedia Publishing Indonesia. <https://books.google.co.id/books?id=1nLVEAAAQBAJ>
- Yogianto, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1720–1728. <https://doi.org/10.33379/gtech.v8i3.4495>
- Yogianto, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1720–1728. <https://doi.org/10.33379/gtech.v8i3.4495>