**SURVEY SENTIMEN ANALYSIS PILKADA JAKARTA 2017 PADA TWITTER**

# PENDAHULUAN

Pemilu pilkada memang suatu kegiatan lima tahunan yang dilakukan disetiap daerah di Indonesia.

Begitupun pemilu pilkada Jakarta dilakukan, dan pada pengolahan data kali ini digunakan dataset twit di twitter pada pilkada Jakarta tahun 2017 dengan menggunakan orange data mining.

# METODOLOGI

* 1. Dataset

Data set yang digunakan adalah twit dari twitter pada tahun 2017. Berikut data yang visualisasi nya.

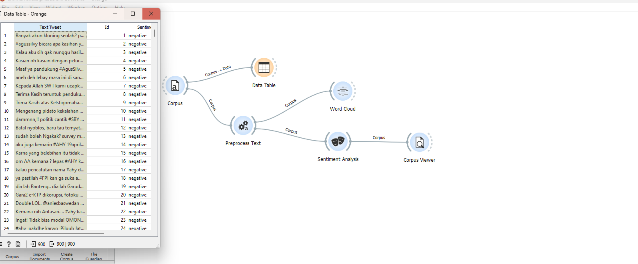


Gambar 1.1 Dataset Pilkada 2017

Kemudian akan dilakukan percobaan pengolahan data dengan menggunakan algoritma text minning sesuai dengan teori pembelajaran pertemuan ke-11.

* 1. Proses data

Tools yang digunakan adalah orange data mining dengan beberapa librarynya. Berikut capture pengolahan datanya:



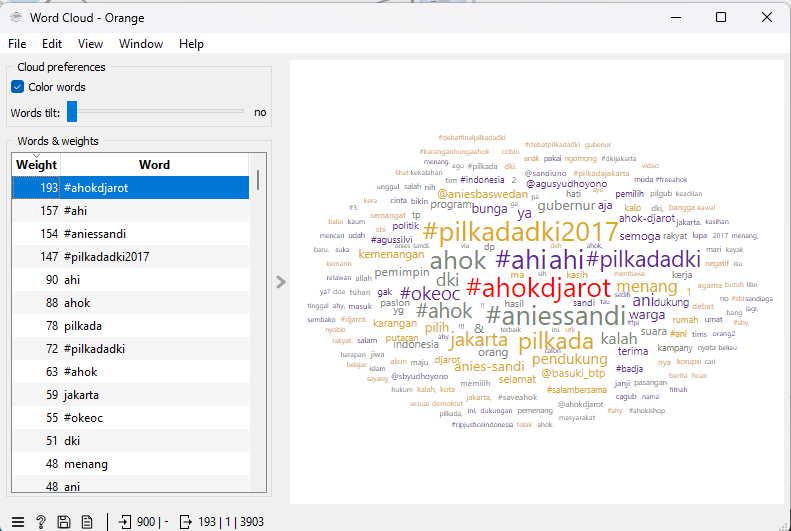
Gambar 1.2 proses pengolahan data orange

# Intepretasi output

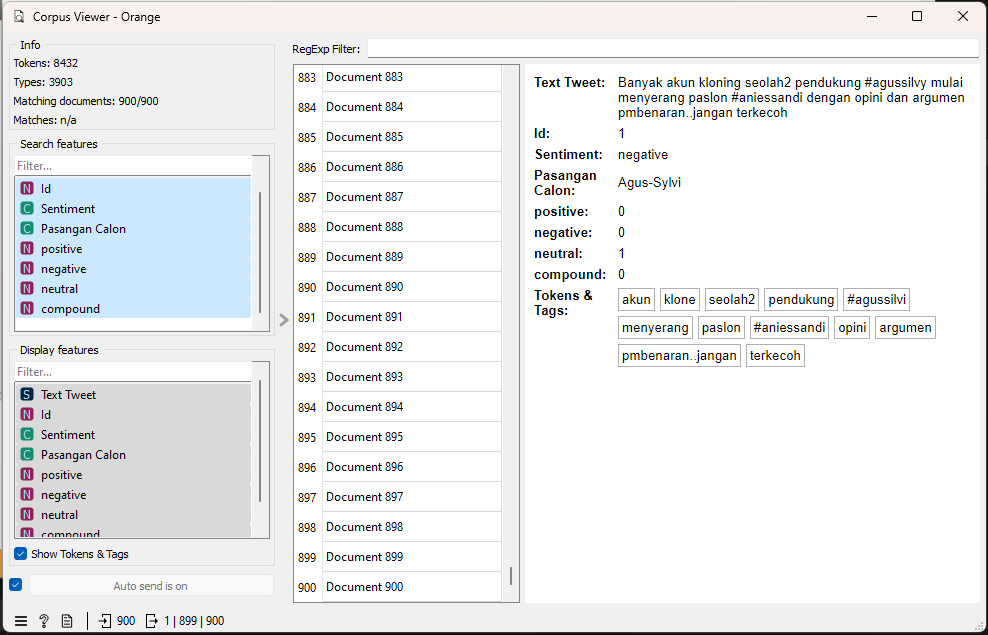
Dari prosesing data diatas, beberapa kolom yang diambil diantaranya:

Kolom tweet pasangan agus-silvy, Basuki cahya puranama – jarot ,anis-sandiaga uno.

Berikut contoh bag word dari pengolahan data



Dari output data bag of word dapat dilihat untuk kata ahokdjarot adalah yang paling banyak muncul dan trenf nya positif



Gambar 1.4 corpus data orange

# KESIMPULAN DAN SARAN

Didapat Nilai silhouette Score 0.35 artinya pengclusteran data kurang baik, dan hal ini dari referensi yang saya dapat bisa dilakukan pengclusteran dengan iterasi beberap kali agar penclusteran lebih optimah, juga bisa dengan metode pendekatan berbeda seperti selain Euclid distance ada manhattan distance dan minkawski distance.

Dari pengolahan data k-mean untuk data titanic pessanger ini lebih akan optimal jika nilai k itu sama dengan 6 sesuai dengan grafik elbow yg didapat.

Jadi bisa disimpulkan secara sederhana untuk hasil pengclusteran data pada data titanic passanger masih kurang baik.

# DAFTAR PUSTAKA

[1] https://www.kaggle.com/code/mrisdal/exploring-survival-on-the-titanic#introduction

Kelebihan dan Kekurangan Pemanfaat K-Mean

|  |  |
| --- | --- |
| Kelebihan | Kekurangan |
| * Sederhana dan Mudah Dipahami   K-Means mudah diimplementasikan dan dipahami. Algoritma ini melibatkan beberapa langkah sederhana seperti inisialisasi centroid, penetapan titik data ke centroid terdekat, dan pembaruan centroid. | * Harus Menentukan Jumlah Cluster (k) di Awal:   Salah satu kelemahan utama adalah bahwa kita harus menentukan jumlah cluster (k) di awal, yang seringkali tidak diketahui sebelumnya dan bisa memerlukan trial and error atau metode lain untuk dipilih. |
| * Efisien dalam Skala Waktu   K-Means memiliki kompleksitas waktu yang rendah (O(n \* k \* d)), di mana n adalah jumlah titik data, k adalah jumlah cluster, dan d adalah dimensi data. Ini membuat K-Means cocok untuk dataset yang besar | * Sensitif terhadap Inisialisasi   Hasil K-Means sangat dipengaruhi oleh inisialisasi awal centroid. Jika centroid awal dipilih dengan buruk, algoritma mungkin memberikan solusi yang buruk atau suboptimal. |
| * Konsisten dan Skalabel:   K-Means dapat dengan mudah diterapkan pada dataset besar dan skala tinggi, karena algoritma ini sangat efisien dalam hal komputasi. | * Bentuk Cluster   K-Means mengasumsikan bahwa cluster berbentuk bulat atau sferis (circular/spherical) dan memiliki ukuran yang serupa. Hal ini membuatnya kurang cocok untuk dataset dengan cluster yang berbentuk atau ukuran yang sangat berbeda. |
| * Fleksibilitas   Algoritma ini dapat digunakan dengan metrik jarak yang berbeda (meskipun biasanya menggunakan Euclidean distance) dan dapat diterapkan dalam berbagai domain dan jenis data | * Tidak Tahan terhadap Outlier   K-Means sangat sensitif terhadap outlier dan titik data yang sangat jauh, yang dapat secara signifikan menggeser centroid dan merusak hasil clustering. |
|  | * Tergantung pada Skala Data   K-Means menggunakan metrik jarak Euclidean, yang sensitif terhadap skala data. Oleh karena itu, normalisasi atau skala data biasanya diperlukan sebelum menerapkan K-Means. |
|  | * Konvergensi ke Minimum Lokal   Algoritma K-Means dapat terjebak dalam minimum lokal, yang berarti mungkin tidak menemukan solusi optimal secara global. Beberapa pengulangan dengan inisialisasi yang berbeda sering kali diperlukan untuk mendapatkan hasil yang baik |

JURNAL PEMANFAATAN TERKAIT K-MEAN

|  |  |
| --- | --- |
| Jurnal 1 | |
| Judul | **A Hybrid Deep Learning Model Using CNN and K-Mean Clustering for Energy Efficient Modelling in Mobile EdgeIoT** |
| Penulis | (Bisen *et al.*, 2023) |
| Tahun | (2023) |
| Hasil dan Pembahasan | 1. Pembentukan Cluster Hemat Energi dalam Mobile Edge Computing:   Pada penelitian ini mengusulkan algoritme baru (E-CFSA) untuk pembentukan cluster hemat energi di MEC. Ini menggabungkan jaringan saraf konvolusi (CNN) untuk pembagian tugas dan metode k-means yang dimodifikasi untuk pemilihan kepala cluster dengan redundansi untuk mengurangi overhead reclustering.   1. Pembelajaran Mesin untuk Penggunaan Energi Optimal:   Penelitian ini juga memperkenalkan model hibrida yang menggunakan pembelajaran mesin (CNN) untuk mengoptimalkan konsumsi energi di MEC dengan membuat keputusan efisien tentang transfer data dan alokasi tugas.   1. Meningkatkan Stabilitas Cluster di Jaringan MEC:   Studi ini membahas tantangan dalam mempertahankan cluster yang stabil di MEC karena seringnya perubahan lokasi. Ini mengusulkan metode menggunakan CNN dan clustering k-means yang dimodifikasi untuk memastikan pemilihan kepala cluster yang tepat dan mengurangi reclustering.   1. E-CFSA vs. Algoritma K-Mean yang Ada:   Penelitian ini membandingkan algoritme E-CFSA baru dengan pendekatan tradisional (WCA dan AB-SEP) untuk pembentukan cluster di MEC. Hasilnya menunjukkan bahwa E-CFSA mencapai kinerja yang lebih baik dalam hal konsumsi energi, overhead, pengiriman paket, dan throughput.   1. Fitur Utama E-CFSA:   Pendekatan ini menggabungkan CNN untuk pembagian tugas dan metode k-means yang dimodifikasi dengan kepala cluster redundan untuk mencapai efisiensi energi dan mengurangi overhead reclustering di jaringan MEC. |
| Kesimpulan Penelitian | Algoritma E-CFSA mencapai kinerja yang lebih baik dalam hal konsumsi energi, overhead, pengiriman paket, dan throughput. |

|  |  |
| --- | --- |
| Jurnal 2 | |
| Judul | **K-Means Cluster Analysis for Image Segmentation** |
| Penulis | (Aqil Burney, Karachi and Humera Tariq, 2014) |
| Tahun | (2014) |
| Hasil dan Pembahasan | * Studi pada jurnal ini berfokus pada pemilihan ruang warna yang optimal untuk segmentasi gambar menggunakan K-Means. * studi ini membandingkan performa K-Means di kedua ruang warna RGB dan Lab\*. * Penelitian ini menyelidiki efektivitas K-Means dalam segmentasi gambar * Mereka membandingkan kinerja K-Means dalam ruang warna RGB dan Lab\* dengan jumlah cluster yang berbeda (2, 3, dan >3). * dan temuan menunjukkan bahwa Lab\* memberikan hasil yang lebih baik, dengan akurasi segmentasi yang lebih tinggi dibandingkan dengan RGB. * Hasilnya menunjukkan bahwa K-Means bekerja lebih baik dalam ruang warna Lab\* dibandingkan dengan RGB, dengan akurasi model antara 30% hingga 65% di Lab\* dan 30% hingga 55% di RGB. * K-Means dan Ruang Warna Lab Lebih Unggul untuk Segmentasi Gambar. |
| Kesimpulan Penelitian | K-Means bekerja lebih baik dalam ruang warna Lab\* dibandingkan dengan RGB, dengan akurasi model antara 30% hingga 65% di Lab\* dan 30% hingga 55% di RGB. |

|  |  |
| --- | --- |
| Jurnal 3 | |
| Judul | **K-mean clustering and local binary pattern techniques for automatic brain tumor detection** |
| Penulis | (Baji, Abdullah and Abdulsattar, 2023) |
| Tahun | (2023) |
| Hasil dan Pembahasan | * pada jurnal ini menggunakan algoritma K-mean untuk memilih bagian cluster yang tepat untuk menggambarkan tumor, * Teknik ini diuji pada 30 gambar MRI, mencapai tingkat akurasi sebesar 87%, menunjukkan efektivitasnya dalam deteksi tumor otak. * dalam jurnal ini menganalisis pengaruh cluster gambar yang berbeda. Setiap cluster kemudian dipecah menjadi bagian kiri dan kanan. Setelah itu, fitur tekstur digambarkan pada setiap bagian. Selanjutnya, ukuran simetri bilateral diterapkan untuk memperkirakan cluster yang berisi tumor. Terakhir, pelabelan komponen terhubung digunakan untuk menentukan kelompok target untuk deteksi tumor otak. Teknik yang dikembangkan diterapkan pada 30 gambar MRI. Akurasi yang menggembirakan diperoleh sebesar 87%. |
| Kesimpulan Penelitian | Algoritma K-mean yang dikembangkan diterapkan pada 30 gambar MRI. Akurasi yang menggembirakan diperoleh sebesar 87%. |

|  |  |
| --- | --- |
| Jurnal 4 | |
| Judul | **Clustering Monovarietal Extra Virgin Olive Oil According to Sensory Profile, Volatile Compounds, and k-Mean Algorithm** |
| Penulis | (Cecchi *et al.*, 2022) |
| Tahun | (2022) |
| Hasil dan Pembahasan | * Penelitian pada jurnal ini tentang pengelompokkan Minyak Zaitun Monovarietas berdasarkan Profil Sensori, minyak zaitun extra virgin monovarietas (EVOO) dengan karakteristik sensori dan kimiawi yang serupa. * pengelompokan data pada penelitian ini menggunakan algoritma K-Mean untuk clustering. * dihipotesiskan bahwa sampel darisuatu kultivar terletak dalam satu cluster, sedangkan sampel non-monovarietal berada ditempatkan secara acak dalam kelompok yang berbeda. * Penelitian ini menganalisis 46 EVOO dari varietas zaitun tunggal dan sekelompok minyak campuran varietas. * Peneliti menemukan dua kelompok berbeda: satu kelompok berisi semua sampel monovarietas dan kelompok lainnya berisi minyak dengan varietas campuran yang tersebar di kedua kelompok. * Deskripsi sensori dan senyawa volatil spesifik adalah faktor utama yang membedakan kelompok, menunjukkan varietas zaitun yang berbeda. |
| Kesimpulan Penelitian | * Peneliti menemukan dua kelompok berbeda: satu kelompok berisi semua sampel monovarietas dan kelompok lainnya berisi minyak dengan varietas campuran yang tersebar di kedua kelompok. |

|  |  |
| --- | --- |
| Jurnal 5 | |
| Judul | **Text Clustering using K-MEAN** |
| Penulis | Chaman Lal1,2, Awais Ahmed1,Reshman Siyal3, Suresh Kumar Beejal3,Shagufta Aftab3,Arshad Hussain2 |
| Tahun | (2021) |
| Hasil dan Pembahasan | * Penelitian ini mengelompokkan Lagu Kebangsaan Pakistan - Studi Kasus Analisis Teks Pendek * Penelitian ini membahas tantangan dalam menerapkan teknik pengelompokkan dokumen ke teks pendek dalam bahasa dengan sumber daya terbatas (seperti bahasa Urdu, bahasa lagu kebangsaan Pakistan) * Eksperimen ini menggunakan lagu kebangsaan Pakistan sebagai studi kasus dan menggunakan pengelompokkan K-Means dengan fitur TF-IDF untuk mencapai pengelompokkan tematik. * Dengan berfokus pada lagu kebangsaan Pakistan, penelitian ini mengeksplorasi kelayakan penggunaan pengelompokkan K-Means dengan fitur TF-IDF untuk mengelompokkan konten serupa di dalam lagu kebangsaan. |
| Kesimpulan Penelitian |  |

|  |  |
| --- | --- |
| Jurnal 6 | |
| Judul | **Study and Implementing K-mean Clustering Algorithm on English Text and**  **Techniques to Find the Optimal Value of K** |
| Penulis | Naeem, Sajid. Wumaier, Aishan |
| Tahun | (2018) |
| Hasil dan Pembahasan | * Pada jurnal ini membahas cara pengoptimalan nilai k pada algoritma k-mean dalam mengolah data text Bahasa inggis * Di Kesimpulan pada jurnal percobaan itu dialami dengan jelas bahwa selain itu ambiguitas dalam nilai True K, k-means juga tinggidipengaruhi oleh pemilihan pusat massa awal, outlier dan noise,prapemrosesan dan dimensi tinggi (data cadangan besar) karena dalam clustering dokumen hasil akhir clustering adalah sangat terpengaruh oleh langkah prapemrosesan itu. Banyak peneliti seperti di [50] telah dirinci lebih baik menggunakan reduksi dimensi meskipun langsung menerapkan k-means in data berdimensi tinggi dan dapat menggunakan PCA (principle komponen analisis) untuk reduksi dimensi. * Nilai K yang sebenarnya sebagian besar dapat dimengerti sementara secara otomatis memilih nilai yang sesuai untuk k adalah hal yang sulit masalah algoritmik. K yang sebenarnya menunjukkan kepada kita berapa banyak cluster harus dibuat dalam kumpulan data kami tetapi K ini sering terjadi ambigu tidak ada jawaban khusus untuk pertanyaan ini sementara banyak varian k-means disajikan untuk memperkirakannya nilai |
| Kesimpulan Penelitian | * Nilai K yang sebenarnya sebagian besar dapat dimengerti sementara secara otomatis memilih nilai yang sesuai untuk k adalah hal yang sulit masalah algoritmik. K yang sebenarnya menunjukkan kepada kita berapa banyak cluster harus dibuat dalam kumpulan data kami tetapi K ini sering terjadi ambigu tidak ada jawaban khusus untuk pertanyaan ini sementara banyak varian k-means disajikan untuk memperkirakannya nilai |

|  |  |
| --- | --- |
| Jurnal 7 | |
| Judul | Data Hiding by Unsupervised Machine Learning Using Clustering K-mean Technique |
| Penulis | Hiba Hamdi Hassan1, Maisa'a Abid Ali Khodher2 (Department of Computer Science, University of Technology, Baghdad, Iraq) |
| Tahun | (2021) |
| Hasil dan Pembahasan | * Artikel kali ini membahas tentang steganografi yang dapat digunakan sebagai metode pembelajaran mesin, ini menyarankan metode baru untuk menyembunyikan data dengan menggunakan mesin tanpa pengawasan pembelajaran (pengelompokan algoritma k-mean). * Sistem ini menggunakan data tersembunyi yang dimasukkan ke dalam gambar sampul terdiri dari tiga langkah: langkah pertama membagi gambar sampul menjadi tiga clustering yang lebih kontras dengan menggunakan cluster k-means, teks atau gambar dipilih untuk dikonversi ke biner dengan menggunakan ASCII kode, langkah ketiga menyembunyikan pesan biner atau gambar biner pada gambar sampul dengan menggunakan sekuensial metode LSB. Setelah itu, sistem diimplementasikan |
| Kesimpulan Penelitian | * Tujuan dari sistem yang disarankan diperoleh, menggunakan Pembelajaran Mesin Tanpa Pengawasan (teknik K-mean) untuk mengirimkan informasi sensitif dengan aman tanpa khawatir tentang serangan man-in-the-middle diusulkan. Metode ini ditandai dengan keamanan dan kapasitas tinggi. Melalui evaluasi, sistem menggunakan PSNR, MSE, Entropy, dan Histogram untuk menyembunyikan pesan rahasia dan gambar rahasia dalam domain spasial pada gambar sampul. |

|  |  |
| --- | --- |
| Jurnal 8 | |
| Judul | Implementation of TF-IDF Algorithm and K-mean Clustering Method to Predict Words or Topics on Twitter |
| Penulis | (Darwis, Tri Pranoto and Eka Wicaksana, 2020) |
| Tahun | (2018) |
| Hasil dan Pembahasan | * Dalam tulisan ini, penulis meneliti clustering dengan k-mean dan tpik data di twiiter * Hasil clustering data tweet menunjukkan prediksiatau mungkin topik pembicaraan yang sedang banyak dibicarakan oleh netizen. Akhirnya data tersebut dapat digunakan untuk membuat keputusan yang memanfaatkan sentimen masyarakat terhadap suatu peristiwa melalui media sosial seperti Twitter. * Hasil pada k-men dengan k = 3 didapatkan clustering corona, Indonesia dan virus * Penulis pengimplementasikan algoritma TF-IDF dalam prediksi data twitter |
| Kesimpulan Penelitian | Hasil clustering dengan menggunakan algotirma k-mean dengan k = 3 yaitu Indonesia corona virus, bisa diartikan data set twitter berbicara ke topik pandemi corona. |

|  |  |
| --- | --- |
| Jurnal 9 | |
| Judul | Detection and Monitoring of Power Line Corridor From Satellite Imagery Using RetinaNet and K-Mean Clustering |
| Penulis | (Mahdi Elsiddig Haroun, Deros and Din, 2021) |
| Tahun | (2021) |
| Hasil dan Pembahasan | * Dalam Jurnal ini, penulis memperkenalkan metode baru untuk memantau koridor saluran listrik dari citra satelit. yaitu * metode yang diusulkan terdiri dari dua tahap.   Pada tahap pertama, Penulis menggunakan RetinaNet yang canggih dan sudah ada model pembelajaran (DL) untuk mendeteksi lokasi TT dari citra satelit. Algoritma perutean telah dibuat dikembangkan untuk membuat jalur antara setiap TT terdeteksi yang berdekatan. Selain algoritma routing, koridor algoritma identifikasi telah dibuat untuk mengekstraksi area koridor saluran listrik.  Pada tahap kedua, algoritma k-mean clustering telah digunakan untuk menyorot wilayah VE dalam area koridor saluran listrik setelah mengubah citra satelit target menjadi ruang warna hue, saturation, dan value (HSV). Yang diusulkan sistem pemantauan mampu mendeteksi TT dari citra satelit dengan mean average presisi (mAP) sebesar 72,45% untuk ambang batas Intersection of Union (IoU) sebesar 0,5 dan 85,21% untuk ambang batas IoU sebesar 0,3. Juga,sistem pemantauan berhasil membedakan wilayah vegetasi dengan kepadatan tinggi dan rendah area koridor saluran listrik. |
| Kesimpulan Penelitian | Untuk penggunaan algoritma k-mean clustering digunakan untuk menyorot wilayah VE dari hasil citra satelit dengan warna blue, hue dan saturation dengan k = 5 |

|  |  |
| --- | --- |
| Jurnal 10 | |
| Judul | **K-MEAN CLUSTERING IN TRANSPORTATION: A WORK ZONE SIMULATOR CASE STUDY** |
| Penulis | (Moradpour and Long, no date) |
| Tahun | (2021) |
| Hasil dan Pembahasan | * Penelitian ini menggunakan simulator untuk membandingkan pola pengemudi dan perilaku ketika membandingkan reaksi terhadap tanda alternatif Departemen Transportasi Missouri (MoDOT). dengan tanda terkini Manual on Uniform Traffic Control Devices (MUTCD). * Metode pengelompokan K-mean digunakan untuk respons driver cluster terhadap konfigurasi tanda zona kerja disajikan di lingkungan simulator dan terungkap pola yang dapat membantu insinyur dalam kegunaan signage zona kerja. Temuan-temuan utama dari penelitian ini akan membantu manajer teknik transportasi membuat keputusan berdasarkan data mengenai keselamatan dan desain zona kerja. |
| Kesimpulan Penelitian | Algortima k-mean pada penelitian ini diterapkan untuk mempartisi data arus lalu lintas. Kekelompokan metode membagi data ke dalam kelompok-kelompok (cluster) berdasarkan persamaan dan perbedaan antar kelompok dan berguna untuk menemukan pola antara sejumlah besar data. Pengelompokan K-mean adalah salah satu metode pengelompokan yang paling umum dan digunakan untuk menentukan pusat cluster. |