

Accuracy Test of K-Means in Predicting Election Participation in the Demographics of Pasuruan District Region

[Uji Akurasi K - Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan]

Dafid Mizta Chulloh¹⁾, Arif Senja Fitriani^{*2)}

¹⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Korespondensi: asfjim@umsida.ac.id

Abstract. *The term Indonesia as a democratic country is legitimately echoed because Indonesia has used elections as a means to change leaders. Participation is a measure of the success of an election. This research aims to predict election participation in the demographics of Pasuruan Regency. The process of predicting or classifying using an algorithm K - Means plus evaluation model Inertia and Silhouette. The results of this research have been carried out in predicting election participation in the demographics of the Pasuruan Regency area which was carried out using the K - K - Means Algorithm with 4 variables, namely 64% election participation with 3 cluster as a comparison and research adventure with 4 cluster as much as 68% of public participation in the election.*

Keywords - Data Mining; K – Means; Participation; Elections; Predictions.

Abstrak. *Sebutan Indonesia sebagai negara demokrasi secara sah digaungkan sebab indonesia telah mendayagunakan pemilu sebagai sarana untuk melakukan pergantian pemimpin. Partisipasi menjadi tolak ukur berhasilnya sebuah pemilu. Pada penelitian ini bertujuan untuk memprediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan. Proses dalam memprediksi atau mengklasifikasikan menggunakan algoritma K - Means ditambah model evaluasi Inertia dan Silhouette. Hasil penelitian ini yang sudah dilakukan dalam memprediksi partisipasi pemilu dalam demografi wilayah Kabupaten Pasuruan yang dilakukan dengan menggunakan Algoritma K – K - Means dengan 4 Variabel yaitu sebanyak 64 % partisipasi pemilu dengan 3 cluster sebagai perbandingan dan petualangan penelitian dengan 4 cluster sebanyak 68 % partisipasi masyarakat dalam pemilu.*

Kata Kunci - Data Mining; K – Means; Partisipasi; Pemilu; Prediksi.

I. PENDAHULUAN

Sebutan Indonesia sebagai negara demokrasi secara sah digaungkan sebab indonesia telah mendayagunakan pemilu sebagai sarana untuk melakukan pergantian pemimpin[1]. Pemilu atau pemilihan umum adalah sebuah pesta demokrasi yang dilakukan dalam proses perwujudan kedaulatan Indonesia[2]. Faktor utama dianggap berhasilnya sebuah pemilu adalah partisipasi[3]. Partisipasi menjadi tolak ukur berhasilnya sebuah pemilu[4]. Mengapa dikatakan seperti itu, Karena sejauh mana jumlah pemilih berpartisipasi dalam pemilu atau pemilihan umum semakin banyak warga yang menggunakan hak pilihannya, semakin besar tingkat partisipasi yang dianggap berhasil[5].

Menurut data yang didapat melalui Komisi Pemilihan Umum (KPU), tingkat partisipasi pemilu terakhir pada tahun 2019 tercatat 77,5% dari penduduk Indonesia. Serta Daftar Pemilih Tetap (DPT) tahun 2019 mencapai angka 199.987.870 jiwa. Sedangkan penduduk indonesia yang menggunakan hak pilihnya sejumlah 158.012.506 jiwa. Hal tersebut dapat dijadikan acuan tolak ukur kesadaran penduduk indonesia dalam melakukan hak dan kewajibannya.

Masalah utama yang sering terjadi dalam pesta demokrasi adalah partisipasi. Kondisi tersebut berdampak pada tingkat partisipasi masyarakat dalam pemilu. Problem yang muncul dalam partisipasi pemilu bermacam - macam dengan berbagai demografi pada suatu wilayah nya seperti, demografi tersebut terpublish dalam bentuk data Badan Pusat Statistik (BPS) yang terupdate setiap tahun. Pada data tersebut meliputi Geografi dan Iklim, Pemerintahan, Penduduk, Sosial dan Kesejahteraan Rakyat, Pertanian, Pariwisata, Transportasi dan Komunikasi, Perbankan, Koperasi dan Perdagangan. Dari kondisi diatas penelitian ini mencoba mengkolaborasi dengan hasil rekapitulasi pemilu.

Pada penelitian ini bertujuan memprediksi partisipasi pemilu dalam demografi suatu wilayah. Manfaat dalam melakukan penulisan atau penelitian tersebut, Meningkatkan pemahaman tentang faktor-faktor yang mempengaruhi partisipasi pemilu. Penelitian ini dapat memberikan gambaran tentang faktor-faktor demografi yang berpengaruh terhadap partisipasi pemilu. Faktor-faktor tersebut dapat digunakan sebagai bahan pertimbangan dalam meningkatkan partisipasi pemilu.

Uji Akurasi K - Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan dapat dilakukan melalui berbagai macam cara dan tahapan guna dapat memprediksi serta menemukan informasi pengetahuan dengan bidang Data Mining[6]. Teknik data mining dapat digunakan untuk mengeksplorasi data, menemukan pola, dan membuat keputusan[7]. Algoritma yang mempunyai keakuratan yang sangat tinggi serta

mempunyai karakteristik dalam memprediksi sebuah permasalahan yaitu K - means[8]. K – means tersendiri terkenal pada umumnya adalah metode yang terbaik dikelasnya dalam pengelompokan (clustering)[9]. Selain itu, K-means juga dapat digunakan untuk mengidentifikasi pola tersembunyi atau kelompok-kelompok yang mungkin tidak terlihat secara kasat mata dalam data[10]. K - means tidak bisa dianggap remeh dalam melakukan prediksi dengan algoritma lain serta dalam keterikatan memberikan pengetahuan memprediksi partisipasi pemilu[11]. Serta membandingkan apakah tingkat partisipasi pemilih dalam pemilihan sebelumnya dapat dijadikan prediktor untuk pemilihan berikutnya. Diharapkan hasil yang diperoleh dapat membantu untuk proses melihat partisipasi pemilu di wilayah kabupaten pasuruan. II. Metode

Penelitian ini terdiri dari 5 proses kerja penelitian sebagaimana ditunjukkan pada gambar 1, yakni : Studi pustaka, Pengumpulan data, Pre – processing, Implementasi algoritma, dan Pengetahuan hasil.



Gambar 1. Proses Kerja Penelitian

A. Studi Pustaka

Dalam tahapan pada penelitian ini studi pustaka tentang memprediksi partisipasi pemilu untuk diteliti khususnya dalam demografi wilayah kabupaten pasuruan.

B. Pengumpulan Data

Penelitian ini melakukan serangkaian kegiatan yang dilakukan untuk mengumpulkan informasi atau data dari berbagai sumber. Pada proses tersebut harus direncanakan dengan cermat untuk memastikan bahwa data yang diperoleh sesuai dengan tujuan penelitian dan metodologi yang digunakan. Teknik pengumpulan data dengan cara mengambil satu per satu disebut teknik pengumpulan data secara manual atau disebut pengumpulan data manual. Teknik ini dilakukan dengan cara mengumpulkan data langsung dari sumber datanya, kemudian menginputkan ke dalam microsoft excel. Data yang dipergunakan yakni data BPS (Badan Pusat Statistik) Kabupaten Pasuruan tahun 2022 dengan hasil rekapitulasi pemilu di Kabupaten Pasuruan tahun 2019.

C. Pre - Processing

Pada proses tersebut telah melibatkan 5 serangkaian proses yang sistematis dan penting. Proses ini mencakup mencakup (Data Integration), (Data Transformation), (Attribute Correlation), dan (Data Cleaning) :

1) Data Integration

Pada tahap integration menggabungkan, menyatukan, dan mengelola data dari berbagai sumber yang berbeda untuk membentuk satu set data yang lebih komprehensif, terstruktur, dan mendukung keakuratan k – means di google colaboratory sesuai dengan tujuan penelitian. Dalam penelitian ini dalam melakukan data integration menggunakan metode pemetaan data.

2) Data Transformation

Tahap transformasi atau perubahan data dilakukan dengan mengubah dari yang asli sesuai sumber kedalam format yang sesuai dan dapat dianalisis menggunakan algoritma K – Means di google colaboratory. Dalam penelitian ini dalam melakukan data transformation menggunakan metode yang berbeda karena menyesuaikan konteks penelitian, metode yang digunakan adalah perubahan format data, normalisasi dan standarisasi.

3) Attribute Correlation

Pada tahap attribute correlation sangat penting sekali karena harus menemukan korelasi data yang menarik serta sefrekuensi akan membuat konteks dalam suatu penelitian akan sejalan dengan imajinasi yang dirancang dalam penelitian. Dalam penelitian ini dalam melakukan attribute correlation menggunakan metode asosiasi karena untuk mengidentifikasi pola-pola asosiasi antara atribut-atribut sesuai konteks penelitian.

4) Data Cleaning

Pada tahap data cleaning ini untuk menghilangkan data atau atribut yang tidak sesuai dalam konteks penelitian. Tanpa Data Cleaning yang baik, hasil analisis dapat terganggu oleh masalah dalam data mentah. Dalam penelitian ini dalam melakukan data cleaning menggunakan metode pemeriksaan visual karena banyak ditemukannya data numerik yang tidak masuk akal atau data teks yang tidak lengkap.

Setelah tahap pengolahan data atau pre - processing selesai dan data set yang siap digunakan telah terbentuk. Data set meliputi Geografi dan Iklim, Pemerintahan, Penduduk, Sosial dan Kesejahteraan Rakyat, Pertanian, Pariwisata,

Transportasi dan Komunikasi, Perbankan, Koperasi dan Perdagangan dengan data set sejumlah 4378 instance dan 83 Atribut.

D. Implementasi Algoritma

Pada tahap implementasi terdapat 2 implementasi yaitu implementasi dengan perhitungan manual dengan menggunakan Microsoft Excel dan Google Colaboratory menggunakan k – means dengan dikolaborasikan 2 teknik yaitu inerti dan silhouette.

E. Pengetahuan Hasil

Pada tahap ini diperoleh informasi berupa pengetahuan hasil cluster dan akurasi pencarian partisipasi pemilu dalam demografi wilayah Pasuruan.

III. HASIL DAN PEMBAHASAN

Variabel merujuk pada sekelompok variabel yang terdiri dari bagian-bagian entitas. Ada 4 atribut yang digunakan sebagai variabel untuk Uji Akurasi K - Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan. Sementara itu, fitur mencakup konten dari variabel tersebut. Informasi mengenai variabel dan fitur yang digunakan dapat diidentifikasi berdasarkan Tabel 2 yang mencantumkan data yang telah diperoleh dari Badan Pusat Statistik (BPS) dan Komisi Pemilihan Umum (KPU). Sebagai pertimbangan mengapa mengambil atribut dijelaskan pada berikut :

1) Pengguna Listrik

Pengguna listrik merupakan salah satu indikator tingkat kesejahteraan masyarakat. Masyarakat yang memiliki akses listrik yang baik cenderung memiliki tingkat kesejahteraan yang lebih tinggi. Tingkat kesejahteraan yang tinggi dapat meningkatkan kesadaran politik masyarakat dan mendorong mereka untuk berpartisipasi dalam pemilu.

2) Jumlah Menara Telepon Seluler

Jumlah menara telepon seluler merupakan indikator ketersediaan infrastruktur telekomunikasi. Ketersediaan infrastruktur telekomunikasi yang baik dapat memudahkan masyarakat untuk mengakses informasi tentang pemilu. Informasi yang lengkap dan akurat dapat meningkatkan kesadaran politik masyarakat dan mendorong mereka untuk berpartisipasi dalam pemilu.

3) Kekuatan Sinyal Telepon Seluler

Kekuatan sinyal telepon seluler merupakan indikator kualitas infrastruktur telekomunikasi. Kualitas infrastruktur telekomunikasi yang baik dapat meningkatkan kenyamanan masyarakat dalam mengakses informasi tentang pemilu. Hal ini dapat meningkatkan kesadaran politik masyarakat dan mendorong mereka untuk berpartisipasi dalam pemilu.

4) Partisipasi

Partisipasi merupakan variabel yang ingin diprediksi. Variabel ini merupakan indikator tingkat partisipasi pemilu dalam suatu wilayah. Variabel ini digunakan untuk menguji pengaruh variabel-variabel demografi terhadap tingkat partisipasi pemilu.

Berdasarkan masalah yang utama mengapa memilih variabel - variabel atau alasan-alasan tersebut, keempat variabel tersebut dapat digunakan untuk memprediksi tingkat akurasi partisipasi pemilu dipengaruhi demografi wilayah kabupaten pasuruan. Keempat variabel tersebut dapat mewakili faktor-faktor demografi yang dapat memengaruhi tingkat partisipasi pemilu.

Table 1. Variabel, Fitur, dan Deskripsi

| No | Nama | Fitur | Deskripsi |
|----|---------------------------------|---|---|
| 1 | Pengguna Listrik | Numerik $\{0,13 < 0,14 < 0,15 < \dots < 0,82\}$ | Persentase, Fitur Numerikal |
| 2 | Jumlah Menara Telepon Seluler | Numerik $\{1, 2, 3, 4, 5, 6, 7\}$ | Jumlah, Fitur Numerikal |
| 3 | Kekuatan Sinyal Telepon Seluler | Kategorikal $\{1, 2, 3\}$ | Sinyal Lemah = 1 Sinyal Kuat = 2 Sinyal Sangat Kuat = 3 |
| 4 | Partisipasi | Kategorikal $\{1, 2\}$ | J1 = $<77,5$ = Rendah = 1 J2 = $>77,5$ = Tinggi = 2 |

A. Implementasi Dengan Perhitungan Manual

1) Data Pengujian

Pada tabel 2, Data pengujian yang telah ditetapkan dalam penelitian yakni merupakan kumpulan informasi yang digunakan untuk mengevaluasi kinerja suatu model atau sistem. Dalam konteks pengujian, data ini berperan penting untuk mengukur sejauh mana model dapat memberikan prediksi atau hasil yang akurat dan dapat diandalkan.

Table 2. Data Pengujian

| No | Pengguna Listrik | Jumlah Menara Telepon Seluler | Kekuatan Sinyal Telepon Seluler | Partisipasi |
|------|------------------|-------------------------------|---------------------------------|-------------|
| 1 | 0,25 | 0 | 3 | 2 |
| 2 | 0,25 | 0 | 3 | 2 |
| 3 | 0,25 | 0 | 3 | 1 |
| 4 | 0,25 | 0 | 3 | 2 |
| 5 | 0,25 | 0 | 3 | 2 |
| ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... |
| 4374 | 0,37 | 0 | 2 | 2 |
| 4375 | 0,37 | 0 | 2 | 2 |
| 4376 | 0,37 | 0 | 2 | 2 |
| 4377 | 0,37 | 0 | 2 | 2 |
| 4378 | 0,37 | 0 | 2 | 2 |

2) Menentukan Centroid awal secara acak atau random

Pemilihan centroid awal secara acak atau random dalam algoritma clustering adalah langkah strategis yang dirancang untuk meningkatkan keberagaman dan kemungkinan menemukan solusi yang lebih baik. Pemilihan acak ini dirancang untuk menghindari jebakan dalam solusi lokal atau suboptimal. Jika selalu memulai dengan centroid yang sama, algoritma cenderung terperangkap dalam pola yang sudah dikenali, tanpa melibatkan variasi yang dapat muncul dalam inisialisasi yang acak. Dengan kata lain, inisialisasi acak membuka pintu menuju eksplorasi yang lebih luas terhadap potensi struktur data yang kompleks.

Table 3. Hasil pemilihan centroid awal secara acak atau random

| Centroid | Pengguna Listrik | Jumlah Menara Telepon Seluler | Kekuatan Sinyal Telepon Seluler | Partisipasi |
|----------|------------------|-------------------------------|---------------------------------|-------------|
| C1 | 0,4042 | 1,0000 | 2,0000 | 2,0000 |
| C2 | 0,3137 | 1,0000 | 2,0000 | 2,0000 |
| C3 | 0,3902 | 2,0000 | 3,0000 | 2,0000 |
| C4 | 0,3902 | 2,0000 | 3,0000 | 1,0000 |

3) Menghitung jarak Centroid terdekat

$$\begin{aligned} \text{Data Ke1 C1} &= \sqrt{(0,25-0,4042)^2+(0-1,0000)^2+(3-2,0000)^2+(2-2,0000)^2} \\ &= 1,4223 \end{aligned}$$

$$\begin{aligned} \text{Data Ke1 C2} &= \sqrt{(0,25-0,3137)^2+(0-1,0000)^2+(3-2,0000)^2+(2-2,0000)^2} \\ &= 1,4155 \end{aligned}$$

$$\begin{aligned} \text{Data Ke1 C3} &= \sqrt{(0,25-0,3902)^2+(0-2,0000)^2+(3-3,0000)^2+(2-2,0000)^2} \\ &= 2,0047 \end{aligned}$$

$$\begin{aligned} \text{Data Ke1 C4} &= \sqrt{(0,25-0,3902)^2+(0-2,0000)^2+(3-3,0000)^2+(2-1,0000)^2} \\ &= 2,2403 \end{aligned}$$

Hasil perhitungan jarak di atas diubah kedalam bentuk tabel maka akan ditunjukkan pada tabel 5

Table 4. Hasil perhitungan jarak Centroid terdekat

| No | C1 | C2 | C3 | C4 | Jarak Terdekat |
|------|--------|--------|--------|--------|----------------|
| 1 | 1,4223 | 1,4155 | 2,0047 | 2,2403 | 1,4155 |
| 2 | 1,4223 | 1,4155 | 2,0047 | 2,2403 | 1,4155 |
| 3 | 1,7387 | 1,7331 | 2,2403 | 2,0047 | 1,7331 |
| 4 | 1,4223 | 1,4155 | 2,0047 | 2,2403 | 1,4155 |
| 5 | 1,4223 | 1,4155 | 2,0047 | 2,2403 | 1,4155 |
| ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... |
| 4374 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 |
| 4375 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 |
| 4376 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 |
| 4377 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 |
| 4378 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 |

Setelah melakukan perhitungan jarak terdekat di atas, kemudian mengelompokan berdasarkan kesamaan karakteristik atau nilai cluster

4) Pengelompokan berdasarkan karakteristik dan nilai

Table 4. Hasil perhitungan jarak Centroid terdekat

| No | C1 | C2 | C3 | C4 | Jarak Terdekat | Kelompok Data |
|------|--------|--------|--------|--------|----------------|---------------|
| 1 | 1,4223 | 1,4155 | 2,0047 | 2,2403 | 1,4155 | Cluster 2 |
| 2 | 1,4223 | 1,4155 | 2,0047 | 2,2403 | 1,4155 | Cluster 2 |
| 3 | 1,7387 | 1,7331 | 2,2403 | 2,0047 | 1,7331 | Cluster 2 |
| 4 | 1,4223 | 1,4155 | 2,0047 | 2,2403 | 1,4155 | Cluster 2 |
| 5 | 1,4223 | 1,4155 | 2,0047 | 2,2403 | 1,4155 | Cluster 2 |
| ... | ... | ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... | ... | ... |
| 4374 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 | Cluster 1 |
| 4375 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 | Cluster 1 |
| 4376 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 | Cluster 1 |
| 4377 | 1,0005 | 1,0018 | 2,2361 | 2,4495 | 1,0005 | Cluster 1 |

4378 1,0005 1,0018 2,2361 2,4495 1,0005 Cluster 1

Jika C1 kurang dari Jarak Terdekat maka dia sebagai Cluster 1, Jika C2 kurang dari Jarak Terdekat maka dia sebagai Cluster 2, Jika C3 kurang dari Jarak Terdekat maka dia sebagai Cluster 3, Jika C4 kurang dari Jarak Terdekat maka dia sebagai Cluster 4.

5) Hasil K – Means iterasi ke - 1

Dari perhitungan di atas, didapat nilai centroid baru pada tabel 7 sebagai berikut :

Table 5. Centroid Baru

| Centroid Baru | Pengguna Listrik | Jumlah Menara Telepon Seluler | Kekuatan Sinyal Telepon Seluler | Partisipasi |
|---------------|------------------|-------------------------------|---------------------------------|-------------|
| Cluster 1 | 0,4176 | 0,5622 | 2,0476 | 1,9233 |
| Cluster 2 | 0,3050 | 0,6246 | 2,1000 | 1,9097 |
| Cluster 3 | 0,3215 | 2,9814 | 2,6966 | 2,0000 |
| Cluster 4 | 0,3251 | 2,2731 | 2,6506 | 1,0000 |

6) K – Means iterasi ke – 6

Table 6. Centroid Baru iterasi ke 6

| Centroid Baru | Pengguna Listrik | Jumlah Menara Telepon Seluler | Kekuatan Sinyal Telepon Seluler | Partisipasi |
|---------------|------------------|-------------------------------|---------------------------------|-------------|
| Cluster 1 | 0,3263 | 0,0293 | 1,7745 | 1,8979 |
| Cluster 2 | 0,3388 | 0,7573 | 2,3802 | 1,8808 |
| Cluster 3 | 0,3202 | 2,7131 | 2,4914 | 2,0000 |
| Cluster 4 | 0,3235 | 2,5463 | 2,5756 | 1,0000 |

Setelah mendapatkan centroid baru, selanjutnya perbandingkan centroid baru tersebut dengan centroid sebelumnya. Apabila nilai keduanya identik atau sama, langkah proses dihentikan. Namun, jika terdapat perbedaan antara keduanya, proses akan dilanjutkan ke iterasi ke 7.

7) Hasil K – Means iterasi ke – 7

Table 6. Centroid Baru iterasi ke 7

| Centroid Baru | Pengguna Listrik | Jumlah Menara Telepon Seluler | Kekuatan Sinyal Telepon Seluler | Partisipasi |
|---------------|------------------|-------------------------------|---------------------------------|-------------|
| Cluster 1 | 0,3263 | 0,0293 | 1,7745 | 1,8979 |
| Cluster 2 | 0,3388 | 0,7573 | 2,3802 | 1,8808 |
| Cluster 3 | 0,3202 | 2,7131 | 2,4914 | 2,0000 |
| Cluster 4 | 0,3235 | 2,5463 | 2,5756 | 1,0000 |

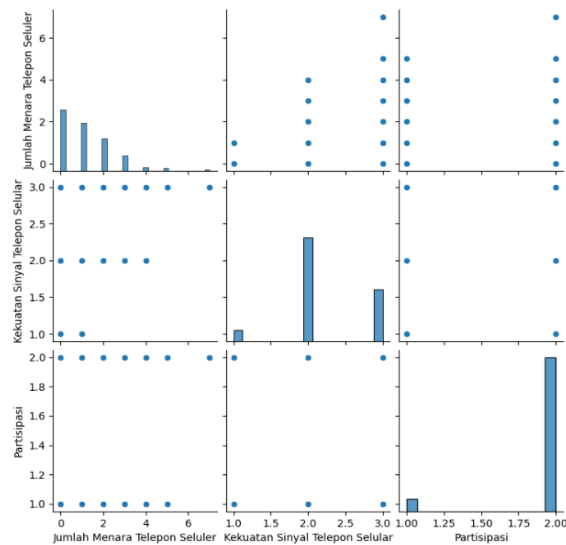
Pada tahapan yang berlangsung ketika hasil sudah mengalami kesamaan maka perhitungan dihentikan pada iterasi ke 7.

Dari hasil centroid baru yang telah dilakukan, penelitian tersebut menunjukkan tingkat variabel yang paling berpengaruh paling tinggi terhadap partisipasi pemilu dalam demografi Kabupaten Pasuruan. Penelitian ini dilakukan dengan berulang kali pengujian dengan rata - rata iterasi berhenti pada iterasi ke 7.

F. Implementasi Dengan Google Colaboratory

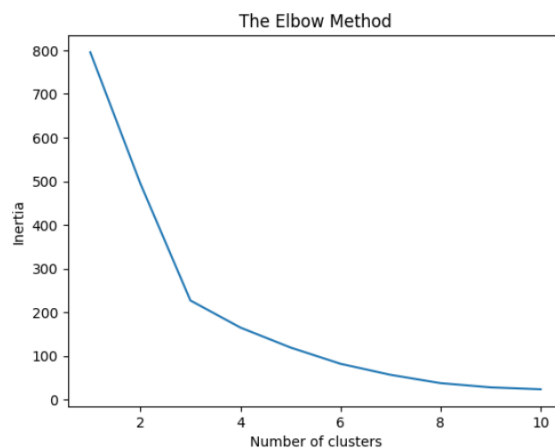
Berikut adalah hasil dari implementasi dari google colaboratory yang ditampilkan pada gambar dibawah ini :

Pada Gambar 2 menampilkan Seaborn plot atau visualisasi data yang menggambarkan hubungan atau correlation jumlah menara telepon seluler, kekuatan sinyal telepon seluler, dan tingkat partisipasi pemilu di Kabupaten Pasuruan.



Gambar 2. Seaborn plot atau visualisasi data

Pada Gambar 3 menyajikan grafik nilai sum of squares untuk berbagai jumlah cluster. Titik siku pada grafik menunjukkan jumlah cluster optimal.



Gambar 3. Hasil metode elbow atau sum of squares

Pada Gambar 4 menampilkan hasil Uji Akurasi K - Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan Menggunakan K-Means dengan 3 cluster. Pada Gambar 5 menampilkan hasil Uji Akurasi K - Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan Menggunakan K-Means dengan 4 cluster.

```
Inertia: 227.18931062726043
The silhouette score is 0.6410339736868782
```

Gambar 3. Hasil inerti dan silhouette dengan cluster 3

```
Inertia: 165.0776963725261
The silhouette score is 0.6783345304898949
```

Gambar 4. Hasil inerti dan silhouette dengan cluster 4

Dengan hasil uji diatas yang dilakukan dari implementasi manual serta implementasi dengan google colaboratory dengan menggunakan 4 variabel menghasilkan sebuah pernyataan cluster yang optimal serta hasil prediksi partisipasi masyarakat dalam pemilu sebesar 64 % partisipasi pemilu dengan 3 cluster sebagai perbandingan dan petualangan penelitian dengan 4 cluster sebanyak 68 % partisipasi masyarakat dalam pemilu.

Pada penelitian ini menemukan bahwa algoritma K - Means dapat digunakan untuk memprediksi partisipasi pemilu. Penelitian ini memberikan kontribusi penting dalam peningkatan partisipasi berikutnya terhadap Komisi Pemilihan Umum (KPU). Hasil penelitian ini juga dapat menjadi dasar bagi Badan Pusat Statistik (BPS) untuk meningkatkan kualitas dan keakuratan dalam penerbitan suatu data agar sesuai dengan yang ada di lapangan.

V. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dalam Uji Akurasi K - Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan. Hasil analisis Penelitian ini menemukan bahwa setelah memplot jumlah cluster terhadap nilai inerti (within-cluster sum of squares), terlihat bahwa grafik menunjukkan titik siku (elbow) pada jumlah cluster 3. Ini menunjukkan bahwa 3 cluster adalah pilihan yang tepat untuk dataset tersebut. Namun, penelitian ini juga mencoba menggunakan 4 cluster yang disebut (k – means model with different clusters) dan memberikan wawasan lebih dalam tentang struktur data. Dalam memprediksi partisipasi pemilu penggunaan 4 cluster menghasilkan tingkat akurasi partisipasi masyarakat dalam pemilu sebesar 68%, sementara 3 cluster menghasilkan tingkat akurasi partisipasi masyarakat dalam pemilu sebesar 64%. Hasil penelitian ini yang sudah dilakukan, maka dapat disimpulkan dalam memprediksi partisipasi pemilu dalam demografi wilayah Kabupaten Pasuruan yang dilakukan dengan menggunakan Algoritma K – K - Means dengan 4 Variabel yaitu sebanyak 64 % partisipasi pemilu dengan 3 cluster sebagai perbandingan dan petualangan penelitian dengan 4 cluster sebanyak 68 % partisipasi masyarakat dalam pemilu.

UCAPAN TERIMA KASIH

Peneliti ingin menyampaikan penghargaan setinggi-tingginya kepada orang tua dan keluarganya yang senantiasa memberikan sokongan sepenuhnya dalam menjalankan riset ini. Serta tidak lupa mengucapkan terima kasih kepada pihak kampus yang telah menyediakan sarana dan fasilitas penelitian. Segala bentuk kontribusi dan dukungan yang diberikan kepada peneliti memiliki nilai yang sangat penting untuk mencapai kesuksesan penelitian ini. Peneliti mengucapkan terima kasih atas doa dan dukungan penuh yang telah diterima selama proses penelitian.

REFERENSI

- [1] [1] I. P. A. P. Wibawa, I. K. A. Purnawan, D. P. S. Putri, and N. K. D. Rusjyanthi, "Prediksi Partisipasi Pemilih dalam Pemilu Presiden 2014 dengan Metode Support Vector Machine," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, vol. 7, no. 3, pp. 182, 2019, doi: 10.24843/jim.2019.v07.i03.p02.
- [2] [2] K. A. Pratama, A. Zarkasi, and Ansorullah, "Analisis Pengaturan Perlengkapan Pemungutan Suara Pemilu Ditinjau Dari Undang-Undang Tentang Pemilihan Umum," *Limbago J. Const. Law*, vol. 3, no. 2, pp. 293–309, 2023.
- [3] [3] I. Kelibay et al., "Sosialisasi Politik Sebagai Upaya Meningkatkan Partisipasi Masyarakat Menjelang Pemilu Serentak Tahun 2024," *J. Masy. Madani Indones.*, vol. 2, no. 4, pp. 442–449, 2023, doi: 10.59025/js.v2i4.155.
- [4] [4] A. Handayani et al., "Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM," *J. Inf. Syst. Res.*, vol. 5, no. 1, pp. 53–63, 2023, doi: 10.47065/josh.v5i1.4379.
- [5] [5] M. Simanjuntak, N. Nurfalinda, and ..., "Penerapan Metode Naive Bayes Untuk Memprediksi Status Kehadiran Masyarakat Dalam Pemilihan Gubernur," *Student Online J. ...*, pp. 152–162, 2022, [Online]. Available: <https://soj.umrah.ac.id/index.php/SOJFT/article/view/1576%0Ahttps://soj.umrah.ac.id/index.php/SOJFT/article/download/1576/1398>
- [6] [6] S. D. Hilda, A. Voutama, and Y. Umidah, "Analisis Daftar Pemilih Tetap Pemilihan Gubernur dan Wakil Gubernur menggunakan Algoritma K-Means," *JATISI (Jurnal Tek. ...)*, vol. 10, no. 3, pp. 398–408, 2023, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/4921%0Ahttps://jurnal.mdp.ac.id/index.php/jatisi/article/download/4921/1600>
- [7] [7] A. Pambudi, "Penerapan Crisp-Dm Menggunakan Mlr K-Fold Pada Data Saham Pt. Telkom Indonesia (Persero) Tbk (Tlkm) (Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022)," *J. Data Min. dan Sist. Inf.*, vol. 4, no. 1, p. 1, 2023, doi: 10.33365/jdmsi.v4i1.2462.
- [8] [8] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 8, no. 1, p. 56, 2023, doi: 10.30998/string.v8i1.16388.

- [9] [9] M. M. abdoel Wahid, "Determining The Location Of RMU, Using K-Means Clustering, Evaluate The Location Of Existing RMU, Using R-Programming," J. Informatics Telecommun. Eng., vol. 6, no. 1, pp. 10–17, 2022, doi: 10.31289/jite.v6i1.6126.
- [10] [10] R. Risawandi and Y. Afrillia, "Geographic Information System Mapping Of Criminality Villed Areas In Lhokseumawe Using K-Means Method," J. Informatics Telecommun. Eng., vol. 5, no. 2, pp. 442–451, 2022, doi: 10.31289/jite.v5i2.6265.
- [11] [11] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar)," J. Nas. Teknol. dan Sist. Inf., vol. 9, no. 2, pp. 122–131, 2023, doi: 10.25077/teknosi.v9i2.2023.122-131.
- [12] [12] A. S. Fitriani, F. Fajrillah, and W. Novarika, "Implementation of Data Mining Using Naïve Bayes Classification Method To Predict Participation of Governor And Vocational Governor Selection In Jemirahan Village, Jabon District," IJICS (International J. Informatics Comput. Sci., vol. 3, no. 2, p. 66, 2019, doi: 10.30865/ijics.v3i2.1391.
- [13] [13] A. S. Fitriani, "JTAM (Jurnal Teori dan Aplikasi Matematika) Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur," vol. 3, no. 2, pp. 98–104, 2019, doi: 10.31764/jtam.v3i2.995.
- [14] [14] M. N. Zarti, E. Sahputra, A. Sonita, and Y. Apridiansyah, "Application Of Data Mining Using The Naïve Bayes Classification Method To Predict Public Interest Participation In The 2024 Elections," J. Komputer, Inf. dan Teknol., vol. 3, no. 1, pp. 105–114, 2023, doi: 10.53697/jkomitek.v3i1.1192.
- [15] [15] D. E. Safitri and A. S. Fitriani, "Implementasi Metode Klasifikasi Dengan Algoritma Support Vector Machine Kernel Gaussian Rbf Untuk Prediksi Partisipasi Pemilu Terhadap Demografi Kota Surabaya," Indones. J. Bus. Intell., vol. 5, no. 1, p. 36, 2022, doi: 10.21927/ijubi.v5i1.2259.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.