

Implementasi Algoritma Naive Bayes Pada Demografi kabupaten Gresik Untuk Mengukur Partisipasi Pemilu

Oleh:

Hawwani Muhdat Azzahrah
Program Studi Informatika

Universitas Muhammadiyah Sidoarjo
2023

Pendahuluan



Pemilu di indonesia



Golput dalam pemilu



Membantu Pemerintahan

Pendahuluan

- Berdasarkan permasalahan yang ada peneliti berharap penelitian ini dapat menjadi tumpuan pemerintah dalam mengatasi masalah kesadaran masyarakat tentang hak-hak pemilu dan memiliki harapan bahwa hak suara yang mereka pilih tidak akan sia-sia.
- Pada pengujian kali ini bertuju pada implementasi Data Mining untuk mengetahui jumlah atau tingkat partisipasi Pemilu Kabupaten Gresik menggunakan Metode klasifikasi Naive Bayes, sehingga nantinya akan mendapatkan menjadi informasi pada pemerintah atau pihak yang bersangkutan yang dapat menjadikan panduan tentang kesadaran masyarakat dengan hak suara golput.

Pertanyaan Penelitian (Rumusan Masalah)

1. Bagaimana cara untuk memprediksi Tingkat partisipasi Pemilu Tahun 2021 pada Kabupaten Gresik, Jawa Timur?
2. Bagaimana proses klasifikasi dengan Metode Naive Bayes berdasarkan data Pemilu terhadap partisipasi Pemilu Kabupaten Gresik?

Batasan Masalah

1. Menggunakan data BPS (Badan Pusat Statistik) Kabupaten Gresik 2021
2. Data Pemilu tahun 2021 pada halaman web KPU
3. Menggunakan metode klasifikasi Naive Bayes
4. Data pemilu KPU Kabupaten Gresik sebanyak 3,738 pada tahun 2021

Tujuan Penelitian

untuk mengetahui jumlah atau tingkat partisipasi Pemilu Kabupaten Gresik menggunakan Metode dengan klasifikasi Naive Bayes, sehingga nantinya akan mendapatkan hasil yang kemudian akan menjadi informasi pada pemerintah atau pihak yang bersangkutan yang dapat menjadikan panduan tentang kesadaran masyarakat dengan hak suara golput.

Manfaat Penelitian

manfaat yaitu memberikan informasi serta pengetahuan dari hasil yang nantinya akan di Klasifikasi, sehingga dapat menjadi acuan untuk meminimalisir tingkat Golput pada suatu daerah.

Landasan Teori

- Pemilu : Adalah proses pemungutan suara dari pilihan rakyat untuk memilih seseorang menjadi pemimpin suatu politik atau wakilnya dalam pemerintahan.
- Data Mining : “Data mining adalah proses penguraian data dari berbagai sumber dan mengubahnya menjadi informasi atau data atau model yang berarti untuk meningkatkan profit, mengurangi anggaran, atau keduanya” (Witten, 2016)

Landasan Teori

- Klasifikasi : Diartikan sebagai kegiatan pengelompokan suatu pustaka berdasarkan ciri yang sama atas dasar tertentu kemudian di tempatkan secara bersama pada suatu tempat (Bafadal, 2009:51).
- Naïve Bayes : Naive Bayes merupakan metode untuk mempelajari probabilitas suatu objek yang memiliki ciri tertentu dari peninjauan sekumpulan data dengan menjumlahkan kombinasi dari dataset untuk menemukan hubungan dan meringkas data yang tidak di duga.

Landasan Teori

Teorema metode naïve bayes :

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)}$$

Keterangan :

X : data dengan klas yang masih belum ditemukan

H : hipotesa data X adalah suatu class yang tertentu

P(H | X) : Probabilitas hipotesa H untuk kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesa H (Prior probabilitas)

P(X | H) : Probabilitas X dalam kondisi hipotesis H

P(X) : Probabilitas

Metode Penelitian

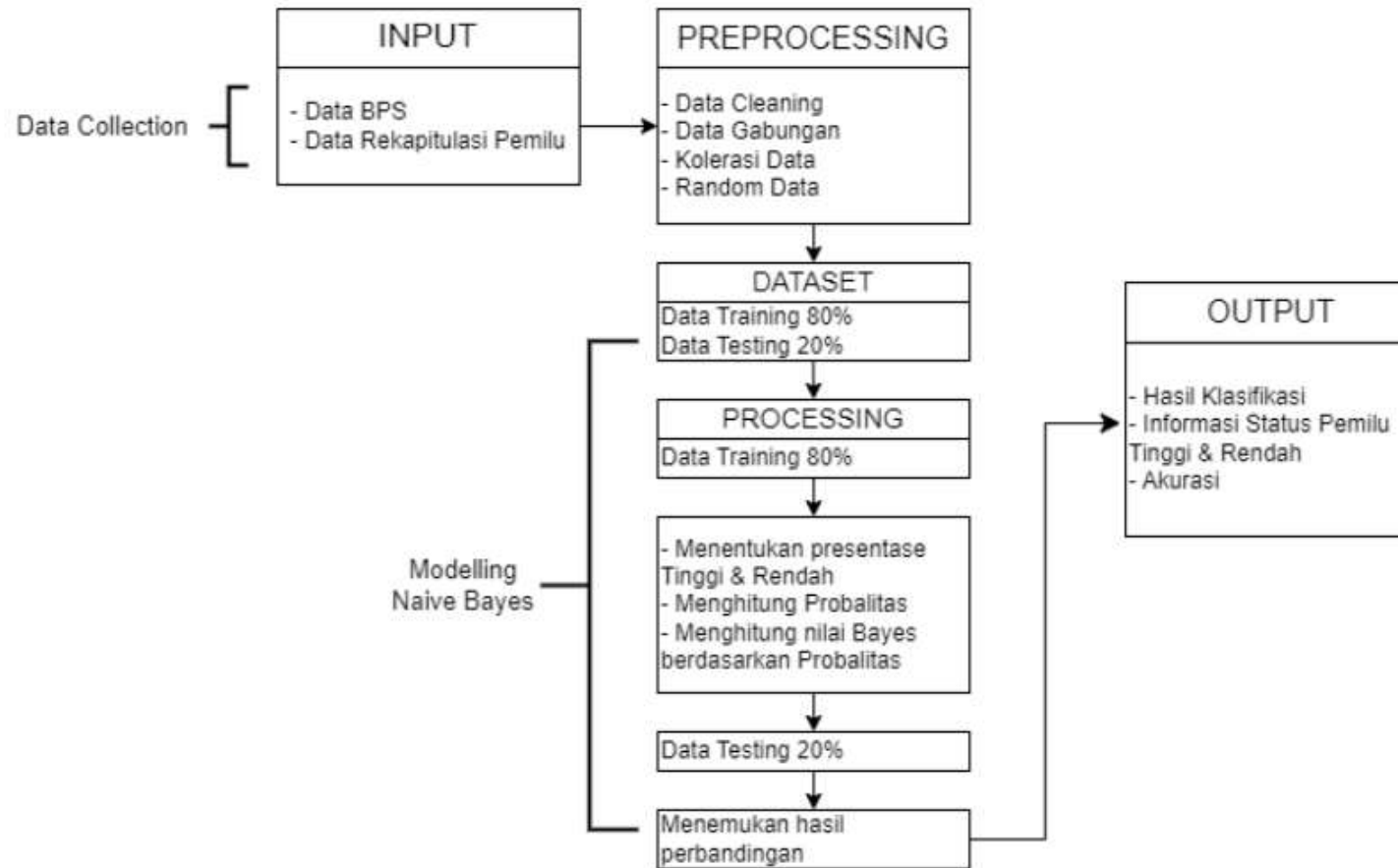


Diagram Klasifikasi

Metode Penelitian

- Pengumpulan data :

1. Studi Literatur Pengumpulan data menggunakan metode studi kepustakaan ini melalui pencarian dari jurnal, buku, internet, dan literatur lainnya yang masih berkaitan dengan penelitian ini.
2. Observasi Metode pengumpulan data dilakukan dengan pencarian dilaman web KPU dan BPS. Kemudian dari data BPS yang tersedia data kemudian didownload, setelah itu untuk pengumpulan data KPU data yang diambil 10 yaitu TPS, DPT, dan Hak Pilih. Informasi yang diperoleh dapat diterapkan dalam penelitian yang dilakukan

Metode Penelitian

- Pada penelitian ini terdapat sebanyak 18 atribut yang diambil dari data BPS dan 3 atribut dari kota Gresik. Atribut memerlukan adanya seleksi lagi, karena ada beberapa komponen yang ada pada tribute memiliki peran penting dalam studi ini

No.	Atribut	Skala	Klarifikasi
1	Sistem Peringatan Dini Bencana Alam	Kategorikal	Ada, Tidak Ada
2	Sistem Peringatan Dini Khusus Tsunami	Kategorikal	Wilayah Berpotensi Tsunami, Bukan Wilayah Potensi Tsunami, Tidak Ada
3	Perlengkapan Keselamatan	Kategorikal	Ada, Tidak Ada
4	Rambu-rambu dan Jalur Evakuasi Bencana Alam	Kategorikal	Ada, Tidak Ada
5	Pembuatan, Perawatan, atau Normalisasi : Sungai, Kanal, Tanggul, Parit, Drainase, Waduk, Pantai, dll	Kategorikal	Ada, Tidak Ada
6	Minimarket/Swalayan	Kategorikal	Banyak, Sedikit
7	Toko/Warung Kelontong	Kategorikal	Banyak, Sedikit
8	Warung/Kedai Makan	Kategorikal	Banyak, Sedikit
9	Koperasi	Kategorikal	Banyak, Sedikit
10	Jumlah Menara Telepon Seluler (BTS)	Kategorikal	Banyak, Sedikit
11	Jumlah Operator Layanan Komunikasi Telepon Seluler yang Menjangkau di Desa/Kelurahan	Kategorikal	Banyak, Sedikit
12	Kondisi Sinyal Telepon Seluler di Sebagian Wilayah Desa/Kelurahan	Kategorikal	Sangat Kuat, Sinyal Kuat, Sinyal Lemah, Tidak Ada
13	Jenis Transportasi	Kategorikal	Darat, Air, Darat dan Air, Udara
14	Keberadaan Angkutan Umum	Kategorikal	Ada, Dengan trayek Tetap. Ada, Tanpa Trayek Tetap Tidak Ada Angkutan Umum
15	Jenis Permukaan Jalan	Kategorikal	Aspal/Beton, Diperkeras (Kerikil, Batu, dll)
16	Dapat Dilalui Kendaraan Roda 4 atau Lebih	Kategorikal	Sepanjang Tahun, Sepanjang Tahun Kecuali Kondisi Tertentu
17	Kantor Pos/ Pos Pembantu/ Rumah Pos	Kategorikal	Beroperasi, Jarang Beroperasi, Tidak Beroperasi, Tidak ada
18	Perusahaan/Agen jasa Ekspedisi Swasta	Kategorikal	Beroperasi, Jarang Beroperasi, Tidak ada

Metode Penelitian

Data Training



Probabilitas Tinggi

$$P(H|data) = \frac{\text{Jumlah data target tinggi}}{\text{Jumlah data target}} = \frac{76}{80} = 0,95$$

Probabilitas Rendah

$$P(H|data) = \frac{\text{Jumlah data target rendah}}{\text{Jumlah data target}} = \frac{4}{80} = 0,05$$

Metode Penelitian

Data testing



Metode Penelitian

Source code Naïve Bayes dengan python

```
# Build a Gaussian Naive Bayes Classifier
model = GaussianNB()

# Model training
model.fit(data_train, target_train)

# Make predictions on the test data
y_pred = model.predict(data_test)

# Calculate accuracy and F1 score
accuracy = accuracy_score(target_test, y_pred)
f1 = f1_score(target_test, y_pred)

# Append metrics to lists
accuracies.append(accuracy)
f1_scores.append(f1)

# Print the results for this iteration
print("Accuracy:", accuracy)
print("F1 Score:", f1)
print(classification_report(target_test, y_pred))
print("\n")

# Decrement the ratio for the next iteration
ratio -= 0.0125

# Print the average metrics across iterations
print("Average Accuracy:", sum(accuracies) / len(accuracies))
print("Average F1 Score:", sum(f1_scores) / len(f1_scores))
```


Hasil dan Kesimpulan

- Pada penelitian ini 3,738 data data berhasil dikumpulkan. Dengan 18 atribut yang berhasil diseleksi dari Data KPU dan BPS kota Gresik pada tahun 2021. Pada Iterasi 20 Ratio terbaik untuk perbandingan data train dan data test sebanyak 83,75% : 16,25%. Dan hasil akurasi yang didapatkan sebanyak 62,84%. Beberapa faktor yang mempengaruhi akurasi adalah data cacat yang dapat menimbulkan gangguan, ketidakakuratan, sehingga melemahkan kemampuan model.

