

## Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks Pembangunan Manusia Di Provinsi Jawa Timur

Muhammad Arifat<sup>1</sup>, Wardiana Adinda Putri<sup>2\*</sup>, Alfin Syairotin M.<sup>3</sup>

<sup>1,2,3</sup> Statistika, Universitas Nahdlatul Ulama Sunan Giri  
[arifatm169@gmail.com](mailto:arifatm169@gmail.com)<sup>1</sup>, [ardianaadinda0025@gmail.com](mailto:ardianaadinda0025@gmail.com)<sup>2\*</sup>,  
[syairotinalfin@gmail.com](mailto:syairotinalfin@gmail.com)<sup>3</sup>

*Diajukan* 16 Mei 2023 *Diperbaiki* 27 Juni 2023 *Diterima* 30 Juni 2023

### Abstrak

**Latar Belakang:** IPM adalah alat ukur pencapaian kualitas hidup suatu negara yang terdiri atas tiga dimensi, yaitu: kesehatan, pengetahuan, dan hidup layak. Terdapat variasi IPM yang cukup signifikan antara kota dan kabupaten. Untuk mengatasi permasalahan ini, perlu adanya klasifikasi IPM di Jawa Timur sebagai acuan pemerataan di seluruh wilayah Jawa Timur.

**Tujuan :** Mendapatkan hasil klasifikasi IPM di Jawa Timur menggunakan metode *Naive Bayes Classifier* (NBC).

**Metode :** Digunakan metode kuantitatif dengan metode NBC dan software Jupyter Notebook untuk mengklasifikasikan data IPM skala nominal yang didapatkan dari BPS Provinsi Jawa Timur. Faktor-faktor yang dianalisis meliputi Pendapatan Per kapita, Angka Harapan Hidup, Harapan Lama Sekolah, Rata-rata Lama Sekolah, Produk Domestik Regional Bruto, Penduduk Miskin, Jumlah Fasilitas Kesehatan, dan Jumlah Tenaga Kesehatan dengan skala rasio.

**Hasil:** Metode klasifikasi NBC berhasil dipakai untuk memprediksi IPM di Jawa Timur. Data *training* dan *testing* yang optimal dengan pembagian 70% dan 30% menghasilkan akurasi 91,6%. Dari 12 data *testing*, model dapat memprediksi IPM dengan keakuratan 92% dan sensitivitas yang baik pada kelas Sangat Tinggi dan Tinggi.

**Kesimpulan:** Disimpulkan bahwa prediksi IPM di Provinsi Jawa Timur cukup akurat dengan persentase keakuratan mencapai 92%. Model juga memiliki nilai *recall* yang baik pada kelas Sangat Tinggi dan Tinggi serta cukup pada kelas Sedang.

**Kata kunci:** IPM, *Naive Bayes Classifier*, Kuantitatif, Jawa Timur

### Abstract

**Background:** HDI is a measure of the quality of life of a country that consists of three dimensions: health, knowledge, and decent living. There are significant variations in HDI between cities and districts. For the problem, it is necessary to classify HDI in East Java as a reference for equalization throughout East Java.

**Objective:** To obtain HDI classification results in East Java using the Naive Bayes Classifier (NBC) method.

**Methods:** Quantitative methods are used with the NBC method and Jupyter Notebook software to classify nominal scale HDI data come from BPS East Java Province. The factors analyzed include Per capita income Average Years of Schooling, Expected Years of Schooling, Life Expectancy, Gross Regional Domestic Product, Poor Population, Number of Health Facilities, and Number of Health Workers with a ratio scale.

**Results:** The NBC classification method successfully predicted HDI in East Java. Optimal training and testing data with a division of 70% and 30% resulted in an accuracy of 91.6%. From 12 testing data, the model can predict HDI with 92% accuracy and good sensitivity in the Very High and High classes.

**Conclusion:** It is concluded that the prediction of HDI in East Java Province is entirely accurate, with an accuracy percentage of 92%. The model also has a good recall value in the Very High and High classes and is sufficient in the Medium category.

**Keywords:** HDI, Naive Bayes Classifier, Quantitative, East Java

## PENDAHULUAN

Indeks Pembangunan Manusia (IPM) merupakan suatu alat ukur yang digunakan untuk menentukan kualitas hidup di suatu negara yang mana terdiri dari tiga dimensi utama, yaitu dimensi pengetahuan, dimensi kehidupan layak, dan dimensi kesehatan (BPS, 2020). IPM merupakan salah satu tolak ukur pembangunan pada kualitas sumber daya manusia (Hasibuan et al., 2020).

Pendapatan per kapita adalah hasil dari membagi pendapatan nasional suatu negara dengan jumlah penduduknya pada periode tertentu (Hapsari, 2019). Tujuan dari pengukuran pendapatan per kapita adalah untuk mengevaluasi tingkat kesejahteraan atau standar hidup suatu negara dari tahun ke tahun, serta menilai kemajuan pembangunan dan dampaknya terhadap masyarakat. Oleh karena itu, pendapatan per kapita dapat menjadi indikator kemajuan yang dicapai dan akibat dari peningkatan tersebut. Variabel pendapatan per kapita merupakan indikator yang termasuk dalam dimensi kehidupan layak

Angka Harapan Hidup (AHH) merupakan salah satu indikator untuk menilai kesehatan dan kualitas hidup penduduk (Hapsari, 2019). Menurut definisi dari Badan Pusat Statistik (BPS) Indonesia, AHH adalah rata-rata umur yang diharapkan akan dicapai oleh bayi yang baru lahir pada tahun tertentu. Artinya, AHH dapat menggambarkan perkiraan rata-rata usia individu dari saat lahir pada suatu tahun tertentu (Hapsari, 2019). AHH termasuk dalam dimensi kesehatan.

Harapan Lama Sekolah (HLS) merupakan suatu indikator yang berguna untuk mengevaluasi tingkat kesetaraan pembangunan pendidikan di Indonesia. HLS mengukur kesempatan individu

untuk memperoleh pendidikan, dimulai dari usia tujuh tahun. Sehingga, HLS dapat memberikan gambaran mengenai kesetaraan akses pendidikan di berbagai wilayah di Indonesia bagi masyarakatnya (Putera, 2020). Variabel HLS termasuk ke dalam dimensi pengetahuan

Rata-rata lama sekolah dapat digunakan sebagai indikator tingkat pendidikan formal yang dicapai oleh penduduk suatu daerah. Semakin tinggi rata-rata lama sekolah, semakin tinggi jenjang pendidikan yang telah dicapai oleh masyarakat (Putera, 2020). Rata-rata lama sekolah dihitung dengan menjumlahkan jumlah tahun yang dihabiskan oleh penduduk yang berusia 25 tahun ke atas untuk menyelesaikan semua jenjang pendidikan formal yang pernah diikuti, kemudian hasilnya dibagi dengan jumlah populasi yang diamati (Rafiqi, 2020). Rata-rata lama sekolah juga variabel yang masuk dalam dimensi pengetahuan.

Produk Domestik Regional Bruto (PDRB) adalah nilai total dari produk akhir barang dan jasa yang dihasilkan oleh berbagai sektor ekonomi di suatu wilayah dalam jangka waktu tertentu (Cholili, 2014). PDRB dapat digunakan untuk menunjukkan kemampuan suatu wilayah dalam memanfaatkan sumber daya alam yang ada di wilayah tersebut. Variabel PDRB merupakan indikator yang termasuk dalam dimensi kehidupan layak.

Fasilitas kesehatan merujuk pada tempat di mana berbagai layanan kesehatan diberikan, mulai dari pelayanan promotif, pencegahan, pengobatan, hingga rehabilitasi (Hikmah, 2022). Variabel fasilitas kesehatan masuk dalam dimensi kesehatan

Tenaga kesehatan adalah individu yang bekerja secara sukarela di bidang kesehatan dan telah mendapatkan pengetahuan dan keterampilan melalui pendidikan kesehatan. Beberapa jenis tenaga kesehatan memiliki kewenangan untuk memberikan layanan kesehatan yang dibutuhkan (Fadhillah, 2018).

Variabel tenaga kesehatan masuk dalam dimensi kesehatan.

Pada penelitian terdahulu, IPM dapat diklasifikasikan dengan Metode *K-Nearest Neighbor* (k-NN) dan Support Vector Machine dengan atribut Umur Harapan Hidup, Jumlah Pengangguran, Produk Domestik Regional Bruto, dan Rata-Rata Lama Sekolah (Pratiwi & Wijayanto, 2019). Pada penelitian lainnya klasifikasi IPM dikukan dengan atribut lain seperti Pengeluaran perkapita, Rata-rata Lama Sekolah, dan Angka Harapan Hidup, Harapan Lama Sekolah (Fauzi et al., 2017). Hal yang hampir sama juga dilakukan pada penelitian yang lain terkait IPM ini digunakan klasifikasi iPM dengan atribut Pengeluaran Per kapita yang Disesuaikan, Harapan Lama Sekolah, Rata-rata Lama Sekolah dan Angka Harapan Hidup (Fauzi, 2017).

Tingkat IPM Provinsi di Pulau Jawa yang termasuk dalam kategori sedang adalah Provinsi Jawa Timur (Joko Ade Nursiyono, 2021). Namun, jika dilihat dari tingkat Kabupaten/Kota di Provinsi Jawa Timur, maka tingkat IPM dapat dipastikan beragam. Sebagai contoh pada tahun 2021, IPM untuk Kabupaten Sampang adalah 62,8%, sedangkan IPM untuk Kota Surabaya mencapai 82,31%. Di samping itu, selisih nilai IPM antara daerah dengan nilai IPM terendah dan tertinggi juga sangat besar. Dari beberapa sumber ditunjukkan bahwa pembangunan di Provinsi Jawa Timur cenderung terfokus pada kota-kota besar (Budiman & Cahyani, 2022). Oleh karena itu, diperlukan klasifikasi nilai Indeks Pembangunan Manusia untuk mengetahui akurasi nilai tersebut.

Salah satu metode klasifikasi yang dapat digunakan adalah metode Naive Bayes. Metode ini berasal dari teorema Bayes yang merupakan metode klasifikasi yang berbasis statistik dan probabilitas.

Teorema Bayes digunakan untuk memprediksi peluang di masa depan yang mana metode ini diusulkan oleh Thomas Bayes (Ramaulidyah et al., 2021). Keuntungan dari metode ini adalah hanya diperlukan ukuran sampel atau data yang tidak besar untuk melakukan proses klasifikasi dengan tujuan prediksi. Selain itu, Metode ini memiliki keunggulan dalam hal kecepatan dan akurasi yang tinggi untuk permasalahan yang kompleks dan data cukup besar, sehingga diperoleh nilai error yang rendah (Putri et al., 2014).

Selain itu, penelitian terdahulu berkaitan tentang klasifikasi IPM yaitu "Klasifikasi Indeks Pembangunan Manusia (IPM) Multi-Class Dengan Metode Artificial Neural Network (ANN)" oleh (Fathurrahman & Qisthi, 2021) menghasilkan perbandingan klasifikasi antara metode ANN dan Support Vector Machine (SVM) yang mana akurasi metode ANN sebesar 97,4% dan 53,25% untuk metode SVM. Yang berarti metode ANN lebih baik dibandingkan SVM. Sehingga berdasarkan penelitian tersebut, maka pada penelitian ini nantinya menggunakan metode Naive Bayes dengan studi kasus yang sama yaitu IPM bertujuan untuk melihat nilai akurasi dan membandingkannya dengan penelitian terdahulu apakah nilai akurasi lebih tinggi dibandingkan metode ANN atau lebih rendah dibandingkan SVM.

Sebelum dilakukan klasifikasi, perlu dianalisis terlebih dahulu terkait statistik deskriptif untuk memahami karakteristik data IPM di Jawa Timur dan variabel yang diduga mempengaruhinya. Langkah selanjutnya adalah digunakan metode Naive Bayes Classifier (NBC) untuk melakukan klasifikasi IPM di Jawa Timur. Dengan demikian, hasil klasifikasi IPM di Jawa Timur dengan menggunakan NBC akan memberikan gambaran tentang klasifikasi tingkat IPM di Jawa Timur beserta atribut-atribut yang berpengaruh.

## METODE

### Desain penelitian

Digunakan pendekatan kuantitatif dalam desain penelitian ini. Metode statistik yang diusulkan adalah Naive Bayes Classifier (NBC). Proses klasifikasi IPM di Jawa Timur dengan menggunakan Metode NBC bertujuan untuk mengetahui berapa persen nilai akurasi klasifikasi IPM dengan memanfaatkan. Software Jupiter Notebook. Pada penelitian ini, digunakan data sekunder berupa data IPM beserta variabel-variabel yang diduga mempengaruhinya yang data pengamatan berasal Badan Pusat Statistik (BPS) Jawa Timur.

### Populasi dan sampel

Populasi yang digunakan pada penelitian ini digunakan IPM yang meliputi 3 komponen yaitu dimensi kesehatan, dimensi pengetahuan, dan dimensi hidup layak. Sampel yang digunakan adalah IPM Jawa Timur pada tahun 2021. Unit penelitian yang digunakan yaitu IPM dan variabel-variabel yang mempengaruhinya untuk tingkat Kabupaten/Kota di Jawa Timur, meliputi Tulungagung, Tuban, Trenggalek, Surabaya, Sumenep, Situbondo, Sidoarjo, Sampang, Probolinggo, Ponorogo, Pasuruan, Pamekasan, Pacitan, Ngawi, Nganjuk, Mojokerto, Malang, Magetan, Madiun, Lumajang, Lamongan, Kediri, Jombang, Jember, Gresik, Bondowoso, Bojonegoro, Blitar, Batu, Banyuwangi, dan Bangkalan.

### Teknik Pengambilan Sampel

Digunakan teknik sampling yaitu purposive sampling dalam penelitian ini dengan mengambil data di Website resmi

BPS Jawa Timur sesuai dengan tujuan dari penelitian.

### Subyek Penelitian

Subyek penelitian yang digunakan adalah Nilai Indeks Pembangunan Manusia beserta variabel prediktornya yang mana disajikan pada Tabel 1, sebagai berikut:

**Tabel 1.** Definisi Variabel Penelitian

Variabel	Nama Variabel	Skala
Y	Indeks Pembangunan Manusia (IPM)	Nominal (1= Sangat Tinggi, 2= Tinggi, 3= Sedang)
X <sub>1</sub>	Pendapatan Perkapita	Rasio
X <sub>2</sub>	Angka Harapan Hidup	Rasio
X <sub>3</sub>	Harapan Lama Sekolah	Rasio
X <sub>4</sub>	Rata-rata Lama Sekolah	Rasio
X <sub>5</sub>	Produk Domestik Regional Bruto (PDRB)	Rasio
X <sub>6</sub>	Penduduk Miskin	Rasio
X <sub>7</sub>	Jumlah Fasilitas Kesehatan	Rasio
X <sub>8</sub>	Jumlah Tenaga Kesehatan	Rasio

### Teknik Analisis Data

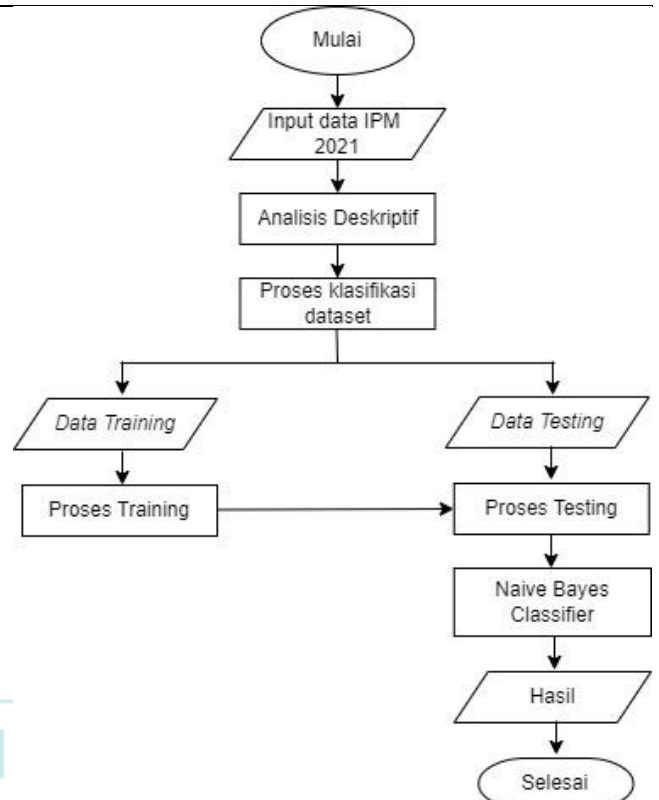
Prosedur yang diterapkan untuk analisis data pada penelitian diberikan secara bertahap sebagai berikut

1. Melakukan analisis deskriptif pada data IPM Jawa Timur pada tahun 2021 untuk memperoleh gambaran umum IPM dan variabel-variabel yang mempengaruhinya.

## Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks....

2. Membagi data pengamatan menjadi dua bagian untuk proses training proses testing dengan melihat tingkat akurasi yang tertinggi, diperoleh pembagian data sebesar 70% data Training dan 30 % data Testing.
3. Membangun model klasifikasi dengan algoritma NBC dari data training dengan langkah-langkah sebagai berikut (Mashfia, 2022).
  - 3.1. Menghitung nilai probabilitas prior dari data training, yaitu probabilitas IPM yang bernilai Sedang, Tinggi, dan Sangat Tinggi.
  - 3.2. Menghitung nilai probabilitas untuk setiap variabel dari data training sesuai kelas yang diberikan.
  - 3.3. Menggunakan data testing dengan mengalikan probabilitas prior dengan variabel pada masing-masing kelas sesuai probabilitas dari data training.
  - 3.4. Mencari nilai probabilitas maksimal pada setiap kelas.
4. Menghitung tingkat akurasi pada proses klasifikasi dengan ukuran jumlah kesalahan (APER), akurasi, sensitifitas, dan spesifisitas.

Prosedur pada penelitian ini dapat digambarkan alurnya dengan diagram alur sebagaimana ditam[pilkan pada Gambar 1 berikut.

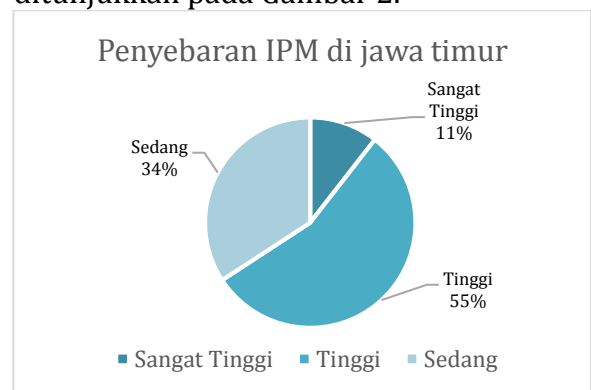


Gambar 1. Diagram Alir

## HASIL DAN PEMBAHASAN

### Statistika Deskriptif

Provinsi Jawa Timur terdiri dari 29 kabupaten dan 9 kota, yang mempunyai wilayah yang luas dan tingkat penyebaran IPM yang berbeda-beda, sehingga faktor spasial atau lokasi dapat mempengaruhi peningkatan IPM. IPM di Jawa Timur memiliki penyebaran yang sangat beragam, sebagaimana ditunjukkan pada Gambar 2.



Gambar 2. Penyebaran IPM di Jawa Timur



## Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks....

Nilai IPM di Provinsi Jawa Timur didominasi oleh IPM Tinggi yaitu sebesar 55%, sedangkan nilai IPM yang Sedang di Provinsi Jawa Timur yaitu 34%, serta sisanya termasuk IPM bernilai Sangat Tinggi hanya sebesar 11% .

Tabel 2 menunjukkan Statistik Deskriptif dari Variabel Pendapatan Perkapita di Provinsi Jawa Timur.

**Tabel 2.** Statistik Deskriptif Pendapatan Perkapita

Pendapatan Perkapita	Nilai
Mean	11.568.526
Standard Deviation	22.162.167
Minimum	8673
Maximum	17862

Berdasarkan tabel di atas, nilai rata-rata pendapatan per kapita di Provinsi Jawa Timur adalah sebesar Rp.11.568.526 dan standar deviasi variabel pendapatan per kapita adalah 2216,2167. Artinya, terdapat variasi yang cukup besar dalam pendapatan per kapita dalam sampel yang diambil. Standar deviasi yang tinggi menunjukkan bahwa data dalam sampel tersebar jauh dari nilai rata-rata, sehingga dapat disimpulkan bahwa pendapatan per kapita dalam populasi yang diwakili oleh sampel ini sangat heterogen atau tidak seragam. Kota Surabaya memiliki nilai pendapatan per kapita tertinggi sebesar Rp.17.862.000/orang/tahun, dipengaruhi oleh statusnya sebagai ibukota Provinsi Jawa Timur dan lokasinya yang strategis sebagai pusat sektor ekonomi dan pariwisata. Kota Surabaya memiliki banyak pabrik, UMKM, serta tempat wisata seperti Kawasan Kota Lama Surabaya dan Makam Sunan Ampel. Oleh karena itu, tidak heran jika Kota Surabaya memiliki pendapatan per kapita tertinggi di Provinsi Jawa Timur. Di sisi lain, Kabupaten Bangkalan memiliki nilai pendapatan per kapita terendah sebesar

Rp.8.673.000/orang/tahun, dikarenakan mayoritas penduduknya bekerja di sektor pertanian dan perikanan, serta infrastrukturnya yang masih terbelakang.

Statistik Deskriptif dari Variabel Angka Harapan Hidup di Provinsi Jawa Timur dapat dilihat pada Tabel 3.

**Tabel 3.** Statistik Deskriptif Angka Harapan Hidup

Umur Harapan Hidup (Tahun)	Nilai
Mean	71,6915
Standard Deviation	2,0148
Minimum	66,8900
Maximum	74,1800

Menurut tabel di atas, rata-rata Angka Harapan Hidup di Provinsi Jawa Timur adalah 71,69 tahun dengan standar deviasi 2,0148483, yang menunjukkan bahwa umur harapan hidup di populasi yang diamati relatif homogen. Kota Surabaya memiliki Angka Harapan Hidup tertinggi di Jawa Timur sebesar 74,18 Tahun, karena faktor-faktor seperti lingkungan, pangan, pendidikan, kebijakan pemerintah, dan perekonomian masyarakat yang baik. Di sisi lain, Kabupaten Bondowoso memiliki Angka Harapan Hidup terendah di Jawa Timur sebesar 66,89 tahun, disebabkan oleh faktor-faktor seperti produksi pangan yang menurun dan hama tanaman. Selain faktor ketahanan pangan, ada banyak faktor lain yang memengaruhi Angka Harapan Hidup di Kabupaten Bondowoso, yang membuatnya menjadi yang terendah di Provinsi Jawa Timur.

Statistik Deskriptif dari Variabel Harapan Lama Sekolah di Provinsi Jawa Timur dapat dilihat pada Tabel 4.

**Tabel 4.** Statistik Deskriptif Harapan Lama Sekolah

Harapan Lama Sekolah (Tahun)	Nilai
Mean	13,3989
Standard Deviation	0,9016
Minimum	11,7300
Maximum	15,7500

## Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks....

Berdasarkan tabel di atas, dapat dilihat bahwa rata-rata Harapan Lama Sekolah di Provinsi Jawa Timur adalah 13,3989 Tahun dan standar deviasinya sebesar 0,90168, menunjukkan variasi yang relatif kecil atau homogen dalam populasi yang diamati. Kota Malang memiliki Harapan Lama Sekolah tertinggi di Jawa Timur, dengan angka 15,75 tahun, sementara Kabupaten Bangkalan memiliki Harapan Lama Sekolah terendah di Jawa Timur, yaitu 11,73 tahun. Hal ini menunjukkan bahwa Kabupaten Bangkalan memiliki sistem pendidikan yang kurang baik dan faktor lain yang mempengaruhi Harapan Lama Sekolah. Harapan Lama Sekolah dapat digunakan untuk menilai kesetaraan akses pendidikan bagi masyarakat di berbagai wilayah di Indonesia.

Statistik Deskriptif dari Variabel Rata-rata Lama Sekolah di Provinsi Jawa Timur dapat dilihat pada Tabel 5.

Tabel 5 Statistik Deskriptif Rata-rata Lama Sekolah

Rata-rata Lama Sekolah (Tahun)	Nilai
Mean	8,0607
Standard Deviation	1,5765
Minimum	4,8600
Maximum	11,3700

Berdasarkan tabel di atas, terlihat bahwa rata-rata lama sekolah di Provinsi Jawa Timur adalah 8,06078947 tahun. Standar deviasi sebesar 1,57657751 menunjukkan bahwa variasi rata-rata lama sekolah di populasi yang diamati relatif kecil atau homogen. Artinya, mayoritas individu dalam populasi memiliki rata-rata lama sekolah yang mendekati nilai rata-rata atau median dari data.

Kota Madiun memiliki rata-rata lama sekolah tertinggi di Jawa Timur, yaitu 11,37 tahun. Hal ini menunjukkan bahwa mayoritas penduduk di kota Madiun telah menempuh pendidikan hingga setidaknya

tingkat pendidikan menengah atas atau setara. Di sisi lain, Kabupaten Sampang memiliki rata-rata lama sekolah terendah di Jawa Timur, yaitu hanya 4,86 tahun. Artinya, mayoritas penduduk di kabupaten Sampang hanya menempuh pendidikan hingga tingkat pendidikan dasar atau setara, dan hanya sedikit yang melanjutkan ke jenjang pendidikan menengah atau perguruan tinggi.

Statistik deskriptif dari variabel PDRB di Provinsi Jawa Timur dapat dilihat pada Tabel 6.

Tabel 6. Statistik Deskriptif PDRB

PDRB	Nilai
Mean	40.742.789
Standard Deviation	49.940.164
Minimum	13469
Maximum	300337

Berdasarkan Tabel 6, nilai rata-rata PDRB di Provinsi Jawa Timur adalah sebesar Rp.40.742.789.000. Dalam konteks umur harapan hidup, nilai standar deviasi sebesar 49.940.164. Standar deviasi yang tinggi menunjukkan adanya variasi yang cukup besar dalam pendapatan per kapita dalam sampel yang diambil. Hal ini menunjukkan bahwa PDRB dalam populasi yang diwakili oleh sampel ini sangat heterogen atau tidak seragam.

Kota Kediri memiliki sektor ekonomi yang paling berkembang dan memberikan kontribusi terbesar terhadap pertumbuhan ekonomi Jawa Timur dengan nilai PDRB sebesar Rp.300.337.000.000. Hal ini menunjukkan bahwa ada banyak aktivitas ekonomi yang terjadi di Kota Kediri dan banyak nilai tambah yang dihasilkan oleh berbagai sektor ekonomi di wilayah tersebut, seperti industri yang berkembang pesat, sektor pariwisata yang maju, atau pusat perdagangan yang ramai. Di sisi lain, Kabupaten Pamekasan menjadi kabupaten

## Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks....

dengan PDRB paling rendah di Provinsi Jawa Timur dengan nilai PDRB hanya sekitar Rp.13.469.000.000.

Statistik Deskriptif dari Variabel Penduduk Miskin di Provinsi Jawa Timur dapat dilihat pada Tabel 7 .

**Tabel 7.** Statistik Deskriptif Penduduk Miskin

Penduduk Miskin	Nilai
Mean	120,3352
Standard Deviation	74,3528
Minimum	8,3700
Maximum	276,5800

Tabel 7 menyajikan statistik deskriptif dari variabel Penduduk Miskin di Provinsi Jawa Timur, yang mengindikasikan bahwa orang yang masuk kategori miskin adalah mereka yang memiliki rata-rata pengeluaran per kapita per bulan di bawah batas kemiskinan. Dari tabel tersebut dapat dilihat bahwa nilai rata-rata dari Penduduk Miskin di Provinsi Jawa Timur adalah sebesar 120,335263 Ribu Orang, sementara nilai standar deviasi sebesar 74,3528824. Hal ini mengindikasikan bahwa terdapat variasi tingkat kemiskinan yang cukup besar di daerah yang diteliti.

Dalam penelitian tersebut, Kabupaten Malang menunjukkan jumlah penduduk miskin tertinggi sebanyak 276,58 ribu orang. Hal ini menunjukkan bahwa tingkat kemiskinan di Kabupaten Malang lebih tinggi dibandingkan dengan daerah lain di Jawa Timur yang diteliti. Di sisi lain, Kota Mojokerto menunjukkan populasi penduduk miskin paling rendah sekitar 8,37 Ribu orang. Hal ini menunjukkan bahwa pemerintah Kota Mojokerto telah berhasil mengurangi angka kemiskinan di wilayahnya. Oleh karena itu, tingkat kemiskinan di suatu daerah dapat dipengaruhi oleh kebijakan dan upaya pemerintah setempat.

Statistik Deskriptif dari Variabel Fasilitas Kesehatan di Provinsi Jawa Timur dapat dilihat pada Tabel 8 .

**Tabel 8.** Statistik Deskriptif Fasilitas Kesehatan

Jumlah Fakes	Nilai
Mean	132,4473
Standard Deviation	69,5991
Minimum	22
Maximum	301

Tabel 8 memberikan gambaran statistik deskriptif dari variabel Fasilitas kesehatan di Provinsi Jawa Timur. Nilai rata-rata Fasilitas di provinsi tersebut adalah 132,44737, dengan jumlah faskes yang terdiri dari Rumah Sakit, Rumah Sakit Bersalin, Poliklinik, Puskesmas, dan Puskesmas Pembantu di setiap kabupaten/kota. Standar deviasi sebesar 69,599121 menunjukkan bahwa ada perbedaan yang signifikan dalam tingkat Fasilitas Kesehatan di setiap daerah yang diteliti.

Kota Surabaya memiliki jumlah faskes tertinggi, yang dapat memberikan akses kesehatan yang lebih baik bagi masyarakatnya, terutama dalam situasi darurat. Sebaliknya, Kota Batu memiliki jumlah faskes terendah di Jawa Timur, yang dapat menghambat pelayanan kesehatan yang memadai bagi masyarakatnya. Pasien di Kota Batu mungkin memilih untuk berobat di Kota Malang karena Kota Malang memiliki fasilitas kesehatan yang lebih lengkap dan berkualitas, termasuk rumah sakit besar, klinik, puskesmas, dan praktik dokter spesialis.

Statistik Deskriptif dari Variabel Tenaga Kesehatan di Provinsi Jawa Timur dapat dilihat pada Tabel 9 .



## Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks....

**Tabel 9.** Statistik Deskriptif Tenaga Kesehatan

Jumlah Nakes	Nilai
Mean	3622,7631
Standard Deviation	3321,6199
Minimum	1030
Maximum	20706

Tabel 9 menunjukkan statistik deskriptif untuk variabel Tenaga Kesehatan di Provinsi Jawa Timur. Nilai rata-rata Tenaga Kesehatan di provinsi tersebut adalah 3622,76. Dalam penelitian ini, Tenaga Kesehatan mencakup dokter, dokter gigi, perawat, dan bidan yang bekerja di setiap kabupaten/kota di Jawa Timur. Standar deviasi untuk variabel ini adalah 3321,62. Nilai standar deviasi yang tinggi dapat mengindikasikan bahwa ada beberapa wilayah di Jawa Timur dengan tingkat kemiskinan yang sangat tinggi, dan wilayah-wilayah tersebut mungkin memiliki kurangnya tenaga kesehatan. Di sisi lain, beberapa wilayah lainnya mungkin memiliki lebih banyak tenaga kesehatan.

Metode klasifikasi Naive Bayes membagi data menjadi data pelatihan dan data pengujian. Data pelatihan digunakan untuk membuat model klasifikasi, sedangkan data pengujian digunakan untuk memprediksi kelas pada data tersebut dan mengukur performa model yang sudah dilatih sebelumnya. Pertama-tama, data pelatihan dan data pengujian digabungkan untuk mendapatkan hasil percobaan yang optimal. Dilakukan empat kali percobaan, dan kombinasi dengan akurasi klasifikasi terbaik dipilih.

**Tabel 10.** Percobaan Pembagian Data *Training* dan Data *Testing*

No	Pesentase Data (%)		Akurasi
	<i>Training</i>	<i>Testing</i>	
1	70	30	0,9160
2	75	25	0,9000
3	80	20	0,8750
4	85	15	0,8330

Setelah itu, langkah berikutnya adalah menghitung probabilitas prior untuk setiap kelas pada data pelatihan, yaitu probabilitas awal sebelum melihat data yang diberikan. Probabilitas prior digunakan sebagai dasar untuk menghitung probabilitas posterior setelah melihat data pelatihan.

Kemudian, dilakukan perhitungan probabilitas kondisional untuk setiap variabel pada data pelatihan terhadap kelas yang sesuai dengan menggunakan metode Maximum Likelihood Estimation (MLE). Probabilitas kondisional ini menggambarkan kemungkinan munculnya nilai tertentu pada variabel saat diketahui kelas targetnya. Probabilitas kondisional ini kemudian digunakan dalam menghitung probabilitas posterior untuk setiap kelas pada data pengujian, sehingga dapat digunakan dalam memprediksi kelas dari data tersebut.

**Tabel 11.** Probabilitas Prior setiap kelas

Probabilitas		
Sedang	Tinggi	Sangat Tinggi
0,1052	0,5526	0,3421

Selanjutnya, probabilitas prior dari setiap kelas pada data pelatihan dihitung dan ditampilkan pada Tabel 10. Probabilitas prior mengindikasikan seberapa sering setiap kelas muncul dalam dataset. Dalam kasus ini, kelas 2 memiliki probabilitas prior yang paling tinggi, diikuti oleh kelas 3 dan kelas 1.

Kemudian, probabilitas setiap variabel pada data pelatihan terhadap kelas yang sesuai ditampilkan pada Tabel 11. Probabilitas ini

## Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks....

digunakan dalam menghitung probabilitas posterior untuk setiap kelas pada data pengujian.

**Tabel 12.** Probabilitas untuk setiap variabel

Variabel/ Kelas	1	2	3
Pendapatan Perkapita	16299,5	11861,23	9640
Umur Harapan Hidup (Tahun)	73,61	72,607	69,623
Harapan Lama Sekolah (Tahun)	14,978	13,52	12,719
Rata-rata Lama Sekolah (Tahun)	10,75	8,498	6,528
PDRB	81658,75	42098,38	25963,5
Penduduk Miskin	84,83	96,548	169,685
Jumlah Fakes	173	127,1	128,6
Jumlah Nakes	9746,5	1935,76	2848,31

Dalam klasifikasi Naive Bayes, probabilitas rata-rata setiap variabel pada data training untuk setiap kelas dihitung dan digunakan untuk mengklasifikasikan data testing. Probabilitas prior adalah probabilitas awal sebelum adanya informasi tambahan, sedangkan atribut pada masing-masing kelas adalah nilai atribut yang dihitung berdasarkan data training untuk setiap kelas. Dengan cara ini, Naive Bayes dapat mengklasifikasikan data testing berdasarkan probabilitas kondisional dari atribut-atribut pada data testing terhadap setiap kelas yang telah dihitung dari data training. Langkah selanjutnya adalah mencari nilai probabilitas maksimum pada setiap kelas, yang dilakukan dengan menghitung nilai probabilitas kondisional dari setiap atribut pada data testing terhadap masing-masing kelas yang telah dihitung sebelumnya berdasarkan data training. Nilai probabilitas kondisional

dari setiap atribut pada data testing akan dikalikan dengan probabilitas prior dari masing-masing kelas, dan kelas dengan nilai probabilitas tertinggi akan dipilih sebagai hasil klasifikasi untuk data testing tersebut.

**Tabel 13.** Hasil perkalian Probabilitas Prior dengan Atribut

No.	Kelas	Posterior Probabilitas
1	2	0,000671919348, 0,999327506, 5,74682726e-07
2	1	1,00000000, 2,47204513e-40, 7.72070330e-97
3	2	8,87782390e-09, 0,585571299, 0,414428692
4	2	2,19263893e-09, 0,502955711, 0,497044287
5	2	0,0143777012, 0,985622292, 6,69407659e-09
6	3	8,84187725e-19, 3,52017539e-05, 0,999964798

Probabilitas maksimum menunjukkan seberapa yakin model dalam memprediksi kelas yang benar untuk setiap sampel pada data testing, dan semakin tinggi nilai probabilitas, semakin besar kepercayaan model dalam prediksi kelas yang benar.

**Tabel 14.** Mencari Nilai maksimum

Baris	Nilai Probabilitas Maksimum
1	0,9993
2	1
3	0,5855
4	0,5029
5	0,9856
6	0,9999

Setelah proses klasifikasi dilakukan, langkah terakhir adalah menghitung tingkat akurasi klasifikasi dengan menggunakan nilai akurasi, sensitivity, dan specificity. Untuk menghitung akurasi klasifikasi, hasil klasifikasi model dibandingkan dengan kelas sebenarnya pada data testing dan dinyatakan dalam persentase dari jumlah data testing yang terklasifikasi dengan benar. Selain akurasi, kita juga dapat menghitung nilai sensitivity dan specificity untuk menggambarkan kemampuan model dalam mengidentifikasi kelas positif dan

## Penerapan Metode Naive Bayes Classifier Untuk Klasifikasi Indeks....

negatif dengan benar. Confusion Matrix digunakan untuk menghitung kedua nilai ini dengan membandingkan hasil klasifikasi model dengan kelas sebenarnya pada data testing. Tabel 14 merupakan contoh Confusion Matrix yang menunjukkan hasil prediksi ketepatan Indeks Pembangunan Manusia menggunakan metode Naive Bayes Classifier.

**Tabel 15.** Confusion Matriks setiap kelas.

Aktual	Prediksi		
	Sangat Tinggi	Tinggi	Sedang
Sangat Tinggi	1	0	0
Tinggi	0	7	0
Sedang	0	1	3

Berdasarkan Tabel 15, hasil prediksi menggunakan Naive Bayes Classifier menunjukkan bahwa 1 data pada kelas Sangat Tinggi dan 7 data pada kelas Tinggi terklasifikasi dengan benar, sedangkan 1 data pada kelas Tinggi terklasifikasi salah sebagai kelas Sedang dan 3 data pada kelas Sedang terklasifikasi dengan benar. Tidak ada data yang terklasifikasi salah sebagai kelas Sangat Tinggi atau Sedang.

**Tabel 16.** Classification report

Kelas	Precision	Recall	F1 Score	Support
Sangat Tinggi	1	1	1	1
Tinggi	0,88	1	0,93	7
Sedang	1	0,75	0,86	4
Accuracy			0,92	12
Macro Avg	0,96	0,92	0,93	12
Weighted Avg	0,93	0,92	0,91	12

Untuk mengevaluasi kesalahan prediksi menggunakan Naive Bayes Classifier, digunakan Classification report yang menghasilkan nilai precision yang sangat baik untuk kelas Sangat Tinggi dan Sedang yaitu 1.00 dan nilai precision yang baik untuk kelas Tinggi yaitu 0.88. Model juga memiliki nilai recall (Sensitivity) yang baik untuk kelas Sangat Tinggi dan Tinggi yaitu 1.00, dan nilai recall yang cukup untuk

kelas Sedang yaitu 0.75. Model juga memiliki nilai F1-score yang sangat baik untuk kelas Sangat Tinggi, baik untuk kelas Tinggi, dan cukup untuk kelas Sedang. Akurasi model mencapai 0.92, menunjukkan bahwa model dapat melakukan klasifikasi dengan baik pada data uji.

## PENUTUP

1. Provinsi Jawa Timur memiliki Indeks Pembangunan Manusia (IPM) dengan persentase tinggi sebesar 55%, sedang sebesar 34%, dan sangat tinggi sebesar 11%. Pendapatan per kapita rata-rata di provinsi ini adalah Rp11,568,526 dengan standar deviasi sebesar 2216,2167. Kota Surabaya memiliki pendapatan per kapita tertinggi, sementara Kabupaten Bangkalan memiliki pendapatan per kapita terendah. Rata-rata Angka Harapan Hidup di Jawa Timur adalah 71,69 tahun dengan standar deviasi sebesar 2,0148. Kota Surabaya memiliki Angka Harapan Hidup tertinggi, sedangkan Kabupaten Bondowoso memiliki Angka Harapan Hidup terendah. Rata-rata Harapan Lama Sekolah di provinsi ini adalah 13,3989 tahun. Nilai rata-rata PDRB di Jawa Timur adalah Rp40.742.789.000 dengan nilai standar deviasi sebesar 49.940.164. Kota Kediri memiliki PDRB tertinggi dan Kabupaten Pamekasan terendah. Jumlah penduduk miskin tertinggi berada di Kabupaten Malang, sementara Kota Mojokerto memiliki jumlah penduduk miskin terendah. Rata-rata fasilitas kesehatan di provinsi ini adalah 132,4473 dengan standar deviasi sebesar 69,5991. Kota Surabaya memiliki jumlah fasilitas kesehatan paling tinggi, sedangkan Kota Batu memiliki jumlah fasilitas kesehatan paling rendah.
2. Dalam melakukan prediksi terhadap Indeks Pembangunan Manusia di Provinsi Jawa Timur menggunakan 12 data testing

dan metode Naive Bayes Classifier, berhasil diperoleh presentase keakuratan sebesar 92%. Model tersebut memiliki nilai recall yang baik pada kelas Sangat Tinggi dan kelas Tinggi sebesar 1.00, serta nilai recall yang cukup pada kelas Sedang sebesar 0.75. Selain itu, model tersebut juga memiliki nilai F1-score yang sangat baik pada kelas Sangat Tinggi dan Tinggi, serta cukup pada kelas Sedang. Model tersebut memiliki akurasi sebesar 0.92, yang menunjukkan kemampuannya dalam melakukan klasifikasi pada data uji.

3. Berdasarkan prediksi tersebut, diperoleh hasil bahwa terdapat 1 Kota/Kabupaten yang memiliki IPM Sangat Tinggi, 7 Kota/Kabupaten yang memiliki IPM Tinggi, dan 3 Kota/Kabupaten yang memiliki IPM Sedang. Namun, terdapat 1 Kota/Kabupaten yang seharusnya berada pada kelas Tinggi namun diprediksi berada pada kelas Sedang.

#### DAFTAR PUSTAKA

- Budiman, M. A., & Cahyani, N. (2022). Pemodelan Regresi Logistik Ordinal Pada Indeks Pembangunan Manusia (IPM) Di Jawa Timur Tahun 2020. *Jurnal Statistika Dan Komputasi*, 1(2), 64-73. <https://doi.org/10.32665/statkom.v1i2.1169>
- Cholili, F. M. (2014). Analisis Pengaruh Pengangguran, Produk Domestik Regional Bruto (PDRB), dan Indeks Pembangunan Manusia (IPM) Terhadap Jumlah Penduduk Miskin. *Journal of Economic and Business*, 1-14.
- Fadhillah, H. (2018). Pengaturan Tentang Tenaga Kesehatan Dalam Peraturan Perundang-Undangan Dan Azas Kepastian Hukum. *Jurnal Hukum Kesehatan*. 21, 1-9. <https://doi.org/10.24167/shk.v5i1.1653>
- Fathurrahman, M., & Qisthi, N. (2021). Klasifikasi Indeks Pembangunan Manusia (IPM) Multi-Class Dengan Metode Artificial Neural Network (ANN). 0, 377-384.
- Fauzi, F., Yamin, M., & Wahyu, T. (2017). Klasifikasi Indeks Pembangunan Manusia Kabupaten / Kota Se-Indonesia dengan Pendekatan Smooth Support Vector Machine (SSVM) Kernel Radial Basis Function (RBF). *Seminar Nasional Pendidikan, Sains Dan Teknologi Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang*, 88-97.
- Hapsari, R. I. (2019). Analisis Pengaruh Indeks Pembangunan Manusia, Pendapatan Perkapita Dan Investasi Terhadap Kemiskinan Di Provinsi Lampung Dalam Perspektif Ekonomi Islam Tahun 2010-2018. In *Progress in Retinal and Eye Research* (Vol. 561, Issue 3).
- Hasibuan, L. S., Rujiman, & Sukardi. (2020). Analisis Determinan Indeks Pembangunan Manusia (IPM) Di Indonesia. *Jurnal Penelitian Pendidikan Sosial Humaniora*, 5(2), 139-141. <https://doi.org/https://doi.org/10.32696/jp2sh.v5i2.470>
- Hikmah, N. (2022). Analisis Spasial Sebaran Fasilitas Kesehatan Tingkat Pertama (Fktp) Di Kota Jambi. In *Braz Dent J.* (Vol. 33, Issue 1).
- Joko Ade Nursiyono, S. (2021). Indeks



- Pembangunan Manusia Provinsi Jawa Timur 2020-2021. Eksponensial, 12(2), 161-165.
- Mashfia, F. R. (2022). Prediksi Ketepatan Waktu Kelulusan Mahasiswa Menggunakan Metode Naive Bayes Classifier. 90. Timur, B. jawa. (2020). Indeks Pembangunan Manusia (IPM) Jawa Timur Tahun 2021. Berita Resmi Statistik, 53(9), 1689-1699.
- Pratiwi, I. A. A. S., & Wijayanto, A. W. (2019). Klasifikasi Indeks Pembangunan Manusia dengan Metode K-Nearest Neighbor dan Support Vector Machine di Pulau Jawa. Jurnal Ilmu Komputer, 15(1), 8-21.
- Putera, R. D. (2020). Analisis Pengaruh Angka Harapan Lama Sekolah, Upah Minimum Provinsi, dan Tingkat Pengangguran Terhadap Kemiskinan di Indonesia. <https://all3dp.com/2/fused-deposition-modeling-fdm-3d-printing-simply-explained/>
- Putri, R. E., Suparti, & Rahmawati, R. (2014). Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbour Pada Analisis Data Status Kerja di Kab.Demak. Jurnal Gaussian, 3, 831-838. <https://doi.org/10.14710/j.gauss.3.4.831-838>
- Rafiqi, A. (2020). Pengaruh Rata-rata Lama Sekolah, Pengeluaran Riil Perkapita, Pertumbuhan Ekonomi dan Pengangguran Terhadap Tingkat Kemiskinan di Provinsi D.I Yogyakarta. Skripsi.
- Ramaulidyah, F. N., Hayati, M. N., & Goejantoro, R. (2021). Perbandingan Metode Klasifikasi Naive Bayes Dan K-Nearest Neighbor Pada Data Status Pembayaran Pajak Pertambahan Nilai Di Kantor Pelayanan Pajak Pratama Samarinda Ulu.