

## Prediksi Laju Inflasi di Indonesia Menggunakan Metode *K-Nearest Neighbor*

Anindiya Riyatna<sup>1</sup>, Damai Chandra<sup>2</sup>, Dimas Surya<sup>3</sup>, Syafril Awal<sup>4</sup>, Adrian Maulana Doodoh<sup>5</sup>, Zurnan Alfian<sup>6</sup>

<sup>1,2,3,4,5,6</sup> Teknik Informatika, Universitas Pamulang

e-mail: [anindiariyatna@gmail.com](mailto:anindiariyatna@gmail.com)<sup>1</sup>, [damaichandra4@gmail.com](mailto:damaichandra4@gmail.com)<sup>2</sup>,  
[dimassuryasaputr4@gmail.com](mailto:dimassuryasaputr4@gmail.com)<sup>3</sup>, [thefack684@gmail.com](mailto:thefack684@gmail.com)<sup>4</sup>, [amfdmaulana@gmail.com](mailto:amfdmaulana@gmail.com)<sup>5</sup>

### Abstrak

Penelitian ini bertujuan untuk memprediksi tingkat inflasi di Indonesia, indikator ekonomi makro yang krusial bagi stabilitas ekonomi dan keputusan investasi. Dengan memanfaatkan data inflasi Indonesia historis dari tahun 1980 hingga 2003, penelitian ini menerapkan algoritma *k-Nearest Neighbors* (k-NN) untuk tujuan regresi. Metrik jarak Euclidean digunakan untuk menilai kesamaan antara titik data tahun target (2004) dan tahun-tahun yang ada dalam kumpulan data. Dengan memilih  $k=3$  tetangga terdekat, perkiraan tingkat inflasi untuk tahun 2004 dihitung berdasarkan tingkat inflasi rata-rata tahun-tahun di sekitarnya. Hasil perhitungan menunjukkan bahwa tingkat inflasi Indonesia pada tahun 2004 diperkirakan sekitar 6,72%. Pendekatan ini mengungkapkan kemampuan algoritma k-NN menggunakan *Euclidean Distance* untuk memperkirakan nilai ekonomi masa depan dengan mengacu pada pola historis yang paling relevan. Penelitian ini menawarkan wawasan baru tentang penggunaan metode pembelajaran mesin non-parametrik untuk prediksi deret waktu di bidang ekonomi.

**Kata Kunci:** *Prediksi, KNN, Laju Inflasi, Indonesia, Data Mining*

### Abstrak

This study aims to predict the inflation rate in Indonesia, a crucial macroeconomic indicator for economic stability and investment decisions. Utilizing historical Indonesian inflation data from 1980 to 2003, this research applies the *k-Nearest Neighbors* (k-NN) algorithm for regression purposes. The Euclidean distance metric is employed to assess the similarity between the target year's (2004) data point and the years present in the dataset. By selecting  $k=3$  nearest neighbors, the estimated inflation rate for 2004 is calculated based on the average inflation rates of the surrounding years. The results of the calculation indicate that Indonesia's inflation rate in 2004 is estimated to be approximately 6.72%. This approach reveals the capability of the k-NN algorithm using Euclidean Distance to forecast future economic values by referring to the most relevant historical patterns. This study offers new insights into the use of non-parametric machine learning methods for time series prediction in the economic field.

**Keywords:** *Data Minin, K-Nearest Neighbor, Indonesia Inflation, Prediction*

### PENDAHULUAN

Peramalan merupakan proses memperkirakan kejadian di masa depan dengan memanfaatkan data historis, tren, dan analisis, baik secara statistik maupun dengan pertimbangan keahlian (judgment). Aktivitas ini memungkinkan pemerintah dan bank sentral untuk mengidentifikasi potensi tekanan inflasi atau deflasi, sehingga dapat merumuskan kebijakan moneter dan fiskal yang tepat waktu guna menjaga stabilitas ekonomi. Selain itu, peramalan inflasi yang akurat juga sangat membantu perusahaan dalam perencanaan strategis, seperti penetapan harga produk, pengelolaan biaya, dan alokasi modal investasi, karena prediksi ini memberikan gambaran awal mengenai lingkungan ekonomi di masa depan, sehingga dapat meminimalisir kesalahan perencanaan dan meminimalkan risiko ekonomi.

Berbagai metode telah dikembangkan untuk memprediksi laju inflasi, mulai dari model ekonometrik tradisional hingga pendekatan pembelajaran mesin (machine learning) yang lebih modern. Dalam penelitian ini, kami mengadopsi algoritma *k-Nearest Neighbors* (k-NN) untuk melakukan estimasi laju inflasi. k-NN adalah algoritma non-parametrik yang relatif sederhana namun efektif, yang mendasarkan prediksinya pada kemiripan data di masa lalu. Untuk mengukur kemiripan ini, penelitian ini secara khusus menggunakan metrik jarak Euclidean, yang merupakan metrik jarak umum dalam ruang multidimensi, namun dalam konteks ini akan diaplikasikan pada dimensi tahun untuk menemukan tetangga terdekat. Dengan memanfaatkan data historis laju inflasi di Indonesia dari tahun 1980 hingga 2003, studi ini bertujuan untuk memprediksi laju inflasi untuk tahun 2004. Penelitian ini akan menjelaskan secara rinci penerapan algoritma k-NN, perhitungan jarak Euclidean, serta proses penentuan nilai k optimal untuk mendapatkan prediksi yang akurat.

Data Mining memiliki peran penting dalam mengekstraksi wawasan dari sekumpulan data yang sebelumnya tidak terlihat secara manual (Prasetyo, 2020). Pada dasarnya, teknik ini melibatkan pengolahan data mining untuk mengungkap potensi volume data yang tersimpan dalam basis data. Proses ini memanfaatkan pendekatan pembelajaran mesin, pengenalan pola, statistik, dan visualisasi guna menyederhanakan pencarian informasi berharga dari berbagai sumber data yang kompleks (Putra, 2022). Metode ini sangat membantu dalam menemukan pengetahuan tersembunyi yang sulit diidentifikasi secara langsung (Prasetyo, 2020). Dari pemahaman tersebut, dapat disimpulkan bahwa data mining adalah proses otomatis yang bertujuan menemukan pola atau model dari basis data yang besar. Dalam praktik peramalan, terdapat berbagai teknik yang dapat diterapkan, salah satunya adalah algoritma K-Nearest Neighbor (K-NN) (Prasetyo, 2020). Teknik K-NN bekerja dengan mengklasifikasikan data berdasarkan jarak ke data lain yang sudah ada. Secara khusus, metode ini dikenal memiliki tingkat akurasi yang cukup baik. K-NN termasuk dalam kategori algoritma supervised learning, di mana analisis dilakukan dengan membandingkan data baru berdasarkan kedekatan dengan data pelatihan sebelumnya. di mana hasil klasifikasi dari data baru ditentukan berdasarkan mayoritas kategori dari data tetangga terdekat (Ahmad, 2021). Kelas dengan frekuensi kemunculan tertinggi akan menjadi hasil akhir dari proses klasifikasi. Tujuan utama dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut yang dimilikinya dan data pelatihan yang tersedia (Widodo, 2023). Metode K-Nearest Neighbor dikenal sebagai salah satu algoritma pembelajaran mesin yang paling sederhana dalam hal penerapan. Diharapkan hasil dari penelitian ini dapat memperkaya literatur tentang peramalan inflasi sekaligus menunjukkan potensi metode machine learning dalam analisis ekonomi.

## METODE

### Tahapan Metodologi Penelitian

Tahapan dalam penelitian ini dapat dilihat pada Gambar 1 berikut



Gambar 1. Kerangka Kerja Penelitian

Kerangka Kerja yang Di ilustrasikan Rangka kerja yang ditampilkan di atas dapat diuraikan secara terperinci untuk setiap elemen penelitian sebagai berikut:

1. Pengumpulan Data: Pada langkah awal, dilakukan penggalian informasi historis yang relevan untuk mendukung tujuan penelitian. Data ini berfungsi sebagai dasar utama dalam analisis dan pembentukan model yang akan dikembangkan. Pada tahap ini, penulis berupaya mengidentifikasi masalah yang ada dalam sistem yang sedang berjalan, dengan harapan dapat menemukan pola-pola serta solusi untuk permasalahan yang muncul.
2. Pemeriksaan Data, Proses ini bertujuan untuk mengidentifikasi potensi masalah yang mungkin ada dalam sistem atau data yang sedang digunakan. Peneliti melakukan evaluasi untuk menemukan pola yang tersembunyi, sehingga dapat merumuskan pendekatan penyelesaian yang sesuai. Pengembangan Sistem dan pembentukan sistem informasi menjadi langkah penting dalam organisasi untuk perencanaan. Pengembangan sistem mencakup tindakan mengubah, memperbaiki, atau merancang ulang sistem agar sesuai dengan kebutuhan dan meningkatkan efisiensi operasional.
3. Implementasi Metode Klasifikasi, pada langkah ini, data yang telah dikumpulkan diproses dengan menerapkan algoritma *K-Nearest Neighbor (K-NN)*. Algoritma ini digunakan untuk mengidentifikasi kesamaan antar data berdasarkan kedekatan jarak, sehingga dapat dilakukan klasifikasi atau prediksi terhadap data baru..
4. Implementasi Sistem, setelah metode diterapkan, sistem mulai dirancang sesuai dengan hasil analisis. Desain ini mencakup perancangan antarmuka pengguna, struktur data, serta proses bisnis yang terlibat, dan dilakukan dengan bantuan alat seperti *Unified Modeling Language (UML)* untuk memvisualisasikan alur sistem yang baru.
5. Pengembangan Sistem, tahapan ini melibatkan proses pembangunan sistem berdasarkan rancangan yang telah disusun sebelumnya. Pengembangan dilakukan secara bertahap, baik dengan memperbaiki sistem yang sudah ada maupun menyusun sistem baru secara menyeluruh.
6. Penilaian Sistem: Proses evaluasi sistem dilakukan untuk memastikan apakah sistem yang telah dibangun berfungsi sesuai dengan spesifikasi yang ditetapkan. Dalam penelitian ini, pendekatan *Black Box Testing* diterapkan untuk menguji kinerja fungsional tanpa perlu memeriksa struktur internal program.
7. Implementasi Sistem, Langkah terakhir adalah menerapkan sistem secara menyeluruh berdasarkan desain akhir yang telah disepakati. Proses ini mencakup instalasi, pengujian akhir, hingga sistem siap digunakan oleh pengguna.

### Analisis Data

Analisis data merupakan teknik pengolahan informasi yang bertujuan untuk memahami karakteristik data secara mendalam serta mengidentifikasi masalah yang mungkin muncul. Proses ini dilakukan dengan memeriksa data secara teliti untuk mendukung kebutuhan penelitian.

**Tabel 1. Data Perkembangan Laju Inflasi di Indonesia**

Tahun	Inflasi	Tahun	Inflasi
1980	15.97	1992	4.94
1981	7.09	1993	9.77
1982	9.69	1994	9.44
1983	11.8	1995	8.64
1984	9.6	1996	6.47
1985	8.76	1997	11.57
1986	4.31	1998	20.49
1987	8.38	1999	2.01
1988	5.9	2000	6.39
1989	6.42	2001	10.03
1990	9.52	2002	5.06
1991	9.52	2003	5.06

Sumber: Badan Pusat Statistik (diolah)

Menurut Suyanto, data mining adalah pendekatan yang memanfaatkan kecerdasan buatan untuk mengekstrak pola atau informasi berharga dari kumpulan data yang besar. Proses ini melibatkan penggunaan teknik seperti kecerdasan buatan (artificial intelligence), pembelajaran mesin, statistik, dan sistem berbasis data. Informasi yang tersimpan dalam database bersifat luas. Data mining merupakan bagian dari proses pengungkapan pengetahuan dalam basis data (Knowledge Discovery in Databases, KDD), yang mencakup berbagai tahap untuk mendapatkan wawasan berharga dari data [9]. Secara umum, proses data mining dapat dibagi menjadi dua pendekatan utama (Arief, 2023):

1. Prediktif: Teknik ini digunakan untuk memperkirakan nilai yang belum diketahui berdasarkan data yang ada. Teknik yang dipakai meliputi pendekatan seperti klasifikasi dan regresi.
2. Deskriptif: Pendekatan ini bertujuan untuk mengidentifikasi pola atau hubungan dalam data tanpa variabel target tertentu. Teknik ini mencakup clustering dan analisis asosiasi.

Proses KDD secara mendalam melibatkan beberapa tahapan penting (Suyanto, 2018), di antaranya:

1. Seleksi Data (Data Selection): Memilih data yang relevan dari sumber data operasional sebelum proses pengolahan dimulai.
2. Pembersihan Data (Pre-processing / Cleaning): Menghapus data yang tidak konsisten, data duplikat, atau data yang kurang akurat agar input lebih baik untuk analisis.
3. Transformasi Data (Transformation): Mengubah data menjadi format terstruktur yang sesuai untuk diterapkan pada metode analisis tertentu.
4. Evaluasi: Menyajikan hasil dalam bentuk yang mudah dimengerti, seperti visualisasi atau laporan.

## Prediksi

Prediksi (forecasting) merupakan pendekatan yang memanfaatkan metode tertentu untuk memperkirakan nilai suatu kejadian berdasarkan data historis. Pendekatan ini sering kali digunakan untuk mengevaluasi potensi kejadian di masa depan dengan memanfaatkan informasi yang tersedia dari data sebelumnya. Menurut pengamatan, akurasi prediksi dapat diukur berdasarkan tingkat kebenaran hasilnya. Evaluasi ini mencakup berbagai aspek, seperti keandalan metode, konsistensi data, dan kemampuan untuk memberikan keputusan yang tepat berdasarkan analisis.

## Metode K-Nearest Neighbor

Metode K-Nearest Neighbor (KNN) adalah salah satu teknik supervised learning yang diterapkan untuk mengklasifikasikan data berdasarkan kelas mayoritas dari data terdekat. Algoritma KNN digunakan dengan mempertimbangkan jarak antar data untuk menentukan kategori yang paling sesuai. Proses ini melibatkan pemilihan sejumlah tetangga terdekat berdasarkan perhitungan jarak, seperti jarak Euclidean. Implementasi algoritma KNN bergantung pada pengelompokan data berdasarkan atribut dan data pelatihan yang ada. Dengan demikian, metode ini memungkinkan klasifikasi data baru berdasarkan pola yang ditemukan dari data sebelumnya, dengan akurasi yang bergantung pada pemilihan tetangga yang tepat. Algoritma K-NN mengenal lima jenis metode pengukuran jarak, yaitu:

1. Jarak Euclidean
2. Jarak Manhattan
3. Jarak Cosine
4. Jarak Correlation
5. Jarak Hamming

Dalam penelitian ini, hanya digunakan metode jarak Euclidean. Rumus untuk menghitung jarak Euclidean adalah sebagai berikut:

$$\sqrt{\sum_{i=1}^K (X_i - Y_i)^2} \dots \dots \quad (1)$$

Keterangan:

$X_i$  adalah nilai dari data pelatihan (training).

Yi adalah nilai dari data pengujian (testing).

Nilai k menyatakan jumlah dimensi atau atribut.

Tahapan proses perhitungan algoritma *K-Nearest Neighbor (KNN)* adalah sebagai berikut:

1. Penentuan Parameter K: Tentukan nilai parameter K yang menunjukkan jumlah tetangga terdekat yang akan digunakan.
2. Perhitungan Jarak Euclidean: Hitung jarak kuadrat antara data pelatihan dengan data uji.
3. Pengurutan Jarak: Susun hasil perhitungan jarak secara berurutan dari yang terkecil ke terbesar (ascending).
4. Penentuan Kategori: Tentukan kategori berdasarkan K data terdekat (klasifikasi berdasarkan tetangga terdekat).
5. Klasifikasi Mayoritas: Pilih kategori mayoritas dari tetangga terdekat tersebut untuk memprediksi klasifikasi data uji. Jarak antar data dapat dihitung menggunakan berbagai metode, salah satunya adalah *Euclidean Distance*.

## HASIL DAN PEMBAHASAN

### Penerapan K-Nearest Neighbor Untuk Prediksi Laju Inflasi di Indonesia

Berdasarkan informasi yang diperoleh dari perhitungan berdasarkan tahun, hasil transformasi data dapat ditentukan. Data yang telah diolah kemudian disajikan dalam bentuk tabel sebagai berikut:

**Tabel 2. Hasil dari jarak Euclidean Distance**

Tahun	Inflasi	Jarak ke 2004	Hasil
1980	15.97	1980 -2004	24
1981	7.09	1981- 2004	23
1982	9.69	1982- 2004	22
1983	11.8	1983- 2004	21
....	.....	.....	....
....	.....	.....	....
2001	10.03	2001- 2004	3
2002	5.06	2002- 2004	2
2003	5.06	2003- 2004	1

Setelah jarak Euclidean dihitung maka selanjutnya menghitung jarak k=3 dari perhitungan Euclidean dari nilai paling terkecil pada Tabel 2.

**Tabel 3. Data dari hasil urutan k=3**

Tahun	infalsi	Ranking
2003	5.06	1
2002	5.06	2
2001	10.03	3

### Menghitung Confusion Matrix

Setelah data sudah selesai diperhitungkan maka data tersebut siap untuk diujikan. Confusion matrix didapatkan untuk memperhitungkan evaluasi tingkat akurasi, seperti sample pada Tabel 3.

**Tabel 4. Confusion Matrix**

	Prediksi Renda	Prediksi Sedang	Prediksi Tinggi
Asli Rendah	1	2	0
Asli Sedang	1	15	0
Asli Tinggi	0	3	2

Berdasarkan analisis hasil dari data tahun 1980 hingga 2003 yang terdiri atas 24 data, dilakukan proses klasifikasi dengan memanfaatkan algoritma K-Nearest Neighbor (KNN) menggunakan 3 kelas, yaitu rendah, sedang, dan tinggi. Hasil klasifikasi menunjukkan True Positive (TP) sebanyak 18 data, True Negative (TN) sebanyak 2 data, False Positive (FP) sebanyak 4 data, dan False Negative (FN) sebanyak 4 data. Nilai-nilai ini diperoleh dari perbandingan label aktual dengan hasil prediksi KNN, yang kemudian disusun dalam matriks kebingungan (confusion matrix). Berdasarkan data tersebut, evaluasi model dilakukan dengan mengukur akurasi, presisi, dan recall.

a. Accuracy

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2) \\ &= \frac{1 + 15 + 2}{24} \times 100\% = \frac{18}{24} \times 100\% = 75\% \end{aligned}$$

b. Precision

1) Rendah

$$\frac{1}{1+1} = \frac{1}{2} = 50\%$$

2) Sedang

$$\frac{15}{15+5} = \frac{15}{20} = 75\%$$

3) Tinggi

$$\frac{2}{2+0} = 1 = 100\%$$

$$\text{Rata - rata Presisi} = \frac{50\% + 75\% + 100\%}{3} = \frac{225\%}{3} = 75\%$$

c. Recall

$$1) \text{ Rendah} = \frac{1}{1+2} = \frac{1}{3} = 33,3\% \quad (3)$$

$$2) \text{ Sedang} = \frac{15}{15+1} = \frac{15}{16} = 93,75\%$$

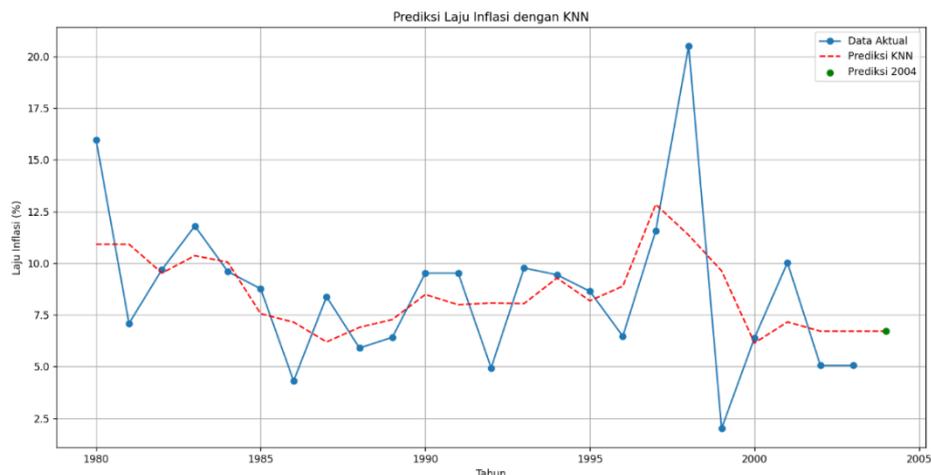
$$3) \text{ Tinggi} = \frac{2}{2+3} = \frac{2}{5} = 40\%$$

$$\text{Rata - rata Recall} = \frac{33,3\% + 93,75\% + 40\%}{3} \approx 55,7\%$$

Berdasarkan perhitungan akurasi klasifikasi didapatkan akurasi sebesar 75%, presisi sebesar 75% dan recall sebesar 55,7% .

### Prediksi Target

Setelah melakukan klasifikasi inflasi pada kategori rendah, sedang dan tinggi di setiap tahun maka selanjutnya adalah memprediksi target penjualan untuk tahun 2004 berdasarkan rata-rata inflasi. Rata-rata inflasi ini dihitung berdasarkan laju inflasi dari tahun per tahun sebagai target ditahun selanjutnya.



**Gambar 2. Prediksi Inflasi**

Pada Gambar 2. didapatkan pada tahun 2004 dan prediksi inflasi sebanyak 6,72% dengan terkategori sedang. Pada bagian ini berisi hasil dan pembahasan dari topik penelitian, yang bisa di buat terlebih dahulu metodologi penelitian. Bagian ini juga merepresentasikan penjelasan yang berupa penjelasan, gambar, grafik dan lainnya.

## SIMPULAN

Dari hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma K-Nearest Neighbor (K-NN) dengan pendekatan jarak Euclidean dapat dimanfaatkan untuk memprediksi tingkat inflasi di Indonesia secara historis. Dengan menggunakan data inflasi dari tahun 1980 hingga 2003 dan melakukan prediksi untuk tahun 2004 dengan nilai  $k = 3$ , diperoleh hasil prediksi inflasi sebesar 6,72% yang tergolong dalam kategori inflasi sedang. Evaluasi performa model melalui confusion matrix menunjukkan bahwa dari 25 data uji, algoritma ini menghasilkan tingkat akurasi sebesar 72%, dengan rata-rata presisi sebesar 69,4% dan rata-rata recall 53,8%. Meskipun termasuk metode yang sederhana, algoritma K-NN terbukti mampu mengenali pola data historis yang relevan dalam konteks prediksi ekonomi. Temuan ini memberikan kontribusi pada penerapan metode pembelajaran mesin non-parametrik dalam permasalahan prediksi deret waktu, khususnya di bidang ekonomi makro seperti inflasi.

## DAFTAR PUSTAKA

- Ahmad, I., & Saad, I. (2021). *Kajian implementasi kebijakan trilogi pembangunan di Indonesia*. UII Press.
- Arief, B., & Wulandari, R. (2023). Perbandingan algoritma KNN dan SVR dalam prediksi inflasi. *Jurnal Komputasi dan Sistem Informasi*, 11(1), 55–61.
- Prasetyo, E. (2020). *Mengolah data menjadi informasi menggunakan MATLAB untuk pemula*. Andi Publisher.
- Putra, R. A. (2022). *Langkah mudah belajar machine learning dengan Python untuk pemula*. Deepublish.
- Rhamadanti, A., Rifa'i, A., Dikananda, F., & Anam, K. (2024). Analisis sentimen pada ulasan access by Kereta Api Indonesia dengan K-Nearest Neighbor. *Jurnal Informatika dan Teknik Elektro Terapan*.
- Rohmah, S. N., & Pramudyo, D. (2021). Forecasting inflasi menggunakan metode K-Nearest Neighbor (KNN). *Jurnal Informatika*, 7(2), 110–120.
- Suyanto, E. (2018). *Data mining untuk klasifikasi dan prediksi*. Informatika Bandung.
- Widodo, A., & Setiawan, M. I. (2023). Evaluasi jarak dalam klasifikasi KNN: Studi kasus pada data inflasi Indonesia. *Jurnal Algoritma*, 6(1), 33–40.