

## Implementation of the K-Nearest Neighbor (K-NN) Algorithm for Predicting Bestseller of Religious Books

Alfan Jamil<sup>1\*</sup>, Hermanto<sup>2</sup>, Fajriyanto<sup>3\*</sup>

<sup>1</sup>Sistem Informasi, Fakultas Sains & Teknologi, Universitas Ibrahimi Situbondo

Jl. KHR. Syamsul Arifin, Sukorejo, Sumberejo, Kec. Banyuputih, Kabupaten Situbondo, Jawa Timur 68374

### Informasi Artikel

Diterima : 11 Juli 2025  
Revisi : 15 Juli 2025  
Publikasi : 30 September 2025

### Kata Kunci:

Data Mining,  
K-Nearest Neighbor  
Prediksi Penjualan  
Kitab dan Buku  
Akurasi Model

### ABSTRAK

Perkembangan teknologi informasi memberikan peluang yang signifikan dalam pengelolaan data untuk membantu membuat keputusan bisnis yang lebih tepat. Salah satu cara yang dipakai adalah data mining, terutama teknik K-Nearest Neighbor (K-NN) yang cukup efektif dalam klasifikasi dan ramalan berdasarkan data masa lalu. Penelitian ini bertujuan untuk menerapkan teknik K-NN guna memprediksi penjualan buku dan kitab terpopuler di Toko Kitab Assyarif, yang berlokasi di Pondok Pesantren Salafiyah Syafi'iyah Sukorejo. Diharapkan metode ini bisa membantu dalam merencanakan stok dan mengenali produk yang paling diminati oleh pelanggan. Penelitian ini menggunakan pendekatan kuantitatif dengan metode observasi, wawancara, dan pengumpulan dokumentasi. Data yang digunakan mencakup harga, jumlah yang terjual, stok awal, serta stok akhir. Hasil analisis dengan menggunakan Euclidean Distance menunjukkan bahwa teknik K-NN mampu memprediksi kategori penjualan (laris atau tidak laris) dengan tingkat keakuratan sebesar 75% yang didapat dari pengujian cross-validation, sehingga dapat menjadi solusi yang efektif untuk mendukung manajemen penjualan dan keputusan berbasis data.

### ABSTRACT

The development of information technology provides significant opportunities in data management to help make more accurate business decisions. One method used is data mining, especially the K-Nearest Neighbor (K-NN) technique which is quite effective in classification and forecasting based on past data. This study aims to apply the K-NN technique to predict sales of the most popular books and scriptures at the Assyarif Book Store, located at the Salafiyah Syafi'iyah Islamic Boarding School in Sukorejo. It is hoped that this method can help in planning stock and identifying products that are most in demand by customers. This study uses a quantitative approach with observation, interviews, and documentation collection methods. The data used include price, quantity sold, initial stock, and final stock. The results of the analysis using Euclidean Distance show that the K-NN technique is able to predict sales categories (best-selling or not-selling) with an accuracy level of 75% obtained from cross-validation testing, so it can be an effective solution to support sales management and data-based decisions.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



### \*Penulis Koresponden

Email: [alfanjamil2003@gmail.com](mailto:alfanjamil2003@gmail.com)

Cara sitasi IEEE:

A. Jamil, Hermanto, & Fajriyanto, "Implementation of the K-Nearest Neighbor (K-NN) Algorithm for Predicting Bestseller of Religious Books," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 3, pp. 1328-1339, September 2025, doi: 10.30811/jaise.v5i3.7352

## 1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat dewasa ini telah membawa dampak besar dalam berbagai aspek kehidupan, termasuk dunia usaha dan perdagangan. Salah satu manfaat signifikan dari kemajuan ini adalah meningkatnya kemampuan dalam mengelola data dalam jumlah besar secara efisien untuk mendukung pengambilan keputusan yang lebih tepat, akurat, dan berbasis data (*data-driven decision making*). Dalam praktik bisnis, kemampuan untuk menganalisis data penjualan, mengidentifikasi tren pasar, dan memprediksi kebutuhan konsumen telah menjadi faktor penting dalam meningkatkan efisiensi operasional dan daya saing[1].

Salah satu pendekatan yang banyak digunakan dalam proses analisis data tersebut adalah data mining. Data mining merupakan proses eksplorasi dan ekstraksi pola tersembunyi dari data berukuran besar yang kompleks, yang dapat menghasilkan informasi atau pengetahuan baru yang bermanfaat. Dalam konteks penjualan, data mining terbukti efektif untuk mendukung strategi pemasaran, personalisasi layanan pelanggan, pengelolaan inventori, hingga prediksi penjualan. Teknik-teknik utama yang digunakan dalam data mining antara lain klusterisasi, asosiasi, regresi, dan klasifikasi.

Salah satu metode klasifikasi yang cukup populer adalah algoritma K-Nearest Neighbor (K-NN), yang bekerja dengan mengklasifikasikan data baru berdasarkan kemiripan dengan sejumlah data tetangga terdekat. Kelebihan K-NN terletak pada kemudahan implementasi dan kemampuannya dalam menangani data numerik. Berbagai studi sebelumnya telah menunjukkan efektivitas algoritma ini dalam mendukung proses prediksi penjualan. Misalnya, Yulia dan Ibnu (2021) menggunakan K-NN untuk memprediksi penjualan beton instan dan memperoleh akurasi tinggi. Sementara itu, Andini et al. (2020) menerapkan metode yang sama dalam analisis pola pembelian di toko ritel untuk membantu pengambilan keputusan pengadaan barang[2].

Namun, sebagian besar penelitian tersebut masih berfokus pada produk ritel modern dan belum menjangkau konteks yang lebih spesifik seperti penjualan kitab dan buku literatur keislaman. Produk-produk seperti kitab kuning dan buku pelajaran agama memiliki pola permintaan yang berbeda, cenderung musiman, dan tergantung pada kalender akademik pesantren. Penelitian dalam konteks ini masih sangat terbatas, padahal toko-toko kitab di lingkungan pesantren juga menghadapi tantangan dalam pengelolaan data penjualan dan perencanaan stok

Toko Kitab Assyarif yang berada di lingkungan Pondok Pesantren Salafiyah Syafi'iyah Sukorejo merupakan unit usaha pesantren yang menyediakan berbagai kitab dan buku keislaman untuk santri, pengajar, dan masyarakat umum. Seiring bertambahnya jumlah santri dan kebutuhan literatur, pengelolaan stok dan penjualan menjadi semakin kompleks. Sayangnya, proses manajemen data penjualan di toko ini masih dilakukan secara manual, sehingga menyulitkan dalam menentukan produk terlaris dan jumlah stok yang perlu disediakan secara optimal[3].

Untuk mengatasi permasalahan tersebut, penelitian ini mengimplementasikan algoritma K-Nearest Neighbor (K-NN) guna memprediksi kategori penjualan kitab dan buku (laris atau tidak laris) berdasarkan data historis. Penelitian ini memanfaatkan atribut seperti harga, jumlah terjual, stok awal, dan stok akhir sebagai variabel prediktif. Output dari sistem ini diharapkan dapat mendukung pengambilan keputusan yang lebih akurat dalam hal perencanaan stok dan strategi penjualan. Adapun kebaruan dari penelitian ini terletak pada Penerapan metode K-NN dalam konteks toko kitab berbasis pesantren yang belum banyak diteliti sebelumnya, Integrasi sistem prediksi K-NN dengan evaluasi performa model menggunakan teknik *cross-validation* untuk mengukur akurasi, Penyajian hasil prediksi dalam bentuk visual interaktif yang memudahkan pihak non-teknis dalam memahami dan menggunakan informasi hasil prediksi.[4].

Berdasarkan latar belakang tersebut, penelitian ini memberikan kontribusi pada pengembangan sistem pendukung keputusan berbasis data mining yang kontekstual dan aplikatif, serta mendukung transformasi digital di lingkungan usaha tradisional seperti pesantren. Penelitian ini juga diharapkan dapat menjadi model yang dapat diadaptasi oleh unit usaha sejenis untuk meningkatkan efisiensi dan akurasi dalam manajemen penjualan dan stok barang.

## 2. METODE

Dalam penyusunan proposal skripsi ini kami melakukan observasi langsung ke tempat penelitian dengan melakukan pengamatan dan mengumpulkan data-data yang diperlukan dalam laporan penelitian ini. Berikut lebih jelasnya tentang metode-metode penelitian yang kami lakukan dalam penelitian ini[5]

### 2.1. Teknik Pengumpulan Data

### 1. Observasi (Pengamatan)

Observasi merupakan teknik pengumpulan data dengan mengamati secara langsung aktivitas yang terjadi di lapangan agar mengetahui alur proses sistem yang berjalan yang bertujuan untuk memperoleh data atau informasi tentang objek penelitian[6].

### 2. Wawancara (Interview)

Wawancara merupakan Teknik pengumpulan data yang melibatkan interaksi langsung antara peneliti dan pihak terkait melalui sebuah pertanyaan yang telah disusun sebelumnya. Wawancara dilakukan untuk mendapatkan alur proses bisnis dan aturan-aturan terkait sistem yang dibangun dari pihak terkait sehingga pembangunan sistem informasi sesuai dengan kebutuhan dan dapat menjadi solusi yang tepat[7].

### 3. Data Sekunder

Data Sekunder adalah bentuk data historis yang berisi informasi mengenai variabel variabel yang telah terkumpul dan diarsipkan sebelumnya oleh entitas lain. Data sekunder berasal dari dalam perusahaan itu sendiri dalam penelitian ini, dimana peneliti menggunakan data sekunder internal yang mencakup data yang dikumpulkan oleh perusahaan secara individu, seperti laporan hasil penjualan untuk tujuan akuntansi, laporan aktivitas pemasaran, dan pengetahuan mengenai pelanggan[8].

## 2.2 Analisis Data

Analisa data adalah suatu metode pengolahan data menjadi informasi untuk membantu memahami karakteristik data dan memecahkan masalah, terutama yang berkaitan dengan penelitian[9].

Tabel 1. Penjualan Produk

Bulan	Nama Barang	Harga	Terjual	Stok Awal	Stok Akhir	Kategori	Jumlah Laku Harga
Januari	Sy Fathul Qorib	130000	50	54	4	Laris	6500000
Januari	Sy Jurumiyah	90000	28	43	15	Tidak Laris	2520000
Januari	Sy Kaelani	65000	42	67	25	Tidak Laris	2730000
Januari	Sy Taklim	70000	35	50	15	Tidak Laris	2450000
Februari	Kitab Safinatun Najah	30000	70	110	40	Laris	2100000
Februari	Buku Bahasa Arab						
Februari	Dasar	40000	55	85	30	Laris	2200000
Februari	Buku Tajwid Lengkap	35000	60	95	35	Tidak Laris	2100000
Februari	Buku Pendidikan						
Maret	Akhlak	38000	45	65	20	Laris	1710000
Maret	Kitab Tijan Darori	33000	50	72	22	Laris	1650000
Maret	Riyadhus Shalihin	80000	31	49	18	Tidak Laris	2480000
April	Kamus Arab-Indonesia	50000	40	60	20	Laris	2000000
April	Kitab Tauhid Dasar	37000	49	80	31	Tidak Laris	1813000
April	Panduan Shalat						
April	Lengkap	32000	65	90	25	Laris	2080000
Mei	Kumpulan Doa Harian	25000	70	100	30	Laris	1750000
Mei	Kitab Tafsir Jalalain	60000	45	80	35	Tidak Laris	2700000
Mei	Kitab Ta'lim						
Mei	Muta'allim	40000	33	60	27	Tidak Laris	1320000
Juni	Buku Saku Ibadah	27000	58	95	37	Laris	1566000
Juni	Kitab Aqidatul Awam	35000	40	75	35	Tidak Laris	1400000
Juli	Kitab Ushul Fiqih	55000	48	85	37	Tidak Laris	2640000
Juli	Buku Dzikir Pagi						
Juli	Petang	30000	68	95	27	Laris	2040000

## 2.3 Prediksi

Prediksi/peramalan (forecasting) adalah menentukan jumlah kebutuhan bulan/tahun mendatang berdasarkan dengan dukungan data historis (historical data) atau serangkaian waktu/periode yang dianalisis sehingga dapat diperhitungkan untuk memprediksi jumlah kebutuhan pada bulan/tahun mendatang. Prediksi juga dapat digunakan dalam pengklasifikasian, tidak hanya untuk memprediksi time series, karena sifatnya yang bisa menghasilkan class berdasarkan atribut yang ada[7].

## 2.4 Data Mining

Data mining merupakan suatu metode untuk mengambil pengetahuan atau informasi berharga dari kumpulan data yang besar dan rumit. Tujuan utama dari data mining adalah untuk menemukan pola, hubungan,

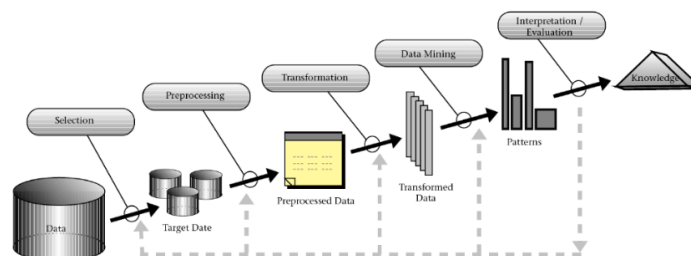
atau informasi yang mungkin tidak tampak jelas dalam data, sehingga dapat memberikan wawasan yang lebih mendalam dan akurat. Proses data mining menggunakan berbagai metode statistik, matematis, dan kecerdasan buatan guna menganalisis data secara sistematis dan otomatis. Hasil dari data mining dapat dimanfaatkan untuk membantu pengambilan keputusan, menemukan tren di pasar, meningkatkan efisiensi operasional, atau merancang strategi bisnis[10].

Umumnya, Data Mining dapat dikelompokkan menjadi dua kategori utama, yaitu:

1. **Prediksi**  
Ini adalah proses yang digunakan untuk mengidentifikasi karakteristik penting dalam sebuah basis data. Metode data mining yang termasuk dalam deskriptif mining meliputi clustering, asosiasi, dan sequential mining[11].
2. **Deskriptif**  
Ini adalah proses yang bertujuan untuk menemukan pola dari data dengan memanfaatkan beberapa variabel di masa depan. Salah satu teknik yang termasuk dalam predictive mining adalah klasifikasi. Secara sederhana, data mining sering disebut sebagai proses untuk menyaring atau “menambang” pengetahuan dari sekumpulan data yang besar[12].

Secara keseluruhan, proses Knowledge Discovery in Databases (KDD), dapat dijelaskan sebagai berikut:

1. **Pemilihan Data** Sebelum informasi diambil dalam KDD, penting untuk memilih data dari kumpulan data operasional. Data yang telah terpilih akan digunakan dalam proses penambangan data.
2. **Pembersihan Data** Sebelum data mining bisa dilakukan, data yang menjadi fokus KDD harus dibersihkan terlebih dahulu. Pembersihan ini termasuk menghapus data yang berulang, memeriksa ketidakkonsistenan, dan memperbaiki kesalahan dalam data.
3. **Pengubah Data** Pengubahan data berarti mengolah data ke dalam format yang tepat untuk ditambang.
4. **Penambangan Data** Penambangan data adalah kegiatan mencari pola atau informasi berharga yang terdapat dalam data terpilih dengan menggunakan teknik atau metode tertentu.
5. **Interpretasi/Evaluasi Pola** informasi yang diperoleh dari proses penambangan data harus disajikan dengan cara yang mudah dipahami oleh pemangku kepentingan. Misalnya, dapat menggunakan visualisasi atau tampilan yang menjelaskan hasil dari sistem.



Gambar 1. Knowledge Discovery in Database

## 2.5 K-Nearest Neighbors (KNN)

K-Nearest Neighbor adalah Algoritma klasifikasi K-Nearest Neighbor adalah metode klasifikasi yang menentukan kategori berdasarkan mayoritas kategori pada K-Nearest Neighbor, dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing, K-Nearest Neighbor merupakan metode untuk melakukan klasifikasi terhadap obyek baru berdasarkan (K) tetangga terdekatnya. K-NN termasuk algoritma supervised learning yang mana hasil dari query instance baru, diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi[13].

Tujuan dari algoritma K-NN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data. Sehingga dengan penerapan algoritma K-Nearest Neighbor dapat mempermudah Toko Kitab Assarif pada penjualan produk dengan mengambil objek baru berdasarkan data yang letaknya terdekat dari data baru tersebut. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidian[14].

Pada algoritma K-NN terdapat 5 (lima) cara untuk mengetahui tetangga terdekat yaitu:

1. Jarak *Euclidean*

2. Jarak *Manhattan*
3. Jarak *Cosine*
4. Jarak *Corelation*
5. Jarak *Hamming*

Pada Penelitian ini penulis hanya menggunakan jarak *euclidean*, maka rumus perhitungan jarak dengan Euclidean seperti dibawah ini.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Keterangan

- x dan y adalah dua buah vektor berdimensi n
- xi dan yi adalah elemen ke-i dari masing-masing vector x dan y
- d(x,y) adalah jarak (Euclidean distance) antara vektor x dan y

Langkah-langkah untuk menghitung algoritma K-NN

1. Menentukan parameter K (jumlah tetangga paling dekat).
2. Menghitung kuadrat jarak euclid (query instance) masing masing objek terhadap training data yang diberikan
3. Kemudian mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak euclid terkecil
4. Mengumpulkan kategori Y (klasifikasi nearest neighbor berdasarkan nilai k)
5. Dengan menggunakan kategori nearest neighbor yang paling mayoritas maka dapat diprediksi objek yang baru. (k) pada algoritma k-nearest neighbor adalah banyaknya tetangga terdekat yang akan digunakan sebagai titik untuk melakukan klasifikasi pada data atau objek baru. Untuk menghitung jarak antar objek data pada algoritma k-nearest neighbor dapat dilakukan dengan beberapa cara, salah satunya adalah dengan Ecludian Distance.

## 2.6. Tahapan Implementasi Metode K-NN

- a. Pengumpulan dan Persiapan Data  
Tahap awal dimulai dengan mengumpulkan data penjualan kitab dan buku dari Toko Kitab Assyarif selama periode tertentu. Data ini mencakup informasi seperti nama barang, harga, jumlah terjual, stok awal, stok akhir, dan periode waktu (bulan). Setelah dikumpulkan, data dibersihkan dan dipersiapkan dalam format yang sesuai untuk proses analisis.
- b. Penentuan Fitur (Feature Selection)  
Fitur-fitur yang dianggap relevan untuk mempengaruhi tingkat penjualan dianalisis dan dipilih, seperti: Bulan, Harga barang, Jumlah stok awal, Jumlah stok akhir Jumlah terjual Fitur-fitur ini akan menjadi dasar perhitungan dalam menentukan kedekatan antar data.
- c. Pemberian Label (Kategori Target)  
Setiap data diberi label berdasarkan hasil penjualannya, misalnya dikategorikan sebagai "Laris" atau "Tidak Laris". Penentuan label ini didasarkan pada nilai ambang tertentu (threshold), seperti jumlah barang terjual melebihi atau di bawah nilai tertentu.
- d. Normalisasi Data (Opsional)  
Jika diperlukan, data dinormalisasi agar skala antar fitur seimbang. Ini penting agar tidak ada satu fitur yang mendominasi perhitungan jarak karena memiliki rentang nilai yang lebih besar.
- e. Penentuan Nilai K (Jumlah Tetangga Terdekat)  
Nilai K menentukan berapa banyak tetangga terdekat yang akan diperhitungkan untuk mengklasifikasikan data baru. Nilai ini dipilih berdasarkan eksperimen dan pengujian akurasi (misalnya K=3, K=5, dst).
- f. Perhitungan Jarak (Distance Calculation)  
Dilakukan perhitungan jarak antara data baru (data yang ingin diprediksi) dengan seluruh data training menggunakan metode jarak tertentu, seperti: Euclidean Distance, Manhattan Distance, Cosine Similarity Dalam penelitian ini digunakan Euclidean Distance, yang merupakan metode paling umum dan sederhana.

Rumus Euclidean Distance:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Keterangan

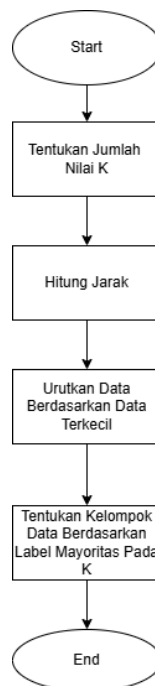
- $d(p, q)$  = jarak antara titik (vektor)  $p$  dan  $q$
- $n$  = jumlah atribut (fitur)
- $p_i, q_i$  = nilai atribut ke- $i$  dari masing-masing titik
- Simbol  $\sum$  (sigma) = menjumlahkan semua selisih kuadrat dari tiap atribut

g. Klasifikasi dan Prediksi

Setelah menghitung jarak, diambil  $K$  data dengan jarak terdekat, lalu dilakukan voting atau pencarian mayoritas label dari tetangga tersebut. Jika mayoritas data tetangga tergolong “Laris”, maka data baru diprediksi juga sebagai “Laris”.

h. Evaluasi dan Validasi Hasil

Hasil prediksi dievaluasi menggunakan metrik akurasi, presisi, dan recall, terutama jika data dibagi menjadi data training dan data uji (testing). Evaluasi ini membantu memastikan bahwa model dapat memberikan hasil prediksi yang andal dan bisa digunakan untuk membantu pengambilan keputusan penjualan di masa depan.



Gambar 2. Flowchart

## 2.7. Tahapan Implementasi Sistem Prediksi

Penelitian ini menciptakan sebuah sistem ramalan yang berbasis web dan dibangun dengan menggunakan framework Streamlit. Sistem tersebut ditujukan untuk meramalkan kategori penjualan buku dan kitab di Toko Kitab Assyarif, yang terbagi menjadi dua jenis yaitu Laris dan Tidak Laris. Proses ramalan memanfaatkan algoritma K-Nearest Neighbor (K-NN) karena efektivitasnya dalam mengklasifikasikan data berdasarkan jarak[15].

Sistem ini dilengkapi dengan berbagai fitur seperti persiapan data, pemrosesan data, pemilihan nilai  $k$ , evaluasi model, dan hasil prediksi. Tahapan implementasi sistem terbagi menjadi beberapa langkah sebagai berikut:

### 1. Persiapan Data

Data yang digunakan untuk pelatihan didapat dari ringkasan penjualan buku dan kitab setiap bulan, mulai dari tahun 2023 hingga tahun 2024. Informasi ini disimpan dalam file bernama `data_training.csv`, yang mencakup atribut berikut:

- Jumlah Terjual
- Stok Awal
- Stok Akhir
- Kategori (Laris atau Tidak Laris)

Untuk meningkatkan ketepatan model, ditambahkan fitur baru yaitu Persentase Terjual, yang dihitung dari Terjual dibagi Stok Awal. Selanjutnya, dilakukan normalisasi pada fitur numerik dengan menggunakan `StandardScaler` dari pustaka `scikit-learn` agar semua data memiliki skala yang uniform sebelum melanjutkan pelatihan model.

## 2. Pemrosesan Data

Data pengujian diambil dari file `data_uji.csv` yang mencakup informasi penjualan dari Agustus hingga Oktober 2024. Tidak ada label kategori dalam data ini, karena tujuan utamanya adalah untuk menguji kemampuan prediksi model K-NN yang telah dilatih sebelumnya. Sistem membaca setiap baris dari data tersebut, menghitung persentase penjualan, melakukan normalisasi, dan menggunakan model untuk mengklasifikasikan kategori penjualan.

## 3. Pemilihan Nilai K

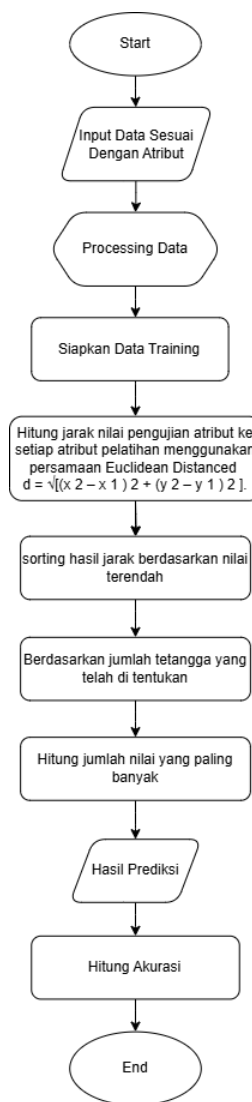
Sistem menawarkan pengaturan interaktif kepada pengguna untuk menentukan nilai K (jumlah tetangga terdekat) dalam algoritma K-NN dengan bantuan komponen slider yang terletak di sidebar. Dalam percobaan ini, nilai K yang digunakan adalah 3, yang menunjukkan hasil terbaik berdasarkan uji validasi silang.

## 4. Evaluasi Model

Penilaian model dilakukan dengan metode 5-Fold Cross Validation. Hasil dari validasi silang menunjukkan akurasi sebesar 0.80 atau 80%, yang menunjukkan bahwa model memiliki kemampuan yang baik dalam memprediksi kategori penjualan. Angka ini ditampilkan secara langsung di sidebar aplikasi sebagai tanda kinerja model.

## 5. Hasil Prediksi

Setelah model selesai dilatih dan divalidasi, sistem digunakan untuk meramalkan data uji. Di bawah ini adalah menunjukkan hasil ramalan untuk data penjualan kitab dan buku menggunakan `streamlit`



**Gambar 3.** Diagram Alir

### 3. HASIL DAN PEMBAHASAN

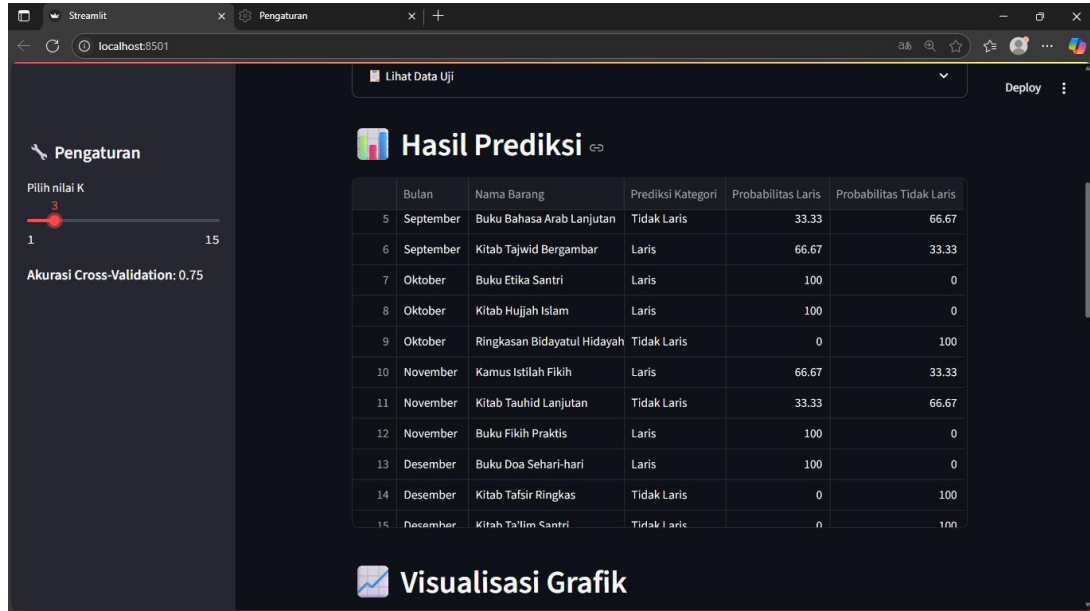
Dalam penelitian ini menghasilkan sebuah sistem prediksi penjualan kitab dan buku menggunakan algoritma K-Nearest Neighbor (K-NN) yang diimplementasikan dalam platform berbasis web menggunakan Streamlit. Tujuan dari sistem ini adalah untuk membantu pengelola Toko Kitab Assyarif PP. Salafiyah Syafi'iyah Sukorejo dalam mengklasifikasikan produk ke dalam kategori "Laris" dan "Tidak Laris" berdasarkan data penjualan historis.

#### 3.1. Akurasi Model

Proses evaluasi dilakukan dengan metode 5-Fold Cross Validation, yang menghasilkan nilai akurasi sebesar 0.75 (75%). Nilai ini menunjukkan bahwa model dapat memprediksi dengan ketepatan 75 dari 100 kasus pada data yang belum dikenali sebelumnya. Nilai akurasi ini dianggap cukup baik mengingat model menggunakan data penjualan historis yang bervariasi dan berbasis atribut numerik seperti stok awal, stok akhir, dan jumlah terjual.

#### 3.2. Hasil Prediksi

Data terdiri dari penjualan bulan Agustus hingga Oktober, dengan masing-masing produk dianalisis probabilitasnya untuk masuk ke dalam kategori “Laris” atau “Tidak Laris”. Hasil prediksi ditampilkan dalam bentuk tabel interaktif yang mencakup kolom: Bulan, Nama Barang, Prediksi Kategori, Probabilitas Laris, dan Probabilitas Tidak Laris.

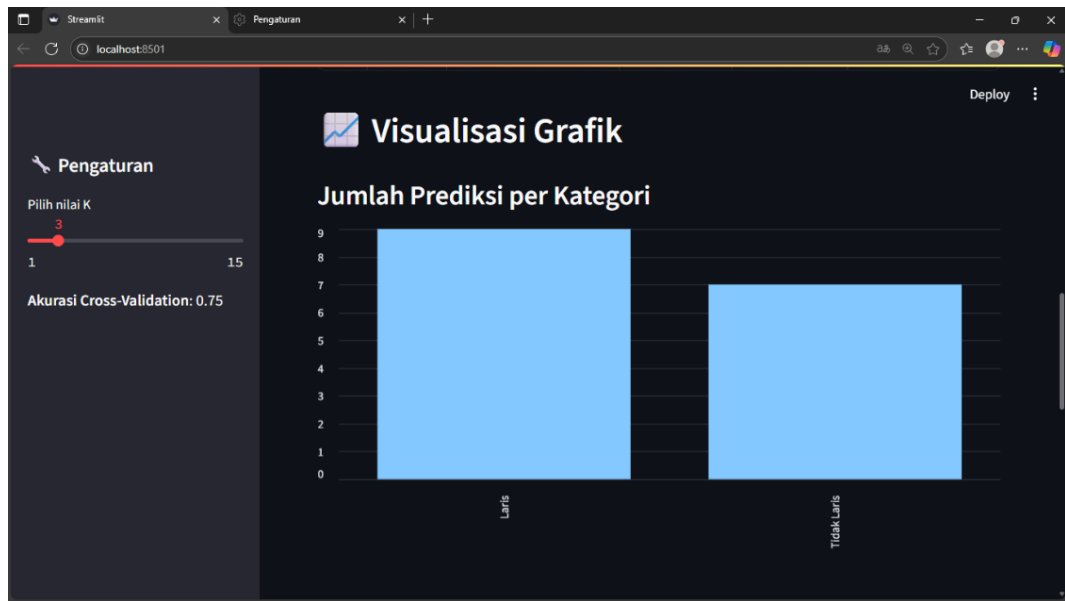


*Gambar 4. Hasil Prediksi*

Prediksi ini menunjukkan bahwa model mampu mengidentifikasi pola penjualan berdasarkan atribut numerik dan memberikan probabilitas keyakinan dari hasil klasifikasi tersebut

### 3.3. Visualisasi Grafik

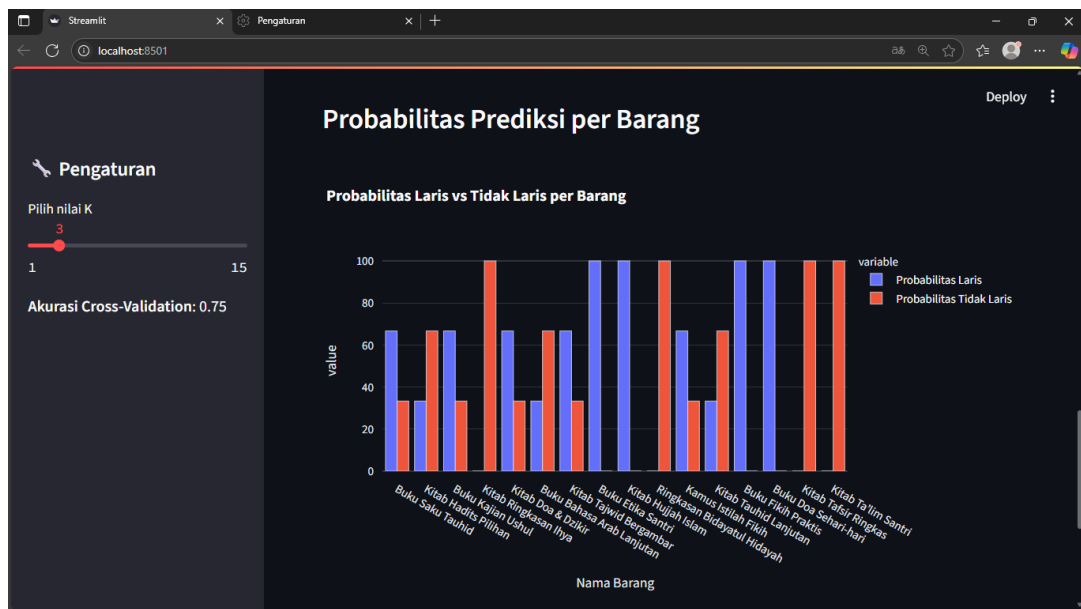
Untuk memperjelas distribusi hasil klasifikasi, sistem menampilkan visualisasi grafik jumlah prediksi per kategori. Berdasarkan grafik tersebut, sebanyak 9 barang diprediksi masuk ke dalam kategori “Laris”, sedangkan 7 barang diprediksi sebagai “Tidak Laris” Visualisasi ini membantu pengguna dalam memahami sebaran hasil prediksi secara cepat tanpa harus membaca seluruh tabel satu per satu. Selain itu, hal ini juga memperlihatkan bahwa model memiliki kecenderungan yang cukup seimbang dalam memetakan produk ke dua kategori tersebut, yang menunjukkan bahwa tidak terjadi overfitting terhadap satu kelas tertentu. Grafik ini juga bermanfaat dalam pengambilan keputusan oleh pihak toko, karena dapat digunakan sebagai indikator awal untuk menentukan strategi stok, promosi, dan pembelian ulang berdasarkan potensi kelarisan produk.



*Gambar 5. Visualisasi Grafik*

### 3.4. Visualisasi Probabilitas Prediksi per Barang

Grafik ini menampilkan probabilitas hasil prediksi dari masing-masing barang terhadap dua kategori: Laris dan Tidak Laris. Probabilitas ini dihitung berdasarkan proporsi jumlah tetangga terdekat ( $K$ ) dari masing-masing kelas dalam algoritma  $K$ -NN. Setiap batang biru menunjukkan tingkat probabilitas bahwa suatu barang masuk kategori Laris, sementara batang merah menunjukkan kemungkinan sebagai Tidak Laris seperti gambar berikut:



*Gambar 6. Visualisasi Probabilitas*

Visualisasi ini tidak hanya menunjukkan hasil klasifikasi akhir, tetapi juga memberikan wawasan tingkat keyakinan model, yang dapat digunakan untuk membantu pengambilan keputusan yang lebih hati-hati oleh pemilik toko, misalnya untuk produk dengan margin tipis atau dengan stok terbatas.

#### 4. KESIMPULAN

Berdasarkan research dan penerapan sistem untuk memprediksi penjualan buku dan kitab dengan menggunakan metode K-Nearest Neighbor (K-NN) di Toko Kitab Assyarif PP. Salafiyah Syafi'iyah Sukorejo, beberapa kesimpulan dapat ditarik sebagai berikut:

1. Metode K-Nearest Neighbor (K-NN) terbukti sebagai alat yang efektif dalam meramalkan kategori penjualan produk (Laris atau Tidak Laris) dengan memanfaatkan data penjualan yang telah ada. Algoritma ini mampu melakukan klasifikasi berdasarkan jarak antar data, menghasilkan prediksi yang cukup tepat.
2. Sistem prediksi yang berbasis web ini, yang dikembangkan dengan bantuan Streamlit, mendukung toko dalam mengklasifikasikan kategori penjualan secara otomatis, cepat, dan interaktif. Fitur utama sistem ini mencakup normalisasi data, pengaturan nilai K, validasi model, serta kemampuan untuk mengeksport hasil prediksi dalam bentuk PDF dan Excel.
3. Hasil evaluasi model dengan metode 5-Fold Cross Validation menunjukkan sistem prediksi memiliki akurasi sebesar 80%, yang mencerminkan kinerja model yang cukup solid dan dapat dimanfaatkan sebagai alat dalam pengambilan keputusan strategis.
4. Prediksi yang diperoleh menunjukkan adanya fluktuasi penjualan dari bulan ke bulan, dengan bulan September 2024 mencatat produk yang lebih laris dibandingkan Agustus dan Oktober. Informasi ini dapat digunakan sebagai dasar untuk merencanakan pengadaan stok, promosi, atau distribusi produk pada bulan-bulan yang akan datang.
5. Dengan adanya sistem prediksi ini, Toko Kitab Assyarif diharapkan bisa mengatur persediaan barang dengan lebih efisien, mengurangi risiko kelebihan atau kekurangan stok, dan meningkatkan pelayanan kepada pelanggan melalui ketersediaan produk yang lebih sesuai.

#### UCAPAN TERIMAKASIH

Puji syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan karunia-Nya sehingga jurnal ini dapat diselesaikan dengan baik.

Penulis menyampaikan terima kasih kepada dosen pembimbing yang telah memberikan bimbingan dan arahan selama proses penyusunan jurnal. Ucapan terima kasih juga ditujukan kepada pimpinan dan staf Toko Kitab Assyarif atas izin dan data yang diberikan sebagai bahan penelitian.

Terima kasih juga kepada seluruh dosen, keluarga tercinta, serta teman-teman yang telah memberikan dukungan, doa, dan semangat selama proses penyusunan jurnal ini. Semoga jurnal ini dapat memberikan manfaat dan menjadi kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang data mining.

#### REFERENSI

- [1] B. Susilo, N. A. Ramdhan, and O. S. Bachri, "Application of the K-Nearest Neighbor Algorithm for Predicting Digital Product Sales Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Produk Digital," vol. 4, no. October, pp. 1466–1476, 2024.
- [2] M. Kafil and F. T. Industri, "Penerapan Metode K-Nearest Neighbors," vol. 3, no. 2, pp. 59–66, 2019.
- [3] L. Suryadi, N. Ngajiyanto, N. E. Pratiwi, F. Ardhy, and P. Riswanto, "Penerapan Data Mining Prediksi Penjualan Mebel Terlaris Menggunakan Metode K-Nearest Neighbor(K-Nn) (Studi Kasus : Toko Zerita Meubel)," *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 7, no. 2, pp. 174–184, 2022, doi: 10.32767/jusim.v7i2.1697.
- [4] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [5] A. Azis, A. T. Zy, and A. S. Sunge, "Prediksi Penjualan Obat Dan Alat Kesehatan Terlaris Menggunakan Algoritma K-Nearest Neighbor," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 1, pp. 117–124, 2024, doi: 10.47233/jteksis.v6i1.1078.
- [6] A. A. Putri, "Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor ( Studi Kasus : PT . Central Brastagi Utama )," vol. 1, no. 6, pp. 354–361, 2021.
- [7] S. Tri Wijaya, I. Hartami Santi, and Z. Wulansari, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Harga Jagung Dengan Pengujian Rmse," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1255–1260, 2023, doi: 10.36040/jati.v7i2.7391.
- [8] D. Ojol, "Implementasi Metode K-Nearest Neighbor Dalam Menentukan Waktu Optimal Penarikan Pesanan Driver Ojol 1,2," vol. 11, no. 2, pp. 255–269, 2024.
- [9] A. Yogiarto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
- [10] M. Bakriansyah, A. Subair, and W. Setiawan, "Penerapan Data Mining Untuk Memprediksi Penjualan Produk Terlaris UD Timbul Jaya Menggunakan Metode K-Nearest Neighbor," vol. 4, no. 2, pp. 327–337, 2025.

- 
- [11] Taufik Hidayat, Yuni Handayani, and Ahmad Syaifudin, "Implementasi Algoritma K-Nearest Neighbor Untuk Meningkatkan Penjualan Produk Meuble dan Furniture," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 9, no. 2, pp. 118–124, 2023.
- [12] K. Neighbor, D. Maulida, and Y. R. Nasution, "Prediksi Penjualan Produk Pepsodent Unilever dengan Algoritma," vol. 5, no. 4, pp. 883–892, 2024, doi: 10.47065/josyc.v5i4.5718.
- [13] A. Ghani Muttaqin, K. Auliasari, and F. Santi Wahyuni, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada Pt.Wika Industri Energy," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 4, no. 2, pp. 1–6, 2020, doi: 10.36040/jati.v4i2.2728.
- [14] F. B. Putra, M. T. Chulkamdi, and F. Febrinita, "IMPLEMENTASI DATA MINING UNTUK MEMREDIKSI DATA STOK FUKUBI OUTFIT MENGGUNAKAN METODE K- NEAREST NEIGHBOR," vol. 8, no. 5, pp. 9907–9914, 2024.
- [15] F. Hasanah, T. Suprpti, N. Rahaningsih, and I. Ali. "Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Buku Berdasarkan Peminatan," *J. Account. Inf. Syst.*, vol. 5, no. 1, pp. 102–111, 2022, doi: 10.32627/aims.v5i1.467.