

BAB II TINJAUAN PUSTAKA

2.1 Teori Dasar

2.1.1 Perguruan Tinggi

Berdasarkan Peraturan Pemerintah Republik Indonesia Nomor 30 Tahun 1990 Tentang Pendidikan Tinggi bahwa "*Pendidikan tinggi adalah satuan pendidikan yang menyelenggarakan pendidikan tinggi yang merupakan jenjang yang lebih tinggi daripada pendidikan menengah di jalur pendidikan sekolah*".

Institusi perguruan tinggi menyediakan berbagai program studi dan penelitian di berbagai disiplin ilmu. Fakultas-fakultas di perguruan tinggi biasanya menawarkan program sarjana, magister, dan doktoral dalam bidang-bidang seperti ilmu sosial, ilmu alam, teknik, kedokteran, dan humaniora. Selain itu, perguruan tinggi berperan penting dalam pengembangan penelitian ilmiah dan kontribusi terhadap kemajuan ilmu pengetahuan dan teknologi. Di banyak kasus, perguruan tinggi juga menjadi tempat untuk kegiatan ekstrakurikuler, aktivitas sosial, dan pengembangan kepribadian mahasiswa.

Selain itu, di dalam Peraturan Pemerintah Republik Indonesia Nomor 30 Tahun 1990 Tentang Pendidikan Tinggi juga memuat tujuan pendidikan tinggi, yaitu:

- "*Menyiapkan peserta didik menjadi anggota masyarakat yang memiliki kemampuan akademik dan/atau profesional yang dapat menerapkan, mengembangkan dan/atau menciptakan ilmu pengetahuan, teknologi dan/atau kesenian*".
- "*Mengembangkan dan menyebarkan ilmu pengetahuan, teknologi dan/atau kesenian serta mengupayakan penggunaannya untuk meningkatkan taraf kehidupan masyarakat dan memperkaya kebudayaan nasional*".

2.1.2 Penerimaan Mahasiswa Baru

Mahasiswa diterima di perguruan tinggi negeri dan swasta melalui proses Penerimaan Mahasiswa Baru (PMB) yang dilakukan. Proses ini melibatkan seleksi oleh perguruan tinggi atau universitas untuk menerima calon mahasiswa ke dalam program studi yang tersedia. Setiap tahun, perguruan tinggi melaksanakan penerimaan mahasiswa baru ini yang umumnya melibatkan beberapa tahapan seperti pengumpulan berkas pendaftaran, ujian tertulis, wawancara, dan evaluasi prestasi akademik. Biasanya, calon mahasiswa baru mendaftar secara online atau offline sesuai dengan prosedur yang berlaku. Mereka juga diminta untuk menyerahkan dokumen seperti transkrip nilai, surat rekomendasi, dan hasil ujian tertentu. (P. Andriani, 2019).

2.1.3 Data

Data adalah pengukuran atau pengamatan yang dikumpulkan sebagai sumber informasi. Data mengacu pada seperangkat nilai yang pada variabel (apa yang diukur) dan unit pengamatan (anggota sampel atau populasi). Data adalah bagian penting dalam statistik, yang perlu dipahami untuk diterapkan pengukuran statistik dengan benar dalam menyimpulkan asumsi tertentu dengan benar. Data mempunyai peran penting dalam bidang statistik. Pengetahuan dan pemahaman tentang data akan membantu dalam menerapkan pengukuran statistik secara akurat dari data mentah untuk menyimpulkan hasil yang bermakna (Handayani, 2022). Jenis data yang berbeda merepresentasikan data yang berbeda juga. Ada dua jenis data berdasarkan sifat diantaranya.

1) Data Kualitatif

Data yang disajikan berbentuk verbal atau naratif merupakan data kualitatif yang dikumpulkan melalui diskusi kelompok, wawancara, kuesioner dengan pertanyaan terbuka, serta situasi lain yang kurang terstruktur. Data kualitatif umumnya memiliki

sifat objektif, sehingga dapat menghasilkan berbagai interpretasi dari setiap pembacanya.

2) **Data Kuantitatif**

Data kuantitatif merupakan data untuk direpresentasikan berbentuk angka, di mana angka-angka tersebut bisa bervariasi besar atau kecil. Angka-angka ini mungkin sesuai dengan kategori atau label tertentu. Data kuantitatif bisa didapatkan melalui survei yang menghasilkan jawaban kaku berupa angka. Contohnya, usia Bambang adalah 21 tahun, dan tinggi rata-rata siswa di kelas A adalah 170 cm.

2.1.4 **Basis Data**

Basis data adalah sebuah lokasi di dalam sebuah organisasi atau bisnis di mana data dihubungkan untuk memfasilitasi pencarian dan penggunaan kembali data yang lebih cepat dan mudah. Sebaliknya, sistem basis data menggunakan komputer untuk mengelola dan mengatur catatan untuk menyimpan dan melestarikan semua data yang tersedia dalam organisasi atau bisnis, memberikan informasi terbaik bagi pengguna untuk mengambil keputusan (Adyanata Lubis, 2016). Berikut ini adalah beberapa bagian yang membentuk sistem basis data:

1) *Hardware*

Perangkat komputer standar yang mendukung pemrosesan data adalah perangkat keras yang dimaksud. Di dalam sistem basis data, perangkat keras yang sering digunakan meliputi server database, hard disk, dan komponen lainnya.

2) *Operating System*

Sistem operasi, manajemen sumber daya, dan eksekusi program semuanya dilakukan oleh program pada komputer.

3) *Database Management System (DBMS)*

Pengguna tidak langsung melakukan operasi fisik pada basis data, tetapi mengatur penyimpanan, modifikasi, dan

pengambilan data menggunakan program yang dikenal sebagai sistem manajemen basis data (DBMS).

4) Pengguna (*User*)

Pengguna menggunakan program yang ditulis dalam bahasa pemrograman untuk berkomunikasi dengan basis data dan mengelola data.

5) Aplikasi atau perangkat lain

Aplikasi ini bergantung pada kebutuhan, di mana pengguna basis data dapat membuat program khusus yang mudah digunakan untuk memasukkan, mengubah, dan mengambil data. Program-program tersebut dapat dijalankan langsung dengan DBMS atau ditulis menggunakan bahasa pemrograman.

2.1.5 Algoritma

Algoritma merupakan metode efektif yang direpresentasikan sebagai serangkaian instruksi terbatas yang telah dijelaskan secara jelas untuk menyelesaikan suatu fungsi. Algoritma pemecahan masalah, sederhananya, adalah serangkaian tahapan logis. Ketika dihadapkan pada suatu masalah, seseorang harus melakukan pendekatan terhadap solusinya secara logis. Prosedur langkah demi langkah yang digunakan untuk perhitungan dikenal sebagai algoritme dalam disiplin ilmu komputer dan matematika (Sigit Susanto, dkk., 2022).

Urutan algoritma dimulai dengan kondisi atau input awal, diikuti oleh instruksi-instruksi yang menjelaskan komputasi. Saat dieksekusi dan diproses melalui urutan kondisi yang telah didefinisikan dengan baik, ini menghasilkan output dan berakhir pada kondisi akhir yang ditentukan. Dalam hal pemrosesan data komputer, algoritme sangat penting. Insinyur perangkat lunak menggunakan algoritme dalam sistem komputer untuk membuat logika yang membuat perangkat lunak lebih mudah digunakan. Ada tiga jenis algoritma yang mendasar, yaitu.

1) *Sequence Algorithm*

Algoritme ini mengikuti serangkaian langkah yang harus diselesaikan secara berurutan. Mendidihkan air adalah ilustrasi langsung dari algoritme ini, di mana setiap langkah harus diselesaikan secara berurutan, mulai dari atas dan terus ke bawah.

2) *Looping Algorithm*

Algoritme yang mengulang serangkaian tahapan yang telah ditentukan sebelumnya. Dalam beberapa situasi, ada kebutuhan untuk melakukan langkah-langkah tertentu berulang kali.

3) *Conditional Algorithm*

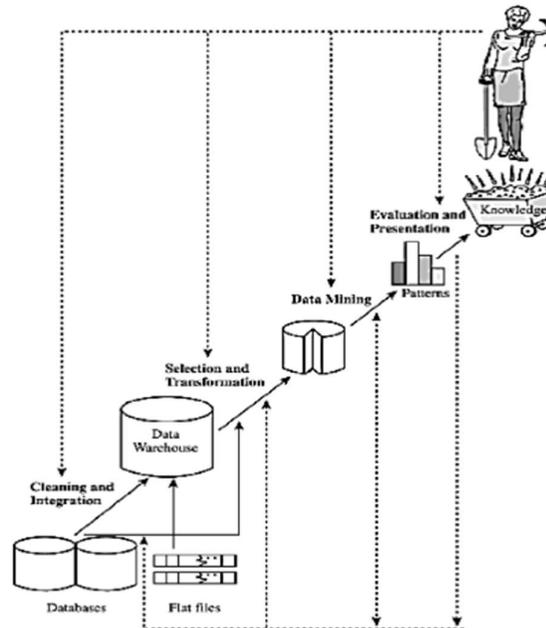
Algoritme ini melanjutkan ke fase berikutnya jika kondisi yang telah ditentukan terpenuhi.

2.1.6 Knowledge Discovery in Database (KDD)

Proses menemukan pengetahuan dalam basis data disebut dengan *Knowledge Discovery in Database (KDD)*. Proses penggalian atau pengenalan pola, pengetahuan, dan informasi yang berpotensi berguna dari kumpulan data yang sangat besar tercakup dalam deskripsi KDD yang komprehensif. Pengetahuan dan data yang dihasilkan oleh KDD harus dapat diandalkan, orisinal, dapat dipahami, dan praktis. Proses menemukan data yang valid, unik, mungkin berguna, dan akhirnya memiliki pola yang masuk akal disebut sebagai penemuan pengetahuan dalam database. Ini bukanlah tugas yang mudah untuk dicapai.

Mengintegrasikan dan menavigasi berbagai sumber data merupakan masalah umum dalam penemuan pengetahuan, terutama ketika data menjadi lebih multidimensi dan membutuhkan metode baru untuk pengenalan pola. Penemuan pengetahuan dalam data, atau KDD, didefinisikan sebagai proses mengeksplorasi dan menganalisis data dalam jumlah besar untuk mendapatkan

pengetahuan dan informasi yang bermanfaat (Fitri Marisa et al., 2021).



Gambar 2.1 Tahapan KDD

Sumber : Han & Kamber (2006)

Gambar 2.1 mengilustrasikan fase-fase utama dari proses KDD, yang meliputi tahapan-tahapan berikut:

1) *Data Cleaning*

Menghilangkan duplikasi, memastikan konsistensi data, dan memperbaiki kesalahan pengetikan adalah bagian dari proses pembersihan data. Secara umum, data yang diperoleh dari hasil uji coba atau database perusahaan sering kali tidak lengkap, mengandung kekurangan seperti data yang hilang atau tidak akurat atau kesalahan ketik. Selain itu, terkadang terdapat properti data yang tidak sesuai dengan teori *data mining* yang sedang diteliti. Dengan semakin sedikitnya data yang diproses dan semakin sedikitnya data yang kompleks yang ditangani, prosedur pembersihan data

ini juga dapat berdampak pada output informasi dari teknik *data mining*.

2) ***Data Integration***

Untuk membuat satu *database* baru yang dibutuhkan KDD, data harus diintegrasikan dari beberapa *database* atau digabungkan dengan informasi terkait lainnya. Integrator data harus menggabungkan beberapa set data dan menghilangkan *noise* dari sumber data untuk menyelesaikan tahap pembersihan dan integrasi KDD.

3) ***Data Selection***

Metode memilih informasi yang relevan untuk diperiksa dari data operasional. Basis data lain berisi data yang dipilih.

4) ***Data Transformation***

Yaitu, mengubah format data ke dalam bentuk tertentu yang dapat digunakan oleh metode *data mining*. Beberapa teknik umum tertentu, seperti analisis asosiasi dan pengelompokan, membutuhkan data masukan dalam format kategorikal.

5) ***Data Mining***

Menggunakan strategi, taktik, atau algoritme tertentu untuk mencari tren atau data yang menarik.

6) ***Pattern Evaluation***

Mengenali tren yang menarik dalam temuan *data mining*. Pada titik ini, keberhasilan hipotesis saat ini ditentukan dengan menganalisis output dari teknik penggalian data seperti membedakan pola atau model prediksi.

7) ***Knowledge Presentation***

Menampilkan pola informasi yang dihasilkan oleh metode *data mining*. Grafik ini memfasilitasi komunikasi yang jelas dan sederhana dari hasil *data mining*.

2.1.7 *Data Warehouse*

Menurut Bill Inmon, yang diakui sebagai bapak *Data Warehouse*, Sebuah "gudang data" adalah sekelompok kumpulan data yang berorientasi pada subjek, gabungan, tidak berubah-ubah, dan varian waktu yang membantu manajer mengambil keputusan. Definisi lain yang terkait dengan *Business Intelligence* (BI) menyatakan bahwa *data warehouse* adalah basis data yang didesain untuk digunakan dalam aktivitas BI, membantu pengguna dalam memahami dan meningkatkan kinerja organisasi mereka (Lane dan Potineni, 2014). *Data warehouse* seringkali dikaitkan dengan BI karena menjadi sumber data untuk analisis bisnis. Fokus *data warehouse* adalah menyediakan data untuk analisis bisnis, bukan untuk proses transaksi atau operasional bisnis. *Data warehouse* juga sering disebut sebagai *database Online Analytical Processing* (OLAP) karena perannya dalam memenuhi kebutuhan analitik (Wardani, dkk., 2019).

Data Warehouse merupakan sumber daya yang memerlukan pengolahan untuk menghasilkan informasi, baik berupa data kualitatif maupun kuantitatif, yang menggambarkan fakta dan memberikan manfaat bagi peneliti atau memberikan gambaran tentang kondisi atau situasi tertentu. Informasi merupakan data yang telah diproses untuk menghasilkan analisis yang diperlukan oleh pengguna (Gunadi & Widiyanto, 2020).

Definisi *data warehouse* berdasarkan fungsinya dalam menyediakan informasi menurut Ponniah (2010) adalah sebagai berikut:

1. Memberikan perspektif yang komprehensif dan terintegrasi terhadap informasi bisnis atau organisasi.
2. Menyediakan akses mudah ke informasi terkini dan historis untuk mendukung pengambilan keputusan strategis.

3. Membuat informasi dari data operasional tersedia tanpa mengganggu sistem operasional utama.
4. Mengintegrasikan informasi dari organisasi atau perusahaan secara konsisten.
5. Menyajikan sumber informasi secara strategis dengan cara yang fleksibel dan interaktif.

Keempat karakteristik data warehouse (terintegrasi, berbasis waktu, berorientasi pada subjek, dan tidak mudah berubah) tetap menjadi panduan dalam pengembangan *data warehouse*. Berorientasi pada subjek menyiratkan bahwa *data warehouse* menyajikan pandangan yang terfokus dan jelas mengenai masalah subjek tertentu dengan mengabaikan data yang tidak relevan dalam proses pengambilan keputusan (Han, Kamber, dan Pei, 2012). Salah satu pendekatan dalam membangun *data warehouse* berorientasi pada subjek adalah penggunaan *data warehouse* untuk memberikan informasi tentang kinerja dosen. Dalam hal ini, dosen adalah subjeknya. Ini berbeda dengan data operasional sehari-hari atau aktivitas dosen yang tercatat di berbagai aplikasi seperti Sistem Informasi Perkuliahan, Sistem Informasi Bimbingan Skripsi, dan Sistem Informasi Kehadiran Pegawai berbasis OLTP (*Online Transactional Processing*). Database OLTP dapat berfungsi sebagai sumber data untuk membangun data warehouse berbasis OLAP (*Online Analytical Processing*).

2.1.8 Data Mining

Teknik mengekstraksi pola dan informasi yang relevan dari data dalam jumlah besar dikenal sebagai *data mining*. Ekstraksi data, analisis, analisis, dan pengumpulan data adalah bagian dari proses ini. Analisis matematis digunakan untuk mengidentifikasi pola dan tren dalam data yang sering kali sulit atau rumit untuk diidentifikasi melalui metode eksplorasi data tradisional, sebagai akibat dari hubungan yang rumit antara data atau volume data yang besar. Menemukan pola yang sebelumnya tidak teridentifikasi adalah

tujuan utama dari *data mining*, yang kemudian dapat diterapkan pada penyelesaian berbagai masalah (M. Arhami & M. Nasir, 2020).

Data mining telah menjadi teknologi yang kuat dengan potensi besar untuk membantu perusahaan mengidentifikasi informasi penting dari data yang mereka kumpulkan mengenai perilaku pelanggan dan prospek. Informasi yang dihasilkan dari *data mining* dapat digunakan sebagai alat bantu pengambilan keputusan untuk mengembangkan bisnis. *Data mining* mampu memberikan solusi atas berbagai pertanyaan bisnis yang kompleks. Selain itu, *data mining* juga dapat digunakan untuk membuat prediksi di masa depan, sehingga pengusaha dapat mengambil keputusan yang efektif, proaktif, dan dinamis. Proses *data mining* secara umum meliputi tiga tahap:

1. Eksplorasi

Prosesnya dimulai dengan menyiapkan banyak data, yang kemudian dibersihkan atau dikurangi sesuai kebutuhan, dan materi duplikat dihapus untuk menyisakan hanya data yang berguna.

2. Pemodelan atau identifikasi pola

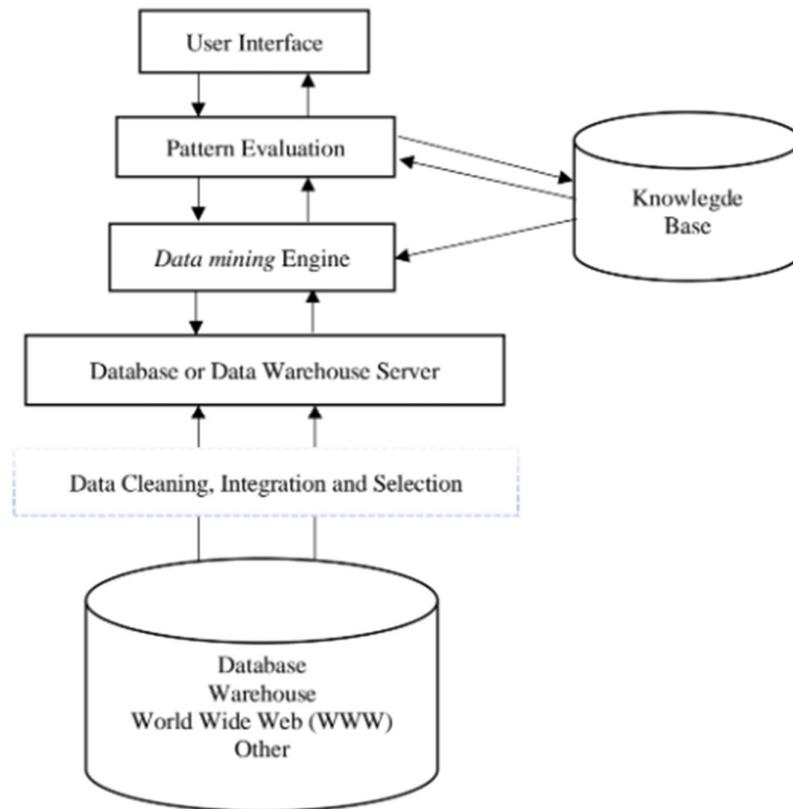
Tahap selanjutnya adalah membangun model statistik yang bertujuan untuk menilai dan menghasilkan prediksi yang paling akurat dan terbaik. Karena perlunya pengujian dan perbandingan temuan yang sering dilakukan, prosedur ini dapat memakan waktu jika menggunakan kumpulan data yang sama untuk model.

3. Penerapan

Dengan menggunakan data pelatihan dan pengujian, model dievaluasi pada langkah terakhir untuk menghasilkan estimasi atau prakiraan hasil yang diinginkan.

Alur proses *data mining* dijelaskan oleh arsitektur *data mining*, yang mencakup segala sesuatu mulai dari administrasi

sumber data hingga interaksi antara pengguna dan sistem *data mining*. Gambar 2.2 menampilkan arsitektur untuk *data mining*.



Gambar 2. 2 Arsitektur *Data Mining*

Sumber : Indriani (2023)

Menurut Efori Buulolo (2020), sumber data untuk *data mining* dapat ditemukan di berbagai lokasi, termasuk *database*, gudang, internet, dan sumber lainnya. Setelah data yang tidak berguna dihilangkan, data dari beberapa sumber ini digabungkan dan dipilih. Selanjutnya, data yang dipilih diformat sedemikian rupa sehingga dapat digunakan untuk penggalian data. Data yang telah dikonversi kemudian diproses untuk menciptakan pengetahuan atau informasi yang dapat ditampilkan kepada pengguna.

2.1.9 Metode Klasifikasi

Klasifikasi adalah proses mengidentifikasi model atau fungsi yang mencirikan dan memisahkan kelas-kelas data atau konsep

untuk memprediksi kelas data yang kelasnya tidak diketahui. Data pelatihan, atau objek data dengan label kelas yang telah ditentukan, dianalisis untuk membangun model. Model akhir dapat ditampilkan dalam berbagai cara, termasuk pohon keputusan dan klasifikasi *IF-THEN*.

Data dikelompokkan menggunakan nilai atribut klasifikasi dan data pelatihan agar algoritme klasifikasi dapat berfungsi. Data baru dikategorikan ke dalam kelompok yang sudah ada sebelumnya dengan menggunakan metode pengelompokan ini. *Decision Tree* adalah salah satu ilustrasi klasifikasi. Dalam sebuah pohon keputusan, setiap *node* menunjukkan tes atribut dari kumpulan data, dan setiap cabang menunjukkan hasil tes. Berdasarkan kondisi *IF THEN outcome*, pohon keputusan yang dibuat dapat dilihat sebagai sekumpulan aturan. Berlawanan dengan label kelas, tujuan pengguna sering kali adalah untuk memperkirakan nilai dari data yang tidak tersedia atau hilang. Prediksi dalam situasi ini terutama berkaitan dengan mengidentifikasi distribusi dari data yang disediakan. (Amril Mutoi & Adam Puspabhuana, 2017).

2.1.10 *K-Nearest Neighbor*

Salah satu pendekatan pembelajaran mesin dan penggalian data yang digunakan untuk regresi dan klasifikasi disebut *K-nearest neighbor*, atau K-NN. K-NN bekerja dengan melihat dataset dan mencari tahu kategori atau nilai tujuan dari data baru dengan melihat sebagian besar data terdekatnya. K-NN menghitung jarak antara data baru dan data lama selama proses berlangsung. Algoritme ini kemudian menggunakan sebagian besar tetangga terdekat dari data baru untuk mengetahui kategori atau nilai tujuannya. Ketika melakukan klasifikasi atau regresi, jumlah tetangga ditentukan oleh nilai K; nilai K yang lebih besar mengindikasikan lebih banyak tetangga untuk digunakan. (Ferdina Kusumah, et al., 2022).

Metode K-NN merupakan bagian dari *Instance-Based Learning* (IBL), yaitu teknik *machine learning* yang mempelajari

saat sebuah masukan diberikan untuk klasifikasi. Berbeda dengan metode fast learner seperti *Decision Tree*, Naïve Bayes, ANN, dan SVM, teknik IBL dikenal sebagai *lazy learner*. Secara umum, metode K-NN terdiri dari dua langkah: menyimpan setiap pola latih saat pelatihan, dan memeriksa semua pola latih untuk menemukan sejumlah k pola terdekat saat melakukan klasifikasi sebuah pola.

Nilai K yang diperoleh dari prosedur K-NN menawarkan akurasi generalisasi terbaik untuk data yang akan datang. Menemukan nilai K yang ideal adalah salah satu kesulitan dalam pelatihan. Bergantung pada jenis atribut data, nilai K dapat dihitung dengan menggunakan jarak, kemiripan, dan perbedaan. Rumus berikut menjelaskan cara menghitung jarak *Euclidean*, yang digunakan untuk menentukan jarak antara dua titik.

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

Gambar 2. 3 Rumus Perhitungan KNN

Langkah-langkah klasifikasi dapat diringkas sebagai berikut:

- 1) Memilih nilai k.
- 2) Menghitung rentang data antara *testing* dan tetangga terdekat nilai k pada *data training*.
- 3) Menentukan kelas yang memiliki jumlah tetangga terdekat terbanyak, dilakukan pemeriksaan dengan cara voting. Kelas yang memiliki jumlah tetangga terdekat terbanyak akan dipilih sebagai hasil klasifikasi.

Performa metode kNN dalam klasifikasi biasanya dievaluasi menggunakan metrik akurasi. Metrik akurasi dapat dihitung dengan menggunakan rumus berikut.

$$\text{Akurasi} = \frac{\text{Jumlah prediksi yang benar}}{\text{Jumlah seluruh prediksi}}$$

Gambar 2. 4 Rumus Akurasi KNN

Ukuran kesederhanaan tertinggi adalah skor akurasi, yang mengukur jumlah prediksi benar dibandingkan dengan total kasus. Saat seluruh prediksi tepat, skor ini mencapai 1.0, tetapi jika semua prediksi salah, skornya menjadi 0. Sayangnya, skor akurasi ini memiliki kekurangan karena tidak memperhitungkan ketidakseimbangan antara *false positive* dan *false negative*.

2.1.11 Python

Python adalah salah satu bahasa komputer yang paling populer. Bahasa ini digunakan di banyak bidang, seperti pembelajaran mesin, pengembangan web, analisis data, dan pengembangan perangkat lunak secara umum. Guido van Rossum adalah penciptanya, dan pada awalnya dirilis pada tahun 1991. *Python* terkenal karena sintaksnya yang sederhana dan mudah dipahami, serta menyediakan banyak pustaka dan modul untuk berbagai keperluan pengembangan perangkat lunak. Selain itu, *Python* memiliki komunitas pengguna yang besar dan aktif, menjadikannya salah satu bahasa pemrograman paling disukai di kalangan pengembang (Irfan Ardiansah & Ryan Hara Permana, 2023).

Sebagai bahasa interpreter, *Python* tidak memerlukan kompilasi (mengubah kode sumber menjadi kode biner). Namun, *Python* mengubah kode sumber ke bentuk intermediate yang dapat dijalankan di berbagai platform komputer. Hal ini membuat *Python* lebih portabel. *Python* termasuk dalam kategori bahasa pemrograman tingkat tinggi yang mendukung berbagai paradigma pemrograman seperti berorientasi objek, imperatif, dan fungsional.

Salah satu keunggulan *Python* adalah sifatnya yang dinamis dengan manajemen memori otomatis (Syamsudin Manai, 2019).

2.2 Tinjauan Studi

Penelitian ini didukung oleh berbagai referensi dari jurnal-jurnal terdahulu yang relevan dengan topik penelitian. Berikut adalah beberapa jurnal tersebut.

1. Jurnal pertama merupakan sebuah penelitian yang dilakukan oleh Adhitya Rahmat D.N, Karina Auliasari, Yosep Agus Pranoto. dengan judul “IMPLEMENTASI METODE *K-NEAREST NEIGHBOR* (KNN) UNTUK SELEKSI CALON KARYAWAN BARU (Studi Kasus: BFI *Finance* Surabaya)” yang dimuat dalam Jurnal JATI (Jurnal Mahasiswa Teknik Informatika) pada tahun 2020. Jurnal ini membahas sebuah penelitian yang bertujuan untuk mengatasi permasalahan dalam proses seleksi calon karyawan di BFI *Finance* Surabaya. Proses seleksi yang melibatkan berkas masuk, tes psikologi, interview, dan pemeriksaan kesehatan seringkali memakan waktu lama dan kurang efektif. Studi kasus dilakukan di BFI *Finance* Surabaya, di mana data calon karyawan baru dikumpulkan dan dianalisis. Hasil penelitian menunjukkan bahwa implementasi KNN dapat membantu dalam mengidentifikasi kandidat yang paling cocok dengan profil yang diinginkan oleh perusahaan. Proses seleksi menjadi lebih efisien dan akurat karena KNN dapat mempertimbangkan multiple atribut dan variabel dalam menentukan kemiripan antara calon karyawan dengan profil yang telah ada. Selain itu, jurnal ini juga mencakup evaluasi kinerja metode KNN dalam konteks seleksi karyawan, seperti tingkat akurasi, sensitivitas, dan spesifisitas. Dengan menggunakan metode ini, BFI *Finance* Surabaya dapat mengoptimalkan proses rekrutmen dan seleksi karyawan baru, meningkatkan keberlanjutan sumber daya manusia, serta mengurangi risiko kesalahan dalam pemilihan calon karyawan. Implementasi KNN sebagai bagian dari sistem seleksi karyawan memberikan kontribusi positif terhadap efektivitas dan efisiensi proses rekrutmen perusahaan.

2. Jurnal kedua merupakan sebuah penelitian yang dilakukan oleh Lusi Suryadi, Ngajiyanto, Novia Eka Pratiwi, Ferly Ardhy dan Pakartika Riswanto. dengan judul “PENERAPAN *DATA MINING* PREDIKSI PENJUALAN MEBEL TERLARIS MENGGUNAKAN METODE *K-NEAREST NEIGHBOR*(K-NN) (STUDI KASUS: TOKO ZERITA MEUBEL)” yang dimuat dalam Jurnal JUSIM (Jurnal Sistem Informasi Musirawas) pada tahun 2022. Jurnal ini mengulas sebuah penelitian yang bertujuan untuk mengatasi masalah prediksi penjualan produk terlaris di Toko Zerita Meubel. Toko ini menjual berbagai jenis produk mebel dan, berdasarkan permintaan konsumen selama setahun terakhir, pemilik toko sudah bisa mengidentifikasi beberapa produk yang laris dan yang kurang laris. Penelitian ini didedikasikan untuk pengembangan sistem prediksi penjualan produk terlaris yang menggunakan teknik *data mining* dan metode *k-nearest neighbour* (K-NN), meskipun pada kenyataannya sistem yang ada saat ini masih bersifat manual dan menghasilkan data yang kurang akurat dan efisien. Di Toko Zerita Meubel, manajemen stok dan efektivitas strategi pemasaran dipengaruhi secara positif oleh *data mining* dengan metode K-NN. Sistem ini meningkatkan efisiensi operasional toko, mengurangi risiko kelebihan atau kekurangan stok, dan memberikan layanan yang lebih personal kepada pelanggan dengan menyediakan produk mebel sesuai preferensi mereka. Penelitian ini turut mengoptimalkan strategi bisnis dan meningkatkan keuntungan toko mebel tersebut.
3. Jurnal ketiga merupakan sebuah penelitian yang dibuat oleh Inna Alvi Nikmatun dan Indra Waspada. dengan judul “IMPLEMENTASI *DATA MINING* UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*” yang dimuat dalam Jurnal Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer pada tahun 2019. Publikasi ini meneliti bagaimana masa studi mahasiswa diklasifikasikan dengan menggunakan teknik penggalian data. Penelitian ini bertujuan memahami dan memprediksi durasi studi mahasiswa berdasarkan atribut nilai mata kuliah mereka. Sebanyak 377

data siswa digunakan untuk proses pengambilan sampel ini; data ini mencakup 72 kualitas nilai mata kuliah dan 1 kelas target yang diwakili oleh periode studi, dan mereka berasal dari kelas yang lulus antara tahun 2007 dan 2011. Penambahan data adalah langkah yang dilakukan setelah penemuan pengetahuan dalam basis data (KDD). Untuk mengklasifikasikan masa studi, algoritma *K-Nearest Neighbour* (K-NN) diterapkan. Cara terbaik untuk mengklasifikasikan masa studi didasarkan pada hasil percobaan yang memilih karakteristik dari semua kelas pilihan dengan tingkat keberhasilan 74,95%. Implikasi penting dalam pemantauan masa studi mahasiswa dimiliki oleh penelitian ini dan dapat membantu perguruan tinggi mengambil tindakan yang tepat untuk meningkatkan efisiensi studi. Langkah-langkah lebih efektif dapat diambil oleh institusi pendidikan untuk memastikan mahasiswa menyelesaikan studi dalam batas waktu yang ditetapkan dengan memahami faktor-faktor yang memengaruhi masa studi.

4. Jurnal keempat merupakan sebuah publikasi yang telah dilaksanakan oleh Aah Sumiah dan Nita Mirantika. dengan judul “Perbandingan Metode *K-Nearest Neighbor* dan Naive Bayes untuk Rekomendasi Penentuan Mahasiswa Penerima Beasiswa pada Universitas Kuningan” yang dimuat dalam Jurnal Buffer Informatika pada tahun 2020. Publikasi ini membandingkan dan mengkontraskan algoritma *K-Nearest Neighbour* (KNN) dan Naive Bayes untuk merekomendasikan penerima beasiswa mahasiswa. Tujuan dari penelitian ini adalah untuk mengidentifikasi metode terbaik yang digunakan oleh Universitas Kuningan untuk memilih kandidat beasiswa. Bagian akademik dapat menggunakan sistem informasi yang dibuat dari data yang telah dianalisa untuk merekomendasikan beasiswa kepada pelamar dengan memanfaatkan *Visual Basic.NET* dan *SQL Server*. Penelitian ini merupakan pengembangan dari penelitian sebelumnya. Dalam studi ini, kedua algoritma dibandingkan untuk menilai efektivitasnya dalam klasifikasi data mahasiswa penerima beasiswa. Pendekatan Naive Bayes

mengungguli metode *K-Nearest Neighbour*, yang memiliki akurasi 40%, menurut hasil penelitian, dengan akurasi 70%. Penelitian ini memberikan wawasan berharga mengenai penerapan kedua algoritma tersebut dalam rekomendasi penerimaan beasiswa, yang dapat mendukung pengambilan keputusan di Universitas Kuningan.

5. Jurnal kelima merupakan sebuah penelitian yang dilakukan oleh Hozairi, Anwari dan Syariful Alim. dengan judul “IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL *K-NEAREST NEIGHBOR*, *DECISION TREE* SERTA NAIVE BAYES” yang dimuat dalam Jurnal *Network Engineering Research Operation* pada tahun 2023. Artikel ini membahas penggunaan pendekatan *data mining* untuk klasifikasi kelulusan siswa, khususnya dengan menggunakan alat *Orange Data Mining*. *Decision Tree*, Naive Bayes, dan *K-Nearest Neighbor* (K-NN) adalah tiga model kategorisasi utama, diintegrasikan oleh penelitian ini untuk meningkatkan akurasi dan keandalan prediksi kelulusan mahasiswa. Pola berdasarkan kemiripan dengan contoh yang ada diidentifikasi oleh metode K-NN, sedangkan aturan keputusan berbentuk pohon dihasilkan oleh *Decision Tree*. Teorema Bayes digunakan oleh model Naive Bayes untuk memperkirakan probabilitas kelulusan berdasarkan fitur-fitur yang diberikan. Penelitian ini dilakukan pada mahasiswa angkatan 2016 Program Studi Teknik Informatika Universitas Islam Madura. Penelitian ini bertujuan untuk menilai kekuatan prediksi dari tiga model - *Decision Tree*, Naive Bayes, dan *K-Nearest Neighbor* (K-NN) - dalam hal prediksi kelulusan mahasiswa. Model evaluasi *Confusion Matrix* dan ROC diimplementasikan selama prosedur pengujian data, yang dilakukan dengan menggunakan *K-Fold Cross Validation* (K=5). Hasil perbandingan antara ketiga model tersebut menunjukkan bahwa Naive Bayes lebih akurat (89% vs 77%) dibandingkan K-NN (77% vs 74%) dan *Decision Tree* 74% vs 74%. Penelitian ini menyarankan untuk menggunakan model Naive Bayes untuk mengklasifikasikan tingkat

kelulusan mahasiswa di Program Studi Teknik Informatika Universitas Islam Madura.

6. Jurnal keenam merupakan sebuah jurnal yang berjudul “Penerapan *Data Mining* untuk Memprediksi Tingkat Penjualan pada Produk Layanan Spesialis Kecantikan Menggunakan Metode *K-Nearest Neighbor*” yang ditulis oleh Richard Parulian Hutapea, Analisis komprehensif dilakukan oleh penulis untuk memprediksi tingkat penjualan produk layanan pada PT. Joanne Studio adalah studio kosmetik yang berfokus pada ekstensi bulu mata. Berbagai masalah dalam manajemen stok dihadapi oleh PT. Joanne Studio karena belum memiliki sistem yang dapat menghasilkan prediksi penjualan yang akurat. Penumpukan barang yang tidak terjual, kerusakan barang karena disimpan terlalu lama, atau bahkan kekurangan stok saat permintaan tinggi sering terjadi sebagai akibatnya. Perusahaan mengalami kerugian sebagai akibat dari semua ini. Penelitian ini bermaksud untuk menggunakan metode *data mining* dan algoritma *K-Nearest Neighbor* (KNN) untuk meramalkan tingkat penjualan untuk memecahkan masalah ini. Pengumpulan data penjualan selama enam bulan terakhir dilibatkan oleh penelitian ini yang kemudian dianalisis dengan teknik klasifikasi *data mining* menggunakan algoritma KNN. Mengolah data penjualan menjadi prediksi yang akurat adalah tujuan utama penerapan metode ini yang diharapkan dapat membantu perusahaan dalam perencanaan stok barang secara lebih efisien. Tingkat akurasi prediksi yang sangat tinggi, yaitu 94 persen, dicapai oleh para peneliti dengan menggunakan KNN. Bahwa model prediksi yang dibangun memiliki kemampuan yang baik dalam memperkirakan tingkat penjualan di masa depan ditunjukkan oleh skor akurasi ini sehingga dapat dijadikan sebagai dasar untuk pengambilan keputusan yang lebih akurat dalam manajemen stok.
7. Penelitian yang dilakukan oleh Qurotul A’yuniyah dan Muhammad Reza pada tahun 2023 berjudul "Penerapan Algoritma *K-Nearest Neighbor* untuk Klasifikasi Jurusan Siswa di SMA Negeri 15 Pekanbaru" bertujuan untuk mengatasi permasalahan yang dihadapi

SMA Negeri 15 Pekanbaru dalam proses penjurusan siswa yang memakan banyak waktu dan tenaga ketika dilakukan secara manual. Algoritma *K-Nearest Neighbor* (KNN) dan alat *RapidMiner* digunakan oleh penelitian ini untuk mengklasifikasikan data siswa secara efisien. Hasil penelitian menunjukkan bahwa dengan parameter $k=3$, akurasi maksimal yang dicapai adalah 93.52%, rata-rata *precision* sebesar 88.14%, dan rata-rata *recall* mencapai 100.00%. Angka-angka ini menunjukkan bahwa algoritma KNN mampu memberikan hasil klasifikasi yang sangat akurat dan andal untuk menentukan jurusan siswa dengan cepat dan tepat. Penggunaan KNN dalam klasifikasi jurusan siswa di SMA Negeri 15 Pekanbaru terbukti oleh kesimpulannya efektif dan efisien, menawarkan solusi praktis bagi sekolah untuk meningkatkan proses penjurusan siswa. Kontribusi penting dalam bidang *Data Mining* dan pendidikan diberikan oleh penelitian ini dengan menunjukkan bahwa teknik klasifikasi otomatis dapat meningkatkan akurasi dan efisiensi proses penjurusan, serta mendukung pengambilan keputusan yang lebih baik di lingkungan sekolah.

8. Penelitian yang dilakukan oleh Ma'ruf Aziz Muzani, M. Iqbal Abdullah Sukri, Syifa Nur Fauziah, Agus Fatkhurohman, dan Dhani Ariatmanto pada tahun 2021 dengan judul "*Data Mining* untuk Klasifikasi Produk Menggunakan Algoritma *K-Nearest Neighbor* pada Toko Online" bertujuan untuk mengatasi tantangan dalam pencarian produk di e-marketplace yang semakin berkembang dan beragam, khususnya di platform Shopee. Investigasi ini menggunakan algoritma *K-Nearest Neighbour* (KNN) untuk mengkategorikan produk dalam tiga kategori: peralatan rumah tangga, pakaian muslim, dan ponsel dan aksesoris. Data diperoleh melalui penggalian web. *Preprocessing* dilakukan terlebih dahulu dalam proses klasifikasi, yang kemudian diikuti dengan pembobotan kata dengan pendekatan *TF-IDF* dan mencari tahu seberapa mirip tulisan satu dengan yang lain menggunakan *cosine similarity*. Nilai k yang ideal ditemukan dengan membuat peringkat

kemiripan kosinus. Setelah melakukan pengujian terhadap sembilan data dengan tiga nilai k yang berbeda, kami menemukan bahwa $k=3$ memberikan hasil terburuk dalam hal akurasi, presisi, dan *recall*. Hasil yang lebih baik, dengan akurasi 88,89%, presisi 83,33%, dan *recall* 100%, diperoleh dengan nilai $k=5$ dan $k=7$. Temuan ini menunjukkan bahwa produk di *e-marketplace* dapat diklasifikasikan secara efektif menggunakan algoritma KNN, memudahkan pengguna dalam mencari produk yang diinginkan dengan lebih cepat dan akurat. Kontribusi penting dalam penerapan *data mining* diberikan oleh penelitian ini untuk meningkatkan pengalaman berbelanja online dengan menyediakan metode yang efisien untuk pengelompokan produk berdasarkan kemiripan atribut.

