

BAB 2 TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian sebelumnya dilakukan dengan tujuan untuk mengumpulkan referensi dan pembandingan yang dapat berguna dalam penelitian yang sedang dilaksanakan. Selain itu, penelitian sebelumnya juga bertujuan untuk menghindari kesan bahwa penelitian yang sedang dilakukan adalah duplikasi, yang pada akhirnya membantu peneliti untuk menyajikan penelitian yang bersifat orisinal. Daftar penelitian terkait topik yang sedang dibahas oleh peneliti dapat di temukan dalam tabel 2.1:

Berdasarkan tinjauan jurnal-jurnal yang telah dianalisis oleh penulis, ada beberapa perbedaan dengan penelitian sebelumnya. dalam penelitian ini, peneliti berfokus pada melakukan komparasi dua algoritma klasifikasi dengan tujuan untuk melihat algoritma mana yang memiliki nilai akurasi terbaik. Algoritma klasifikasi yang digunakan dalam penelitian ini merupakan algoritma yang sering digunakan dalam proses klasifikasi secara umum.

2.2 Data Mining

Penambangan data atau yang dikenal sebagai Data Mining merupakan proses analisis data dari berbagai perspektif dengan tujuan untuk mengidentifikasi informasi yang berharga. Informasi ini dapat digunakan untuk meningkatkan keuntungan, mengurangi biaya transaksi, atau bahkan mencapai keduanya. Secara teknis, penambangan data dapat dijelaskan sebagai metode untuk menemukan

korelasi atau pola dalam data yang terdapat dalam basis data dunia nyata yang besar. Data mining merupakan proses mengekstrak sejumlah data yang sebelumnya tidak diketahui atau bisa didefinisikan sebagai bagian dari proses penggalian pengetahuan dari database[3].

Kemampuan data mining untuk mengungkap informasi dari basis data yang luas sering diibaratkan seperti menambang emas di wilayah yang kaya. Teknologi ini memiliki berbagai aplikasi termasuk:

- a) Prediksi tren dan karakteristik bisnis, dengan data mining yang otomatis mencari informasi prediktif dalam basis data besar
- b) Mengungkap pola-pola yang belum pernah di ketahui sebelumnya, melalui data mining yang secara menyeluruh mengeksplorasi basis data dan mengidentifikasi pola-pola yang sebelumnya tersembunyi.

Secara teknis, data mining dapat di jelaskan sebagai pendekatan yang digunakan untuk mengidentifikasi sebagian kecil dari data yang sangat besar. Proses pengumpulan dan analisis data seperti itu dapat diimplementasikan melalui perangkat lunak komputasi statistik, komputasi matematis, atau bahkan kecerdasan buatan (AI). Dalam data mining, ada lima peran utama, yaitu estimasi, prediksi, pengidentifikasian asosiasi, klasifikasi, dan pengelompokan data.

Data mining adalah rangkaian Langkah untuk mengeksplorasi nilai tambah berupa informasi yang sebelumnya tidak diketahui secara manual dari sekumpulan data[4].

Data Mining memiliki sejumlah keunggulan sebagai alat analisis diantaranya:

- a) Mampu mengelola data dalam jumlah besar dan dengan tingkat kompleksitas yang tinggi
- b) Dapat menangani beragam jenis atribut data
- c) Kemampuan dalam mencari dan memproses data secara otomatis, meskipun dalam beberapa teknik data mining, beberapa parameter perlu diatur secara manual oleh pengguna.
- d) Mampu memanfaatkan pengalaman serta kesalahan sebelumnya untuk meningkatkan kualitas dan hasil analisis dan menghasilkan hasil yang optimal

Sedangkan data mining memiliki kekurangan dalam pencarian data, yaitu tidak melakukan pencarian secara individu, melainkan mengklasifikasikan data berdasarkan kriteria tertentu.

Data mining adalah satu tahap dalam Knowledge Discovery in Database (KDD). [5]. Selama proses penambangan data ini, kita dapat melakukan klasifikasi, prediksi, dan pengkategorian untuk mengungkap informasi yang bermanfaat dalam kumpulan data yang sangat besar. [6]. Penggunaan istilah “Data Mining” dan “KDD” seringkali digunakan secara bergantian untuk menjelaskan proses yang lebih lanjut dalam menggali informasi yang tersembunyi dalam basis data besar[7].

Tabel 2.1 Literatur Review / Penelitian Terkait

Peneliti	Judul Penelitian	Tahun	Hasil Penelitian	Kelebihan	Kelemahan
Aisha alfani, Fahrur Rozi, Farid Sukmana	Prediksi penjualan produk unilever menggunakan metode k-nearest neighbor. [8]	2021	Haisil dari penelitian ini didasarkan pada tingkat akurasi tertinggi dalam mengklasifikasikan penjualan produk, yaitu sekitar 86,66%, sementara tingkat akurasi terendah dalam mengklasifikasikan penjualan produk adalah sekitar 40%	Hasil pengujian telah menunjukkan bahwa akurasi tertinggi tercapai pada nilai k sebesar 86,66%. Ini terjadi dalam uji coba yang dilakukan pada baris data ke-80 dengan nilai k=30	Tidak memberikan dorongan untuk melakukan penelitian lebih lanjut
Lusi suryadi, Ngajiyanto, Novia eka P, Ferly ardhy, Pakartika Riswanto	Penerapan data mining prediksi penjualan mebel terlaris menggunakan metode K-nearest neighbor. [9]	2022	Penelitian ini mengembangkan sebuah algoritma K-Nearest Neighbor dalam konteks Teknik data mining yang di gunakan untuk memprediksi penjualan produk populer di Zerita Meubel	Hasil akurasi yang di dapatkan dari proses klasifikasi penjualan produk terlaris yakni sebesar 87,51%	Dapat dipertimbangkan untuk mengimplementasikan data mining dengan mengkombinasikan dua Teknik algoritma data mining lainnya. Hal ini diharapkan dapat menghasilkan hasil yang lebih beragam dan tentunya dapat meningkatkan nilai akurasi yang di peroleh.

Peneliti	Judul Penelitian	Tahun	Hasil Penelitian	Kelebihan	Kelemahan
Muhammad taufik hidayat, nana suarna, nining rahaningsih	Implementasi algoritma naïve bayes untuk prediksi persediaan barang PT. Dilmoni citra mebel Indonesia. [10]	2023	Dari penerapan model algoritma Naive Bayes, dicapai tingkat akurasi sebesar 95,60%, class precision sebesar 96,56% dan class recall sebesar 97,02%	Menghasilkan nilai akurasi yang cukup besar	Tidak memberikan penjelasan tentang ilustrasi gambar pada jurnal.
Norma ayuningtyas, nining R, Fadhil M Basyyar	Penerapan data mining pada penjualan produk MS Glow menggunakan metode naïve bayes untuk strategi pemasaran. [11]	2022	Dengan pengumpulan data penjualan tahun 2021 yang mencakup 240 entri data berdasarkan atribut seperti Kode Buku, Tanggal Buku, Bulan, Nama Produk, Harga, Jumlah Terjual, Stok Awal, Stok Masuk, Stok Akhir, dan Restok. Hasil perhitungan prediksi menggunakan algoritma Naïve Bayes menghasilkan tingkat akurasi prediksi sekitar 92,50%. Ketepatan klasifikasi untuk kategori "YA" adalah sekitar 95,71% sementara untuk	Penggunaan Teknik Naïve Bayes dalam menganalisis data penjualan produk MS Glow menghasilkan tingkat keberhasilan yang sangat tinggi, yakni mencapai 98,50%	Diperlukan sebuah algoritma pembandingan untuk mengevaluasi dan membandingkan tingkat akurasi diantara berbagai algoritma

Peneliti	Judul Penelitian	Tahun	Hasil Penelitian	Kelebihan	Kelemahan
			kategori “TIDAK” adalah sekitar 88,00%. Sedangkan tingkat keberhasilan dalam mengidentifikasi kategori “YA” adalah sekitar 91,78% dan untuk kategori “TIDAK” adalah sekitar 93,62%		
Reyhan dzickrillah laksamana, edy santoso, bayu rahayudi	Prediksi penjualan roti menggunakan metode exponential smoothing (studi kasus : harum bakery) [12]	2019	Presentase Kesalahan Absolut Terkecil (MAPE) dalam metode Single Exponential Smoothing mencaoi nilai minimum Ketika $\alpha = 0,1$ dengan MAPE sebesar 27,4039%, Sementara itu, dalam metode Double Exponential Smoothing, nilai MAPE terkecil juga didapatkan pada saat nilai $\alpha = 0,1$ dengan hasil sekitar 25,124%, dan yang terakhir, dalam metode Triple Exponential Smoothing,	Metode penelitian mudah dipahami	Tidak terdapat penjelasan mengenai grafik hasil prediksi

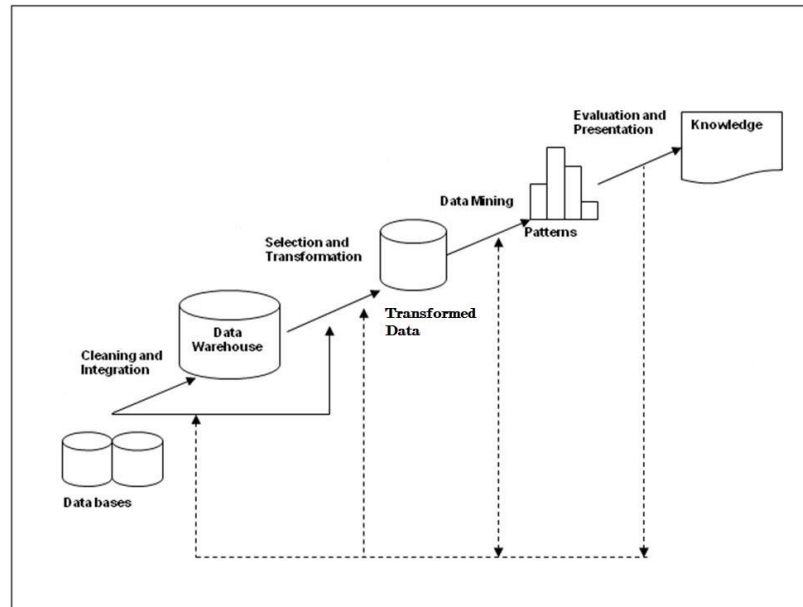
Peneliti	Judul Penelitian	Tahun	Hasil Penelitian	Kelebihan	Kelemahan
			MAPE terendah tercapai Ketika nilai $\alpha = 0,1$, $\beta = 0,1$, dan $\gamma = 0,4$ dengan nilai MAPE sekitar 25,303%.		
Rustam, sidik rahmatullah, supriyanto, sri wahyuni	Penerapan data mining untuk prediksi penjualan produk triplek pada PT puncak menara hijau mas. [13]	2020	Metode Clustering menggunakan algoritma K-Means dapat di terapkan untuk mengelompokan data penjualan triplek berdasarkan volume penjualan, yaitu kelompok 'Laris', 'Kurang Laris', dan 'paling laris'. Dengan demikian, pihak pabrik dapat mengambil Langkah-langkah untuk meningkatkan popularitas produk triplek yang kurang laris.	Penjelasan dalam jurnal dituliskan secara rinci	Ringkasan (Abstrak) tidak memberikan gambaran lengkap seluruh penelitian

Peneliti	Judul Penelitian	Tahun	Hasil Penelitian	Kelebihan	Kelemahan
Sahara abdy, Erika Roberta Br gultom, sri ramadhany, Afifuddin	Prediksi penjualan sparepart mobil terlaris menggunakan metode K-Nearest neighbor. [14]	2022	Penelitian mencapai prediksi terbaik dalam penjualan di divisi Mitshubishi dengan akurasi sekitar 80,00% dan dengan tingkat akurasi sekitar 66,67% untuk divisi Isuzu. Ini memungkinkan untuk menetapkan target yang dapat meningkatkan akurasi dalam mengklasifikasikan produk terlaris	Hasil studi mencapai tingkat akurasi yang tinggi yakni mencapai 80%	Tidak memberikan dorongan untuk melakukan penelitian lebih lanjut
Rismala, Irfan ali, ade rizki rinaldi	Penerapan metode k-nearest neighbor untuk prediksi penjualan sepeda motor terlaris. [15]	2023	Tingkat akurasi dalam mengolah data penjualan sepeda motor PT. Sumber Rejeki Jabar pada periode Januari hingga Desember 2022 dengan menggunakan parameter K senilai 5 mencapai 96,15%	Nilai akurasi yang di dapatkan cukup tinggi yakni dengan nilai K=5 mencapai 96,15%	Tidak menjelaskan symbol yang terdapat pada rumus
Rara iriane, nurfaizah	Penerapan data mining untuk	2023	Dalam penelitian ini, ditemukan tingkat akurasi	Nilai akurasi yang di dapatkan dalam	Tidak ada penjelasan mengenai simbol-simbol dalam rumus,

Peneliti	Judul Penelitian	Tahun	Hasil Penelitian	Kelebihan	Kelemahan
	prediksi penjualan produk pangan hewan menggunakan metode k-nearest neighbor. [16]		sekitar 80,4% dalam mengklasifikasikan kategori produk dengan menggunakan Euclidean Distance. Prediksi menunjukkan bahwa produk terlaris yang paling banyak terjual adalah Whiskas 80 gram Junior Tuna, dengan prediksi penjualan mencapai 6963 unit. Sementara produk yang diprediksi memiliki penjualan paling sedikit adalah Whiskas Adult 1,2kg dengan prediksi sebanyak 8 Unit	penggunaan algoritma KNN untuk prediksi produk pangan hewan cukup tinggi yakni sekitar 80,4%	dan masih memungkinkan untuk melakukan penelitian lebih lanjut dengan menerapkan algoritma lainnya.
Muhammad rizal lubis	Analisa prediksi penjualan produk dengan menggunakan metode C45 (studi kasus PT kawan	2019	Dalam penerapan algoritma C4.5 membaca seluruh sampel data training dari penyimpanan dan mengangkatnya kedalam memori, memungkinkan algoritma ini untuk	Penjabaran rumus di lakukan secara rinci	di dalam abstrak tidak mendeskripsikan penelitian secara keseluruhan.

Peneliti	Judul Penelitian	Tahun	Hasil Penelitian	Kelebihan	Kelemahan
	lama ace hardware. [17]		mengonstruksi pohon keputusan dengan mengakses data training yang ada dalam memori.		
Yahya, Winda Puspita Hidayanti	Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Efektivitas Penjualan Vape(Rokok Elektrik) pada "Lombok Vape On". [18]	2020	Hasil dari penelitian ini adalah berdasarkan implementasi KNN dari pengolahan dengan K-Fold Validation 6 aalah percobaan yang menghasilkan nilai akurasi tertinggi yakni diperoleh sekitar 86,48% dan AUC sebesar 0,874.	Nilai akurasi yang di dapatkan cukup tinggi yakni sekitar 86,48%	Tidak di lakukan perhitungan manual dari sampel data yang bertujuan untuk menguji apakah model yang di bangun dengan menggunakan rapid miner sudah sesuai.
Odi Nurdiawan, Noval Salim	Penerapan Data Mining pada Penjualan Barang Menggunakan Metode Metode Naive Bayes Classifier untuk Optimasi Strategi Pemasaran. [19]	2018	Penerapan metode Naive Bayes yang ada di dalam penelitian ini adalah di gunakan pada data set penjualan barang dan mencapai tingkat akurasi sebesar 97,22%	Nilai akurasi yang di dapatkan cukup tinggi dan adanya perhitungan manual sehingga dapat memudahkan pembaca dalam memahami metode naive bayes	Dalam abstrak tidak mendeskripsikan penelitian secara keseluruhan.

Gambar 2.1 di bawah ini mengilustrasikan proses ekstraksi data menjadi pengetahuan:



Gambar 2.1 Proses KDD

2.2.1 Data Selection

Sebelum memulai proses penambangan data, langkah awal adalah menyelesaikan proses pemilihan data dari seluruh Kumpulan data operasional. Pada tahap ini, keputusan mengenai pemilihan dataset atau fokus pada subset variabel tertentu sangat penting, karena proses ini memerlukan waktu. Hasil seleksi data yang akan digunakan dalam proses penambangan data akan disimpan dalam dokumen terpisah yang berbeda dari basis data operasional.

2.2.2 Preprocessing atau cleaning

Setelah menyelesaikan tahap pemilihan data, Langkah berikutnya dalam data mining adalah *preprocessing* data. Tahap ini melibatkan proses pembersihan data untuk mengidentifikasi dan menghapus data ganda, data dengan kesalahan sintaksis

dan memaksimalkan hasil analisis. *Preprocessing* data merupakan tahap yang sangat penting dalam proses data mining, karena kualitas hasil data mining sangat dipengaruhi oleh proses *preprocessing* data. Preprocessing data merupakan tahapan sebelum proses pengklasifikasian yang di perlukan untuk membersihkan, menghilangkan, mengubah bentuk data yang tujuannya agar model dapat terbentuk dari data yang baik [20]

2.2.3 Transformasi

Transformasi data adalah Langkah yang mengacu pada mengubah bentuk dari satu bentuk ke dalam bentuk lainnya. Tujuan utama dari transformasi data adalah untuk meningkatkan kualitas data dan mengoptimalkan kinerja algoritma pemrosesan data. Beberapa contoh umum dari transformasi data meliputi:

- a) Normalisasi: Proses mengubah data kedalam skala yang seragam, sehingga memungkinkan perbandingan dan analisis yang lebih mudah.
- b) Penghapusan duplikat: proses menghilangkan data yang sama dalam satu data set
- c) Imputasi: proses mengisi nilai yang hilang dalam data dengan nilai yang sesuai
- d) Pemfilteran: proses pemilihan subset data berdasarkan kriteria tertentu.
- e) Encoding: proses mengubah format data dari suatu bentuk ke bentuk lain, seperti mengonversi data nominal menjadi data numerik
- f) Reduksi dimensi: proses mengurangi jumlah fitur atau variabel dalam data

Transformasi data memiliki peran sangat penting dalam pemrosesan data, karena dapat membantu meningkatkan kualitas data dan mengoptimalkan kecepatan pemrosesan data.

2.2.4 Proses Penambangan Data

Proses penambangan data adalah tahap dimana pola atau informasi yang sebelumnya tidak diketahui dapat diidentifikasi dengan metode khusus. Berbagai Teknik, metode, dan algoritma yang digunakan dalam proses ini sangat bervariasi. Pemilihan metode dan algoritma bergantung pada tujuan dan konteks keseluruhan penambangan data. Proses ini seringkali memanfaatkan metode matematika, statistika dan juga kecerdasan buatan.

2.2.5 Interpretation atau Evaluation

Hasil dari proses penambangan data harus disajikan dalam format yang mudah dimengerti oleh berbagai pemangku kepentingan yang tertarik dengan hasil penambangan data tersebut. Pada tahap ini, dilakukan pemeriksaan pola informasi berdasarkan fakta-fakta yang telah ada sebelumnya.

2.3 Klasifikasi

Klasifikasi adalah Salah satu fitur dalam data mining yang melibatkan pembuatan model untuk meramalkan atau mengelompokkan objek-objek dalam basis data ke dalam kelas atau kategori tertentu [21]. Klasifikasi adalah Tindakan

mengelompokan informasi berdasarkan atribut-atribut yang diberikan label, sehingga algoritma-algoritma yang digunakan untuk menangani permasalahan klasifikasi termasuk dalam kategori *administered learning* atau pembelajaran yang diawasi. Hasil dari proses ini adalah mencapai tingkat akurasi atau presisi yang spesifik.

2.4 Teori Dasar

Dasar teori dalam penelitian ini berfokus pada penerapan algoritma Naive Bayes dan K-Nearest Neighbor. Algoritma-algoritma ini dipilih karena kemampuannya untuk mengklasifikasikan data penjualan leopard gecko yang belum di ketahui hanya dengan menggunakan data yang ada. Data set ini akan digunakan sebagai data latih dan data uji dalam rangka mengklasifikasi penjualan berdasarkan kriteria dan platform penjualan.

Penjelasan lebih lanjut mengenai algoritma klasifikasi Naive Bayes dan K-Nearest Neighbor dapat di deskripsikan di bawah ini:

a) Naive Bayes

Naive Bayes adalah sebuah metode klasifikasi yang menggunakan prinsip probabilitas sederhana. Algoritma ini menghitung probabilitas berdasarkan kombinasi atribut dan frekuensi nilai-nilai dalam data. Dengan memanfaatkan teorema Bayes, algoritotma ini memperkirakan probabilitas berbagai atribut yang independen dan dapat memprediksi kelas variabel berdasarkan informasi ini[22].

Naive Bayes adalah salah satu metode pembelajaran mesin yang efektif [23]. Algoritma ini mengandalkan probabilitas dan statistik dan dikembangkan berdasarkan teorema Bayes yang di usulkan oleh ilmuwan Inggris, Thomas Bayes. Prinsip dasar dari Naive Bayes adalah memprediksi peristiwa dimasa yang akan datang berdasarkan pengalaman di masa sebelumnya. Metode Naive Bayes sangat cocok untuk mengklasifikasikan data dalam format biner atau multi-kelas. Metode ini menerapkan Teknik klasifikasi objek di masa yang akan datang dengan menggunakan probabilitas bersyarat berdasarkan pengalaman masa lalu. Probabilitas ini adalah perkiraan kemungkinan kejadian tertentu berdasarkan pengalaman serupa yang telah terjadi (karena asumsi, praduga, pernyataan, atau sebab lainnya.) rumus dasar teorema ini adalah sebagai berikut:

$$P(A | B) = P(B | A)P(A)P(B) \dots \dots \dots (1)$$

Keterangan:

$P(A|B)$: Probabilitas A dengan bukti bahwa B telah terjadi (probabilitas)

$P(B|A)$: Probabilitas B terjadi dengan bukti bahwa A telah terjadi

$P(A)$: peluang terjadinya A

$P(B)$: Peluang terjadinya B

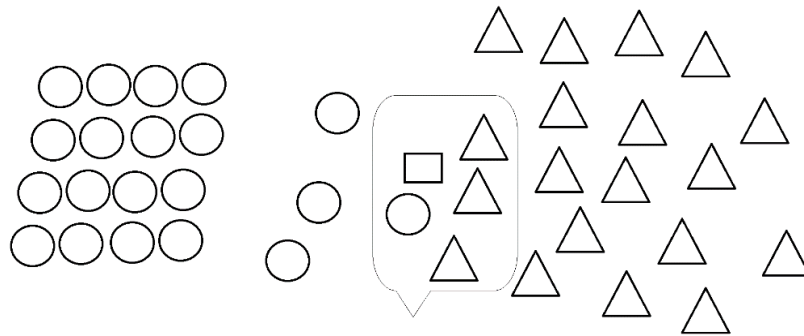
Kemudian timbul pertanyaan tentang bagaimana menerapkan teorema Bayes untuk melakukan klasifikasi data? Atau lebih spesifik, bagaimana mengestimasi nilai-nilai ini? Ini dilakukan dengan menggunakan metode Naive Bayesian Classifier, yang bekerja dengan mencari maksimum nilai $P(A|B)$

b) K-NN

K-Nearest Neighbor atau KNN adalah suatu algoritma yang memanfaatkan data latih (training record) yang diperoleh dari tetangga terdekat (nearest neighbor) dalam melakukan klasifikasi data, dengan jumlah tetangga terdekat yang di tentukan oleh nilai K [24]. Algoritma KNN adalah algoritma klasifikasi atau regresi non parametrik dalam bidang pengenalan pola[25]. Algoritma KNN mengikuti serangkaian tahap kerja yang meliputi:

- 1) Langkah pertama adalah menentukan jumlah tetangga yang ingin digunakan
- 2) Selanjutnya menghitung jarak antara data baru dengan seluruh tetangga yang ada, menggunakan rumus perhitungan seperti Euclidean Distance atau metode lainnya
- 3) Setelah itu, memilih k tetangga terdekat dengan jarak terkecil untuk digunakan dalam proses pengambilan keputusan prediksi berdasarkan hasil perhitungan jarak yang telah dilakukan sebelumnya.

K-Nearest Neighbor (KNN) adalah suatu algoritma pembelajaran mesin yang berfokus pada pembelajaran non-parametrik dan konsep Lazy Learning. Berikut ini adalah gambar yang di temukan dalam buku “Data Mining untuk Klasifikasi Data” yang ditulis oleh Dr. Suyanto mengenai konsep dasar KNN:



Gambar 2.2 Konsep dasar KNN

Tujuan algoritma ini adalah untuk melakukan klasifikasi pada objek-objek baru berdasarkan atribut-atribut dan data latih. Pada tahun 1968, Cover dan Hart memperkenalkan algoritma K-Nearest Neighbor (KNN) sebagai algoritma yang dikenal sebagai algoritma “malas” karena pendekatannya melibatkan penyimpanan semua data pelatihan dan pembentukan model yang ditunda hingga data uji diberikan kepada algoritma ini untuk diprediksi. [26]. Rumus yang mendasari proses KNN adalah sebagai berikut:

$$distance = \sqrt{\sum_{i=1}^n (x_{training}^i - x_{testing})^2} \dots \dots \dots (2)$$

Keterangan:

$x_{training}^i$: data training ke-i

$x_{testing}$: data testing

i : record data ke- i dari tabel

n : jumlah data training

secara sederhana, proses K-NN dapat diibaratkan sebagai sistem pemungutan suara dalam klasifikasi data. Nilai K dalam K-NN dapat diperoleh dengan mengukur jarak, kemiripan atau ketidakmiripan antara data. Ketika melakukan klasifikasi, K-NN akan memeriksa semua pola yang ada dalam data latih untuk menemukan pola klasifikasi yang sesuai. Hal yang menarik, algoritma K-NN beroperasi secara lokal, hanya mempertimbangkan nilai K terdekat, sehingga sesuai dengan kelompok data yang terdekat.

2.5 Leopard Gecko

Leopard gecko (*Eublepharis Macularius*) adalah varian dari tokek yang di temukan di wilayah Pakistan, India, dan Iran. Leopard gecko adalah tokek hias yang berasal dari wilayah Pakistan dan banyak di gemari karena memiliki warna yang menarik dengan karakter yang sangat jinak dan mudah untuk di pelihara [27]. Salah satu ciri khas dari Leopard Gecko adalah ekornya yang menyerupai daun dengan kulit kasar yang memiliki pola bitnik-bintik mirip dengan tekstur bola basket pada warna dasar yang berkisar antara kuning pucat hingga kuning lemon. Gambar 2.3 adalah contoh gambar dari leopard gecko dengan sub species *eublepharis macularius*:



Gambar 2.3 Leopard gecko

2.6 Rapid Miner

Dalam penelitian ini, penulis memutuskan untuk menggunakan aplikasi Rapid Miner. Alasan pemilihan Rapid Miner adalah karena kemampuannya untuk menggabungkan visualisasi, statistik, dan informasi penting tentang model kedalam laporan dengan cepat dan mudah. Rapid Miner juga menyediakan pelaporan yang informatif, yang memungkinkan pengguna untuk melihat Riwayat alur kerja untuk setiap elemen dan visualisasi langsung dari laporan tersebut. Kelebihan lainnya adalah antarmuka pengguna yang intuitif, yang memungkinkan pengguna untuk fokus pada analisis data daripada harus terjebak dalam pengkodean yang rumit. Rapid miner juga dapat membuat pembuatan *pipeline* analisis data yang kompleks menjadi lebih sederhana. RapidMiner adalah sebuah aplikasi yang populer

digunakan untuk melakukan pemrosesan data mining. Aplikasi ini terkenal karena menyediakan beragam algoritma untuk tugas seperti klasifikasi, pengelompokan, dan analisis regresi dalam proses penambangan data. RapidMiner dirancang dengan antarmuka pengguna yang ramah, terdiri dari tiga perspektif utama yakni *welcome*, *design* dan *result* [28]. Pengguna dapat memulai dengan mengimpor data kedalam aplikasi, kemudian data tersebut dapat dihubungkan dengan operator, termasuk operator untuk mengatasi *missing values*. Data yang lengkap akan dimodelkan dengan menggunakan tampilan *design* yang terhubung dengan operator untuk menjalankan proses penambangan data.