

## BAB II TINJAUAN PUSTAKA

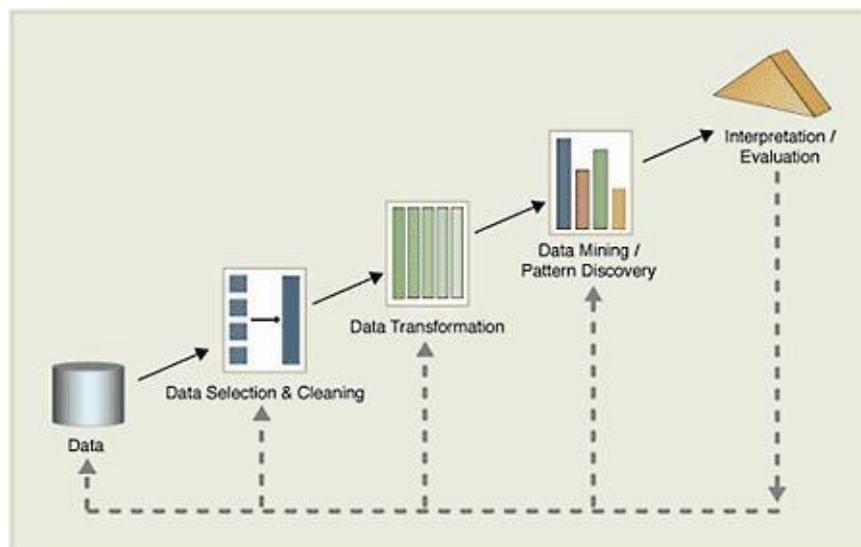
### 2.1 Tinjauan Pustaka Terkait Teknik atau Metode

#### 2.1.1 Data Mining

Analisis *observasi* data yang sangat besar untuk menemukan hubungan yang belum diketahui sebelumnya dan dua metode baru untuk meringkas data sehingga lebih mudah dipahami untuk pemilih data (Jefri, 2013).

*Detesis, estimasi, prediksi, klasifikasi, clustering*, dan *asosiasi* adalah fungsi yang dapat digunakan bisnis untuk mempertahankan dan meningkatkan fungsi dan menggunakan pengetahuan yang diperoleh (Nurjoko dan Kurniawan, 2016). Mengekstraksi nilai tambahan dari kumpulan data dengan cara manual adalah proses yang dikenal sebagai data

*Data mining*, juga dikenal sebagai Penemuan Pengetahuan dalam *Database* sejarah untuk menemukan pola, keteraturan, atau hubungan dalam banyak (Santoso, 2007)



Gambar 2.1 Tahapan Data Mining

#### 2.1.2. Tahapan Penggalan Informasi

*Data mining* adalah proses menemukan informasi yang menarik seperti pola, aturan, hubungan, aturan data lainnya.

**a. Data Selection (Seleksi Data)**

Sebelum tahap dalam atau dipilih dari sekumpulan data operasional. Data yang dipilih untuk proses data mining disimpan dalam suatu berkas yang berbeda. *Fase* ini bertujuan untuk mengekstrak gudang yang besar.

**b. Pre-Processing / Cleaning (Pemilihan Data)**

Sebelum proses data mining dapat dimulai, data KDD harus dibersihkan. Menghapus duplikat adalah beberapa contoh proses perbaikan data. Proses *enhriment* juga dilakukan, yang berarti "memperkaya", seperti data atau data dari sumber luar. akurat adalah bagian dari proses ini.

**c. Transformation (Transformasi Data)**

*Coding* dalam KDD adalah bergantung pada jenis atau pola informasi yang akan dicari dalam basis data. Ini dapat digunakan digunakan proses data.

**d. Data Mining**

Proses menemukan informasi yang m dalam kumpulan data tertentu dengan menggunakan metode atau teknik tertentu. Data mining memiliki banyak teknik, metode, dan algoritma yang berbeda. tepat. dibangun di atas *algoritma* yang digunakan.

**e. Interpretasi/Pemeriksaan (Interpretasi dan Pemeriksaan)**

Ditunjukkan sehingga dapat memahaminya. Tahap ini merupakan bagian dari proses *interpretasi* KDD. Tahap ini mencakup menilai *hipotesis* atau fakta sebelumnya. Tahap ini berakhir dengan pemilihan model bisnis.

**2.1.3. Regresi Linear Berganda**

*Regresi linear* berganda adalah metode *statistik* yang digunakan untuk menganalisis hubungan antara satu *variabel dependen* (terikat) dengan dua atau lebih *variabel independen* (bebas). Metode ini merupakan perluasan dari *regresi linear* sederhana, yang hanya melibatkan satu *variabel independen*. Tujuan utama dari *regresi linear* berganda adalah untuk memahami dan memodelkan

pengaruh beberapa *variabel independen* terhadap *variabel dependen* secara *simultan*.

#### 2.1.4 Komponen Utama dalam *Regresi Linear Berganda*:

1. ***Variabel Dependen (Y)***: *Variabel* yang ingin kita prediksi atau jelaskan. Contohnya, dalam penelitian tentang pengaruh pupuk terhadap pertumbuhan tanaman, tinggi tanaman bisa menjadi *variabel dependen*.
2. ***Variabel Independen (X1, X2, X3, ...)***: *Variabel-variabel* yang digunakan untuk memprediksi *variabel dependen*. Dalam contoh yang sama, jenis pupuk (kimia, organik, biopupuk) bisa menjadi *variabel independen*.
3. ***Koefisien Regresi ( $\beta_0, \beta_1, \beta_2, \dots$ )***: Nilai yang menggambarkan seberapa besar pengaruh setiap variabel independen terhadap *variabel dependen*. *Koefisien* ini menunjukkan perubahan rata-rata dalam *variabel dependen* untuk setiap satu unit perubahan dalam *variabel independen*, dengan asumsi *variabel* lain tetap *konstan*.
4. ***Intercept ( $\beta_0$ )***: Nilai awal dari *variabel dependen* ketika semua *variabel independen* bernilai nol. Ini merupakan titik di mana garis *regresi* memotong sumbu Y.
5. ***Error Term ( $\epsilon$ )***: Komponen yang menangkap *variabilitas variabel dependen* yang tidak dijelaskan oleh *variabel independen*. Ini mencakup faktor-faktor lain yang tidak dimasukkan dalam model tetapi mempengaruhi *variabel dependen*.

#### 2.1.5 Persamaan *Regresi Linear Berganda*

*Regresi linier* berganda adalah perkembangan dari *regresi* sederhana yang menjelaskan hubungan antara *variabel* terikat dengan lebih dari satu *variabel* bebas (Freund, Wilson, & Sa, 2006 : 73). Tujuan utama analisis *regresi linier* berganda sama seperti dengan *regresi* sederhana, yaitu menggunakan hubungan antara *variabel* terikat (*response*) dan *variabel* bebas (*factor*) untuk memprediksi atau

menjelaskan karakteristik dari *variabel* (Freund, Wilson, & Sa, 2006 : 73). Secara umum model *regresi linier* berganda dapat dituliskan sebagai berikut:

Bentuk persamaan *regresi linear* berganda adalah:

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \dots\dots\dots (1)$$

Keterangan:

$Y$  =  $Y$  variabel terikat

$X_1$  = variabel bebas

$\beta_0$  = Konstanta regresi

$\beta_1$  = koefisien regresi

$i = 1, 2, 3, \dots, n$  (jumlah pengamatan)

Untuk mencari nilai *konstanta*  $a$  dan *koefisien regresi*  $b$  dapat dihitung dengan metode *matriks*, persamaan metode normal (*alternatif*) dan *kuadrat* terkecil.

$$Y = a + b_1 X_1 + b_2 X_2 \dots\dots\dots (2)$$

$Y$  = nilai

$A$  = *intercept*,

$b_1, b_2$  = koefisien  $X_1$  dan  $X_2$

Maka  $b_1$  dan  $b_2$  dicari dengan menggunakan *substitusi* untuk mencari koefisien  $b_1, b_2, b_3$  dengan perkalian persamaan  $a$  dan persamaan  $b$ , dan juga persamaan selanjutnya.

$$\sum X = an + b_1 \sum X_1 + b_2 \sum X_2$$

$$\sum X_1 Y = a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2$$

$$\sum X_2 Y = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2$$

### 2.1.6 Proses Analisis dalam Regresi Linear Berganda:

1. **Pengumpulan Data:** Data yang mencakup *variabel dependen* dan semua *variabel independen* dikumpulkan.
2. **Estimasi Koefisien:** Koefisien regresi dihitung menggunakan metode *kuadrat terkecil (least squares method)* yang meminimalkan jumlah *kuadrat* dari selisih antara nilai *observasi* dan nilai prediksi.

3. **Uji Signifikansi:** Uji statistik, seperti uji t dan uji F, digunakan untuk menentukan apakah *variabel independen* memiliki pengaruh signifikan terhadap *variabel dependen*.
4. **Diagnostik Model:** Asumsi-asumsi klasik *regresi* (normalitas, *homoskedastisitas*, *multikolinearitas*, dan *independensi*) diperiksa untuk memastikan *validitas* model.
5. **Interpretasi Hasil:** *Koefisien regresi* diinterpretasikan untuk memahami pengaruh masing-masing *variabel independen* terhadap *variabel dependen*. Model juga dievaluasi menggunakan indikator seperti *R-squared* untuk mengukur seberapa baik model menjelaskan *variabilitas variabel dependen*.

## 2.2 Pupuk

Pupuk adalah dan ditambahkan ke media tanam mendukung pertumbuhannya dan memaksimalkan hasilnya.

Kehidupan tumbuhan secara alami mendapatkan banyak hara dari Bumi dan *atmosfer* di atasnya. Meskipun demikian, sebanding dengan kebutuhan tanaman, sehingga *intervensi* manajemen *biosfer* diperlukan, dicapai, termasuk

Untuk metabolisme yang sempurna, tanaman membutuhkan tiga belas unsur hara *makro* unsur hara *mikro* (*Cl, Fe, Mn, Zn, Cu, B, Mo*), dan unsur hara *non-esensial* atau *fungsional* (*Na, Co, V, Si, Ni*), sementara itu lima bahan hara yang tidak penting, Dalam beberapa situasi, dapat bertindak sebagai pengganti atau memenuhi fungsi beberapa komponen hara yang sangat penting. penting mengurangi, Tanaman hanya dapat menghasilkan jumlah hara yang rendah.

Pupukan berfungsi sebagai sumber hara untuk pertumbuhan. digunakan untuk dapat ditanam.

## 2.3 Proses Penjualan

Penjualan adalah perjanjian suatu barang akan membayar untuk menghitung harga jual barang tersebut. Dalam penjualan, kita harus mempertimbangkan semua opsi yang tersedia untuk mengubah prospek menjadi pelanggan yang membayar, yang

tidak sederhana. Untuk memahami kebutuhan pelanggan dan melakukan penjualan, Anda perlu berinteraksi dengan mereka secara teratur, mengetahui suka dan tidak suka mereka, dan menjual produk atau layanan Anda dengan harga yang tepat. Jika mereka sudah menjadi pelanggan Anda, Anda perlu memelihara hubungan dengan tindak lanjut dan memperbarui mereka tentang layanan atau barang baru.

## **2.4 Kelompok tani**

Kelompok petani, peternak, yang saling mengenal dan bekerja sama untuk memperluas bisnis mereka. Dengan peran mereka sebagai pelaku utama dalam pembangunan

### **a. Karakteristik Kelompok Petani**

Anggota yang saling mengenal, akrab, dan percaya satu sama lain memiliki perspektif. Mereka Berdasarkan kesepakatan bersama, anggota dibagi tugas dan tanggung jawab.

### **b. Beberapa Fungsi Dari Kelompok Tani :**

1. Media untuk belajar di mana anggota mengajar satu sama lain untuk mereka kualitas hidup mereka.
2. Tempat untuk meningkatkan sesama kelompok tani, dan pihak lain. Ini membantu usaha tani menjadi dan lebih siap dan penyelesaian permasalahan di kelompok tersebut.
3. Bisnis produk dari hasil pertanian sehingga petani dapat mengetahui harga serta tanaman yang akan di tanam.

## **2.5 RapidMiner**

*RapidMiner* merupakan perangkat lunak yang bersifat terbuka (*open source*).

### **a. *RapidMiner* memiliki beberapa sifat sebagai berikut:**

1. Ditulis dengan bahasa *pemrograman Java* sehingga dapat dijalankan di berbagai sistem operasi.

**b. Beberapa Fitur dari *RapidMiner*, antara lain:**

1. Menyediakan *prosedur* untuk pengolahan data dan pembelajaran mesin yang mencakup: *preprocessing* data, *visualisasi*, model, dan evaluasi
2. Dideteksi dengan XML, dan dilengkapi dengan GUI.
3. Statistika R

## 2.6 Tinjauan Pustaka Terkait Masalah Penelitian

Tabel 2.1 Tinjauan Pustaka.

Nama Peneliti	Judul	Tahun	Pola Yang Dianalisa	Kekuatan Penelitian	Kelemahan Penelitian
Adji Dwi Alfarizi & Andri	Pemanfaatan Data Mining Dalam Memprediksi Produksi Pada PT Pupuk Sriwidjaja Palembang Menggunakan Algoritma <i>Regresi Linier</i> Berganda	2018	Dengan menggunakan data historis dan pola yang ditemukan, penelitian ini memprediksi jumlah produksi pupuk di masa depan. Algoritma regresi linier berganda membantu memperkirakan produksi di tahun 2021 berdasarkan pola yang terdeteksi dari data tahun 2019-2020	Dengan memberikan prediksi yang akurat tentang produksi di masa depan, penelitian ini memberikan alat yang berguna bagi manajemen untuk membuat keputusan strategis, seperti perencanaan produksi, alokasi sumber daya, dan peningkatan efisiensi	Meskipun model ini dapat menghasilkan prediksi yang bermanfaat untuk jangka pendek hingga menengah, kemampuan untuk memprediksi dalam jangka panjang mungkin terbatas, terutama jika terjadi perubahan signifikan dalam teknologi, kebijakan, atau lingkungan bisnis
Elvrida Jesika Purba	Penerapan Data Mining Untuk Prediksi Stok Produk	2021	Dalam penelitian yang menggunakan metode <i>regresi linear</i> berganda	kemampuannya untuk memberikan wawasan mendalam, membuat	<i>Variabilitas</i> tinggi dalam data penjualan atau jumlah produk yang masuk bisa

	Susu Pada Pt.Ps Maju Bersama Menggunakan Metode <i>Regresi Linear</i> Berganda		untuk memprediksi stok produk susu di PT. PS Maju Bersama, pola yang dianalisis melibatkan hubungan antara <i>variabel bebas (independent variables)</i> dan <i>variabel terikat (dependent variable)</i> . Berikut adalah pola-pola yang dianalisis, Hubungan antara Jumlah Masuk dan Stok Akhir, bagaimana perusahaan dapat mengoptimalkan stok produk.	prediksi yang akurat, dan menyediakan informasi yang berguna untuk pengambilan keputusan yang lebih baik dalam pengelolaan stok produk susu.	menyebabkan ketidakstabilan dalam model prediksi, membuat hasilnya kurang dapat diandalkan..
Dewi Hastuti, Hendri Wibowo, Endah Subekti, Proyoga Aditama	Analisis Produksi <i>Cobb Douglas</i> Dengan Metode <i>Regresi Linier</i>	2022	Dengan menggunakan pola analisis ini, penelitian bertujuan untuk memberikan pemahaman	Penelitian berhasil mengidentifikasi faktor-faktor produksi yang <i>signifikan</i> seperti luas lahan,	Dengan hanya 25 petani sebagai <i>responden</i> , ada kemungkinan bahwa hasil penelitian tidak sepenuhnya

	Berganda Pada Usaha Tani Bawang Daun ( <i>Allium Fistulosum L</i> ) (Studi Kasus Desa Sidomukti Kecamatan Bandungan Kabupaten Semarang)		yang mendalam tentang faktor-faktor yang mempengaruhi produksi bawang daun dan kondisi skala produksi yang terjadi di petani bawang daun Desa Sidomukti.	pupuk kandang, dan pupuk urea yang mempengaruhi hasil produksi, memberikan panduan praktis bagi petani dalam mengelola usahatannya.	mewakili <i>populasi</i> petani bawang daun yang lebih luas. Ini dapat membatasi <i>generalisasi</i> temuan ke area yang lebih luas atau ke petani dengan kondisi yang berbeda.
Harsiti1, Zaenal Muttaqin2, Ela Srihartini	Penerapan Metode <i>Regresi Linier</i> Sederhana Untuk <i>Prediksi</i> Persediaan Obat Jenis <i>Tablet</i>	2022	Dalam penelitian ini, pola yang dianalisis adalah pola penjualan dan persediaan obat jenis <i>tablet</i> (khususnya <i>ibuprofen</i> ) di Klinik Teluk Banten. Peneliti menggunakan metode <i>regresi linier</i> sederhana untuk memprediksi persediaan obat di masa mendatang berdasarkan	Hasil <i>prediksi</i> menggunakan metode <i>regresi linier</i> sederhana menunjukkan tingkat akurasi sebesar <b>98,505%</b> , yang berarti model ini memiliki kemampuan prediksi yang sangat baik. Ini menunjukkan bahwa metode <i>regresi linier</i> sederhana mampu menangkap pola data	kelemahan utama dari metode <i>regresi linier</i> sederhana adalah keterbatasannya dalam menangkap <i>kompleksitas</i> nyata dari data penjualan obat dan ketergantungannya pada pola <i>linear</i> , yang mungkin tidak selalu ada dalam <i>konteks klinik</i> atau rumah sakit yang menghadapi

			data penjualan dari periode sebelumnya.	dengan baik dan memprediksi jumlah persediaan obat yang dibutuhkan dengan tingkat kesalahan yang rendah.	variasi permintaan yang lebih dinamis
Wiga Maulana Baihaqi1 , Melia Dianingrum2 , Kurnia Aswin Nuzul Ramadhan3	<i>Regresi Linier</i> Sederhana untuk Memprediksi Kunjungan Pasien di Rumah Sakit Berdasarkan Jenis Layanan dan Umur Pasien	2019	Pola yang dianalisis dalam penelitian ini bertujuan memberikan wawasan yang lebih dalam tentang faktor-faktor yang mempengaruhi kunjungan pasien di RSUD Cilacap. Dengan memahami pola-pola ini, pihak manajemen dapat merencanakan strategi pelayanan yang lebih baik dan menyesuaikan sumber daya sesuai dengan kebutuhan pasien	Model <i>regresi linier</i> sederhana mudah dipahami dan diinterpretasikan, memungkinkan manajemen rumah sakit untuk dengan cepat mendapatkan wawasan dari data kunjungan pasien	Terdapat banyak faktor lain yang dapat mempengaruhi kunjungan pasien, seperti tren kesehatan masyarakat, penyakit musiman, atau kebijakan pemerintah, yang mungkin tidak diperhitungkan dalam model.