

BAB II

LANDASAN TEORI

2.1 Tinjauan Studi

Sebelum penelitian ini dilakukan, sudah terdapat beberapa penelitian yang menjadi dasar untuk menyelesaikan penelitian ini, penelitian tersebut diantaranya sebagai berikut:

Yuliana, dosen Universitas Widya Dharma program studi Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan dengan judul penelitian “Penerapan Model Regresi Linear Robust Dengan Estimasi M Pada Data Nilai Kalkulus II Mahasiswa Universitas Widya Dharma Klaten”. Penelitian ini bertujuan untuk menunjukkan bahwa regresi linear robust dengan estimasi M dapat mengatasi suatu data yang mengandung pengamatan outlier. Setelah itu, peneliti mendapatkan persamaan model regresi linear robust:

$$Y = 0.177 X_1 + 0.655 X_2 - 0.947$$

Dengan Y menunjukkan nilai kalkulus II, X_1 menunjukkan nilai kalkulus I, dan X_2 menunjukkan nilai trigonometri.

Berdasarkan hasil penelitian dan pembahasan, penelitian ini dapat disimpulkan bahwa:

1. Melalui regresi linear robust dengan estimasi M diperoleh suatu estimasi parameter regresi yang konvergen tanpa harus membuang pengamatan *outliernya*. Hal ini berarti regresi linear robust dengan estimasi M dapat digunakan untuk mengatasi suatu data yang mengandung pengamatan *outlier*.
2. Dari model regresi robust yang telah didapat tersebut diperoleh suatu model regresi robust persamaan $Y = 0.177 X_1 + 0.655 X_2 - 0.947$. Dari model regresi linear robust ini dapat digunakan untuk memprediksikan suatu nilai kalkulus II secara tepat[5].

Ali Fikri, mahasiswa program studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Dian Nuswantoro dengan judul penelitian “Penerapan Data Mining Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression”. Penelitian ini menggunakan teknik data mining dengan metode estimasi yang bertujuan mempermudah dalam menghitung kekuatan beton yang nantinya akan dihasilkan pada saat pembangunan.

Kesimpulan dari penelitian ini adalah dengan adanya aplikasi perhitungan kekuatan beton yang nantinya dihasilkan saat pembangunan maka akan mempermudah pihak CV. Sinar Harapan Semarang dalam meningkatkan kualitas beton yang bagus dan memiliki ketahanan yang kuat. Sehingga akan bisa mengetahui hasilnya sesuai dengan rencana. Dimana penggunaan metode *Linear Regression* sangat baik untuk pemecahan kasus perhitungan kekuatan beton yang akan dihasilkan berdasarkan komponen yang digunakan. Hal ini menjadikan *Linear Regression* menjadi alternatif lain sebagai metode yang layak dijadikan acuan untuk mengembangkan model estimasi pada kasus-kasus lain. Serta dapat dikembangkan menjadi sebuah sistem pendukung keputusan untuk menghitung kekuatan beton yang akan dihasilkan[6].

Eggy Inaidi Andana Warih, mahasiswa program studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Dian Nuswantoro dengan judul penelitian “Penerapan Data Mining Untuk Menentukan Estimasi Produktivitas Tanaman Tebu Dengan Menggunakan Algoritma Linear Regresi Berganda di Kabupaten Rembang”. Penelitian ini bertujuan untuk membantu memperoleh hasil seberapa besar produksi tanaman tebu di Kabupaten Rembang sebagai acuan petani untuk menambah hasil panen tebu di tiap tahunnya.

Dari perhitungan penelitian, di dapat bahwa hasil estimasi produktivitas mendapatkan 15.132,00067 Kg/Ha yang sebelumnya lahannya adalah 147313 hektar (ha), berarti produktivitas tahun depannya mengalami peningkatan dari tahun sebelumnya[7].

Yuniarsi Rahayu, S.Si., M.Kom., dosen program studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Dian Nuswantoro dengan judul penelitian “Penerapan Metode Numerik Pada Peramalan Untuk Menghitung Koefisien-koefisien Pada Garis Regresi Linear Berganda”. Penelitian ini membahas tentang salah satu penerapan dalam metode numerik, yaitu pada masalah penerapan pada peramalan untuk menghitung koefisien-koefisien pada garis regresi linear berganda.

Penelitian ini menghasilkan koefisien-koefisien pada persamaan regresi linear berganda dengan menggunakan ke-3 metode (Metode Cramer, Metode Eliminasi Gauss-Jordan, dan Metode Matriks Balikan) adalah sama yaitu menghasilkan persamaan regresi linear berganda sebagai berikut:

$$Y = 9.9958 + 0.5502 X_1 + 0.0552 X_2 + 0.4609 X_3$$

Dari pembahasan dan analisis data, penelitian ini dapat disimpulkan sebagai berikut:

1. Permasalahan yang dibahas ini, adalah studi kasus dengan satu perubah terikat (Y) dan tiga perubah X_{1i} , X_{2i} , X_{3i} .
2. Penggunaan Metode Numerik dalam menghitung koefisien-koefisien pada regresi linear berganda.
3. Metode yang digunakan disini adalah Metode Cramer, Metode Eliminasi Gauss-Jordan, dan Metode Matriks Balikan yang menghasilkan 4 persamaan linear dengan 4 variabel.
4. Penggunaan Matlab dalam perhitungan.
5. Hasil perhitungan koefisien-koefisien regresi linear berganda yang diperoleh dari penggunaan 3 metode tersebut adalah sama yaitu $a_0 = 9.9958$; $a_1 = 0.5502$; $a_2 = 0.0552$; $a_3 = 0.4609$ sehingga persamaan regresi linear berganda adalah $Y = 9.9958 + 0.5502 X_1 + 0.0552 X_2 + 0.4609 X_3$
6. Kesalahan baku (*standard error*) regresi adalah 2.0792[8]

Tabel 2.1 Ringkasan Tinjauan Studi

No	Peneliti	Judul Penelitian	Tahun	Metode / Algoritma	Hasil Penelitian
1	Yuliana	Penerapan Model Regresi Linear Robust Dengan Estimasi M Pada Data Nilai Kalkulus II Mahasiswa Universitas Widya Dharma Klaten	2014	Model Regresi Linear Berganda (Model Regresi Linear Robust)	Melalui regresi linear robust dengan estimasi M diperoleh suatu estimasi parameter regresi yang konvergen tanpa harus membuang pengamatan <i>outliernya</i> . Hal ini berarti regresi linear robust dengan estimasi M dapat digunakan untuk mengatasi suatu data yang mengandung pengamatan <i>outlier</i> . Dari model

					<p>regresi robust yang telah didapat tersebut diperoleh suatu model regresi robust persamaan $Y = 0.177 X_1 + 0.655 X_2 - 0.947$. Dari model regresi linear robust ini dapat digunakan untuk memprediksikan suatu nilai kalkulus II secara tepat.</p>
2	Ali Fikri	Penerapan Data Mining Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode	2013	Model Regresi Linear Berganda	<p>Dengan adanya aplikasi perhitungan kekuatan beton yang nantinya dihasilkan saat pembangunan maka akan mempermudah pihak CV.</p>

		Estimasi Menggunakan Linear Regression			Sinar Harapan Semarang dalam meningkatkan kualitas beton yang bagus dan memiliki ketahanan yang kuat. Sehingga akan bisa mengetahui hasilnya sesuai dengan rencana. Dimana penggunaan metode <i>Linear Regression</i> sangat baik untuk pemecahan kasus perhitungan kekuatan beton yang akan dihasilkan berdasarkan komponen yang digunakan.
--	--	---	--	--	---

3	Eggy Inaidi Andana Warih	Penerapan Data Mining Untuk Menentukan Estimasi Produktivitas Tanaman Tebu Dengan Menggunakan Algoritma Linear Regresi Berganda di Kabupaten Rembang	2015	Model Regresi Linear Berganda	Hasil estimasi produktivitas mendapatkan 15.132,00067 Kg/Ha yang sebelumnya lahannya adalah 147313 hektar (ha), berarti produktivitas tahun depannya mengalami peningkatan dari tahun sebelumnya.
4	Yuniarsi Rahayu, S.Si., M.Kom.	Penerapan Metode Numerik Pada Peramalan Untuk Menghitung Koefisien-koefisien Pada Garis Regresi Linear Berganda	2011	Model Regresi Linear Berganda (Metode Cramer, Metode Eliminasi Gauss-Jordan, dan Metode Matriks	Penelitian ini menghasilkan koefisien-koefisien pada persamaan regresi linear berganda dengan menggunakan ke-3 metode (Metode Cramer, Metode

				Balikan)	<p>Eliminasi Gauss-Jordan, dan Metode Matriks Balikan) adalah sama yaitu menghasilkan persamaan regresi linear berganda sebagai berikut:</p> $Y = 9.9958 + 0.5502 X_1 + 0.0552 X_2 + 0.4609 X_3$
--	--	--	--	----------	--

2.2 Tinjauan Pustaka

2.2.1 Data

Data merupakan keterangan atau fakta yang dikumpulkan dari suatu populasi yang akan digunakan untuk menerangkan ciri-ciri populasi yang bersangkutan. Data harus memenuhi kriteria sebagai berikut[14]:

- Objektif, ialah data yang sama dengan keadaan sebenarnya.
- Mewakili populasi
- Galat baku (standard error) kecil
- Tepat waktu
- Relevan

Data dibedakan atas beberapa bagian sebagai berikut[14]:

a. Menurut Sifatnya

1. Data kualitatif

Data kualitatif adalah data bukan dalam bentuk bilangan (non numerik).

2. Data kuantitatif

Data kuantitatif adalah data dalam bentuk bilangan.

b. Menurut Cara Memperolehnya

1. Data primer

Data primer merupakan data yang langsung diperoleh dari lapangan melalui percobaan, survey, dan observasi.

2. Data sekunder

Data sekunder diperoleh dari data primer, biasanya dalam publikasi misalnya data statistik hasil riset dari surat kabar atau majalah.

c. Menurut Waktu

1. Data silang

Data silang merupakan data yang dikumpulkan dalam waktu yang sifatnya temporer.

2. Data berkala

Data berkala merupakan data yang dikumpulkan setiap periode.

d. Menurut Sumber

1. Data internal

Data internal merupakan data yang dikumpulkan oleh unit kerja dalam lingkungannya.

2. Data eksternal

Data eksternal merupakan data yang diambil dari unit lain.

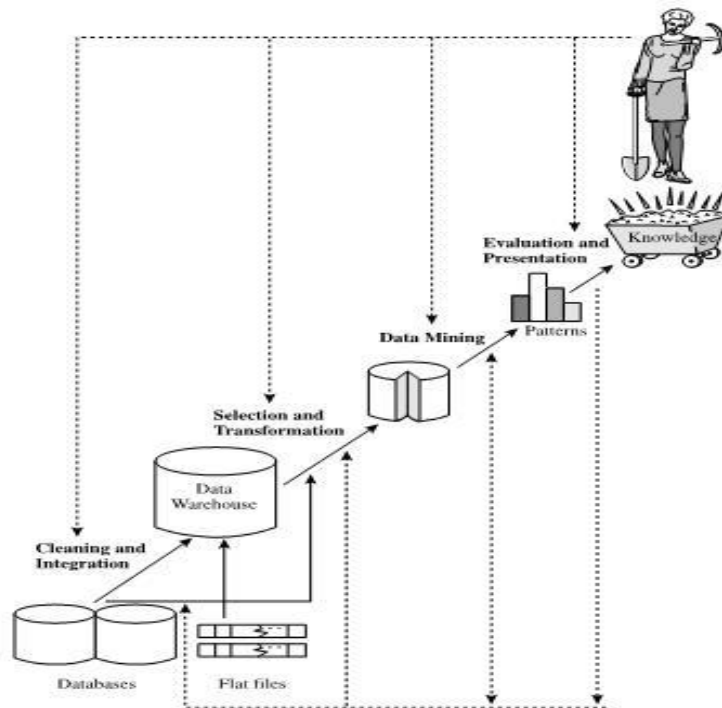
2.2.2 Data Mining

Data mining adalah proses menemukan pola yang berguna dan tren di set data yang besar[9]. Sejak tahun 1990-an, gagasan data mining, biasanya dilihat sebagai proses “pertambangan/mining” data, telah muncul di banyak lingkungan, dari bidang akademik, bisnis atau kegiatan medis, khususnya. Data mining masih diperdebatkan oleh beberapa bidang ilmiah, termasuk Daryl Pregibons yang menyatakan bahwa “data mining adalah campuran statistic, *Artificial Intelligence* (kecerdasan buatan), dan penelitian database”[10].

Terdapat dua jenis tujuan data mining, antara lain:

1. Prediktif (misalnya, klasifikasi, regresi, bias/deteksi anomaly, dll), menggunakan beberapa variabel yang ada untuk memprediksi nilai masa depan (belum diketahui) dari variabel lain.
2. Deskriptif (misalnya, *clustering*, *association rule*, pola sekuensial, dll), identifikasi pola yang menggambarkan data dan dapat dengan mudah dipahami oleh pengguna[10].

Beberapa orang menyebutkan data mining sebagai sinonim dari istilah *knowledge discovery from data* (KDD), sementara yang lain melihat data mining hanya sebagai langkah penting dalam proses penemuan pengetahuan[11]. Sedangkan suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap, sebagai berikut:



Gambar 2.1 Tahapan Data Mining

Gambar 2.1 merupakan gambaran tahapan yang dilakukan proses data mining, dengan penjelasan sebagai berikut:

1. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh, bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi data (*data selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analisis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh, beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa

angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

5. Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

6. Presentasi pengetahuan (*knowledge presentation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat[12].

2.2.3 Model Regresi

Sir Francis Galton (1822-1911), gagasan yang diluncurkan pada akhir abad ke-19 “regresi terhadap mean”, prinsip menurutnya. Diberikan dua pengukuran, nilai taksiran pengukuran kedua adalah lebih dekat dengan mean dari nilai yang diamati terhadap pengukuran pertama (misalnya, ayah yang tinggi memiliki anak lebih tinggi, anak-anak tinggi regresi dengan rata-rata tinggi). Analisis regresi berarti model matematika yang menetapkan (konkret oleh persamaan regresi) hubungan antara nilai-nilai variabel tertentu (respon/hasil/variabel dependen) dan nilai-nilai dari variabel lain (prediktor/variabel independen).

Analisis regresi berkaitan pada prinsip untuk:

1. Penentuan hubungan kuantitatif antara beberapa variabel.
2. Peramalan nilai-nilai variabel sesuai dengan nilai-nilai dari variabel lain (menentukan pengaruh dari “variabel prediktor” pada “variabel respon”)[10].
 - Variabel Respon disebut juga variabel dependen yaitu variabel yang dipengaruhi oleh variabel lainnya dan dinotasikan dengan variabel Y.
 - Variabel Prediktor disebut juga variabel independen yaitu variabel yang bebas (tidak dipengaruhi oleh variabel lainnya) dan dinotasikan dengan variabel X.

Model regresi terbagi menjadi dua, yaitu:

1. Model regresi linear sederhana, yaitu hubungan antara satu variabel dependen dan variabel independen.
2. Model regresi linear berganda, yaitu hubungan antara dua atau lebih variabel dependen dan variabel independen.

Regresi linear adalah teknik alami untuk mempertimbangkan atau menyelesaikannya ketika hasil atau kelas adalah numerik dan semua atribut adalah numerik. Ini adalah metode pokok dalam statistik. Berikut adalah bobot yang telah ditentukan dalam proses penyelesaian dengan menggunakan model regresi linear berganda[2]:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

Dimana:

- Y = kelas (variabel dependen/tidak bebas)
 a = konstanta

b_1, b_2, \dots, b_k = bobot atau koefisien regresi
 X_1, X_2, \dots, X_3 = nilai atribut (variabel independen/bebas)

Model atau metode regresi linear merupakan metode yang cukup populer dan biasanya digunakan untuk menemukan persamaan dari sebuah data yang dimana data tersebut saling berhubungan antara variabel satu dengan variabel yang lain dalam satu database yang cukup besar. Informasi yang dihasilkan dari data mining dengan metode regresi linear bisa dijadikan sebagai dasar untuk pengambilan keputusan[6].

2.2.4 Root Mean Square Error (RMSE)

Dalam regresi linear berganda, nilai a, b_1, b_2, \dots, b_k dapat dihitung menggunakan Metode Kuadrat Terkecil (MKT), salah satunya yaitu *Root Mean Square Error (RMSE)*. RMSE adalah ukuran yang sering digunakan dari perbedaan antara nilai-nilai prediksi oleh model atau estimator. RMSE berfungsi untuk menghitung dan menjadi ukuran besaran kesalahan dalam prediksi[6].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$

Keterangan:

Y_i = data awal (data sebenarnya)

\hat{Y}_i = data akhir (data hasil estimasi)

n = jumlah data

Keakuratan pengukuran estimasi ditunjukkan dengan adanya hasil RMSE yang kecil (mendekati nol). RMSE yang lebih kecil dikatakan lebih akurat dibandingkan RMSE yang lebih besar[15].

Standar error yang baik adalah kisaran 0.0 – 1.0, lebih dari 1.0 sudah tergolong besar[16].

2.2.5 Estimasi

Estimasi adalah suatu metode dimana kita dapat memperkirakan nilai populasi (suatu objek yang diteliti) dengan memakai nilai sampel (contoh objek yang diambil untuk dijadikan penelitian)[13].

Salah satu contoh tugas estimasi dalam penelitian adalah memperkirakan nilai rata-rata (IPK) dari seorang mahasiswa pascasarjana berdasarkan IPK mahasiswa sarjana[9].

2.3 Kerangka Pemikiran

