

Penerapan Data Mining Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression

Ali Fikri A11.2009.04692
 Program Studi Teknik Informatika – S1
 Fakultas Ilmu Komputer
 Universitas Dian Nuswantoro, Jl. Nakula I No. 5-11, Semarang
alifikri84@gmail.com

ABSTRAK

CV.Sinar Harapan merupakan sebuah instansi yang bergerak dibidang teknik sipil. Dimana dalam teknik sipil membutuhkan persiapan yang sangat baik sebelum melakukan pekerjaan yang dilakukan, salah satunya dalam memperkirakan kekuatan beton yang nantinya akan dihasilkan. Karena kekuatan beton merupakan hal yang sangat penting dalam sebuah pembangunan sebuah gedung maupun jalan dan dengan perkembangan teknologi yang sangat pesat maka dalam dunia teknik sipil pun mau tidak mau harus mengikuti apa yang sekarang berkembang dengan memanfaatkan data yang banyak dan melimpah agar menjadi sebuah informasi yang berguna untuk CV.Sinar Harapan. Dengan memanfaatkan perkembangan teknologi sekarang maka dengan melimpahnya data tersebut penulis menggunakan teknik data mining dengan metode estimasi untuk menghitung kekuatan beton yang dihasilkan dengan menggunakan algoritma linear regression yang dimana algoritma ini dipilih sesuai dengan data yang ada yaitu komponen yang digunakan yang digunakan untuk menghitung kekuatan beton yang dihasilkan. Untuk metode pengujiannya menggunakan Cross Validation dan evaluasinya menggunakan Root Mean Square Error (RMSE) untuk mengetahui seberapa besar kesalahan yang dihasilkan dari Linear Regression tersebut.

Kata kunci : Kekuatan Beton, *Data Mining*, Teknik *estimasi*, *Linear Regression*, *Cross Validation*, *Root Mean Square Error*.

I. PENDAHULUAN

1.1 Latar Belakang

CV. Sinar Harapan Semarang merupakan salah satu perusahaan yang bergerak dibidang pembangunan atau teknik sipil. Dimana dalam dunia teknik sipil sekarang berkembang sangat pesat. Oleh karena itu setiap perusahaan harus memiliki strategi dalam menentukan kualitas produk yang dihasilkannya. Salah satunya dalam bagaimana mengembangkan informasi yang sudah ada menjadi informasi yang lebih berguna. Tapi didalam perusahaan biasanya informasi yang dibutuhkan tidak sesuai dengan harapan atau kurang memadai dalam hal penyajian informasi yang dibutuhkan. Dengan kemampuan teknologi informasi untuk mengumpulkan data dan informasi saat ini sedang gencar untuk digali data tersebut untuk mendapatkan sebuah informasi yang cepat dan akurat tentunya sesuai harapan yang diinginkan oleh yang membutuhkan informasi tersebut.

CV. Sinar Harapan Semarang dalam hal ini dihadapkan pada sebuah kasus dalam mengetahui perhitungan tingkat kekuatan beton yang dihasilkan pada saat pembangunan dimana hal ini dipengaruhi oleh berbagai

macam bahan yang digunakan. Pemanfaatan data yang ada dalam sistem informasi untuk menunjang keputusan yang dimana tidak hanya mengandalkan data operasional yang ada, sehingga memanfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna dalam mengambil keputusan. Dan untuk mengatasi masalah penggalian informasi dari data dalam jumlah besar, yang disebut data mining. Dimana dengan penggunaan teknik data mining diharapkan dapat memberikan pengetahuan – pengetahuan yang sebelumnya tersembunyi didalam gudang data sehingga menjadi informasi yang bisa menghasilkan informasi tentang perhitungan tingkat kekuatan beton yang dihasilkan oleh CV Sinar Harapan Semarang.

Dengan adanya perhitungan tingkat kekuatan beton yang dihasilkan oleh perusahaan CV Sinar Harapan Semarang ini maka akan mempermudah perusahaan dalam menghitung komponen atau bahan – bahan yang akan digunakan. dengan memanfaatkan data pembangunan yang selama ini sudah ada dari berbagai bangunan yang sudah dibuat oleh perusahaan CV Sinar Harapan Semarang. Karena dimana pada setiap perusahaan dituntut untuk memiliki keunggulan bersaing dalam hal kualitas. Salah satu hal yang sekarang masih

berkembang yaitu dengan memanfaatkan semua sumber daya yang dimiliki. Selain sumber daya sarana, prasarana dan juga manusia, sistem informasi adalah salah satu sumber daya yang dapat digunakan untuk meningkatkan keunggulan bersaing. Sistem informasi dapat digunakan untuk mendapatkan, mengolah dan menyebarkan informasi untuk menunjang kegiatan operasional sehari – hari sekaligus menunjang kegiatan pengambilan keputusan strategis.

Dengan menggunakan metode estimasi yang ada pada data mining yang sesuai dengan keperluan perusahaan yaitu untuk mengetahui tingkat kekuatan beton yang nantinya akan dihasilkan. Menggunakan linear regression yang dimana nanti dengan memanfaatkan data yang sudah ada pada CV Sinar Harapan Semarang maka nanti untuk menghasilkan persamaan yang digunakan untuk menghitung tingkat kekuatan beton yang nantinya akan dihasilkan.

Karena selama ini belum ada aplikasi untuk menghitung tingkat kekuatan beton yang dihasilkan dengan bahan baku yang digunakan pada CV Sinar Harapan Semarang. Sehingga dengan ini penulis tertarik untuk melakukan “Penerapan Data Mining Untuk Mengetahui Tingkat Kekuatan Beton Yang Dihasilkan Dengan Metode Estimasi Menggunakan Linear Regression (studi kasus di CV Sinar Harapan Semarang)”.

1.2 Tujuan

Dengan aplikasi ini akan mempermudah CV.Sinar harapan dalam menghitung kekuatan beton yang nantinya akan dihasilkan pada saat pembangunan.

1.3 Batasan Masalah

Mengingat luasnya ruang lingkup sistem pendukung keputusan untuk penerimaan pegawai ini, maka penulis memberikan batasan permasalahan pada pembuatan Tugas Akhir ini pada :

- a. Membuat suatu aplikasi yang dapat membantu perhitungan kekuatan beton yang sesuai dengan rencana.
- b. Penerapan metode estimasi untuk mengetahui persamaan yang digunakan untuk menghitung tingkat kekuatan beton.
- c. Pembuatan aplikasi ini hanya sampai pada pengujian sistem

II. TINJAUAN PUSTAKA

2.1 Pengertian Data Mining

Data mining adalah ekstraksi pola yang menarik dari data dalam jumlah besar [4]. Dengan kata lain, data mining merupakan rangkaian proses untuk menghasilkan suatu informasi yang bernilai dari sekumpulan data yang tidak bernilai. Jadi data mining juga bisa dikatakan sebagai pengolahan data menjadi suatu informasi yang memberikan pengetahuan.

Istilah data mining dan Knowledge Discovery in Databases (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda akan tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining.

2.2 Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu :

2.2.1 Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2.2.2 Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada kearah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai variabel dari target sebagai nilai prediksi, selanjutnya pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

2.2.3 Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.

2.2.4 Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori, sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, sedang dan rendah.

2.2.5 Pengklusteran
 Pengklusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan lainnya dan memiliki ketidakmiripan dengan record-record dalam kluster lain.

2.2.6 Asosiasi
 Tugas Asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis disebut dengan analisis keranjang belanja.

Dari macam macam pengelompokan data mining kita dapat menentukan apa yang akan kita pilih untuk melakukan sebuah penelitian.

Metode Data Mining

Ada beberapa klasifikasi metode data mining, sebagai berikut.

1. Association rule learning

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh aturan asosiatif dari analisa pembelian di suatu pasar swalayan adalah dapat diketahuinya berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan susu. Dengan pengetahuan tersebut pemilik pasar swalayan dapat mengatur penempatan barangnya atau merancang kampanye pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

Analisis asosiasi menjadi terkenal karena aplikasinya untuk menganalisa isi keranjang belanja di pasar swalayan. Analisis asosiasi juga sering disebut dengan istilah market basket analysis.

Analisis asosiasi dikenal juga sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola frekuensi tinggi (frequent pattern mining) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien.

2. Clustering

Clustering adalah metode penganalisaan data, yang sering dimasukkan sebagai salah satu metode Data Mining, yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama ke suatu 'wilayah' yang sama dan data dengan karakteristik yang berbeda ke 'wilayah' yang lain.

3. Decision tree

Dalam decision tree tidak menggunakan vector jarak untuk mengklasifikasikan obyek. Seringkali data observasi mempunyai atribut-atribut yang bernilai nominal. Seperti yang diilustrasikan pada, misalkan obyeknya adalah sekumpulan buah-buahan yang bisa dibedakan berdasarkan atribut bentuk, warna, ukuran dan rasa. Bentuk, warna, ukuran dan rasa adalah besaran nominal, yaitu bersifat kategoris dan tiap nilai tidak bisa dijumlahkan atau dikurangkan. Dalam atribut warna ada beberapa nilai yang mungkin yaitu hijau, kuning, merah. Dalam atribut ukuran ada nilai besar, sedang dan kecil. Dengan nilai-nilai atribut ini, kemudian dibuat decision tree untuk menentukan suatu obyek termasuk jenis buah apa jika nilai tiap-tiap atribut diberikan

4. Linear Regression

Linier Regresi ini diasumsikan bahwa terdapat hubungan antara variable yang ingin diramalkan (variabel tak bebas) dengan variabel lain (variabel bebas). Selanjutnya peramalan ini didasarkan pada asumsi bahwa pola pertumbuhan data historis yang bersifat linier, walaupun sebenarnya tidak 100% linier. Pola pertumbuhan ini didekati dengan suatu model yang menggambarkan hubungan-hubungan yang terkait dalam suatu keadaan

Metode Linear Regression itu sendiri 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 Linear Regression ini sendiri juga bisa dijadikan sebagai dasar untuk pengambilan keputusan. Metode data mining ini biasanya dipakai dalam data perhitungan sebuah waktu yang dibutuhkan seseorang untuk mengantar pizza kepada konsumennya, yang nantinya akan menunjukkan jarak yang akan ditempuh oleh pengantar pizza itu, jumlah pesanan yang

dipesan oleh konsumen dan juga banyaknya lampu lalu lintas yang akan dilewatinya. Metode ini jika diaplikasikan pada aplikasi data mining perhitungan kekuatan beton yang nantinya akan dihasilkan dengan berdasarkan bahan baku yang nantinya akan digunakan. Jadi misalnya kita memiliki semen, air dan juga batu kecil yang nanti kita pakai, sehingga nantinya dengan menggunakan teknik data mining ini akan dapat menghasilkan persamaan yang nantinya bisa kita gunakan untuk menghitung ketahanan atau kekuatan beton yang telah kita buat oleh CV. Sinar Harapan Semarang.

2.3 Metode Linear Regression

Algoritma Linear Regression termasuk jenis aturan classification and regression pada data mining selain Linear Regression yang termasuk pada golongan ini adalah Support Vector Machine, Logistic Regression dan lain-lain. Analisis Linear Regression adalah teknik data mining untuk menentukan bahwa terdapat hubungan antara variabel yang ingin diramalkan (variabel tak bebas) dengan variabel lain (variabel bebas). Selanjutnya peramalan ini didasarkan pada asumsi bahwa pola pertumbuhan data historis yang bersifat linier, walaupun sebenarnya tidak 100% linier. Pola pertumbuhan ini didekati dengan suatu model yang menggambarkan hubungan-hubungan yang terkait dalam suatu keadaan

Diagram yang menggambarkan nilai-nilai observasi peubah takbebas dan peubah bebas.

- Nilai peubah bebas ditulis pada sumbu X (sumbu horizontal)
- Nilai peubah takbebas ditulis pada sumbu Y (sumbu vertikal)
- Nilai peubah takbebas ditentukan oleh nilai peubah bebas.

Bentuk Umum Regresi Linier Berganda

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8$$

Y : Kekuatan Beton

a : konstanta

X1 : Semen

b1 : kemiringan ke-1

X2 : Biji Besi

b2 : kemiringan ke-2

X3 : Abu Terbang

b3 : kemiringan ke-3

X4 : Air

b4 : kemiringan ke-4

X5 : Superplasticizer

b5 : kemiringan ke-5

X6 : Batu Kerikil

b6 : kemiringan ke-6

X7 : Pasir

b7 : kemiringan ke-7

X8 : Umur

b8 : kemiringan ke-8

a, b1 dan b2 didapatkan dengan menyelesaikan tiga persamaan Normal berikut:

$$(i) \quad n a + b_1 \sum_{i=1}^n x_{1i} + b_2 \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i$$

$$(ii) \quad a \sum_{i=1}^n x_{1i} + b_1 \sum_{i=1}^n x_{1i}^2 + b_2 \sum_{i=1}^n x_{2i} x_{1i} = \sum_{i=1}^n x_{1i} y_i$$

$$(iii) \quad a \sum_{i=1}^n x_{2i} + b_1 \sum_{i=1}^n x_{2i} x_{1i} + b_2 \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n x_{2i} y_i$$

n : banyak pasangan data

yi : kekuatan beton

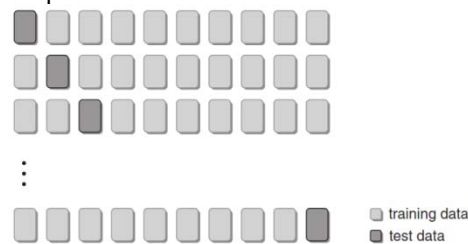
x1i : komponen1

x2i : komponen2

2.4 Cross Validation

Cross Validation adalah cara menemukan parameter terbaik dari suatu model dengan cara menguji besarnya error pada data test. Dalam cross validation, data akan dibagi menjadi k sampel dengan ukuran yang sama. Kemudian k-1 sampel digunakan untuk training dan 1 sampel sisanya untuk testing. Cara ini sering disebut validasi k-fold [9]. Kemudian dilakukan proses silang dimana data testing dijadikan sebagai data training dan sebaliknya data training sebelumnya dijadikan sebagai data testing.

Dalam cross validation kita harus menetapkan jumlah partisi atau fold, standar yang biasa dan terkenal digunakan untuk memperoleh estimasi kesalahan terbaik adalah 10 kali partisi atau tenfold cross-validation.



Gambar 2.4: Ilustrasi tenfold cross validation

2.5 Root Mean Square Error

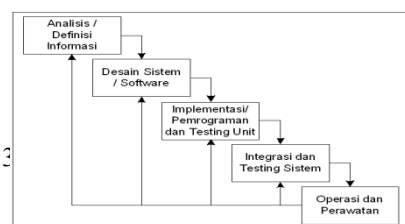
Root Mean Square Error (RMSE) adalah ukuran yang sering digunakan dari perbedaan antara nilai-nilai diprediksi oleh model atau estimator dan nilai-nilai sebenarnya diamati. Perbedaan-perbedaan individual disebut residual ketika perhitungan dilakukan atas sampel data yang digunakan untuk estimasi, dan

disebut kesalahan prediksi ketika dihitung out-of-sample. Rmsd berfungsi untuk agregat besaran kesalahan dalam prediksi untuk berbagai kali menjadi ukuran tunggal daya prediksi. Rmsd adalah ukuran akurasi yang baik, tapi hanya untuk membandingkan kesalahan peramalan model yang berbeda untuk variabel tertentu dan tidak antara variabel, karena skala-dependent, sebagai berikut rumus pada RMSE.

$$RMSE = \sqrt{\sum_{i=1}^N \sum_{d=1}^N \left[\frac{(\hat{T}_{i,d} - T_{i,d})^2}{N \cdot (N - 1)} \right]}$$

2.6 Metode Pengembangan Sistem

Perancangan Sistem dalam Penelitian ini akan menggunakan model WaterFall sehingga model inilah yang akan dibahas, langkah – langkah metode waterfall yaitu pengumpulan data yang digunakan dalam penulisan laporan tugas akhir ini:



Gambar 2.5 Model Waterfall

1. Analisis / definisi kebutuhan.
 - a. Ada dua aktifitas pada tahap ini yaitu:
 - a. Analisis Kebutuhan, yang menghasilkan garis besar kebutuhan.
 - b. Definisi kebutuhan, yang menghasilkan dokumen kebutuhan. Dalam tahap ini, sistem analis harus menggali informasi mengenai fungsi, sifat, tujuan dan kendala-kendala yang ada di dalam sistem, yang kemudian dituangkan menjadi definisi kebutuhan yang jelas. Tujuan dari tahap ini adalah untuk mengetahui bagaimana sebenarnya sistem yang akan dikembangkan.
2. Desain sistem dan software
 - a. Tahap ini terdiri atas:
 - a. Desain Sistem, yang menghasilkan konfigurasi sistem secara keseluruhan termasuk

penjelasan kebutuhan hardware dan software.

b. Desain Software, yang merupakan proses multistep berfokus pada 4 atribut pemrograman: struktur data, arsitektur software, spesifikasi interface, dan algoritma yang digunakan.

Tujuan tahap ini adalah untuk menerjemahkan kebutuhan menjadi representasi software yang bisa diukur, sebelum dilakukan

pemrograman/pengkodean.

3. Implementasi / Testing Unit

Dalam tahap ini, desain yang telah dibuat diterjemahkan dalam bentuk kode program yang dapat dieksekusi dan dimengerti oleh mesin. Setelahss tahap penerjemahan dalam coding program maka tahapan selanjutnya adalah pengujian untuk mencari error dalam penulisan kode tersebut sehingga dihasilkan unit program yang valid.
4. Integrasi dan testing sistem

Tahapan ini diterapkan jika sistem terbagi dalam sub – sub sistem yang lebih kecil. Dalam tahap ini unit program yang telah dibuat dan valid akan diintegrasikan dengan unit program lainnya, dan kemudian dilakukan pengujian secara keseluruhan sehingga siap untuk diaplikasikan / digunakan oleh user.
5. Operasi dan perawatan

Tujuan dari perawatan sistem ialah agar sistem yang telah dikembangkan dapat mengakomodasi perubahan-perubahan yang terjadi pada lingkungan sistem, sehingga kegiatan operasional dapat berjalan dengan baik. Di tahap perawatan, fase-fase awal pengembangan sistem diterapkan kembali.

III. PERANCANGAN DAN IMPLEMENTASI

3.1 Metode Linear Regression

Metode Linear Regression adalah metode statistika yang digunakan untuk

membentuk model hubungan antara variabel terikat (dependen; respon; Y) dengan satu atau lebih variabel bebas (independen, predictor, X). Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai regresi linier sederhana, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai regresi linier berganda.

Analisis regresi setidaknya-tidaknya memiliki 3 kegunaan, yaitu untuk tujuan deskripsi dari fenomena data atau kasus yang sedang diteliti, untuk tujuan control, serta untuk tujuan prediksi. Regresi mampu mendeskripsikan fenomena data melalui terbentuknya suatu model hubungan yang bersifat numeric. Regresi juga dapat digunakan untuk melakukan pengendalian (kontrol) terhadap suatu kasus atau hal-hal yang sedang diamati melalui penggunaan model regresi yang diperoleh. Selain itu, Model regresi juga dapat dimanfaatkan untuk melakukan prediksi untuk variabel terikat. Namun yang perlu diingat, prediksi didalam konsep regresi hanya boleh melakukan prediksi untuk variabel terikat. Namun yang perlu diingat, prediksi di dalam konsep regresi hanya boleh dilakukan di dalam rentang data dari variabel-variabel bebas yang digunakan untuk membentuk model regresi tersebut. Misal, suatu model regresi diperoleh dengan mempergunakan data variabel bebas yang memiliki rentang antara 5 s.d. 25, maka prediksi hanya boleh dilakukan bila suatu nilai yang digunakan sebagai input untuk variabel X berada di dalam rentang tersebut. Konsep ini disebut sebagai interpolasi.

Data untuk variabel independen X pada regresi linier bisa merupakan data pengamatan yang tidak ditetapkan sebelumnya oleh peneliti (observational data) ataupun data yang telah ditetapkan (dikontrol) oleh peneliti sebelumnya (experimental or fixed data). Perbedaannya adalah bahwa dengan menggunakan fixed data, informasi yang diperoleh lebih kuat dalam menjelaskan hubungan sebab akibat antara variabel X dan variabel Y. Sedangkan pada observational data, informasi yang diperoleh belum tentu merupakan hubungan sebab-akibat. Untuk fixed data, peneliti sebelumnya telah memiliki beberapa nilai variabel X yang diteliti. Sedangkan pada observational data, variabel X yang diamati bisa berapa saja, tergantung keadaan dilapangan. Biasanya,

fixed data diperoleh dari percobaan laboratorium, dan observational data diperoleh dengan menggunakan kuisioner.

Di dalam suatu model regresi kita akan menemukan koefisien-koefisien. Koefisien pada model regresi sebenarnya adalah nilai duga parameter di dalam model regresi untuk kondisi yang sebenarnya (true condition), sama halnya dengan statistik mean (rata-rata) pada konsep statistika dasar. Hanya saja, koefisien-koefisien untuk model regresi merupakan suatu nilai rata-rata yang berpeluang terjadi pada variabel Y (variabel terikat) bila suatu nilai X (variabel bebas) diberikan. Koefisien regresi dapat dibedakan menjadi 2 macam, yaitu:

Metode yang biasa digunakan untuk menemukan persamaan yang akan digunakan untuk menghitung kekuatan beton yang akan dihasilkan. Persamaan yang akan digunakan ini merupakan hasil dari sebuah perhitungan estimasi yang dimana dengan menganalisis data tersebut untuk mendapatkan sebuah persamaan perhitungan kekuatan beton. Persamaan ini digunakan untuk memperhitungkan kekuatan beton yang nantinya akan dihasilkan dengan menggunakan algoritma linear regression dan beberapa teknik data mining lainnya.

Analisa regresi linier berganda merupakan pengembangan lanjut dari uraian diatas, khususnya pada kasus yang mempunyai lebih banyak peubah bebas dan parameter b. Hal ini sangat diperlukan dalam realita yang didefinisikan suatu proses untuk menunjuata guna lahan secara simultan ternyata mempengaruhi pergerakan.

Persamaan bentuk umum metode analisis regresi linier berganda

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + \dots + b_zX_z$$

Y = Kekuatan Beton

b₀ = Konstanta

X₁ = Semen

b₁...b_z = Koefisien regresi

X₂ = Biji Besi

X₃ = Abu Terbang

X₄ = Air

X₅ = Superplasticizer

X₆ = Batu Kerikil

X₇ = Pasir

X₈ = Umur

Melihat dari persamaan bentuk umum pada regresi linier berganda maka bentuk persamaaan untuk menghitung

kekuatan beton yang nantinya akan dihasilkan.

$$Y = b_0 + b_1 (\text{Semen}) + b_2 (\text{Biji Besi}) + b_3 (\text{Abu Terbang}) + b_4 (\text{Air}) + b_5 (\text{Superplasticizer}) + b_6 (\text{Batu Kerikil}) + b_7 (\text{Pasir}) + b_8 (\text{Umur})$$

Analisis regresi linier berganda adalah suatu metode statistic. Untuk menggunakannya, terdapat beberapa asumsi yang perlu diperhatikan :

- 1) Nilai peubah khususnya peubah bebas yaitu: Semen, Biji Besi, Abu Terbang, Air, Superplasticizer, Batu Kerikil, Pasir, Umur mempunyai nilai yang didapat dari hasil survey tanpa kesalahan berarti.
- 2) Peubah tidak bebas (Y) yaitu Kekuatan Beton harus mempunyai hubungan korelasi linier dengan peubah bebas (X). Jika hubungan tersebut tidak linier, transformasi linier harus dilakukan, meskipun batasan ini akan mempunyai implikasi lain dalam analisis residual.
- 3) Efek peubah bebas (X) pada peubah tidak bebas merupakan penjumlahan dan harus tidak ada korelasi yang kuat antara sesama peubah bebas.
- 4) Variansi peubah tidak bebas (Y) terhadap garis regresi harus sama untuk semua nilai peubah bebas (X)
- 5) Nilai peubah tidak bebas (Y) harus tersebar normal atau minimal mendekati normal.
- 6) Nilai peubah bebas (X) sebaiknya merupakan besaran yang relative mudah diproyeksi.

Dimana nilai variabel $b_0, b_1, b_2, \dots, b_8$ dapat dihitung dengan menggunakan metode kuadrat terkecil. Berikut adalah rumus persamaan umum untuk metode kuadrat terkecil :

$$\begin{aligned} b_0 + b_1 \sum X_1 + b_2 \sum X_2 + \dots + b_k \sum X_k &= \sum Y \\ b_0 \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 + \dots + b_k \sum X_1 X_k &= \sum X_1 Y \\ b_0 \sum X_2 + b_1 \sum X_2 X_1 + b_2 \sum X_2^2 + \dots + b_k \sum X_2 X_k &= \sum X_2 Y \\ \vdots & \vdots \\ b_0 \sum X_k + b_1 \sum X_k X_1 + b_2 \sum X_k X_2 + \dots + b_k \sum X_k^2 &= \sum X_k Y \end{aligned}$$

Untuk menghitung nilai $b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8$ sesuai dengan kasus yang dihadapi maka dapat dihitung dengan menggunakan analisis regresi linier berganda. Nilai $b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8$ bisa didapat dengan menggunakan Metode Kuadrat Terkecil seperti persamaan yaitu sebagai berikut:

- $b_0 n + b_1 \sum X_1 + b_2 \sum X_2 + b_3 \sum X_3 + b_4 \sum X_4 + b_5 \sum X_5 + b_6 \sum X_6 + b_7 \sum X_7 + b_8 \sum X_8 = \sum Y$
- $b_0 \sum X_1 + b_1 \sum (X_1)^2 + b_2 \sum (X_1 X_2) + b_3 \sum (X_1 X_3) + b_4 \sum (X_1 X_4) + b_5 \sum (X_1 X_5) + b_6 \sum (X_1 X_6) + b_7 \sum (X_1 X_7) + b_8 \sum (X_1 X_8) = \sum X_1 Y$
- $b_0 \sum X_2 + b_1 \sum (X_1 X_2) + b_2 \sum (X_2)^2 + b_3 \sum (X_2 X_3) + b_4 \sum (X_2 X_4) + b_5 \sum (X_2 X_5) + b_6 \sum (X_2 X_6) + b_7 \sum (X_2 X_7) + b_8 \sum (X_2 X_8) = \sum X_2 Y$
- $b_0 \sum X_3 + b_1 \sum (X_1 X_3) + b_2 \sum (X_2 X_3) + b_3 \sum (X_3)^2 + b_4 \sum (X_3 X_4) + b_5 \sum (X_3 X_5) + b_6 \sum (X_3 X_6) + b_7 \sum (X_3 X_7) + b_8 \sum (X_3 X_8) = \sum X_3 Y$
- $b_0 \sum X_4 + b_1 \sum (X_1 X_4) + b_2 \sum (X_2 X_4) + b_3 \sum (X_3 X_4) + b_4 \sum (X_4)^2 + b_5 \sum (X_4 X_5) + b_6 \sum (X_4 X_6) + b_7 \sum (X_4 X_7) + b_8 \sum (X_4 X_8) = \sum X_4 Y$
- $b_0 \sum X_5 + b_1 \sum (X_1 X_5) + b_2 \sum (X_2 X_5) + b_3 \sum (X_3 X_5) + b_4 \sum (X_4 X_5) + b_5 \sum (X_5)^2 + b_6 \sum (X_5 X_6) + b_7 \sum (X_5 X_7) + b_8 \sum (X_5 X_8) = \sum X_5 Y$
- $b_0 \sum X_6 + b_1 \sum (X_1 X_6) + b_2 \sum (X_2 X_6) + b_3 \sum (X_3 X_6) + b_4 \sum (X_4 X_6) + b_5 \sum (X_5 X_6) + b_6 \sum (X_6)^2 + b_7 \sum (X_6 X_7) + b_8 \sum (X_6 X_8) = \sum X_6 Y$
- $b_0 \sum X_7 + b_1 \sum (X_1 X_7) + b_2 \sum (X_2 X_7) + b_3 \sum (X_3 X_7) + b_4 \sum (X_4 X_7) + b_5 \sum (X_5 X_7) + b_6 \sum (X_6 X_7) + b_7 \sum (X_7)^2 + b_8 \sum (X_6 X_8) = \sum X_7 Y$
- $b_0 \sum X_8 + b_1 \sum (X_1 X_8) + b_2 \sum (X_2 X_8) + b_3 \sum (X_3 X_8) + b_4 \sum (X_4 X_8) + b_5 \sum (X_5 X_8) + b_6 \sum (X_6 X_8) + b_7 \sum (X_7 X_8) + b_8 \sum (X_8)^2 = \sum X_8 Y$

Kemudian setelah hasil dari invers ini diketahui maka hasilnya dikalikan dengan $\sum Y, \sum X_1 Y, \sum X_2 Y, \sum X_3 Y, \sum X_4 Y, \sum X_5 Y, \sum X_6 Y, \sum X_7 Y, \sum X_8 Y$ sehingga hasil dari perkalian tersebut adalah sebagai berikut :

N	$\sum X_1$	$\sum X_2$	$\sum X_3$	$\sum X_4$	$\sum X_5$	$\sum X_6$	$\sum X_7$	$\sum X_8$
$\sum X_1$	$\sum X_1^2$	$\sum X_1 X_2$	$\sum X_1 X_3$	$\sum X_1 X_4$	$\sum X_1 X_5$	$\sum X_1 X_6$	$\sum X_1 X_7$	$\sum X_1 X_8$
$\sum X_2$	$\sum X_1 X_2$	$\sum X_2^2$	$\sum X_2 X_3$	$\sum X_2 X_4$	$\sum X_2 X_5$	$\sum X_2 X_6$	$\sum X_2 X_7$	$\sum X_2 X_8$
$\sum X_3$	$\sum X_1 X_3$	$\sum X_2 X_3$	$\sum X_3^2$	$\sum X_3 X_4$	$\sum X_3 X_5$	$\sum X_3 X_6$	$\sum X_3 X_7$	$\sum X_3 X_8$
$\sum X_4$	$\sum X_1 X_4$	$\sum X_2 X_4$	$\sum X_3 X_4$	$\sum X_4^2$	$\sum X_4 X_5$	$\sum X_4 X_6$	$\sum X_4 X_7$	$\sum X_4 X_8$
$\sum X_5$	$\sum X_1 X_5$	$\sum X_2 X_5$	$\sum X_3 X_5$	$\sum X_4 X_5$	$\sum X_5^2$	$\sum X_5 X_6$	$\sum X_5 X_7$	$\sum X_5 X_8$
$\sum X_6$	$\sum X_1 X_6$	$\sum X_2 X_6$	$\sum X_3 X_6$	$\sum X_4 X_6$	$\sum X_5 X_6$	$\sum X_6^2$	$\sum X_6 X_7$	$\sum X_6 X_8$
$\sum X_7$	$\sum X_1 X_7$	$\sum X_2 X_7$	$\sum X_3 X_7$	$\sum X_4 X_7$	$\sum X_5 X_7$	$\sum X_6 X_7$	$\sum X_7^2$	$\sum X_7 X_8$
$\sum X_8$	$\sum X_1 X_8$	$\sum X_2 X_8$	$\sum X_3 X_8$	$\sum X_4 X_8$	$\sum X_5 X_8$	$\sum X_6 X_8$	$\sum X_7 X_8$	$\sum X_8^2$

$$X \begin{pmatrix} \Sigma Y^2 \\ \Sigma X_1 Y \\ \Sigma X_2 Y \\ \Sigma X_3 Y \\ \Sigma X_4 Y \\ \Sigma X_5 Y \\ \Sigma X_6 Y \\ \Sigma X_7 Y \\ \Sigma X_8 Y \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \end{pmatrix}$$

Sehingga setelah nilai dari variabel $b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8$ maka persamaan regresi linier bergandanya untuk menghitung kekuatan beton yang akan dihasilkan yaitu dengan menggunakan rumus sebagai berikut :

$$Y = -23,16375581 + 0,119785255(X_1) + 0,103847249(X_2) + 0,087943082(X_3) + (-0,150297904)(X_4) + 0,291(X_5) + 0,018030184(X_6) + 0,020154456(X_7) + 0,11422562(X_8)$$

Dimana :

Y = Concrete compressive strength(MPa, megapascals)

X1 = Semen (component 1)(kg in a m³ mixture)

X2 = Biji Besi (component 2)(kg in a m³ mixture)

X3 = Abu Terbang (component 3)(kg in a m³ mixture)

X4 = Air (component 4)(kg in a m³ mixture)

X5 = Superplasticizer (component 5)(kg in a m³ mixture)

X6 = Batu Kerikil (component 6)(kg in a m³ mixture)

X7 = Pasir (component 7)(kg in a m³ mixture)

X8 = Umur (day)

Untuk mengetahui seberapa besar tingkat persentase yang dihasilkan persamaan yang sudah kita temukan yaitu maka akan lebih kompleks ada beberapa hal harus dipertimbangkan, seperti :

a. Koefisien korelasi : koefisien korelasi ini digunakan untuk menentukan korelasi antara peubah tidak bebas dengan peubah bebas atau sesama peubah bebas. Koefisien korelasi ini dapat dihitung dengan berbagai cara yang salah satunya adalah sebagai berikut :

$$r = \frac{N \sum_{i=1}^N (X_i Y_i) - \sum_{i=1}^N (X_i) \cdot \sum_{i=1}^N (Y_i)}{\sqrt{[N \sum_{i=1}^N (X_i^2) - (\sum_{i=1}^N (X_i))^2][N \sum_{i=1}^N (Y_i^2) - (\sum_{i=1}^N (Y_i))^2]}}$$

Sehingga untuk menyesuaikan dengan kasus pada perhitungan kekuatan beton yang dihasilkan maka rumus untuk setiap korelasi Nilai $r = 1$ berarti bahwa korelasi antara peubah y dan x adalah positif (meningkatnya nilai x akan mengakibatkan meningkatnya nilai y). sebaliknya, jika $r = -1$, berarti korelasi antara peubah y dan x adalah negative (meningkatnya nilai x akan mengakibatkan menurunnya nilai y). Nilai $r = 0$ menyatakan tidak ada korelasi antarpeubah.

b. Koefisien determinasi : Jika koefisien korelasi berganda dikuadratkan, diperoleh koefisien determinasi berganda yang disimbolkan dengan R^2 . Dimana koefisien determinasi digunakan untuk mengukur besarnya sumbangan dari beberapa variabel X ((Semen (component 1)(kg in a m³ mixture), Biji Besi (component 2)(kg in a m³ mixture), Biji Besi (component 2)(kg in a m³ mixture), Air (component 4)(kg in a m³ mixture), Superplasticizer (component 5)(kg in a m³ mixture), Batu Kerikil (component 6)(kg in a m³ mixture), Pasir (component 7)(kg in a m³ mixture), Umur (day) terhadap naik turunnya tingkat kekuatan beton yang akan dihasilkan (Y) atau Concrete compressive strength(MPa, megapascals). Jika nilai koefisien determinasi dikalikan 100% diperoleh persentase sumbangan variabel X terhadap naik turunnya variabel Y atau kekuatan beton itu sendiri.

$$\overline{R^2} = \frac{[R^2 - \frac{K}{N-1}]}{[\frac{(N-1)}{(N-K-1)}]}$$

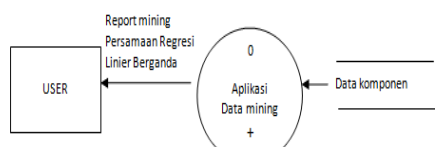
N adalah ukuran sampel dan K adalah jumlah peubah.

c. Mencari hasil dari koefisien determinasi maka bisa juga menggunakan hasil dari koefisien korelasi kemudian nilai dari koefisien korelasi tersebut kita akar 2 atau (R^2)
d. Uji t-tes : Uji t-tes dapat digunakan untuk 2 tujuan yaitu, untuk menguji

signifikansi nilai koefisien korelasi (r) dan untuk menguji signifikansi nilai koefisien regresi. Setiap peubah (X) yang mempunyai koefisien regresi yang tidak signifikan secara statistic harus dibuang dari model.

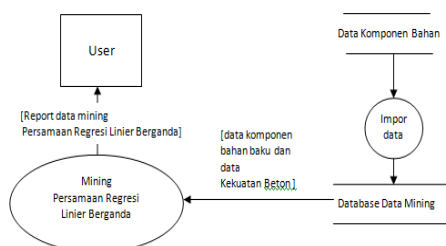
3.1.1 Perancangan dan Analisa Sistem

Pemodelan fungsi digambarkan dengan DCD (Data Context Diagram), DFD (Data Flow Diagram) dan kamus data (Data Dictionary).



Gambar 3.1 DFD level-0

Gambar 3.1 merupakan DCD / DFD level-0 pada Aplikasi Data Mining yang terdiri dari 1 input dan 1 output. External entity berupa pengguna atau user dan 1 filedata yaitu data komponen.



Gambar 3.2 DFD Level-1

Gambar 4.2 merupakan DFD Level-1 dari Aplikasi Data Mining yang dipecah menjadi beberapa proses kecil guna menjelaskan fungsi-fungsi dan arus data yang mengalir pada Aplikasi Data Mining. Berikut proses-proses yang terdapat pada Aplikasi Data Mining :

1. Import Data

Proses import data adalah proses load data dari data rekam medis mahasiswa ke database data mining. Dalam hal ini data rekam medis yang akan diimport sudah di cleaning, integrasi, selection dan di transformasi.

2. Mining data komponen bahan baku

Proses mining komponen bahan baku ini merupakan analisis untuk menemukan persamaan yang akan digunakan untuk menghitung kekuatan beton yang nantinya akan dihasilkan

3.1.2 Desain Database

Tabel 4.7 : Tabel dt_mining

Field Name	Datatype	Len
x_1	varchar	10
x_2	varchar	10
x_3	varchar	10
x_4	varchar	10
x_5	varchar	10
x_6	varchar	10
x_7	varchar	10
x_8	varchar	10
y_1	varchar	10
y	varchar	10
* n	varchar	10

Tabel 4.8 : Tabel Bantu

x1x2	varchar	20
x1x3	varchar	11
* x1x4	varchar	20
x1x5	varchar	20
x1x6	varchar	20
x1x7	varchar	20
x1x8	varchar	20
x1y	varchar	15
x2x3	varchar	15
x2x4	varchar	15
x2x5	varchar	15
x2x6	varchar	15
x2x7	varchar	15
x2x8	varchar	15
x2y	varchar	15
x3x4	varchar	15
x3x5	varchar	15
x3x6	varchar	15

3.2 Implementasi

3.2.1 Tampilan Menu Import

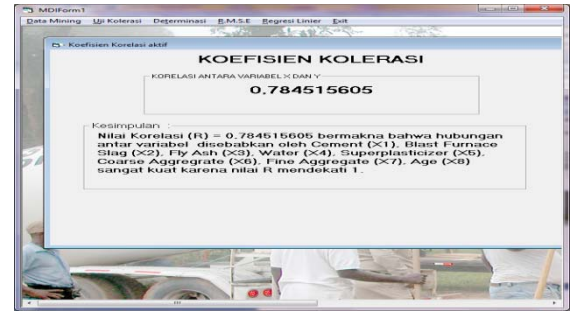


Gambar 4.1 Menu import

Pada proses ini adalah pengambilan data yang akan dilakukan proses data mining dan dimana data ini merupakan data komponen yang digunakan pada pembuatan beton yang sudah pernah dilakukan.

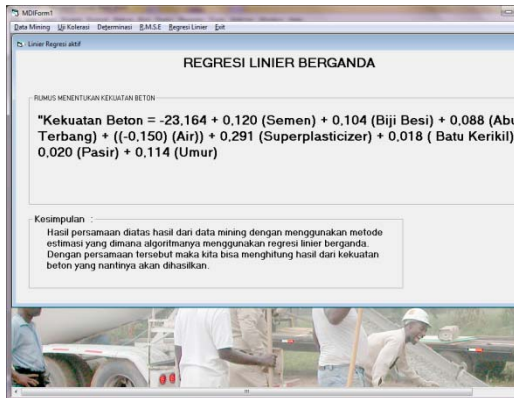
Cement (cc)	Blast Furnace	Fly Ash (cc)	Water (cm)	Superplastic	Coarse Aggr.	Fine Aggr.	Age (day)	Concrete
446	24	79	162	11.61	967	712	56	61.96689
446	24	79	162	11.64	967	712	56	56.14493
446	24	79	162	11.64	967	712	56	55.26460
446	24	79	162	10.3	967	712	56	54.76507
387	20	94	157	14.32	938	845	28	59.29225
387	20	94	157	13.93	938	845	28	46.68441
387	20	94	157	11.61	938	845	28	46.68441
387	20	94	157	14.32	938	845	3	22.75270
387	20	94	157	13.93	938	845	3	25.51061
387	20	94	157	11.61	938	845	3	34.71027
387	20	94	157	14.32	938	845	7	36.83070
387	20	94	157	13.93	938	845	7	45.89841
387	20	94	157	11.61	938	845	7	41.65503
387	20	94	157	14.32	938	845	56	56.33708
387	20	94	157	13.93	938	845	56	47.96884
387	20	94	157	11.61	938	845	56	61.65989
355	19	97	145	13.13	967	871	28	44.02993
355	19	97	145	12.25	967	871	28	55.45455
491	26	123	210	3.93	882	699	28	55.61106
491	26	123	210	3.93	822	699	28	57.91596
491	26	123	210	3.93	882	699	3	25.60910
491	26	123	210	3.93	882	699	7	31.48893

Gambar 4.2 Data berhasil import



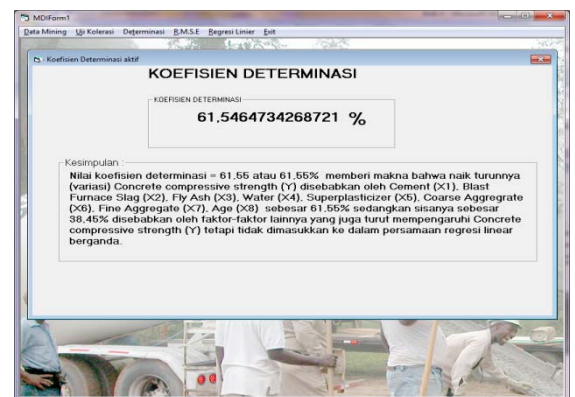
Gambar 4.5 menu koefisien korelasi

Koefisien korelasi merupakan hasil hubungan semua variabel apakah memiliki hubungan yang bagus atau tidak dan semakin mendekati 1 maka akan semakin bagus hasilnya.



Gambar 4.3 menu data mining

Pada proses data mining ini merupakan hasil dari metode linear regression yang bisa digunakan untuk mengetahui berapa besar tingkat kekuatan beton yang dihasilkan.



Gambar 4.6 menu determinasi

Pada determinasi ini akan mengetahui seberapa besar pengaruh yang akan dihasilkan dari variabel bebas dengan variabel tak bebas.

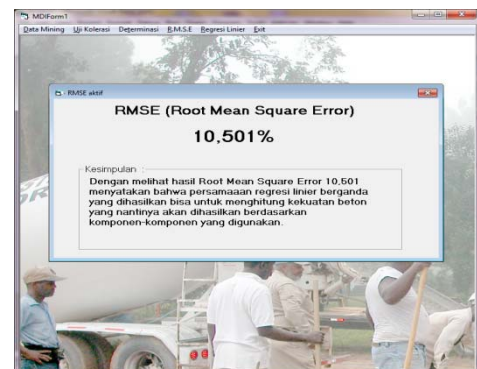
KOLERASI DATA MINING

KORELASI DARI VARIABEL X:

Semen dan Biji Besi	0.279770089426	Semen dan Batu Kerikil	0.10207708807867
Semen dan Abu Terbang	4.3374303118729E	Semen dan Pasir	0.22044838836437
Semen dan Air	0.1441173762537	Semen dan Umur	0.194990604112E
Semen dan Superplasticizer	0.2862720889393E		
Semen dan Kekuatan Beton	0.497926797469E		

Gambar 4.4 menu korelasi

Pada korelasi ini untuk mengetahui hubungan antara setiap variabel dengan variabel bebas dan juga variabel tak bebas dan juga sesama variabel tak bebas.



Gambar 4.7 RMSE

RMSE merupakan salah satu cara mengetahui kesalahan atau error pada persamaan yang sudah dihasilkan dari proses data mining ini dengan metode linier regresi tersebut.

Gambar 4.8 Halaman Linier Regresi

Perhitungan kekuatan beton yang dihasilkan bisa dihitung dengan aplikasi ini sehingga nanti akan mendapatkan hasil yang lebih akurat dan efisien.

4.1 Pemeliharaan Sistem/Perawatan

1. Selama masa pemakaian aplikasi data mining ini, perawatan sistem sangat dibutuhkan dikarenakan beberapa alasan, yaitu:
2. 1. Menjaga keterbaruan aplikasi
3. Dengan berjalannya waktu, pastinya akan membuat aplikasi menjadi usang dan mengharuskan untuk dilakukannya modifikasi baik dari segi perancangan maupun segi aplikasinya.
4. 2. Meningkatkan aplikasi
5. Perkembangan teknologi perangkat keras maupun perangkat lunak yang sangat pesat menuntut adanya peningkatan perangkat keras dan perangkat lunak agar kinerja aplikasi semakin lama akan semakin bassssik dan meningkat.

5 KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dengan adanya aplikasi perhitungan kekuatan beton yang nantinya dihasilkan saat pembangunan maka akan memudahkan 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.

5.2 Saran

Proses penelitian ini mendapatkan banyak hambatan seperti terbatasnya data penelitian dan perangkat keras yang digunakan, untuk penelitian selanjutnya terdapat beberapa saran sebagai berikut :

1. Untuk penelitian selanjutnya dapat dikembangkan dengan metode Support Vector Machine atau dengan metode Nueral Network.
2. Penelitian ini digunakan untuk menyelesaikan kasus perhitungan kekuatan beton yang dihasilkan berdasarkan komponen yang digunakan, dalam penelitian selanjutnya dapat diaplikasikan pada data set yang berbeda seperti data perbankan untuk menganalisa promosi dan harga apakah berpengaruh terhadap konsumen membeli produk tersebut dan lain-lain.

Daftar Pustaka

1. Kusriani, dan Luthfy, E. T. 2009. "Algoritma Data Mining". Yogyakarta: Andi Publishing.
2. Santosa, B. 2007. "Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis". Yogyakarta: Graha Ilmu.
3. Larose, D.T. 2005. "Data Mining Methods And Models". Canada: John Wiley & Sons, Inc.
4. Widodo, P.B. et al. 2013. "Penerapan Data Mining Dengan Visual Basic 6". Bandung: Penerbit Rekayasa Sains.
5. Prasetyo, E. 2012. "Data Mining Konsep dan Aplikasi Menggunakan Visual Basic 6". Yogyakarta: Penerbit Andi.
6. Larose, D. T. (2005). Discovering Knowledge in Data. New Jersey: John Wiley & Sons, Inc.
7. Gorunescu, F. (2011). Data Mining Concepts, Models And Techniques. Verlag Berlin Heidelberg: Springer.

8. Widarjono, A. 2007. Analisis Regresi Terapan.: Ekonisia Fakultas Ekonomi Universitas Islam Indonesia.