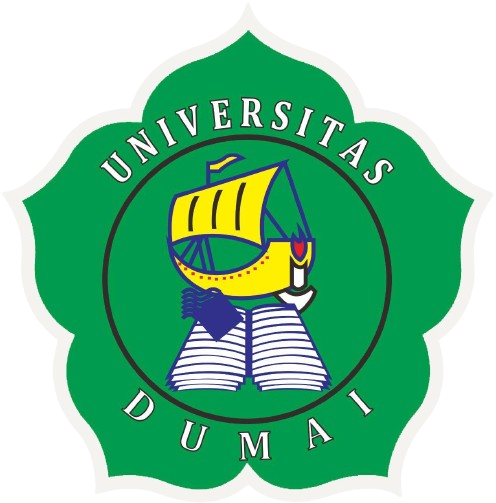
**KARYA TULIS ILMIAH**

**REGRESSION SEBAGAI METODE DATA MINING**

Dosen Pembimbing : Sukri Adrianto,M.Kom



**Disusun Oleh :**

**Anita Carlolina 2213002**

**Nur Afiqah 2213008**

**Desrita Romauli Sinambela 2213017**

**Novi Dianti 2213028**

**Program Studi Sistem Informasi**

**Fakultas Ilmu Komputer**

**Universitas Dumai**

**TP. 2024**

**UNIVERSITAS DUMAI**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI SISTEM INFORMASI (SI)**

**Lembar Pengesahan Karya Tulis Ilmiah (KTI)**

**Tanda Persetujuan Diberikan Kepada**

**Kelompok 3:**

**1. Novi Dianti 2213028**

**2. Anita Carlolina 2213002**

**3. Nur Afiqah 2213008**

**4. Desrita Romauli Sinambela 2213017**

**Metode Regression (KTI)**

**Disetujui untuk menjadi sebuah Karya Tulis Ilmiah (KTI) Fakultas Ilmu Komputer Program Studi Sistem Informasi Universitas Dumai**

**Menyetujui,**

|  |  |
| --- | --- |
| **Pembimbing** | **Ketua Kelompok** |
| **Sukri Adrianto, M. Kom** | **Novi Dianti** |
|  |  |
|  |  |
|  |  |

**KATA PENGANTAR**

Puji syukur kami panjatkan kehadirat Tuhan Yang Maha Esa atas segala rahmat, hidayah, dan karunia-Nya yang telah melimpahkan berkahnya kepada kami sehingga dapat menyelesaikan karya tulis ilmiah ini dengan judul "Regression Sebagai Metode Data Mining". Dokumen ini disusun sebagai salah satu tugas mata kuliah Data Mining yang kami jalani di bawah bimbingan dan dukungan Sukri Adrianto, M.Kom, dosen yang berdedikasi dalam memandu kami dalam pemahaman konsep-konsep penting terkait infrastruktur sistem terintegrasi.

Akhir kata, kami memohon maaf apabila terdapat kekurangan dalam penyusunan dokumen ini. Kami berharap karya tulis ilmiah ini dapat memberikan wawasan yang bermanfaat bagi pembaca.

Penulis

**DAFTAR ISI**

**LEMBAR PENGESAHAN ii**

**KATA PENGANTAR iii**

**DAFTAR ISI iv**

**BAB I PENDAHULUAN 1**

1.1 Latar Belakang 1

1.2 Rumusan Masalah 1

1.3 Tujuan 1

**BAB II PEMBAHASAN 2**

* 1. Regression 2
  2. Jenis-Jenis Dari Metode Regression 2
  3. Langkah-langkah dalam Regression sebagai Metode Data Mining 2
  4. Aplikasi Regresi dalam Data Mining 3
  5. Analisis Cloud Computing Untuk Penyimpanan Dokumen Terhadap Proses Pembelajaran Menggunakan Algoritma Regresi Linear Berganda (STUDI KASUS) 3

**BAB III PENUTUP 6**

3.1 Kesimpulan 6

3.2 Saran 6

**BAB IV DISKUSI KELOMPOK 8**

**DAFTAR PUSTAKA 12**

**BAB I**

**PENDAHULUAN**

1. **Latar Belakang Masalah**

Data mining adalah proses ekstraksi pola dan pengetahuan yang bermanfaat dari data yang beragam dan kompleks. Dalam konteks ini, regresi adalah salah satu metode yang paling umum digunakan dalam data mining. Regresi memungkinkan kita untuk memodelkan hubungan antara variabel input dan output, serta membuat prediksi berdasarkan pola yang ditemukan dalam data.

Dengan demikian, pemahaman yang mendalam tentang regresi sebagai metode data mining sangat penting dalam konteks analisis data modern. Ini tidak hanya membantu organisasi dan perusahaan dalam membuat keputusan yang lebih baik berdasarkan data, tetapi juga membuka pintu untuk penemuan pengetahuan baru yang dapat meningkatkan kinerja dan efisiensi.

1. **Rumusan Masalah**
   * + 1. Apa yang dimaksud dengan Regression ?
       2. Apa jenis-jenis dari metode Regression ?
       3. Bagaimana langkah-langkah dalam Regresi sebagai Metode Data Mining ?
       4. Apa saja aplikasi Regresi dalam Data Mining?
       5. Bagaimana Menganalisis Cloud Computing Untuk Penyimpanan Dokumen Terhadap Proses Pembelajaran Menggunakan Algoritma Regresi Linear Berganda (STUDI KASUS) ?
2. **Tujuan**
3. Untuk mengetahui apa itu Regression.
4. Untuk mengetahui dan memahami jenis-jenis dari metode Regression.
5. Untuk mengidentifikasi dan memahami langkah-langkah dalam Regresi sebagai Metode Data Mining .
6. Untuk mengetahui aplikasi yang ada dalam metode regresi data mining

**BAB II**

**PEMBAHASAN**

1. **Regression**

Regression adalah metode yang bertujuan mencari pola dan menentukan sebuah nilai numerik. Linear regression dan logistic regression merupakan teknik yang populer digunakan dalam metode regression. Linear regression adalah analisis regresi dengan menggunakan satu atau dua variabel independen untuk memprediksi sebuah nilai. Sedangkan logistic regression adalah analisis regresi yang cukup rumit dan biasanya digunakan untuk menentukan probabilitas suatu kejadian. Regression sering digunakan dalam memecahkan permasalahan di dunia bisnis seperti menentukan metode distribusi, kapasitas distribusi,dan memprediksi musim (kecepatan angin, kelembaban, dan tekanan udara).

1. **Jenis-jenis dari metode Regression**

Terdapat beberapa jenis regresi yang umum digunakan dalam data mining, antara lain:

1. Regresi Linear: Regresi linear adalah jenis regresi yang digunakan ketika hubungan antara variabel independen dan dependen diasumsikan bersifat linier. Contohnya adalah regresi linier sederhana dan regresi linier berganda.
2. Regresi Logistik: Regresi logistik digunakan ketika variabel dependen adalah biner (dua kategori) dan hubungan antara variabel independen dan dependen diestimasi menggunakan fungsi logistik.
3. Regresi Non-linear: Regresi non-linear digunakan ketika hubungan antara variabel independen dan dependen tidak bersifat linier. Model regresi non-linear dapat berupa polinomial, eksponensial, atau jenis fungsi lainnya.
4. **Langkah-langkah dalam Regression sebagai Metode Data Mining**

Proses regresi dalam data mining melibatkan beberapa langkah, di antaranya:

1. Pemilihan Variabel: Langkah pertama adalah memilih variabel-variabel yang akan digunakan dalam analisis regresi. Variabel independen dipilih berdasarkan relevansinya dengan variabel dependen.
2. Pengumpulan Data: Data yang diperlukan untuk analisis regresi dikumpulkan dari sumber yang tersedia. Data harus lengkap dan representatif untuk mencapai hasil yang akurat.
3. Pembersihan Data: Data sering kali mengandung noise, outlier, atau missing value yang perlu dibersihkan sebelum proses analisis. Pembersihan data melibatkan identifikasi dan penanganan masalah-masalah tersebut.
4. Pemodelan Regresi: Proses ini melibatkan pemilihan model regresi yang sesuai dan estimasi parameter-parameter model menggunakan teknik seperti metode kuadrat terkecil.
5. Evaluasi Model: Setelah model regresi dibangun, langkah selanjutnya adalah mengevaluasi kinerja model tersebut. Evaluasi dilakukan dengan menggunakan metrik-metrik seperti RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), atau R-squared.
6. Validasi Model: Validasi model dilakukan untuk memastikan bahwa model yang dibangun dapat digeneralisasi dengan baik ke data baru yang tidak terlihat. Metode validasi yang umum digunakan antara lain validasi silang (cross-validation) dan pembagian data menjadi data latih dan data uji.
7. **Aplikasi Regresi dalam Data Mining**

Regresi memiliki berbagai aplikasi dalam data mining, termasuk:

1. Pengembangan Prediksi: Regresi digunakan untuk memprediksi nilai variabel dependen berdasarkan variabel independen yang tersedia. Contohnya termasuk prediksi harga saham, penjualan produk, atau harga rumah.
2. Analisis Penyebab: Regresi dapat digunakan untuk mengevaluasi hubungan sebab-akibat antara variabel-variabel. Misalnya, regresi dapat digunakan untuk mengidentifikasi faktor-faktor yang mempengaruhi kinerja siswa di sekolah.
3. Optimasi Proses: Dalam konteks bisnis, regresi digunakan untuk mengoptimalkan proses-proses seperti rantai pasokan, produksi, atau pelayanan pelanggan dengan memahami faktor-faktor yang memengaruhi kinerja.
4. **Analisis Cloud Computing Untuk Penyimpanan Dokumen Terhadap Proses Pembelajaran Menggunakan Algoritma Regresi Linear Berganda (STUDI KASUS)**

Penelitian ini membahas pemanfaatan teknologi cloud computing dalam proses pembelajaran di Sekolah Chandra Kusuma Jakarta Utara. Namun, kendala muncul karena penyimpanan dokumen rendah dan akses internet lambat karena keterbatasan perangkat komputer server. Penelitian menggunakan algoritma regresi linier berganda untuk mengevaluasi hubungan antara variabel cloud computing (X1) dan proses pembelajaran (X2) terhadap variabel penyimpanan dokumen (Y).

Penelitian ini menggunakan algoritma regresi linear berganda untuk menganalisis hubungan antara variabel independen X1 dan X2 dengan variabel dependen Y. Data dikumpulkan melalui kuesioner dengan skala Likert dan diuji validitas serta reliabilitas. Populasi penelitian terdiri dari 57 responden, termasuk guru dan siswa SMA Chandra Kusuma Jakarta Utara. Analisis regresi dilakukan untuk menentukan hubungan antara variabel-variabel tersebut serta koefisien determinasi digunakan untuk melihat seberapa besar hubungan secara simultan. Pengujian hipotesis dilakukan secara simultan (uji F) dan parsial (uji t) untuk menguji signifikansi hubungan antara variabel.

Hasil dan pembahasan :

1. Data Responden: Penelitian melibatkan 57 responden yang terdiri dari guru dan siswa dari beberapa kelas di SMA Chandra Kusuma Jakarta Utara. Mayoritas responden adalah siswa kelas X.
2. Uji Validitas: Dilakukan uji validitas untuk masing-masing variabel, yaitu cloud computing (X1), penyimpanan dokumen (Y), dan proses pembelajaran (X2). Semua item pertanyaan dari ketiga variabel tersebut dinyatakan valid.
3. Uji Reliabilitas: Dilakukan uji reliabilitas untuk mengukur konsistensi jawaban dari responden terhadap item pertanyaan. Semua variabel menunjukkan nilai reliabilitas yang memenuhi persyaratan.
4. Analisis Regresi Linear Berganda: Dilakukan analisis regresi linear berganda untuk mengetahui hubungan antara variabel independen (X1 dan X2) dengan variabel dependen (Y). Hasilnya menunjukkan bahwa kedua variabel independen memiliki pengaruh positif terhadap variabel dependen.
5. Koefisien Korelasi Ganda Antar Variabel: Dilakukan perhitungan koefisien korelasi ganda antar variabel. Hasilnya menunjukkan adanya hubungan yang signifikan antara variabel X1, X2, dan Y.
6. Analisis Korelasi dan Koefisien Determinasi: Dilakukan analisis untuk mengetahui besarnya koefisien determinasi (R^2) untuk setiap prediktor yang digunakan. Hasilnya menunjukkan bahwa variabel independen (X1 dan X2) menjelaskan sekitar 36,3% variasi dalam variabel dependen (Y).
7. Hasil Uji Hipotesis: Dilakukan uji-F (uji simultan) dan uji-t (uji parsial) untuk menguji signifikansi hubungan antara variabel-variabel tersebut. Hasilnya menunjukkan bahwa secara simultan maupun secara parsial, variabel X1 dan X2 memiliki pengaruh yang signifikan terhadap variabel Y.
8. Dengan demikian, hasil analisis ini memberikan pemahaman yang lebih dalam tentang hubungan antara cloud computing, penyimpanan dokumen, dan proses pembelajaran dalam konteks penelitian yang dilakukan.

Kesimpulan:

Dari hasil pembahasan yang telah diuraikan diatas, maka kesimpulan disesuaikan dengan penetapan tujuan penelitian ini yaitu :

1. Berdasarkan penelitian secara simultan, hasil uji F sebesar 15,387 dengan menggunakan taraf signifikan 5%, dan nilai koefiesien determinasi 36,30%, sedangkan sisanya 63,70% yang ditentukan oleh variabel lain. Sehingga diperoleh kesimpulan bahwa terdapat hubungan yang signifikan antara cloud computing untuk penyimpanan dokumen terhadap proses pembelajaran.

2. Hasil penelitian secara parsial cloud computing untuk penyimpanan dokumen mempunyai hubungan yang signifikan karena uji t dengan taraf signifikan 5%, mempunyai nilai t hitung3,211 dan lebih besar dari t tabel.

3. Hasil penelitian secara parsial penyimpanan dokumen terhadap proses pembelajaran mempunyai hubungan yang signifikan karena uji t dengan taraf signifikan 5%, mempunyai nilai t hitung 3,824 dan lebih besar dari t tabel.

**BAB III**

**PENUTUP**

* 1. **KESIMPULAN**

Regresi adalah salah satu metode yang digunakan untuk mencari pola dan memprediksi nilai numerik berdasarkan hubungan antara variabel independen dan dependen. Jenis-jenis regresi yang umum digunakan termasuk regresi linear, regresi logistik, dan regresi non-linear.

Proses regresi dalam data mining melibatkan langkah-langkah seperti pemilihan variabel, pengumpulan data, pembersihan data, pemodelan regresi, evaluasi model, dan validasi model. Aplikasi regresi dalam data mining mencakup pengembangan prediksi, analisis penyebab, dan optimasi proses dalam berbagai konteks bisnis.

1. Hubungan antara Variabel: Penelitian menunjukkan adanya hubungan yang signifikan antara penggunaan teknologi cloud computing untuk penyimpanan dokumen dan proses pembelajaran di Sekolah Chandra Kusuma Jakarta Utara.
2. Pengaruh Variabel Independen: Variabel independen, yaitu cloud computing untuk penyimpanan dokumen dan proses pembelajaran, memiliki pengaruh positif terhadap variabel dependen, yaitu proses pembelajaran.
3. Koefisien Determinasi: Koefisien determinasi sebesar 36,30% menunjukkan bahwa sebagian besar variasi dalam proses pembelajaran dapat dijelaskan oleh variabel independen yang diteliti, sementara sisanya dipengaruhi oleh faktor-faktor lain yang tidak dimasukkan dalam penelitian ini.
   1. **SARAN**
4. Optimasi Penggunaan Cloud Computing: Sekolah Chandra Kusuma Jakarta Utara dapat mempertimbangkan untuk lebih mengoptimalkan penggunaan teknologi cloud computing dalam proses pembelajaran guna meningkatkan efisiensi dan efektivitas penyimpanan dokumen serta memfasilitasi akses yang lebih cepat dan mudah bagi guru dan siswa.
5. Peningkatan Pengetahuan dan Keterampilan: Penting bagi pihak sekolah untuk memberikan pelatihan dan pembekalan kepada guru dan siswa terkait penggunaan teknologi cloud computing agar mereka dapat memanfaatkannya secara maksimal dalam mendukung proses pembelajaran.
6. Pemantauan dan Evaluasi: Sekolah perlu melakukan pemantauan secara berkala terhadap penggunaan teknologi cloud computing dan dampaknya terhadap proses pembelajaran. Evaluasi rutin dapat membantu dalam mengidentifikasi area-area yang perlu ditingkatkan atau dioptimalkan lebih lanjut.
7. Penelitian Lanjutan: Dianjurkan untuk melakukan penelitian lanjutan yang lebih mendalam, mungkin dengan melibatkan lebih banyak variabel atau populasi yang lebih luas, guna memperoleh pemahaman yang lebih komprehensif tentang pengaruh teknologi cloud computing dalam konteks pendidikan.

**BAB IV**

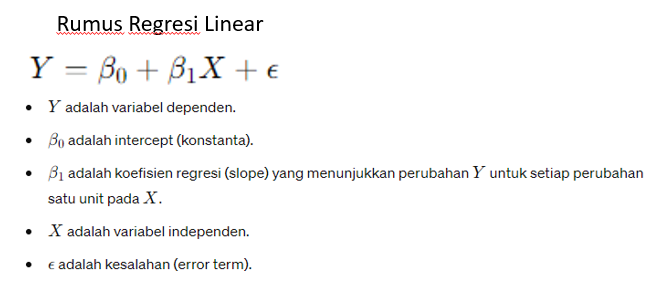
**DISKUSI KELOMPOK**

**Pertanyaan Kelompok:**

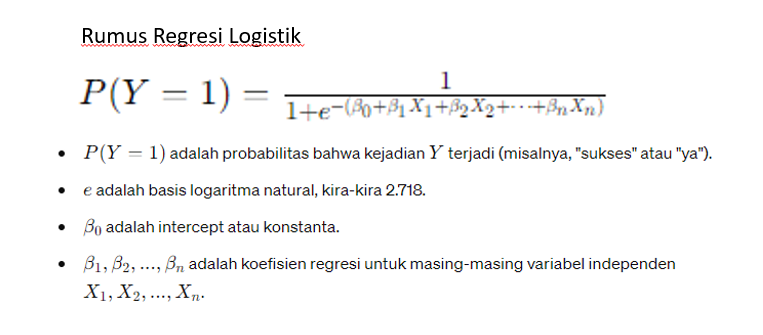
Dari Anistari (Kelompok 2)

Pertanyaan : Berikan contoh dari regresi linier dan regresi logistik beserta rumus !

Dijawab oleh Novi Dianti :



Contohnya memprediksi harga rumah berdasarkan berbagai faktor seperti luas bangunan, jumlah kamar tidur, lokasi, dan usia bangunan. Misalnya, dengan menggunakan data historis, kita bisa membuat model regresi linier yang membantu memprediksi harga rumah baru berdasarkan luas dan jumlah kamar tidurnya.



Contohnya memprediksi penyakit, Dokter atau peneliti kesehatan menggunakan regresi logistik untuk memprediksi kemungkinan seorang pasien menderita penyakit tertentu berdasarkan faktor-faktor seperti usia, jenis kelamin, indeks massa tubuh (BMI), tekanan darah, dan hasil tes laboratorium lainnya. Misalnya, regresi logistik dapat digunakan untuk memprediksi apakah seseorang berisiko tinggi menderita diabetes berdasarkan data medisnya.

Adinda Mayang Zakia (Kelompok 5)

Pertanyaan: Apa variabel independen dan dependen?

Dijawab oleh Destita Romauli Sinambela :

* Variabel independen (variabel bebas) merupakan variabel yang dipengaruhi atau menjadi sebab perubahannya atau timbulnya variabel dependen (terikat).
* Variabel dependen (variabel terikat) merupakan variabel yg dipengaruhi atau yg menjadi akibat, karena adanya variabel bebas.

Dari Studi Kasus Kami, Variabel Independen :

* Cloud Computing (x1)
* Proses Pembelajaran (x2)

Variabel Dependen : Penyimpanan Dokumen (Y)

Eka Wahyu Ningsih (Kelompok 6)

Pertanyaan : Data sering kali mengandung noise, outlier, atau missing value yang perlu dibersihkan sebelum proses analisis. Maksudnya apa dan berikan contohnya?!

Dijawab oleh Anita Carlolina

1. Noise

Noise dalam data adalah komponen acak atau variasi acak yang tidak diinginkan yang mengaburkan sinyal atau informasi yang sebenarnya.

Contoh:

* Data sensor: Sensor suhu yang terpengaruh oleh gangguan elektromagnetik sehingga mencatat suhu yang tidak konsisten atau tidak akurat.
* Data suara: Rekaman audio yang berisi suara bising seperti angin atau suara kendaraan yang mengganggu percakapan.

1. Outlier

Outlier adalah nilai data yang jauh berbeda dari nilai-nilai lainnya dalam dataset. Mereka bisa menjadi indikasi kesalahan atau variasi yang sangat besar dalam proses yang diamati.

Contoh:

* Data medis: Dalam pengukuran tekanan darah, jika sebagian besar pengukuran berkisar antara 120/80 mmHg hingga 140/90 mmHg, tetapi ada satu pengukuran yang tercatat sebagai 200/150 mmHg, itu bisa menjadi outlier.

1. Missing Value

Missing value adalah nilai yang hilang atau tidak tercatat dalam dataset. Hal ini bisa terjadi karena berbagai alasan, termasuk kesalahan dalam pencatatan data atau ketidakterisian formulir.

Contoh:

Data survei: Dalam survei, beberapa responden mungkin melewatkan pertanyaan tertentu, menyebabkan missing value dalam dataset.

Endar Dimayanti (Kelompok 4)

Pertanyaan : masalah apa saja yang ditemukan dari uji hasil regresi linier berganda ?

Dijawab oleh Desrita Romauli Sinambela

Masalah yang biasa muncul dalam pengujian regresi linear berganda adanya Multikolinieritas. Dalam pengujian regresi linear salah satu asumsinya adalah tidak boleh ada multikolinieritas.

Multikolinearitas terjadi ketika terdapat korelasi tinggi antar variabel bebas. Hal ini dapat menyebabkan koefisien regresi menjadi tidak stabil dan tidak dapat diinterpretasikan dengan mudah.

Anisa Salsabila (Kelompok 1)

Pertanyaan : apa peran variabel dependen dan independen dalam regresi ?

Dijawab oleh Anita Carlolina

* Variabel Independen (Predictors/Features/X):Variabel ini adalah variabel yang digunakan untuk memprediksi nilai variabel lain. Mereka adalah input dalam model regresi. Contoh: dalam studi pengaruh suhu terhadap penjualan es krim, suhu adalah variabel independen.
* Variabel Dependen (Response/Outcome/Y):Variabel ini adalah variabel yang diprediksi atau dijelaskan oleh variabel independen. Ini adalah output dalam model regresi. Contoh: dalam studi yang sama, penjualan es krim adalah variabel dependen.

**KESIMPULAN :**

Metode Regression adalah metode data mining yang dihasilkan dari survey berdasarkan variabel Variabel Independen (Cloud Computing (x1) dan Proses Pembelajaran (x2)) dan Variabel Dependen (Penyimpanan Dokumen (Y)).

Tujuan utama dari regresi adalah untuk memahami seberapa besar pengaruh variabel independen terhadap variabel dependen dan untuk membuat prediksi berdasarkan model yang dihasilkan.

**DAFTAR PUSTAKA**

[**https://journal.ibrahimy.ac.id/index.php/JIMI/article/view/1871/1337**](https://journal.ibrahimy.ac.id/index.php/JIMI/article/view/1871/1337)

1. Bagian Atas Formulir