

PREDIKSI PELAPORAN PENCURIAN KENDARAAN BERMOTOR BERBASIS LINIER REGRESI BERGANDA DI KOTA SEMARANG

Brenda Charmelita

Program Studi Teknik Informatika, Universitas Dian Nuswantoro

Jl. Nakula I No. 5-11, Semarang, INDONESIA

Email: brendacharmelita@gmail.com

Kendaraan bermotor merupakan salah satu jenis alat transportasi kemajuan teknologi di era modern. Pencurian kendaraan bermotor merupakan salah satu tindak kejahatan yang sering terjadi di kota Semarang. Berdasarkan masalah tersebut maka penelitian ini dilakukan menggunakan data mining yang dimodelkan dengan algoritma linier regresi berganda yang bertujuan untuk mengestimasi jumlah data pelaporan, sehingga kedepannya mampu meningkatkan jumlah penyelesaian pencurian kendaraan bermotor. Variabel yang digunakan dalam analisis linier regresi adalah variabel dependen Y dan variabel independen X1 dan X2. Dari hasil yang diperoleh dengan akurasi tingkat kepercayaan 95%, maka diperoleh koefisien korelasi berganda (R) adalah 0,9651 dan koefisien determinasi (R²) adalah 0,9314. Sedangkan nilai hasil estimasi hampir mendekati data yang sebenarnya yaitu 8297,9943 dengan Root Mean Square Error (RMSE) sebesar 26,93420972.

Kata Kunci : data mining, estimasi, linier regresi berganda, pencurian kendaraan bermotor

Motor vehicles are one of the types of transportation technological advances of the modern era. Motor vehicle theft is a crime that often occurs in the city of Semarang. Based on these issues, this study was performed using data mining algorithms that are modeled with a linear regression that aims to estimate the amount of data reporting, so that future settlement is able to increase the number of motor vehicle theft. Variables used in the linear regression analysis is the dependent variable Y and independent variables X1 and X2. From the results obtained with an accuracy rate of 95%, then the multiple correlation coefficient (R) is 0.9651 and the coefficient of determination (R²) is 0.9314. While nearing the estimated value of the actual data that is 8297.9943 with Root Mean Square Error (RMSE) of 26.93420972

Keywords: data mining, estimation, doubled linear regression, theft of motor vehicles.

I. PENDAHULUAN

Kota Semarang merupakan Ibukota Jawa Tengah, sebagai kota metropolitan terbesar ke-5 di Indonesia, persaingan untuk meraih kehidupan yang layak sangatlah tinggi. Tidak sedikit masyarakat yang memilih untuk melakukan tindakan kriminal demi mendapatkan uang lebih agar kehidupan ekonominya lebih sejahtera. Kriminal atau yang biasa disebut kejahatan merupakan permasalahan yang terjadi di lingkungan masyarakat dikarenakan oleh faktor ekonomi yang mengakibatkan keresahan dalam kehidupan masyarakat.

Pencurian kendaraan bermotor adalah tindakan kejahatan terhadap harta benda yang menimbulkan kerugian dan meresahkan di lingkungan masyarakat. Ada beberapa faktor yang menyebabkan tindak pidana kejahatan pencurian bermotor seperti kelalaian pemilik kendaraan bermotor dalam menjaga kendaraan bermotor miliknya. Semakin meningkat kepemilikan kendaraan bermotor semakin meningkat pula tindak pidana pencurian sepeda motor di daerah kota Semarang [1].

Dari uraian diatas dapat dilakukan analisa terhadap data tindak kriminal dan data penyelesaian tindak kriminal di Polda Jawa Tengah dengan menggunakan data mining.

Proses data mining yang akan dilakukan dalam penelitian ini adalah metode estimasi dengan menggunakan algoritma Linear Regression. Metode Estimasi merupakan fungsi tambahan yang ada pada proses data mining yang digunakan untuk memperkirakan data penyelesaian tindak pidana pencurian kendaraan bermotor yang bisa disebut "data populasi" dengan memakai data tindak pidana pencurian kendaraan bermotor yang bisa disebut "data sampel" [2].

Algoritma Linear Regression merupakan teknik data mining terhadap hubungan satu variabel atau lebih untuk menentukan relasi variabel yang diramalkan dimana data variabel saling berhubungan antara satu dengan variabel lain mencakup dalam database yang besar sehingga dapat diasumsikan sebagai pola pengambilan suatu keputusan [2].

II. TINJAUAN STUDI

Penelitian tugas akhir ini dibuat dengan melakukan studi kasus terlebih dahulu dari penelitian-penelitian sebelumnya. Dari beberapa studi keputusan tersebut, penelitian ini mengangkat tema dengan topik yang terkait dalam penelitian diatas, antara lain mengenai algoritma dan objek yang digunakan akan diangkat dalam penelitian ini. Referensi lain dalam penelitian tugas akhir ini diambil dari buku-buku serta beberapa skripsi yang

berkaitan dengan tindak pidana dan algoritma Linear Regression.

2.1. Tindak Pidana Pencurian

Tindak pidana pencurian kendaraan bermotor merupakan perbuatan kejahatan pada barang milik orang lain yang dapat menghasilkan hasil yang menguntungkan untuk tersangka dan hukuman dari pencurian kendaraan bermotor tidak berat bahkan sanksi yang didapat tersangka sangat kecil sekali sehingga tidak membuat tersangka perbuatan pencurian kendaraan bermotor menjadi jera.

2.2. Data Mining

Menurut Gartner Group data mining adalah proses ekstraksi pola dalam penyimpanan kumpulan data dengan jumlah besar dengan teknik statistic dan matematika yang digunakan untuk mengolah pengetahuan menjadi sebuah informasi.

2.3. Metode Estimasi

Metode estimasi dalam proses data mining memiliki variabel numerik berdasarkan nilai variabel prediksi. Estimasi merupakan metode untuk memperkirakan nilai populasi dengan menggunakan nilai sampel. Populasi yaitu suatu objek yang diteliti, dalam hal ini data penyelesaian tindak pidana pencurian. Sedangkan sampel yaitu objek yang dijadikan penelitian, dalam hal ini terdapat 19.472 kasus tindak pidana pencurian. Estimasi yang biasa disebut perkiraan maka diperlukan rata-rata Populasi (μ) yang diambil dari data sampel penyelesaian tindak pidana pencurian. Rata-rata populasi (σ^2) diestimasi dengan varians sampel yang dimiliki [2].

2.4. Algoritma Linear Regression

Analisis linear regression merupakan proses teknik data mining pada alat statistic yang menjelaskan pola hubungan variable independen merupakan variabel dimana dia mempengaruhi variabel lain yang bersifat berdiri sendiri. Variabel independen ini dinotasikan dengan "X". Sedangkan variabel dependen merupakan variabel yang dimana keberadaannya dipengaruhi dengan variabel lain yang memiliki sifat tidak berdiri sendiri. Variabel dependen ini dinotasikan dengan "Y" [3]. Linear Regression sederhana memiliki variabel bebas hanya ada satu, sedangkan linier regression berganda memiliki variabel bebas lebih dari satu.

2.5. Algoritma Linear Regression Berganda

Algoritma linear regression berganda (multiple linear regression algorithm) atau biasa disebut dengan algoritma linier regression klasik adalah model regression yang digunakan untuk memprediksi dan mengestimasi hubungan dua atau lebih variabel bebas pada variabel terikat terhadap parameter dengan menggunakan metode seperti metode kemungkinan maksimum (maximum likelihood estimation/ML) dan metode kuadrat terkecil (ordinary least square/OLS).

Persamaan umum garis regresi linier berganda dengan dua variabel, adalah:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Persamaan umum garis regresi linier berganda dengan tiga variabel, adalah:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \dots + b_nX_n \quad (2)$$

Algoritma linier regression terdapat tiga persamaan yang digunakan adalah sebagai berikut [4]:

Persamaan umum garis regresi linier berganda dengan tiga variabel, adalah:

$$\sum(Y) = na + b_1\sum(X_1) + b_2\sum(X_2) \quad (3)$$

$$\sum(X_1Y) = a\sum(X_1) + b_1\sum(X_1^2) + b_2\sum(X_1X_2) \quad (4)$$

$$\sum(X_2Y) = a\sum(X_2) + b_1\sum(X_1X_2) + b_2\sum(X_2^2) \quad (5)$$

a) Koefisien Determinasi (R^2)

Koefisien determinasi atau biasa disebut kuadrat koefisien korelasi digunakan untuk mengukur kemampuan model dalam mengetahui presentase pengaruh yang terjadi pada variabel bebas dan variabel terikat [5]. Rumus untuk menghitung koefisien determinasi sebagai berikut :

$$R^2 = \frac{a_1\sum X_1.Y + b_2\sum X_2.Y}{\sum Y^2} \quad (6)$$

b) Koefisien Korelasi Berganda (R)

Rumus untuk menghitung angka koefisien korelasi berganda sebagai berikut :

$$R = \frac{a_1\sum X_1.Y + b_2\sum X_2.Y}{\sum Y^2} \quad (7)$$

c) Uji Linieritas (F)

Uji linieritas (F) untuk menunjukkan variabel independen terhadap variabel dependen dalam menentukan hubungan linier (signifikan) atau tidak signifikan. Rumus untuk menghitung Fhitung sebagai berikut :

$$F_{hitung} = \frac{(R^2N - k - 1)}{k(1 - R^2)} \quad (8)$$

d) Pengujian Model

Pengujian model dilakukan untuk menguji seberapa besar kesalahan yang terjadi pada proses persamaan prediksi pada linier regresi berganda. Metode yang digunakan dalam proses pengujian model adalah RMSE (Root Mean Square Error) dimana membedakan nilai-nilai yang diprediksi dengan nilai-nilai sebenarnya dengan menggunakan ukuran yang asli yang akan menghasilkan ukuran indikator kesalahan.

$$RMSE = \sum_{i=1}^n \frac{\sqrt{(X_i - f_i)^2}}{n} [21] \quad (9)$$

III. IMPLEMENTASI

3.1 Proses Pemodelan menggunakan Linier Regresi Berganda

Data dari lampiran ke dua dengan memasukkan nilai-nilai tersebut dalam bentuk persamaan linier regression berganda, yaitu :

$$a_n + b_1 X_1 + b_2 X_2 = \sum Y \quad (10)$$

$$a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 = \sum X_1 Y \quad (11)$$

$$a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 = \sum X_2 Y \quad (12)$$

Dari persamaan diatas maka didapatkan keterangan :

n = jumlah keseluruhan data yaitu 72

a, b1, b2 = nilai yang harus ditentukan berdasarkan data hasil pengamatan.

$\sum X_1$ = jumlah tindak pidana pencurian kendaraan bermotor yaitu 17.074

$\sum X_2$ = jumlah penyelesaian tindak pidana pencurian kendaraan bermotor yaitu 7.543

$\sum Y$ = jumlah data pelapor tindak pidana pencurian kendaraan bermotor yaitu 8.298

$\sum X_1 X_2$ = jumlah perkalian jumlah tindak pidana dengan penyelesaian pencurian kendaraan bermotor yaitu 2.218.642

$\sum X_1 Y$ = jumlah perkalian jumlah tindak pidana dengan data pelapor pencurian kendaraan bermotor yaitu 2.563.883

$\sum X_2 Y$ = jumlah perkalian penyelesaian dengan data pelapor pencurian kendaraan bermotor yaitu 1.074.205

$\sum X_1^2$ = jumlah kuadrat tindak pidana pencurian kendaraan bermotor yaitu 5.225.122

$\sum X_2^2$ = jumlah kuadrat penyelesaian tindak pidana pencurian kendaraan bermotor yaitu 990.915

$\sum Y^2$ = jumlah kuadrat data pelapor tindak pidana pencurian kendaraan bermotor yaitu 1.351.060

Sehingga menjadi :

$$72(a) + 17.074(b_1) + 7543(b_2) = 8.298 \quad (13)$$

$$17.074(a) + 5.225.122(b_1) + 2.218.642(b_2) = 2.563.883 \quad (14)$$

$$7543(a) + 2.218.642(b_1) + 990.915(b_2) = 1.074.205 \quad (15)$$

Persamaan diatas dihitung dengan menggunakan Eliminasi Gauss. Metode Eliminasi Gauss merupakan operasi perhitungan eliminasi dan substitusi variabel – variabelnya sampai salah satu persamaan hanya memiliki satu bilangan tidak diketahui dan menggunakan teknik substitusi.

3.2 Pengujian Hipotesis

Koefisien korelasi berganda adalah ukuran relative dari ukuran $-1 \leq R \leq +1$ dari asosiasi pada dua variabel yang dimanfaatkan untuk mengetahui kuatnya suatu hubungan antara variabel dependen dengan variabel independen. $R = +1$ menyatakan variabel dependen dan variabel independen mempunyai nilai yang sama-sama kecil atau nilai yang sama-sama besar dan memiliki hubungan linier sempurna langsung. $R = -1$ menyatakan variabel dependen mempunyai nilai besar dan nilai kecil yang bersamaan dengan variabel independen yang mempunyai nilai besar atau nilai kecil dan memiliki hubungan linier sempurna tidak langsung.

Rumus Koefisien Korelasi Berganda (R) :

$$R = \sqrt{\frac{a_1 \sum X_1 \cdot Y + a_2 \sum X_2 \cdot Y}{\sum Y^2}} \quad (16)$$

Maka menjadi :

$$R = \sqrt{\frac{(0,6159 \cdot 2563883) + (-0,2985 \cdot 1074205)}{1351060}} \quad (17)$$

$$R = \sqrt{\frac{1.579.095,54 - 320.650,1925}{1351060}}$$

$$R = \sqrt{\frac{1.258.445,348}{1351060}}$$

$$R = \sqrt{0,9314} = 0,9651$$

Dari hasil koefisien korelasi berganda diatas bahwa nilai koefisien korelasi berganda adalah 0,9651 menyatakan nilai mendekati 1 sehingga dapat disimpulkan bahwa nilai variabel dependen yaitu data pelapor tindak pidana pencurian kendaraan bermotor dengan nilai variabel independen yaitu tindak pidana dan penyelesaian pencurian kendaraan bermotor memiliki hubungan yang sempurna (cocok).

3.3 Koefisien Determinasi (R²)

Koefisien determinasi (kuadrat koefisien korelasi) adalah suatu ukuran persamaan linier regression berganda pada data digunakan untuk mengukur kemampuan model dalam mengetahui presentase pengaruh yang terjadi pada

variabel bebas dan variabel terikat yang menyatakan kecocokan pengukuran garis linier regresi mendekati nilai data asli oleh model.

Koefisien determinasi (R^2) mempunyai nilai interval antara 0 sampai 1, jika nilai (R^2) semakin mendekati 1 maka hasil model regresi semakin baik dan jika nilai (R^2) semakin mendekati 0 maka hasil variabel independen tidak dapat menerangkan variabel dependen.

Rumus untuk menghitung koefisien determinasi sebagai berikut :

$$R = \left(\sqrt{\frac{a_1 \sum X_1 \cdot Y + a_2 \sum X_2 \cdot Y}{\sum Y^2}} \right)^2$$

$$R^2 = (R)^2$$

$$R^2 = (0,9651)^2$$

$$R^2 = 0,9314$$
(18)

Hasil yang didapat koefisien determinasi yaitu 0,9314 maka menyatakan linier regression berganda yang dihasilkan pada penelitian ini garis linier regresi cocok dengan data pada penelitian ini.

3.4 Uji Linieritas (F)

Uji linieritas (F) untuk menunjukkan variabel independen terhadap variabel dependen dalam menentukan hubungan linier (signifikan) atau tidak signifikan. Dalam mengambil keputusan uji linieritas (F) dengan menghitung Fhitung dan mencari nilai Ftabel untuk tingkat signifikan 0,05% atau (5%) dan tingkat kepercayaan sebesar (95%).

Rumus untuk menghitung Fhitung yaitu :

$$F_{hitung} = \frac{R^2 (n-k-1)}{k (1-R^2)}$$
(19)

IV. HASIL & PEMBAHASAN

Data yang digunakan pada penelitian ini merupakan data dari Kepolisian Daerah Jawa Tengah. Data yang diperoleh berkisar pada tahun 2010 sampai dengan tahun 2015.

Data pada lampiran pertama yang telah di dapat dari Kepolisian Daerah Jawa Tengah diketahui data yang digunakan adalah 72 data. Variabel Independen yaitu tindak pidana pencurian kendaraan bermotor diasumsikan X_1 dan penyelesaian tindak pidana pencurian kendaraan bermotor diasumsikan X_2 . Variabel Dependen yaitu data pelapor tindak pidana pencurian kendaraan bermotor diasumsikan Y . Dalam perhitungan dalam analisis linier regresi berganda yang terdapat pada lampiran ke lima melakukan perhitungan jumlah $X_1, X_2, Y, X_1 Y, X_2 Y, X_1 X_2, X_1^2, X_2^2, dan Y^2$.

Estimasi jumlah data pelaporan pencurian kendaraan bermotor diketahui dengan menghitung nilai (Y'). Nilai X_1 (Tindak pidana pencurian kendaraan bermotor) dan X_2 (Penyelesaian tindak pidana pencurian kendaraan bermotor) ke dalam persamaan $Y = 0,4681 + 0,6159 X_1$

$- 0,2985 X_2$ dan dimasukkan sebanyak data 72 data sebagai berikut :

$$(Y')_1 = 0,4681 + 0,6159 (50) - 0,2985 (29) = 22,6066$$

$$(Y')_2 = 0,4681 + 0,6159 (54) - 0,2985 (35) = 23,2792$$

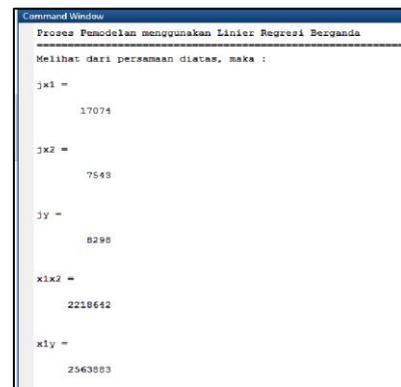
$$(Y')_3 = 0,4681 + 0,6159 (46) - 0,2985 (24) = 21,6355$$

$$(Y')_4 = 0,4681 + 0,6159 (46) - 0,2985 (28) = 20,4415$$

$$(Y')_5 = 0,4681 + 0,6159 (49) - 0,2985 (32) = 21,0952$$

Hasil perhitungan estimasi data pelaporan pencurian kendaraan bermotor terdapat pada lampiran ke - 3. Nilai estimasi jumlah data pelaporan pencurian kendaraan bermotor mencapai 8297,9943.

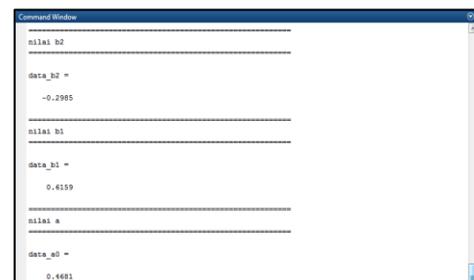
Dari hasil estimasi menyatakan metode estimasi yang dilakukan pada penelitian ini dapat digunakan untuk mengetahui pengaruh prediksi seberapa banyak data pelaporan tindak pidana pencurian kendaraan bermotor di Kepolisian Daerah Jawa Tengah.



Gambar 4.1 Hasil Perhitungan Jumlah $X_1, X_2, Y, X_1 X_2, X_1 \cdot Y, X_2 \cdot Y$



Gambar 4.2 Hasil Perhitungan Jumlah $X_2 \cdot Y, (X_1)^2, (X_2)^2, dan (Y)^2$



Gambar 4.3 Hasil Perhitungan $a, b_1, dan b_2$

```
Command Window
-----
Menghitung Uji Hipotesis
-----
ans =

Ho Ditolak, Ha Diterima

ans =

Ha :  $\alpha = 1 - \beta = 0$  Variabel Tindak Pidana dan Penyelesaian Pencurian Kendaraan Bermotor berpengaruh signifikan terhadap Data Pelapor
```

Gambar 4.4 Hasil Perhitungan Uji Hipotesis

V. PENUTUP

Berdasarkan pada hasil perhitungan pengestimasi data pelapor pencurian kendaraan bermotor dengan menggunakan algoritma linier regression berganda akan memudahkan pihak Kepolisian Daerah Jawa Tengah dalam meningkatkan penyelesaian pencurian kendaraan bermotor sehingga akan mengetahui seberapa banyak terjadinya pencurian kendaraan bermotor, dan dapat mengetahui pelaporan pencurian kendaraan bermotor untuk tahun selanjutnya mengalami peningkatan atau penurunan.

Pengestimasi menggunakan algoritma linier regression berganda dengan cara Eliminasi Gauss yang memiliki tingkat akurasi 95% sehingga dapat diperoleh hasil koefisien korelasi berganda (R) yaitu 0,9651, koefisien determinasi (R²) yaitu 0,9315, dan mendapatkan hasil pengestimasi dari persamaan linier regression berganda yaitu 8297,9943 dengan tingkat kesalahan terjadinya error (RMSE) sebesar 26,9342 [5].

REFERENSI

- [1] G. Y. Hilman, B. Sasmito dan A. P. Wijaya, "Pemetaan Daerah Rawan Kriminalitas di Wilayah Hukum Poltabes Semarang Tahun 2013 dengan Menggunakan Metode Clustering".
- [2] A. Fikri, "Penerapan Data Mining untuk Mengetahui Tingkat Kekuatan Beton yang Dihasilkan dengan Metode Estimasi Menggunakan Linier Regression".
- [3] E. S. Tataming, "Analisa Besar Kontribusi Hambatan Samping Terhadap Kecepatan Dengan Menggunakan Model Regresi Linier Berganda (Studi Kasus: Ruas Jalan Dalam Kota Segmen Ruas Jalan Serapung)," *Sipil Statistik*, vol. 2, 2014.
- [4] Y. H. Ngumar, "Aplikasi Metode Numerik dan Matrik dalam Perhitungan Koefisien-koefisien Regresi Linier Multiple untuk Peramalan," dalam *Konferensi Nasional Sistem Dan Informatika*, 2008.
- [5] I. Ghozali, *Analisa Multifariat Dengan Program SPSS*, Semarang: Badan Penerbit UNDIP, 2005.