

PENERAPAN ALGORITMA REGRESI LINIER BERGANDA PADA DATA PABRIK GULA RENDENG KUDUS

Karina Dian Ariani

Jurusan Teknik Informatika FIK UDINUS, Jl. Nakula No. 5-11 Semarang-50131

karinaariani.off@gmail.com

Abstrak - Tebu merupakan salah satu jenis tanaman yang dapat ditanam didaerah beriklim tropis. Komoditas tebu merupakan salah satu subsektor unggulan di area perkebunan. Tanaman tebu merupakan bahan utama pembuatan gula yang kini sedang mengalami krisis. Kondisi kurangnya pasokan produksi gula yang dikarenakan terjadinya peningkatan permintaan gula. Peningkatan penduduk yang melonjak menyebabkan produktivitas gula semakin rendah. Berdasarkan masalah tersebut maka penelitian ini dilakukan menggunakan data mining yang dimodelkan dengan algoritma regresi linier berganda yang bertujuan untuk mengestimasi jumlah gula yang mampu dihasilkan oleh para petani tebu, sehingga kedepannya mampu meningkatkan jumlah produksi gula. Variabel yang digunakan dalam analisis regresi linier adalah variabel *dependen* dan variabel *independen*. Dari data yang diperoleh maka yang menjadi variabel *dependen* adalah jumlah gula (SHS) yang disimbolkan dengan Y sedangkan variabel *independen* adalah tebu yang digiling yang disimbolkan dengan X_1 dan rendemen sebagai X_2 . Dari hasil yang diperoleh dengan akurasi tingkat kepercayaan 95%, maka diperoleh koefisien korelasi berganda (R) adalah 0.990740634 dan koefisien determinasi (R^2) adalah 0.981567003. Sedangkan nilai hasil estimasi hampir mendekati data yang sebenarnya yaitu 180.8261724 Kuintal (Ku) dengan *Root Mean Square Error* (RMSE) sebesar 0.0624.

Kata Kunci : produksi gula, data mining, estimasi, linier regresi berganda

I. PENDAHULUAN

Tebu merupakan salah satu komoditas perkebunan yang penting ditanam karena sebagai bahan baku utama gula. Peran gula juga sangat berpengaruh dalam perekonomian, mengingat kebutuhan gula yang semakin tinggi membuat Indonesia menjadi Negara pengimpor gula. Ada beberapa faktor yang menyebabkan Indonesia menjadi Negara pengimpor gula. Salah satu faktor utamanya adalah meningkatnya permintaan gula oleh masyarakat yang kurang mampu dipenuhi oleh industri gula dalam negeri. Dengan demikian perlunya meningkatkan keefektifan dalam pengolahan produksi

untuk meningkatkan hasil produksi yang lebih baik [2].

Dari uraian diatas dapat dilakukan analisa terhadap data produksi gula di daerah Kabupaten Kudus dengan menggunakan metode *data mining*. *Data mining* adalah proses menggali nilai tambah berupa informasi yang tidak diketahui secara manual [3].

Salah satu proses *data mining* yang akan digunakan adalah metode estimasi dengan algoritma Linier Regresi. Estimasi merupakan perkiraan suatu hal dalam sejumlah sample. Estimasi adalah fungsi tambahan yang ada pada *data mining*. Algoritma regresi linier merupakan teknik

data mining untuk menentukan bahwa terdapat hubungan antara variabel yang ingin diramalkan dengan variabel lain [4].

Beberapa penelitian yang telah dilakukan sebelumnya, Edy Susanto Tataming pada penelitiannya menyatakan bahwa analisis regresi merupakan sebuah alat statistika yang memberikan penjelasan tentang pola hubungan (model) antar dua variabel atau lebih [5]. Penelitian lainnya Sarita Permata Dewi menyatakan bahwa analisis regresi linier berganda digunakan untuk memprediksi pengaruh dua variabel bebas atau lebih terhadap satu variabel terikat [6]. M. Fathurrahman & Haeruddin menyatakan bahwa analisis regresi linier berganda merupakan salah satu teknik analisis data yang sering digunakan untuk mengkaji hubungan antar beberapa variabel dan meramal suatu variabel [7]. Berdasarkan penelitian tersebut maka penelitian yang akan dibuat dilakukan menggunakan metode yang sama yakni analisis regresi linier berganda karena metode tersebut lebih umum digunakan pada studi kasus yang sedang diteliti.

II. TEORI PENUNJANG

2.1 Estimasi

Estimasi adalah memperkirakan suatu hal dari sejumlah sample, lebih condong ke pengklasifikasian namun variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Pembangunan sistem dilakukan menggunakan record yang lengkap yang menyediakan nilai dari variabel target sebagai prediksi. Sehingga nilai variabel target akan dibuat sesuai dengan nilai variabel prediksi

2.2 Algoritma Regresi Linier Berganda

Analisa regresi linier berganda adalah teknik untuk menentukan hubungan antara variabel yang ingin diestimasi atau diramalkan (variabel

tak bebas) dengan variabel lain (variabel bebas). Dalam analisa regresi dikenal dua jenis variabel yakni variabel tergantung disebut juga variabel *dependen* yaitu variabel yang keberadaannya dipengaruhi oleh variabel lainnya yang sifatnya tidak dapat berdiri sendiri dan dinotasikan dengan Y, dan variabel bebas yang disebut sebagai variabel *independen* merupakan variabel yang mempengaruhi variabel yang lain yang sifatnya berdiri sendiri dan dinotasikan dengan X [5]. Persamaan umum algoritma regresi linier berganda, adalah :

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n$$

Keterangan :

Y = variabel terikat / variabel *dependent* (nilai yang diprediksi)

a_0, a_1, a_2, a_n = koefisien regresi

X_1, X_2, \dots, X_n = variabel bebas / variabel *independent*

Model koefisien regresi yang digunakan dalam perhitungan adalah menggunakan metode Eliminasi Gauss. Metode ini mempunyai prinsip yaitu mengurangi sistem persamaan ke dalam bentuk segitiga atas sedemikian hingga salah satu dari persamaan – persamaan tersebut hanya mengandung satu bilangan tak diketahui. Persamaan yang digunakan adalah sebagai berikut [16] :

$$\sum Y = a_0n + a_1 \sum X_1 + a_2 \sum X_2$$

$$\sum X_1Y = a_0 \sum X_1 + a_1 \sum X_1^2 + a_2 \sum X_1X_2$$

$$\sum X_2Y = a_0 \sum X_2 + a_1 \sum X_1X_2 + a_2 \sum X_2^2$$

a. Koefisien Korelasi Berganda (R)

Koefisien korelasi Berganda (R) adalah suatu ukuran relatif dari asosiasi diantara dua variabel.

Koefisiensi korelasi bertujuan untuk mengetahui kuatnya hubungan antara variabel *dependen* dengan variabel *independen* [5].

Rumus yang digunakan adalah sebagai berikut :

$$R = \sqrt{\frac{a_1 \sum X_1 Y + a_2 \sum X_2 Y}{\sum Y^2}}$$

b. Koefisien Determinasi

Koefisiensi determinasi digunakan untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel *dependen* [18].

Nilai dari koefisiensi determinasi (R^2) mempunyai interval antara 0 sampai 1 ($0 \leq R^2 \leq 1$). Jika nilai R^2 mendekati 1, semakin baik untuk model regresi tersebut dan apabila nilai R^2 mendekati 0 maka variabel *independen* tidak dapat menjelaskan variabel *dependen* secara keseluruhan [19].

$$R^2 = \sqrt{\frac{a_1 \sum X_1 Y + a_2 \sum X_2 Y}{\sum Y^2}}^2$$

c. Uji F (Uji Linieritas)

Uji F digunakan untuk mengetahui hubungan antara variabel *independent* dengan variabel *dependen* mempunyai hubungan linier (signifikan) atau tidak (tidak signifikan) [6]. Pengambilan keputusan dilakukan dengan cara membandingkan antara nilai F_{hitung} dan F_{tabel} dengan tingkat signifikansi yaitu alpha (α) sebesar 5% (0,05), dengan tingkat kepercayaan pada umumnya sebesar 95% [14].

Dengan rumus F_{hitung} :

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

Keterangan :

N = jumlah sampel data

k = jumlah variabel bebas (variabel *independent*)

Perhitungan F_{tabel} dilakukan dengan tabel bantu distribusi F dengan nilai probabilitas 0.05% dengan perhitungan derajat kebebasan (Dk) pembilang = k, dan derajat kebebasan (Dk) penyebut = n-k-1. Serta hipotesis yang digunakan adalah

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_n = 0$$

Model regresi linier berganda tidak signifikan atau tidak memiliki hubungan linier antara variabel *independen* terhadap variabel *dependen*.

$$H_1 : \beta_1 \neq \beta_2 \neq \dots \neq \beta_n \neq 0$$

Model regresi linier berganda signifikan atau memiliki hubungan linier antara variabel *independen* terhadap variabel *dependen*.

Pengambilan keputusan sebagai berikut :

$$F_{hitung} < F_{tabel} = H_0 \text{ diterima}$$

$$F_{hitung} > F_{tabel} = H_0 \text{ ditolak,}$$

$$H_1 \text{ diterima [19]}$$

2.3 Pengujian Model

Tahap pengujian model adalah tahap evaluasi dimana model hasil persamaan linier regresi diprediksi seberapa besar kesalahannya. Metode yang digunakan dalam pengujian model

adalah *Root Mean Square Error* (RMSE). *Root Mean Square Error* (RMSE) adalah ukuran yang digunakan sebagai pembeda antara nilai – nilai yang diprediksi dengan nilai – nilai yang sebenarnya [4]

$$RMSE = \frac{\sum_{i=1}^n \sqrt{(X_i - f_i)^2}}{n}$$

Dimana semakin besar nilai RMSE yang dihasilkan maka keakuratan suatu model semakin sedikit atau tidak akurat, sedangkan semakin kecil nilai RMSE maka semakin baik akurasi suatu model regresi linier tersebut [20].

III. HASIL & IMPLEMENTASI

Variabel yang digunakan dalam penelitian ini adalah tebu yang digiling, rendemen dan jumlah gula (SHS). Jumlah gula (SHS) digolongkan dalam variabel *dependen* karena di pengaruhi oleh variabel *independen* yaitu tebu yang digiling dan rendemen. Dari data yang telah didapat diketahui bahwa data yang digunakan adalah 144 data. Dari variabel yang telah didefinisikan sebelumnya maka yang menjadi variabel *dependen* yaitu jumlah gula (SHS) diasumsikan sebagai Y, sedangkan variabel *independen* yaitu tebu dan rendemen diasumsikan sebagai X_1 dan X_2 .

Hasil yang didapat dari perhitungan menggunakan analisis regresi linier berganda adalah menghasilkan model linier $Y = 0.0104 + 0.0628 X_1 - 0.00005249 X_2$. Dari persamaan dapat diketahui hasil estimasi produksi gula yaitu 180.8261724 Kuintal (Ku) gula. Serta besar hubungan linieritas yang dihasilkan dengan tingkat kepercayaan 95%, maka diperoleh koefisien korelasi berganda (R) adalah 0.990740634 dan koefisien determinasi (R^2) adalah

0.981567003. Sedangkan besar tingkat kesalahan yang dihitung menggunakan metode *Root Mean Squared Error* (RMSE) adalah sebesar 0.0624.

IV. PENUTUP

Berdasarkan penelitian yang telah dilakukan dalam pengestimasian produksi gula dari Pabrik Gula Rendeng Kabupaten Kudus menggunakan algoritma *Multiple Linier Regression* dapat dimanfaatkan sebagai salah satu acuan untuk mengetahui seberapa banyak gula yang dapat dihasilkan serta berguna untuk mengetahui seberapa besar potensi komoditas gula di daerah Kabupaten Kudus.

Variabel yang digunakan adalah variabel *dependen* dan variabel *independen*. Variabel *dependen* dalam data yang digunakan adalah jumlah gula (SHS) yang diasumsikan dengan Y serta variabel *independennya* adalah tebu yang digiling yang diasumsikan sebagai X_1 dan rendemen yang diasumsikan sebagai X_2 . Hasil dari pengestimasian produksi gula menggunakan pendekatan Eliminasi Gauss dengan tingkat kepercayaan 95%, maka diperoleh koefisien korelasi berganda (R) adalah 0.990740634 dan koefisien determinasi (R^2) adalah 0.981567003. Sedangkan hasil estimasi dari model persamaan linier regresi yakni 180.8261724 Kuintal (Ku) gula dan tingkat *Root Mean Squared Error* (RMSE) sebesar 0.0624.

DAFTAR PUSTAKA

- [1] S. Purnama, "Website Dinas Perkebunan Provinsi Jawa Barat," 28 Agustus 2014. [Online]. Available: <http://disbun.jabarprov.go.id/index.php/artikel/detailartikel/39>. [Accessed 23 Desember 2014].

- [2] E. R. Anandita, "Klasifikasi Tebu dengan Menggunakan Algoritma Naive Bayes Classification pada Dinas Kehutanan dan Perkebunan Pati".
- [3] E. T. L. Kursini, *Algoritma Data mining*, Yogyakarta: Andi, 2009.
- [4] A. Fikri, "Penerapan *Data mining* untuk Mengetahui Tingkat Kekuatan Beton yang Dihasilkan dengan Metode Estimasi Menggunakan Linier Regression".
- [5] 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.
- [6] S. P. Dewi, "Pengaruh Pengendalian Internal dan Gaya Kepemimpinan terhadap Kinerja Karyawan SPBU Yogyakarta (Studi Kasus pada SPBU Anak Cabang Perusahaan RB.Group)," *Nominal*, vol. 1, 2012.
- [7] M. Fathurahman and H. , "Pemodelan Regresi Linier untuk Data Deret Waktu," *Ekspansional*, vol. 2, 2011.
- [8] S. Sigilipu, "Pengaruh Penerapan Informasi Akuntansi Manajemen dan Sistem Pengukuran Kinerja terhadap Kinerja Manajerial," *EMBA*, vol. 1, 2013.
- [9] H. Susanto and S. , "*Data mining* untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa Lalu," *Pendidikan Vokasi*, vol. 4, 2014.
- [10] F. Rushdy, "repository.usu.ac.id," 2012. [Online]. Available: <http://repository.usu.ac.id/bitstream/123456789/30937/4/Chapter%20II.pdf>. [Accessed 24 Desember 2014].
- [11] L. Ernawati and E. Suryani, "Analisa Faktor Produktivitas Gula Nasional dan Pengaruhnya terhadap Harga Gula Domestik dan Permintaan Gula Impor dengan Menggunakan Sistem Dinamik," *Teknik POMITS*, vol. 1, 2013.
- [12] D. T. Larose, *Discovering Knowledge in Data : An Introduction to Data mining*, Jhon Willey & Sons.Inc, 2005.
- [13] D. Kurniawan, "Regresi Linier," 2008.
- [14] J. Sarwono, *12 Jurus Ampuh SPSS untuk Riset Skripsi*, Jakarta: Elexmedia Komputindo , 2013.
- [15] J. T. S. H. Nur Setiaji Pamungkas, "Model Regresi Linier Pengaruh Komposisi Kendaraan Terhadap Tingkat Kecelakaan pada Jalan Tol Surabaya-Gempol".
- [16] Y. H. Ngumar, "Aplikasi Metode Numerik dan Matrik dalam Perhitungan Koefisien - Koefisien Regresi Linier Multiple untuk Peramalan," *Konferensi Nasional Sistem dan Informatika*, 2008.
- [17] R. Istiarini and S. , "Pengaruh Sertifikasi Guru dan Motivasi Guru terhadap Kinerja Guru SMA Negeri 1 Sentolo Kabupaten Kulon Progo Tahun 2012," *Jurnal Pendidikan Akutansi Indonesia*, vol. X, pp. 98-113, 2012.
- [18] I. Ghozali, *Analisa Multifariat dengan Program SPSS*, Semarang: Badan Penerbit UNDIP, 2005.
- [19] W. Sulaiman, *Analisis Regresi menggunakan SPSS Contoh Kasus & Pemecahannya*, Yogyakarta: Andi, 2004.
- [20] A. Yusuf, H. Ginardi and I. Arieshanti, "Pengembangan Perangkat Lunak Prediktor Nilai Mahasiswa Menggunakan Metode Spectral Clustering dan Bagging Regresi Linier," *Jurnal Teknik ITS*, vol. 1, 2012.
- [21] Sudjana, *Metoda Statistika*, Bandung: Tarsinto, 2005.