**Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunaka Metode Linier Regression**

**Application of Data Mining in prediction of palm oil production of PT Borneo Ketapang Indah using linear method Regression**

**Nuning Rusmilawati**

Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta, Jl. Wates Km 10. Yogyakarta 55753, Indonesia

Email : nuningrusmila@gmail.com

**ABSTRAK**

PT Borneo Ketapang Indah merupakan suatu perusahaan yang bergerak dalam bidang perkebunan produksi kelapa sawit yang terletak di pulau Kalimantan. Dalam memprediksi hasil produksi kelapa sawit dapat membantu perusahaan untuk melihat factor apa saja yang mempengaruhi dalam proses produksi kelapa sawit pada masa yang akan datang dengan menggunakan data tiga tahun terakhir 2017-2019 berdasarkan data yang ada seperti Luas Lahan, Curah Hujan, Pemupukan dan Jumlah Produksi sebagai variable yang akan diprediksi. Oleh karena itu, peneliti melakukan analisi Data Mining Hasil Produksi kelapa sawit agar menjadi informasi yang dapat di gunakan oleh PT Borneo Ketapang Indah. Dalam melakukan analisis peneliti menggunakan tools Rapidminer, metode yang digunakan dalam penelitian ini adalah Liniear Regression. Setelah dilakukan analisis prediksi Liniear Regression pada tools rapidminer maka di dapatlah 2 variable factor yang mempengaruhi Hasil Produksi kelapa sawit yaitu variable X1 (Luas Lahan) dan variable X2 (Curah Hujan) dengan nilai coefficient masing-masing variable sebesar 0,646 untuk variable Luas Lahan, variable Curah Hujan sebesar 0,530.

Kata Kunci : *Data Mining, Liniear Regression, Rapidminer*

**ABSTRACT**

PT Borneo Ketapang Indah is a company engaged in palm oil production plantations located in the island of Borneo. In predicting the results of palm oil production can help the company to see what factors affect in the process of production of oil palm in the future using the data of the last three years 2017-2019 based on existing data such as land area, rainfall, fertilization and production amount as a variable that will be predicted. Therefore, researchers analyze the Data Mining of palm oil production results to be the information that can be used by PT Borneo Ketapang Indah. In conducting research analysis using Rapidminer tools, the method used in this study is Liniear Regression. After analysis of the prediction Liniear Regression on RapidMiner tools, there are 2 variable factors that affect the yield of palm oil production is variable X1 (land area) and variable X2 (rainfall) with a value coefficient each variable of 0.646 for the variable land area, the rainfall variable of 0.530.

Keywords: *Data Mining, Liniear Regression, Rapidminer*

**1. PENDAHULUAN**

**PENERAPAN DATA MINING DALAM PREDIKSI HASIL PRODUKSI KELAPA SAWIT PT BORNEO KETAPANG INDAH MENGGUNAKAN METODE LINIER REGRESSION**

Tanaman Kelapa Sawit saat ini merupakan salah satu jenis tanaman perkebunan yang menduduki posisi penting disektor pertanian umumnya, dan sektor perkebunan khususnya, hal ini disebabkan karena dari sekian banyak tanaman yang menghasilkan minyak atau lemak, kelapa sawit yang menghasilkan nilai ekonomi terbesar per hektarnya di dunia. Peramalan atau forecasting merupakan seni dan ilmu yang dilakukan untuk mengetahui atau memperkirakan apa yang terjadi pada masa yang akan datang. Peramalan menjadi sangat penting karena penyusunan suatu rencana diantaranya didasarkan pada suatu proyeksi atau ramalan.(Setyamodjaja, 2006)

PT Borneo Ketapang Indah merupakan suatu perusahaan yang bergerak dalam bidang perkebunan produksi kelapa sawit yang terletak di pulau Kalimantan. Dalam memprediksi jumlah produksi kelapa sawit dapat membantu perusahaan untuk melihat factor apa saja yang mempengaruhi dalam proses produksi kelapa sawit pada masa yang akan datang dengan menggunakan data yang ada seperti Luas Lahan, Curah Hujan, Pemupukan dan Jumlah Produksi sebagai variable yang akan diprediksi.

Berdasarkan urai-uraian diatas, penulis tertarik untuk mengangkat skripsi dengan judul “Penerapan Data Mining dalam Prediksi Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah Menggunakan Metode Linear Regression”. Dimana penerapan Data Mining diharapkan dapat mengatasi permasalahan.

**2. TINJAUAN PUSTAKA**

Penelitian pertama (Kurniadi et al., 2020) oleh yang berjudul “Penerapan Metode Regresi Linier untuk Memprediksi Kebiasaan Pelanggan Studi Kasus: PT. Mensa Binasukses”. PT. Mensa Binasukses merupakan distributor usaha yang bergerak di bidang pelayanan kesehatan, khususnya dalam hal penjualan obat-obatan, baik itu penjualan obat melalui resep dokter maupun apotek.

Permasalahan yang muncul adalah bagaimana mengolah data transaksi pelanggan yang sudah ada menjadi sebuah prediksi kebiasaan pelanggan. Hasil prediksi ini akan digunakan sebagai strategi perusahaan untuk memberi rekomendasi pembelian produk dalam bentuk paket (untuk pelanggan) dan strategi pemasaran serta pembuatan katalog (untuk perusahaan).

Proses prediksi dilakukan dengan cara menggali informasi hubungan asosiasi antar item produk yang terjadi pada setiap transaksi. Salah satu Metode yang dapat digunakan untuk menggali informasi potensial ini adalah dengan pendekatan data mining menggunakan metode regresi linier. Penelitian ini bertujuan untuk mengimplementasikan sebuah perangkat lunak untuk memprediksi kebiasaan pelanggan menggunakan metode regresi linier dengan memanfaatkan data transaksi pelanggan yang sudah ada.

Penelitian kedua oleh (Herwanto et al., 2019) yang berjudul “Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi”. Hasil panen padi sangat berpengaruh dalam pemenuhan kebutuhan pangan pokok beras. Dikarenakan kebutuhan beras tiap tahunnya selalu mengalami peningkatan, diperlukan prediksi hasil panen untuk memperkirakan penanaman yang akan dilakukan, sehingga dapat mencukupi pemenuhan kebutuhan beras. Untuk memperoleh hasil panen, diperlukan informasi mengenai luas lahan, jumlah bibit yang akan di tanam, dan pupuk yang digunakan sebagai penunjang kebutuhan tanaman padi.

**3. METODOLOGI PENELITIAN**

Pada penelitian Penerapan Data Mining Dalam Prediksi Jumlah Hasil Produksi Kelapa Sawit PT Borneo Ketapang Indah menggunakan metode Linear Regression akan melalui beberapa proses, berikut diagram alur proses jalan penelitian pada gambar.



Gambar 1. Flowchart alur penelitian

**4. PEMBAHASAN**

Data yang digunakan dalam penelitian ini yaitu data jumlah hasil produksi kelapa sawit berdasarkan data historis 3 tahun terakhir yaitu 2017, 2018 dan 2019 data tersebut berasal dari PT Borneo Ketapang Indah. Berdasarkan data yang diperoleh langsung dari PT Borneo ketapang Indah yang dimana data tersebut telah melalui proseses selection yang terdiri dari beberapa atribut yaitu Bulan, Luas Lahan, Curah Hujan, Pupuk dan Jumlah produksi yang dapat dilihat pada table dibawah.

| Bulan | Luas Lahan | Curah Hujan | Pupuk (kg) | Jumlah Produksi |
| --- | --- | --- | --- | --- |
| Jan-17 | 6,28 | 31140 | 7500 | 338535 |
| Feb-17 | 11,01 | 22170 | 7500 | 395119 |
| Mar-17 | 10,13 | 41070 | 7550 | 450037 |
| Apr-17 | 17,71 | 22470 | 8250 | 480269 |
| Mei-17 | 18,89 | 24090 | 8250 | 53638 |
| Juni-17 | 15,35 | 18730 | 7810 | 507942 |

Tabel 1. Atribut

**PERHITUNGAN MANUAL**

Adapun **rumus Liniear Regression** adalah sebagai berikut :

Ƴ՚= ɑ + b₁X₁ + b₂X₂+…..bₙXₙ

Keterangan :

a. Ƴ՚ = Variable Dependen (Nilai yang diprediksikan)

b. X₁, X₂, Xₙ = Variable Independen

c. Konstanta (nilai Ƴ՚ apabila X₁, X₂, ….. Xₙ=0)

d. b₁, b₂, bₙ = Koefisien Regresi

**Rumus Liniear Regression kedua**, untuk mencari a dan b dapat di peroleh dengan menggunakan rumus dibawah ini :

a. ∑ Ƴ₁ = nɑ+ ∑b₁X₁ + ∑b₂X₂ + ∑b₃X₃

b. ∑X₁Y₁ = ɑ∑x₁ + b₁∑(X₁)² + b₂∑X₁X₂ + b₃∑X₁X₃

c. ∑X₂Y₁ = ɑ∑x₂ + b₁∑X₁X₂ + b₂∑(X₂)² + b₃∑X₂X₃

d. ∑X₃Y₁ = ɑ∑x₃ + b₁∑X₁X₃ + b₂∑X₂X₃ + b₃∑(X₃)²

Tahapan perhitungan manual Liniear Regression.

a. Siapkan data \*prediksi jumlah produksi

b. Hitung berapa jumlah data set (n). \*n=19

c. Tentukan mana yang X dana mana Y. Y disini adalah nilai yang akan diprediksi. \*X1=Luas Tanah, X2= Curah Hujan, X3=Pupuk, Y= Jumlah Produksi

d. Hitung (X₁)², (X₂)², (X₃)³

e. Hitung X₁\*X₂, X₁\*X₃, X₂\*X₃, X₁\*Y, X₂\*Y, X₃\*Y

f. Hitung ∑X₁, ∑X₂, ∑X₃, Ƴ

g. Hitung ∑(X₁)², ∑(X₂)², ∑(X₃)²

h. Hitung ∑X₁X₂, ∑X₁X₃, ∑X₂X₃, ∑X₁Y, ∑X₂Y, ∑X₃Y

i. Masukan perhitungan kedalam rumus persamaan mencari a, b1,b2, b3

j. Hitung nilai a, b1, b2, b3 menggunakan perhitungan matrixs

k. Buat persamaan regregsi liniear

Perhitungan Manual Liniear Regression dengan membagi data training dan testing dapat dilihat pada table dibawah ini yang dimana atribut Luas Lahan menjadi variable X1, atribut Curah Hujan menjadi variable X2, atribut Pupuk menjadi variable X3 dan atribut Jumlah Produksi menjadi variable Y yang nantinya akan diprediksi.



Gambar 2. Data Training



Gambar 3. Data Testing

Hitung $∑X\_{1 }$, $∑X\_{2}$ , $∑X\_{3}$, $∑$ $Y$



Gambar 4. Jumlah Variable

Hitung $(X\_{1})^{2}$, $(X\_{2})^{2}$, $(X\_{3})^{2}$, dan $∑(X\_{1})^{2}$, $∑(X\_{2})^{2}$, $∑(X\_{3})^{2}$



Gambar 5. Jumlah pengkuadratan

Hitung $X\_{1}$\*$X\_{2}$ , $X\_{1}$\*$X\_{3}$ , $X\_{1}$\*$X\_{₃}$ , $X\_{2}$\*$X\_{3}$ ,$ ,X\_{1}$\* $Y$ , $X\_{2}$\* $Y$ , $X\_{3}$\* $Y$ , dan $∑X\_{1 }X\_{2 }$, $∑X\_{1 }X\_{3 }$, $∑X\_{1 }X\_{3}$, $∑X\_{2 }X\_{3 }$, $∑X\_{1 }Y$ , $∑X\_{2 }Y$ , $∑X\_{3 }Y$

Hitung nilai a, b1, b2, b3 menggunakan perhitungan matriks. Cara dari perhitungan matriks secara manual dapat dilihat pada pembahasan nomor 4, maka setelah memasukan rumus matriks yang di peroleh hasil matriks sebagai berikut.

$$\sum\_{}^{}y^{1}=na+ ∑b₁^{}X₁^{}+\sum\_{}^{}b₂X₂+∑b₃X₃$$

|  | 19 |  | 8.180.008 |
| --- | --- | --- | --- |
| A | 276,36 | B | 121.152.068 |
|  | 387.889,00 |  | 169.533.345.850 |
|  | 148.890 |  | 64.198.775.700 |

Tabel 2. Perhitungan Matriks

|  |  |
| --- | --- |
| A | -177.630,76 |
| b1 | 11.310,59 |
| b2 | 3,96 |
| b3 | 46,28 |

Tabel 3. Hasil Perhitungan Matriks



Gambar 6. Output perhitungan manual

**Pengujian Menggunakan Rapidminer**



Gambar 7. Antar muka halaman utama

Menghubungkan panel susunan operator dapat dilihat pada gambar dibawah dan akan dijelaskan fungsi-fungsinya, panel operators retrieve yang mana telah dijelaskan sebelumnya berfungsi untuk memproses objek data, liniear regression yaitu model algoritma yang berfungsi untuk menghasilkan model berdasarkan algoritma yang telah dipilih yaitu liniear regression, apply model yang dimana panel operator ini berfungsi untuk menerapkan model yang digunakan.



Gambar 8. Panel Operators

Setelah semua panel operator terhubung dengan mengklik icon play maka akan muncul tab result hasi prediksi dengan jumlah produksi yang diprediksikan dalam 3 bulan kedepan yakni meliputi bulan April diprediksikan menghasilkan jumlah produksi sebesar : 350862.531, bulan Mei sebesar : 304206.794, Jun sebesar : 389668.603.



Gambar 9. Hasil Prediksi

Dapat dilihat pada gambar dibawah ini tampilan nilai error, yang dimana factor yang mempengaruhi produksi kelapa sawit di PT Borneo Ketapang Indah adalah Luas Lahan dan Curah Hujan dengan nilai Coefficient masing-masing variable 0,646 variable X1 (Luas Lahan), 0,530 variable X2 (Curah Hujan).



Gambar 10. Tampilan Nilai *error*

**5. KESIMPULAN**

Berdasarkan hasil dan pembahasan penelitian mengenai Penerapan Data Mining Dalam Prediksi Hasil Produksi Kelapa Sawit Pada PT Borneo Ketapang Indah (BKI) dengan menggunakan proses data mining dengan metode Linear Regression menghasilkan kesimpulan sebagai berikut :

1. Metode Linear Regression dapat diterapkan untuk memprediksi faktor apa saja yang berpengaruh pada hasil produksi kelapa sawit PT. BORNEO Ketapang Indah (BKI) dengan menggunakan hasil produksi tahun 2017-2019.

2. Model diuji dengan menggunakan tools Rapidminer dengan membagi data Testing dan Training. Data training di ambil dari tahun 2017-2018, sedangkan data testing diambil dari data tahun 2019. Dari data yang telah diuji menggunakan tools rapidminer maka dapat disimpulkan faktor-faktor apa saja yang mempengaruhi hasil produksi kelapa sawit menggunakan Algoritma Liniear Regression.

3. Hasil prediksi jumlah produksi kelapa sawit pada PT Borneo Ketapang Indah dalam 3 bulan kedepan meliputi bulan April adalah sebesar : 350862.531, Mei : 304206.794, dan Juni sebesar : 389668.603.

4. Dari tiga variable yang digunakan untuk melakukan prediksi jumlah produksi kelapa sawit, meliputi Luas Lahan(X1), Curah Hujan(X2) dan Pupuk(X3). Maka variable Curah Hujan dan Luas lahan yang paling berpengaruh pada jumlah produksi kelapa sawit dengan masing-masing nilai coefficient Luas Lahan (X1) sebesar 0,646 dan Curah Hujan (X2) sebesar 0,530. Dengan perhitungan metode Linier Regression dapat menjadi solusi bagi perusahaan dalam mengatasi masalah ketidaksatabilan dalam hasil produksi kelapa sawit.

**6. UCAPAN TERIMA KASIH**

Ucapan terima kasih sebesar-besarnya kepada PT Borneo Ketapang Indah yang telah mengizinkan saya untuk melakukan penelitian, serta semua pihak yang turut berperan dalam pelaksanaan penelitian.

**DAFTAR PUSTAKA**

Ayuni, G. N., & Fitrianah, D. (2020). *Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ*. *14 no*.

Harlan, J. (2018). *Analisis Regresi Linier*. Gunadarma.

Herwanto, H. W., Widiyaningtyas, T., & Indriana, P. (2019). *Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi*. *8*, *No. 4,*.

Hermawati, & Fajar, A. (2013). *Data Mining*. Penerbit Andi.

Kurniadi, A., Jasmir, & Novianto, Y. (2020). *Penerapan Metode Regresi Linier untuk Memprediksi Kebiasaan Pelanggan Studi Kasus: PT. Mensa Binasukses*. *2*, *No.2*, *J*.

Kurniawan, D. (2008). *REGRESSI LINIEAR*.

Kusrini, & Taufiq, L. E. (2009). *Algoritma Data Mining*. Andi:Yogyakarta.

Larose, D. T. (2005). *Discovering Knowledge in Data : An Itroduction to Data Mining*.

Nurzaman, F. (2017). *PENERAPAN ALGORITMA REGRESI LINIER UNTUK PREDIKSI JUMLAH KLAIM PADAASURANSI KESEHATAN*.

Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). *Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda*. *No. 1*, *Feb*.

Pramudiono, I. (2007). *engantar Data Mining: Menambang Permata Pengetahuan di Gunung Data.* http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip.

Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis Teori dan Aplikasi*. Graha Ilmu Yogyakarta.

Siradjuddin, I. (2018). *Machine Learning: Liniear Regression*. Politeknik Negeri Malang.

Setyamodjaja, D. (2006). *Kelapa sawit : teknik budi daya, panen dan pengolahan / Djoehana Setyamidjaja* (Ed. rev.,). Jakarta : Kanisius, 2006.