Jurnal Ilmu Teknik dan Komputer
Vol. 6 No. 2 Juli 2022
ISSN 2548-740X
E-ISSN 2621-1491

Komparasi Algoritma Regresi Linear dan Algoritma C 4.5 Untuk Memprediksi Penjualan Sayur Mayur (PT. Kebun Sayur Segar)

Muhamad Fahri Naufal*, Vina Ayumi

Teknik Informatika, Universitas Mercu Buana, Jakarta
Jl. Meruya Selatan No. 1, Kembangan, Jakarta Barat
*fahrinaufal18@gmail.com, vina.ayumi@mercubuana.ac.id

ABSTRACT

Sayur-mayur adalah makanan pokok selain nasi yang sering dikonsumsi sehari-hari. Sayur-mayur juga makanan sehat yang mengandung kadar air tinggi, vitamin dan serat yang banyak sehingga baik untuk tubuh. Adapun contoh sayur yang sering di konsumsi sehari-hari seperti Bayam yang mengandung vitamin A, B, C dan E, ada juga Wortel yang merupakan sayuran yang baik bagi mata karena mengandung vitamin A, dan masih banyak lagi sayur yang banyak mengandung vitamin dan baik untuk tubuh. PT. Kebun Sayur Segar menanam sendiri sayur-mayur yang di jual sehingga kondisi sayur sampainya diprodusen masih segar, namun sering kali ada beberapa sayur yang layu dan busuk dikarenakan stok terlalu banyak. Hal ini mengakibatkan kerugian untuk PT. Kebun Sayur Segar. Oleh karena itu dalam penelitian ini peneliti ingin melakukan prediksi terhadap penjualan sayuran pada PT. Kebun Sayur Segar untuk dapat memprediksi terhadap stok sayur kedepannya untuk dapat mengurangi kerugian oleh karena itu dibutuhkan prediksi dengan menggunakan algoritma regresi linear dan algoritma C4.5 untuk memprediksi. Penelitian ini mendapatkan hasil dari grafik penjualan sayur-mayur yaitu jenis sayur yang diminati customer adalah hydroponic. Bedasarkan dari pengujian dari semua percobaan ini dimana pengujian menggunakan parameter MAE,MSE,R2 Score dengan hasil pengujian regresi linear, Bedasarkan parameter yang dipakai mendapatkan hasil akurasi nilai standar, dimana algoritma regresi linear mendapatkan akurasi parameter MAE 42762.92, MSE 14888137878.77 dan R2 Score 0.64. Untuk perhitungan menggunakan algoritma C 4.5 dengan parameter MAE 32729.73, MSE 15607295869.63, dan R2 Score 0.62.

Kata Kunci: prediksi, algoritma regresi linear, algoritma C4.5

PENDAHULUAN

Sayur-mayur adalah makanan pokok selain nasi yang sering dikonsumsi sehari-hari. Sayur-mayur juga makanan sehat yang mengandung kadar air tinggi, vitamin dan serat yang banyak sehingga baik untuk tubuh. Adapun contoh sayur yang sering di konsumsi sehari-hari seperti Bayam yang mengandung vitamin A, B, C dan E, ada juga Wortel yang merupakan sayuran yang baik bagi mata karena mengandung vitamin A, dan masih banyak lagi sayur yang banyak mengandung vitamin dan baik untuk tubuh.

Penjualan Sayur di Indonesia sangat banyak dikarenakan indonesia adalah negara tropis yang dimana sayur-mayur tumbuh dengan baik di indonesia. Salah satu distributor sayur-mayur yang banyak menjual ke beberapa produsen yaitu PT. Kebun Sayur Segar. PT. Kebun Sayur Segar menanam sendiri sayur-mayur yang di jual sehingga kondisi sayur sampainya diprodusen masih segar, namun sering kali ada beberapa sayur yang layu dan busuk dikarenakan stok terlalu banyak. Hal ini mengakibatkan kerugian untuk PT. Kebun Sayur Segar. Oleh karena itu dalam penelitian ini peneliti ingin melakukan prediksi terhadap penjualan sayuran pada PT. Kebun Sayur Segar untuk dapat memprediksi terhadap stok sayur kedepannya untuk dapat mengurangi kerugian.

Pada penelitian sebelumnya yang berjudul "Penerapan Data Mining Untuk Memprediksi Penjualan Kain Tenun Mnggunakan Regresi Linear" bahwa dari hasil prediksi pada salah satu produk penjualan didapatkan pada bulan januari jumla penjualan produk Rang sebanyak 12 lebih produk yang akan terjual, sedang dari bulan pebruari sampai bulan agustus sebanyak 26 lebih produk rang yang akan terjual serta dari bulan september sampai bulan

desember sebanyak 25 lebih produk yang akan terjual(Firda Widiastutia, Wafiah Murniatib, 2022).

Pada penelitian lainnya yang berjudul "Analisa Algoritma C 4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri" bahwa analisis prediksi penjualan menggunakan algoritma C 4.5 dapat dipergunakan untuk memprediksi penjualan obat pertanian di periode yang akan datang pada Toko Dewi Sri berdasarkan data penjualan dari bulan Oktober sampai bulan November tahun 2019. Dengan adanya sistem prediksi penjualan menggunakan algoritma C 4.5 dapat memperkecil terjadinya kesalahan dalam menentukan stok obat untuk periode berikutnya(Rosita Dewi & Farouq Mauladi, 2020).

STUDI LITERATUR

1. Prediksi

Prediksi adalah sebuah ramalan untuk mendapatkan informasi yang akan terjadi di masa depan tentang suatu peristiwa. Prediksi menunjukkan informasi apa yang akan terjadi pada suatu keadaan atau situasi tertentu dengan probalitas kejadian terbesar untuk menjadi masukan atau input untuk proses perencanaan keputusan. Proses prediksi bisa dengan cara kualitatif yaitu melalui pendapat ahli atau juga dengan cara kuantitatif dengan perhitungan matematis. Salah satu metode prediksi kuantitatif adalah menggunakan analisis deret waktu (time series)(Wanto & Windarto, 2017).

2. Algoritma Regresi Linear

Metode Regresi Linear ini biasa digunakan untuk melakuan prediksi secara matematis melalui garis lurus antara, variabel independen (x) dan dependen (y). Metode Regresi Linear ini juga salah satu metodek statistik untuk digunakan sebagai produksi untuk melakukan peramalan atau prediksi tentang karakteristik kualiatas ataupun kuantitas(Ginting et al., 2019).

3. Algoritma C 4.5

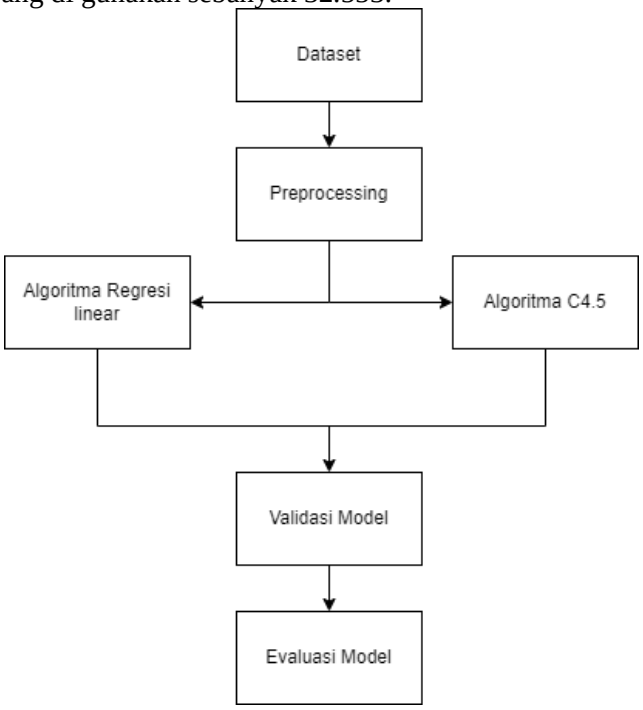
Algoritma C 4.5 adalah algoritma yang digunakan untuk membentuk pohon keputusan yang menyerupai seperti sebuah flowchart, dimana masing-masing internal node-nya akan dinyatakan menjadi sebuah atribut pengujian, setiap cabang mewakili output dari pengujian, dan setiap node daun (terminal node) menentukan label class. Node paling atas dari sebuah pohon adalah node akar(Septiani, 2017).

METODOLOGI

Metode pada penelitian ini adalah sebagai berikut,

1. Dataset

Data yang digunakan adalah *dataset* yang di peroleh dari PT. Kebun Sayur Segar pada bulan Juni sampai dengan bulan September, dengan total *dataset* yang di gunakan sebanyak 32.553.



Gambar 1. Proses modeling

2. Preprocessing Data

Pada proses *data cleaning* ini dengan jumlah *dataset* 32.553 peneliti menghapus seluruh data yang tidak di gunakan pada kolom yang kosong. Dan pada tahapan ini setelah di lakukan *data cleaning* di peroleh 30.751 data.

3. Implementasi Algoritma

a. Algoritma Regresi Linear

Regresi Linier merupakan cara untuk mengukur hubungan korelasi antara dua variabel atau lebih yang digunakan untuk data prediksi melalui garis lurus(Suryanto, 2019).

b. Algoritma C4.5

C 4.5 adalah Algoritma yang digunakan untuk membentuk sebuah pohon keputusan. Pohon keputusan merupakan metodeklasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusa mengubah fakta yang sangat besar menjadi pohon keputusanyang merepresentasikan aturan. Aturan dapat dengan mudah dipahamidengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentukbahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu(Eska, 2018).

4. Validasi Model

Sebuah proses tahapan untuk mengambil keputusan untuk mencari model terpercaya yang di gunakan sebagai bagian pengambil sebuah keputusan, untuk membandingkan *output* ukuran kinerja pada simulasi

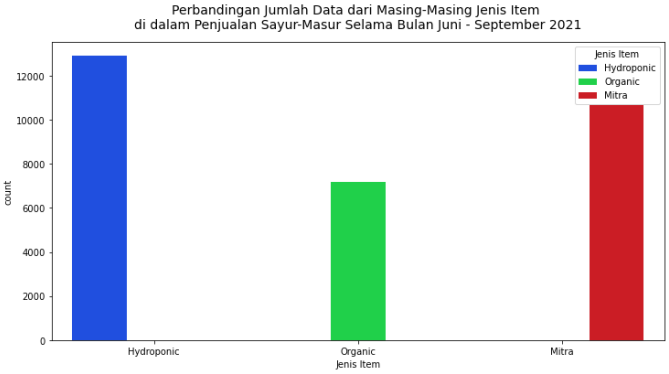
dengan ukuran yang sesuai oleh sistem dan dapat melihat nilai – nilai sampingan atau variabel.

5. Evaluasi Model

Metode ini untuk menganalisa dan mengukur sejauh mana keakuratan hasil yang telah dicapai oleh model dengan menggunakan evaluasi akan diberikan, fase yang dilakukan menggunakan perbandingan kuantitatif yang terdiri dari MSE, MAE dan R2 Score

HASIL DAN PEMBAHASAN

Berikut ini adalah grafik dari jenis item pada penjualan sayur-mayur selama bulan juni – september 2021.

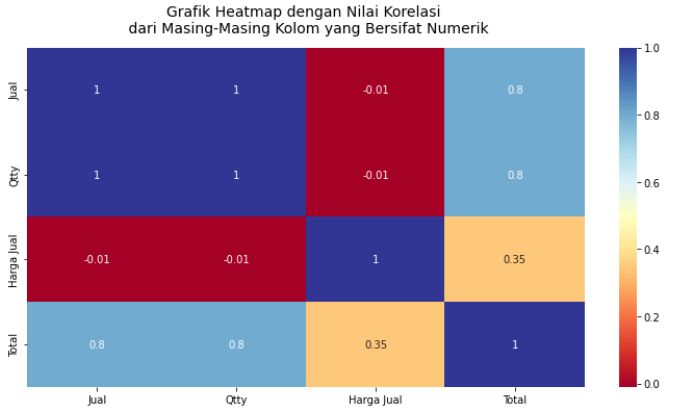


Gambar 2. Grafik Jenis Item Penjualan

Berdasarkan hasil grafik batang diatas dapat diketahui bahwasanya menunjukkan perbandingan data kategorik di dalam kolom "Jeni Item" yang ada pada "data_penjualan_sayur_mayur" yang dijual selama bulan Juni - September 2021. Harus diingat bahwa grafik batang diatas hanya menunjukkan nilai total dari masing-masing data yang ada di kolom "Jenis Item" tanpa adanya pengaruh dari variabel/kolom lain. Berdasarkan grafik diatas saya dapat mengambil kesimpulan bahwasanya jenis item hidroponik adalah yang peling tertinggi jumlah nilainya dari pada jenis item organik maupun mitra.

Nilai total yang ditunjukkan pada bagian grafik hidroponik adalah lebih dari 12.000 data. Selanjutnya adalah mitra yang memiliki nilai total terbanyak setelah hidroponik yakni sebanyak lebih dari 10.000 data sedangkan jenis item organik memiliki nilai total paling sedikit diantara kedua jenis item sebelumnya (hidroponik dan mitra) yang memiliki nilai total yang lebih dari 6.000 data atau mungkin tepatnya berada disekitar range nilai total yakni 7.000 - 7.190'an.

Bedasarkan dari grafik *heatmap* pada Gambar 3, melihat hubungan korelasi atau hubungan timbal balik dari antar masing kolom-kolom.



Gambar 3. Grafik *Heatmap*

1. Variabel "Jual" yang ada di SUMBU Y memiliki nilai korelasi dengan variabel "Harga Jual" yang ada di SUMBU X sebesar 0.0011 yang dimana kedua variabel ini tidak terlalu memiliki hubungan timbal-balik yang bagus bahkan cenderung tidak memiliki hubungan timbal-balik.
2. Variabel "Harga Jual" yang ada di SUMBU Y memiliki nilai korelasi dengan variabel "Total" yang ada di SUMBU X adalah 0.32 yang dimana kedua variabel ini memiliki hubungan timbal balik yang cukup bagus.
3. Dan contoh terakhir adalah variabel "Total" yang ada di SUMBU Y yang memiliki nilai korelasi atau hubungan timbal-balik dengan variabel "Jual" yang ada di SUMBU X adalah sebesar 0.86 yang dimana kedua variabel ini (antara Total dengan Jual atau sebaliknya) memiliki hubungan timbal balik yang sangat bagus yang dibuktikan dengan nilai korelasi antar masing-masing variabel.

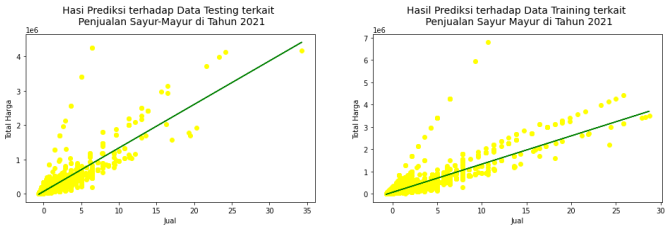
Pada penelitian ini akan melakukan eksperimen yang terdiri dari, data *training* sebesar 70% dan data *testing* sebesar 30%. Pada pengujian yang dilakukan menggunakan parameter MAE (*Mean absolute error*), MSE (*Mean Squared error*) dan R2 *Score*.

a. Algoritma Regresi Linier

Dari tabel 1 peneliti dapat hasil dari pengujian dan pelatihan, di dapatkan hasil MAE pada *data testing* sebesar 42762.92 dan *pada data training* sebesar 41846.53. Sedangkan hasil dari MSE pada data testing sebesar 14888137878.77 dan pada *data training* sebesar 17993298320.13. Terakhir hasil dari R2 *Score* 0.64 untuk *data testing* dan 0.64 pada *data training*.

Tabel 1. Hasil Akurasi MAE,MSE,dan R2 *Score* Algoritma Regresi Linear

	MAE	MSE	R2 *Score*
Data Testing	42762.92	14888137878.77	0.64
Data Training	41846.53	17993298320.13	0.64



Gambar 4. Grafik Regresi Linear

Gambar 4 adalah hasil dari *data testing* & *data training*, keduanya memiliki hubungan yang linear serta juga memiliki hubungan yang positif karena data-data yang ada di dalam variabel variabel dependen (kolom "Total") mengalami peningkatan yang disertai dengan nilai yang ada di dalam variabel variabel independen (kolom "Jual").

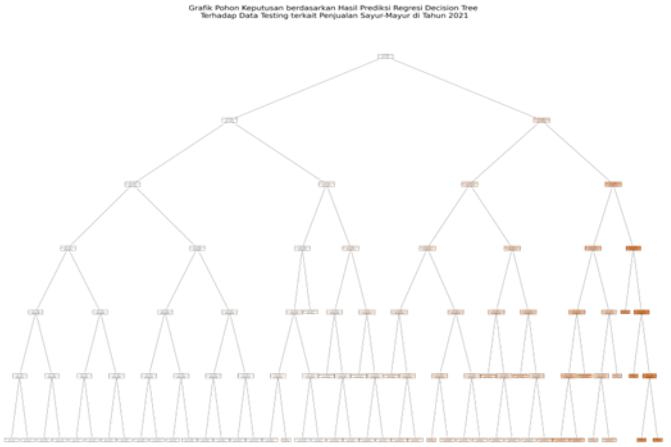
Meskipun kedua grafik diatas linear (karena garis bertambah nilainya),tetapi terdapat perbedaan yang cukup jelas seperti dalam grafik pada *data testing* yang garisnya mengalami peningkatan secara lurus keatas sedangkan pada grafik di data training, garisnya memang mengalami peningkatan tetapi tidak meningkat secara lurus keatas. Selain perbedaan pada garis juga terdapat perbedaan dalam persebaran titik data yang dimana pada hasil prediksi di data testing, persebaran datanya cukup menyebar sedangkan persebaran data di *data training* bersifat lebih berkumpul dari pada yang ada di grafik linear *data testing*.

b. Algoritma C 4.5

Dari tabel 2 peneliti dapat hasil dari pengujian dan pelatihan, di dapatkan hasil MAE pada *data testing* sebesar 32729.73 dan pada *data training* sebesar30106.84. Sedangkan hasil dari MSE pada *data testing* sebesar 15607295869.63 dan pada *data training* sebesar 17691258678.72. Terakhir hasil dari R2 *Score* 0.62 untuk *data testing* dan 0.65 pada *data training*.

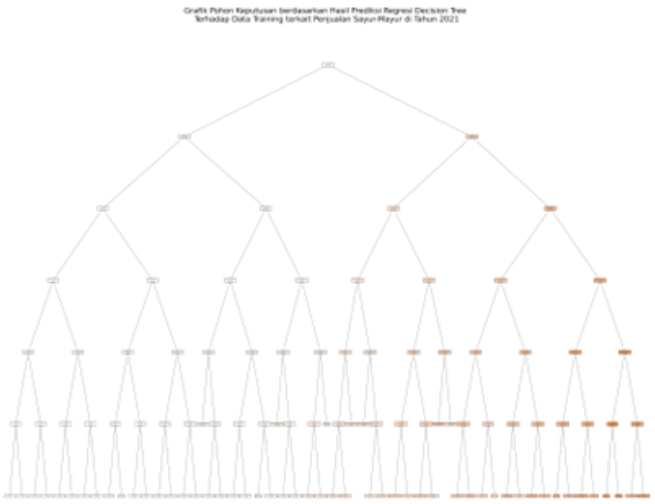
Tabel 2. Hasil Akurasi MAE,MSE,dan R2 *Score* Algoritma C 4.5

	MAE	MSE	R2 *Score*
Data Testing	32729.73	15607295869.63	0.62
Data Training	30106.84	17691258678.72	0.65



Gambar 5. Pohon Keputusan Data Training

Dari gambar 5 visualisasi pohon keputusan untuk kebutuhan penjualan sayur mayur di ketahui bahwa, atribut nilai *gain* tertinggi sebesar 121142.138 dikatakan sebagi *root node*, untuk nilai *spliting* dengan 6089 samples dengan hasil sebesar 96985.846, *decision node* terhadap *data testing* untuk nilai jual dengan 4360 samples hasil yang didaptkan 58971.423, *terminal node* dengan 2 samples dengan menghasilakn nilai 2500.0



Gambar 6. Pohon Keputusan Data Testing

Dari gambar 6 visualisasi pohon keputusan untuk kebutuhan penjualan sayur mayur di ketahui bahwa, atribut nilai *gain* tertinggi sebesar 119356.319 dikatakan sebagi *root node*, untuk nilai spliting dengan 2613 samples dengan hasil sebesar 97586.272, *decision node* terhadap data testing untuk nilai jual dengan 2271 samples hasil yang didaptkan 73113.178, *terminal node* dengan 3 samples dengan menghasilakn nilai 63000.0

KESIMPULAN

Bedasarkan haasil dari penelitian yang telah dilakukan untuk memprediksi penjualan sayur-mayur terdapat grafik dari penjualan yang paling diminati oleh *customer* pada bulan juni-september 2021 yaitu jenis sayur *hydroponic*, dari penjualan jenis sayur *hydroponic* terdapat 12.000 data pembelian. Penelitian ini juga menguji dari semua percobaan ini dimana pengujian menggunakan parameter MAE,MSE,R2 Score dengan hasil pengujian regresi linear yang standar. Bedasarkan parameter yang dipakai mendapatkan hasil akurasi nilai standar, dimana algoritma regresi linear mendapatkan akurasi parameter MAE 42762.92, MSE 14888137878.77 dan R2 Score 0.64. Untuk perhitungan menggunakan algoritma C 4.5 dengan parameter MAE 32729.73, MSE 15607295869.63, dan R2 Score 0.62. Maka dapat disimpulkan bahwa algoritma Regresi linear mendapatkan nilai akurasi lebih baik untuk memperkuat penjualan sayur-mayur.

DAFTAR PUSTAKA

Eska, J. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5*. *2*. https://doi.org/10.31227/osf.io/x6svc

Firda Widiastutia, Wafiah Murniatib, S. (2022). *Penerapan Data Mining Untuk Memprediksi Penjualan Kain Tenun Mnggunakan Regresi Linear*. *2*(1).

Ginting, F., Buulolo, E., & Siagian, E. R. (2019). Implementasi Algoritma Regresi Linear Sederhana Dalam Memprediksi Besaran Pendapatan Daerah (Studi Kasus: Dinas Pendapatan Kab. Deli Serdang). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *3*(1), 274–279. https://doi.org/10.30865/komik.v3i1.1602

Rosita Dewi, K., & Farouq Mauladi, K. (2020). Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri. *Seminar Nasional Inovasi Teknologi*, *25*(2020), 2580–3336.

Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, *13*(1), 76–84. https://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/149/126

Suryanto, A. A. (2019). Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi. *Saintekbu*, *11*(1), 78–83. https://doi.org/10.32764/saintekbu.v11i1.298

Wanto, A., & Windarto, A. P. (2017). Analisis Prediksi Indeks Harga Konsumen Berdasarkan Kelompok Kesehatan Dengan Menggunakan Metode Backpropagation. *Jurnal & Penelitian Teknik Informatika Sinkron*, *2*(2), 37–43. https://zenodo.org/record/1009223#.Wd7norlTbhQ