
ANALISIS MULTIVARIATE ADAPTIVE REGRESSION SPLINES DALAM PREDIKSI PENCAPAIAN TARGET PROFIT PERUSAHAAN DISTRIBUTOR

Suparmadi¹, Rohminatin², Rahmadita³
Sistem Informasi, Universitas Royal, Kisaran
e-mail: ¹suparmadi43@gmail.com

Abstract: *An objective performance evaluation process that can describe the company's condition is through the achievement of targets from each factor that influences profit achievement. The many variables that influence this profit achievement require in-depth and systematic analysis to obtain complete, fast and accurate information. However, so far the Gerai Mustika distributor company has only focused on sales factors, and ignored other factors. Furthermore, the large amount of financial data that has been computerized so far at Gerai Mustika is only limited to weekly, monthly and annual reports, has not been optimally utilized to help management evaluate performance. The study aims to build a prediction model for achieving profit targets from company financial data using Multivariate Adaptive Regression Splines (MARS) based on machine learning. Based on the analysis results, the MARS model is proven to be superior to conventional multiple regression in predicting Gerai Mustika's profit, indicated by a higher R-squared value (0.756), as well as lower MAE and MAPE, so it is able to capture more complex data patterns and provide more accurate predictions; While multiple regression is still relevant for simple linear relationships, it is less than optimal in representing non-linear data variations.*

Keywords: *Prediction; MARS Method; Distributor; Profit; Machine Learning*

Abstrak: Proses evaluasi kinerja yang objektif yang dapat menggambarkan kondisi perusahaan yaitu melalui pencapaian target-target dari setiap faktor yang mempengaruhi pencapaian profit. Banyaknya variabel yang mempengaruhi pencapaian profit ini maka dibutuhkan analisa yang mendalam dan sistematis agar mendapatkan informasi yang utuh, cepat dan akurat. Akan tetapi, selama ini perusahaan distributor Gerai Mustika hanya fokus pada faktor penjualan saja, dan mengabaikan faktor yang lain. Selanjutnya, jumlah data keuangan yang besar yang telah terkomputerisasi selama ini di Gerai Mustika baru sebatas untuk laporan mingguan, bulanan dan tahunan saja, belum dimanfaatkan secara optimal untuk membantu manajemen mengevaluasi kinerja. Penelitian bertujuan untuk membangun model prediksi pencapaian target profit dari data keuangan perusahaan menggunakan Multivariate Adaptive Regression Splines (MARS) berbasis machine learning. Berdasarkan hasil analisis, model MARS terbukti lebih unggul dibandingkan regresi berganda konvensional dalam memprediksi profit Gerai Mustika, ditunjukkan oleh nilai R-squared yang lebih tinggi (0,756), serta MAE dan MAPE yang lebih rendah, sehingga mampu menangkap pola data yang lebih kompleks dan memberikan prediksi yang lebih akurat; sementara regresi berganda masih relevan untuk hubungan linear sederhana, namun kurang optimal dalam merepresentasikan variasi data yang non-linear.

Kata kunci: *Prediksi;Metode_MARS;Distributor;Profit;Machine_Learning*

PENDAHULUAN

Pemanfaatan kemajuan teknologi big data sangat penting bagi perusahaan distributor, yang bertujuan untuk

mencapai profit. Manajemen sering menetapkan target profit sebagai ukuran kinerja karyawan. Evaluasi kinerja perusahaan dapat dilakukan melalui laporan keuangan dari perhitungan

laba/rugi, di mana keuntungan bersih (profit) menjadi indikator utama (Puspita et al., 2023). Perhitungan laba/rugi melibatkan beberapa variabel seperti penjualan, pembelian, persediaan awal, persediaan akhir, retur penjualan, retur pembelian, dan biaya operasional (Adiyanto, 2023). Karena banyaknya variabel yang mempengaruhi hasil perhitungan tersebut, diperlukan analisis mendalam untuk mendapatkan informasi yang lengkap. Banyaknya variabel yang mempengaruhi profit tersebut, maka dibutuhkan analisa yang mendalam agar mendapatkan informasi yang utuh. Prediksi dibutuhkan agar manajemen dapat segera mengambil langkah strategis untuk memperbaiki kinerja sebelum terjadinya resiko.

Penelitian ini dilakukan di perusahaan distributor pakaian Gerai Mustika. Peneliti menemukan masalah dalam menganalisis pencapaian target profit pada laporan laba/rugi, di mana evaluasi hanya terfokus pada omset penjualan. Manajemen percaya bahwa mencapai target keuntungan hanya bergantung pada tingginya omset penjualan. **Akan tetapi**, fokus hanya pada omset penjualan dapat mengabaikan variabel lain yang juga memengaruhi profit. Akibatnya, penyebab pencapaian atau ketidakcapaian target profit tidak dapat dipahami secara menyeluruh. Hal ini

METODE

Pendekatan pemecahan masalah dimulai pemahaman masalah yaitu kesulitan dalam menganalisis dan memprediksi pencapaian target profit menggunakan laporan keuangan. Selanjutnya, identifikasi variabel dengan mengidentifikasi variabel-variabel yang relevan yang mempengaruhi pencapaian target profit. Kemudian pengumpulan data keuangan yang diperlukan dari perusahaan distributor. Dilanjutkan dengan praproses data dengan membersihkan dan mengatasi outlier. Untuk mengatasi outlier dalam penelitian ini, dataset akan dilakukan normalisasi

berpotensi mengurangi keuntungan perusahaan bahkan dapat menyebabkan kerugian yang dapat mengarah pada kebangkrutan. Selanjutnya, jumlah data keuangan yang besar dan telah terkomputerisasi selama ini di Gerai Mustika baru sebatas untuk laporan mingguan, bulanan dan tahunan saja, belum dimanfaatkan secara optimal untuk membantu manajemen mengevaluasi kinerja.

Berangkat dari permasalahan di atas, penelitian ini bertujuan untuk membangun model prediksi pencapaian target profit melalui penerapan teknologi big data berbasis machine learning dengan metode Multivariate Adaptive Regression Splines (MARS). Metode MARS adalah metode statistik yang digunakan untuk regresi yang fleksibel dan adaptif terhadap hubungan antara variabel dependen dan independen dengan kombinasi dari fungsi-fungsi dasar sederhana yang disebut spline (Abed et al., 2023; Adriyanov & Advisor, 2020; Bekar Adiguzel & Cengiz, 2023). Metode MARS dipilih karena secara otomatis dapat menangkap interaksi antara variabel, yang memungkinkan model untuk memperhitungkan efek gabungan dari beberapa variabel input yang hasilnya seringkali mudah diinterpretasikan oleh pengguna (Huang et al., 2020).

data untuk mengurangi dampak outlier yang dapat menyebabkan model prediksi menjadi tidak stabil atau bahkan menyimpang menggunakan Z-Score. Z-score adalah sebuah metode statistik yang digunakan untuk menormalkan atau standarisasi data (JM Chinawa, AT Chinawa, FA Ujunwa, CL Odimegwu, JT Onyia, K Chukwu, BF Chukwu, Aronu AE, 2019; Singh & Singh, 2020). Selanjutnya pemodelan dengan MARS untuk membangun model prediksi pencapaian target profit yang melibatkan proses penyesuaian model ke data yang ada dan mengoptimalkan parameter-parameter model untuk mendapatkan hasil yang terbaik. Selanjutnya melakukan validasi model untuk memastikan bahwa

model yang dikembangkan dapat memberikan prediksi yang akurat dan konsisten. Kemudian dilakukan interpretasi hasil untuk memahami faktor-faktor atau gabungan faktor yang paling berpengaruh terhadap pencapaian target

profit dan memberikan wawasan kepada manajemen perusahaan distributor. Kemudian implementasi dalam praktik perusahaan untuk membantu dalam evaluasi kinerja perusahaan dan pengambilan keputusan.

HASIL DAN PEMBAHASAN

Berdasarkan hasil pengumpulan data didapati hasil dataset dengan variabel:

1. Periode

- Menunjukkan bulan dan tahun pengamatan, mulai dari Januari 2020 sampai Desember 2024.
- Formatnya menunjukkan bulan dan tahun (misal Jan-20, Feb-20, dst.).

2. Penjualan

- Nilai penjualan dalam satuan tertentu (kemungkinan dalam juta atau ribuan, tergantung konteks data).
- Nilai berkisar dari sekitar 981.5 sampai 1963.4, dengan tren meningkat dari tahun ke tahun.

3. Pembelian

- Nilai pembelian dalam satuan yang sama dengan penjualan.
- Nilai bervariasi dari sekitar 990.4 sampai 1609.9, juga menunjukkan tren peningkatan secara umum.

4. Biaya

- Nilai biaya operasional dan lain-lain, berkisar dari sekitar 47.4 sampai 67.4.
- Nilai ini relatif lebih stabil dengan variasi kecil, menunjukkan pengeluaran yang terkendali.

5. Stok Gudang

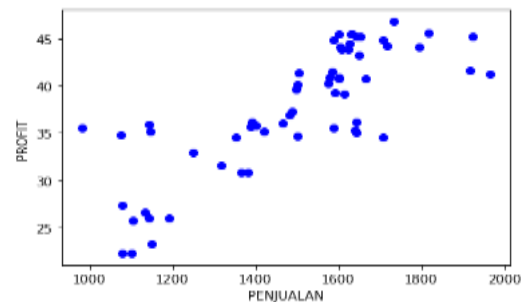
- Jumlah stok gudang yang tersedia dalam satuan tertentu.
- Rata-rata berkisar di sekitar 300-an sampai 900-an, menunjukkan variasi seiring perubahan penjualan dan pembelian.

6. Profit

- Nilai keuntungan bersih dalam satuan tertentu.
- Berkisar dari sekitar 22.3 sampai 46.8, dengan peningkatan secara

progresif selama periode pengamatan.
- Beberapa bulan menunjukkan profit tertinggi di atas 45, terutama pada tahun terakhir dan bulan-

Visualisasi Data

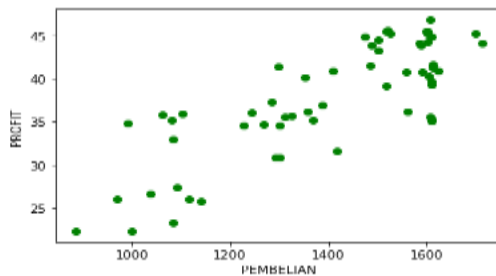


Gambar 1. Visualisasi data Scatter Penjualan

Visualisasi scatter plot penjualan ini menampilkan sebaran titik data berwarna biru, di mana sumbu-X merepresentasikan variabel independen (Penjualan) dan sumbu-Y (Profit) merepresentasikan nilai penjualan.

Dari pola titik-titik tersebut, terlihat:

1. Sebagian besar titik membentuk kecenderungan naik: nilai penjualan cenderung meningkat seiring bertambahnya nilai pada sumbu-X.
2. Terdapat beberapa kelompok (cluster) data, terutama pada nilai penjualan tinggi, yang menunjukkan adanya periode atau kondisi tertentu dengan penjualan yang konsisten tinggi.
3. Beberapa titik berada jauh dari pola utama, yang berpotensi merupakan *outlier* atau kejadian penjualan yang tidak biasa.



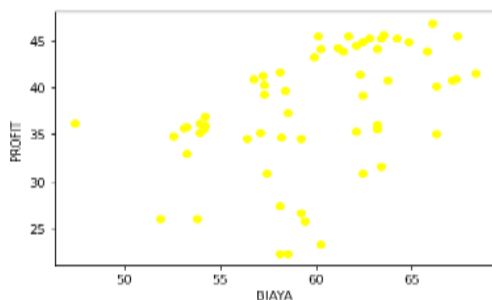
Gambar 2. Visualisasi data Scatter Pembelian

Visualisasi scatter plot pembelian ini menampilkan titik-titik data berwarna hijau yang merepresentasikan hubungan antara variabel pada sumbu-X (Pembelian) dengan variabel pada sumbu-Y (Profit).

Dari pola sebaran tersebut terlihat bahwa:

1. Terdapat kecenderungan hubungan positif: nilai pembelian cenderung meningkat seiring bertambahnya nilai pada sumbu-X.
2. Sebaran data membentuk beberapa kelompok (*cluster*), terutama pada nilai pembelian menengah hingga tinggi, yang dapat menunjukkan adanya segmen pasar atau periode dengan pola pembelian serupa.
3. Meskipun sebagian besar titik mengikuti pola naik, ada juga titik-titik yang menyebar di luar tren utama, yang bisa menjadi indikasi variasi perilaku pembelian atau faktor eksternal yang memengaruhi.

Secara keseluruhan, scatter plot ini menunjukkan bahwa variabel pada sumbu-X berpengaruh terhadap jumlah pembelian, dengan indikasi korelasi positif yang cukup jelas.

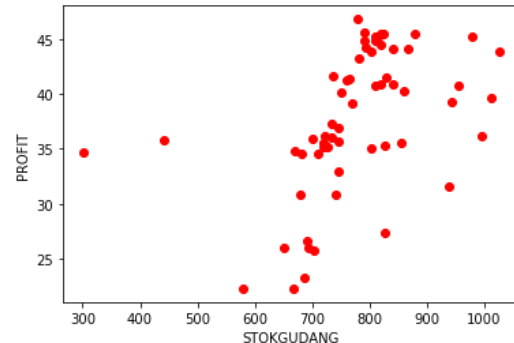


Gambar 3. Visualisasi data Scatter Biaya

Visualisasi scatter plot biaya ini menampilkan sebaran titik data berwarna kuning yang merepresentasikan hubungan antara variabel pada sumbu-X (biaya) dengan variabel pada sumbu-Y (Profit).

Berdasarkan pola sebaran:

1. Terdapat indikasi bahwa nilai biaya cenderung meningkat seiring bertambahnya nilai pada sumbu-X, meskipun penyebarannya cukup bervariasi.
2. Titik-titik tersebar dengan jarak yang berbeda, menunjukkan adanya variasi biaya pada kondisi yang mirip, yang bisa disebabkan oleh faktor eksternal atau perbedaan efisiensi.
3. Beberapa kelompok titik berada di area biaya tinggi, yang mengindikasikan adanya kondisi tertentu yang memicu peningkatan biaya secara signifikan.
4. Tidak semua titik mengikuti pola linear, sehingga kemungkinan ada faktor lain yang memengaruhi biaya selain variabel pada sumbu-X.



Gambar 4. Visualisasi data Scatter Stok Gudang

Visualisasi scatter plot stok gudang ini menampilkan sebaran titik data berwarna merah yang merepresentasikan hubungan antara variabel pada sumbu-X (stok gudang) dengan variabel pada sumbu-Y (Profit).

Berdasarkan pola sebaran:

1. Terlihat kecenderungan bahwa stok cenderung meningkat pada nilai sumbu-X tertentu, membentuk pola positif meskipun tidak sepenuhnya linear.

2. Titik-titik data banyak terkonsentrasi pada area stok tinggi, yang menunjukkan periode atau kondisi di mana gudang menyimpan barang dalam jumlah besar.
 3. Beberapa titik berada jauh dari kelompok utama, mengindikasikan adanya anomali atau situasi khusus seperti lonjakan stok atau penurunan drastis.
 4. Pola penyebaran juga memperlihatkan adanya beberapa kelompok (*cluster*), yang dapat mengindikasikan siklus persediaan atau strategi pengisian stok yang berbeda.
- Berikut adalah tabel perbandingan kinerja antara regresi berganda konvensional dan metode MARS berbasis machine learning.

Tabel 1. Perbandingan Performance Regresi dan MARS

Metode	MSE	RMSE	R-squared	MAE	MAPE (%)
Regresi Berganda Konvensional	4.58	4.276	0.520	3.58	9.58
Metode MARS	10.15	3.186	0.756	2.73	7.31

1. Regresi berganda memiliki nilai MSE yang lebih rendah (4.58) dibandingkan MARS (10.15), menunjukkan bahwa prediksi dari regresi konvensional lebih mendekati data nyata secara rata-rata kuadrat, meskipun RMSE dari MARS lebih kecil (3.19) dibandingkan RMSE regresi (4.28). Hal ini bisa terjadi karena distribusi kesalahan dan sensitivitas terhadap outlier berbeda.
2. Model MARS memiliki R-squared yang lebih tinggi (0.756) dibandingkan regresi konvensional (0.520), menandakan bahwa MARS mampu menjelaskan persentase variasi variabel target yang lebih besar, dan lebih baik dalam menangkap pola data.
3. MARS juga menunjukkan MAE dan MAPE yang lebih baik (lebih kecil), yang berarti secara rata-rata prediksi dari model ini lebih mendekati nilai aktual dan memberikan estimasi yang lebih akurat dalam persentase.

SIMPULAN

Model MARS menunjukkan performa prediksi yang lebih baik dibandingkan regresi berganda konvensional pada dataset ini, khususnya dalam hal kemampuan menjelaskan variasi data (R-squared) dan keakuratan prediksi (MAPE). Regresi berganda mungkin cukup baik secara linear, tetapi model MARS mampu menangkap hubungan yang lebih kompleks dan non-linear, menghasilkan performa prediksi yang lebih unggul secara umum.

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih yang setinggi-tingginya kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia (Kemendikbudristek) atas dukungan pendanaan melalui skema Penelitian Dasar Pemula yang disetujui dalam Program Penelitian Nasional Tahun 2025.

DAFTAR PUSTAKA

- Abed, M. S., Kadhim, F. J., Almusawi, J. K., Imran, H., Bernardo, L. F. A., & Henedy, S. N. (2023). Utilizing Multivariate Adaptive Regression Splines (MARS) for Precise Estimation of Soil Compaction Parameters. *Applied Sciences*, *13*(21), 11634. <https://doi.org/10.3390/app132111634>
- Adiyanto, M. R. (2023). PERENCANAAN LABA UMKM PASUNDAN Triana. *Jurnal Ekonomi Kreatif Dan Manajemen Bisnis Digital*, *2*(2), 210–228. Retrieved from <https://transpublika.co.id/ojs/index.php/JEKOMBITAL/article/view/543>
- Adriyanov, B., & Advisor, K. (2020). *Multivariate Adaptive Regression Splines (MARS) for non-linear time series dependence: An application in biostatistics*. *Interuniversity Master in Statistics and Operations Research*. Universitat Politècnica de Catalunya - BarcelonaTech. Retrieved from <https://upcommons.upc.edu/handle/2117/348071>
- Bekar Adiguzel, M., & Cengiz, M. A. (2023). Model selection in multivariate adaptive regressions splines (MARS) using alternative information criteria. *Heliyon*, *9*(9), e19964. <https://doi.org/10.1016/j.heliyon.2023.e19964>
- Gogineni, A., Rout, M. K. D., & Shubham, K. (2024). Evaluating machine learning algorithms for predicting compressive strength of concrete with mineral admixture using long short-term memory (LSTM) Technique. *Asian Journal of Civil Engineering*, *25*(2), 1921–1933. <https://doi.org/10.1007/s42107-023-00885-x>
- Guo, D., Chen, H., Tang, L., Chen, Z., & Samui, P. (2022). Assessment of rockburst risk using multivariate adaptive regression splines and deep forest model. *Acta Geotechnica*, *17*(4), 1183–1205. <https://doi.org/10.1007/s11440-021-01299-2>
- Huang, H., Ji, X., Xia, F., Huang, S., Shang, X., Chen, H., ... Mei, K. (2020). Multivariate adaptive regression splines for estimating riverine constituent concentrations. *Hydrological Processes*, *34*(5), 1213–1227. <https://doi.org/10.1002/hyp.13669>
- Imron, M. A., & Prasetyo, B. (2020). Improving Algorithm Accuracy K-Nearest Neighbor Using Z-Score Normalization and Particle Swarm Optimization to Predict Customer Churn. *Journal of Soft Computing Exploration*, *1*(1), 56–62. <https://doi.org/10.52465/josce.v1i1.7>
- JM Chinawa, AT Chinawa, FA Ujunwa, CL Odimegwu, JT Onyia, K Chukwu, BF Chukwu, Aronu AE, A. I. (2019). for an Uncommon Neurosurgical Emergency in a Developing Country. *Nigerian Journal of Clinical Practice*, *22*, 1070–1077. <https://doi.org/10.4103/njcp.njcp>
- Juwita, Safii, M., & Efendi Damanik, B. (2022). Algoritma Naïve Bayes Untuk Memprediksi Penjualan Pada Toko VJCakes Pematang Siantar. *Journal of Machine Learning and Artificial Intelligence*, *1*(4), 337–346. <https://doi.org/10.55123/jomlai.v1i4.1674>
- Kao, L. J., & Chiu, C. C. (2020). Application of integrated recurrent neural network with multivariate adaptive regression splines on SPC-EPC process. *Journal of Manufacturing Systems*, *57*(April), 109–118. <https://doi.org/10.1016/j.jmsy.2020.07.020>
- Mushtaq, K., Zou, R., Waris, A., Yang, K., Wang, J., Iqbal, J., & Jameel, M. (2023). Multivariate wind power

- curve modeling using multivariate adaptive regression splines and regression trees. *PLoS ONE*, 18(8 August), 1–25. <https://doi.org/10.1371/journal.pone.0290316>
- Prihanditya, H. A. (2020). The Implementation of Z-Score Normalization and Boosting Techniques to Increase Accuracy of C4.5 Algorithm in Diagnosing Chronic Kidney Disease. *Journal of Soft Computing Exploration*, 1(1), 63–69. <https://doi.org/10.52465/josce.v1i1.8>
- Priya Varshini, A. G., Anitha Kumari, K., Bhalavaishnavi, V. S., Kalpana Devi, C., & Pranesh, S. (2022). Machine Learning approach for Software Effort Estimation using Combination of Principal Component Regression and Neural Network. *Journal of Physics: Conference Series*, 2325(1). <https://doi.org/10.1088/1742-6596/2325/1/012049>
- Puspita et al. (2023). Analisis Rasio Profitabilitas dan Aktivitas Sebagai Dasar Penilaian Kinerja Keuangan pada PT HM Sampoerna , Tbk yang Terdaftar di Bursa Efek Indonesia Pesatnya kemajuan dari teknologi yang semakin hari semakin meningkat dengan drastis serta signifikan men. *Jurnal Media Wahana Ekonmika*, 19(4), 696–708. <https://doi.org/https://doi.org/10.31851/jmwe.v19i4.11033>
- Sembiring, M. A. (2019). Prediksi Kinerja Pencapaian Hasil Usaha Menggunakan Decision Tree. *Journal of Science and Social Research*, 2(August), 74–81. Retrieved from <https://jurnal.goretanpena.com/index.php/JSSR/article/view/378>
- Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97, 105524. <https://doi.org/10.1016/j.asoc.2019.105524>
- Suparmadi, S., & Ramadhani, A. (2022). Sistem Estimasi Pencapaian Target Profit Menggunakan Model Regresi Berbasis Machine Learning. *Journal of Science and Social Research*, 5(3), 703. <https://doi.org/10.54314/jssr.v5i3.1042>