

# ANALISIS REGRESI LINEAR DALAM DATA MINING UNTUK PREDIKSI SIJIL OFF DI KSOP KELAS I SAMARINDA

Fajri Harits Muzaki\*<sup>1</sup>  
Wawan Joko Pranoto<sup>2</sup>

<sup>1,2</sup>Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur,  
\*e-mail: [2011102441042@umkt.ac.id](mailto:2011102441042@umkt.ac.id)<sup>1</sup>, [wjp337@umkt.ac.id](mailto:wjp337@umkt.ac.id)<sup>2</sup>

## Abstrak

Penelitian ini bertujuan untuk menganalisis penerapan regresi linear dalam data mining sebagai metode prediksi untuk Sijil Off di Kantor Kesyahbandaran dan Otoritas Pelabuhan (KSOP) Kelas I Samarinda. Dengan menggunakan data historis Sijil Off, penelitian ini mengembangkan model regresi linear untuk memprediksi jumlah Sijil Off yang akan dikeluarkan. Metode regresi linear dipilih karena kemampuannya dalam menangani data yang memiliki hubungan linier dan kemudahan interpretasi hasilnya. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, pembuatan model regresi linear, dan evaluasi model menggunakan metrik Root Mean Squared Error (RMSE). Hasil penelitian menunjukkan bahwa model regresi linear dapat menjadi alat yang efektif dalam memprediksi jumlah Sijil Off dengan tingkat akurasi yang memuaskan. Rekomendasi untuk penelitian selanjutnya termasuk penggunaan variabel independen tambahan untuk meningkatkan akurasi model dan penerapan model regresi linear berganda. Penelitian ini memberikan kontribusi pada praktik pengambilan keputusan di KSOP Kelas I Samarinda dengan menyediakan model prediktif yang dapat membantu dalam perencanaan dan pengelolaan Sijil Off.

**Kata kunci:** Regresi Linear, Data Mining, Prediksi, Sijil Off, KSOP Samarinda

## Abstract

The study aims to analyze the application of linear regression in data mining as a predictive method for Certificate Off at the Samarinda Class I Municipal and Port Authority (KSOP). Using historical Off Certificate data, the study developed a linear regression model to predict the number of Off Certificates to be issued. The linear regression method was chosen because of its ability to handle data that has a linear relationship and the ease of interpreting the results. The research phases include data collection, data pre-processing, linear regression modeling, and model evaluation using Root Mean Squared Error metrics. (RMSE). The results show that the linear regression model can be an effective tool in predicting the number of Off certificates with a satisfactory accuracy level. Recommendations for further research include the use of additional independent variables to improve model accuracy and the application of dual linear regression models. This research contributes to decision-making practices at KSOP Class I Samarinda by providing predictive models that can help in the planning and management of the Off Certificate.

**Keywords:** Linear Regression, Data Mining, Predictions, Certificate Off, KSOP Samarinda

## PENDAHULUAN

KSOP merupakan Kantor Syahbandar dan Otoritas Pelabuhan, yang bertanggung jawab dalam pengelolaan kegiatan operasional dan keamanan di pelabuhan. KSOP juga memiliki tugas dan kewajiban untuk memastikan bahwa setiap kapal yang berlabuh di pelabuhan memenuhi persyaratan dan standar keamanan yang ditetapkan. Dalam melaksanakan fungsi dan tugasnya maka syahbandar memiliki kewenangan Melaksanakan sijil awak kapal (Mursidi, 2023). Untuk memaksimalkan pelayanan di butuhkan prediksi yang akurat mengenai permohonan, salah satunya yaitu permohonan sijil OFF di KSOP Samarinda.

Langkah untuk mengatasi permasalahan tersebut adalah dengan menggunakan metode regresi linier dalam data mining. Metode regresi linier adalah salah satu teknik statistik yang digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen (Simanungkalit dkk., t.t.). Metode regresi linier digunakan untuk memprediksi atau mengestimasi nilai variabel dependen berdasarkan variabel independen yang sudah diketahui (Simanungkalit dkk., t.t.). Dalam konteks prediksi sijil OFF di KSOP Samarinda, variabel independen yang dapat

digunakan dalam metode regresi linier dapat meliputi faktor seperti jumlah permohonan sijil OFF pada bulan sebelumnya.

Alasan utama memilih metode regresi linier dalam penelitian ini adalah karena metode ini memiliki beberapa keunggulan yang relevan dengan tujuan penelitian. Pertama, regresi linier merupakan metode yang sederhana dan mudah untuk diinterpretasikan, yang memungkinkan peneliti dan praktisi untuk memahami hubungan antara variabel independen dan variabel dependen dengan lebih baik (Indriani dkk., 2022). Metode ini efektif dalam memprediksi nilai variabel dependen berdasarkan variabel independen yang telah diketahui, yang telah terbukti dalam berbagai penelitian sebelumnya (Ardyanti & Abdriando, 2023). Regresi linier dapat memberikan estimasi yang akurat, yang ditunjukkan oleh nilai Mean Absolute Percentage Error (MAPE) yang rendah dalam penelitian-penelitian sebelumnya yang menggunakan metode ini (Indriani dkk., 2022). Metode ini memiliki kemampuan untuk diintegrasikan ke dalam aplikasi berbasis web, yang memungkinkan hasil yang konsisten dengan perhitungan manual (Indriani dkk., 2022). Regresi linier berganda cocok untuk data time series dengan persebaran data yang linear, yang sesuai dengan jenis data yang sering ditemukan dalam permohonan sijil OFF di KSOP Samarinda (Ardyanti & Abdriando, 2023).

Berdasarkan hasil penelitian sebelumnya yang menggunakan metode regresi linier menunjukkan bahwa metode ini efektif dalam memprediksi nilai variabel dependen berdasarkan variabel independen yang telah diketahui. Misalnya, dalam konteks prediksi jumlah kunjungan pasien BPJS, penelitian yang dilakukan di RS Vita Insani Pematang Siantar menggunakan teknik pengumpulan data langsung untuk mengumpulkan data yang relevan dengan jumlah kunjungan pasien (Simanungkalit dkk., t.t.). Dalam penelitian lain yang berkaitan dengan prediksi produksi padi, data yang diperoleh dari BPS Jawa Timur dari tahun 2007 hingga 2021 digunakan untuk memprediksi produksi padi dengan menggunakan metode Simple Regresi Linear, dan data tahun 2021 digunakan sebagai nilai validasi (Putra dkk., 2023).

Selain itu, dalam penelitian yang berkaitan dengan prediksi data obat, tahapan yang digunakan dalam melakukan prediksi melibatkan pengumpulan data, pembersihan data, dan pemodelan data untuk mempermudah penelitian dan memastikan penelitian berjalan dengan baik dan sistematis (Setiawan dkk., 2022). Hasil dari penelitian ini menunjukkan bahwa regresi linier dapat memberikan insight yang berguna untuk memahami pola yang ada dari data yang dianalisis, seperti dalam kasus penelitian yang mengeksplorasi pola jumlah penduduk dari data BPS (Ardyanti & Abdriando, 2023).

Tujuan dari penelitian adalah untuk mengembangkan model prediksi yang dapat secara efektif memperkirakan jumlah permohonan sijil OFF di Kantor Kesyahbandaran dan Otoritas Pelabuhan (KSOP) Kelas I Samarinda. Penelitian ini bertujuan untuk menerapkan metode regresi linier dalam data mining sebagai alat analisis untuk mengidentifikasi faktor-faktor yang mempengaruhi permohonan sijil OFF dan untuk menentukan seberapa besar pengaruh variabel independen terhadap variabel dependen. Dengan menggunakan model regresi linier, penelitian ini diharapkan dapat memberikan wawasan yang berguna bagi KSOP Samarinda dalam merencanakan dan mengoptimalkan layanan mereka, serta meningkatkan efisiensi operasional dan kepuasan pelanggan.

## **METODE**

### **Jenis Penelitian**

Metode penelitian yang digunakan penelitian merupakan adalah Penelitian Terapan (applied research). Penggunaan jenis penelitian ini bertujuan untuk memecahkan masalah-masalah praktis, baik masalah praktis dari individu atau kelompok. yang nantinya hasil riset atau penelitian terapan tersebut dapat langsung digunakan, serta dengan menggunakan jenis penelitian research and development yang dilakukan untuk mengembangkan sebuah produk baru yang teruji secara empiris. Atau dengan kata lain penelitian terapan merupakan dirancang untuk menjawab pertanyaan spesifik yang bertujuan memecahkan masalah praktis bukan untuk memperoleh pengetahuan baru atau teori baru.

**Tempat dan Waktu Penelitian**

Penelitian ini dilaksanakan disaat melakukan PKL (Praktik Kerja Lapangan) di sebuah instansi pemerintahan yang ada di Kota Samarinda yang dimulai dari bulan Juli 2023 sampai bulan September tahun 2023.

**Dataset**

Dataset yang digunakan dalam proses penelitian ini diolah dari Kantor Syahbandar Dan Otoritas Pelabuhan Kelas I Samarinda, LAPORAN TAHUNAN 2023.xlsx berisi 45 data dan terdiri dari 3 kolom. Ke tiga kolom tersebut terdiri dari 1 kolom kategori (Bulan, dan 2 kolom numerik (X) yaitu SIJIL ON, dan (Y) SIJIL OFF.

**Pre-processing Data**

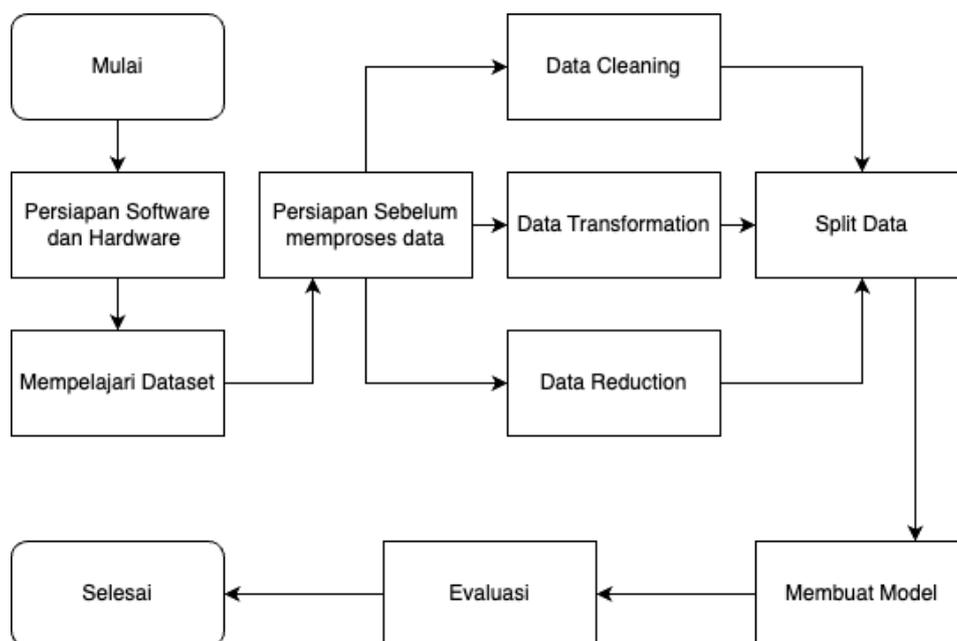
Pre-processing dilakukan sebelum pengolahan data dilakukan agar data yang digunakan dapat diolah dengan benar dan menghindari data yang salah. Data awal yang digunakan masih dianggap sebagai data mentah, tetapi data yang diperlukan akan diformat dengan cara tertentu selama proses pengolahan. Data dibersihkan (Cleaning) untuk memperbaiki atau menghapus data yang rusak. Langkah selanjutnya adalah melakukan uji standarisasi dan outlier untuk mengetahui kualitas data yang dibersihkan (Cleaning).

**Analisis Data**

Dalam proses penelitian, analisis data menggunakan regresi linier, diolah dan di terapkan menggunakan software Rapid Minner.

Proses pertama analisis data adalah membersihkan kumpulan data dari jenis data yang tidak diinginkan, seperti data kosong, data di luar ambang batas, dan tipe data yang tidak sesuai. Hasil dari proses pembersihan data akan dipilah menjadi dua bagian, yaitu data yang digunakan sebagai pelatihan dan data yang digunakan sebagai penilaian. Pelatihan digunakan untuk melatih algoritma, dan penilaian digunakan untuk mengetahui bagaimana algoritma yang telah dilatih sebelumnya berfungsi. Komposisi pembagian ini adalah 70% dari data yang ada untuk data training dan 30% untuk data testing.

Untuk memulai penelitian, diperlukan perangkat lunak dan perangkat keras. Di antara perangkat lunak yang diperlukan adalah kumpulan data yang digunakan dalam proses simulasi data, serta paket dan bibliotek yang digunakan dalam proses olah data. Datasheet memeriksa dan menghapus data yang tidak diperlukan dari kumpulan data tertentu. Langkah tambahan adalah menentukan persentase untuk data pelatihan dan pengujian. Sebagai hasilnya, proses pengujian keakuratan data dilakukan dengan data yang tidak termasuk dalam datasheet. Gambar berikut menunjukkan alur penelitian yang dilakukan.



**Gambar 1. Tahapan Data Mining**

**HASIL DAN PEMBAHASAN**

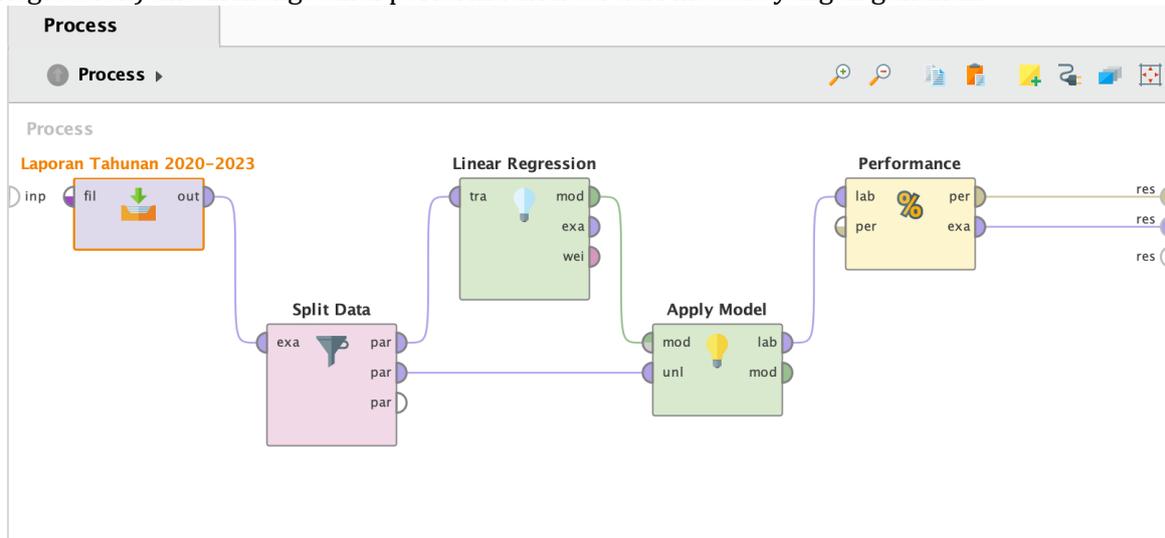
Pada proses implementasi regresi linear dalam estimasi proyeksi siji OFF di KSOP Kelas I Samarinda, data yang telah di terima dihitung terlebih dahulu pada tahun 2020-2023. Hasil data yang diperoleh berdasarkan 5 variabel yang berpengaruh dalam penelitian ini yaitu proyeksi Siji OFF dan Siji ON tahun 2023 dapat dilihat pada table berikut berikut :

**Tabel 1. Data Mentah Jumlah Permohonan Siji**

BULAN	SIJIL OFF	SIJIL ON
01/01/20	320	1827
01/02/20	242	1449
01/03/20	242	1302
01/04/20	57	1236
01/05/20	61	780
01/06/20	102	1420
01/07/20	281	1757
01/08/20	320	1786
01/09/20	357	1718
01/10/20	370	1621
01/11/20	360	1687
01/12/20	459	1760
01/01/21	320	1827
01/02/21	242	1449
01/03/21	242	1302
01/04/21	57	1236
01/05/21	61	780
01/06/21	102	1420
01/07/21	281	1757
01/08/21	320	1786
01/09/21	357	1718
01/10/21	370	1621
01/11/21	360	1687
01/12/21	459	1760
01/01/22	137	425
01/02/22	105	276
01/03/22	182	384
01/04/22	129	275
01/05/22	182	380
01/06/22	145	308
01/07/22	67	251
01/08/22	163	267
01/09/22	156	234
01/10/22	133	265
01/11/22	155	206
01/12/22	134	212
01/01/23	143	354
01/02/23	114	349
01/03/23	172	413
01/04/23	103	216
01/05/23	191	326
01/06/23	122	280
01/07/23	124	214
01/08/23	146	187

### Input Data

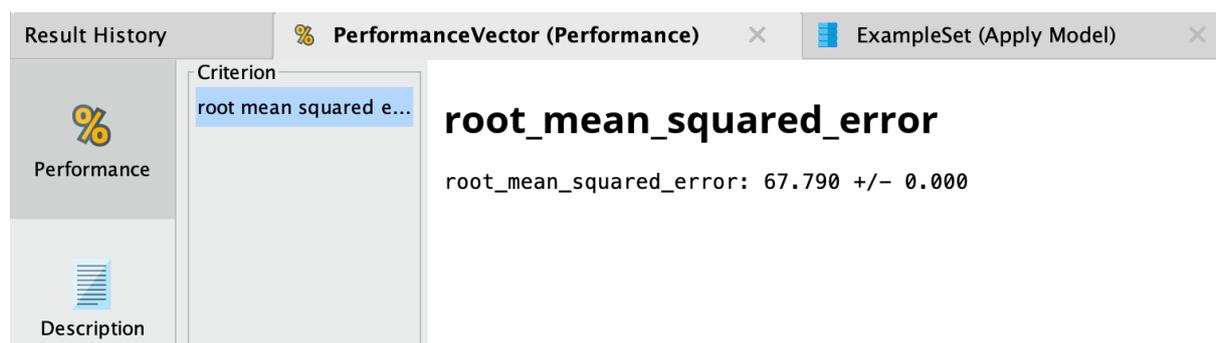
Perolehan data mentah yang sudah ada, kemudian data akan diolah menggunakan aplikasi RapidMiner. Data masukan berupa excel, kemudian masukan model data pilih menu Operators pilih operator Read Excel yang berfungsi untuk membaca data yang akan di analisis, selanjutnya menggunakan operator Split Data yang berfungsi untuk membagi data testing dan data training dengan perbandingan , selanjutnya menggunakan operator Linear Regression untuk memproses data training, selanjutnya menggunakan operator Apply Model yang berfungsi untuk memproses data yang sudah di training dan data testing, dan menggunakan operator Performance (Regression) untuk mengetahui performa akurasi dari metode yang di gunakan.



**Gambar 2. Proses Pengujian Dengan Melihat Performance**

### Output Data

Nilai RMSE (Root Mean Square Error) rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai obeservasinya. RMSE (Root Mean Square Error) menghitung seberapa berbedanya seperangkat nilai dari suatu objek hasil prediksi. Semakin kecil nilai RMSE (Root Mean Square Error), semakin dekat nilai yang diprediksi dan diamati (Setiawan dkk., 2022). Dengan menerapkan model tersebut melalui aplikasi RapidMiner Studio menghasilkan angka RMSE sebesar 67.790 +/- 0.000, nilai yang cukup baik dengan aplikasi RapidMiner Studio terhadap data uji yang disiapkan.



**Gambar 3. Hasil Luaran Performance**

### KESIMPULAN

Dari hasil penelitian yang telah dilakukan, beberapa kesimpulan dapat diambil mengenai penerapan regresi linier dalam data mining untuk prediksi Sijil OFF di KSOP Kelas I Samarinda. Model regresi linier yang dikembangkan menunjukkan kemampuan yang baik dalam memprediksi jumlah Sijil OFF yang akan dikeluarkan oleh KSOP Samarinda. Hal ini terlihat dari tingkat akurasi model yang memuaskan, yang diukur menggunakan metrik Root Mean Squared Error (RMSE). Proses pra-pemrosesan data yang meliputi pembersihan data dan pemilihan

variabel yang tepat terbukti penting dalam membangun model prediksi yang efektif. Penggunaan aplikasi RapidMiner dalam proses pra-pemrosesan dan pembuatan model regresi linier memudahkan analisis dan meningkatkan efisiensi waktu penelitian. Penelitian ini menunjukkan bahwa regresi linier merupakan metode yang cocok untuk diterapkan dalam kasus prediksi di sektor kepelabuhanan, khususnya dalam memprediksi permohonan Sijil OFF. Metode ini memberikan wawasan yang berguna bagi KSOP Samarinda dalam merencanakan dan mengoptimalkan layanan mereka. Sebagai rekomendasi untuk penelitian selanjutnya, disarankan untuk mempertimbangkan penggunaan variabel independen tambahan yang mungkin mempengaruhi jumlah permohonan Sijil OFF. Selain itu, penerapan model regresi linier berganda juga dapat dipertimbangkan untuk meningkatkan akurasi prediksi. Penelitian ini memberikan kontribusi pada praktik pengambilan keputusan di KSOP Kelas I Samarinda dengan menyediakan model prediktif yang dapat membantu dalam perencanaan dan pengelolaan Sijil OFF, sehingga dapat meningkatkan efisiensi operasional dan kepuasan pelanggan.

#### DAFTAR PUSTAKA

- Ardyanti, A. A. P., & Abdriando, A. (2023). PENERAPAN DATA MINING UNTUK MENGESTIMASI LAJU PERTUMBUHAN PENDUDUK DENPASAR MENGGUNAKAN METODE REGRESI LINIER BERGANDA. *JBASE - Journal of Business and Audit Information Systems*, 6(1). <https://doi.org/10.30813/jbase.v6i1.4317>
- Indriani, I., Siregar, D., & Windarto, A. P. (2022). Penerapan Metode Linear Regression dalam Mengestimasi Jumlah Penduduk. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 1112. <https://doi.org/10.30865/jurikom.v9i4.4676>
- Mursidi, M. (2023). Analisis Faktor Yang Mempengaruhi Keselamatan Pelayaran (Studi Pada KSOP Tanjung Emas Semarang). *JURNAL APLIKASI PELAYARAN DAN KEPELABUHANAN*, 14(1), 94–106. <https://doi.org/10.30649/japk.v14i1.106>
- Putra, R. N., Aziz, A., & Zaini, A. (2023). Implementasi Metode Simple Regresi Linear dan Single Exponential Smoothing untuk Memprediksi Produksi Padi Jawa Timur. 5(2).
- Setiawan, D., Surojudin, N., & Hadikristanto, W. (2022). *Prediksi Penjualan Obat Dengan Algoritma Regresi Linear*.
- Simanungkalit, E., Erwansyah, K., Kom, S., Kom, M., & Kustini, R. (t.t.). *Implementasi Data Mining untuk Mengestimasi Jumlah Kunjungan Pasien RS Vita Insani Pematang Siantar dengan Menggunakan Metode Regresi Linear Berganda*.