

# Implementation of the Support Vector Regression Algorithm And Particle Swarm Optimization In Sales Forecasting

Rahmawati<sup>1</sup>, Riyad Sabilul Muminin<sup>2</sup>, Irdham Denni<sup>1</sup>, Rizi Ramdhani<sup>1</sup>

<sup>1</sup>Program Studi Sistem Informasi, Fakultas Ilmu Terapan Dan Sains, Institut Pendidikan Indonesia, Jl. Terusan Pahlawan No.32, Sukagalih, Garut, Jawa Barat 44151, Indonesia

<sup>2</sup>Program Studi Bisnis Digital, Fakultas Ekonomi, Universitas Garut, Jl. Hampor No.52A, Mekarwangi, Garut, Jawa Barat 44151, Indonesia

*\*Penulis koresponden, e-mail : rahmawati@institutpendidikan.ac.id*

---

**Abstract:** Sales forecasting is an important element in planning business strategies that have the possibility of happening in the future. But how to improve the accuracy of the current forecasting process is still a big question mark for companies. This study uses the Support Vector Regression algorithm because it is one of the forecasting techniques that is categorized as good compared to other techniques. The Particle Swarm Optimization algorithm is integrated for attribute optimization so that forecasting accuracy can be better. Based on the test results, it was found that SVR-PSO produced an RMSE value of 9.40.

**Keywords:** sales forecasting; support vector regression; particle swarm optimization.

**Abstrak:** Peramalan penjualan merupakan elemen penting dalam perencanaan strategi bisnis yang memiliki kemungkinan terjadi di masa depan. Tetapi bagaimana meningkatkan ketepatan proses peramalan saat ini masih menjadi tanda tanya besar bagi perusahaan. Penelitian ini menggunakan algoritme Support Vector Regression karena merupakan salah satu teknik peramalan yang dikategorikan baik dibandingkan dengan teknik yang lainnya. Algoritme Particle Swarm Optimization diintegrasikan untuk optimasi atribut agar akurasi peramalan dapat lebih baik. Berdasarkan hasil pengujian, didapatkan SVR-PSO menghasilkan nilai RMSE 9.40.

**Kata kunci :** peramalan penjualan; support vector regression; particle swarm optimization.

---

## PENDAHULUAN

Tujuan utama dari peramalan (sales forecasting) adalah untuk menemukan informasi yang memiliki kemungkinan terjadi di masa depan. Menurut (Heizer, 2017) peramalan adalah salah satu bidang seni dan keilmuan matematis untuk memprediksikan peristiwa di masa depan dengan melibatkan data historis. Dengan menggunakan teknik peramalan, pengguna dapat menganalisis dan menemukan pengetahuan baru untuk membuat strategi, keputusan yang tepat dan merencanakan berbagai sumber daya dengan efektif dan efisien. Teknik peramalan telah banyak digunakan dalam berbagai bidang salah satunya adalah bidang bisnis.

Peramalan bisnis menjadi sangat penting bagi perusahaan guna memenuhi permintaan pasar yang mengalami kenaikan pada tahun-tahun berikutnya. Tetapi bagaimana meningkatkan ketepatan proses peramalan saat ini masih menjadi tanda tanya besar bagi perusahaan (Narkevičius, 2016). Untuk data yang mengandung pola tren, musiman atau yang dapat disebut dengan data time series, sering kali rentan mengalami kegagalan atau peramalan yang jauh dengan data aktual dikarenakan

pengaruh penjualan yang sangat fluktuatif. Banyak metode yang dapat digunakan untuk membuat peramalan, dimulai dari metode statistik tradisional sampai metode atau algoritme untuk data mining modern seperti machine learning yang terkomputasi. Namun metode statistik tradisional seperti exponential smoothing, ARIMA dan regression memiliki akurasi lebih rendah dibandingkan dengan metode pembelajaran mesin seperti support vector regression (SVR), long short-term memory (LSTM) dan artificial neural network (ANN) (Namini, 2018) (Zhang, 2014) (Kandananond, 2012). Oleh karena itu, pembelajaran mesin sangat berpeluang dalam mencapai kualitas peramalan yang baik.

Support Vector Regression (SVR) telah banyak diimplementasikan dalam kasus peramalan. Penelitian yang dilakukan oleh (Rifqi, 2018) menggunakan algoritme SVR untuk peramalan permintaan darah PMI kota Malang, menghasilkan bahwa nilai rata-rata kesalahan (average error) pengujian berkisar 3,8%, yang di mana jika hasil rata-rata kesalahan yang dihasilkan <10% dapat dikategorikan baik untuk peramalan. Penelitian lainnya (Hasan, 2015) (Lu, 2011) (Wen, 2014) menunjukkan jika algoritme SVR memiliki kemampuan generalisasi dalam peramalan yang baik.

Kehandalan performa algoritme SVR sangat ditentukan dengan pemilihan parameter yang digunakan (Kridanto, 2015). Tidak ada aturan umum dalam pemilihan parameter-parameter tersebut. Optimasi atribut dilakukan untuk memberikan atribut yang optimal. Particle Swarm Optimization (PSO) merupakan salah satu algoritme optimasi berbasis populasi yang dikembangkan oleh Eberhart dan Kennedy.

Dalam penelitian ini akan dilakukan pencarian suatu model yang bisa melakukan peramalan jumlah penjualan dengan menggunakan teknik Data Mining. Berdasarkan fungsi dari dua algoritme yang telah dibahas sebelumnya, penelitian ini akan menggunakan algoritme SVR dalam membuat model peramalan yang akan dikombinasikan dengan algoritme PSO sebagai optimasi atribut yang digunakan. Model yang dihasilkan dituangkan dalam bentuk visualisasi data yang diharapkan bisa digunakan sebagai sumber informasi bagi pihak eksekutif dalam melakukan pengambilan keputusan. Pengambilan keputusan yang diambil dapat berupa penentuan target untuk operasional, penentuan strategi marketing, atau sebagai acuan produksi dalam menyiapkan produk yang sesuai dengan kebutuhan.

## **KAJIAN PUSTAKA**

Tujuan utama peramalan (forecasting) adalah menghasilkan prediksi yang akurat. Pada dunia bisnis, ketika peramalan terlalu rendah berdampak pada kurangnya inventory. Begitu pun sebaliknya ketika peramalan terlalu tinggi berdampak pada inventory yang berlebihan dan memakan budget yang tidak efektif. Peramalan (Forecasting) merupakan seni dan ilmu untuk memprediksi peristiwa masa depan. Peramalan mungkin melibatkan pengambilan data historis (seperti rekap

penjualan masa lalu) dan memproyeksikannya ke masa depan dengan model matematika (Heizer, 2017).

Dari definisi di atas bisa disimpulkan bahwa peramalan sangat penting bagi perusahaan dalam mencapai tujuan perusahaan, karena peramalan berdampak pada keputusan manajemen ketika melakukan perencanaan ke depannya. Perencanaan yang efektif baik dalam jangka pendek maupun jangka panjang tergantung pada perkiraan permintaan untuk produk-produk perusahaan.

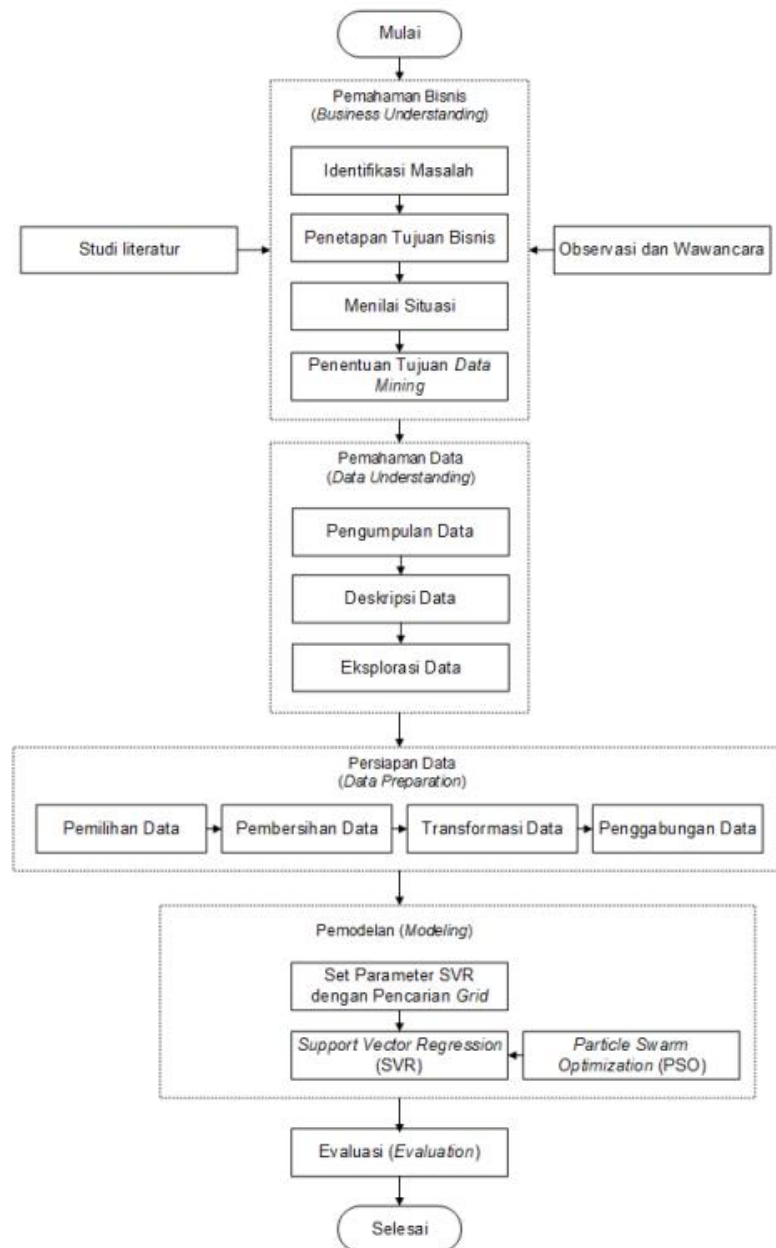
Cabang ilmu data mining muncul dari pertemuan ilmu statistik dan pembelajaran mesin (machine learning) yang menyediakan teknologi untuk membantu menganalisis dan memahami informasi yang terkandung dalam basis data (Palmer, 2011). Data mining dapat dikatakan sebagai hasil evolusi alami dari teknologi informasi. Secara garis besar data mining merupakan proses menemukan pola atau tren yang berguna di dalam dataset. data mining menggunakan kombinasi dari basis pengetahuan eksplisit, dan keterampilan analisis yang canggih, untuk mengungkap tren dan pola tersembunyi. Data mining digunakan untuk memecahkan berbagai masalah bisnis, seperti deteksi kecurangan, market basket analysis, dan analisis churn pelanggan. Umumnya, perusahaan menggunakan alat data mining pada volume besar data terstruktur, seperti basis data manajemen hubungan pelanggan (CRM) atau inventory suku cadang pesawat. Tujuan dari data mining adalah untuk menjelaskan dan memahami data. Data mining tidak dimaksudkan untuk membuat prediksi atau mendukung hipotesis (Hurwitz, 2018).

SVR merupakan adaptasi dari paradigma klasifikasi berdasarkan statistik / pembelajaran mesin yang baru diperkenalkan yaitu, Support Vector Machine (SVM) (Hsieh, 2011). Seperti halnya SVM, SVR bertujuan untuk menciptakan hyperplane dan support vector untuk menentukan nilai prediksi. Konsep SVR didasarkan pada mengestimasi suatu fungsi dengan cara meminimalkan batas atas dari kegagalan generalisasi, sehingga SVR mampu mengatasi overfitting (Awad, 2015). Namun kehandalan performa algoritme SVR sangat ditentukan dengan pemilihan parameter yang digunakan (Kridanto, 2015).

Algoritme Particle Swarm Optimization (PSO) adalah teknik komputasi stokastik yang secara iteratif mengoptimalkan solusi alternatif suatu masalah sampai mencapai sasaran fitness atau kualitas (Awad, 2015). Algoritme PSO pertama kali diperkenalkan oleh Dr. Kennedy dan Dr. Eberhart pada tahun 1995 yang terinspirasi oleh perilaku sosial dari pergerakan sekawanan burung atau ikan (bird flocking atau fish schooling).

## **METODE PENELITIAN**

Metodologi yang digunakan dalam penelitian ini adalah dasar metodologi CRISP-DM yang telah dimodifikasi berupa keterangan tambahan yang sesuai dengan kebutuhan penelitian. Berikut desain metodologi penelitian yang digambarkan pada gambar 1



**Gambar 1. Metodologi Penelitian**

## HASIL DAN PEMBAHASAN

Data dipilih berdasarkan hasil pencarian yang relevan antara objek penelitian dengan penelitian sebelumnya yang terkait. Dipilih algoritma Support Vector Regression sebagai algoritma peramalan dengan teknik regresi. Kinerja SVR sangat dipengaruhi oleh pengaturan parameter complement (C) sebagai batas nilai penalti toleransi kesalahan, nilai epsilon ( $\epsilon$ ) yang mengatur batas kesalahan fungsi dan nilai gamma ( $\gamma$ ) yang menyatakan lebar dari fungsi kernel RBF. Maka dari itu, terlebih dahulu dilakukan pengujian untuk menentukan parameter yang sesuai. Pengujian tersebut diantaranya menentukan nilai C, nilai  $\epsilon$  dan nilai  $\gamma$ . Untuk setiap pengujian diulang sebanyak 10

kali dengan menggunakan pencarian grid dan tenfold cross validation untuk mendapatkan parameter dengan akurasi yang baik. Dalam pengujian digunakan parameter awal sebagai berikut (Rusmalawati, 2018) :

$C : 1-200$

$\varepsilon : 0,00001-0,1$

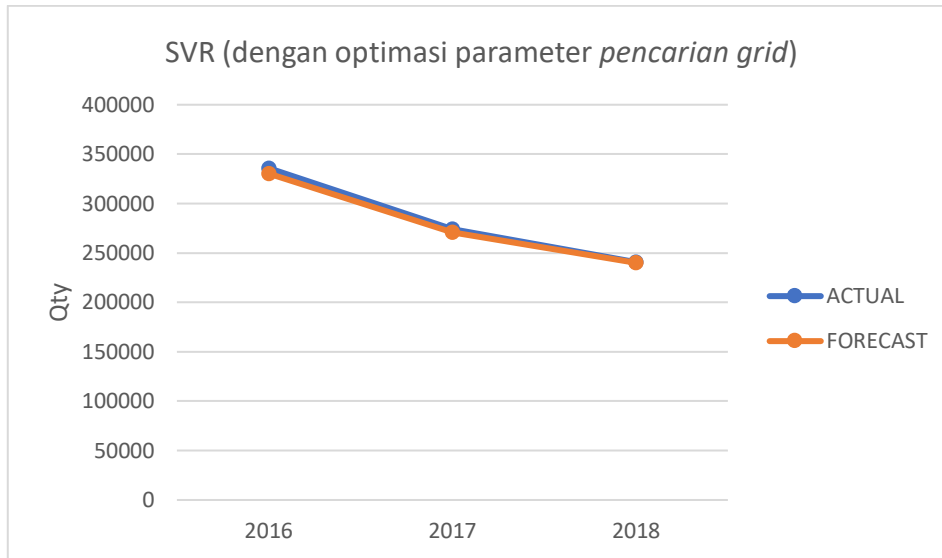
$\gamma : 0,1-1$

Dari beberapa hasil pengujian parameter maka dapat ditetapkan rentang nilai parameter yang akan digunakan dalam tahap selanjutnya adalah  $C = 1000-10000$ ,  $\varepsilon = 0.0001-0.1$  dan  $\gamma = 0.000001-0.01$ . Rentang nilai tersebut akan diuji kembali menggunakan dataset yang sama dan metode yang sama, yaitu pencarian grid dan tenfold cross-validation. Maka hasil yang didapat untuk nilai parameter terbaik dari pencarian grid adalah sebagai berikut :

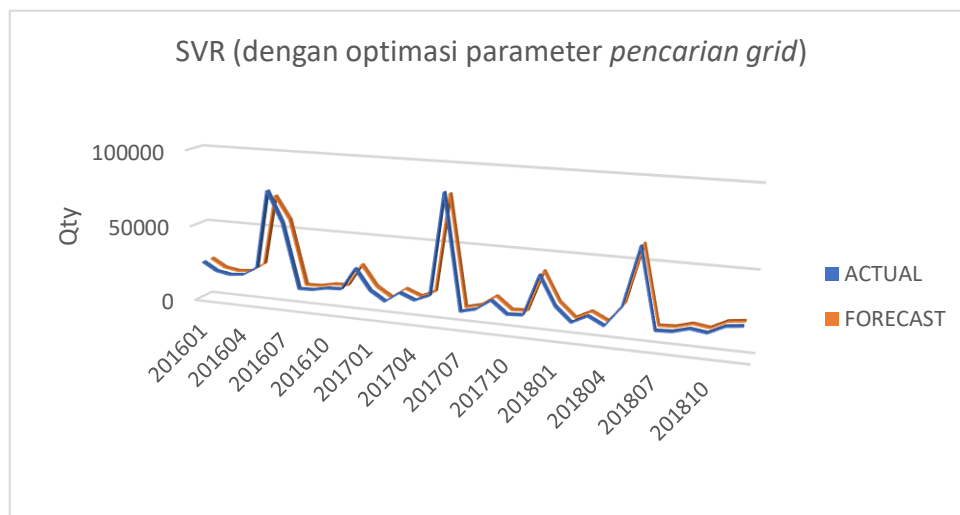
**Tabel 1. Nilai Parameter SVR**

C	$\varepsilon$	$\gamma$
7300	0.02008	0.000001

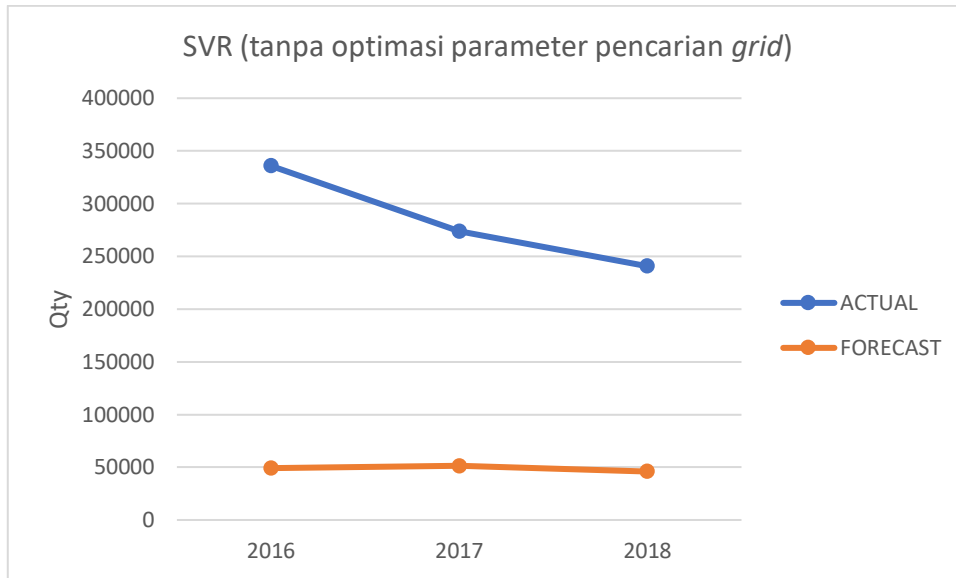
Setelah menemukan nilai parameter yang dapat dilihat pada pada tabel 1, pengujian model dilanjutkan untuk meramalkan penjualan. Berikut merupakan hasil dari pengujian dataset terhadap model SVR dalam meramalkan penjualan brand abc. Berdasarkan pengujian algoritma SVR dengan optimasi parameter pencarian grid diperoleh hasil RMSE 35,83. Untuk melihat seberapa efektif hasil parameter yang dilakukan, maka dilakukan pengujian dengan parameter tanpa dioptimalkan. Pada tools rapidminer angka default untuk semua parameter SVR adalah  $C = 0.0$ ,  $\varepsilon = 0.001$  dan  $\gamma = 0.0$ . Hasil pengujian didapatkan hasil RMSE sebesar 303,63. Dengan melihat hasil RMSE dari kedua pengujian tersebut dapat disimpulkan bahwa optimasi parameter pencarian grid terbukti efektif dalam mencari parameter optimal untuk SVR, dengan hasil RMSE atau tingkat kesalahan peramalan yang semakin mengecil. Grafik pengujian model SVR dengan optimasi parameter pencarian grid dapat dilihat pada gambar 2 dan gambar 3. sedangkan untuk grafik pengujian model SVR tanpa optimasi parameter dapat dilihat pada gambar 4 dan gambar 5.



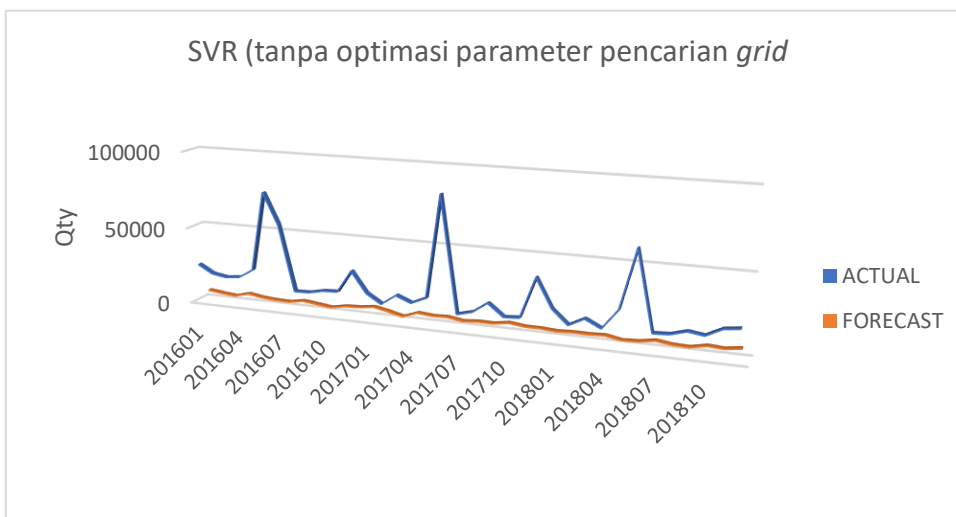
**Gambar 2. Grafik pengujian model SVR dengan optimasi parameter pencarian grid (pertahun)**



**Gambar 3. Grafik pengujian model SVR dengan optimasi parameter pencarian grid (perbulan)**



**Gambar 4. Grafik pengujian model SVR tanpa optimasi parameter pencarian grid (pertahun)**



**Gambar 5. Grafik pengujian model SVR tanpa optimasi parameter pencarian grid (perbulan)**

Pengujian selanjutnya melakukan proses PSO kepada model SVR yang telah dibentuk. Pertama, dilakukan pengujian pada nilai parameter population size dan maximum number of generation yang mempunyai nilai default pada rapidminer 5 dan 30. Selanjutnya dilakukan kembali uji coba dengan mengubah nilai-nilai untuk parameter PSO. Pada tabel 2 merupakan hasil dari pengujian yang dilakukan.

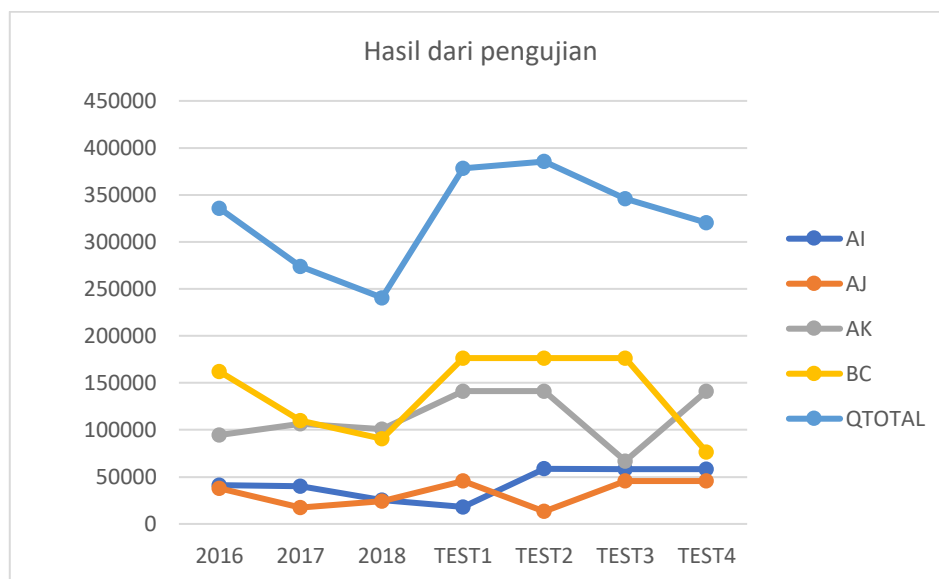
**Tabel 2. Hasil pengujian parameter population size dan maximum number of generation**

Parameter PSO	RMSE
Population size = 5, max generate = 30	13,19
Population size = 5, max generate = 40	11,63
Population size = 5, max generate = 50	13,10
Population size = 10, max generate = 30	12,48

Parameter PSO	RMSE
Population size = 10, max generate = 40	9,40
Population size = 10, max generate = 50	9,49
Population size = 15, max generate = 30	12,82
Population size = 15, max generate = 40	9,9
Population size = 15, max generate = 50	9,63
Population size = 20, max generate = 30	9,44
Population size = 20, max generate = 40	9,45
Population size = 20, max generate = 50	9,47

Berdasarkan tabel 2 pada parameter population size dengan nilai 10 dan parameter maximum number of generation dengan nilai 40 telah terjadi konvergen yang artinya performa akurasi telah maksimal. Dengan demikian dapat diketahui SVR-PSO menghasilkan error rate terbaik pada saat population size 10 dan maximum number of generation 40 dengan hasil prediksi RMSE 9,40.

Pengujian dilanjutkan dengan mengujikan data baru sebagai nilai-nilai prediktor, agar pemodelan dapat digunakan untuk meramalkan penjualan di masa depan. Nilai-nilai prediktor didapatkan dari rentang nilai penjualan perproduk pada tahun 2016-2018. Terdapat 4 dataset berbeda, dimana setiap dataset meminimalkan nilai pada rentang salah satu produk dan yang lainnya memaksimalkan nilai pada rentang produk. Tujuan dibentuk 4 dataset tersebut adalah untuk melihat kombinasi dari 4 produk yang paling mempengaruhi terhadap total peramalan penjualan. Gambar 6 menunjukkan gambaran jelas tentang produk yang diramalkan akan terjadi penurunan dan kenaikan. Dari 4 dataset yang dilakukan uji coba, dataset TEST2 menghasilkan total peramalan penjualan yang paling tinggi. Dengan demikian kombinasi produk yang paling optimal didapatkan pada dataset TEST2.



**Gambar 6 Grafik hasil pengujian**



Tahap selanjutnya, dilakukan percobaan untuk meramalkan 6 bulan ke depan untuk tahun 2020 dengan kombinasi produk pada dataset TEST 2. Selain perproduk, hasil dari peramalan juga ditampilkan berdasarkan wilayah, tipe cust dan tipe toko.

## **KESIMPULAN DAN SARAN**

Berdasarkan hasil pembahasan yang telah dijelaskan sesuai dengan tahapan penelitian, maka dapat diambil beberapa kesimpulan bahwa algoritme Support Vector Regression (SVR) dapat dikombinasikan dengan algoritme Particel Swarm Optimization (PSO) dalam melakukan peramalan penjualan. Hasil dari mengkombinasikan algoritme SVR-PSO didapatkan hasil akurasi peramalan menjadi lebih baik, dimana nilai RMSE yang dihasilkan SVR tanpa PSO sebesar 35,83, sedangkan hasil kombinasi dari SVR-PSO didapatkan hasil prediksi RMSE sebesar 9,40. Pemodelan SVR-PSO dapat digunakan untuk meramalkan penjualan. Berdasarkan hasil pengujian dalam meramalkan penjualan produk dapat dilihat bahwa penjualan dari berbagai produk akan mengalami kenaikan dan penurunan pada tahun-tahun berikutnya. Hal-hal tersebut dapat dijadikan pengetahuan atau sumber informasi bagi pihak eksekutif dalam mengambil keputusan. Hasil dari peramalan tersebut juga dapat dipakai sebagai acuan pihak produksi dalam mempersiapkan inventory yang sesuai. Adapun beberapa saran yang disampaikan agar hasil penelitian ini dapat bermanfaat dan dapat lebih baik pada penelitian selanjutnya adalah Menambahkan atribut lebih banyak dan jumlah data yang lebih besar, sehingga hasil pengukuran yang akan didapatkan lebih baik lagi, dikarenakan konsep dari mesin pembelajaran adalah semakin banyak data yang diolah maka semakin banyak mesin dapat belajar dan menemukan pola yang lebih efektif. Menggunakan metode lainnya seperti deep learning untuk penanganan kapabilitas lebih kompleks untuk mempelajari, mencerna, dan mengklasifikasikan dalam data penjualan.

## **DAFTAR PUSTAKA**

- Awad, M., & Khanna, R. (2015). Efficient learning machines: Theories, concepts, and applications for engineers and system designers. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*.
- Hasan, N., Nath, N. C., & Rasel, R. I. (2015). A Support Vector Regression Model for Forecasting Rainfall. *IEEE, (Eict)*. <https://doi.org/10.1109/EICT.2015.7392014>
- Heizer, J., Render, B., & Munson, C. (2017). *T Welf T H Edit Ion Oper At Ions* (12th ed.).
- Hsieh, H., & Lee, T. (2011). *A Hybrid Particle Swarm Optimization and Support Vector Regression Model for Financial Time Series Forecasting*. 2(2), 48–56. <https://doi.org/10.5430/ijba.v2n2p48>
- Hurwitz, J., & Kirsch, D. (2018). Machine Learning for Dummies. In *Journal of the American Society for Information Science* (IBM Limite, Vol. 35). United States of America: John Wiley & Sons, Inc.

- Kandananond, K. (2012). *International Journal of Engineering Business Management A Comparison of Various Forecasting Methods for Autocorrelated Time Series Regular Paper*. 4, 1–6. <https://doi.org/10.5772/51088>
- Kridanto, A., & Buliali, J. L. (2015). Metode Hibrida FCM dan PSO-SVR untuk Prediksi Data Arus Lalu Lintas. *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(3), 302–311. <https://doi.org/10.28932/jutisi.v1i3.414>
- Lu, X., & Geng, X. (2011). Car Sales Volume Prediction Based on Particle Swarm Optimization Algorithm and Support Vector Regression. *IEEE*, (1), 71–74. <https://doi.org/10.1109/ICICTA.2011.25>
- Narkevičius, R., & Šeškauskis, Z. (2016). *Sales Forecasting Management*. University of Borås.
- Palmer, A., Jimenez, R., & Gervill, E. (2011). Data Mining: Machine Learning and Statistical Techniques. *ResearchGate*, 3(May 2014). <https://doi.org/10.5772/13621>
- Rifqi, M. R., Setiawan, B. D., & Bachtihar, F. A. (2018). *Support Vector Regression Untuk Peramalan Permintaan Darah : Studi Kasus Unit Transfusi Darah Cabang – PMI Kota Malang*. 2(10), 3332–3342.
- Rusmalawati, V., & Furqon, M. T. (2018). Peramalan Harga Saham Menggunakan Metode Support Vector Regression ( SVR ) Dengan Particle Swarm Optimization ( PSO ). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(5), 1980–1990.
- Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1394–1401. <https://doi.org/10.1109/ICMLA.2018.00227>
- Wen, Q., Mu, W., Sun, L., Hua, S., & Zhou, Z. (2014). Daily Sales Forecasting for Grapes by Support Vector Machine. *IFIP*, 351–360. [https://doi.org/10.1007/978-3-642-54341-8\\_37](https://doi.org/10.1007/978-3-642-54341-8_37)
- Zhang, X., Zhang, T., Young, A. A., & Li, X. (2014). Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data. *PLOS ONE*, 9(2), 1–16. <https://doi.org/10.1371/journal.pone.0088075>