

# MEMPREDIKSI PRODUCTIVITY SCORE BERDASARKAN HOURS PER WEEK

Fabian Darrell Widyadhana Reswara <sup>1)</sup>

<sup>1)</sup> Teknik Informatika Universitas Tarumanagara  
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia  
email : [fdarrellworks@gmail.com](mailto:fdarrellworks@gmail.com)

## ABSTRACT

*Individual productivity is an important indicator in assessing performance both in the work and educational environment. This research aims to predict productivity scores based on the hours per week variable, namely the number of hours a person spends working per week. Using several approaches using regression methods, we explore the relationship between the number of hours worked per week and productivity levels. The dataset used in this research was taken from Kaggle. The best prediction accuracy is found in the SVR RBF algorithm using training data and test data of 70%:30% and this data is most suitable using SVR RBF using training data and test data of 70%:30%. This research provides insight into the optimal limits of working hours that can support productivity. The best MAE value is using the Linear SVR algorithm with training data and test data of 70%:30% of 8.801. A good RMSE value using the SVR RBF algorithm with training data and test data of 70%:30% is 11.391. The best R2 value is found in SVR RBF with training data and test data of 70%:30% of 0.062. The worst performance, worst prediction accuracy, and least suitable data use the SVR RBF algorithm with training data and test data 50%:50%. The MAE value that has poor performance is the SVR RBF algorithm with training data and test data of 50%:50% of 10.286. The RMSE value which has poor performance uses the SVR RBF algorithm with training data and test data of 50%:50% with a value of 13.046. The R2 value that has poor performance is the SVR RBF algorithm with training data and test data of 50%:50% with a value of 0.017.*

## Key words

*Productivity score, hours per week, regression, SVR*

## 1. Pendahuluan

Produktivitas adalah elemen penting yang menjadi perhatian utama di berbagai sektor, baik di dunia kerja maupun pendidikan. Produktivitas kerja dapat dianggap sebagai sikap mental [1]. Di era modern yang penuh dengan tantangan efisiensi, kemampuan untuk mengukur dan memprediksi tingkat produktivitas menjadi semakin krusial. Lingkungan kerja memainkan peran sangat

penting dalam meningkatkan produktivitas karyawan [2]. Salah satu faktor yang sering dikaitkan dengan produktivitas adalah waktu yang dihabiskan untuk bekerja atau belajar, yang biasanya diukur dalam jam per minggu. Ada anggapan umum bahwa semakin lama seseorang bekerja atau belajar, semakin tinggi pula produktivitas yang dihasilkan.

Dengan perkembangan teknologi dan tuntutan dunia kerja yang semakin kompetitif, penting untuk memahami apakah penambahan jam kerja selalu berdampak positif terhadap produktivitas, atau ada titik optimal di mana produktivitas mulai menurun meskipun jam kerja meningkat. Oleh karena itu, penelitian ini bertujuan untuk memprediksi skor produktivitas berdasarkan jumlah jam kerja per minggu menggunakan pendekatan statistik dan pembelajaran mesin. Melalui analisis data empiris, penelitian ini diharapkan dapat memberikan wawasan mengenai hubungan antara jam kerja per minggu dan produktivitas, serta memberikan rekomendasi berbasis data bagi perusahaan, institusi pendidikan, dan individu untuk mengelola waktu kerja secara lebih efektif.

Penelitian ini relevan dengan tren global seperti kerja jarak jauh, di mana fleksibilitas waktu kerja meningkat, tetapi produktivitas tetap harus dipertahankan. Dengan mengkaji hubungan antara jam kerja dan produktivitas, penelitian ini mencoba mencari tahu apakah produktivitas selalu meningkat seiring bertambahnya jam kerja.

## 2. Metode Penelitian

### 2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari Kaggle melalui link <https://www.kaggle.com/datasets/mrsimple07/remote-work-productivity>. Data tersebut mengandung kolom yang berisi Employee\_ID, Employee\_Type yang bisa Remote atau In-Office, Hours\_Worked\_Per\_Week yang berisi berapa jam setiap karyawan bekerja setiap minggu, Productivity\_Score untuk nilai produktivitas, dan Well\_Being\_Score yaitu nilai kesejahteraan. Data ini membantu untuk menganalisa dampak lingkungan kerja terhadap nilai produktivitas dan nilai kesejahteraan.

## 2.2 Model Support Vector Regression (SVR)

Support Vector Regression (SVR) merupakan penerapan dari metode Support Vector Machine (SVM). SVM kemudian dikembangkan untuk menyelesaikan kasus regresi yang disebut SVR [3]. SVR dapat digunakan untuk kasus regresi untuk mengatasi kasus data non-linear [4]. SVR ditemukan oleh Vladimir N. Vapnik pada tahun 1999 [5]. SVR digunakan untuk menemukan suatu fungsi yang memiliki deviasi paling besar. Tujuan dari SVR adalah menemukan suatu garis pemisah atau yang disebut hyperplane terbaik [6]. Prinsip kerja dari metode ini adalah mencari ruang pemisah yang paling optimal dari suatu dataset dalam kelas yang berbeda [7]. SVR akan dipakai untuk memprediksi nilai produktivitas berdasarkan jam tiap minggu karyawan bekerja. Berikut ini adalah algoritma yang akan dipakai.

SVR Linear merupakan kernel linear diterapkan untuk menangkap hubungan linier antara hours per week dan productivity score. Persamaan regresi yang digunakan sederhana dan memodelkan hubungan linier langsung. Fungsi ini sederhana dan tidak melakukan transformasi nonlinier pada data, hanya memetakan data ke dalam dimensi fitur yang sama [8]. SVM linear digunakan ketika data dapat dipisahkan dengan garis lurus. Contohnya, jika kita memiliki dua kelas dalam dua dimensi, kita dapat menggunakan garis lurus untuk memisahkan kelas-kelas tersebut.

SVR RBF (Radial Basis Function) merupakan kernel RBF merupakan salah satu kernel yang digunakan untuk menentukan parameter terbaik. Kernel ini menggunakan fungsi Gaussian (fungsi basis radial) untuk mengukur kesamaan antara dua vector input dalam ruang fitur [9]. RBF adalah kernel yang sangat umum digunakan dan efektif untuk data yang tidak terpisah secara linear. Kernel ini memetakan data ke ruang fitur yang lebih tinggi secara otomatis, memungkinkan pemisahan yang lebih baik.

SVR Polinomial merupakan kernel polinomial memungkinkan model untuk memodelkan interaksi data dengan derajat polinomial tertentu, memberikan fleksibilitas dalam menangani data yang memiliki hubungan non-linier yang lebih terstruktur [10]. SVM polinomial menggunakan fungsi kernel polinomial, yang cocok untuk dataset di mana hubungan antar fitur mengikuti pola polinomial.

## 2.3 Evaluasi Model

Mean Absolute Error (MAE) adalah salah satu metrik evaluasi dalam machine learning yang digunakan untuk menilai sejauh mana prediksi model mendekati nilai sebenarnya [11]. Metrik ini menghitung rata-rata dari nilai absolut selisih antara hasil prediksi dengan nilai aktual, tanpa memperhitungkan arah kesalahan (baik positif maupun negatif). Semakin kecil nilai MAE, maka semakin baik kualitas model tersebut. Kesalahan absolut menghilangkan tanda negatif dari selisih antara nilai aktual dan prediksi, sehingga setiap kesalahan

dihitung secara positif. MAE menghitung rata-rata dari semua kesalahan absolut, memberikan gambaran tentang seberapa besar rata-rata kesalahan prediksi model dalam unit yang sama dengan variabel yang diprediksi. MAE memberikan ukuran seberapa jauh prediksi model dari nilai sebenarnya, tanpa memberikan bobot lebih besar untuk kesalahan yang lebih besar (tidak seperti MSE yang mengkuadratkan selisih). Kelebihan MAE adalah mudah diinterpretasikan serta tidak terlalu sensitive terhadap outlier. Kekurangan MAE adalah tidak memberikan bobot lebih pada kesalahan besar. Rumus MAE dapat dituliskan sebagai berikut:

$$\frac{1}{n} \sum_{i=1}^n |Y_i - F(X_i)| \quad (1)$$

Mean Squared Error (MSE) merupakan metrik yang mengukur rata-rata dari selisih kuadrat antara nilai prediksi dan nilai sebenarnya. MSE digunakan untuk mengevaluasi kinerja metode peramalan [12]. Dalam metode ini, setiap kesalahan data dikuadratkan, kemudian dijumlahkan dan dibagi dengan jumlah periode pengamatan [13]. Nilai MSE yang lebih kecil tidak selalu menjamin akurasi yang lebih tinggi [14]. Namun biasanya, nilai MSE yang lebih kecil menandakan performa model yang baik. MSE menghitung seberapa jauh rata-rata prediksi dari model dibandingkan dengan nilai sebenarnya. MSE cenderung lebih sensitif terhadap outlier karena selisih yang besar akan menjadi lebih besar setelah dikuadratkan. Disini, MSE akan dipakai untuk menghitung nilai RMSE. Rumus MSE adalah sebagai berikut:

$$\frac{1}{n} \sum_{i=1}^n (Y_i - F(X_i))^2 \quad (2)$$

Root Mean Squared Error (RMSE) adalah metrik yang digunakan untuk menilai seberapa baik model dalam melakukan prediksi. Metrik ini sering digunakan untuk mengevaluasi hasil peramalan [15]. RMSE menunjukkan seberapa besar kesalahan yang dihasilkan oleh prediksi [16]. RMSE juga digunakan untuk melihat seberapa dekat nilai prediksi regresi yang dihasilkan oleh model dengan nilai sebenarnya [17]. RMSE adalah akar kuadrat dari perbedaan kuadrat rata-rata antara prediksi data dan pengamatan aktual [18]. Metrik ini dihasilkan dengan menghitung akar kuadrat dari nilai Mean Squared Error (MSE). RMSE menghitung seberapa jauh rata-rata prediksi dari model dibandingkan dengan nilai sebenarnya, tetapi dalam unit yang sama dengan variabel target. Dengan mengambil akar kuadrat dari MSE, RMSE memberikan nilai error yang lebih mudah diinterpretasikan, karena tidak dikuadratkan seperti pada MSE. RMSE sensitif terhadap outlier, karena error besar tetap memiliki pengaruh yang signifikan. Rumus RMSE adalah sebagai berikut:

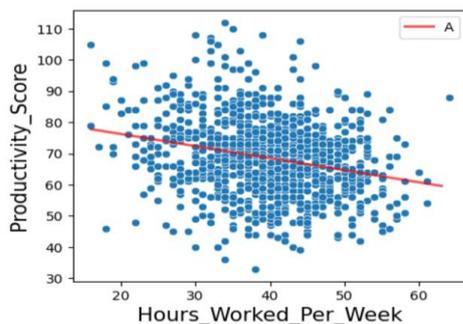
$$\sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \tag{3}$$

R<sup>2</sup> (R-Squared), atau koefisien determinasi, digunakan untuk mengetahui sejauh mana variabel bebas berkontribusi terhadap variabel terikat [19]. Pengujian koefisien determinasi bertujuan untuk mengukur besarnya pengaruh variabel bebas terhadap variabel terikat [20]. Dalam regresi, R<sup>2</sup> mengukur proporsi variasi total dari variabel dependen yang dapat dijelaskan oleh variabel independen (fitur) menggunakan model regresi. Jika R<sup>2</sup> = 1, berarti model dapat menjelaskan 100% variasi dalam data, atau dengan kata lain, prediksi model sempurna. Jika R<sup>2</sup> = 0, berarti model tidak mampu menjelaskan variasi dalam data. Prediksinya sama dengan menggunakan rata-rata y sebagai prediksi untuk semua data. Jika R<sup>2</sup> negatif, yang jarang terjadi, itu menunjukkan bahwa model lebih buruk daripada garis rata-rata horizontal. Dalam kasus ini, model tidak hanya gagal, tetapi juga menambah error pada prediksi. Kelebihannya adalah mudah diinterpretasikan karena memberikan proporsi variasi yang dapat dijelaskan oleh model serta sangat umum digunakan dalam regresi. Kekurangannya adalah tidak mempertimbangkan kompleksitas model. Model yang sangat kompleks (overfitting) mungkin memiliki nilai R<sup>2</sup> yang tinggi, meskipun kinerjanya buruk pada data baru.

Hasil dari evaluasi kinerja model dibandingkan untuk ketiga kernel SVR (linear, RBF, dan polinomial) untuk menentukan model mana yang paling sesuai dalam memprediksi productivity score.

### 3. Hasil Percobaan

Saat menggunakan regresi linear, jika jam tiap minggu adalah 40 jam, prediksi nilai produktivitas adalah 68.49. R<sup>2</sup> model regresi linear adalah 0.580. Saat menggunakan metode SVR linear, prediksi nilai produktivitas yang didapatkan jika jam tiap minggu adalah 40 adalah 67.024. Berikut ini adalah nilai MAE, RMSE, dan R<sup>2</sup> untuk algoritma SVR Linear, SVR RBF, dan SVR Polinomial serta plot untuk prediksi data linear.



Gambar 1. Plot untuk prediksi data linier

Tabel 1. Perbandingan antar algoritma

Data Latih: Data Uji	Algoritma	Nilai parameter kernel	MAE	RMS E	R2
70%:30%	SVR Linear	1	8.801	11.419	0.058
	SVR RBF	1	8.814	11.391	0.062
	SVR Polinomial	1	8.851	12.752	0.044
60%:40%	SVR Linear	1	10.043	12.861	0.052
	SVR RBF	1	10.129	12.861	0.035
	SVR Polinomial	1	10.129	12.861	0.035
50%:50%	SVR Linear	1	10.137	12.865	0.045
	SVR RBF	1	10.286	13.046	0.017
	SVR Polinomial	1	10.146	12.854	0.045

Melalui tabel di atas, nilai MAE terendah terdapat pada algoritma SVR Linear dengan nilai latih dan nilai uji 70%:30% yaitu sebesar 8.801. Nilai RMSE terendah terdapat pada algoritma SVR RBF dengan nilai latih dan nilai uji 70%:30%. Nilai R<sup>2</sup> terbesar terdapat pada algoritma SVR RBF dengan nilai latih dan nilai uji 70%:30% sebesar 0.062. Ini menandakan bahwa performa yang paling baik ada di SVR Linear dengan menggunakan data latih dan data uji 70%:30%. Akurasi prediksi paling baik terdapat pada algoritma SVR RBF dengan menggunakan data latih dan data uji 70%:30% dan data ini paling cocok menggunakan SVR RBF dengan menggunakan data latih dan data uji 70%:30%.

Melalui tabel di atas, nilai MAE terbesar terdapat pada algoritma SVR RBF dengan nilai latih dan nilai uji 50%:50% yaitu sebesar 10.286. Nilai RMSE terbesar terdapat pada algoritma SVR RBF dengan nilai latih dan nilai uji 50%:50% yaitu sebesar 13.046. Nilai R<sup>2</sup> terendah terdapat pada algoritma SVR RBF dengan nilai latih dan nilai uji 50%:50% yaitu sebesar 0.017. Ini menandakan bahwa performa paling buruk, akurasi prediksi paling buruk, dan data paling tidak cocok menggunakan algoritma SVR RBF dengan data latih dan data uji 50%:50%.

#### 4. Kesimpulan

Akurasi prediksi paling baik secara keseluruhan terdapat pada algoritma SVR RBF dengan menggunakan data latih dan data uji 70%:30% dan data ini paling cocok menggunakan SVR RBF dengan menggunakan data latih dan data uji 70%:30%. Ini berarti algoritma SVR RBF dengan data latih dan data uji 70%:30% bisa memprediksi nilai produktivitas terbaik. Nilai MAE yang rendah berarti performa yang lebih baik. Ini berarti algoritma terbaik untuk nilai MAE adalah SVR Linear dengan data latih dan data uji 70%:30% sebesar 8.801. Nilai RMSE yang rendah menandakan performa yang lebih baik. Ini berarti algoritma SVR RBF merupakan algoritma terbaik untuk nilai RMSE dengan data latih dan data uji 70%:30% sebesar 11.391. Nilai R2 yang lebih tinggi menandakan performa yang lebih baik. Ini berarti performa terbaik terdapat pada SVR RBF untuk nilai R2 dengan data latih dan data uji 70%:30% sebesar 0.062. Performa paling buruk, akurasi prediksi paling buruk, dan data paling tidak cocok menggunakan algoritma SVR RBF dengan data latih dan data uji 50%:50%. Ini berarti algoritma SVR RBF dengan data latih dan data uji 50%:50% tidak cocok untuk memprediksi nilai produktivitas. Nilai MAE yang tinggi berarti performa yang kurang baik. Ini berarti algoritma SVR RBF dengan data latih dan data uji 50%:50% sebesar 10.286 merupakan performa terburuk untuk MAE. Nilai RMSE yang tinggi menandakan performa yang kurang baik. Ini berarti algoritma SVR RBF dengan data latih dan data uji 50%:50% dengan nilai 13.046 merupakan nilai terburuk untuk RMSE. Nilai R2 yang rendah berarti performa yang kurang baik. Ini berarti algoritma SVR RBF dengan data latih dan data uji 50%:50% dengan nilai sebesar 0.017 merupakan algoritma terburuk untuk nilai R2. Kekurangan dari data ini adalah nilai MAE, RMSE, dan R2 yang cukup besar sehingga tidak terlalu menandakan ketergantungan. Kemungkinan untuk pengembangan selanjutnya adalah menggunakan metode-metode regresi lain untuk mencari nilai MAE, RMSE, dan R2 yang lebih bagus.

#### REFERENSI

- [1] I. Malikhah, A. P. Nst and G. P. Kaban, "Analisis Kompetensi Sdm Dan Pemanfaatan Teknologi Informasi Terhadap Produktivitas Kerja Pelaku Usaha Di Desa Pematang Serai Kabupaten Langkat," *MANEGGIO: Jurnal Ilmiah Magister Manajemen*, vol. 6, no. 1, 2023.
- [2] J. F. Fau and P. Buulolo, "Pengaruh Lingkungan Kerja terhadap Produktivitas Kerja Pegawai di kantor Samsat Kabupaten Nias Selatan," *Remik: Riset dan E-Jurnal Manajemen Informatika Komputer*, vol. 7, no. 1, 2023.
- [3] A. Umiyati, D. Dasari and F. Agustina, "Peramalan Harga Batubara Acuan," *Jurnal EurekaMatika*, vol. 9, no. 1, 2021.
- [4] R. Z. Rais, M. and A. S. Ahmar, "Implementation of the Support Vector Regression," *ARRUS Journal of Mathematics and Applied Science*, vol. 4, no. 1, 2024.
- [5] R. E. Cahyono, J. P. Sugiono and S. Tjandra, "Analisis Kinerja Metode Support Vector Regression (SVR)," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 1, no. 2, 2019.
- [6] L. M. Ginting, M. M. Sigiuro, E. D. Manurung and J. J. P. Sinurat, "Perbandingan Metode Algoritma Support Vector Regression dan," *Journal of Applied Technology and Informatics*, vol. 1, no. 2, 2021.
- [7] H. Apriyani and K. , "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *Journal of Information Technology Ampera*, vol. 1, no. 3, 2020.
- [8] S. D. Wahyuni and R. H. Kusumodestoni, "Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunning," *Bulletin Of Information Technology (BIT)*, vol. 5, no. 2, 2024.
- [9] S. Rabbani, D. Safitri, N. Ramadhani, A. A. F. Sani and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, 2023.
- [10] M. M. Siregar, R. Hizria and D. Pardede, "Perbandingan Kinerja Kernel SVM dalam Klasifikasi Kategori Kanker Kulit Menggunakan Transfer Learning," *DSI: Jurnal Data Science Indonesia*, vol. 4, no. 1, 2024.
- [11] I. Amansyah, J. Indra, E. Nurlaelasari and A. R. Juwita, "Prediksi Penjualan Kendaraan Menggunakan Regresi Linear.," *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 4, 2024.
- [12] S. Sautomo and H. F. Pardede, "Prediksi Belanja Pemerintah Indonesia Menggunakan Long Short-Term Memory (LSTM)," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, 2021.
- [13] D. Dardanella, A. P. Hidayat, S. H. Santosa and R. Siskandar, "PERAMALAN HARGA JUAL CABAI MERAH DI PASAR RAKYAT KEMANG PERUSAHAAN UMUM DAERAH PASAR TOHAGA KABUPATEN BOGOR," *Indonesia Journal of Science*, vol. 3, no. 1, 2022.
- [14] H. Elvaningsih, E. F. Tawakal and M. , "Prediksi Stok Obat Menggunakan Metode Backpropagation," *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 2021.
- [15] M. Muharrom, "Analisis Komparasi Algoritma Data Mining Naive Bayes, K-Nearest Neighbors dan Regresi Linier Dalam Prediksi Harga Emas," *Bulletin Of Information Technology (BIT)*, vol. 4, no. 4, 2023.
- [16] A. T. Nurani, A. Setiawan and B. Susanto, "Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda," *Jurnal Sains dan Edukasi Sains*, vol. 6, no. 1, 2023.
- [17] R. E. E. L. Lubis, A. Priansyah, Y. R. W. N. and T. Meutia, "PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAHAN PANGAN DI INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *JURNAL MAHASISWA AKUNTANSI SAMUDRA (JMAS)*, vol. 4, no. 4, 2023.
- [18] A. D. Milniadi and N. O. Adiwijaya, "ANALISIS PERBANDINGAN MODEL ARIMA DAN LSTM DALAM PERAMALAN HARGA PENUTUPAN SAHAM (STUDI KASUS: 6 KRITERIA KATEGORI SAHAM MENURUT PETER LYNCH)," *Jurnal Ilmiah*

*Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, vol. 2, no. 6, 2023.

- [19] A. Harun and N. N. A. Salmah, "Determinan Keputusan Pembelian Pelanggan Berdasarkan Kepercayaan dan Kemudahan Melalui Aplikasi Shopee," *Jurnal Media Wahana Ekonomika*, vol. 17, no. 2, 2020.
- [20] H. N. Febriani and M. Annas, "Determinasi Minat Beli Ulang Konsumen Pada Aplikasi TikTok," *Dynamic Management Journal*, vol. 7, no. 1, 2023.