



OPTIMASI PRODUKTIVITAS TENAGA KERJA MELALUI PREDIKSI BERBASIS *MACHINE LEARNING*

Adelina Octavia Sihombing
Badan Pusat Statistik Provinsi Sumatera Utara
adelina@bps.go.id

INFO ARTIKEL

Riwayat Artikel :

Diterima : 18 Juni 2025

Disetujui : 9 Juli 2025

Kata Kunci :

Model Prediksi, Tenaga Kerja,
Pembelajaran Mesin, *Random Forest*.

ABSTRAK

Kesejahteraan tenaga kerja ditandai dengan meningkatnya upah yang diterima seorang pekerja, ukuran ini bisa digambarkan sebagai produktivitas tenaga kerja yang dapat dipengaruhi oleh beberapa variabel seperti pendidikan, pengalaman kerja, penanaman modal, belanja modal, kepadatan penduduk, pendapatan asli daerah, dana alokasi khusus, ketinggian wilayah dan infrastruktur. Penelitian ini akan menggunakan metode *machine learning* untuk menemukan model terbaik yang akan memprediksi produktivitas tenaga kerja di Indonesia. Dari lima model yang diuji yaitu *Linear Regression*, *K-Nearest Neighbors (KNN)*, *Support Vector Regression (SVR)*, *Random Forest*, dan *Gradient Boosting* diperoleh hasil bahwa model *Random Forest* memiliki nilai RMSE terkecil yaitu sebesar 0,270. Sehingga jika ingin melakukan prediksi terhadap produktivitas tenaga kerja dapat digunakan model *Random Forest* tersebut. Dilihat berdasarkan *feature importance* diperoleh hasil bahwa pendidikan memiliki pengaruh yang paling besar terhadap produktivitas tenaga kerja.

ARTICLE INFO

Article History :

Received : Jun 18, 2025

Accepted : Jul 9, 2025

Keywords:

Prediction Model, Manpower, Machine Learning, Random Forest

ABSTRACT

Labor welfare is commonly reflected in the increase of wages received by workers, which serves as an indicator of labor productivity. Labor productivity can be influenced by various factors including education, work experience, capital investment, capital expenditure, population density, regional original income, special allocation funds, altitude, and infrastructure. This study applies machine learning techniques to identify the most accurate model for predicting labor productivity in Indonesia. Five models were evaluated: Linear Regression, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Random Forest, and Gradient Boosting. Among these, the Random Forest model demonstrated the lowest Root Mean Square Error (RMSE) value of 0.270, indicating superior predictive performance. Consequently, the Random Forest model is recommended for labor productivity prediction. Feature importance analysis revealed that education is the most influential factor affecting labor productivity.

1. PENDAHULUAN

Pemenuhan tujuan pembangunan berkelanjutan demi kesejahteraan masyarakat dapat tercapai salah satunya melalui pekerjaan layak dan pertumbuhan ekonomi yang merupakan tujuan ke-8 diantara 17 tujuan lainnya. Hal ini sejalan dengan Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 yang bertujuan untuk mengurangi kemiskinan, meningkatkan layanan pendidikan dan kesehatan, revitalisasi industri, dan meningkatkan daya saing bisnis lokal. Harapannya peningkatan produktivitas tenaga kerja akan mempercepat pembangunan di Indonesia, sehingga pertumbuhan ekonomi juga semakin membaik.

Pada era transformasi digital masalah terkait ketenagakerjaan menempatkan perusahaan dan pemerintah berada di situasi dilema antara meningkatkan keuntungan dengan mengadopsi teknologi digital terbaru atau meningkatkan produktivitas tenaga kerja untuk mengatasi pengangguran. Padahal pertumbuhan ekonomi di suatu negara memang tidak lepas dari peranan tenaga kerja dalam menggerakkan roda perekonomian. Salah satu peranan tersebut dapat dilihat melalui indikator produktivitas tenaga kerja. Harapannya semakin produktif tenaga kerja di suatu wilayah maka pertumbuhan ekonominya pun semakin baik.

Produktivitas tenaga kerja biasanya dikaitkan dengan upah pekerja, semakin produktif seorang pekerja maka semakin tinggi pula upah yang diterima oleh pekerja tersebut. Namun untuk melihat sebenarnya variabel apa yang mampu meningkatkan upah pekerja masih terdapat banyak kemungkinan variabelnya yang bisa saja berasal dari dalam diri pekerja itu sendiri maupun dari lingkungannya. Misalnya variabel pendidikan, pengalaman kerja, keterampilan, kesehatan, teknologi, lingkungan pekerjaan, ataupun variabel yang berasal dari dalam diri pekerja itu sendiri seperti motivasi kerja, budaya, hubungan dengan pekerja lainnya dan mungkin masih banyak variabel lainnya. Model *machine learning* masih minim digunakan pada penelitian regresi terkait produktivitas tenaga kerja di Indonesia, metode yang paling sering digunakan biasanya dalam bentuk regresi panel dengan aplikasi *Eviews* seperti pada penelitian (Fadillah, et al., 2021)

tentang determinan produktivitas tenaga kerja dilihat dari variabel indeks pembangunan manusia, infrastruktur ekonomi, upah minimum provinsi, tingkat partisipasi angkatan kerja, belanja daerah, dan pendapatan asli daerah. Selain menggunakan variabel indikator makro, penelitian untuk melihat produktivitas tenaga kerja juga dapat dilakukan dengan menggunakan faktor yang lebih spesifik seperti penggunaan Artificial Intelligence pada perusahaan (Roy, et al., 2021).

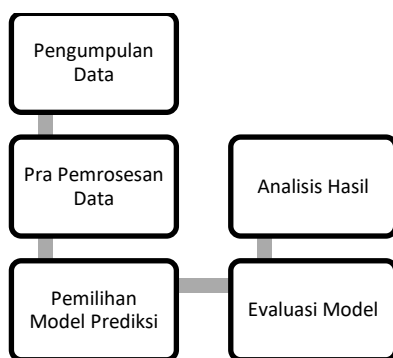
Penelitian tentang penggunaan *machine learning* dalam melakukan prediksi produktivitas tenaga kerja seperti pada penelitian (Lim, et al., 2024) yang mengulas secara kritis terkait penggunaan *machine learning* untuk melihat produktivitas tenaga kerja konstruksi. Berbagai model yang umum digunakan untuk produktivitas tenaga kerja yaitu *Random Forest*, (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), *Artificial Neural Network* (ANN), dan *Deep Learning* (CNN, LSTM, YOLO, YOWO), namun tidak ada satu model terbaik yang dapat dipilih secara agregat tergantung data yang digunakan pada setiap penelitian. Oleh karena itu pada penelitian ini akan dilihat model mana yang terbaik dalam memprediksi produktivitas tenaga kerja.

Pada penelitian ini ingin dilihat bagaimana produktivitas tenaga kerja di Indonesia dari sisi pendidikan, pengalaman kerja, penanaman modal, belanja modal, kepadatan penduduk, pendapatan asli daerah, dana alokasi khusus, ketinggian wilayah dan infrastruktur. Metode yang digunakan yaitu metode *machine learning*, mengingat besarnya data yang akan digunakan dengan banyaknya provinsi dan kabupaten/kota yang ada di Indonesia serta tahun penelitian yang digunakan. Harapannya akan diperoleh model terbaik mampu memprediksi produktivitas tenaga kerja di Indonesia. Penerapan *machine learning* dipilih karena model ini mampu menganalisis data dengan lebih mendalam (Mahesh, 2020).

2. METODE

Penelitian ini menggunakan metode *Machine Learning* untuk memprediksi produktivitas tenaga kerja. Hal ini dilakukan dengan beberapa tahapan mulai dari persiapan data, pemrosesan data, pemodelan, prediksi data,

dan akhirnya dilakukan analisis terhadap hasil pengolahan data. Data yang digunakan pada penelitian ini bersumber dari Badan Pusat Statistik (BPS), Direktorat Jenderal Perimbangan Keuangan, dan *Database for Policy and Economic Research* (INDO-DAPOER) menurut provinsi di Indonesia. Beberapa variabel yang digunakan pada penelitian ini yaitu produktivitas tenaga kerja yang diukur dengan transformasi logaritma natural dari upah pekerja per jam, rata-rata lama sekolah dalam tahun untuk melihat tingkat pendidikan pekerja, usia sebagai indikator pengalaman kerja, Penerimaan Modal Dalam Negeri (PMDN), belanja modal, kepadatan penduduk dalam satuan jiwa per kilometer persegi, Pendapatan Asli Daerah (PAD), Dana Alokasi Khusus (DAK), ketinggian wilayah dalam satuan meter di atas permukaan air laut, dan infrastruktur yang dilihat melalui persentase rumah tangga yang dapat mengakses listrik, sanitasi layak, dan air bersih. Pengolahan selanjutnya dilakukan dengan menggunakan *software Google Colab* dengan bahasa pemrograman Python.



Gambar 1. Alur Penelitian

Pada penelitian yang menggunakan metode *machine learning* biasanya melalui tahapan pra-pemrosesan data untuk memastikan apakah data yang akan digunakan sudah bebas dari missing value dan sudah dilakukan validasi tertentu agar hasil yang diberikan lebih maksimal. Pada tahap ini juga ingin dilihat apakah terdapat nilai nol (*zero value*) yang dapat mengganggu proses pengolahan data. Jika ada maka dimungkinkan untuk tetap digunakan ataupun dihapus dari dataset yang ada. Selanjutnya dilakukan transformasi data pada data yang memiliki satuan yang berbeda-beda tujuannya agar mudah

membandingkan antar variabel bebas yang digunakan dalam model.

Reduksi data dilakukan pada penelitian ini untuk memberikan hasil yang lebih akurat pada pemilihan model prediksi terbaik. Hal ini dilakukan dengan melihat hubungan yang erat antara variabel pada penelitian ini dengan produktivitas tenaga kerja, namun memiliki hubungan yang kurang erat dengan sesama variabel bebas lainnya.

Metode *machine learning* dilakukan untuk menganalisis produktivitas tenaga kerja untuk memudahkan pengambilan keputusan dimana dataset yang besar dengan 33 provinsi, pada 372 kabupaten, dan 17 tahun sehingga membentuk 6.324 data dengan 12 variabel. Pengolahan data dilakukan dengan menerapkan beberapa algoritma *machine learning* yaitu *Linear Regression*, *K-Nearest Neighbors* (KNN), *Support Vector Regression* (SVR), *Random Forest*, dan *Gradient Boosting*.

Linear Regression

Regresi linier sederhana sering digunakan sebagai model regresi yang paling sederhana untuk digunakan. Pada penelitian ini metode ini digunakan untuk melihat kondisi awal dataset dalam melihat gambaran variabel produktivitas tenaga kerja. Pada regresi linier hubungan variabel independen dalam hal ini rata-rata lama sekolah, usia, PMDN, belanja modal, kepadatan penduduk, PAD, DAK, ketinggian wilayah, listrik, sanitasi, air bersih terhadap produktivitas tenaga kerja. Adapun algoritma yang digunakan yaitu sebagai berikut:

Algoritma *Linear Regression* (1)

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
# Building the regression model
model = LinearRegression()
model.fit(X_train, y_train)
# Making predictions
y_pred = model.predict(X_test)
# Evaluating the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

K-Nearest Neighbors (KNN)

Model K-Nearest Neighbors (KNN) biasa digunakan untuk melihat klasifikasi data yang memperhatikan hubungan kedekatan data



dengan nilai yang tersedia di dalam dataset. Pada model ini akan terbentuk sebanyak K tetangga terdekat dari suatu data di dalam model. Berikut ini algoritma yang digunakan:

Algoritma *K-Nearest Neighbors* (KNN) (2)

```
from sklearn.metrics import
mean_absolute_error, mean_squared_error
# Initialize the KNN model
knn_model =
KNeighborsRegressor(n_neighbors=5)
# Train the model
knn_model.fit(X_train, y_train)
# Predict on the test set
y_pred_knn = knn_model.predict(X_test)
import numpy as np
# Evaluate the model again with the numpy
import
mae_knn = mean_absolute_error(y_test,
y_pred_knn)
rmse_knn =
np.sqrt(mean_squared_error(y_test,
y_pred_knn))
mae_knn, rmse_knn
```

Support Vector Regression (SVR)

Model *Support Vector Regression* (SVR) biasanya digunakan untuk meminimalkan kesalahan pada model prediksi dengan mengabaikan kesalahan-kesalahan kecil yang dibuatnya. Harapannya fungsi yang terbentuk memiliki kesalahan yang tidak terlalu jauh dari suatu nilai tertentu. Adapun algoritma untuk model SVR adalah sebagai berikut:

Algoritma *Support Vector Regression* (SVR) (3)

```
from sklearn.svm import SVR
# Initialize the SVR model
svr_model = SVR(kernel='rbf')
# Train the model
svr_model.fit(X_train, y_train)
# Predict on the test set
y_pred_svr = svr_model.predict(X_test)
# Evaluate the model
mae_svr = mean_absolute_error(y_test,
y_pred_svr)
rmse_svr =
np.sqrt(mean_squared_error(y_test,
y_pred_svr))
mae_svr, rmse_svr
```

Random Forest

Model *Random Forest* sangat dikenal sebagai metode ensemble learning dengan menggabungkan beberapa pohon keputusan yang akurat menggambarkan dataset. Adapun algoritma yang digunakan (Zhang & Ma, 2012) yaitu sebagai berikut:

Algoritma *Random Forest* (4)

```
from sklearn.ensemble import
RandomForestRegressor
```

```
# Initialize the Random Forest model
rf_model =
RandomForestRegressor(n_estimators=100,
random_state=42)
# Train the model
rf_model.fit(X_train, y_train)
# Predict on the test set
y_pred_rf = rf_model.predict(X_test)
# Evaluate the model
mae_rf = mean_absolute_error(y_test,
y_pred_rf)
rmse_rf =
np.sqrt(mean_squared_error(y_test,
y_pred_rf))
mae_rf, rmse_rf
```

Gradient Boosting

Model *Gradient Boosting* dikenal dengan metodenya yang membangun model secara bertahap dimana model yang baru akan memperbaiki kesalahan dari model sebelumnya demikian seterusnya sampai terbentuk model paling signifikan, dengan algoritma berikut:

Algoritma *Gradient Boosting* (5)

```
from sklearn.ensemble import
GradientBoostingRegressor
# Initialize and train the Gradient Boosting
Regressor model
gb_model=GradientBoostingRegressor(n_estim
ators=100, random_state=42)
gb_model.fit(X_train, y_train)
# Predict on the test set with Gradient
Boosting Regressor
y_pred_gb = gb_model.predict(X_test)
# Evaluate the Gradient Boosting Regressor
model
mae_gb = mean_absolute_error(y_test,
y_pred_gb)
rmse_gb =
np.sqrt(mean_squared_error(y_test,
y_pred_gb))
mae_gb, rmse_gb
```

Dari beberapa model pada metode machine learning tersebut akan dipilih model terbaik dengan menggunakan ukuran MAE dan RMSE, kemudian ukuran R² digunakan untuk melihat bagaimana variasi model dalam menjelaskan variabel produktivitas tenaga kerja.

3. HASIL DAN PEMBAHASAN

Penelitian ini dimulai dengan melakukan tahapan pra pemrosesan data terlebih dahulu untuk menghasilkan data set yang akan digunakan dalam pemodelan. Berdasarkan dataset pada penelitian ini memang tidak terdapat *missing value* namun ada beberapa nilai nol pada beberapa variabel. Pada variabel PMDN 437 data bernilai nol, variabel DAK ada 74 data, variabel belanja modal 4 data, dan variabel PAD 1 data. Selanjutnya dilakukan

transformasi data pada data yang memiliki satuan yang berbeda-beda seperti variabel PMDN, PAD, DAK, dan upah pekerja yang dalam satuan rupiah sementara terdapat variabel rata-rata lama sekolah dan usia dalam satuan tahun. Transformasi logaritma natural dilakukan pada variabel upah pekerja sebagai produktivitas tenaga kerja, PMDN, PAD, belanja modal, kepadatan penduduk, dan DAK. Namun pada variabel yang memiliki data bernilai nol tidak dilakukan transformasi logaritma natural (tetap bernilai nol).

Hubungan antar variabel bebas dan tidak bebas dalam model regresi biasanya diharapkan memiliki hubungan yang kuat namun antar variabel bebasnya tidak memiliki hubungan yang kuat. Untuk melihat hal ini dibentuk matriks korelasi antar variabel bebas dan tidak bebas yaitu rata-rata lama sekolah, usia, PMDN, belanja modal, kepadatan penduduk, PAD, DAK, ketinggian wilayah, listrik, sanitasi, air bersih, dan produktivitas tenaga kerja. Pada lampiran 1 dapat dilihat bahwa variabel ketinggian wilayah, akses listrik, dan kepadatan penduduk memiliki korelasi yang lemah dengan produktivitas tenaga kerja. Ketinggian wilayah jika dilihat hubungannya dengan variabel lain juga memiliki hubungan yang lemah. Sementara itu akses listrik dan kepadatan penduduk memiliki hubungan yang kuat dengan variabel bebas lainnya. Akses listrik memiliki hubungan yang kuat dengan kepadatan penduduk, PAD, sanitasi, air, dan PMDN. Kepadatan penduduk memiliki hubungan yang kuat dengan PAD, listrik, dan air. Oleh karena itu pada penelitian ini variabel ketinggian wilayah, akses listrik, dan kepadatan penduduk tidak digunakan pada pemilihan model dengan machine learning.

	tahun	padat	dak	pad	modal	listrik	\
tahun	1.000000	0.042640	0.496234	0.641039	0.530484	0.389470	
padat	0.042640	1.000000	-0.090359	0.428947	-0.064579	0.487354	
dak	0.496234	-0.090359	1.000000	0.257605	0.337649	0.104991	
pad	0.641039	0.428947	0.257605	1.000000	0.595739	0.572234	
modal	0.530484	-0.064579	0.337649	0.595739	1.000000	0.225463	
listrik	0.389470	0.487354	0.104991	0.572234	0.225463	1.000000	
sanitasi	0.482561	0.321703	0.075818	0.514177	0.298837	0.518466	
air	0.445820	0.505909	0.080933	0.564363	0.208089	0.578167	
pmdn	0.474559	0.230708	0.151326	0.530512	0.355653	0.482973	
tinggi	-0.000204	-0.149898	0.141688	-0.133783	-0.097843	-0.124234	
umur	0.330971	0.126909	0.173380	0.203490	0.066059	0.184384	
sekolah	0.210289	0.017654	0.164807	0.137274	0.140242	0.102595	
upah	0.707659	-0.176175	0.349038	0.361779	0.410027	0.161412	
	sanitasi	air	pmdn	tinggi	umur	sekolah	upa
tahun	0.482561	0.445820	0.474559	-0.000204	0.330971	0.210289	0.707659
padat	0.321703	0.505909	0.230708	-0.149898	0.126909	0.017654	-0.176175
dak	0.075818	0.080933	0.151326	0.141688	0.173380	0.164807	0.349038
pad	0.514177	0.564363	0.530512	-0.133783	0.203490	0.137274	0.361779
modal	0.298837	0.208089	0.355653	-0.097843	0.066059	0.140242	0.410027
listrik	0.518466	0.578167	0.482973	-0.124234	0.184384	0.102595	0.161412
sanitasi	1.000000	0.545317	0.382198	-0.278565	0.211693	0.101718	0.323177
air	0.545317	1.000000	0.290213	-0.163766	0.214070	0.219377	0.275650
pmdn	0.382198	0.290213	1.000000	-0.104935	0.178340	0.042245	0.246584
tinggi	-0.278565	-0.163766	-0.104935	1.000000	0.115361	0.114454	0.021595
umur	0.211693	0.214070	0.178340	0.115361	1.000000	0.188564	0.335650
sekolah	0.101718	0.219377	0.042245	0.114454	0.188564	1.000000	0.305208
upah	0.323177	0.275650	0.246584	0.021595	0.335654	0.305208	1.000000

Gambar 2. Matriks korelasi antara variabel

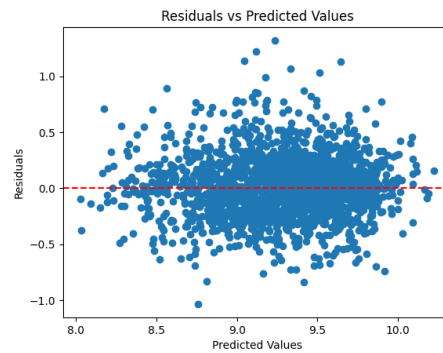
Pada tabel 1 dapat dilihat bagaimana performa metode regresi linier dengan nilai RMSE yang masih tinggi 0,350 dan nilai R2 yang rendah yaitu 0,455. Harapannya pada sebuah pemilihan metode terbaik yang perlu dilihat adalah nilai RMSE yang kecil mendekati nol dan nilai R2 yang tinggi mendekati satu. Sehingga dirasa metode Linear Regression masih belum tepat untuk menggambarkan hubungan antar variabel bebas dengan produktivitas tenaga kerja pada penelitian ini. Jika dilihat pada *K-Nearest Neighbors* (KNN) diperoleh nilai RMSE 0,343 dengan R2 0,477. Sudah lebih baik dari *Linear Regression* namun berdasarkan tabel dapat dilihat bahwa metode yang memiliki nilai RMSE terkecil atau mendekati nol yaitu metode *machine learning* Random Forest dengan nilai RMSE 0,270 dan nilai R2 0,676. Nilai R2 digunakan untuk melihat bagaimana variasi data pada variabel bebas (rata-rata lama sekolah, usia, PMDN, belanja modal, PAD, DAK, sanitasi, dan air bersih) yang digunakan untuk menjelaskan variabel produktivitas tenaga kerja. Nilai R2 0,676 memiliki penjelasan bahwa variabel bebas pada penelitian ini mampu menggambarkan 67 persen variabel produktivitas tenagar kerja, sementara 33 persen lainnya dipengaruhi oleh variabel lain di luar model.

Tabel 1. Hasil Pengujian Metode *Machine Learning*

Machine Learning	MAE	RMSE	R2
Linear Regression	0.278	0.350	0.455
K-Nearest Neighbors (KNN)	0.267	0.343	0.477
Support Vector Regression (SVR)	0.261	0.333	0.506
Random Forest	0.205	0.270	0.676
Gradient Boosting	0.223	0.286	0.634

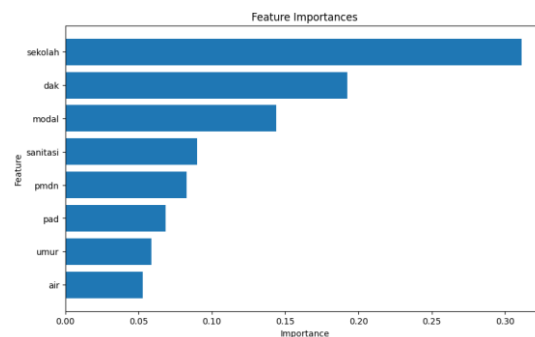
Agar model yang dipilih memiliki konsistensi nilai prediksi yang baik dilakukan *cross validation* terhadap data. Hasil *cross validation* menunjukkan performa model *Random Forest* melalui pengukuran *Root Mean Squared Error* (RMSE) pada lima fold (lipatan) yang berbeda dengan nilai RMSE fold 1 sampai 5 secara berurutan yaitu 0.282; 0.308; 0.384; 0.279; 0.303 dan nilai rata-rata RMSE dari kelima fold tersebut adalah 0.311 nilai ini menunjukkan performa model yang stabil pada data yang tidak terlihat selama pelatihan. Standard deviation sebesar 0.0381 menunjukkan bahwa RMSE pada setiap fold tidak terlalu bervariasi, yang berarti model memiliki performa yang stabil.

Jika ingin melihat bagaimana residual dari model yang terpilih maka dapat dilihat pada gambar 1 dimana dilihat nilai residuala yang berkisar pada nilai 0 artinya hasil prediksi sangat dekat dengan nilai yang sebenarnya. Dan jarak antar eror yang tidak jauh menunjukkan model sudah baik dalam hubungan antar variabel bebas dengan variabel produktivitas tenaga kerja.



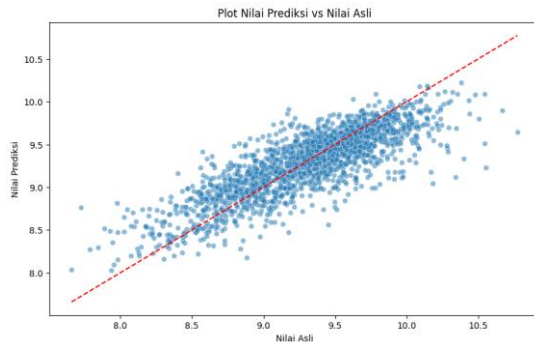
Gambar 1. Plot Residual model *Random Forest*

Selanjutnya akan dilihat bagaimana variabel bebas yang memiliki pengaruh paling besar di dalam model menggunakan *feature importance*, seperti pada gambar 2. Hasil yang diperoleh bahwa variabel rata-rata lama sekolah sebagai indikator pendidikan memiliki pengaruh yang paling besar secara berurutan diikuti dengan variabel DAK, belanja modal, sanitasi, PMDN, PAD, umur, dan air.



Gambar 2. Hasil pengecekan *feature importance*

Seaborn dapat digunakan untuk melihat bagaimana plot antara nilai prediksi dengan nilai dataset yang sebenarnya. Pada gambar 3 dapat dilihat bagaimana perbandingan nilai asli dengan nilai prediksi berdasarkan model *Random Forest*. Plot yang saling mengumpul menunjukkan bahwa antara nilai prediksi dengan nilai aslinya tidak berbeda jauh.



Gambar 3. Plot Perbandingan Nilai Prediksi dengan Nilai Asli dataset

Penelitian yang dilakukan (Bijalwan, et al., 2024) membandingkan enam algoritma machine learning untuk menilai produktivitas tenaga kerja perkotaan menghasilkan algoritma *XG Boost* yang memberikan perbandingan terbaik dibandingkan lima algoritma lainnya. Hasil tersebut berbeda dengan penelitian ini dimana model *Random Forest* merupakan model yang paling sesuai untuk melihat produktivitas tenaga kerja. Selain kinerja tenaga kerja, model *machine learning* juga dapat digunakan untuk melihat produktivitas mahasiswa. Penelitian yang dilakukan (Indahyanti, et al., 2022) menggunakan pendekatan *Ensemble Learning* untuk memprediksi produktivitas mahasiswa, hasilnya algoritma *Random Forest* menghasilkan prediksi yang paling akurat dalam menggambarkan produktivitas mahasiswa. Algoritma yang sama juga menjadi pilihan terbaik pada penelitian (Nawawi, et al., 2024) dalam melakukan prediksi ketepatan penempatan karir seorang tenaga kerja di Indonesia.

Produktivitas tenaga kerja tidak hanya dapat dilihat secara makro seperti pada penelitian ini, namun juga dapat dilihat secara mikro pada sebuah perusahaan tertentu. Seperti penelitian yang dilakukan (Huja, et al., 2025) menggunakan model machine learning untuk memprediksi kinerja buruh pabrik garmen menggunakan algoritma regresi linier. Penelitian yang dilakukan (Hidayah & Hidayah, 2025) juga secara mikro pada PT. Zhafran Putra Utama Kabupaten Soppeng untuk melihat prediksi gaji karyawan dengan model *machine learning* menggunakan algoritma regresi linier.

Hal yang menarik pula jika masalah produktivitas tenaga kerja dilihat juga dari sisi

kualitas tenaga kerja secara personal seperti penelitian yang dilakukan (Maehendrayuga, et al., 2024) dalam melihat turnover karyawan menggunakan variabel demografi, status pekerjaan, dan tingkat kepuasan karyawan. Pada penelitian ini algoritma *Random Forest* juga menjadi algoritma yang paling tepat untuk memprediksi turnover karyawan.

4. PENUTUP

4.1. Kesimpulan

Produktivitas tenaga kerja sebagai indikator penting dalam pertumbuhan ekonomi untuk mencapai tujuan pembangunan berkelanjutan dan selaras dengan rencana pembangunan jangka menengah dapat dilihat pada penelitian ini. Dari beberapa variabel bebas yang diteliti diperoleh kesimpulan bahwa dengan menggunakan model *Random Forest* variabel yang memiliki pengaruh terbesar yaitu pendidikan, DAK, dan belanja modal terhadap produktivitas tenaga kerja di Indonesia.

4.2. Saran

Berdasarkan hasil yang diperoleh pada penelitian ini bahwa variabel yang paling berpengaruh terhadap produktivitas tenaga kerja yaitu pendidikan sehingga pemerintah dapat melakukan perbaikan pada sistem pendidikan nasional yang mampu mendorong peningkatan tingkat pendidikan penduduk Indonesia terutama pada mereka yang berada di usia kerja. Selain itu variabel dana alokasi khusus (DAK) juga memberikan pengaruh terkuat berikutnya sehingga pada penentuan alokasi DAK, pemerintah perlu mempertimbangkan beberapa hal agar alokasi yang disampaikan ke daerah lebih merata sesuai dengan kebutuhan daerah tersebut.

Pada penelitian ini produktivitas tenaga kerja dilihat secara makro berdasarkan indikator-indikator ekonomi, pada penelitian selanjutnya mungkin dapat melihat produktivitas tenaga kerja langsung kepada personal tenaga kerja tersebut menggunakan indikator yang bersifat kualitatif.

5. DAFTAR PUSTAKA

Bijalwan, P. et al., 2024. Predicting the Productivity of Municipality Workers:

- A Comparison of Six Machine Learning Algorithms. *Economies*, 12(16).
- Fadillah, A., I. & R., 2021. Determinants of Labor Productivity and Economic Growth in Indonesia. *IOSR Journal of Economics and Finance (IOSR-JEF)*, 12(2), pp. 49-57.
- G. D., Roy, V. V. & Vertesy, D., 2021. The impact of artificial intelligence on labor productivity. *Eurasian Business Review* , Band 11, pp. 1-25.
- Huja, W. N., Hidayat, M. & Asnawi, M. F., 2025. Analisis Model Prediksi Kinerja Karyawan pada Sebuah Perusahaan Garmen menggunakan Machine Learning. *TECHNOMEDIA : Informatics and Computer Science*, 2(1), pp. 9-14.
- I. & Hidayah, A., 2025. Implementasi Machine Learning Dengan Metode Regresi Linear Untuk Prediksi Gaji Karyawan Berdasarkan Masa Kerja. *Jurnal RISTER : Riset Sistem Cerdas*, 2(1), pp. 1-7.
- Indahyanti, U., Azizah, N. L. & Setiawan, H., 2022. Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa. *Jurnal Sains dan Informatika* , 8(2), pp. 160-169.
- Lim , Y. T., Yi , W. & Wang, H., 2024. Application of Machine Learning in Construction Productivity at Activity Level: A Critical Review. *Applied Sciences*, 14(10605).
- Maehendrayuga, A., Setyanto, A. & K., 2024. Analisa Prediksi Turnover Karyawan menggunakan Machine Learning. *Bit-Tech(Binary Digital - Technology)*, 7(2), pp. 648-659.
- Mahesh, B., 2020. Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 9(1).
- Nawawi, H. M., Hikmah, A. B., Mustopa, A. & Wijaya, G., 2024. Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir. *Jurnal Saintekom : Sains, Teknologi, Komputer dan Manajemen*, 14(1), pp. 13-25.