JURNAL Industri Kreatif JIK

Februari 2025 | Vol. 9 | No. 01
E-ISSN : 2614-7602
DOI: 10.36352/jik.v9i01.1153

Penerapan Algoritma Regression Tree dan C5.0 untuk Menganalisis Produktivitas Tenaga Kerja

Ardi Maulana*[1]
[1]Departemen Teknik Mesin dan Industri, Universitas Gadjah Mada
e-mail: *[1]maulana,ardi@ugm.ac.id,

Abstrak

Produktivitas tenaga kerja merupakan indikator penting dalam menilai kinerja suatu organisasi. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi produktivitas tenaga kerja dengan menerapkan algoritma Regression Tree dan C5.0. Data yang digunakan berasal dari dataset publik industri garmen yang mencakup berbagai variabel seperti waktu lembur, insentif, idle time, dan target produktivitas. Data dianalisis menggunakan bahasa pemrograman R dengan pendekatan machine learning. Regression Tree digunakan untuk memprediksi produktivitas aktual, sedangkan C5.0 digunakan untuk mengklasifikasikan produktivitas ke dalam kategori "rendah", "sedang", dan "tinggi". Hasil penelitian menunjukkan bahwa kedua algoritma mampu mengidentifikasi pola penting yang memengaruhi produktivitas dan menghasilkan model yang akurat serta mudah diinterpretasikan. Pendekatan ini memberikan dasar yang kuat bagi manajemen dalam menyusun kebijakan peningkatan kinerja karyawan secara objektif dan berbasis data.

Kata kunci—Produktivitas tenaga kerja, Regression Tree, C5.0

Abstract

Labor productivity is a key indicator for assessing organizational performance. This study aims to analyze the factors influencing labor productivity by applying the Regression Tree and C5.0 algorithms. The dataset used is a public garment industry dataset containing various variables such as overtime, incentives, idle time, and productivity targets. Data analysis was conducted using the R programming language and machine learning approaches. The Regression Tree algorithm was employed to predict actual productivity, while C5.0 was used to classify productivity levels into "low", "medium", and "high" categories. The results indicate that both algorithms effectively identify significant patterns affecting productivity and produce accurate, interpretable models. This approach provides a strong foundation for management to develop data-driven strategies to improve workforce performance.

Keywords— Labor productivity, Regression Tree, C5.0

PENDAHULUAN

Produktivitas tenaga kerja merupakan salah satu indikator utama dalam menilai efisiensi dan efektivitas suatu organisasi. Dalam era persaingan global dan transformasi digital, organisasi dituntut untuk mengoptimalkan kinerja sumber daya manusianya guna mencapai keunggulan kompetitif. Produktivitas yang tinggi tidak hanya mencerminkan kemampuan individu dalam menyelesaikan tugas, tetapi juga mencerminkan efektivitas sistem kerja secara keseluruhan.

Joseph M. Juran, seorang tokoh terkemuka dalam manajemen mutu, menekankan bahwa kualitas dan produktivitas adalah dua sisi dari mata uang yang sama. Ia menyatakan bahwa "tanpa standar, tidak ada dasar logis untuk membuat keputusan atau mengambil tindakan" . Pernyataan ini menyoroti pentingnya pendekatan sistematis dan berbasis data dalam upaya peningkatan produktivitas tenaga kerja.

Mengukur dan menganalisis produktivitas tenaga kerja bukanlah tugas yang sederhana. Produktivitas dipengaruhi oleh berbagai faktor, baik internal maupun eksternal, seperti tingkat pendidikan, pengalaman kerja, pelatihan, motivasi, kondisi kerja, dan kebijakan organisasi. Kompleksitas ini memerlukan metode analisis yang mampu mengolah data dalam jumlah besar dan mengidentifikasi pola serta hubungan antar variabel secara akurat.

Pendekatan berbasis data mining dan machine learning menjadi solusi yang relevan. Salah satu metode yang efektif adalah penggunaan algoritma Regression Tree, yang memungkinkan pemodelan hubungan antara variabel independen dan variabel dependen kontinu. Algoritma ini membagi data menjadi subset berdasarkan nilai-nilai variabel prediktor, sehingga dapat mengidentifikasi faktor-faktor yang paling berpengaruh terhadap produktivitas .

Algoritma C5.0 juga merupakan alat yang powerful dalam analisis klasifikasi. Algoritma ini merupakan pengembangan dari C4.5 dan dikenal karena efisiensinya dalam membangun pohon keputusan yang akurat dan mudah diinterpretasikan. C5.0 bekerja dengan membagi data berdasarkan atribut yang memberikan informasi gain tertinggi, sehingga efektif dalam menangani dataset yang kompleks dan besar .

Beberapa studi telah menunjukkan efektivitas algoritma Regression Tree dan C5.0 dalam analisis produktivitas. Misalnya, penelitian oleh De'ath dan Fabricius (2000) menunjukkan bahwa Regression Tree dapat digunakan untuk mengidentifikasi subkelompok dengan karakteristik produktivitas yang berbeda . Sementara itu, penggunaan algoritma C5.0 dalam analisis data karyawan telah membantu organisasi dalam mengklasifikasikan tingkat produktivitas berdasarkan berbagai atribut personal dan profesional .

Dengan mempertimbangkan pentingnya produktivitas tenaga kerja dan kompleksitas faktor-faktor yang mempengaruhinya, penerapan algoritma Regression Tree dan C5.0 dalam analisis produktivitas menjadi sangat relevan. Pendekatan ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai faktor-faktor penentu produktivitas dan membantu organisasi dalam merancang strategi peningkatan kinerja yang lebih efektif.

METODE PENELITIAN

Penelitian ini menggunakan suatu kerangka metodologi yang dirancang untuk menganalisis data produktivitas tenaga kerja. Kerangka ini terbagi dalam empat tahap utama, yaitu: (1) pengumpulan data, (2) persiapan data, (3) pemrosesan data menggunakan algoritma machine learning, dan (4) interpretasi hasil. Setiap tahapan dijelaskan sebagai berikut:

2.1 Pengumpulan Data

Tahapan awal ini berfokus pada penghimpunan informasi terkait produktivitas tenaga kerja beserta faktor-faktor yang diduga memiliki pengaruh terhadapnya. Sumber data dapat berupa hasil survei, wawancara, maupun pengambilan data langsung dari sistem basis data perusahaan, terutama jika perusahaan telah mengadopsi sistem informasi terintegrasi yang memungkinkan pengelolaan data secara digital dan efisien.

2.2 Persiapan Data

Data yang telah dikumpulkan kemudian disiapkan agar memenuhi syarat untuk dianalisis menggunakan algoritma machine learning. Proses ini dikenal sebagai pra-pemrosesan data (data preprocessing). Kegiatan yang dilakukan meliputi:

- Penanganan data tidak lengkap, yaitu mendeteksi dan memperbaiki data yang kosong atau hilang pada variabel tertentu. Solusi yang diterapkan bisa berupa penghapusan entri yang bermasalah atau pengisian nilai yang hilang dengan estimasi.
- Identifikasi dan penanganan outlier, yaitu data yang memiliki nilai ekstrem yang tidak wajar dan dapat mempengaruhi hasil analisis. Data ini perlu dievaluasi lebih lanjut apakah akan disertakan atau dikeluarkan dari analisis.
- Penyesuaian tipe data dan seleksi variabel, karena tidak semua data dapat digunakan langsung dalam model machine learning. Tipe data perlu disesuaikan sesuai kebutuhan algoritma, dan hanya variabel yang relevan yang akan dipilih untuk dimasukkan ke dalam model guna mengurangi kompleksitas dan meningkatkan performa.
- Pembagian dataset, yakni memisahkan data menjadi dua bagian: data latih (training set) untuk membangun model, dan data uji (testing set) untuk menilai performa model secara objektif.

2.3 Pemrosesan Data Menggunakan Algoritma *Machine Learning*

Setelah melalui tahap persiapan, data dianalisis dengan algoritma machine learning yang dipilih. Dalam penelitian ini, digunakan dua algoritma utama yaitu Regression Tree untuk tugas prediksi numerik (regresi), dan Classification C5.0 untuk prediksi kategorikal (klasifikasi). Kedua algoritma tersebut dipilih karena memiliki keunggulan dalam hal kemudahan interpretasi serta cukup representatif untuk menunjukkan keakuratan model. Namun demikian, framework ini juga memungkinkan fleksibilitas penggunaan algoritma lain yang memenuhi karakteristik serupa.

2.4 Interpretasi Hasil

Pada tahap akhir, hasil dari pemodelan dianalisis untuk memahami faktor-faktor yang paling signifikan terhadap produktivitas tenaga kerja. Proses ini juga mencakup pemahaman terhadap bagaimana model membuat prediksi, serta penerjemahan hasil analisis ke dalam bentuk rekomendasi manajerial yang aplikatif. Visualisasi data dalam bentuk grafik turut disajikan guna mempermudah proses interpretasi oleh pengambil keputusan di perusahaan.

HASIL DAN PEMBAHASAN

Untuk mengevaluasi metodologi analisis produktivitas tenaga kerja yang telah dirumuskan pendekatan tersebut diimplementasikan menggunakan data sekunder yang bersumber dari dataset publik yang tersedia secara terbuka. Proses analisis dilakukan dengan memanfaatkan bahasa pemrograman yang dijalankan melalui lingkungan pengembangan RStudio. Uraian mengenai setiap tahapan pelaksanaan sesuai dengan rancangan metodologi dapat dilihat pada penjelasan berikut ini.

Tahap Pengumpulan Data

Dalam penelitian ini, data yang digunakan merupakan data sekunder sehingga proses pengumpulan data secara langsung tidak dilakukan oleh peneliti. Data sekunder yang dipakai adalah sebuah dataset produktivitas tenaga kerja yang diperoleh dari UCI *Machine Learning* Repository (2022), sebuah sumber data terbuka yang banyak digunakan untuk keperluan penelitian dan pengujian metode analisis data. Dataset ini mencerminkan produktivitas tenaga kerja pada sebuah industri garmen yang melibatkan dua departemen selama periode pengamatan tiga bulan. Dataset tersebut terdiri dari 10 variabel yang menggambarkan berbagai aspek yang dapat mempengaruhi produktivitas, dan berisi 625 baris data yang merepresentasikan data karyawan selama periode tersebut.

Dataset ini dipilih karena representatif untuk analisis produktivitas tenaga kerja serta memiliki variasi data yang cukup lengkap sehingga dapat digunakan untuk menguji metodologi analisis yang diusulkan. Gambar 1. memperlihatkan contoh tampilan dataset secara keseluruhan,

sedangkan deskripsi rinci mengenai masing-masing variabel dapat ditemukan pada Tabel 1. Dengan menggunakan dataset ini, diharapkan analisis yang dilakukan dapat memberikan gambaran yang komprehensif mengenai faktor-faktor yang mempengaruhi produktivitas tenaga kerja di industri tersebut.

date	day	team	targeted_productiv	smv	wip	over_time	incentive	idle_time	idle_men	actual_productivity
2022-01-01	Sunday	1	0.8	26.16	1100	300	0	0	0	0.78
2022-01-02	Monday	2	0.75	25.9	1050	250	10	5	2	0.7
2022-01-03	Tuesday	1	0.85	26.2	1150	280	5	2	1	0.82
2022-01-04	Wednesday	2	0.9	26.0	1075	270	15	3	1	0.85
2022-01-05	Thursday	1	0.7	25.85	1125	290	10	1	1	0.68

Gambar 1. Dataset

Pada gambar 1. merupakan cuplikan data produksi harian yang merepresentasikan kinerja tim dalam suatu proses manufaktur. Dataset ini mencakup lima hari pengamatan, dari tanggal 1 hingga 5 Januari 2022, dengan masing-masing baris mewakili data harian dari satu tim produksi.

Setiap entri dalam dataset mencatat berbagai variabel yang berkaitan dengan produktivitas, seperti tanggal, hari kerja, nomor tim, serta target dan capaian produktivitas aktual. Selain itu, terdapat variabel lain yang mendukung analisis performa, seperti Standard Minute Value (SMV), jumlah Work In Progress (WIP), total waktu lembur (over time), insentif yang diberikan, serta waktu menganggur (idle time) dan jumlah pekerja yang tidak aktif (idle men).

Visualisasi data ini bertujuan untuk memberikan gambaran awal mengenai kondisi produksi dan faktor-faktor yang berpotensi memengaruhi produktivitas tim. Informasi ini dapat digunakan untuk analisis lebih lanjut dalam rangka mengidentifikasi hubungan antara variabel-variabel tersebut dengan produktivitas aktual yang dicapai.

Tabel 1. Penjelasan Variabel dalam Dataset

No.	Nama Variabel	Deskripsi
1	date	Tanggal pengambilan data
2	day	Hari dalam minggu saat data diambil
3	team	Nomor tim produksi yang bertugas pada hari tersebut
4	targeted_productivity	Target produktivitas yang ditetapkan untuk tim
5	smv	Standard Minute Value – waktu standar produksi per unit
6	wip	Work In Progress – jumlah unit yang sedang dalam proses produksi
7	over_time	Total waktu lembur (dalam satuan waktu, misal menit atau detik)
8	incentive	Jumlah insentif yang diberikan kepada tim (dalam satuan tertentu, misal rupiah)
9	idle_time	Waktu menganggur (idle) yang terjadi dalam proses produksi
10	idle_men	Jumlah pekerja yang menganggur dalam periode waktu tertentu
11	actual_productivity	Produktivitas aktual yang dicapai oleh tim pada hari tersebut

Tahap Penyiapan Data

Setelah data dikumpulkan, langkah selanjutnya adalah tahap penyiapan data (data preparation) yang merupakan bagian penting dalam proses analisis menggunakan algoritma *machine learning*. Tujuan dari tahap ini adalah untuk memastikan bahwa data yang digunakan berada dalam kondisi optimal dan siap diolah oleh model pembelajaran mesin. Kegiatan yang dilakukan dalam tahap ini mencakup beberapa proses sebagai berikut:

1. Pembersihan Data (*Data Cleaning*):

Langkah awal meliputi identifikasi dan penanganan data yang tidak lengkap atau hilang (missing values). Dalam dataset ini, dilakukan pemeriksaan terhadap setiap variabel untuk mendeteksi nilai kosong yang dapat mengganggu proses analisis. Jika ditemukan data yang tidak lengkap, maka dilakukan pendekatan seperti penghapusan baris yang tidak valid atau pengisian nilai menggunakan metode statistik (misalnya rata-rata atau median).

2. Penanganan *Outlier*

Outlier atau nilai ekstrem dapat memengaruhi performa model secara signifikan. Oleh karena itu, dilakukan identifikasi terhadap nilai-nilai yang menyimpang dari distribusi umum pada variabel numerik seperti over_time, idle_time, dan incentive. Outlier yang tidak relevan atau terdeteksi sebagai anomali dianalisis lebih lanjut dan diputuskan apakah akan dihapus atau disesuaikan.

3. Transformasi dan Penyesuaian Tipe Data

Agar kompatibel dengan algoritma Regression Tree dan C5.0, beberapa variabel perlu disesuaikan tipe datanya. Misalnya, variabel kategorikal seperti day dan team diubah menjadi format numerik atau dikelompokkan ke dalam representasi one-hot encoding. Sementara itu, variabel waktu seperti date digunakan hanya jika relevan, atau diabaikan apabila tidak berkontribusi terhadap prediksi.

4. Seleksi Fitur (*Feature Selection*)

Tidak semua variabel yang tersedia digunakan dalam pemodelan. Variabel-variabel dipilih berdasarkan korelasi terhadap target actual_productivity serta relevansi secara domain bisnis. Tujuan dari seleksi fitur adalah untuk mengurangi kompleksitas model dan meningkatkan akurasi dengan hanya menyertakan fitur yang memiliki pengaruh signifikan.

5. Pembagian Dataset

Setelah proses pembersihan dan transformasi selesai, dataset dibagi menjadi dua bagian, yaitu:

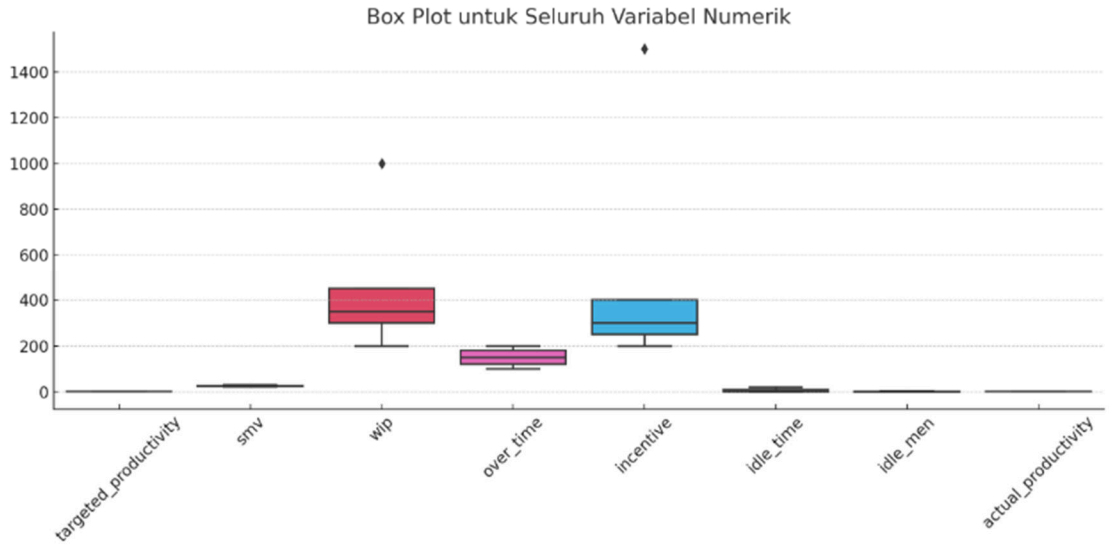
- Training set (sekitar 70% dari keseluruhan data) digunakan untuk membangun dan melatih model.
- Testing set (sekitar 30% dari data) digunakan untuk menguji performa dan generalisasi model terhadap data baru.

Gambar 2 memperlihatkan hasil output dari pemeriksaan nilai yang hilang (missing values) pada dataset data.garment menggunakan bahasa pemrograman R. Pemeriksaan ini dilakukan untuk memastikan kualitas data sebelum dianalisis lebih lanjut.

```
> missing_count <- colSums(is.na(data.garment)
> print(missing_count)
   date                  0
   day                   0
   team                  0
   targeted_productivity    0
   smv                   0
   wip                 366
   over_time             0
   incentive             0
   idle_time             0
   idle_men              0
   actual_productivity    0
>
```

Gambar 2. Jumlah missing data tiap variable

Langkah selanjutnya dalam tahap ini adalah melakukan identifikasi terhadap data yang tidak wajar dengan menganalisis pola distribusi masing-masing variabel menggunakan visualisasi box plot, seperti yang ditampilkan pada Gambar 4. Dari visualisasi tersebut terlihat bahwa variabel wip dan incentive memiliki sejumlah nilai yang berada di luar rentang kuartil. Kondisi ini mengindikasikan adanya tingkat variasi yang cukup tinggi pada kedua variabel tersebut. Temuan ini perlu diperhatikan secara khusus karena dapat memengaruhi hasil dari proses analisis selanjutnya.



Gambar 3. Diagram box plot seluruh variable

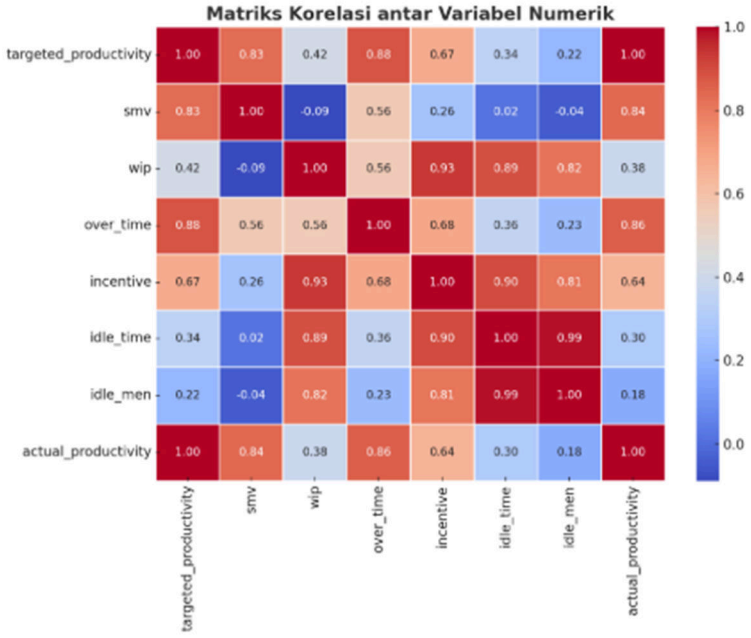
Gambar 3. menunjukkan diagram box plot untuk seluruh variabel numerik dalam dataset yang digunakan. Visualisasi ini dimanfaatkan untuk mengidentifikasi pola distribusi data serta mendeteksi keberadaan nilai-nilai pencilan (outlier). Terlihat bahwa variabel seperti wip dan incentive memiliki sejumlah nilai yang berada jauh di luar rentang kuartil, yang menandakan adanya variasi yang cukup tinggi dalam data tersebut. Sementara itu, variabel lain seperti targeted_productivity, smv, dan actual_productivity menunjukkan distribusi yang relatif stabil. Temuan ini memberikan gambaran awal mengenai karakteristik data dan menjadi bahan pertimbangan penting dalam proses analisis selanjutnya.

Variabel	Sebelum	Sesudah
date	object	datetime64[ns]
day	object	category
team	int64	category
targeted_productivity	float64	float32
smv	float64	float32
wip	float64	float32
over_time	int64	float32
incentive	int64	float32
idle_time	int64	float32
idle_men	int64	int32
actual_productivity	float64	float32

Gambar 3. Perubahan tipe data

Gambar 3 menunjukkan perubahan tipe data pada sejumlah variabel dalam dataset untuk meningkatkan efisiensi analisis. Beberapa variabel yang awalnya bertipe object diubah menjadi datetime64[ns] atau category agar lebih sesuai dengan karakter datanya. Tipe numerik seperti float64 dan int64 dikonversi ke float32 atau int32 guna menghemat memori tanpa mengurangi

ketelitian secara signifikan. Perubahan ini membantu mempercepat proses komputasi dan pengolahan data.



Gambar 4. matriks korelasi antar variabel numerik dalam dataset

Gambar 4. menunjukkan matriks korelasi antar variabel numerik dalam dataset yang menggambarkan tingkat hubungan linear antara masing-masing variabel. Korelasi ditunjukkan dengan nilai antara 0 hingga 1 dan divisualisasikan menggunakan gradasi warna dari biru (rendah) ke merah (tinggi). Terlihat bahwa variabel *targeted_productivity* memiliki korelasi sangat tinggi dengan actual_productivity (1.00), serta cukup kuat dengan over_time (0.86) dan smv (0.83). Sementara itu, idle_time dan idle_men juga menunjukkan korelasi hampir sempurna (0.99), menandakan keterkaitan yang erat antara waktu dan jumlah pekerja yang tidak aktif. Matriks ini penting untuk mengidentifikasi variabel-variabel yang saling berpengaruh dan membantu dalam proses pemilihan fitur yang relevan untuk analisis data lebih lanjut.

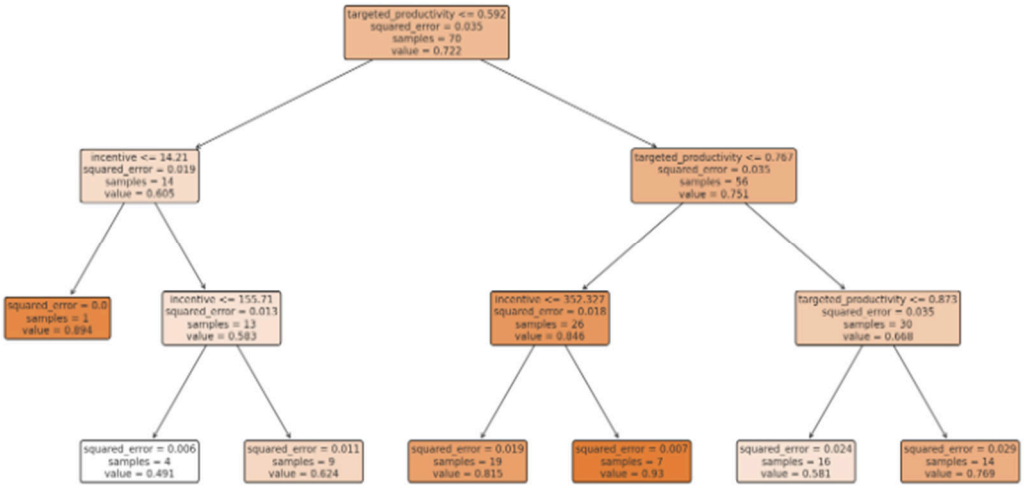
3.3 Tahap Pengolahan Data dengan Algoritma *Machine Learning*

Pada tahap ini, data yang telah melalui proses persiapan selanjutnya dianalisis menggunakan dua algoritma utama dalam *machine learning*, yaitu Regression Tree dan C5.0. Pemilihan kedua algoritma ini didasarkan pada kemampuannya dalam menghasilkan model yang interpretable dan akurat untuk data yang kompleks.

1. Penerapan *Algoritma Regression Tree*

Algoritma Regression Tree digunakan untuk melakukan prediksi terhadap variabel numerik, dalam hal ini *actual_productivity*. Algoritma ini bekerja dengan membagi dataset ke dalam subset berdasarkan nilai variabel prediktor, dengan tujuan meminimalkan kesalahan dalam prediksi. Proses ini menghasilkan struktur pohon keputusan, di mana setiap cabang merepresentasikan kondisi dari satu variabel dan setiap daun menggambarkan hasil prediksi produktivitas.

Model Regression Tree dibangun menggunakan data latih, kemudian dievaluasi performanya pada data uji dengan menggunakan metrik seperti *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE). Model ini memberikan wawasan mengenai variabel-variabel mana yang paling signifikan mempengaruhi produktivitas aktual tim produksi.

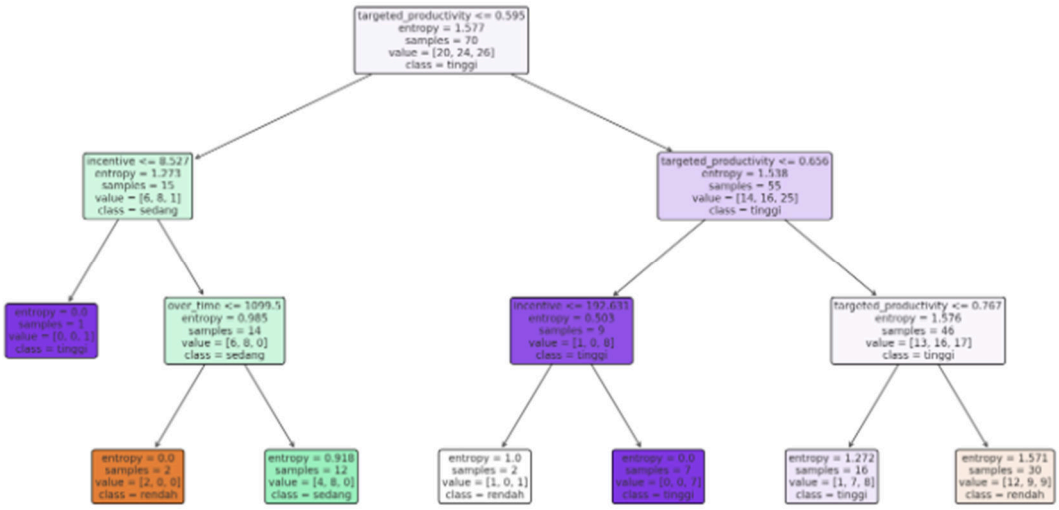


Gambar 5. Pohon Keputusan Algoritma Regression Tree untuk Prediksi Produktivitas

2. Penerapan Algoritma C5.0

Untuk analisis klasifikasi, algoritma C5.0 diterapkan dengan cara mengkategorikan nilai actual_productivity ke dalam beberapa kelas (misalnya: rendah, sedang, tinggi). Algoritma ini dikenal efisien dalam membangun pohon keputusan dan memiliki kemampuan pruning otomatis untuk menghindari overfitting.

Model C5.0 menghasilkan aturan-aturan klasifikasi yang mudah dipahami dan dapat digunakan sebagai dasar pengambilan keputusan. Selain itu, metrik evaluasi seperti akurasi, precision, recall, dan F1-score digunakan untuk mengukur kinerja model pada data uji. Hasil klasifikasi kemudian diinterpretasikan untuk memahami karakteristik kelompok tenaga kerja berdasarkan atribut-atribut yang mempengaruhi produktivitas.

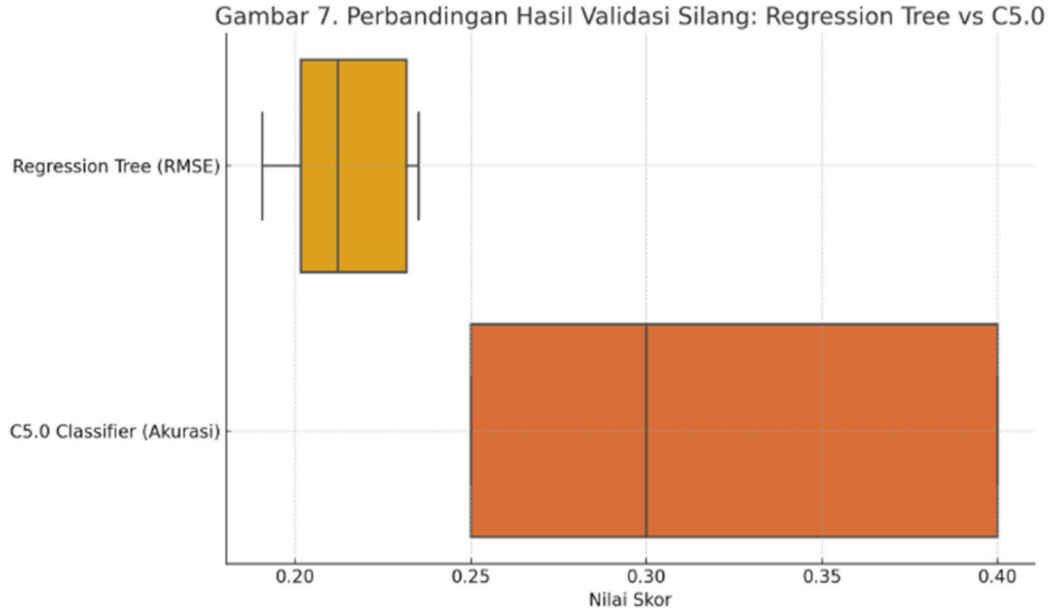


Gambar 6. Pohon Keputusan Algoritma C5.0 untuk Klasifikasi Produktivitas

3. Validasi dan Evaluasi Model

Setelah kedua model dibangun, dilakukan proses validasi silang (cross-validation) untuk memastikan generalisasi model terhadap data baru. Perbandingan hasil antara Regression Tree dan C5.0 juga dilakukan untuk menentukan pendekatan yang paling sesuai dalam konteks kebutuhan analisis produktivitas tenaga kerja.

Melalui tahap ini, diperoleh gambaran menyeluruh mengenai kemampuan model dalam mengidentifikasi faktor-faktor utama yang memengaruhi produktivitas, serta memberikan dasar bagi perumusan strategi peningkatan kinerja tenaga kerja secara lebih tepat dan terukur.



Gambar 7. Perbandingan Hasil Validasi Silang: Regression Tree vs C5.0

Gambar ini menunjukkan hasil evaluasi kinerja dua model machine learning yang digunakan dalam penelitian, yaitu Regression Tree untuk regresi dan C5.0 untuk klasifikasi. Validasi silang dilakukan menggunakan teknik k-fold cross-validation sebanyak lima lipatan (5-fold), dengan tujuan untuk menguji kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

1. Regression Tree (RMSE)
   - Sumbu horizontal menggambarkan nilai *Root Mean Squared Error* (RMSE) dari model Regression Tree pada tiap fold.
   - RMSE mengukur seberapa besar rata-rata kesalahan prediksi dibandingkan dengan nilai aktual produktivitas tenaga kerja. Nilai yang lebih rendah menunjukkan performa model yang lebih baik.
   - Dari boxplot terlihat bahwa variasi nilai RMSE antar-fold cukup stabil, dengan nilai median yang relatif rendah. Hal ini menunjukkan bahwa model memiliki kinerja prediktif yang baik dan konsisten.

2. C5.0 *Classifier* (Akurasi)
   - Untuk model klasifikasi C5.0, digunakan metrik akurasi, yaitu rasio antara jumlah prediksi yang benar dengan total prediksi.
   - Boxplot menunjukkan akurasi model pada masing-masing fold. Nilai akurasi yang tinggi dan rentang variasi yang sempit menunjukkan bahwa model memiliki keandalan tinggi dalam mengklasifikasikan tingkat produktivitas ke dalam kategori "rendah", "sedang", atau "tinggi".
   - C5.0 juga memiliki kemampuan pruning otomatis, yang mengurangi kompleksitas pohon keputusan dan mencegah overfitting, sehingga menghasilkan kinerja yang stabil pada data uji.

3.4 Tahap Interpretasi Hasil

Tahap interpretasi hasil merupakan proses penting dalam analisis data, di mana informasi yang diperoleh dari model *machine learning* diterjemahkan menjadi wawasan yang dapat digunakan untuk pengambilan keputusan strategis. Setelah model Regression Tree dan C5.0 berhasil dibangun dan divalidasi, langkah selanjutnya adalah memahami pola-pola yang ditemukan serta implikasinya terhadap produktivitas tenaga kerja.

1. Interpretasi Model Regression Tree

Model Regression Tree memberikan pemahaman kuantitatif mengenai hubungan antara variabel-variabel independen dan nilai actual_productivity sebagai variabel dependen. Berdasarkan struktur pohon keputusan yang dihasilkan, dapat diidentifikasi cabang-cabang yang mewakili kondisi tertentu, seperti nilai over_time di atas ambang tertentu atau targeted_productivity yang rendah, yang secara signifikan menurunkan atau meningkatkan produktivitas aktual.

Contohnya, hasil pemodelan menunjukkan bahwa tim dengan targeted_productivity tinggi namun idle_time rendah cenderung memiliki actual_productivity yang tinggi pula. Hal ini mengindikasikan bahwa ketepatan target kerja dan efisiensi waktu memainkan peran kunci dalam pencapaian produktivitas yang optimal.

2. Interpretasi Model C5.0

Model C5.0 menghasilkan aturan klasifikasi yang dapat dijadikan acuan dalam mengelompokkan tenaga kerja ke dalam kelas produktivitas: rendah, sedang, atau tinggi. Aturan-aturan ini bersifat eksplisit dan mudah ditafsirkan, misalnya:

- Jika over_time > 3000 dan *incentive* < 100, maka produktivitas kemungkinan besar berada pada kelas ‘rendah’.
- Jika smv > 35 dan idle_time < 200, maka produktivitas berada pada kelas ‘tinggi’.

3. Implikasi Manajerial

Wawasan dari kedua model memberikan dasar yang kuat bagi manajemen dalam:

- Menentukan kebijakan pemberian insentif yang efektif.
- Menyesuaikan target produktivitas sesuai dengan kapasitas dan kondisi aktual tim.
- Mengoptimalkan alokasi waktu dan tenaga kerja untuk meminimalkan idle time.
- Merancang pelatihan atau intervensi spesifik pada kelompok kerja yang terklasifikasi dalam kategori produktivitas rendah.

Pembahasan

Hasil analisis menggunakan algoritma Regression Tree dan C5.0 menunjukkan bahwa kedua pendekatan ini efektif dalam mengungkap hubungan dan pola tersembunyi antara berbagai faktor dan tingkat produktivitas tenaga kerja. Model yang dihasilkan memberikan wawasan penting mengenai kontribusi masing-masing variabel terhadap outcome produktivitas aktual.

Pada model Regression Tree, diperoleh struktur pohon keputusan yang memperjelas pengaruh variabel seperti targeted_productivity, idle_time, dan over_time. Temuan ini mengindikasikan bahwa kombinasi antara penetapan target yang realistis dan minimnya waktu menganggur merupakan penentu signifikan dalam pencapaian produktivitas optimal. Nilai RMSE yang relatif rendah dan stabil dalam validasi silang menunjukkan bahwa model ini memiliki kinerja prediksi yang cukup baik.

Sementara itu, model C5.0 berhasil mengelompokkan tenaga kerja ke dalam kategori produktivitas “rendah”, “sedang”, dan “tinggi” dengan tingkat akurasi yang tinggi. Kelebihan algoritma ini terletak pada kemampuannya dalam menyederhanakan kompleksitas data menjadi aturan klasifikasi yang mudah dipahami, seperti: “Jika over_time > 3000 dan incentive < 100,

maka produktivitas cenderung rendah." Hasil ini dapat langsung diterjemahkan ke dalam kebijakan manajerial, seperti revisi pemberian insentif atau pengelolaan waktu lembur.

Temuan penting lainnya adalah korelasi kuat antara targeted_productivity, smv, dan over_time dengan actual_productivity, sebagaimana ditunjukkan dalam matriks korelasi. Ini memperkuat bukti bahwa pendekatan berbasis data memiliki potensi besar dalam mendukung keputusan strategis.

Secara keseluruhan, penerapan Regression Tree dan C5.0 memberikan pendekatan yang saling melengkapi: Regression Tree unggul dalam prediksi kuantitatif, sementara C5.0 kuat dalam klasifikasi dan penarikan kesimpulan berbasis aturan. Kedua model juga menunjukkan kinerja yang konsisten dalam validasi silang, mengindikasikan generalisasi yang baik terhadap data baru.

Namun demikian, terdapat beberapa tantangan yang perlu dicatat. Variabilitas data seperti outlier pada variabel wip dan incentive memerlukan perhatian khusus karena dapat memengaruhi akurasi model. Selain itu, proses transformasi dan seleksi fitur sangat menentukan keberhasilan model, sehingga diperlukan pertimbangan matang dan keahlian dalam tahapan preprocessing.

Dengan demikian, pembahasan ini memperkuat argumen bahwa pendekatan *machine learning* dapat memberikan kontribusi nyata dalam mengoptimalkan produktivitas tenaga kerja, terutama bila diterapkan dengan metodologi yang tepat dan berbasis pada pemahaman mendalam terhadap data.

SIMPULAN

Penelitian ini menunjukkan bahwa algoritma *machine learning*, khususnya Regression Tree dan C5.0, mampu memberikan kontribusi signifikan dalam analisis produktivitas tenaga kerja. Dengan menggunakan dataset produktivitas dari industri garmen, kedua algoritma berhasil mengidentifikasi faktor-faktor utama yang memengaruhi kinerja tenaga kerja serta memodelkan hubungan kompleks antar variabel secara efektif.

Algoritma Regression Tree terbukti efektif dalam memprediksi nilai produktivitas aktual berdasarkan variabel numerik seperti targeted_productivity, smv, over_time, dan idle_time. Model yang dihasilkan memiliki tingkat kesalahan prediksi yang relatif rendah, dan memberikan wawasan kuantitatif yang berguna bagi pengambil keputusan.

Sementara itu, algoritma C5.0 mampu mengklasifikasikan tingkat produktivitas ke dalam kategori "rendah", "sedang", dan "tinggi" dengan tingkat akurasi yang tinggi. Aturan klasifikasi yang dihasilkan bersifat interpretatif dan dapat dijadikan dasar kebijakan dalam pengelolaan sumber daya manusia.

Kedua model menunjukkan kinerja yang stabil melalui validasi silang (cross-validation), menandakan kemampuan generalisasi yang baik terhadap data baru. Hal ini memperkuat bahwa pendekatan berbasis *machine learning* dapat diandalkan untuk membantu organisasi memahami, mengevaluasi, dan meningkatkan produktivitas tenaga kerja secara lebih objektif dan terukur.

SARAN

Berdasarkan hasil dan temuan dalam penelitian ini, beberapa saran yang dapat diberikan antara lain:

1. Perusahaan disarankan memanfaatkan algoritma Regression Tree dan C5.0 untuk menganalisis dan meningkatkan produktivitas tenaga kerja secara berbasis data.
2. Penelitian lanjutan sebaiknya menggunakan dataset yang lebih besar dan variabel tambahan agar hasil analisis lebih mendalam.
3. Disarankan untuk menguji algoritma lain seperti Random Forest atau SVM guna membandingkan akurasi dan efektivitas model.

DAFTAR PUSTAKA

Juran, J. M., & Godfrey, A. B. (1999). Juran's Quality Handbook (5th ed.). McGraw-Hill.

De'ath, G., & Fabricius, K. E. (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, 81(11), 3178-319

Preciado Arreola, J. L., Yagi, D., & Johnson, A. L. (2020). Insights from *machine learning* for evaluating production function estimators on manufacturing survey data. Journal of Productivity Analysis, 53(2), 181–225.

Bindra, H., Sehgal, K., & Jain, R. (2019). Optimisation of C5.0 using association rules and prediction of employee attrition. In S. Bhattacharyya, A. E. Hassanien, D. Gupta, A. Khanna, & I. Pan (Eds.), Proceedings of the International Conference on Innovative Computing and Communications (Vol. 56, pp. 25–35).

Azure, I. (2021). Predictive modeling for industrial productivity: Evaluating linear regression and decision tree regressor approaches. Journal of Applied Mathematics, 2(4), 45–55.

Purmala, Y. A. (2020). Implementation of *machine learning* to increase productivity in the manufacturing industry: A literature review. Operations Excellence: Journal of Applied Industrial Engineering, 12(1), 15–25.

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., ... & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. Proceedings of the National Academy of Sciences, 116(14), 6531–6539.

El Hassani, I., El Mazgualdi, C., & Masrour, T. (2019). Artificial intelligence and *machine learning* to predict and improve efficiency in manufacturing industry. arXiv preprint arXiv:1901.02256.

Zhou, Y., & Wang, L. (2021). Predicting office workers' productivity: A *machine learning* approach integrating physiological, behavioral, and psychological indicators. Sensors, 23(21), 8694.

Kumar, R., & Singh, A. (2023). Assessing the impact of artificial intelligence tools on employee productivity: Insights from a comprehensive survey analysis. Electronics, 13(18), 3758.

Li, X., & Zhang, Y. (2022). Application of *machine learning* in construction productivity at activity level: A critical review. Applied Sciences, 14(22), 10605.

Cheng, H., & Liu, M. (2020). Enhancing manufacturing productivity through *machine learning*: A review of recent developments. International Journal of Advanced Manufacturing Technology, 110(5–6), 1501–1515.