

Prediksi Gaji Pemain Bola Berdasarkan Statistiknya Menggunakan Algoritma Machine Learning Random Forest

Irkham Ramadhani

Program Studi : Sistem Informasi, Universitas Darwan Ali

Email : irkhamramadhani@gmail.com

ABSTRACT— This research aims to predict football players' salaries based on their statistical abilities and career track records, addressing the significant salary disparities within the industry. A data mining approach is employed, with the *Random Forest Regression* algorithm as the primary method, chosen for its capability to handle data complexity and its robustness against large outliers. A secondary dataset from Kaggle, comprising 17,954 rows and 51 columns of football player data, was processed through stages of handling missing values, feature selection, and target variable (*wage_euro*) designation. To mitigate issues of extreme outliers and skewed data distribution in salaries, a *logarithmic transformation* was applied to the *wage_euro* variable. The model was trained using 70% of the training data and evaluated with 30% of the testing data using RapidMiner software. Evaluation results show excellent model performance after the logarithmic transformation. The model achieved a Relative Error (MAPE) of 6.94% +/- 5.45% , indicating that salary predictions, on average, deviate by less than 7% from the actual values. Furthermore, a Root Mean Squared Error (RMSE) is 0.310 and an Absolute Error (MAE) of 0.246 +/- 0.189 (both on a logarithmic scale) demonstrate high accuracy and consistency. This superior performance is attributed to Random Forest's effectiveness in managing data complexity, while the logarithmic transformation successfully stabilized the salary scale. The developed model is capable of providing realistic and reliable salary estimations, aligning with the research objective to support efficient cost allocation and avoid under-prediction of salaries for the majority of players

Keywords— Salary Prediction, Football Players, Random Forest Regression, Logarithmic Transformation, Data Mining, RapidMiner

ABSTRAK— Penelitian ini bertujuan untuk memprediksi gaji pemain sepak bola berdasarkan statistik kemampuan dan rekam jejak mereka, guna mengatasi disparitas gaji yang signifikan di industri ini. Pendekatan data mining digunakan dengan algoritma *Random Forest Regresi* sebagai metode utama, dipilih karena kemampuannya dalam menangani kompleksitas data dan ketahanannya terhadap outlier besar. Dataset sekunder dari Kaggle yang mencakup 17.954 baris dan 51 kolom data pemain sepak bola diproses melalui tahapan penanganan nilai kosong, seleksi fitur, dan penentuan label (*wage_euro*). Untuk mengatasi masalah outlier ekstrem dan distribusi data gaji yang miring, *transformasi logaritmik* diterapkan pada variabel target (*wage_euro*). Model dilatih menggunakan 70% data training dan dievaluasi dengan 30% data testing menggunakan perangkat lunak RapidMiner. Hasil evaluasi model ML didapatkan performa yang lebih baik setelah penerapan *transformasi logaritmik*. Model mencapai *Relative Error* (MAPE) sebesar 6.94% +/- 5.45%, mengindikasikan bahwa prediksi gaji rata-rata meleset kurang dari 7% dari nilai aktual. Selain itu, *Root Mean Squared Error* (RMSE) 0.310 dan *Absolute Error* (MAE) 0.246 +/- 0.189 (dalam skala logaritmik) menunjukkan akurasi dan konsistensi yang tinggi. Performa superior ini dicapai karena *Random Forest* efektif menangani kompleksitas data, sementara *transformasi logaritmik* menstabilkan skala gaji. Model yang dikembangkan mampu memberikan estimasi gaji yang realistis dan andal, sejalan dengan tujuan penelitian untuk mendukung alokasi biaya yang efisien dan menghindari prediksi gaji yang terlalu rendah untuk sebagian besar pemain.

Kata kunci— Prediksi Gaji, Pemain Sepak Bola, Random Forest Regresi, Transformasi Logaritmik, Data Mining, RapidMiner.

I. PENDAHULUAN

Sepak bola adalah salah satu olahraga yang sangat di sukai secara global, dan menarik minat masyarakat dari berbagai kalangan dan usia. Kepopuleran ini tidak hanya sebatas hobi, namun juga menjanjikan prospek karier yang menarik, terutama dari sisi finansial, Benefit finansial yang didapatkan pemain sepak bola profesional memang sangat menarik, dengan nominal bayaran yang tidak kecil.

Fenomena gaji pemain sepak bola sering kali menunjukkan disparitas yang signifikan. Sebagai contoh, seorang pemain bintang dunia seperti *Cristiano Ronaldo* memiliki bayaran yang fantastis berkat kemampuan dan segudang prestasinya. Namun, pemain lain seperti *Scott McTominay*, yang juga bermain di klub besar sekelas *Manchester United*, menerima bayaran yang relatif lebih rendah dibandingkan pemain bintang di klub yang sama [1].

Perbedaan nominal gaji ini menimbulkan pertanyaan fundamental mengenai parameter apa saja yang secara objektif digunakan untuk menentukan besaran kompensasi seorang pemain. Kondisi fisik pemain, teknik yang dikuasai, kemampuan taktis, dan mental pemain seringkali menjadi faktor penentu prestasi yang berujung pada besaran gaji [2].

Untuk mengatasi permasalahan ini, penelitian ini dilaksanakan dengan harapan untuk bisa mengidentifikasi dan menentukan nominal gaji pemain sepak bola secara objektif berdasarkan berbagai parameter relevan, seperti kemampuan individu pemain dan rekam jejak pengalamannya.

Pendekatan ini diharapkan dapat memberikan dasar perhitungan yang akurat dan konsisten, serta menghindari potensi keberpihakan dalam penilaian gaji pemain.

Dalam rangka mencapai akurasi dan konsistensi, penelitian ini memanfaatkan pendekatan data mining, yang merupakan bagian dari machine learning. Algoritma *Random Forest* akan digunakan sebagai metode utama untuk melakukan prediksi ini.

Data mining atau dikenal juga sebagai *Knowledge Discovery in Data*, adalah sebuah proses untuk mengidentifikasi pola atau informasi tersembunyi dari sebuah dataset atau *data warehouse*. Umumnya, data mining bertujuan untuk melakukan mitigasi risiko, menyusun strategi yang lebih efektif, atau untuk mengurangi biaya operasional perusahaan [3]. Algoritma *Random Forest* akan digunakan sebagai metode utama untuk melakukan prediksi ini.

Random Forest merupakan salah satu model dalam *Machine Learning* yang umum digunakan untuk mengatasi studi kasus *klasifikasi* maupun *regresi*. Seperti yang dilakukan dalam penelitian dengan judul Prediksi potensial gempa bumi menggunakan Metode *Random forest* [4].

Model ini digunakan karena memiliki performa yang baik di bandingkan algoritma lainnya seperti yang di lakukan dalam penelitian berjudul Prediksi Kelulusan Predikat *Cum Laude* dengan Pendekatan *Machine Learning*, dimana peneliti tersebut membandingkan beberapa Algoritma ML seperti *Naive Bayes*, *C4.5*, *Logistic Model Trees*, *J48* dan *ID3*. Yang mana *random forest* memiliki tingkat akurasi sebesar 96% [5].

Bahkan sebuah penelitian dengan tujuan yang sama dengan judul untuk pengembangan Sistem Pendukung Keputusan untuk Menentukan Transfer Pemain Sepak Bola Menggunakan Metode AHP (*Analytical Hierarchy Process*) yang mana peneliti tersebut mengembangkan sistem pendukung keputusan

transfer setiap pemain sepak bola menggunakan metode AHP dan mendapatkan akurasi yang tinggi sebesar 80% [6].

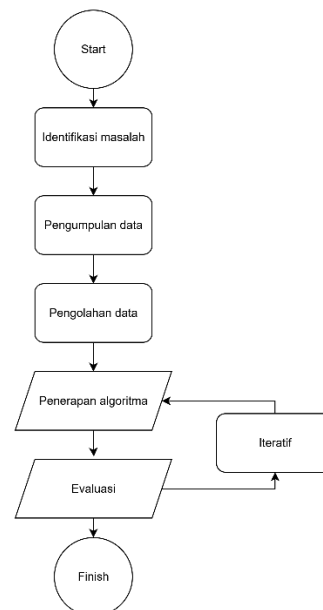
Pada akhir nya, penulis menggunakan *Random Forest* karena terbukti dapat bertahan di dataset memiliki *outlier* yang besar, namun yang menjadi perhatian adalah algoritma ini memiliki kekurangan seperti terlalu mudah mengalami *overfitting*. Maka dari itu perlu penyesuaian tentang parameter apa saja yang digunakan [7][8].

Perhitungan ini akan menggunakan bantuan tools bernama *Rapidminer*, sebuah perangkat lunak yang menyediakan sebuah lingkungan untuk melakukan penelitian hingga pelatihan menggunakan algoritma machine learning yang lengkap meliputi, persiapan data, visualisasi, pelatihan model hingga validasi dengan berbagai metrik [9]. Semua hal itu di bungkus ke dalam satu aplikasi yang mudah di operasikan bahkan bagi yang bukan background AI. Dan penulis menggunakan *Rapidminer* agar perhitungan lebih mudah dan meminimalisir kesalahan.

Dan penting bagi penulis untuk memastikan data yang digunakan dalam adalah data yang konkret dan terpercaya untuk menghasilkan keluaran yang akurat.

II. METODOLOGI PENELITIAN

Badan bagian ini menjelaskan tahapan-tahapan yang dilalui dalam melakukan penelitian, dimulai dari identifikasi masalah hingga evaluasi model. Alur penelitian ini dapat terlihat melalui diagram alir berikut:



Gambar 1. Diagram alir penelitian

A. Identifikasi Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, permasalahan utama dalam penelitian ini adalah menentukan besaran gaji pemain sepak bola secara objektif. Tujuan utamanya adalah membangun model

yang dapat memprediksi gaji pemain dengan mempertimbangkan berbagai parameter krusial seperti kemampuan fisik (misalnya kecepatan gerak atau respons yang lincah) dan predikat/prestasi pemain yang didapatkan selama karier mereka. Harapannya, hasil prediksi ini dapat membantu dalam efisiensi alokasi biaya, memastikan bayaran yang diterima sesuai dengan kapabilitas pemain, tidak terlalu besar maupun terlalu kecil.

B. Pengumpulan Data

Untuk mencapai tujuan penelitian, pengumpulan data yang relevan mengenai kemampuan dan prestasi pemain adalah langkah krusial. Data yang digunakan dalam penelitian ini diperoleh dengan jerih payah hingga dari sumber sekunder, yaitu dataset yang tersedia di platform Kaggle. Dataset ini berisi informasi komprehensif tentang pemain sepak bola dunia yang telah dikumpulkan oleh peneliti lain dan disediakan untuk umum. Penting untuk memastikan integritas dan keandalan data sekunder yang digunakan guna menjamin akurasi hasil akhir.

Dataset awal yang digunakan terdiri dari 17.954 baris dan 51 kolom dengan beragam tipe data (numerik, tekstual, boolean). Dataset ini dapat diakses melalui tautan berikut <https://www.kaggle.com/datasets/maso0dahmed/fotball-players-data>. Adapun penjelasan detail mengenai setiap kolom dalam dataset adalah sebagai berikut:

Tabel 1
 Penjelasan tiap kolom

Nama kolom	Penjelasan
<i>name</i>	Nama pemain
<i>full_name</i>	Nama lengkap pemain
<i>birth_date</i>	Tanggal lahir pemain
<i>age</i>	Usia pemain
<i>height_cm</i>	Tinggi pemain dalam sentimeter
<i>weight_kgs</i>	Berat pemain dalam kilogram
<i>positions</i>	Posisi yang bisa dimainkan pemain
<i>nationality</i>	Kewarganegaraan pemain
<i>overall_rating</i>	Penilaian keseluruhan pemain di FIFA
<i>potential</i>	Penilaian potensi pemain di FIFA
<i>value_euro</i>	Nilai pasar pemain dalam euro
<i>wage_euro</i>	Upah mingguan pemain dalam euro
<i>preferred_foot</i>	Kaki yang lebih disukai pemain
<i>international_reputation(1-5)</i>	Penilaian reputasi internasional dari 1 hingga 5
<i>weak_foot(1-5)</i>	Penilaian kaki yang lebih lemah pemain dari 1 hingga 5
<i>skill_moves(1-5)</i>	Penilaian gerakan keterampilan dari 1 hingga 5
<i>body_type</i>	Tipe tubuh pemain
<i>release_clause_euro</i>	Klausul rilis pemain dalam euro

<i>national_team</i>	Tim nasional pemain
<i>national_rating</i>	Penilaian di tim nasional
<i>national_team_position</i>	Posisi di tim nasional
<i>national_jersey_number</i>	Nomor jersey di tim nasional
<i>crossing</i>	Penilaian untuk kemampuan crossing
<i>finishing</i>	Penilaian untuk kemampuan finishing
<i>heading_accuracy</i>	Penilaian untuk akurasi heading
<i>short_passing</i>	Penilaian untuk kemampuan passing pendek
<i>volleys</i>	Penilaian untuk voli
<i>dribbling</i>	Penilaian untuk dribbling
<i>curve</i>	Penilaian untuk tembakan melengkung
<i>freekick_accuracy</i>	Penilaian untuk akurasi tendangan bebas
<i>long_passing</i>	Penilaian untuk passing panjang
<i>ball_control</i>	Penilaian untuk kontrol bola
<i>acceleration</i>	Penilaian untuk akselerasi
<i>sprint_speed</i>	Penilaian untuk kecepatan lari
<i>agility</i>	Penilaian untuk kelincahan
<i>reactions</i>	Penilaian untuk reaksi
<i>balance</i>	Penilaian untuk keseimbangan
<i>shot_power</i>	Penilaian untuk kekuatan tembakan
<i>jumping</i>	Penilaian untuk melompat
<i>stamina</i>	Penilaian untuk stamina
<i>strength</i>	Penilaian untuk kekuatan
<i>long_shots</i>	Penilaian untuk tembakan jarak jauh
<i>aggression</i>	Penilaian untuk agresi
<i>interceptions</i>	Penilaian untuk intersepsi
<i>positioning</i>	Penilaian untuk penempatan
<i>vision</i>	Penilaian untuk visi
<i>penalties</i>	Penilaian untuk penalti
<i>composure</i>	Penilaian untuk ketenangan
<i>marking</i>	Penilaian untuk penandaan
<i>standing_tackle</i>	Penilaian untuk tekel berdiri
<i>sliding_tackle</i>	Penilaian untuk tekel sliding

C. Pengolahan Data

Tahap pengolahan data bertujuan untuk menyiapkan data mentah agar siap digunakan dalam pemodelan, meminimalkan kesalahan, dan meningkatkan akurasi hasil akhir. Proses ini meliputi beberapa langkah:

- 1) Penanganan Nilai Kosong: Sebanyak 246 baris data yang mengandung nilai kosong dihapus dari total 17.954 baris, sehingga menyisakan 17.708 baris data bersih.

- 2) Seleksi Fitur (Kolom): Penulis melakukan seleksi fitur dengan menghilangkan kolom-kolom yang dianggap tidak relevan atau kurang berpengaruh terhadap nominal gaji, serta kolom dengan tipe data non-numerik yang tidak akan diproses lebih lanjut tanpa encoding tambahan. Kolom-kolom yang tidak digunakan antara lain: *birth_date*, *body_type*, *full_name*, *national_jersey_number*, *national_rating*, *national_team*, *national_team_position*, *nationality*, *positions*, *preferred_foot*, *release_clause_euro*, dan *value_euro*.
- 3) Penentuan Label (Variabel Target): Kolom *wage_euro* ditetapkan sebagai label atau variabel target, yang merepresentasikan besaran gaji mingguan pemain dalam *Euro*. Pemilihan ini didasarkan pada tujuan penelitian untuk memprediksi gaji, sehingga kolom seperti *release_clause_euro* dan *value_euro* dianggap kurang sesuai.
- 4) Penentuan ID: Kolom *name* dijadikan sebagai ID untuk memudahkan identifikasi pemain.
- 5) Normalisasi: Hal ini dilakukan agar dapat membantu meningkatkan akurasi dari hasil prediksi dan metode yang digunakan akan disesuaikan dengan kebutuhan dari dataset.
- 6) Jumlah Fitur Akhir: Setelah proses seleksi, dataset yang akan digunakan untuk pemodelan terdiri dari 39 kolom (fitur) yang relevan dan satu kolom label.
- 7) Pembagian Data: Data kemudian dibagi menjadi dua kelompok: data training (70%) untuk melatih model dan data testing (30%) untuk menguji performa model.

D. Penerapan Algoritma

Setelah data melalui tahap pengolahan, data siap untuk diterapkan pada algoritma *machine learning*. Dalam penelitian ini, algoritma *Random Forest Regresi* dipilih untuk membangun model yang mampu memprediksi gaji pemain sepak bola dengan hasil yang mendekati akurat. *Random Forest* adalah model *ensemble* yang terbentuk dari beberapa model *Decision Tree*, menggunakan metode *bootstrap* dan *random feature selection* untuk membentuk setiap *Decision Tree*. Algoritma *Random Forest* melibatkan beberapa tahapan [10]:

- 1) Menentukan nilai 'k' atau jumlah pohon (tree) yang akan dibentuk. Dalam *Rapidminer*, nilai ini dimasukkan secara manual pada parameter model.
- 2) Memilih sampel secara acak (*bagging*) untuk setiap pohon yang akan diolah menggunakan algoritma *Decision Tree*.
- 3) Mengulangi proses hingga mencapai nilai 'k' yang telah ditentukan.
- 4) Pada kasus prediksi regresi, hasil prediksi *Random Forest* merupakan nilai rata-rata dari prediksi seluruh model *Decision Tree*. Sedangkan pada saat kasus klasifikasi, hasil didasarkan pada mayoritas prediksi kelas itu sendiri.

E. Evaluasi model

Dilakukan untuk mengukur seberapa baik model yang telah dilatih mampu memprediksi gaji pemain sepak

bola. Evaluasi ini dilakukan secara *iteratif*, di mana model akan dilatih dan dioptimasi berulang kali hingga mencapai tingkat akurasi yang optimal. [11] Metrik *regresi* yang akan digunakan untuk menilai performa model adalah sebagai berikut.

- 1) Root Mean Squared Error (RMSE): Merupakan nilai akar dari rata-rata selisih antara nilai aktual dan nilai hasil prediksi. Semakin kecil nilai RMSE, semakin baik model dalam memprediksi.
- 2) Mean Absolute Error (MAE): Merupakan nilai rata-rata absolut dari selisih antara nilai aktual dan nilai hasil prediksi. Semakin kecil nilai MAE, semakin baik model dalam memprediksi.
- 3) Mean Absolute Percentage Error (MAPE): Merupakan nilai rata-rata dari selisih nilai aktual dengan nilai hasil prediksi dalam bentuk persentase. Dalam *RapidMiner*, metrik ini dikenal sebagai *relative_error*. Semakin kecil nilai MAPE, semakin baik model dalam memprediksi..

III. DESAIN, HASIL DAN PEMBAHASAN

A. Pengumpulan dataset

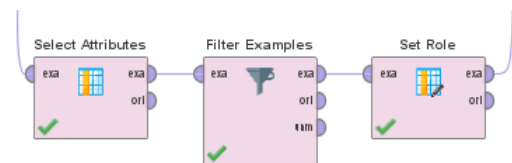


Gambar 2. Operator Retrieve

Operator *Retrieve* digunakan untuk menginput data *fifa_players* ke dalam runtime *RapidMiner*, sehingga data dapat diproses ke tahapan selanjutnya. Pada tahap ini, jumlah dataset masih 17.954 baris dan 51 kolom.

B. Pengolahan data

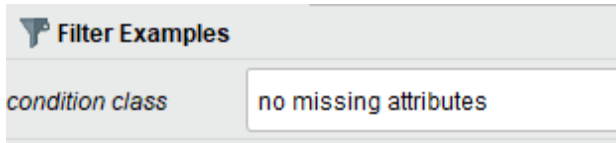
Selanjutnya, pada tahap pra-pemrosesan data, dilaksanakan pembersihan data dari nilai null, seleksi fitur, serta pemilihan ID dan label data. Alur ini direpresentasikan oleh operator *Select Attributes*, *Filter Examples*, dan *Set Role*.



Gambar 3. Operator pengolahan data

Operator *Select Attributes*, digunakan untuk menyeleksi fitur/kolom yang akan dipertahankan atau dihapus tanpa melalui file nya secara langsung.

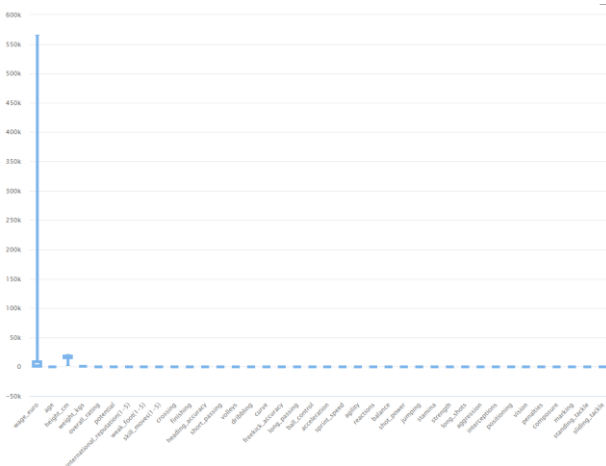
Operator *Filter Examples*, Memiliki beberapa fungsi yang bisa digunakan untuk memfilter data dengan query tertentu, akan tetapi disini digunakan untuk menyeleksi baris data yang mengandung nilai null/kosong.



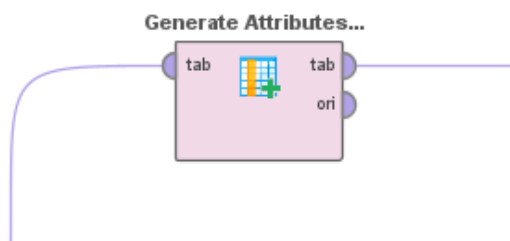
Gambar 4. Pengaturan parameter Filter Examples

Operator *Set Role*, digunakan untuk mengatur tipe apa saja pada beberapa kolom, misal nya di kolom *wage_euro* akan di ubah tipe nya menjadi label. Ini perlu dilakukan karena model memerlukan penanda sebuah label untuk belajar.

Kemudian bagian analisis data, bisa dilihat bahwa dataset memiliki data *outlier* ekstrem pada kolom *wage_euro*. kemungkinan karena ada pemain yang memiliki bayaran yang sangat besar dan artinya tidak ada yang salah dengan dataset, walaupun *decision tree* bisa bertahan dengan data *outlier*. Tetap harus dilakukan normalisasi agar hasil akhir yang lebih akurat.

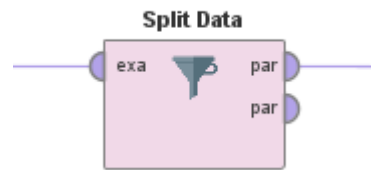


Maka dari itu penulis menggunakan teknik *Logarithmic Transformations* untuk mengurangi dampak dari data outlier [12], dengan menggunakan operator *Generate attribute*.



Gambar 5. Logarithmic Transformations

Operator *Generate attribute* berfungsi untuk menambahkan parameter khusus sehingga memungkinkan untuk membuat data baru bisa berupa angka random atau perubahan pada data seperti teknik *Logarithmic Transformations*.

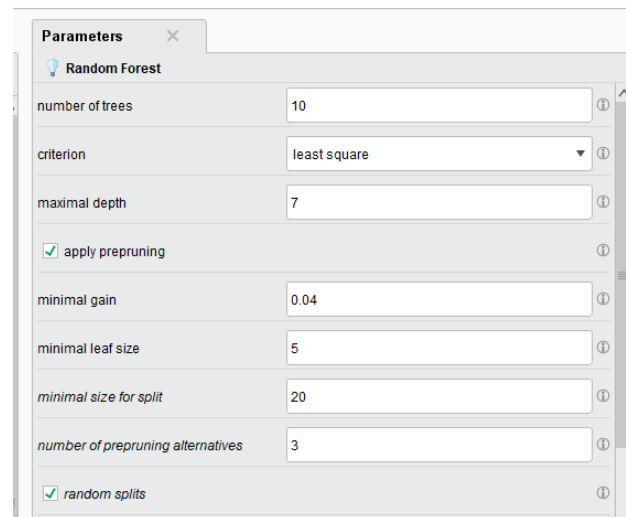


Gambar 6. Logarithmic Transformations

Operator *Split Data*, digunakan untuk membagi data menjadi beberapa sesuai kemauan bagian yaitu data training dan testing.

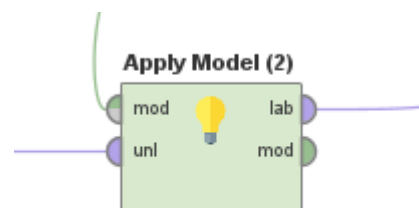
C. Penerapan Algoritma

Operator *Random Forest*, merupakan model ML yang akan digunakan untuk menganalisis data yang sebelumnya sudah di persiapkan dan dilakukan beberapa penyesuaian pada parameter yang digunakan seperti berikut.



Gambar 7. Parameter pada algoritma

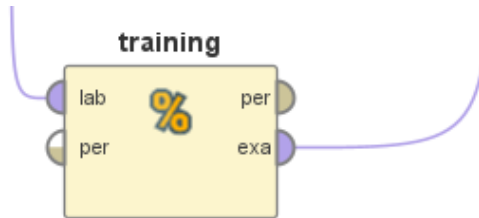
Kemudian data training dari operator *split data* akan dihubungkan agar model ML bisa berlatih dengan data yang memiliki label.



Gambar 8. Pelatihan model selesai

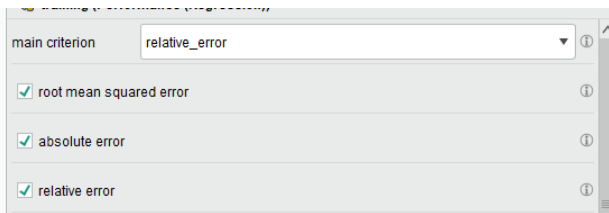
Operator Apply Model, model yang sebelumnya dilatih sekarang akan menggunakan data testing untuk di uji seberapa akurat hasil prediksi nya setelah berlatih menggunakan data training.

D. Validasi model



Gambar 9. Pelatihan model selesai

Operator Performance, disini model akan di nilai apakah hasil prediksi nya memiliki performa yang baik dan sesuai dengan tahapan awal. Bagian ini akan di ulang kembali sampai hasil performa yang baik di dapatkan. Operator ini memerlukan beberapa penyesuaiannya dengan metrik evaluasi yang sebelum nya di tentukan, yaitu:



Gambar 10. Metrik yang digunakan

Rapid miner memiliki beberapa perbedaan dalam penamaan metrik nya namun yang penulis pahami adalah absolute error adalah MAE sedangkan relative error adalah MAPA.

E. Iteratif

Melakukan pengulangan hingga akurasi model bisa lebih baik, hasil validasi model akan dijelaskan melalui tabel berikut.

Percobaan	Validasi model		
	RMSE	MAE	MAPE
Parameter Default	11007.379 +/- 0.000	4820.889 +/-	92.93% +/-
Apply prepruning	12237.374 +/- 0.000	5497.636 +/-	201.68% +/-
Logarithmic Transformations	0.283 +/- 0.000	0.218 +/- 0.180	6.13% +/-
			5.32%

IV. KESIMPULAN

Berdasarkan hasil evaluasi model setelah penerapan transformasi logaritmik pada variabel target wage_euro, diperoleh metrik kinerja sebagai berikut:

- 1) Root mean Squared Error (RMSE): 0.310. Nilai ini berada dalam skala logaritmik. Angka yang kecil ini menunjukkan bahwa rata-rata kesalahan kuadrat model sangat rendah pada skala log, mengindikasikan prediksi yang presisi.
- 2) Absolute Error (MAE): 0.246 +/- 0.189. Sama seperti RMSE, nilai ini juga dalam skala log. Rata-rata kesalahan absolut yang sangat rendah (0.246) dengan standar deviasi yang relatif kecil (0.189) menunjukkan bahwa model konsisten dalam meminimalkan kesalahan absolut di seluruh data.
- 3) Relative Error (MAPE): 6.94% +/- 5.45%. Ini adalah metrik kunci yang menunjukkan akurasi model dalam persentase. Dengan rata-rata kesalahan relatif hanya 6.94%, ini berarti prediksi gaji model secara rata-rata meleset kurang dari 7% dari nilai gaji sebenarnya. Standar deviasi yang kecil (5.45%) menunjukkan konsistensi yang sangat baik di seluruh pengujian, mengindikasikan bahwa model tidak hanya akurat tetapi juga stabil. Hasil ini merupakan peningkatan drastis dibandingkan dengan kondisi awal data yang memiliki relative_error lebih dari 200%

Performa yang sangat baik ini dicapai berkat kombinasi penggunaan algoritma Random Forest yang robust dan implementasi transformasi logaritmik pada variabel target wage_euro. Transformasi logaritma secara efektif "menjinakkan" distribusi gaji yang sangat miring dan mengurangi dampak ekstrem dari outlier (pemain bergaji sangat tinggi), sehingga memungkinkan Random Forest untuk belajar pola data secara lebih efektif dan akurat. Model ini kini mampu memberikan estimasi gaji yang realistis dan dapat diandalkan untuk sebagian besar pemain, sejalan dengan tujuan penelitian untuk memberikan "harga kasar" dan menghindari prediksi gaji yang terlalu kecil.

V. DAFTAR PUSTAKA

[1] Y. M. Sawitri, "5 Pemain Top Ini Gajinya Terlalu Rendah! Bayangkan Bayaran Scott McTominay Kalah Jauh dibanding Phil Jones," bola.com. Accessed: Jul. 28, 2025. [Online]. Available: <https://www.bola.com/dunia/read/4635125/5-pemain-top-ini-gajinya-terlalu-rendah-bayangkan-bayaran-scott-mctominay-kalah-jauh-dibanding-phil-jones>

[2] O. Maliki, H. Hadi, and I. F. Royana, "Analisis Kondisi Fisik Pemain Sepakbola Klub PERSEPU UPGRIS Tahun 2016," *Jendela Olahraga*, vol. 2, no. 2, pp. 1–8, 2017.

[3] D. Wardhani et al., *Pengantar Data Mining*, vol. 6, no. 3. 2023.

- [4] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection," *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 83–89, 2023, doi: 10.36080/idealism.v6i2.3036.
- [5] F. Indra Kusuma Budiyanoto, I. Hermadi, and M. K. D. Hardhienata, "Prediksi Performa Akademik Mahasiswa untuk Kelulusan Predikat Cum Laude dengan Pendekatan Machine Learning," *J. Ilmu Komput. dan Agri-Informatika*, vol. 11, no. 1, pp. 39–49, 2024, doi: 10.29244/jika.11.1.39-49.
- [6] H. P. K. Negara, E. Santoso, and H. Nurwasito, "Sistem Pendukung Keputusan Menentukan Transfer Pemain Sepak Bola Menggunakan Metode AHP (Analytical Hierarchy Process)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 7, pp. 2670–2678, 2018.
- [7] C. E. Murwaningtyas *et al.*, "Analisis Pengaruh Media Sosial terhadap Produktivitas Akademik Mahasiswa menggunakan Metode Decision Tree dan Random Forest," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 2, pp. 499–509, 2024.
- [8] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *J. Appl. Comput. Sci. Technol.*, vol. 4, no. 1, pp. 58–64, 2023, doi: 10.52158/jacost.v4i1.491.
- [9] H. Rizqifaluthi and M. A. Yaqin, "Process Mining Akademik Sekolah menggunakan RapidMiner," *Matics*, vol. 10, no. 2, p. 47, 2019, doi: 10.18860/mat.v10i2.5158.
- [10] T. Purwa, "Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017)," *J. Mat. Stat. dan Komputasi*, vol. 16, no. 1, p. 58, 2019, doi: 10.20956/jmsk.v16i1.6494.
- [11] Trivusi, "Perbedaan MAE, MSE, RMSE, dan MAPE pada Data Science," *trivusi.web.id*. Accessed: Jul. 28, 2025. [Online]. Available: <https://www.trivusi.web.id/2023/03/perbedaan-mae-mse-rmse-dan-mape.html>
- [12] K. Benoit, "Linear Regression Models with Logarithmic Transformations," *London Sch. Econ.*, pp. 1–8, 2011, [Online]. Available: <http://www.kenbenoit.net/courses/ME104/logmodels2.pdf>