Jurnal Pseudocode Vol 13 (1) Februari 2026, hal 69-74

ISSN 2355-5920 | e-ISSN 2655-1845
https://ejournal.unib.ac.id/pseudocode/index

# Analisis Prediksi Probabilitas Otomatisasi Pekerjaan Tahun 2030 Menggunakan Algoritma *Linear Regression* dan *Gradient Boosting*

Nicholas Leonardo, Ahmad Zidane Arrasyid, Muhammad Arif Billah Natagama*, Hafidz Muhammad Dzaky, Vitri Tundjungsari

Prodi Teknik Informatika, Universitas Esa Unggul
Jl. Arjuna Utara No.9, Duri Kepa, Kec. Kb. Jeruk, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta 11510

Informasi Naskah:
Diterima: 19 Januari 2026/ Direview: 24 Januari 2026/ Direvisi: 31 Januari 2026/ Disetujui Terbit: 05 Februari 2026
DOI: 10.33369/pseudocode.13.1.69-74
*Korespondensi: gama82059@student.esaunggul.ac.id

***Abstract***

*The rapid development of artificial intelligence and automation is expected to significantly impact future employment. This study aims to predict job automation probability in 2030 using supervised learning methods. A public* dataset *containing job types, education levels, and automation probabilities was utilized. Linear Regression and* XGBoost *Regressor were employed to build and compare predictive models. The research process included* data *preprocessing, training–testing* data *split, model training, and performance evaluation using Root Mean Square Error (RMSE) and coefficient of determination ($R^2$). Experimental results indicate that* XGBoost *outperforms Linear Regression by achieving lower RMSE and higher $R^2$ values. This study provides insights into automation risks and may support workforce skill development planning.*

*Keywords: job automation; supervised learning; linear regression;* XGBoost

## 1. Pendahuluan

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*/AI) dan sistem otomatisasi telah membawa perubahan signifikan terhadap struktur dan karakteristik dunia kerja modern. Berbagai sektor industri mulai mengadopsi teknologi berbasis AI untuk meningkatkan efisiensi, produktivitas, dan akurasi proses bisnis. Transformasi ini berdampak langsung pada pola kebutuhan tenaga kerja, di mana beberapa jenis pekerjaan mengalami pergeseran peran, sementara pekerjaan lain berpotensi mengalami pengurangan atau bahkan tergantikan oleh sistem otomatis. Fenomena tersebut menimbulkan kekhawatiran global terkait keberlanjutan pekerjaan manusia di masa depan, khususnya menjelang tahun 2030 [1], [2].

Sejumlah penelitian menunjukkan bahwa otomatisasi cenderung memengaruhi pekerjaan yang bersifat rutin, berulang, dan terstruktur, baik pada sektor manufaktur maupun jasa. Penerapan teknologi AI, *machine learning*, dan robotika telah mempercepat proses otomatisasi pada aktivitas yang sebelumnya membutuhkan intervensi manusia [3]. Kondisi ini tidak hanya berdampak pada jumlah lapangan kerja, tetapi juga pada perubahan kompetensi dan keterampilan yang dibutuhkan oleh tenaga kerja agar tetap relevan di era digital. Oleh karena itu, analisis terhadap risiko otomatisasi pekerjaan menjadi isu penting dalam perencanaan sumber daya manusia dan kebijakan ketenagakerjaan [4], [5], [6], [7].

Pendekatan berbasis data menjadi salah satu metode yang efektif untuk memahami dan memprediksi risiko otomatisasi pekerjaan. Dengan ketersediaan data historis dan proyeksi masa depan, metode *machine learning* dapat digunakan untuk memodelkan hubungan kompleks antara karakteristik pekerjaan, tingkat pendidikan, dan probabilitas otomatisasi. *Supervised learning*, sebagai salah satu cabang pembelajaran mesin, memungkinkan model untuk mempelajari pola dari data berlabel sehingga dapat menghasilkan prediksi yang lebih akurat dibandingkan pendekatan statistik konvensional [8]. Pendekatan ini sangat relevan untuk permasalahan prediksi numerik, seperti estimasi probabilitas otomatisasi pekerjaan.

Beberapa studi terdahulu telah membahas dampak AI dan otomatisasi terhadap dunia kerja. Frey dan Osborne mengemukakan bahwa banyak profesi memiliki probabilitas tinggi untuk terotomatisasi akibat kemajuan teknologi komputer. Penelitian lain juga menunjukkan bahwa algoritma pembelajaran mesin mampu meningkatkan akurasi prediksi dalam berbagai permasalahan sosial ekonomi, termasuk analisis risiko dan prediksi perilaku tenaga kerja. Model regresi sering digunakan sebagai *baseline* karena kemudahan interpretasinya, sementara metode *ensemble learning* terbukti lebih unggul dalam menangani hubungan non-linear dan interaksi antar variabel [9], [10].

*Linear Regression* merupakan metode statistik yang umum digunakan untuk memodelkan hubungan linier antara variabel independen dan variabel dependen. Metode ini sering dijadikan sebagai model dasar dalam penelitian prediksi karena kesederhanaan dan kemudahan interpretasinya. Namun, keterbatasan *Linear Regression* terletak pada asumsi hubungan linier antar variabel, sehingga kurang optimal dalam menangkap pola data yang kompleks. Sebaliknya, metode *ensemble* seperti *Extreme Gradient Boosting* (XGBoost) mampu memodelkan hubungan non-linear melalui kombinasi sejumlah pohon keputusan yang dioptimasi secara bertahap, sehingga menghasilkan performa prediksi yang lebih baik pada dataset dengan karakteristik kompleks [11], [12].

Jurnal Pseudocode Vol 13 (1) Februari 2026, hal 69-74

Meskipun penelitian mengenai otomatisasi pekerjaan dan penggunaan *machine learning* telah banyak dilakukan, kajian yang secara spesifik membandingkan kinerja model linier dan model *ensemble* dalam memprediksi probabilitas otomatisasi pekerjaan di masa depan, khususnya dengan mempertimbangkan tingkat pendidikan dan jenis pekerjaan, masih relatif terbatas. Oleh karena itu, diperlukan penelitian yang mengkaji efektivitas berbagai algoritma *supervised learning* dalam memprediksi risiko otomatisasi pekerjaan secara kuantitatif.

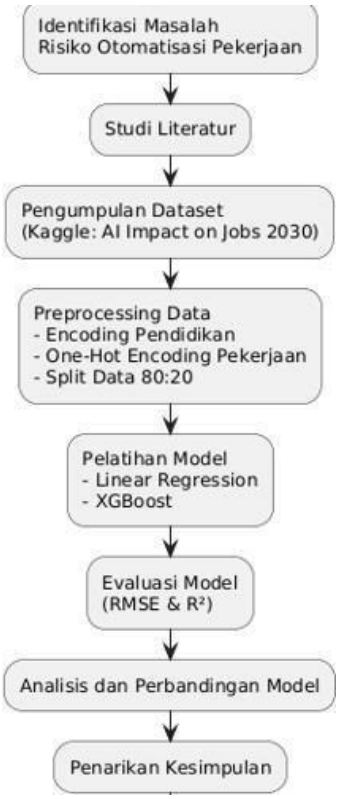
Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk memprediksi probabilitas otomatisasi pekerjaan pada tahun 2030 menggunakan pendekatan *supervised learning*. Dua algoritma *machine learning*, yaitu *Linear Regression* dan XGBoost *Regressor*, digunakan untuk membangun dan membandingkan model prediksi berdasarkan karakteristik pekerjaan dan tingkat pendidikan. Evaluasi performa model dilakukan menggunakan metrik *Root Mean Square Error* (RMSE) dan koefisien determinasi ($R^2$). Hasil penelitian ini diharapkan dapat memberikan gambaran awal mengenai risiko otomatisasi berbagai jenis pekerjaan serta menjadi referensi bagi individu, institusi pendidikan, dan pembuat kebijakan dalam merancang strategi pengembangan keterampilan tenaga kerja di masa depan.

## 2. Metodologi Penelitian

Tujuan penelitian ini adalah mengembangkan model prediksi risiko otomatisasi pekerjaan pada tahun 2030 berbasis supervised learning, dengan tahapan metodologis yang disusun secara sistematis sehingga memungkinkan replikasi oleh peneliti selanjutnya. Proses penelitian meliputi tahap pengumpulan dataset, *preprocessing* data, pembagian data latih dan data uji, pelatihan model menggunakan algoritma *Linear Regression* dan XGBoost [13], serta evaluasi performa model menggunakan metrik *Root Mean Square Error* (RMSE) dan koefisien determinasi ($R^2$). Proses pelatihan model menerapkan algoritma *Linear Regression* dan XGBoost, dengan evaluasi performa dilakukan menggunakan metrik *Root Mean Square Error* (RMSE) serta koefisien determinasi ($R^2$).

### 2.1. Rancangan Penelitian

Penelitian ini menggunakan pendekatan eksperimental yang didasarkan pada metode *supervised learning* dalam proses pengembangan model. Model prediksi dibangun untuk menganalisis dan memprediksi probabilitas otomatisasi pekerjaan pada tahun 2030 berdasarkan karakteristik pekerjaan dan tingkat pendidikan. Dua algoritma *machine learning*, yaitu *Linear Regression* dan XGBoost, digunakan untuk membandingkan kinerja model dalam memprediksi risiko otomatisasi pekerjaan. Alur penelitian secara umum meliputi pengumpulan data, *preprocessing* data, pelatihan model, evaluasi model, dan analisis hasil, sebagaimana ditunjukkan pada Gbr. 1.




Gbr. 1. Diagram Alur Penelitian

### 2.2. Teknik Pengumpulan Data

Sumber data dalam penelitian ini berasal dari data sekunder yang diperoleh melalui platform Kaggle. Dataset yang dipakai berjudul "*AI Impact on Jobs 2030*" yang berisi informasi mengenai jenis pekerjaan, tingkat pendidikan, kategori risiko, serta probabilitas otomatisasi pekerjaan pada tahun 2030. Dataset ini digunakan secara khusus untuk keperluan analisis dan penelitian akademik.

### 2.3. *Preprocessing* Data

Tahap *preprocessing* data bertujuan untuk meningkatkan kualitas data sehingga dapat digunakan secara optimal dalam proses analisis. Adapun tahapan *preprocessing* yang dilakukan meliputi:

1. Pembersihan Data, merupakan tahapan awal preprocessing yang dilakukan untuk memastikan kualitas data dengan cara menangani data yang tidak lengkap (*missing values*) dan mengeliminasi data yang tidak relevan terhadap tujuan penelitian.
2. *Encoding* Data Kategorikal, dengan mengubah atribut kategorikal seperti tingkat pendidikan dan jenis pekerjaan menjadi representasi numerik menggunakan teknik *encoding* yang sesuai.
3. Normalisasi Data Numerik, yaitu tahapan penyesuaian skala fitur numerik untuk memastikan kesetaraan rentang nilai, sehingga proses pelatihan model dapat berjalan secara optimal tanpa bias.

Jurnal Pseudocode Vol 13 (1) Februari 2026, hal 69-74

4. P**e**mbagian Data, merupakan tahapan *preprocessing* yang bertujuan untuk membagi dataset menjadi data latih (*training data*) dan data uji (*testing data*) dengan rasio 80:20 sebagai dasar evaluasi performa model.

### 2.4. Metode *Linear Regression* dan XGBoost *Regressor*

Metode Prediksi yang digunakan dalam penelitian ini terdiri dari dua pendekatan *supervised learning*, yaitu *Linear Regression* dan XGBoost *Regressor* [11]. *Linear Regression* diterapkan sebagai model *baseline* untuk memodelkan hubungan linier antara variabel independen dan probabilitas otomatisasi pekerjaan. Sementara itu, XGBoost diterapkan untuk memodelkan hubungan non-linear dan interaksi antar fitur yang kompleks, yang tidak dapat ditangani secara optimal oleh model linier [10], [13]. Kedua model dilatih pada data latih dan selanjutnya dievaluasi menggunakan data uji guna mengukur performa prediksi.

### 2.5. Implementasi Sistem

Eksperimen pada penelitian ini dilakukan menggunakan platform Google Colaboratory (Google Colab) dengan dukungan bahasa pemrograman Python. Proses eksperimen mencakup tahapan pemuatan dataset, *preprocessing* data, pelatihan model menggunakan algoritma *Linear Regression* dan XGBoost, serta evaluasi performa model berdasarkan metrik *Root Mean Square Error* (RMSE) dan koefisien determinasi ($R^2$).

### 2.6. Proses Eksperimen dan Pemodelan

Proses eksperimen pada penelitian ini diawali dengan pemuatan dataset "*AI Impact on Jobs 2030*" yang diperoleh dari platform Kaggle ke dalam lingkungan Google Colaboratory (Google Colab). Dataset yang digunakan kemudian diperiksa untuk memastikan kelengkapan atribut dan konsistensi data. Data yang tidak lengkap atau tidak relevan terhadap tujuan penelitian diidentifikasi dan ditangani pada tahap awal sebelum proses pemodelan dilakukan.

Tahap selanjutnya adalah *preprocessing* data yang bertujuan untuk menyiapkan dataset agar sesuai untuk diproses oleh algoritma *machine learning*. Pada tahap ini, atribut tingkat pendidikan dikodekan menggunakan teknik *ordinal encoding*, sedangkan atribut jenis pekerjaan dikodekan menggunakan *one-hot encoding*. Setelah proses pengkodean, variabel fitur (X) dan variabel target (y) dipisahkan, dengan probabilitas otomatisasi pekerjaan sebagai variabel target. Dataset kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 untuk mendukung proses evaluasi performa model secara objektif.

Setelah *preprocessing* selesai, dilakukan pelatihan model *machine learning* menggunakan dua algoritma *supervised learning*, yaitu *Linear Regression* dan XGBoost *Regressor*. Kedua model dilatih secara terpisah menggunakan data latih untuk mempelajari hubungan antara karakteristik pekerjaan, tingkat pendidikan, dan probabilitas otomatisasi pekerjaan.

### 2.7. Evaluasi Model dan Penyajian Hasil

Model yang telah melalui tahap pelatihan selanjutnya dievaluasi menggunakan data uji. Evaluasi performa model dilakukan dengan menggunakan metrik *Root Mean Square Error* (RMSE) dan koefisien determinasi ($R^2$). Nilai RMSE digunakan untuk mengukur tingkat kesalahan prediksi, sedangkan nilai $R^2$ digunakan untuk menilai kemampuan model dalam menjelaskan variasi data target.

Hasil evaluasi dari masing-masing model digunakan untuk melakukan perbandingan kinerja antara *Linear Regression* dan XGBoost. Selain itu, hasil prediksi probabilitas otomatisasi pekerjaan dari kedua model dianalisis dengan membandingkan nilai prediksi terhadap nilai aktual. Ringkasan hasil evaluasi dan analisis perbandingan model disajikan sebagai dasar dalam pembahasan dan penarikan kesimpulan penelitian.

## 3. Hasil dan Pembahasan

Bab ini memaparkan hasil implementasi dan evaluasi model prediksi risiko otomatisasi pekerjaan pada tahun 2030 dengan pendekatan *Linear Regression* dan XGBoost. Fokus pembahasan pada bab ini meliputi hasil pengolahan data, analisis performa masing-masing model berdasarkan metrik *Root Mean Square Error* (RMSE) dan koefisien determinasi ($R^2$), serta kajian perbandingan kinerja kedua algoritma dalam memprediksi probabilitas otomatisasi pekerjaan.

### 3.1. Karakteristik Dataset

Penelitian ini menggunakan dataset berjudul *AI Impact on Jobs 2030* yang diperoleh dari platform Kaggle. Dataset berhasil dimuat dan terdiri dari sejumlah data pekerjaan yang merepresentasikan berbagai jenis profesi dan tingkat pendidikan. Setiap data memiliki beberapa atribut utama, antara lain jenis pekerjaan (*job title*), tingkat pendidikan *(education level*), kategori risiko, serta probabilitas otomatisasi pekerjaan pada tahun 2030. Penelitian ini memfokuskan pada atribut jenis pekerjaan, tingkat pendidikan, dan probabilitas otomatisasi sebagai variabel utama dalam proses pelatihan dan evaluasi model prediksi. Struktur atribut dataset yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Struktur Atribut Dataset *AI Impact on Jobs 2030*

| No | Atribut | Deskripsi |
|---|---|---|
| 1 | Job_Title | Jenis atau nama pekerjaan |
| 2 | Education_Level | Tingkat pendidikan yang dibutuhkan |

Jurnal Pseudocode Vol 13 (1) Februari 2026, hal 69-74

| | | |
|---|---|---|
| 3 | Automation_Probability_2030 | Probabilitas otomatisasi pekerjaan |
| 4 | Risk_Category | Kategori tingkat risiko otomatisasi |
| 5 | Industry | Sektor industri pekerjaan |

### 3.2. Identifikasi Atribut Dataset

Berdasarkan hasil pemrosesan awal dataset, sistem secara otomatis mengidentifikasi atribut atribut utama yang digunakan dalam penelitian. Atribut Job_Title digunakan sebagai representasi jenis pekerjaan [11], Education_Level sebagai karakteristik latar belakang pendidikan, dan Automation_Probability_2 030 sebagai variabel target (label) yang menjadi sasaran prediksi model. Identifikasi atribut bertujuan untuk menjamin kesesuaian dan konsistensi data yang digunakan dalam tahapan *preprocessing* dan pemodelan, sehingga meminimalkan potensi kesalahan pada proses pelatihan model serta evaluasi hasil prediksi.

### 3.3. *Preprocessing* Data

Tahap *preprocessing* data dilakukan untuk mempersiapkan dataset sebelum digunakan dalam proses pelatihan model machine learning. Dataset yang digunakan masih mengandung data kategorikal dan numerik yang tidak dapat langsung diproses oleh algoritma prediksi. Pada tahap awal, dataset disajikan dalam bentuk mentah untuk merepresentasikan kondisi awal data. Selanjutnya, dilakukan proses pengkodean Tingkat Pendidikan menggunakan *ordinal encoding* dan pengkodean jenis pekerjaan menggunakan *one hot encoding*, dengan tujuan mengonversi data kategorikal menjadi bentuk numerik yang sesuai untuk pemrosesan oleh algoritma *machine learning*. Pada tahap ini dilakukan pemisahan variabel fitur (X) dan variabel target (y), dengan probabilitas otomatisasi pekerjaan sebagai variabel target prediksi. Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan rasio 80:20. Contoh hasil *preprocessing* data sebelum dan sesudah *encoding* disajikan pada Tabel 2.

Tabel 2. Contoh Data Sebelum dan Sesudah *Preprocessing*

| No | Job Title | Education Level | Education Encoded | Automation Probability |
|---|---|---|---|---|
| 1 | Factory Worker | High School | 0 | 0.78 |
| 2 | Data Analyst | Bachelor’s | 1 | 0.42 |
| 3 | Software Engineer | Master’s | 2 | 0.21 |
| 4 | Administrative Staff | Bachelor’s | 1 | 0.55 |
| 5 | Machine Operator | High School | 0 | 0.68 |

Tabel 2 menampilkan contoh data sebelum dan sesudah dilakukan *preprocessing*. Berdasarkan tabel tersebut, dapat dilihat bahwa sebelum *preprocessing* data masih mengandung atribut kategorikal dan format yang belum dapat diproses secara langsung oleh algoritma *machine learning*. Setelah *preprocessing*, data telah diubah ke dalam bentuk numerik melalui proses *encoding* serta disusun secara terstruktur, sehingga lebih konsisten dan siap digunakan dalam proses pemodelan. Hasil *preprocessing* ini memastikan bahwa dataset telah berada dalam format yang sesuai untuk tahap selanjutnya, yaitu proses pelatihan dan evaluasi model prediksi risiko otomatisasi pekerjaan menggunakan metode *Linear Regression* dan XGBoost.

### 3.4. Hasil Prediksi Probabilitas Otomasi Pekerjaan

Setelah dataset melalui tahap *preprocessing*, sistem melakukan proses prediksi probabilitas otomatisasi pekerjaan menggunakan dua algoritma *supervised learning*, yaitu *Linear Regression* dan XGBoost [14]. Tahap ini bertujuan untuk mengestimasi tingkat risiko otomatisasi pada berbagai jenis pekerjaan berdasarkan karakteristik pekerjaan dan tingkat pendidikan yang terdapat dalam dataset.Pada pengujian yang dilakukan, sistem menggunakan data uji untuk menghasilkan nilai prediksi probabilitas otomatisasi pekerjaan. Setiap data pekerjaan diproses oleh kedua model untuk memperoleh nilai prediksi yang kemudian dibandingkan dengan nilai aktual [15]. Hasil prediksi tersebut digunakan untuk mengevaluasi kinerja masing-masing model.

Tabel 3. menyajikan contoh lima data pekerjaan beserta nilai probabilitas otomatisasi aktual dan hasil prediksi yang dihasilkan oleh model *Linear Regression* dan XGBoost.

Tabel 3. Contoh Hasil Prediksi Probabilitas Otomatisasi Pekerjaan

| No | Jenis Pekerjaan | Tingkat Pendidikan | Nilai Aktual | Prediksi Linear Regression | Prediksi XGBoost |
|---|---|---|---|---|---|
| 1 | Factory Worker | High School | 0,78 | 0,75 | 0,80 |
| 2 | Data Analyst | Bachelor’s | 0,42 | 0,45 | 0,43 |
| 3 | Software Engineer | Master’s | 0,21 | 0,25 | 0,23 |
| 4 | Administrati | Bachelor | 0,55 | 0,52 | 0,56 |

Jurnal Pseudocode Vol 13 (1) Februari 2026, hal 69-74

| | | | | | |
|---|---|---|---|---|---|
| | ve Staff | 's | | | |
| 5 | Machine Operator | High School | 0,68 | 0,65 | 0,70 |

Berdasarkan hasil prediksi, dapat disimpulkan bahwa kedua model memiliki kemampuan yang baik dalam memprediksi probabilitas otomatisasi dengan nilai yang mendekati nilai aktual. Model XGBoost cenderung memberikan prediksi yang lebih dekat terhadap nilai aktual pada pekerjaan dengan pola data yang kompleks [12], [16], sementara *Linear Regression* menunjukkan performa yang stabil pada data dengan hubungan prediksi. Hasil penelitian ini menunjukkan bahwa pendekatan *supervised learning* memiliki efektivitas yang baik dalam memprediksi risiko otomatisasi pekerjaan, serta menjadi landasan bagi analisis perbandingan kinerja model pada tahap selanjutnya.

### 3.5. Pembahasan

Hasil prediksi probabilitas otomatisasi pekerjaan pada tahap sebelumnya menunjukkan bahwa pendekatan *supervised learning* mampu mengidentifikasi tingkat risiko otomatisasi pada berbagai jenis pekerjaan secara efektif. Berdasarkan hasil pengujian, kedua model prediksi yang digunakan dalam penelitian, yaitu *Linear Regression* dan XGBoost, berhasil menghasilkan nilai prediksi yang mendekati nilai aktual[17]. Hasil evaluasi menggunakan metrik *Root Mean Square Error* (RMSE) dan koefisien determinasi ($R^2$) menunjukkan bahwa model XGBoost memiliki kinerja yang lebih baik dibandingkan model *Linear Regression*. Perbedaan performa antara kedua model dipengaruhi oleh karakteristik data pekerjaan yang memiliki hubungan kompleks antara tingkat pendidikan, jenis pekerjaan, dan probabilitas otomatisasi. *Linear Regression* mengasumsikan hubungan linier antar variabel, sehingga model ini cenderung menghasilkan prediksi yang stabil namun kurang fleksibel dalam menangkap variasi pola data yang bersifat nonlinear [18]. Visualisasi hasil menunjukka n pola prediksi yang membentuk klister tertentu, yang mencerminkan keterbatasan model linier dalam merepresentasikan kompleksitas data dunia nyata. Sebaliknya, model XGBoost memiliki kemampuan yang lebih unggul dalam menangkap hubungan non linear dan interaksi antar fitur melalui pendekatan *ensemble learning* berbasis *boosting*. Kemampuan XGBoost dalam mengombinasikan banyak pohon keputusan memungkinkan model ini untuk mengurangi kesalahan prediksi secara bertahap [16], Kondisi ini menghasilkan nilai RMSE yang lebih rendah serta nilai $R^2$ yang lebih tinggi, selaras dengan hasil penelitian yang dilaporkan oleh Chen dan Guestrin [3] yang beranggapan bahwa XGBoost unggul dalam menangani dataset dengan pola kompleks dan heterogen.

Pemilihan metode *supervised learning* dalam penelitian ini didasarkan pada ketersediaan data berlabel berupa probabilitas otomatisasi pekerjaan. Pendekatan ini berbeda dengan metode klasifikasi berbasis aturan atau analisis statistik sederhana yang cenderung kurang adaptif terhadap perubahan pola data. Dengan memanfaatkan *supervised learning*, sistem mampu mempelajari hubungan antara variabel pendidikan dan pekerjaan secara langsung dari data, sehingga menghasilkan prediksi yang lebih representatif [16].

Jika dibandingkan dengan model linier, penggunaan metode *ensemble* seperti XGBoost memberikan keunggulan dari sisi akurasi prediksi [18]. Namun demikian, *Linear Regression* tetap memiliki kelebihan dalam hal kemudahan interpretasi, sehingga masih relevan digunakan sebagai model *baseline*. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pendekatan *supervised learning*, khususnya menggunakan XGBoost, dapat dijadikan solusi yang efektif dalam memprediksi risiko otomatisasi pekerjaan di masa depan dan dapat dimanfaatkan sebagai dasar pengambilan keputusan dalam perencanaan pengembangan keterampilan tenaga kerja.

## 4. Kesimpulan

Penelitian ini bertujuan untuk memprediksi probabilitas otomatisasi pekerjaan pada tahun 2030 menggunakan pendekatan *supervised learning* dengan membandingkan algoritma *Linear Regression* dan XGBoost *Regressor*. Berdasarkan hasil eksperimen yang telah dilakukan, kedua model mampu memprediksi probabilitas otomatisasi pekerjaan dengan tingkat akurasi yang cukup baik. Namun demikian, hasil evaluasi menunjukkan bahwa model XGBoost memiliki kinerja yang lebih unggul dibandingkan *Linear Regression*, yang ditunjukkan oleh nilai *Root Mean Square Error* (RMSE) yang lebih rendah serta nilai koefisien determinasi ($R^2$) yang lebih tinggi.

Keunggulan XGBoost dalam penelitian ini dipengaruhi oleh kemampuannya dalam menangkap hubungan non-linear dan interaksi antar fitur, khususnya pada data pekerjaan yang memiliki karakteristik kompleks antara jenis pekerjaan dan tingkat pendidikan. Sementara itu, *Linear Regression* tetap menunjukkan performa yang stabil dan mudah diinterpretasikan, sehingga masih relevan digunakan sebagai model *baseline* dalam analisis prediksi risiko otomatisasi pekerjaan.

Hasil penelitian ini mengindikasikan bahwa tingkat pendidikan dan jenis pekerjaan memiliki pengaruh terhadap probabilitas otomatisasi di masa depan. Pekerjaan yang bersifat rutin dan membutuhkan tingkat pendidikan yang relatif rendah cenderung memiliki risiko otomatisasi yang lebih tinggi, sedangkan pekerjaan yang menuntut keterampilan analitis dan tingkat pendidikan yang lebih tinggi menunjukkan risiko otomatisasi yang lebih rendah. Temuan ini memberikan gambaran awal mengenai dampak perkembangan kecerdasan buatan terhadap struktur pasar tenaga kerja.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan berasal dari satu sumber data publik, sehingga karakteristik dan asumsi yang terkandung di dalamnya dapat memengaruhi hasil prediksi. Selain itu, variabel yang digunakan masih terbatas pada jenis pekerjaan dan tingkat pendidikan, sehingga belum sepenuhnya merepresentasikan faktor lain seperti kreativitas, kemampuan sosial, dan adaptabilitas tenaga kerja. Penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih beragam,

menambahkan variabel pendukung, serta menerapkan metode pembelajaran mesin lain untuk meningkatkan akurasi dan generalisasi model prediksi.

**Ucapan Terima Kasih**

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis menyampaikan terima kasih kepada Program Studi Teknik Informatika, Universitas Esa Unggul, atas dukungan fasilitas dan lingkungan akademik yang diberikan selama proses penelitian. Ucapan terima kasih juga disampaikan kepada Ibu Vitri Tundjungsari atas bimbingan, arahan, dan masukan yang sangat berharga dalam penyusunan naskah penelitian ini. Selain itu, penulis mengapresiasi penyedia dataset pada platform Kaggle yang telah menyediakan data publik sehingga penelitian ini dapat terlaksana.

## Referensi

[1] Dr. P. Ros and Dr. B. Loeung, "ANALYZING THE IMPACT OF ARTIFICIAL INTELLIGENCE ON JOB ROLES, SKILLS REQUIREMENTS, AND EMPLOYMENT PATTERNS," *Journal of Social Sciences and Humanities*, pp. 72–90, Mar. 2025, doi: 10.56943/jssh.v4i1.699.

[2] K. A'yuni and Fahrurrozi, "The Impact of Artificial Intelligence (AI) Implementation on the Labor Market from the Perspective of Maqashid Syariah," *Journal of Business Management and Islamic Banking*, pp. 181–198, Dec. 2025, doi: 10.14421/jbmib.2025.0402- 05.

[3] Muh. Arfah, S. Suherlan, and S. A. Pramono, "Eksplorasi Transformasi Digital dalam MSDM: Dampak Integrasi Artificial Intelligence dan Big Data Analytics terhadap Pengambilan Keputusan Strategis," *Jurnal Minfo Polgan*, vol. 14, no. 1, pp. 183– 192, Mar. 2025, doi: 10.33395/jmp.v14i1.14673.

[4] "The Future of Jobs Report 2020 O C T O B E R 2 0 2 0," 2020.

[5] C. Khalaf, G. Michaud, and G. J. Jolley, "Predicting declining and growing occupations using supervised machine learning," *J. Comput. Soc. Sci.*, vol. 6, no. 2, pp. 757–780, Oct. 2023, doi: 10.1007/s42001-023-00211-0.

[6] Y. Shen and X. Zhang, "The impact of artificial intelligence on employment: the role of virtual agglomeration," *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, p. 122, Jan. 2024, doi: 10.1057/s41599-024-02647-9.

[7] A. Ramadian, D. Z. Nurlinda, and N. Ramadhina, "Digital transformation in HR Planning: Adaptation strategies and challenges in the industrial revolution 5.0 era," *Jurnal Akuntansi dan Manajemen*, vol. 22, no. 1, pp. 39–52, Apr. 2025, doi: 10.36406/jam.v22i1.137.

[8] V. Tundjungsari and S. Wijaya, "SENTIMEN ANALISIS PADA APLIKASI E-BRANCH BCA MENGGUNAKAN METODE NAÏVE BAYES", doi: 10.37817/ikraith-teknologi.v8i1.

[9] J. Nartey, "AI Job Displacement Analysis (2025-2030)," 2025. doi: 10.2139/ssrn.5316265.

[10] A. Y. Timur, "SISTEM REKOMENDASI LAGU INDONESIA MENGGUNAKAN METODE CONTENT-BASED FILTERING DAN COSINE SIMILARITY," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5949.

[11] Riska Rismaya, Dwi Yuniarto, and David Setiadi, "Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa," *Router : Jurnal Teknik Informatika dan Terapan*, vol. 3, no. 1, pp. 15–23, Feb. 2025, doi: 10.62951/router.v3i1.389.

[12] H. Zhang and W. Zhang, "Advancing enterprise risk management with deep learning: A predictive approach using the XGBoost- CNN-BiLSTM model," *PLoS One*, vol. 20, no. 4, p. e0319773, Apr. 2025, doi: 10.1371/journal.pone.0319773.

[13] S. Hakkal and A. A. Lahcen, "XGBoost To Enhance Learner Performance Prediction," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100254, Dec. 2024, doi: 10.1016/j.caeai.2024.100254.

[14] A. Sulami, V. Atina, and N. Nurmalitasari, "Penerapan Metode Content Based Filtering dalam Sistem Rekomendasi Pemilihan Produk Skincare," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 9, no. 2, p. 172, Dec. 2024, doi: 10.30998/string.v9i2.24066.

[15] E. Ayuningrum, Y. Azhar, and G. I. Marthasari, "Sistem Rekomendasi Produk Skincare Korea Berbasis Web Menggunakan Metode Collaborative Filtering," *Jurnal Repositor*, vol. 4, no. 4, Feb. 2024, doi: 10.22219/repositor.v4i4.32284.

[16] Y. Lin *et al.*, "An interpretable XGBoost model for risk prediction of progression from sepsis-associated acute kidney injury to chronic kidney disease," *Inform. Med. Unlocked*, vol. 58, p. 101685, 2025, doi: 10.1016/j.imu.2025.101685.

[17] S. S. Hema Harsha, "Navigating the Future of Work: The Impact of Artificial Intelligence on Jobs, Skills, and Workforce Dynamics," *International Journal of Business and Management Invention*, vol. 14, no. 5, pp. 67–79, May 2025, doi: 10.35629/8028-14056779.

[18] I. M. Siregar, D. Pratama, and C. Himawan, "Penggunaan Jaccard Similarity Coefficient dalam Optimasi Proses Rekrutmen Karyawan Berbasis Profil dan Kompetensi," *SINTECH (Science and Information Technology) Journal*, vol. 7, no. 2, pp. 101–111, Aug. 2024, doi: 10.31598/sintechjournal.v7i2.1617.