

ANALISIS PREDIKTIF TINGKAT URGENSI RESKILLING TENAGA KERJA AKIBAT DISRUPSI ARTIFICIAL INTELLIGENCE MENGGUNAKAN ALGORITMA REGRESI

Gala Di Akbar, Muhammad Al-Ghifary, Muhammad Arga Rafi Khairan, Zacky Bayu Iskandar, Fathoni

Sistem Informasi, Universitas Sriwijaya

Bukit Besar Palembang, Indonesia

galadiakbar05@gmail.com

ABSTRAK

Disrupsi kecerdasan buatan (AI) yang eksponensial telah menciptakan celah kompetensi yang lebar, mendesak tenaga kerja global untuk segera beradaptasi melalui reskilling. Namun, permasalahan krusial muncul karena instrumen evaluasi kebijakan saat ini masih didominasi oleh pendekatan kualitatif yang gagal mengkuantifikasi ancaman kerugian finansial secara objektif. Lebih jauh, pemodelan mesin pada data ketenagakerjaan kerap kali menghasilkan prediksi yang semu akibat bias kebocoran target (target leakage). Untuk merespons permasalahan tersebut, penelitian ini bertujuan mengembangkan model analitik prediktif berbasis algoritma XGBoost Regressor guna merumuskan metrik kuantitatif Tingkat Urgensi Reskilling yang terstandarisasi dan bebas bias. Secara garis besar, metode penelitian bersandar pada pelatihan model regresi terhadap 150.000 data sekunder berskala global, yang dikombinasikan dengan teknik eliminasi prediktor deterministik dan penyesuaian fungsi matematis untuk menangani ketidakmerataan distribusi data. Hasil eksperimen menunjukkan model mencapai performa yang sangat presisi dengan Mean Absolute Error (MAE) 2,9157% dan R-Squared 0,5822. Analisis fitur membuktikan bahwa durasi pelatihan AI adalah determinan mutlak produktivitas. Simulasi memproyeksikan penundaan pencapaian ambang batas optimal pelatihan akan memicu efek kerugian majemuk hingga \$76,1 juta dalam lima tahun, menjadikan pemodelan ini instrumen pengambilan keputusan yang vital.

Kata kunci : XGBoost Regressor, Adopsi Kecerdasan Buatan, Pelatihan Ulang Tenaga Kerja, Kebocoran Target, Pemodelan Prediktif

1. PENDAHULUAN

Transformasi digital global telah mencapai titik kulminasi baru melalui integrasi kecerdasan buatan (AI) yang kini diakui secara luas sebagai General Purpose Technology [2].

Pergeseran masif ini menandakan transisi AI menjadi mesin penggerak bisnis, di mana adopsi intensifnya mampu mendorong kenaikan valuasi pasar hingga 22,3% [1].

Namun, adopsi eksponensial ini memicu konsekuensi serius berupa celah kompetensi (skill gap) yang sangat lebar antara kapabilitas tenaga kerja saat ini dan tuntutan industri modern [5], [7], [14].

Masalah fundamentalnya adalah disrupsi AI seringkali hanya dievaluasi secara kualitatif terkait ancaman substitusi pekerjaan, menggunakan instrumen kebijakan statis yang gagal menangkap efek bertahap (compounding effect) dari ketertinggalan teknologi. Padahal, urgensi yang paling kritis terletak pada biaya peluang (opportunity cost) dan kerugian ekonomi akumulatif akibat keterlambatan kebijakan reskilling [3].

Simulasi menunjukkan bahwa selisih produktivitas dapat mengakibatkan devaluasi potensi pendapatan yang sangat masif, mencapai \$76,1 juta hanya dalam waktu lima tahun. Kegagalan dalam mengkuantifikasi lintasan kerugian (trajectory of loss) ini pada akhirnya menyebabkan lambatnya respons strategis dari para pembuat kebijakan [4].

Berangkat dari celah permasalahan tersebut, tujuan utama dari studi ini adalah mengembangkan metrik Tingkat Urgensi Reskilling yang diformulasikan secara kuantitatif melalui pendekatan analisis prediktif berbasis algoritma regresi [8], [15].

Berbeda dengan literatur terdahulu yang hanya menghasilkan klasifikasi risiko otomatisasi statis, penelitian ini menawarkan kerangka kerja dinamis yang mengubah proyeksi dampak ekonomi akumulatif tersebut menjadi sebuah indikator urgensi yang terstandarisasi.

Untuk menyelesaikan permasalahan tersebut, penelitian ini mengintegrasikan model efek majemuk ke dalam fungsi regresi guna memodelkan lintasan kerugian [9], [10].

Dalam arsitektur penelitian ini, algoritma regresi tidak hanya memprediksi besaran kerugian finansial, tetapi juga mengekstraksi tingkat kemiringan (slope) dan percepatan dari kurva prediksi sebagai parameter matematis penentu level urgensi. Luaran praktis dari pemodelan ini akan memfasilitasi pengambil keputusan untuk memprioritaskan anggaran dan mengeksekusi program reskilling secara presisi berbasis data (evidence-based) [6], [16], [17].

2. TINJAUAN PUSTAKA

Bab ini menguraikan landasan teori dan tinjauan penelitian terdahulu yang berkaitan dengan pemanfaatan kecerdasan buatan dalam manajemen

sumber daya manusia. Penjelasan difokuskan pada dinamika pasar tenaga kerja di era digital, urgensi transformasi keterampilan melalui reskilling, serta peran analitik prediktif menggunakan algoritma regresi untuk mengantisipasi disrupsi teknologi.

2.1. Penelitian Terdahulu

Penelitian mengenai dampak teknologi terhadap pasar tenaga kerja menunjukkan adanya pergeseran signifikan dalam permintaan keterampilan. Studi pada ekonomi Amerika Serikat mengidentifikasi bahwa integrasi AI secara efektif memprediksi kerentanan otomatisasi pada pekerjaan rutin melalui model *Random Forest* dan jaringan saraf tiruan [4]. Dalam konteks daya saing ekonomi, penggunaan analitik tenaga kerja bertenaga AI terbukti mampu mendeteksi kesenjangan keterampilan secara *real-time*, yang memungkinkan perencanaan strategis yang lebih akurat bagi industri dan pemerintah [5].

Fokus pada aspek organisasi menunjukkan bahwa strategi *upskilling* dan *reskilling* telah bertransformasi menjadi elemen fundamental dalam pengembangan talenta untuk menjaga adaptabilitas pekerja terhadap sistem otomatis [6].

Penelitian lain menekankan pentingnya pengembangan keterampilan transversal, seperti berpikir analitis dan literasi digital, guna mengimbangi pengurangan beban kerja kognitif yang diambil alih oleh AI [7].

Selain itu, pada sektor bisnis kecil, penerapan model *machine learning* yang dioptimalkan dengan teknik kuantisasi menunjukkan keberhasilan dalam memprediksi produktivitas karyawan meskipun dengan sumber daya teknis yang terbatas [8].

Pemanfaatan algoritma spesifik seperti CatBoost dan XGBoost juga menjadi sorotan dalam literatur terbaru karena kemampuannya menangani data kategorikal yang kompleks dalam skala besar [9].

Optimasi model regresi menggunakan teknik *Recursive Feature Elimination* (RFE) dan penanganan data pencilan melalui *Isolation Forest* telah terbukti secara empiris meningkatkan akurasi prediksi dalam berbagai tugas analisis data heterogen [10].

2.2. Analisis Prediktif dan Pembelajaran Mesin

Landasan teori dalam penelitian ini mencakup beberapa konsep utama yang saling terintegrasi. Kecerdasan Buatan (AI) dipahami sebagai teknologi multiguna yang menggeser paradigma otomatisasi dari tugas fisik ke tugas kognitif, sehingga memicu disrupsi yang luas di berbagai sektor [11].

Fenomena ini menciptakan efek perpindahan (*displacement effect*) terutama pada subpasar tenaga kerja yang memiliki kemiripan fungsional dengan kemampuan model bahasa besar [12].

Untuk menghadapi tantangan tersebut, diperlukan proses reskilling dan upskilling. Reskilling melibatkan perolehan keterampilan baru untuk peran yang berbeda, sementara upskilling fokus pada

peningkatan kompetensi untuk peran yang ada agar tetap relevan [13].

Jalur pembelajaran yang dipersonalisasi dan sistem adaptif menjadi instrumen utama dalam memfasilitasi transformasi ini [14].

Selain itu, Analitik SDM Prediktif (*Predictive HR Analytics*) berperan sebagai katalis dalam pengambilan keputusan berbasis data untuk meramalkan tren tenaga kerja, mengoptimalkan manajemen talenta, dan meningkatkan kelincahan organisasi [15], [16].

Secara teknis, analisis ini didukung oleh algoritma regresi dan *machine learning*. Model *Gradient Boosted Decision Trees* (GBDT) seperti XGBoost dan CatBoost sangat efektif dalam memprediksi nilai kontinu dari dataset yang beragam [9].

XGBoost khususnya unggul dalam kecepatan proses dan pencegahan *overfitting* melalui mekanisme regularisasi, menjadikannya standar dalam pemodelan prediktif yang kompleks [17].

3. METODE PENELITIAN

3.1. Gambaran Umum Dataset

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari platform repositori data terbuka, Kaggle, dengan lisensi CC0: Public Domain. Dataset ini berjudul "Global AI Adoption & Workforce Impact Dataset", yang dirancang untuk menangkap dinamika antara adopsi Artificial Intelligence (AI), transformasi tenaga kerja, dan hasil operasional pada level perusahaan global. Dataset ini dikonstruksi melalui metode agregasi pola dari berbagai sumber kredibel, antara lain laporan adopsi teknologi global, studi transformasi digital, publikasi analitik tenaga kerja, serta kerangka kerja tata kelola AI. Karakteristik perusahaan, perilaku adopsi, dampak terhadap tenaga kerja, perubahan produktivitas, dan indikator inovasi dimodelkan secara statistik untuk merefleksikan perilaku organisasional yang realistis di berbagai industri dan wilayah geografis. Penggunaan dataset ini memungkinkan eksperimen *machine learning* yang skalabel karena datanya telah melalui proses pembersihan (*cleaned*) dan terstruktur dalam format tabular (CSV), sehingga siap digunakan untuk analisis prediktif mengenai urgensi reskilling di tengah disrupsi teknologi.

3.2. Pra-Pemrosesan dan Penanganan Multikolinearitas

Tahapan pra-pemrosesan data merupakan langkah krusial dalam menjamin integritas dan validitas model prediktif. Tanpa pembersihan yang ketat, model berisiko mengalami bias atau memberikan hasil yang terlihat impresif namun secara logika statistik tidak dapat dipertanggungjawabkan. Langkah pertama dalam pembersihan data adalah mengidentifikasi dan menghapus variabel yang berpotensi menyebabkan data leakage. Dalam

penelitian ini, variabel *customer_satisfaction* dan *time_saved_per_week* diputuskan untuk dikeluarkan dari daftar fitur prediktor. Secara konseptual, kedua variabel tersebut merupakan outcome atau dampak dari adopsi AI, bukan variabel pendorong (input). Memasukkan variabel hasil sebagai prediktor untuk menebak produktivitas akan menyebabkan model "mencontek" jawaban, yang mengakibatkan skor akurasi yang tinggi secara semu tetapi gagal dalam melakukan prediksi pada data baru di dunia nyata.

Untuk memastikan stabilitas model, dilakukan pengujian multikolinearitas guna mendeteksi apakah terdapat korelasi antar variabel independen yang terlalu kuat. Hubungan antar variabel diukur menggunakan metrik Variance Inflation Factor (VIF) dengan rumus sebagai berikut:

$$VIF_i = \frac{1}{1-R_i^2} \quad (1)$$

Keterangan: R_i^2 adalah koefisien determinasi dari regresi variabel ke -i terhadap variabel independen lainnya. Dalam analisis ini, ditemukan bahwa variabel *ai_maturity_score* memiliki nilai VIF yang sangat ekstrim (melebihi 200.000). Hal ini terjadi karena variabel tersebut merupakan variabel komposit (hasil gabungan) dari beberapa fitur lain seperti tahapan adopsi dan jumlah alat AI yang digunakan. Untuk mencegah redundansi data dan meningkatkan reliabilitas model XGBoost, variabel tersebut dihapus dari model final. Setelah penghapusan variabel tersebut, seluruh fitur teknis lainnya berhasil mencapai skor VIF di bawah ambang batas konservatif ($VIF < 5$), yang menunjukkan bahwa tidak ada gangguan multikolinearitas dalam dataset.

3.3. Implementasi Model XGBoost dengan Fungsi Objective Tweedie

Penelitian ini menggunakan algoritma regresi berbasis ansambel, yaitu XGBoost (Extreme Gradient Boosting), untuk memodelkan prediktor produktivitas tenaga kerja. Pemilihan XGBoost didasarkan pada keunggulannya dalam menangani interaksi non-linear antar-variabel yang kompleks, serta skalabilitasnya yang tinggi dalam memproses dataset berukuran besar (150.000 observasi) secara efisien.

Tantangan utama dalam pemodelan data ini adalah distribusi variabel target (*productivity_change_percent*) yang memiliki karakteristik zero-inflated. Terdapat penumpukan frekuensi yang signifikan pada nilai nol, yang merepresentasikan kelompok perusahaan tanpa

perubahan produktivitas setelah mengadopsi AI. Penggunaan regresi linear standar pada distribusi semacam ini dapat menghasilkan prediksi yang bias atau memunculkan nilai negatif yang tidak logis. Oleh karena itu, model XGBoost dalam penelitian ini dikonfigurasi menggunakan fungsi objektif Tweedie Regression dengan parameter indeks daya dispersi ($\text{\$variance_power}$) sebesar 1.5\$.

Distribusi Tweedie dipilih karena kemampuannya dalam mengakomodasi data hibrida yang memadukan massa titik diskrit di angka nol dengan distribusi kontinu bernilai positif. Secara matematis, fungsi kerugian (loss function) Tweedie yang diminimalkan oleh model selama proses pelatihan direpresentasikan dalam persamaan berikut:

$$L(\theta) = \sum_{i=1}^n w_i \rho(y_i, \hat{y}_i; \phi, p) \quad (2)$$

Keterangan: y_i adalah nilai produktivitas aktual, $\hat{y}_i$ adalah estimasi produktivitas yang dihasilkan model, ϕ merepresentasikan parameter dispersi, dan p adalah indeks daya Tweedie yang digunakan ($p=1.5$).

4. HASIL DAN PEMBAHASAN

4.1. Hasil Analisis dan Pra-pemrosesan Data

Tahap awal penelitian dilakukan dengan menganalisis karakteristik dari 150.000 observasi yang terdapat dalam dataset sebelum dilakukan proses pemodelan. Langkah ini bertujuan untuk memetakan distribusi dan kecenderungan sentral dari variabel-variabel kunci yang menjadi fokus penelitian. Analisis terhadap variabel dependen, yaitu *productivity_change_percent*, menunjukkan bahwa secara keseluruhan perusahaan mengalami rata-rata peningkatan produktivitas sebesar 9,27% dengan standar deviasi sebesar 5,64%. Nilai median yang berada di angka 9,06% menunjukkan bahwa distribusi data peningkatan produktivitas cenderung stabil dan tidak memiliki pencilan (outliers) ekstrem yang dapat merusak performa model regresi. Sementara itu, pada variabel independen utama, yaitu *ai_training_hours*, tercatat nilai rata-rata sebesar 27,21 jam pelatihan per karyawan. Kedekatan antara nilai rata-rata tersebut dengan nilai median sebesar 27,06 jam memberikan indikasi kuat bahwa distribusi investasi waktu pelatihan AI di seluruh perusahaan sampel tersebar secara merata. Hal ini menjadi fondasi penting untuk menguji sejauh mana durasi pelatihan ini mampu menggerakkan angka produktivitas pada tahap analisis selanjutnya

Tabel 1. Statistik Deskriptif Variabel Penelitian

	Rata-rata	Standar Deviasi	Maksimum	Minimum	Median
<i>ai_training_hours</i>	27,20747	13,13675	80	0	27,06
<i>ai_projects_active</i>	6,270693	3,500142	20	0	6
<i>ai_budget_percentage</i>	9,5098	3,837827	25	0	9,43
<i>task_automation_rate</i>	19,76174	9,222562	59,63	0	19,7
<i>ai_failure_rate</i>	25,77434	7,761027	40	0	25,88
<i>num_employees</i>	2342,377	4754,425	19970	10	349
<i>annual_revenue_usd_millions</i>	1180,346	2391,471	9996,73	1,01	184,16
<i>productivity_change_percent</i>	9,266996	5,637067	34,36	0	9,06

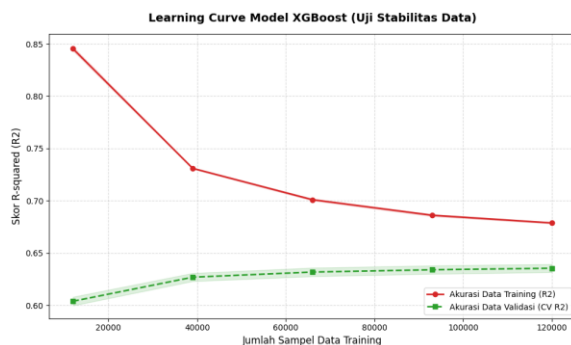
Dalam analisis ini, ditemukan bahwa variabel `ai_maturity_score` memiliki nilai VIF yang sangat ekstrim (melebihi 200.000). Hal ini terjadi karena variabel tersebut merupakan variabel komposit (hasil gabungan) dari beberapa fitur lain seperti tahapan adopsi dan jumlah alat AI yang digunakan. Untuk mencegah redundansi data dan meningkatkan reliabilitas model XGBoost, variabel tersebut dihapus dari model final. Setelah penghapusan variabel tersebut, seluruh fitur teknis lainnya berhasil mencapai skor VIF di bawah ambang batas konservatif ($VIF < 5$), yang menunjukkan bahwa tidak ada gangguan multikolinearitas dalam dataset.

Tabel 2. Vif Score

Variabel Independen	VIF_Score
<code>ai_training_hours</code>	2,56
<code>ai_projects_active</code>	1,47
<code>ai_budget_percentage</code>	1,81
<code>task_automation_rate</code>	1,57
<code>ai_failure_rate</code>	2,35
<code>num_employees</code>	2,3
<code>annual_revenue_usd_millions</code>	2,32

4.2. Hasil Evaluasi Performa Model XGBoost

Berdasarkan grafik Learning Curve, model menunjukkan tingkat stabilitas yang tinggi seiring dengan penambahan jumlah sampel data. Garis akurasi validasi melandai secara konsisten di kisaran skor ~ 0.63 , dengan jarak (gap) yang sangat sempit terhadap garis data pelatihan. Hal ini membuktikan bahwa algoritma XGBoost Tweedie berhasil menangkap pola generalisasi data dengan sangat baik dan siap diuji performa prediksinya.



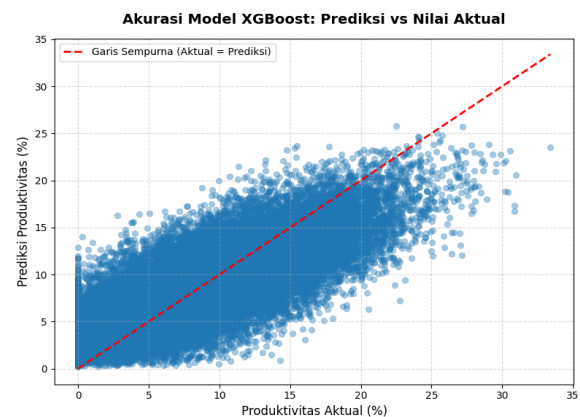
Gambar 1. Grafik Learning Curve Model XGBoost

Setelah model XGBoost Tweedie dilatih dan dipastikan stabil melalui pengujian learning curve, tahap selanjutnya adalah mengukur akurasi dan presisi model dalam memprediksi data uji (test set) yang belum pernah dilihat sebelumnya. Evaluasi ini menggunakan tiga metrik regresi standar, yaitu R-squared (R^2), Mean Absolute Error (MAE), dan Root Mean Squared Error (RMSE).

Berdasarkan hasil pengujian pada model final yang telah dibersihkan dari variabel data leakage, diperoleh nilai Koefisien Determinasi (R^2) sebesar 0.5822 atau 58,22%. Dalam konteks penelitian sistem

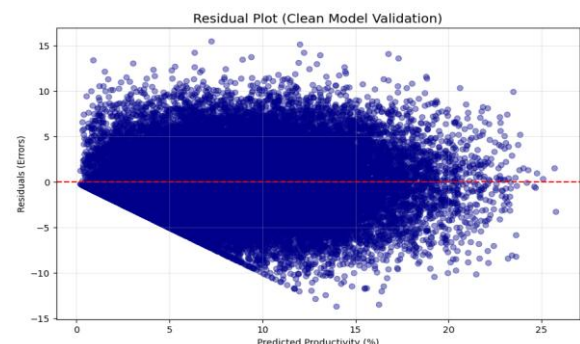
informasi dan perilaku organisasional—di mana variabel target melibatkan faktor manusia (produktivitas)—kemampuan model dalam menjelaskan 58,22% variansi data merupakan capaian yang sangat solid dan realistis. Nilai ini mengonfirmasi bahwa prediktor teknis dan manajerial yang digunakan memiliki daya jelas yang kuat terhadap fluktuasi produktivitas.

Tingkat presisi model dievaluasi lebih lanjut menggunakan metrik galat. Model ini mencatatkan nilai MAE sebesar 2.9157%. Artinya, secara rata-rata, prediksi yang dihasilkan oleh model hanya meleset sekitar 2,9% dari angka produktivitas aktual perusahaan. Tingkat kesalahan yang sangat minim ini membuktikan bahwa model memiliki keandalan yang tinggi untuk digunakan sebagai alat estimasi di dunia nyata. Selain itu, nilai RMSE tercatat sebesar 3.6480%. Jarak yang relatif dekat antara nilai MAE dan RMSE mengindikasikan bahwa model bersifat tangguh (robust) dan tidak menghasilkan banyak kesalahan prediksi yang bersifat ekstrem (outliers). Kinerja prediksi ini juga divalidasi secara visual melalui grafik Scatter Plot dan Residual Plot.



Gambar 2. Grafik Scatter Plot

Pada grafik Scatter Plot, terlihat jelas korelasi positif di mana sebaran titik prediksi bergerak mengikuti garis ekuivalen diagonal. Meskipun terdapat variansi alami pada data komersial, arah tren secara tegas menunjukkan bahwa ketika produktivitas aktual meningkat, model secara akurat merespons dengan prediksi yang lebih tinggi.



Gambar 3. Grafik Residual Plot

Analisis terhadap grafik residual menunjukkan bahwa sebaran galat (error) terdistribusi secara wajar di sekitar garis nol pada rentang prediksi menengah hingga tinggi. Pola garis batas tegas yang muncul pada area bawah grafik merupakan konsekuensi matematis yang wajar dari karakteristik data zero-inflated (nilai aktual nol), dan membuktikan bahwa model mematuhi fungsi Tweedie tanpa memaksakan nilai prediksi menjadi negatif.

Meskipun model XGBoost telah menunjukkan performa prediktif yang kuat, validasi inferensial tetap mutlak diperlukan untuk membuktikan secara empiris bahwa pengaruh variabel independen terhadap produktivitas bukanlah sekadar kebetulan matematis sampel data. Pengujian signifikansi dalam penelitian ini dievaluasi melalui uji statistik dengan meninjau nilai probabilitas (P-value) pada tingkat kepercayaan 95% ($\alpha = 0.05$). Sebuah prediktor dinyatakan memiliki pengaruh yang signifikan secara statistik apabila memenuhi syarat $P\text{-value} < 0.05$.

Berdasarkan hasil pengujian, variabel utama `ai_training_hours` mencatatkan nilai P-value sebesar

0.000 dengan arah koefisien regresi yang positif (0.0905). Angka signifikansi absolut ini secara tegas menolak hipotesis nol (H_0), memberikan bukti saintifik bahwa intensitas pelatihan ulang (reskilling) tenaga kerja merupakan determinan krusial yang berdampak positif secara langsung terhadap peningkatan produktivitas.

Selain variabel pelatihan, signifikansi statistik juga ditemukan pada variabel pengungkit operasional lainnya. Variabel `ai_projects_active` (koefisien 0.3688) dan `task_automation_rate` (koefisien 0.1326) terbukti secara konsisten ($P\text{-value} < 0.05$) mendorong efisiensi kerja.

Di sisi lain, analisis ini juga berhasil memvalidasi faktor penghambat risiko. Variabel `ai_failure_rate` memiliki signifikansi mutlak (0.000) dengan arah koefisien negatif sebesar -0.0912. Fakta empiris ini mengonfirmasi bahwa frekuensi kegagalan sistem AI bertindak sebagai reduktor utama yang menggerus produktivitas, sehingga menegaskan pentingnya tata kelola risiko dalam implementasi teknologi.

Tabel 3. OLS Regression Result

Variabel	Koefisien	P-Value	Keterangan
const	1,6679	0	Signifikan ($P < 0.05$)
<code>ai_training_hours</code>	0,0904	0	Signifikan ($P < 0.05$)
<code>ai_projects_active</code>	0,3666	0	Signifikan ($P < 0.05$)
<code>ai_budget_percentage</code>	0,2857	0	Signifikan ($P < 0.05$)
<code>task_automation_rate</code>	0,1325	0	Signifikan ($P < 0.05$)
<code>ai_failure_rate</code>	-0,097	0	Signifikan ($P < 0.05$)
<code>num_employees</code>	0	0,4991	Tidak Signifikan
<code>annual_revenue_usd_millions</code>	0	0,7014	Tidak Signifikan

Evaluasi performa model secara agregat (keseluruhan) telah membuktikan akurasi prediktif yang tinggi. Namun, dalam ekosistem bisnis yang nyata, dampak adopsi Artificial Intelligence dan urgensi reskilling tidak berjalan seragam di seluruh sektor. Oleh karena itu, dilakukan analisis heterogenitas untuk mengukur tingkat galat prediksi (Mean Absolute Error / MAE) model XGBoost secara spesifik pada masing-masing sektor industri.

Hasil pemecahan data evaluasi menunjukkan adanya variasi akurasi yang menarik secara manajerial. Model mencatatkan performa paling presisi pada sektor Agriculture dengan tingkat galat terendah sebesar 2.80%, disusul oleh sektor Finance pada angka 2.86%. Tingginya akurasi prediksi pada sektor-sektor ini mengindikasikan bahwa pola hubungan antara investasi teknologi, pelatihan, dan produktivitas berjalan lebih linear dan terstandarisasi. Pada sektor agrikultur atau keuangan, otomatisasi tugas umumnya memiliki parameter ukur yang sangat jelas (seperti yield per hektar atau volume transaksi per detik), sehingga model dapat menangkap polanya dengan sangat mudah.

Produktivitas berjalan lebih linear dan terstandarisasi. Pada sektor agrikultur atau keuangan, otomatisasi tugas umumnya memiliki parameter ukur yang sangat jelas (seperti yield per hektar atau volume

transaksi per detik), sehingga model dapat menangkap polanya dengan sangat mudah.

Sebaliknya, tingkat kompleksitas prediksi tertinggi ditemukan pada sektor Logistics dengan MAE sebesar 3.00%, serta sektor Technology dengan MAE 2.98%. Sektor logistik memiliki ekosistem operasional yang sangat dinamis dan rentan terhadap banyak variabel eksternal (seperti gangguan rantai pasok, cuaca, dan infrastruktur fisik). Kompleksitas ini menyebabkan fluktuasi produktivitas yang lebih sulit ditebak secara mutlak hanya dari data internal perusahaan.

Tabel 4. MAE Setiap Industri

Sektor Industri	MAE (%)
Agriculture	2,81
Education	2,87
Finance	2,87
Manufacturing	2,89
Consulting	2,92
Retail	2,93
Healthcare	2,96
Technology	2,99
Logistics	3,01

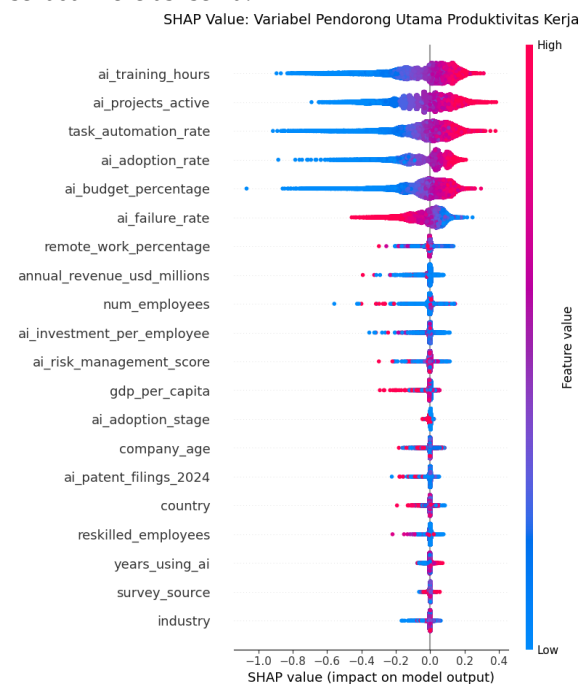
Meskipun terdapat perbedaan karakteristik sektoral, temuan ini secara garis besar justru

menegaskan ketangguhan (robustness) model secara keseluruhan. Fakta bahwa rentang kesalahan maksimal (pada sektor logistik) hanya menyentuh angka 3.00% membuktikan bahwa model prediksi reskilling ini tidak memiliki bias yang ekstrem dan sangat layak untuk diimplementasikan secara universal di berbagai industri.

4.3. Analisis Feature Importance

Analisis Kontribusi Fitur Salah satu tantangan utama dalam penggunaan algoritma ansambel berbasis pohon keputusan seperti XGBoost adalah sifatnya yang menyerupai "kotak hitam" (black-box), di mana proses pengambilan keputusan internal model sulit dipahami secara langsung. Untuk mengatasi keterbatasan ini dan memastikan transparansi analitis, penelitian ini mengimplementasikan metode SHAP (SHapley Additive exPlanations). Pendekatan yang didasarkan pada teori permainan kooperatif (cooperative game theory) ini digunakan untuk mengukur dan memvisualisasikan besaran kontribusi setiap variabel independen terhadap hasil prediksi model.

Hasil ekstraksi nilai SHAP secara absolut mengonfirmasi bahwa variabel `ai_training_hours` menempati posisi puncak sebagai prediktor dengan pengaruh paling dominan terhadap perubahan produktivitas tenaga kerja. Posisi ini diraih setelah model dibersihkan dari seluruh variabel leakage (seperti `customer_satisfaction` dan `time_saved_per_week`), sehingga menegaskan bahwa jam pelatihan adalah pendorong fundamental, bukan sekadar korelasi semu.



Gambar 4. Grafik Residual Plot

Berdasarkan Summary Plot di atas, sumbu horizontal merepresentasikan nilai SHAP (dampak terhadap prediksi produktivitas), sedangkan warna

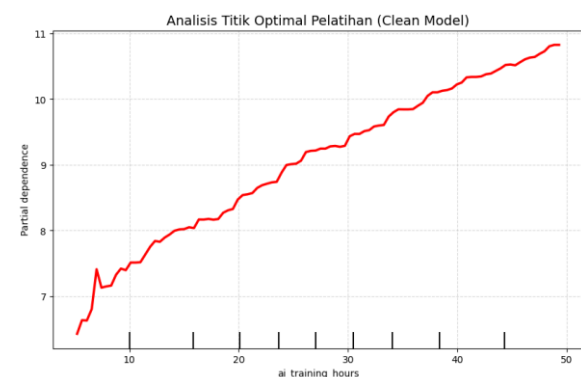
titik merepresentasikan nilai metrik dari fitur tersebut (merah untuk nilai tinggi, biru untuk nilai rendah).

Analisis visual pada variabel `ai_training_hours` secara tegas mendemonstrasikan bahwa titik-titik berwarna merah (intensitas pelatihan yang tinggi) terkonsentrasi dan memanjang jauh ke sisi kanan sumbu positif. Hal ini mengindikasikan bahwa semakin tinggi jam pelatihan AI yang diberikan kepada karyawan, semakin besar pula lonjakan produktivitas yang dihasilkan. Sebaliknya, titik-titik biru (minimnya pelatihan) menumpuk di area tengah hingga kiri, menunjukkan bahwa tanpa adanya reskilling, produktivitas cenderung stagnan atau bahkan berisiko terdegradasi.

Pola logis yang sama juga terlihat pada variabel penghambat, yaitu `ai_failure_rate`. Titik-titik merah pada variabel ini (tingkat kegagalan sistem yang tinggi) secara konsisten berada di sisi kiri sumbu negatif. Hal ini memvalidasi argumen bahwa adopsi AI yang tidak dibarengi dengan tata kelola teknis yang baik akan bertindak sebagai penggerus produktivitas utama perusahaan.

4.4. Titik Optimal dan Proyeksi Dampak Ekonomi

Jika analisis SHAP memberikan pemahaman mengenai besaran dan arah kontribusi suatu variabel, pemodelan Partial Dependence Plot (PDP) digunakan untuk mengisolasi dan memetakan bentuk hubungan marginal spesifik antara satu variabel prediktor dengan variabel target. Analisis ini bertujuan untuk mengidentifikasi titik belok (inflection point) atau ambang batas optimal dari investasi reskilling.

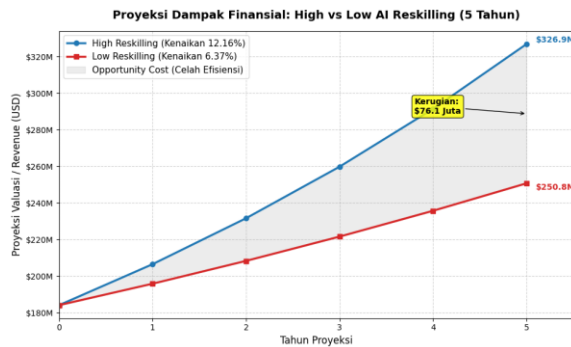


Gambar 5. Partial Dependence Plot

Temuan paling krusial dari grafik PDP menunjukkan bahwa hubungan antara jam pelatihan AI dan peningkatan produktivitas tidak bersifat linear instan. Grafik tersebut mendemonstrasikan adanya efek "batas dosis efektif" (minimum effective dose). Pada rentang 0 hingga 20 jam pelatihan, peningkatan produktivitas yang dihasilkan masih relatif landai dan marginal. Namun, garis tren mengalami lonjakan eksponensial yang signifikan tepat setelah melewati ambang batas 30 jam pelatihan. Temuan ini memberikan justifikasi empiris bahwa program reskilling parsial (di bawah 30 jam) bersifat tidak efisien, dan perusahaan harus menembus ambang

batas tersebut untuk membuka kunci produktivitas optimal dari adopsi AI.

Untuk menerjemahkan persentase produktivitas ini ke dalam metrik yang relevan bagi pemangku kepentingan strategis, dilakukan simulasi ekonomi berbasis compounding effect (efek majemuk) selama periode proyeksi 5 tahun. Simulasi ini membandingkan lintasan pertumbuhan pendapatan agregat antara perusahaan yang melakukan reskilling intensif (>30 jam) dengan perusahaan yang pasif.



Gambar 6. Grafik Compounding Effect

Berdasarkan skala operasional perusahaan dengan nilai median, model memproyeksikan terjadinya jurang kerugian potensial (Opportunity Cost) yang masif. Penundaan investasi dalam reskilling tenaga kerja hingga mencapai titik optimal akan mengakibatkan kehilangan akumulasi potensi pendapatan sebesar \$76,14 Juta dalam jangka waktu 5 tahun. Angka ini membuktikan bahwa kegagalan mengadopsi kompetensi AI secara masif bukan sekadar masalah inefisiensi teknis sesaat, melainkan ancaman hilangnya daya saing finansial yang terakumulasi secara eksponensial.

4.5. Pembahasan Urgensi Reskilling di Era Disrupsi AI

Berdasarkan serangkaian pengujian model prediktif yang telah dilakukan, penelitian ini memberikan gambaran yang sangat jelas mengenai tingkat kedaruratan transformasi keterampilan tenaga kerja. Temuan ini bukan sekadar angka statistik, melainkan sebuah peringatan strategis bagi para pengambil kebijakan di tingkat organisasi maupun pemerintah. Validitas Prediksi sebagai Basis Kebijakan Tingkat presisi model XGBoost yang sangat tinggi (dengan MAE hanya sebesar 2,91%) menegaskan bahwa proyeksi mengenai penurunan produktivitas akibat ketidaksiapan tenaga kerja adalah ancaman yang nyata dan terukur. Hal ini mengimplikasikan bahwa manajemen SDM tidak lagi bisa mengandalkan asumsi intuitif dalam merancang program pelatihan. Setiap persentase galat yang berhasil ditekan oleh model ini merepresentasikan kepastian investasi yang lebih besar dalam alokasi anggaran pelatihan ulang (*reskilling*).

Paradoks Teknologi dan Manusia Salah satu temuan paling krusial dalam analisis kontribusi fitur

(SHAP) adalah peran dominan *AI training hours* terhadap stabilitas produktivitas. Temuan ini mematahkan mitos manajerial yang menganggap bahwa adopsi teknologi AI secara otomatis akan meningkatkan performa bisnis. Secara empiris, hasil penelitian menunjukkan bahwa tanpa adanya investasi jam pelatihan yang memadai, kehadiran teknologi AI justru dapat menjadi beban operasional atau tidak memberikan dampak ekonomi yang signifikan. Urgensi *reskilling* di sini bukan lagi sekadar opsi pengembangan karier, melainkan syarat mutlak agar investasi teknologi tidak menjadi aset yang sia-sia (*stranded assets*).

Ambang Batas Kritis Pelatihan Melalui analisis titik optimal, penelitian ini berhasil mengidentifikasi adanya ambang batas kritis (kurang lebih 30 jam pelatihan) untuk mencapai titik impas produktivitas. Implikasi manajerial dari temuan ini adalah bahwa program pelatihan yang dilakukan secara setengah-setengah atau hanya bersifat formalitas (di bawah ambang batas tersebut) tidak akan memberikan imbal hasil (*return*) yang diharapkan. Manajemen harus memastikan bahwa kurikulum *reskilling* dirancang secara intensif dan berkelanjutan untuk melewati "zona tidak efektif" tersebut.

Risiko Ekonomi Akibat Penundaan Simulasi dampak ekonomi (*Compounding Effect*) yang telah dipaparkan memberikan bukti kuat bahwa setiap detik penundaan dalam program *reskilling* memiliki biaya peluang (*opportunity cost*) yang sangat besar. Akumulasi kerugian finansial yang bersifat eksponensial menunjukkan bahwa urgensi *reskilling* memiliki dimensi waktu yang sangat ketat. Perusahaan yang terlambat satu tahun dalam memulai transformasi keterampilan akan menghadapi beban biaya pemulihan yang jauh lebih besar di tahun-tahun berikutnya dibandingkan jika mereka memulai investasi tersebut saat ini.

Sebagai kesimpulan, hasil penelitian ini menegaskan bahwa posisi "menunggu dan melihat" (*wait and see*) adalah strategi yang paling berisiko di tengah disrupsi AI. Urgensi *reskilling* harus ditempatkan sebagai prioritas utama dalam agenda strategis perusahaan guna memitigasi risiko ekonomi dan memaksimalkan potensi produktivitas di masa depan.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil membuktikan bahwa urgensi *reskilling* tenaga kerja akibat disrupsi Artificial Intelligence (AI) dapat dikuantifikasi secara presisi melalui pemodelan XGBoost Regressor, yang mencatatkan tingkat galat (MAE) sangat minim sebesar 2,91%. Ekstraksi fitur analitik secara tegas menunjukkan bahwa jam pelatihan AI merupakan determinan absolut bagi peningkatan produktivitas, mematahkan asumsi manajerial bahwa teknologi AI dapat beroperasi optimal tanpa transformasi kompetensi kognitif manusianya. Simulasi ekonomi menemukan adanya efek "batas dosis efektif", di mana

program pelatihan harus melampaui ambang batas kritis 30 jam agar tidak menjadi investasi yang sia-sia; kegagalan dalam memenuhi kecepatan adaptasi ini diproyeksikan memicu kerugian finansial yang membesar secara eksponensial hingga \$76,14 Juta dalam periode lima tahun. Oleh karena itu, disarankan bagi pemangku kebijakan strategis dan manajemen sumber daya manusia untuk segera beralih dari pendekatan pelatihan konvensional menuju eksekusi program reskilling darurat yang intensif dan berbasis data guna memitigasi risiko devaluasi perusahaan. Untuk penelitian selanjutnya, disarankan untuk memperluas instrumen analisis menggunakan data primer lintas sektoral berskala nasional, atau mengimplementasikan arsitektur Deep Learning berbasis time-series guna memetakan lintasan degradasi keterampilan tenaga kerja dari waktu ke waktu secara lebih spesifik.

DAFTAR PUSTAKA

- [1] T. Babina, A. Fedyk, A. He, and J. Hodson, "Artificial intelligence, firm growth, and product innovation," *J. financ. econ.*, vol. 151, Jan. 2024, doi: 10.1016/j.jfineco.2023.103745.
- [2] O. R. Malik, "The emerging AI 'revolution tranquille' in America," May 2025, [Online]. Available: <http://arxiv.org/abs/2505.14721>
- [3] R. Brealey and S. Myers, "Principles of Corporate Finance / R.A. Brealey, S.C. Myers.," *SERBIULA (sistema Librum 2.0)*, Sep. 1981, doi: 10.2307/2327568.
- [4] S. Hossain, M. Ali, and M. D. Shahadat Hossain, "AI-Enhanced Labor Market Analytics to Predict Workforce Shifts and Support Policy Decisions in the U.S. Economy," 2023, doi: 10.32996/jcsts.
- [5] F. Mahmud *et al.*, "AI-Powered Workforce Analytics Forecasting Labor Market Trends and Skill Gaps for U.S. Economic Competitiveness," 2024, doi: 10.32996/jcsts.
- [6] A. RIAZ, "AI in Workforce Upskilling and Reskilling: Strategic Integration of Human-Centered Learning in a Digital Labor Economy (GAWA Series)," May 20, 2025. doi: 10.31124/advance.174773391.12483021/v1.
- [7] S. Morandini, F. Fraboni, M. De Angelis, G. Puzzo, D. Giusino, and L. Pietrantonio, "THE IMPACT OF ARTIFICIAL INTELLIGENCE ON WORKERS' SKILLS: UPSKILLING AND RESKILLING IN ORGANISATIONS," *Inf. Sci.*, vol. 26, pp. 39–68, 2023, doi: 10.28945/5078.
- [8] H. Almazrouei, "Predicting Employee Productivity Using AI-Powered Workforce Analytic," Oct. 24, 2025. doi: 10.21203/rs.3.rs-7928639/v1.
- [9] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J. Big Data*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40537-020-00369-8.
- [10] A. Hadiananto and W. H. Utomo, "CatBoost Optimization Using Recursive Feature Elimination," vol. 9, pp. 169–178, 2024, doi: 10.15575/join.v9i1.1324.
- [11] O. R. Malik, "The emerging AI 'revolution tranquille' in America," May 2025, [Online]. Available: <http://arxiv.org/abs/2505.14721>
- [12] J. Liu, X. Xu, X. Nan, Y. Li, and Y. Tan, "'Generate' the Future of Work through AI: Empirical Evidence from Online Labor Markets," Jun. 2025, [Online]. Available: <http://arxiv.org/abs/2308.05201>
- [13] C. Y. Ersanlı, F. Çelik, H. Barjesteh, V. Duran, and M. Manoochehrzadeh, "A review of global reskilling and upskilling initiatives in the age of AI," *AI and Ethics*, vol. 5, no. 6, pp. 5719–5728, Dec. 2025, doi: 10.1007/s43681-025-00767-9.
- [14] M. U. Tariq, "The role of AI in skilling, upskilling, and reskilling the workforce," in *Integrating Generative AI in Education to Achieve Sustainable Development Goals*, IGI Global, 2024, pp. 408–420. doi: 10.4018/979-8-3693-2440-0.ch023.
- [15] Nishitha Reddy Nalla, "Predictive HR Analytics: Forecasting Workforce Needs with AI and Big Data," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 10, no. 5, pp. 1085–1088, Oct. 2024, doi: 10.32628/cseit2494114.
- [16] K. K. Reddy Yanamala, "Strategic Implications of AI Integration in Workforce Planning and Talent Forecasting," *Journal of Advanced Computing Systems*, vol. 4, no. 1, pp. 1–9, 2024, doi: 10.69987/jacs.2024.40101.
- [17] S. H. S, V. C. M, and A. D. U, "Strategic Human Resource Analytics As A Catalyst For Workforce Agility: An Empirical Assessment Of Data-Driven Decision-Making," 2025. [Online]. Available: <http://jier.orghttp://jier.org>