Prediksi Promosi Pegawai Menggunakan Metode *Extremely Randomized Trees*

**Meilany Nonsi Tentua1, Saptaningsih Sumarmi2**

1Jurusan Teknik Informatika, Universitas PGRI Yogyakarta

2Jurusan Manajemen, Universitas PGRI Yogyakarta

e-mail: **1meilany@upy.ac.id**, 2sapta@upy.ac.id

**Intisari**

Seiring perkembangan teknologi informasi dan analisis data, peluang baru muncul dalam pengambilan keputusan di bidang sumber daya manusia. Penerapan metode *ensemble*, seperti *extremely randomized trees*, menjanjikan dalam meramalkan keputusan promosi dengan akurasi yang lebih tinggi.

Meskipun telah ada banyak penelitian di bidang ini, ada kebutuhan untuk lebih mendalam dalam mengadaptasi metode *ensemble* ke konteks promosi pegawai, mempertimbangkan berbagai faktor seperti keterampilan, pengalaman, dan pencapaian individu. Penelitian ini bertujuan mengembangkan model prediksi promosi pegawai yang efektif, menggabungkan metode extremely randomized trees dengan data sumber daya manusia yang relevan.

Model yang dihasilkan dari penelitian ini mencapai tingkat akurasi sebesar 99% pada data pelatihan, menunjukkan prediksi yang sangat tepat. Penggunaan data uji mengonfirmasi performa model dengan skor akurasi 98%, menggambarkan kemampuannya dalam memprediksi secara konsisten. Faktor-faktor seperti avg\_training\_score, department, total\_score, age, dan sum\_metric diidentifikasi sebagai paling berpengaruh dalam pembentukan model, memberikan pandangan penting bagi manajemen sumber daya manusia.

**Kata kunci**— *Extremely Randomized Trees*, Promosi Pegawai, *Machine Learning*

***Abstract***

*With the advancement of information technology and data analytics, new opportunities arise in decision-making within the field of human resources. The implementation of ensemble methods, such as extremely randomized trees, holds promise in predicting promotion decisions with higher accuracy. Despite numerous studies in this domain, there is a need for a deeper exploration in adapting ensemble methods to the context of employee promotions, considering various factors like skills, experience, and individual achievements. This research aims to develop an effective predictive model for employee promotions by integrating extremely randomized trees with relevant human resources data.*

*The model derived from this study achieved an accuracy rate of 99% on the training data, indicating highly precise predictions. The use of test data confirmed the model's performance with an accuracy score of 98%, illustrating its consistent predictive capabilities. Factors such as avg\_training\_score, department, total\_score, age, and sum\_metric were identified as the most influential in shaping the model, providing crucial insights for human resources management.*

***Keywords—*** *Extremely Randomized Trees, Employee Promotion, Machine Learning*

PENDAHULUAN

Dalam dunia bisnis yang kompetitif, pengelolaan sumber daya manusia menjadi krusial bagi perusahaan dalam mencapai keunggulan kompetitifnya [1]. Salah satu aspek penting dalam pengelolaan sumber daya manusia adalah kebijakan promosi pegawai [2]. Promosi pegawai tidak hanya memengaruhi kepuasan dan motivasi individu tetapi juga berdampak langsung pada kinerja organisasi secara keseluruhan [1].

Pada saat yang sama, perkembangan teknologi informasi dan analisis data telah membuka peluang baru dalam pengambilan keputusan di berbagai bidang [3], termasuk sumber daya manusia. Dengan memanfaatkan teknik-teknik analisis data yang canggih, perusahaan dapat membuat keputusan promosi yang lebih akurat dan berbasis bukti [4].

Salah satu pendekatan yang menjanjikan dalam meramalkan keputusan promosi pegawai adalah metode *extremely randomized trees*[[1]](#footnote-1) yang merupakan salah satu metode *ensamble*. Metode *ensemble* menggabungkan prediksi dari beberapa model atau algoritma yang berbeda untuk meningkatkan akurasi dan keandalan prediksi [5]. Penerapan metode *ensemble* dalam konteks prediksi promosi pegawai dapat membantu perusahaan mengidentifikasi pegawai yang berpotensi berkembang dan memberikan kontribusi signifikan pada tingkat lebih tinggi.

Namun, meskipun banyak penelitian telah dilakukan dalam domain ini [6], [7], masih ada kebutuhan untuk menggali lebih dalam dan mengadaptasi metode ensemble dengan lebih baik ke dalam konteks promosi pegawai. Selain itu, penelitian lebih lanjut dapat membahas berbagai faktor yang mempengaruhi keputusan promosi, termasuk keterampilan, pengalaman kerja, dan pencapaian individu.

Dengan menggabungkan konsep metode *extremely randomized trees* dengan data sumber daya manusia yang relevan, penelitian ini bertujuan untuk mengembangkan model prediksi promosi pegawai yang lebih efektif dan dapat diandalkan. Hasil dari penelitian ini diharapkan dapat memberikan panduan berharga bagi manajemen sumber daya manusia dalam membuat keputusan promosi yang lebih tepat, mengoptimalkan kekuatan kerja, dan secara keseluruhan meningkatkan kinerja organisasi.

METODE PENELITIAN

Pada bagian ini akan dijelaskan tentang langkah pengumpulan data, metode ensambles, pemrosesan data, pembagian dataset, pelatihan model ensambles, validasi model dan evaluasi kinerja model.

## Pengumpulan Data

Data yang digunakan pada penelitian merupakan data sekunder yang diambil pada *website kaggle* dengan judul “*Employees Evaluation for Promotion”*[[2]](#footnote-2)*.* Dataset terdiri dari 54.808 data latih dan 23.490 data uji dengan 13 fitur, yaitu: (1) employee\_id: ID pegawai; (2) department: Department dari pegawai; (3) region: Region dari pegawai; (4) education: tingkat pendidikan; (5) gender: jenis kelamin pegawai; (6) recruitment\_channel: tempat *recruitment* pegawai; (7) no\_ of\_ trainings: banyaknya pelatihan *soft skills* dan *technical skills*; (8) age: umur pegawai; (9) previous\_ year\_ rating: Rating pegawai tahun sebelumnya; (10) length\_ of\_ service: masa kerja; (11) awards\_ won: penghargaan yang diperoleh (0 atau 1); (12) avg\_ training\_ score: nilai rata-rata pelatihan; (13) is\_promoted: rekomendasi untuk promosi.

## Metode *Extremely Randomized Trees*

Metode yang digunakan pada penelitian adalah *extremely randomized trees*, merupakan metode *ensemble* yang menggabungkan beberapa pohon keputusan untuk membuat prediksi yang lebih akurat dan stabil. Rumus umum untuk prediksi *Random Forest* (klasifikasi)[8] dapat direpresentasikan pada persamaan (1).

$\hat{Y}\_{RF}=Model(\hat{Y}\_{1}, \hat{Y}\_{2}, …, \hat{Y}\_{n})$ (1)

dimana:

$\hat{Y}\_{RF}$ adalah prediksi akhir dari Random Forest.

 $\hat{Y}\_{1}, \hat{Y}\_{2}, …, \hat{Y}\_{n}$ adalah prediksi dari setiap pohon keputusan

$\hat{Y\_{RF}}= \frac{1}{n}\sum\_{i=1}^{n}\hat{Y\_{i}}$ (2)

$\hat{Y}\_{RF}$ adalah prediksi akhir dari *Random Forest*.

 $\hat{Y}\_{i}$ adalah prediksi dari pohon keputusan ke-*i.*

*n* adalah jumlah pohon keputusan dalam *Random Forest*.

Misalkan *K* adalah jumlah kelas dalam prediksi promosi pegawai, dan $Count\_{ik}$ adalah jumlah observasi yang diberikan kelas *k* di simpul *i* dalam pohon keputusan ke-*i*, maka prediksi dari setiap pohon keputusan $\hat{Y}\_{i}$ adalah $ \hat{Y}\_{i}=argmax\_{k}(Count\_{ik})$ [9].

Dalam *extremely randomized trees*, unsur keacakan diterapkan lebih lanjut dalam proses pemilihan fitur. Mirip dengan metode *random forest*, subset acak dari fitur kandidat digunakan, tetapi perbedaannya terletak pada langkah-langkah pengambilan keputusan terkait pemisahan [10]. Daripada mencari ambang batas yang paling diskriminatif, ambang batas diambil secara acak untuk setiap fitur kandidat, dan kemudian ambang batas terbaik yang dihasilkan secara acak dipilih sebagai aturan pemisahan. Pendekatan ini umumnya dapat mengurangi variasi model sedikit lebih banyak, walaupun dengan konsekuensi peningkatan bias yang sedikit lebih besar.

## Pemrosesan Data

Dataset yang telah tersedia tersebut merupakan dataset yang tidak seimbang antara label “tidak dipromosikan” dan “dipromosikan”, sehinggan untuk mengatasi masalah ketidakseimbangan kelas tersebut, diterapkan random oversampling. Pengambilan sampel berlebih secara acak melibatkan pemilihan contoh secara acak dari kelas minoritas, dengan penggantian, dan menambahkannya ke kumpulan data pelatihan.

## Evaluasi Kinerja Model

Mengukur kinerja model menggunakan metrik seperti akurasi, presisi, recall, F1-score, pada subset pengujian. Model yang digunakan untuk klasifikasi diukur dengan menghitung jumlah kelas yang diprediksi dengan benar (true positive), jumlah diprediksi bukan termasuk kelas tersebut dan benar (true negative) dan yang salah prediksi (false positive or false negative).

Tabel 12. Confusion Matrik

Nilai metrik tersebut dapat dihitung dari confusion matrix dengan persamaan sebagai berikut :

Accuracy = $\frac{TP + TN + TNet}{TP + FP + TN + FN + TNet + FNet }$ (3)

Precision = $\frac{TP}{TP + FP }$ (4)

Recall = $\frac{TP}{TP + FN + FNet }$ (5)

F score =2\* $\frac{Precision \* Recall}{Precision + Recall }$ (6)

HASIL DAN PEMBAHASAN

Sampel data latih dan data uji yang terdiri dari 13 kolom disajikan pada gambar 1.



Gambar 1. Sampel data latih dan data uji

Pada data latih, banyaknya label 1 (pegawai yang dipromosi) pada kolom is\_promoted sebanyak 4.668 data, sedangkan label 0 (pegawai yang tidak dipromosi) sebanyak 50.140 data. Gambar 2. Memperlihatkan distribusi data pada data latih yang digunakan.

 

Gambar 2. Distribusi data

Banyaknya pelatihan yang diikuti oleh pegawai adalah hanya satu sampai dua pelatihan. Umur pegawai paling banyak ada pada rentang 25-35 tahun. Penghargaan yang didapat oleh pegawai pada tahun sebelumnya kebanyakan adalah 3 penghargaan. Lamanya pegawai bekerja kebanyakan ada direntang satu sampai 5 tahun. Kebanyakan pegawai tidak mendapatkan reward pada tahun sebelumnya. Paling banyak pegawai mendapatkan nilai 50-60 untuk pelatihan yang telah dikutinya. Gambar 3 memperlihatkan distribusi label promosi pada data latih berdasarkan gender.



Gambar 3. Distribusi data promosi berdasarkan gender

Banyaknya pegawai bergender perempuan lebih sedikit dibandingkan laki-laki. Terlihat juga bahwa data pegawai yang dipromosikan tidak seimbang dengan data pegawai yang dipromosi. Gambar 4 memperlihatkan distribusi umur pegawai berdasarkan tingkat pendidikan.



Gambar 4. Distribusi umur pegawai berdasarkan tingkat pendidikan

Data memperlihatkan pegawai yang ada pada data latih paling banyak memiliki tingkat pendidikan sarjana. Gambar 5 memperlihatkan distribusi lama bekerja pegawai berdasarkan umur.



Gambar 5. Distribusi lama bekerja pegawai berdasarkan umur

 Grafik strip menunjukkan bahwa semakin lama layanan di perusahaan, penyebaran usia juga meningkat, menunjukkan adanya hubungan positif antara keduanya. Dengan kata lain, pegawai yang memiliki masa kerja lebih lama cenderung memiliki usia yang lebih tua. Hal ini sesuai dengan logika, mengingat pegawai yang lebih muda memiliki waktu kerja yang lebih singkat di perusahaan dan, oleh karena itu, memiliki investasi waktu yang lebih sedikit.

 Pada gambar 2 terlihat bahwa data pegawai yang dipromosikan lebih sedikit dibandingkan data pegawai yang tidak dipromosi, hal ini dapat menyebabkan bias pada model. Oleh karena itu, data harus diseimbangkan dengan menggunakan RandomOverSampler. Setelah dilakukan random over sampling data pegawai yang dipromosi dan tidak, masing-masing berjumlah 50.140 data. Gambar 6 memperlihatkan *confusion* matriks dari evaluasi model prediksi yang dibangun menggunakan metode *extremely randomized trees.*



Gambar 6. *Confusion* matriks

 Data pelatihan mencapai tingkat akurasi sebesar 99%, menandakan bahwa 99% dari entri dalam data pelatihan diprediksi dengan benar. Walaupun nilai ini sangat tinggi, kekhawatiran mengenai overfitting dapat diabaikan karena data pengujian juga memberikan tingkat akurasi yang tinggi, yaitu sebesar 98%. Secara lebih rinci, dari data pelatihan, 9.613 target yang tidak dipromosikan (1) diprediksi dengan benar, dan 10.073 dari target yang dipromosikan (0) juga diprediksi dengan benar. Sementara itu, dari data uji, 98% dari entri diprediksi dengan benar.

 Selain itu, matriks klasifikasi menunjukkan hasil yang lebih rinci, dengan 9.613 entri yang tidak dipromosikan (1) diprediksi dengan benar, dan 10.073 entri dari target yang dipromosikan (0) diprediksi dengan benar. Namun, perlu dicatat bahwa terdapat 2 entri yang seharusnya tidak dipromosikan tetapi salah diklasifikasikan sebagai dipromosikan, serta 368 entri yang seharusnya dipromosikan tetapi salah diklasifikasikan sebagai tidak dipromosikan. Gambar 7 memperlihatkan performa model pada data uji.



Gambar 7. Performa model

 Penggunaan data uji memperlihatkan model yang telah dibangun memiliki skor akurasi 0,98. Artinya, dalam set validasi, model memprediksi dengan tepat sebesar 98%. Gambar 8 memperlihatkan fitur yang paling berpengaruh pada pembentukan model.



Gambar 8. Fitur yang berpengaruh pada model prediksi

 Pada gambar terlihat fitur disajikan berdasarkan pengaruhnya pada prediksi, mulai dari yang memiliki pengaruh tertinggi hingga yang terendah. Grafik ini mempertimbangkan nilai absolut *SHapley Additive* exPlanations (SHAP)[[3]](#footnote-3), sehingga tidak memperhatikan apakah fitur tersebut mempengaruhi prediksi secara positif atau negatif. Grafik ini memberikan informasi tentang fitur utama yang memengaruhi prediksi dari satu observasi, serta magnitudo nilai SHAP untuk setiap fitur tersebut.

 Fitur avg\_training\_score memiliki nilai SHAP nilai absolut tertinggi. Artinya, mempunyai pengaruh paling tinggi terhadap prediksi promosi. Hal ini dapat diartikan bahwa pegawai. kemampuan belajar berperan dalam promosi. Departemen pegawai memainkan peran besar dalam memprediksi promosi. Hal ini dapat terjadi karena departemen tempat pegawai tersebut ditugaskan bergantung pada program studi dan spesialisasi yang dipelajari. Ada beberapa keterampilan yang diperlukan untuk posisi yang lebih tinggi di sebuah perusahaan.

 Fitur total\_score memiliki kontribusi tertinggi ketiga terhadap prediksi karena alasan yang sama dengan avg\_training\_score. usia juga memiliki pengaruh yang tinggi terhadap prediksi kemungkinan untuk dipromosikan. Ini karena fitur no\_of\_trainings memiliki nilai bentuk absolut terendah. Hal ini disebabkan banyaknya pelatihan yang memperhatikan produktivitas pegawai.

KESIMPULAN

Model prediksi yang dibangun menggunakan metode *extremely randomized trees* memiliki tingkat akurasi sebesar 99% pada data latih, menandakan bahwa 99% dari entri dalam data pelatihan diprediksi dengan benar. Penggunaan data uji memperlihatkan model yang telah dibangun memiliki skor akurasi 0,98. Artinya, dalam set validasi, model memprediksi dengan tepat sebesar 98%. Fitur yang paling berpengaruh dalam pembentukan model prediksi menunjukkan bahwa avg\_training\_score, department, total\_score, age dan sum\_metric memiliki pengaruh tertinggi dalam memprediksi promosi karyawan. Hasilnya juga menunjukkan bahwa pengaruh prediksi tersebut bervariasi berdasarkan gender.

SARAN

Pengembangan model prediksi untuk promosi pegawai dapat menggunakan metode-metode ensamble yang lain seperti *gradient-boosted trees* atau *bagging meta-estimator*.

DAFTAR PUSTAKA

[1] A. L. A and M. C. . Wilson, “Human Resource Systems and Sustained Competitive Advantage : A Competency-Based Perspective,” *Acad. Manag. Rev.*, vol. 19, no. 4, pp. 699–727, 1994.

[2] G. Demo, E. R. Neiva, I. Nunes, and K. Rozzett, “Human resources management policies and practices scale (HRMPPS): exploratory and confirmatory factor analysis,” *BAR - Brazilian Adm. Rev.*, vol. 9, no. 4, pp. 395–420, 2012, doi: 10.1590/s1807-76922012005000006.

[3] Y. K. Dwivedi, N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja, H. Albanna, M. A. Albashrawi, A. S. Al-Busaidi, J. Balakrishnan, Y. Barlette, *et al.*, “‘So what if ChatGPT wrote it?’ Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy,” *Int. J. Inf. Manage.*, vol. 71, no. March, 2023, doi: 10.1016/j.ijinfomgt.2023.102642.

[4] F. Paetz, W. J. Steiner, and H. Hruschka, “Advanced data analysis techniques with marketing applications,” *J. Bus. Econ.*, vol. 92, no. 4, pp. 557–561, 2022, doi: 10.1007/s11573-022-01101-z.

[5] R. Kalule, H. A. Abderrahmane, W. Alameri, and M. Sassi, “Stacked ensemble machine learning for porosity and absolute permeability prediction of carbonate rock plugs,” *Sci. Rep.*, vol. 13, no. 1, pp. 1–17, 2023, doi: 10.1038/s41598-023-36096-2.

[6] Y. Chen, X. Lin, and K. Zhan, *The Employee Promotion Decision based on the Randomforest Algorithm and the Analytic Hierarchy Process*, no. Isemss. Atlantis Press SARL, 2023. doi: 10.2991/978-2-38476-126-5\_185.

[7] M. Ilwani, G. Nassreddine, and J. Younis, “Machine Learning Application on Employee ‎Promotion,” *Mesopotamian J. Comput. Sci.*, vol. 2023, pp. 106–120, 2023, doi: 10.58496/mjcsc/2023/013.

[8] L. Rokach and O. Maimon, *Data Mining With Decision Tree*. 2014.

[9] W.-Y. Loh, “Classification and regression trees CLASSIFICATION TREES,” *C*, vol. 1, pp. 14–23, 2011, doi: 10.1002/widm.8.

[10] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.

1. https://scikit-learn.org/stable/modules/ensemble.html [↑](#footnote-ref-1)
2. https://www.kaggle.com/datasets/muhammadimran112233/employees-evaluation-for-promotion/data [↑](#footnote-ref-2)
3. https://shap.readthedocs.io/en/latest/ [↑](#footnote-ref-3)