

Prediksi Pengunduran Diri Karyawan Dengan Pendekatan Deep Neural Network Dalam Machine Learning

Satya Wirapati^{a1}, Luh Gede Astuti^{a2}, I Made Widiartha^{a3}, I Wayan Santiyasa^{a4}

^aProgram Studi Informatika, FMIPA,
Universitas Udayana

Jl. Raya Kampus Unud No.9, Jimbaran, Kec. Kuta Sel., Kabupaten Badung, Bali 80361, Country

¹satyathewise@email.com

²lg.astuti@unud.ac.id

³madewidiartha@unud.ac.id

⁴santiyasa@unud.ac.id

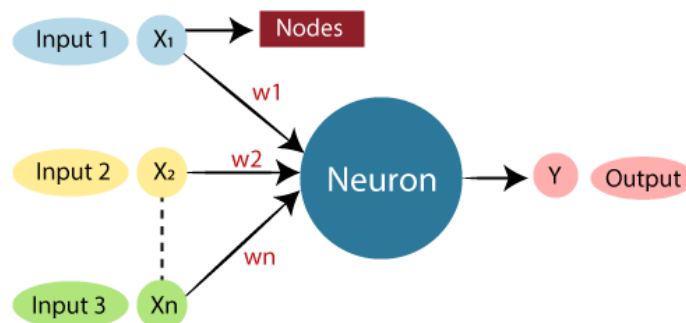
Abstract

In the current digital era, technological advancements are rapidly progressing and impacting various sectors, including the workforce. Employees are crucial assets for companies to achieve their business goals. Therefore, companies need to take steps to retain quality and potential employees, as well as avoid unwanted employee turnover. Employee turnover can incur significant costs for companies, such as recruitment expenses, training costs, and loss of productivity. Hence, companies need to have the ability to predict which employees are likely to resign, so they can take appropriate actions to retain them. In this regard, machine learning techniques can be an effective solution for predicting employee turnover. By using the right machine learning algorithms, companies can identify the factors influencing employees' decisions to resign and take appropriate actions to retain them. Therefore, this research aims to apply machine learning techniques in predicting employee turnover. It is hoped that the results of this research can help companies identify the factors influencing employee turnover and take appropriate actions to retain them, thereby improving overall company performance. In this research, a machine learning approach using a Deep Neural Network (DNN) was implemented to predict employee turnover. The research utilized the IBM HR Analytics Employee Attrition & Performance dataset, a secondary dataset sourced from Kaggle. Preprocessing steps included cleaning data, handling missing values, encoding categorical variables, normalizing data, and addressing class imbalances through resampling with ADASYN. The DNN model achieved an accuracy of 86% in predicting employee resignations. This result demonstrates the effectiveness of machine learning techniques in identifying patterns and providing actionable insights for companies to reduce employee turnover.

Keywords: machine learning, ANN, attrition, regularization, resampling

1. Pendahuluan

Penelitian ini membahas dua topik utama: **Pengunduran Diri Karyawan** dan **Machine Learning**, khususnya penerapannya dalam memprediksi pengunduran diri karyawan.



Gambar 1 Ilustrasi JST

Gambar 1 Ilustrasi JST menjelaskan secara singkat apa itu jaringan syaraf tiruan, yang dimana x_1, x_2 dan sampai seterusnya merupakan nodes input dari input layer, banyak input layer merupakan banyaknya kolom dataset yang kita pakai, lalu setelah itu nodes akan disambungkan ke nodes di layer yang selanjutnya menggunakan *weight* atau bobot, yang dimana nilai akan dikeluarkan melalui nodes output pada output layer

Pengunduran Diri Karyawan didefinisikan sebagai perubahan status pekerjaan ketika seorang karyawan meninggalkan organisasi. Menurut artikel "Understanding employee turnover: A meta-analysis of its antecedents and consequences," pengunduran diri ini dipengaruhi oleh faktor organisasi, personal, dan situasional. Artikel dan jurnal terkait memberikan wawasan untuk memahami dan mengatasi masalah pengunduran diri karyawan.

Machine Learning didefinisikan sebagai teknik komputasi yang memungkinkan sistem untuk belajar dari data tanpa harus secara eksplisit diprogram. Artikel [1] membahas bagaimana machine learning dapat diterapkan pada pemeliharaan prediktif. Selain itu, machine learning telah diterapkan dalam berbagai bidang lain seperti pengobatan kanker. Algoritma machine learning dikategorikan ke dalam empat jenis: Supervised Learning, Unsupervised Learning, Semi-supervised Learning, dan Reinforcement Learning, dengan masing-masing memiliki metode dan penerapannya sendiri.

Artificial Neural Networks (ANN), sebuah teknologi yang didasarkan pada studi tentang otak dan sistem saraf, meniru jaringan saraf biologis dan digunakan untuk memproses input guna menghasilkan output yang diprediksi. Menurut [2], ANN terdiri dari elemen-elemen pemrosesan yang dihubungkan dan diatur dalam lapisan-lapisan tertentu, yang memungkinkan simulasi aktivitas listrik otak dan sistem saraf.

Deep Neural Networks (DNN), jenis ANN yang digunakan dalam penelitian ini, memiliki beberapa keunggulan dalam menangani data yang kompleks dan non-linear serta dalam memproses berbagai jenis data. DNN juga fleksibel dalam menyesuaikan parameter-parameter untuk mendapatkan hasil prediksi yang akurat.

Feature Selection adalah teknik penting dalam machine learning yang bertujuan untuk meningkatkan performa model, mengurangi dimensi data, meningkatkan interpretabilitas, mengurangi overfitting, dan menghemat waktu serta sumber daya [3] mencatat bahwa feature selection membantu memilih fitur-fitur yang paling relevan dan mengeliminasi yang tidak penting. Recursive Feature Elimination (RFE) adalah salah satu metode yang digunakan untuk memilih fitur-fitur penting dalam dataset, seperti yang dibahas oleh [4].

Penelitian-penelitian empiris yang dirangkum menunjukkan bahwa model machine learning, seperti logistic regression, decision tree, random forest, dan neural networks, mampu memprediksi pengunduran diri karyawan dengan akurasi yang baik. Faktor-faktor seperti umur, jenis kelamin, gaji, pekerjaan, masa kerja, kinerja, dan kepuasan kerja memainkan peran penting dalam prediksi tersebut. Penelitian ini memberikan wawasan yang dapat membantu organisasi dalam mengambil tindakan preventif terhadap pengunduran diri karyawan.

Retensi karyawan merupakan salah satu tantangan utama dalam manajemen sumber daya manusia, terutama di era digital saat ini di mana persaingan tenaga kerja semakin ketat. Tingginya tingkat turnover karyawan dapat menyebabkan dampak negatif yang signifikan bagi perusahaan, termasuk biaya rekrutmen dan pelatihan, hilangnya produktivitas, serta penurunan moral karyawan yang tersisa. Oleh karena itu, diperlukan langkah strategis untuk memprediksi pengunduran diri karyawan sehingga perusahaan dapat mengambil tindakan preventif guna mempertahankan karyawan berkualitas.

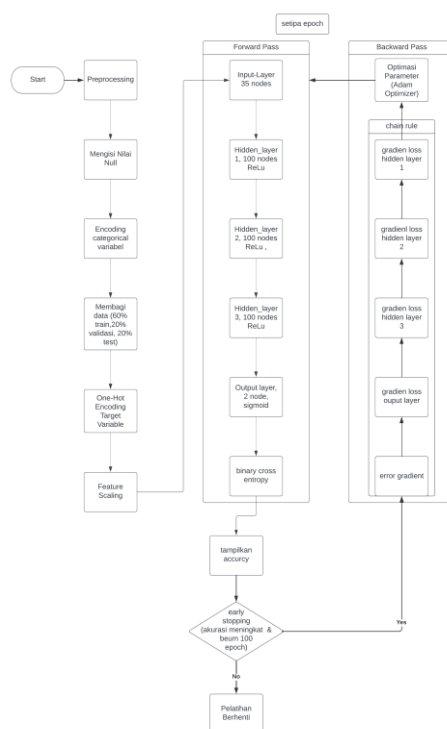
Berbagai penelitian telah dilakukan untuk memprediksi pengunduran diri karyawan menggunakan pendekatan machine learning. Misalnya, penelitian oleh Ali Raza et al. (2022) menggunakan model logistic regression, decision tree, random forest, dan support vector machine, yang menunjukkan bahwa metode ini dapat memprediksi pengunduran diri dengan tingkat akurasi yang baik. Studi lain oleh Parv N. Gandhi et al. (2020) [5] membandingkan logistic regression dengan random forest, di mana random forest mencapai akurasi 97.5% dalam memprediksi turnover karyawan. Selain itu, pendekatan berbasis jaringan saraf tiruan (Artificial Neural Network/ANN) juga telah menunjukkan keunggulan dalam memodelkan hubungan kompleks dalam data karyawan, seperti yang diungkapkan oleh Shawni Dutta dan Samir Kumar Bandyopadhyay (2020) [6]. Namun, metode ANN tradisional sering kali memerlukan pengoptimalan lebih lanjut untuk meningkatkan kinerjanya.

Urgensi penelitian ini terletak pada kebutuhan untuk mengatasi keterbatasan metode sebelumnya dengan memanfaatkan Deep Neural Network (DNN). DNN dipilih karena kemampuannya dalam menangkap pola yang kompleks dan hubungan non-linear antar variabel, seperti yang dibuktikan dalam berbagai penelitian sebelumnya. Dengan kemampuan representasi yang lebih kuat dibandingkan ANN tradisional, DNN mampu menghasilkan prediksi yang lebih akurat, terutama dalam dataset dengan struktur dan distribusi data yang kompleks. Dalam penelitian ini, pendekatan DNN diterapkan pada dataset *IBM HR Analytics Employee Attrition & Performance* untuk memprediksi pengunduran diri karyawan. Hasilnya diharapkan tidak hanya memberikan akurasi prediksi yang tinggi, tetapi juga memberikan wawasan yang dapat diimplementasikan dalam strategi manajemen sumber daya manusia perusahaan.

Penelitian ini bertujuan untuk menjawab tantangan tersebut dengan mengembangkan model prediksi yang efektif dan andal, yang dapat membantu perusahaan memitigasi risiko turnover karyawan secara proaktif. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam upaya pengelolaan tenaga kerja yang berbasis data.

2. Metodologi Penelitian

2.1 Deskripsi Umum



Gambar 2 Diagram Flowchart Sistem

Penelitian ini berfokus pada pembangunan model prediksi pengunduran diri karyawan menggunakan pendekatan *Deep Neural Network (DNN)* dalam *machine learning*. seperti pada *flowchart* yang dapat dilihat pada Gambar 2 Diagram *Flowchart* sistem, Model ini dirancang dengan beberapa lapisan jaringan saraf buatan untuk menangkap pola kompleks dalam data karyawan. Proses dimulai dengan inisialisasi model menggunakan *Sequential* dari Keras, dengan lapisan input dan tiga lapisan tersembunyi yang masing-masing terdiri dari 100 neuron, serta lapisan output dengan satu neuron menggunakan fungsi aktivasi sigmoid. Model dikompilasi menggunakan optimizer Adam dan fungsi *loss binary_crossentropy*. Untuk mencegah *overfitting*, digunakan *Early Stopping* dan metode regularisasi L1 jika diperlukan. Preprocessing data mencakup penghapusan kolom tidak relevan, penanganan nilai hilang, *encoding* variabel kategorikal, normalisasi data, penghapusan *outliers*, dan *resampling* data untuk mengatasi ketidakseimbangan kelas.

2.2 Data dan Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah dataset "IBM HR Analytics Employee Attrition & Performance" dari repositori online Kaggle. Data ini mencakup berbagai informasi tentang karyawan, pekerjaan, lingkungan kerja, kinerja, dan faktor lainnya yang relevan dengan prediksi pengunduran diri.

2.3 Prapemrosesan

Tahapan *preprocessing* mencakup penghapusan kolom yang tidak diperlukan, mengisi nilai yang hilang, *encoding* variabel kategorikal, penghapusan *outliers* menggunakan Z-score, normalisasi data menggunakan metode Z-Score, serta *resampling* menggunakan metode ADASYN untuk mengatasi ketidakseimbangan kelas. Data kemudian dibagi menjadi subset untuk pelatihan, pengujian, dan validasi model dengan rasio 70:30 pada setiap tahap.

2.4 Kerangka Model

Penelitian ini melibatkan empat kerangka model DNN yang berbeda untuk mengevaluasi kinerja prediksi:

- a Model tanpa resampling dan tanpa regularisasi L1.
- b Model dengan resampling tanpa regularisasi L1.
- c Model tanpa resampling dengan regularisasi L1.
- d Model dengan resampling dan regularisasi L1.

2.5 Implementasi Early Stopping

Early Stopping digunakan untuk mencegah *overfitting* dan menghemat waktu komputasi dengan menghentikan pelatihan jika tidak ada peningkatan performa pada data validasi setelah 5 epoch berturut-turut.

2.6 Evaluasi Metode Komputasi

Evaluasi model dilakukan menggunakan *Confusion Matrix* untuk memahami kinerja prediksi dengan metrik-metrik seperti akurasi, *precision*, *recall*, dan F1 score. *Confusion Matrix* memberikan gambaran detail tentang kinerja model dalam menangani prediksi pengunduran diri karyawan.

2.7 Perancangan GUI

Dibuat sebuah Graphical User Interface (GUI) menggunakan Python dengan library Tkinter untuk memudahkan pengguna, terutama pihak HR, dalam mengoperasikan model yang telah dibuat. GUI ini dirancang agar sederhana dan efisien untuk penggunaan praktis.

Dengan desain dan metode ini, model DNN diharapkan dapat memberikan prediksi yang akurat dan andal mengenai pengunduran diri karyawan, membantu perusahaan dalam pengambilan keputusan terkait manajemen sumber daya manusia.

3. Hasil dan Pembahasan

3.1 Pembuatan Model Machine Learning

Proses pembuatan model *Deep Neural Network* (DNN) untuk memprediksi pengunduran diri karyawan dimulai dengan mengimpor pustaka-pustaka Python yang diperlukan seperti Pandas, Numpy, Scikit-learn, TensorFlow, dan Keras. Setelah itu, dataset di-*load* menggunakan Pandas dan dilakukan *preprocessing*, yang meliputi pembersihan data, pengisian nilai yang hilang, *encoding* variabel kategorikal, penghapusan *outliers*, normalisasi data, *resampling* menggunakan metode ADASYN, dan pembagian data menjadi *training*, *validation*, dan *test set*.

3.2 Kerangka dan Pelatihan Model

Model DNN dibangun dengan menggunakan beberapa lapisan *dense* dengan 100 neuron dan fungsi aktivasi ReLU untuk lapisan tersembunyi, serta fungsi sigmoid untuk lapisan output. Model dikompilasi dengan optimizer Adam dan loss function *binary crossentropy*. Model dilatih dengan menggunakan *early stopping* untuk mencegah *overfitting*.

3.3 Evaluasi Model

Evaluasi model dilakukan menggunakan metrik akurasi, loss, dan confusion matrix. Beberapa skenario evaluasi dilakukan:

Default: Model yang dilatih tanpa resampling dan regularisasi menunjukkan *overfitting* dengan akurasi pelatihan tinggi (0.9497) tetapi akurasi validasi lebih rendah (0.8608). Confusion matrix menunjukkan bahwa model lebih baik dalam memprediksi karyawan yang tidak akan resign daripada yang akan resign.

Resampling: Model dengan data yang di-resample menggunakan ADASYN menunjukkan peningkatan akurasi validasi (0.8959) dan performa yang lebih baik dalam memprediksi kelas minoritas (karyawan yang resign). Confusion matrix menunjukkan peningkatan signifikan pada jumlah true positive.

Regularisasi: Dengan penerapan batch normalization, dropout, dan regularisasi L1, masalah *overfitting* berhasil diatasi dengan selisih akurasi pelatihan dan validasi sekitar 5%. Confusion matrix menunjukkan bahwa regularisasi tidak mengatasi masalah imbalance dataset.

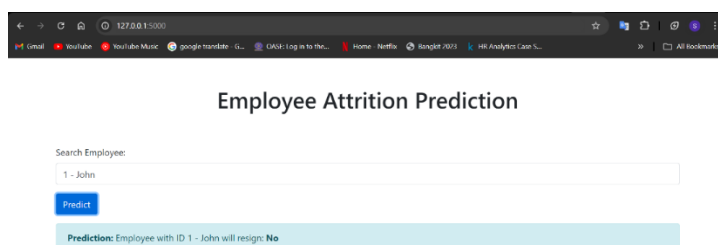
Resampling & Regularisasi: Kombinasi resampling dan regularisasi menghasilkan model dengan akurasi pelatihan (0.9075) dan akurasi validasi (0.9022) yang sangat dekat, tanpa indikasi *overfitting*. Confusion matrix menunjukkan prediksi yang lebih baik untuk kedua kelas.

3.4 Evaluasi Akhir dengan Data Testing

Setelah menyelesaikan masalah *overfitting* dan imbalance dataset, model diuji pada data yang belum pernah dilihat sebelumnya, menghasilkan akurasi sebesar 86%. Confusion matrix menunjukkan performa yang baik dalam mengklasifikasikan data dengan precision 82.4%, recall 87.9%, dan specificity 83.4%.

3.5 Antarmuka Pengguna

Sebuah aplikasi berbasis web dikembangkan menggunakan Python dan Flask untuk memungkinkan pengguna (HR) melakukan prediksi pengunduran diri karyawan. Aplikasi ini menyediakan antarmuka yang mudah digunakan untuk mengunggah file CSV, mencari karyawan berdasarkan ID atau nama, dan memprediksi apakah karyawan akan resign atau tidak.



Gambar 3 UI Prediksi Karyawan

Pada Gambar 3 UI Prediksi Karyawan ditampilkan *UI* web hasil dari penelitian ini, pertama diatas terdapat nama aplikasi web tersebut, yaitu “Employee Attrition Prediction”, selanjutnya pada form pencarian kita bisa memasukkan id dari karyawan atau nama karyawan tersebut, setelah itu kita tekan tombol “Prediksi” dan akan muncul hasil prediksinya, apakah karyawan akan *resign* atau tidak.

4. Kesimpulan

Penelitian ini berhasil mengimplementasikan *Deep Neural Network* (DNN) untuk memprediksi pengunduran diri karyawan menggunakan dataset *IBM HR Analytics Employee Attrition & Performance*. Proses preprocessing data meliputi penghapusan data tidak relevan, normalisasi, encoding variabel kategorikal, serta resampling dengan metode ADASYN untuk mengatasi ketidakseimbangan kelas. Model DNN yang dirancang memiliki tiga *hidden layer* dengan fungsi aktivasi ReLU dan output layer dengan fungsi sigmoid. Optimasi model dilakukan menggunakan algoritma Adam dengan fungsi loss *binary crossentropy*, disertai regularisasi L1 dan teknik *early stopping* untuk mencegah *overfitting*.

Hasil pengujian menunjukkan bahwa model mencapai akurasi **86%** dalam memprediksi pengunduran diri karyawan. Evaluasi menggunakan matriks konfusi (Confusion Matrix) menunjukkan performa yang baik dalam mengidentifikasi karyawan yang berisiko mengundurkan diri. Akurasi ini membuktikan bahwa DNN efektif dalam menangkap pola kompleks dalam data karyawan, termasuk faktor demografi, kepuasan kerja, dan kinerja. Dengan demikian, penelitian ini memberikan kontribusi signifikan bagi perusahaan untuk mengambil langkah proaktif dalam mempertahankan karyawan mereka.

Daftar Pustaka

- [1] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE transactions on industrial informatics*, vol. 11, no. 3, pp. 812-820, 2014.
- [2] S. Walczak and N. Cerpa, "Artificial Neural Networks," *Encyclopedia of Physical Science and Technology*, Vols. -, no. -, pp. 631-645, 2003.
- [3] A. Elisseeff and I. Guyon, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, no. 3, pp. 1157-1182, 2003.
- [4] J. Brownlee, "How to choose a feature selection method for machine learning," *Machine Learning Mastery*, vol. 10, no. -, pp. 1-7, 2019.
- [5] P. N. Gandhi, A. Jangale and G. S. Mhaske, "Employee Attrition Prediction using Machine," *JETIR*, vol. 7, no. 9, p. 9, 2020.
- [6] S. Dutta and S. Bandyopadhyay, "Employee attrition prediction using neural network cross validation method," *International Journal of Commerce and Management Research*, vol. 6, no. 3, pp. 80-85, 2020.