



Klasifikasi Ketidakhadiran di Tempat Kerja Menggunakan Metode *Support Vector Machine*

Natasya Lexbert Sumeisey, Titik Misriati, Imam Nawawi*

Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

Abstrak: Penelitian ini bertujuan untuk mengklasifikasikan ketidakhadiran karyawan di tempat kerja menggunakan metode Support Vector Machine (SVM), dengan menggunakan dataset yang mencakup informasi demografi, pekerjaan, dan faktor-faktor lain yang terkait dengan ketidakhadiran. Dataset yang digunakan berisi catatan ketidakhadiran karyawan sebuah perusahaan kurir di Brasil dari *UCI Machine Learning*. Penelitian ini mengimplementasikan model SVM dengan kernel Radial Basis Function (RBF), yang dipilih karena kemampuannya dalam menangani data non-linier. Hasil evaluasi model menunjukkan kinerja yang sangat baik, dengan AUC sebesar 0,995, akurasi mencapai 98,1%, dan skor F1 sebesar 0,981, yang menunjukkan keseimbangan yang sangat baik antara presisi dan ingatan. Model tersebut berhasil memprediksi sebagian besar ketidakhadiran karyawan secara akurat, dengan kesalahan prediksi yang minimal. Namun, masih ada beberapa kesalahan kecil dalam memprediksi ketidakhadiran, yang dapat diperbaiki dengan menyetel hiperparameter dan menambahkan fitur tambahan yang terkait dengan faktor-faktor yang memengaruhi ketidakhadiran. Secara keseluruhan, penelitian ini menunjukkan bahwa model SVM merupakan alat yang efektif dan efisien untuk memprediksi ketidakhadiran karyawan, dengan hasil yang dapat diterapkan dalam mengelola ketidakhadiran di organisasi.

Kata Kunci: Ketidakhadiran Karyawan, Klasifikasi, Prediksi Ketidakhadiran, Pembelajaran Mesin

DOI:

<https://doi.org/10.53697/jkomitek.v5i2.2618>

*Correspondence: Imam Nawawi

Email: imam.imw@bsi.ac.id

Received: 23-04-2025

Accepted: 23-05-2025

Published: 23-06-2025



Copyright: © 2025 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: This study aims to classify employee workplace absence using the Support Vector Machine (SVM) method, using a dataset that includes demographic information, occupation, and other factors related to absence. The dataset used contains employee absence records a courier company in Brazil from UCI Machine Learning. This study implemented an SVM model with a Radial Basis Function (RBF) kernel, which was chosen for its ability to handle non-linear data. The model evaluation results showed excellent performance, with AUC of 0.995, accuracy reaching 98.1%, and F1-score of 0.981, indicating an excellent balance between precision and recall. The model managed to predict most employee absences accurately, with minimal prediction error. However, there are still some small errors in predicting absenteeism, which can be improved by tuning the hyperparameters and adding additional features related to factors that influence absenteeism. Overall, this study shows that SVM models are effective and efficient tools for predicting employee absenteeism, with results that can be applied in managing absenteeism in organizations.

Keywords: Support Vector Machine (SVM), Employee Absenteeism, Classification, Absenteeism Prediction, Machine Learning Support Vector Machine (SVM)

Pendahuluan

Ketidakhadiran karyawan di tempat kerja adalah masalah yang sering kali dihadapi oleh berbagai organisasi dan dapat berdampak negatif terhadap produktivitas serta keberlanjutan operasional perusahaan. Ketidakhadiran tidak hanya mengurangi efektivitas kerja, tetapi juga menambah beban biaya yang signifikan, baik dalam hal penggantian pekerja sementara, pengurangan output, maupun penurunan moral di antara rekan kerja yang terpengaruh oleh absennya individu tersebut (Traoré, 2024). Oleh karena itu, pemahaman yang lebih mendalam tentang faktor-faktor yang memengaruhi ketidakhadiran sangat penting agar organisasi dapat mengambil langkah-langkah yang tepat untuk meminimalkan dampaknya.

Ketidakhadiran dapat disebabkan oleh berbagai faktor yang sangat beragam, termasuk faktor kesehatan, faktor pribadi, faktor organisasi, serta faktor eksternal seperti transportasi atau kondisi cuaca. Beberapa penelitian menunjukkan bahwa ketidakhadiran sering kali dipengaruhi oleh kondisi kesehatan fisik dan mental karyawan, serta faktor-faktor lain yang terkait dengan stres kerja, kepuasan kerja, dan hubungan interpersonal di tempat kerja (Nawata, 2024). Oleh karena itu, penting bagi organisasi untuk memantau ketidakhadiran karyawan dengan cara yang lebih sistematis dan proaktif untuk dapat meminimalkan risiko ketidakhadiran yang tidak terkendali.

Seiring dengan pesatnya perkembangan teknologi informasi, perusahaan semakin mengandalkan data untuk membuat keputusan yang lebih tepat. Salah satu pendekatan yang dapat digunakan untuk memprediksi ketidakhadiran adalah dengan menerapkan teknik pembelajaran mesin (machine learning). Pembelajaran mesin, khususnya teknik klasifikasi, memungkinkan analisis data yang lebih mendalam dan dapat memberikan prediksi yang lebih akurat dibandingkan dengan metode konvensional (Skorikov et al., 2020). Dengan memanfaatkan algoritma pembelajaran mesin, perusahaan dapat mengidentifikasi pola-pola yang tidak mudah terdeteksi dalam data ketidakhadiran karyawan, serta faktor-faktor yang berkontribusi pada ketidakhadiran tersebut.

Metode pembelajaran mesin menawarkan potensi besar dalam menangani masalah ketidakhadiran. Salah satu algoritma pembelajaran mesin yang terkenal adalah Support Vector Machine (SVM), yang telah digunakan dalam berbagai aplikasi klasifikasi, seperti pengenalan wajah, klasifikasi teks, dan analisis citra. SVM bekerja dengan cara menemukan hyperplane yang memisahkan data dalam ruang berdimensi tinggi, yang memungkinkan klasifikasi data yang kompleks dengan akurasi yang tinggi (Rajath & Pandita, 2022). Keunggulan utama dari SVM adalah kemampuannya untuk menangani data yang tidak dapat dipisahkan secara linear melalui penggunaan kernel yang tepat, sehingga membuat SVM sangat cocok untuk digunakan dalam analisis data ketidakhadiran yang seringkali melibatkan variabel yang kompleks dan beragam (Du et al, 2024).

Support Vector Machine memiliki berbagai keunggulan yang menjadikannya pilihan tepat untuk masalah klasifikasi ketidakhadiran karyawan. Salah satu keuntungan utama SVM adalah kemampuannya untuk bekerja dengan data yang memiliki dimensi tinggi dan jumlah fitur yang banyak (Nath et al, 2022). Dalam konteks ketidakhadiran karyawan, data yang digunakan mungkin melibatkan berbagai jenis variabel, seperti faktor demografis (usia, jenis kelamin, status pernikahan), kondisi kesehatan, kepuasan kerja, dan faktor

organisasi lainnya (sistem pengelolaan waktu, budaya perusahaan, dan sebagainya). SVM dapat menangani variabel-variabel ini dengan efektif dan menghasilkan model yang mampu memprediksi ketidakhadiran berdasarkan pola yang terdeteksi dalam data.

Metode ini menggunakan prinsip margin yang memisahkan kelas-kelas data dengan jarak yang maksimum. Dengan demikian, SVM tidak hanya memberikan klasifikasi yang akurat, tetapi juga memungkinkan untuk mengidentifikasi titik data yang paling kritis atau yang paling sulit diprediksi, yang sering kali memberikan wawasan lebih dalam tentang faktor-faktor yang memengaruhi ketidakhadiran. Salah satu keunggulan SVM lainnya adalah kemampuannya untuk menghindari overfitting, terutama ketika menggunakan kernel yang tepat. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi dan mengevaluasi potensi SVM dalam memprediksi ketidakhadiran karyawan di tempat kerja, serta untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap ketidakhadiran tersebut (Ratnayake & Udawatta, 2021).

Penelitian ini bertujuan untuk mengembangkan model klasifikasi ketidakhadiran karyawan menggunakan metode Support Vector Machine (SVM), dengan fokus pada faktor-faktor yang dapat mempengaruhi ketidakhadiran. Data yang digunakan dalam penelitian ini akan mencakup faktor-faktor seperti data demografis, riwayat kesehatan, kepuasan kerja, serta variabel organisasi yang relevan. Diharapkan, model SVM yang dikembangkan dapat memberikan prediksi yang lebih akurat dan mendalam mengenai ketidakhadiran karyawan. Selain itu, penelitian ini juga bertujuan untuk memberikan wawasan lebih dalam mengenai hubungan antara faktor-faktor yang mempengaruhi ketidakhadiran, yang dapat digunakan oleh perusahaan untuk merumuskan kebijakan yang lebih baik dalam mengelola absensi karyawan.

Metodologi

Penelitian ini bertujuan untuk mengklasifikasikan ketidakhadiran karyawan di tempat kerja dengan menggunakan metode Support Vector Machine (SVM). Untuk mencapai tujuan tersebut, kami menggunakan aplikasi Orange Data Mining, sebuah platform visual untuk analisis data dan pembelajaran mesin, yang memungkinkan pemrosesan data multivariat serta implementasi model klasifikasi yang efisien. Berikut adalah langkah-langkah yang digunakan dalam penelitian ini:

A. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset Absenteeism at Work yang berisi catatan ketidakhadiran karyawan yang di dapatkan dari UCI Machine Learning Data ini meliputi catatan ketidakhadiran karyawan dari Juli 2007 hingga Juli 2010. Setiap record dalam dataset menggambarkan ketidakhadiran seorang karyawan beserta atribut-atribut terkait, seperti alasan ketidakhadiran, waktu ketidakhadiran, kondisi fisik dan sosial karyawan, serta faktor-faktor lain yang berpotensi memengaruhi absensi mereka.

Dataset ini mencakup 21 atribut, yang meliputi informasi mengenai alasan ketidakhadiran, bulan ketidakhadiran, hari dalam minggu, musim, biaya transportasi, jarak dari rumah ke tempat kerja, usia, beban kerja, status disiplin, pendidikan, jumlah anak, status sosial (peminum dan perokok sosial), jumlah hewan peliharaan, serta data fisik

karyawan seperti berat badan, tinggi badan, dan indeks massa tubuh (BMI). Tabel 1 berikut memberikan gambaran mengenai atribut dalam dataset.

Tabel 1. Deskripsi Dataset

Atribut	Deskripsi
ID	Identifikasi individu karyawan.
Alasan ketidakhadiran (ICD)	Stratifikasi alasan ketidakhadiran berdasarkan Kode Penyakit Internasional (ICD), yang mencakup 21 kategori penyakit (I hingga XXI).
Bulan ketidakhadiran	Bulan ketidakhadiran karyawan.
Hari dalam minggu	Hari dalam minggu (Senin hingga Jumat).
Musim	Musim saat ketidakhadiran terjadi (Musim panas, gugur, dingin, semi).
Biaya transportasi	Biaya transportasi yang dikeluarkan oleh karyawan untuk menuju tempat kerja.
Jarak dari rumah ke tempat kerja (km)	Jarak rumah karyawan ke tempat kerja dalam kilometer.
Lama bekerja	Lama bekerja di perusahaan (dalam tahun).
Usia	Usia karyawan.
Beban kerja rata-rata per hari	Rata-rata beban kerja karyawan per hari.
Target yang tercapai	Target yang tercapai oleh karyawan.
Gagal disiplin (ya=1, tidak=0)	Indikator apakah karyawan pernah gagal disiplin.
Pendidikan	Tingkat pendidikan karyawan (SMA, Sarjana, Pascasarjana, Master dan Doktor).
Jumlah anak	Jumlah anak yang dimiliki oleh karyawan.
Peminum sosial (ya=1, tidak=0)	Indikator apakah karyawan merupakan peminum sosial.
Perokok sosial (ya=1, tidak=0)	Indikator apakah karyawan merupakan perokok sosial.
Jumlah hewan peliharaan	Jumlah hewan peliharaan yang dimiliki oleh karyawan.
Berat badan	Berat badan karyawan.
Tinggi badan	Tinggi badan karyawan.
Indeks massa tubuh (BMI)	Indeks massa tubuh karyawan.
Waktu ketidakhadiran (dalam jam) (target)	Waktu ketidakhadiran dalam jam (target yang akan diprediksi).

B. Pra-pemrosesan Data

Sebelum membangun model klasifikasi, tahap pra-pemrosesan data dilakukan untuk memastikan bahwa data siap digunakan dalam analisis. Langkah-langkah pra-pemrosesan ini bertujuan untuk membersihkan data, menangani nilai yang hilang, mengubah tipe data, dan memastikan data dalam format yang sesuai (Tuppada & Patil, 2023). Proses ini dilakukan menggunakan beberapa komponen yang tersedia di Orange Data Mining. Tabel 2 berikut adalah langkah-langkah pra-pemrosesan yang dilakukan.

Tabel 2. Pra-pemrosesan Data pada Orange

Langkah	Deskripsi
Import Data	Data diimpor menggunakan komponen File untuk memuat dataset ke dalam Orange.
Informasi Data	Komponen Data Info digunakan untuk memeriksa informasi dasar dari dataset, seperti jumlah instance dan atribut.
Penyandian Kategorikal	Menggunakan Edit Domain untuk mengubah atribut kategorikal (seperti hari dalam minggu dan musim) menjadi format yang dapat digunakan oleh model SVM.
Seleksi Kolom	Komponen Select Columns digunakan untuk memilih atribut yang relevan dan menghapus atribut yang tidak diperlukan untuk analisis lebih lanjut.

C. Pemilihan Model: Support Vector Machine (SVM)

Setelah data siap, model klasifikasi Support Vector Machine (SVM) dipilih untuk menganalisis data. SVM adalah metode pembelajaran mesin yang efektif dalam menangani data yang non-linear dan berdimensi tinggi (Al-Hamadani, 2023). Pada tahap ini, komponen SVM dalam Orange Data Mining digunakan untuk membangun model dengan parameter yang dioptimalkan. Langkah-langkah yang dilakukan dalam pemilihan model ditunjukkan pada Tabel 3.

Tabel 3. Langkah-langkah Pemilihan Model

Langkah	Deskripsi
Pemilihan Kernel	Kernel Radial Basis Function (RBF) dipilih karena kemampuannya dalam menangani data yang tidak dapat dipisahkan secara linear.
Optimasi Parameter	Parameter C dan Gamma dioptimalkan menggunakan komponen Test and Score untuk memastikan model memaksimalkan margin pemisah dan menghindari overfitting.
Pelatihan Model	Data pelatihan digunakan untuk melatih model SVM dan membangun fungsi pemisah terbaik berdasarkan hyperplane yang ditemukan oleh SVM.

D. Evaluasi Model

Evaluasi model dilakukan menggunakan beberapa metrik utama, seperti akurasi, precision, recall, dan F1-score, untuk menilai kinerja model dalam mengklasifikasikan ketidakhadiran karyawan. Komponen Test and Score pada Orange digunakan untuk menghitung nilai metrik ini.

Tabel 4. Metrik Evaluasi Model

Metrik Evaluasi	Deskripsi
Akurasi	Persentase prediksi yang benar dibandingkan dengan jumlah total prediksi.
Precision	Proporsi prediksi positif yang benar dibandingkan dengan semua prediksi positif.
Recall	Proporsi prediksi positif yang benar dibandingkan dengan semua kasus positif yang sebenarnya.
F1-Score	Rata-rata harmonis antara precision dan recall, yang memberikan gambaran keseimbangan keduanya.
Confusion Matrix	Matriks yang menggambarkan distribusi hasil prediksi antara kelas yang benar dan yang salah.

E. Pengujian dan Implementasi

Setelah model dilatih, komponen Test and Score digunakan untuk menguji kinerja model pada data pengujian. Hasil prediksi akan dievaluasi menggunakan Confusion Matrix, yang akan menunjukkan jumlah prediksi yang benar dan salah (Rainio et al, 2024). Visualisasi hasil ini memberikan gambaran tentang kualitas klasifikasi model dalam memprediksi ketidakhadiran karyawan.

F. Analisis Hasil

Hasil evaluasi dari model klasifikasi akan dianalisis untuk mengidentifikasi faktor-faktor yang paling memengaruhi ketidakhadiran karyawan. Analisis ini akan memberikan wawasan yang dapat digunakan oleh manajer untuk mengoptimalkan kebijakan dan strategi dalam menangani ketidakhadiran. Faktor-faktor yang berpengaruh signifikan akan dijadikan bahan pertimbangan dalam pengambilan keputusan strategis terkait manajemen sumber daya manusia.

Hasil dan Pembahasan

Konfigurasi Model SVM

Konfigurasi model SVM yang diterapkan menggunakan aplikasi Orange Data Mining melibatkan beberapa parameter penting yang diatur untuk membangun model klasifikasi yang optimal. Tabel 5 berikut adalah rincian konfigurasi yang digunakan.

Tabel 5. Konfigurasi Model SVM

Parameter	Nilai
SVM Type	SVM
Cost (C)	1.00
Regression Loss Epsilon (e)	0.10
v-SVM	Tidak digunakan
Kernel	Radial Basis Function (RBF)
C (penalty)	1.00
Gamma	Auto
Numerical Tolerance	0.001
Iteration Limit	100

Model Support Vector Machine (SVM) yang digunakan dalam penelitian ini dirancang dengan konfigurasi yang bertujuan untuk mengoptimalkan kemampuan model dalam mengklasifikasikan ketidakhadiran karyawan dengan efektif. Dalam konfigurasi ini, model SVM standar dipilih karena kemampuannya dalam menangani masalah klasifikasi yang kompleks, seperti klasifikasi ketidakhadiran berdasarkan berbagai faktor.

Parameter Cost (C) diatur ke 1.00, yang mengatur trade-off antara margin yang lebih besar dan kesalahan klasifikasi. Nilai C yang lebih tinggi akan memaksa model untuk meminimalkan kesalahan klasifikasi, namun dapat meningkatkan risiko overfitting jika terlalu besar. Dengan nilai C yang moderat seperti 1.00, model ini berupaya menemukan keseimbangan antara kompleksitas model dan akurasi prediksi.

Parameter Regression Loss Epsilon (ϵ) diatur ke 0.10, yang bertujuan untuk mengatur margin kesalahan dalam klasifikasi. Meskipun nilai ini lebih umum digunakan dalam model regresi, dalam klasifikasi, nilai epsilon ini digunakan untuk memperkenalkan margin kesalahan yang memungkinkan model untuk lebih fleksibel dan tidak terlalu sensitif terhadap outlier. Kernel Radial Basis Function (RBF) dipilih untuk menangani data non-linear, yang umumnya ditemukan dalam masalah klasifikasi dunia nyata. Kernel RBF bekerja dengan cara memetakan data ke ruang berdimensi tinggi, sehingga dapat menemukan pemisah (hyperplane) yang lebih baik antara kelas yang berbeda. Gamma, yang diatur ke auto, memungkinkan sistem untuk secara otomatis menyesuaikan pengaruh data pada satu sama lain, yang membantu dalam menemukan pemisah yang optimal dalam ruang berdimensi tinggi.

Sementara itu, Numerical Tolerance diatur ke 0.001 untuk memastikan proses optimasi dilakukan dengan ketelitian tinggi. Ini memungkinkan model untuk mencapai konvergensi lebih akurat dalam jumlah iterasi yang lebih sedikit. Iteration Limit diatur ke 100, yang membatasi jumlah iterasi dalam proses pelatihan model untuk menghindari perhitungan yang berlarut-larut dan menghemat waktu.

Hasil Kinerja Model

Setelah model dilatih dan diuji, beberapa metrik kinerja digunakan untuk menilai efektivitas model dalam mengklasifikasikan ketidakhadiran karyawan. Tabel 6 berikut adalah hasil evaluasi model yang diperoleh.

Tabel 5. Hasil Kinerja Model

Metrik	Nilai
AUC	0.995
Accuracy (CA)	0.981
F1-Score (F1)	0.981
Precision (Prec)	0.981
Recall	0.982
MCC	0.820

Hasil evaluasi model menunjukkan bahwa Support Vector Machine (SVM) yang diterapkan dalam penelitian ini memiliki kinerja yang sangat baik dalam mengklasifikasikan ketidakhadiran karyawan. Salah satu metrik kinerja yang paling mencolok adalah AUC (Area Under Curve), yang mencapai 0.995. Nilai AUC yang sangat tinggi ini mengindikasikan bahwa model memiliki kemampuan yang luar biasa dalam membedakan antara ketidakhadiran dan tidak ketidakhadiran, dengan daya prediksi yang kuat. AUC yang mendekati angka 1 menunjukkan bahwa model dapat memisahkan kedua kelas dengan sangat baik, sehingga dapat diandalkan dalam aplikasi dunia nyata.

Selain itu, accuracy (CA) model mencapai 98.1%, yang menunjukkan bahwa sebagian besar prediksi yang dilakukan oleh model adalah benar. Dengan tingkat akurasi yang tinggi ini, model berhasil mengklasifikasikan ketidakhadiran karyawan dengan tingkat kesalahan yang sangat rendah, menandakan bahwa model memiliki kemampuan

klasifikasi yang sangat baik. Metrik lainnya, yaitu F1-Score, memperoleh nilai 0.981, yang mencerminkan keseimbangan yang sangat baik antara precision dan recall. Hal ini menunjukkan bahwa model ini tidak hanya dapat mendeteksi ketidakhadiran dengan akurat, tetapi juga menghindari kesalahan prediksi yang tidak diinginkan, dengan sedikit false positives maupun false negatives.

Dengan nilai precision sebesar 98.1%, model menunjukkan bahwa sebagian besar prediksi positif (ketidakhadiran) yang dilakukan adalah benar, yang berarti bahwa model sangat sedikit membuat kesalahan dalam memprediksi ketidakhadiran. Recall, yang mencapai 98.2%, menunjukkan bahwa model ini berhasil mendeteksi sebagian besar ketidakhadiran yang terjadi dalam data pengujian. Ini menunjukkan bahwa model memiliki kemampuan yang sangat baik untuk mendeteksi ketidakhadiran yang sebenarnya, meskipun masih ada beberapa kasus yang tidak terdeteksi.

Terakhir, Matthews Correlation Coefficient (MCC) yang diperoleh adalah 0.820, yang mengindikasikan adanya korelasi yang kuat antara hasil prediksi dan label sebenarnya. MCC adalah metrik yang memberikan gambaran menyeluruh tentang kualitas model, dan dengan nilai 0.820, dapat disimpulkan bahwa model ini memberikan kinerja yang sangat solid secara keseluruhan.

Hasil Analisis

Hasil dari konfigurasi dan kinerja model Support Vector Machine (SVM) yang diterapkan dalam penelitian ini menunjukkan bahwa model ini sangat efektif dalam mengklasifikasikan ketidakhadiran karyawan berdasarkan dataset yang tersedia. Evaluasi model melalui berbagai metrik memberikan indikasi yang kuat bahwa SVM merupakan alat yang tepat untuk masalah klasifikasi ini, dengan tingkat akurasi dan kemampuan diskriminasi yang sangat tinggi.

1. Keberhasilan Model dalam Mengklasifikasikan Ketidakhadiran

Salah satu metrik kinerja utama yang diperoleh adalah AUC (Area Under Curve) yang mencapai 0.995. AUC yang sangat tinggi ini menunjukkan bahwa model memiliki kemampuan diskriminasi yang sangat baik, yaitu kemampuannya untuk membedakan dengan jelas antara kelas ketidakhadiran dan tanpa ketidakhadiran. AUC yang mendekati angka 1 mengindikasikan bahwa model SVM dapat dengan mudah memisahkan kedua kelas tersebut, yang menjadikannya alat yang sangat andal dalam konteks klasifikasi ketidakhadiran. Sebagai tambahan, accuracy (CA) model mencapai 98.1%, yang berarti bahwa model ini mampu mengklasifikasikan ketidakhadiran karyawan dengan tingkat kesalahan yang sangat rendah. Sebagian besar prediksi yang dihasilkan oleh model ini adalah benar, menunjukkan kualitas prediksi yang sangat tinggi.

2. Keseimbangan antara Precision dan Recall

Nilai F1-Score yang tinggi (0.981) menunjukkan bahwa model ini berhasil menjaga keseimbangan yang sangat baik antara precision dan recall. Dalam konteks ini, precision yang tinggi (98.1%) berarti bahwa sebagian besar prediksi yang dilakukan oleh model untuk ketidakhadiran adalah benar, dan hanya sedikit prediksi yang salah. Recall yang tinggi

(98.2%) menunjukkan bahwa model ini sangat baik dalam menangkap ketidakhadiran yang sebenarnya, dengan sedikit ketidakhadiran yang terlewatkan atau tidak terdeteksi. Kombinasi precision dan recall yang tinggi ini adalah hasil yang sangat diinginkan dalam aplikasi nyata, di mana tujuan utama adalah untuk mendeteksi sebanyak mungkin ketidakhadiran tanpa menghasilkan banyak prediksi palsu. Ini sangat penting, terutama dalam konteks pengelolaan sumber daya manusia, di mana ketidakhadiran karyawan perlu diprediksi secara akurat untuk pengambilan keputusan yang lebih baik.

3. Interpretasi Matriks Kebingungannya

Dari Confusion Matrix yang dihasilkan, dapat dilihat bahwa model SVM berhasil memprediksi 692 dari 700 ketidakhadiran yang sebenarnya dengan benar, yang menunjukkan bahwa sebagian besar ketidakhadiran yang tercatat dalam data dapat diprediksi dengan akurat oleh model. Namun, model juga menghasilkan 6 kesalahan positif, yaitu kasus ketidakhadiran yang tidak terdeteksi, dan 8 kesalahan negatif, yaitu ketidakhadiran yang salah diprediksi sebagai tanpa ketidakhadiran. Meskipun ada beberapa kesalahan, angka-angka ini menunjukkan bahwa model SVM bekerja dengan sangat baik, dengan 99% ketidakhadiran terdeteksi dengan benar dan hanya sekitar 1% kesalahan prediksi. Sebagian besar kesalahan terjadi pada data yang lebih sedikit, menunjukkan bahwa kesalahan model lebih sering muncul pada kasus yang jarang atau mungkin ambigu dalam data.

4. Peluang Perbaikan dan Pengembangan Selanjutnya

Meskipun hasil yang diperoleh sangat memuaskan, ada beberapa area di mana model ini masih bisa ditingkatkan. Salah satu langkah untuk memperbaiki kinerja model adalah dengan melakukan tuning hyperparameter, seperti mencoba menggunakan kernel lain seperti linear kernel atau polynomial kernel, dan mengoptimalkan parameter C dan Gamma lebih lanjut. Eksplorasi dengan kernel yang berbeda mungkin dapat memberikan hasil yang lebih baik tergantung pada sifat data yang ada.

Selain itu, pengumpulan lebih banyak data juga dapat membantu meningkatkan kinerja model. Misalnya, faktor-faktor kesehatan mental dan fisik karyawan, yang dapat mempengaruhi ketidakhadiran, mungkin tidak sepenuhnya tercakup dalam dataset yang digunakan. Dengan menambahkan fitur tambahan yang berkaitan dengan kesejahteraan psikologis, stres kerja, atau kepuasan kerja, model dapat memperoleh pemahaman yang lebih komprehensif mengenai penyebab ketidakhadiran, yang pada akhirnya dapat meningkatkan akurasi dan kemampuan prediksi model.

Secara keseluruhan, meskipun model SVM sudah memberikan hasil yang sangat baik dengan akurasi tinggi dan kemampuan klasifikasi yang solid, ada ruang untuk perbaikan yang dapat dilakukan dengan meningkatkan pengumpulan data dan melakukan optimasi lebih lanjut pada model. Dengan langkah-langkah ini, diharapkan model dapat terus disempurnakan dan memberikan hasil yang lebih baik lagi dalam mengklasifikasikan ketidakhadiran karyawan.

Simpulan

Penelitian ini berhasil membangun model Support Vector Machine (SVM) yang efektif dalam mengklasifikasikan ketidakhadiran karyawan berdasarkan dataset yang mencakup berbagai faktor demografis, pekerjaan, dan kesehatan. Berdasarkan hasil evaluasi, model SVM menunjukkan kinerja yang sangat baik dengan AUC 0.995, yang mengindikasikan kemampuan diskriminasi yang luar biasa antara kelas ketidakhadiran dan tanpa ketidakhadiran. Dengan accuracy sebesar 98.1%, model ini mampu mengklasifikasikan ketidakhadiran dengan tingkat kesalahan yang sangat rendah, menunjukkan bahwa sebagian besar prediksi yang dihasilkan adalah benar. Selain itu, F1-Score yang mencapai 0.981 menunjukkan keseimbangan yang sangat baik antara precision dan recall, memungkinkan model mendeteksi ketidakhadiran dengan efektif sambil menghindari kesalahan prediksi yang tidak diinginkan. Meskipun model ini telah menunjukkan hasil yang sangat baik, beberapa kesalahan kecil dalam Confusion Matrix, seperti 6 kesalahan positif dan 8 kesalahan negatif, masih ada. Namun, angka-angka ini tetap mencerminkan kinerja yang sangat baik, dengan sebagian besar ketidakhadiran berhasil diprediksi dengan akurat. Selain itu, masih ada beberapa ruang untuk perbaikan, seperti melakukan tuning hyperparameter lebih lanjut dan menambahkan lebih banyak fitur yang dapat memperkaya pemahaman tentang faktor-faktor yang mempengaruhi ketidakhadiran. Secara keseluruhan, model SVM yang diterapkan dalam penelitian ini terbukti efektif dalam mengklasifikasikan ketidakhadiran karyawan dengan tingkat akurasi yang tinggi dan kemampuan untuk membedakan dengan jelas antara kedua kelas. Dengan tingkat prediksi yang sangat baik, model ini dapat digunakan dalam aplikasi praktis untuk membantu organisasi dalam mengelola ketidakhadiran karyawan, serta memberikan wawasan yang berguna bagi pengambilan keputusan yang lebih baik. Dengan beberapa perbaikan dan pengembangan lebih lanjut, model ini dapat terus ditingkatkan untuk mencapai hasil yang lebih optimal dalam memprediksi ketidakhadiran di masa depan.

Daftar Pustaka

- Abd-Ellah, M. K. (2016). Classification of brain tumor MRIs using a kernel support vector machine. *Communications in Computer and Information Science*, 636, 151-160, ISSN 1865-0929, https://doi.org/10.1007/978-3-319-44672-1_13
- Al-Hamadani, M. N. A. (2023). Classification and analysis of the MNIST dataset using PCA and SVM algorithms. *Vojnotehnički Glasnik*, 71(2), 221–238. <https://doi.org/10.5937/vojtehg71-42689>
- Bargam, B. (2024). Potential of Support Vector Machine Fed by ERA5 for Predicting Daily Discharge in the High Atlas of Morocco. *Advances in Science Technology and Innovation*, 79-82, ISSN 2522-8714, https://doi.org/10.1007/978-3-031-47079-0_18
- Basile, P. (2013). Super-sense tagging using support vector machines and distributional features. *Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, 7689, 176-185, ISSN 0302-9743, https://doi.org/10.1007/978-3-642-35828-9_19

- Cao, J. (2019). Adaptive Bad Pixel Correction Method for Interference-Modulated Images Based on Weighted Least Squares Support Vector Machines (WLS-SVM). *Applied Spectroscopy*, 73(4), 454-463, ISSN 0003-7028, <https://doi.org/10.1177/0003702819830776>
- Das, P. (2020). sigFeature: Novel Significant Feature Selection Method for Classification of Gene Expression Data Using Support Vector Machine and t Statistic. *Frontiers in Genetics*, 11, ISSN 1664-8021, <https://doi.org/10.3389/fgene.2020.00247>
- Du, K.-L., Jiang, B., Lu, J., Hua, J., & Swamy, M. N. S. (2024). Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions. *Mathematics*, 12(24), 3935. <https://doi.org/10.3390/math12243935>
- Jan, S.U. (2018). Performance Analysis of Support Vector Machine-Based Classifier for Spectrum Sensing in Cognitive Radio Networks. *Proceedings 2018 International Conference on Cyber Enabled Distributed Computing and Knowledge Discovery Cyberc 2018*, 385-389, <https://doi.org/10.1109/CyberC.2018.00075>
- Nath, G., Wang, Y., Coursey, A., Saha, K. K., Prabhu, S., & Sengupta, S. (2022). Incorporating a Machine Learning Model into a Web-Based Administrative Decision Support Tool for Predicting Workplace Absenteeism. *Information*, 13(7), 320. <https://doi.org/10.3390/info13070320>
- Nawata, K. (2024). Evaluation of physical and mental health conditions related to employees' absenteeism. *Frontiers in Public Health*. <https://doi.org/10.3389/fpubh.2023.1326334>
- Nuranisah (2020). Analysis of algorithm support vector machine learning and k-nearest neighbor in data accuracy. *Iop Conference Series Materials Science and Engineering*, 725(1), ISSN 1757-8981, <https://doi.org/10.1088/1757-899X/725/1/012118>
- Rainio, O., Teuvo, J., & Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Dental Science Reports*, 14. <https://doi.org/10.1038/s41598-024-56706-x>
- Rajath, J., & Pandita, D. (2022). Machine Learning as a Data Science Tool to Predict Absenteeism for Factory Workers. 261–265. <https://doi.org/10.1109/ICDABI56818.2022.10041623>
- Ratnayake, V., & Udawatta, S. (2021). Influencing Factors of Absenteeism of a Small Scale Garment Factory (Case Study). *Moratuwa Engineering Research Conference*. <https://doi.org/10.1109/MERCON52712.2021.9525775>
- Saji, P. (2022). An Efficient Method to Localize and Quantify Axial Displacement in Transformer Winding Using Support Vector Machines. *2022 IEEE Global Conference on Computing Power and Communication Technologies Globconpt 2022*, <https://doi.org/10.1109/GlobConPT57482.2022.9938221>
- Skorikov, M., Hussain, M., Khan, M. R., Akbar, M. K., Momen, S., Mohammed, N., & Nashin, T. (2020). Prediction of Absenteeism at Work using Data Mining Techniques. *International Conference on Information Technology*. <https://doi.org/10.1109/ICITR51448.2020.9310913>
- Solarz, A. (2017). Automated novelty detection in the WISE survey with one-class support vector machines. *Astronomy and Astrophysics*, 606, ISSN 0004-6361, <https://doi.org/10.1051/0004-6361/201730968>

-
- Traoré, A. (2024). Effect of Absenteeism on the Economic Performance of Small and Medium-sized Enterprises-SMEs. *International Journal of Scientific Research and Management*, 12(12), 8089–8096. <https://doi.org/10.18535/ijstrm/v12i12.em07>
- Tuppad, A., & Patil, S. D. (2023). Data Pre-processing Issues in Medical Data Classification. *Journal of Advanced Zoology*, 44(S6), 1079–1084. <https://doi.org/10.17762/jaz.v44is6.2361>
- Wang, M. (2014). Energy field filling of neic broadband radiated energy catalogue based on support vector machine regression model. *Applied Mechanics and Materials*, 687, 1514–1517, ISSN 1660-9336, <https://doi.org/10.4028/www.scientific.net/AMM.687-691.1514>