

Perbandingan Metode Seleksi Fitur untuk Mengoptimasi Model Support Vector Machine dalam Memprediksi Turnover Pegawai

Ahmad Syafiq Abiyyu¹, Kemas Muslim Lhaksana²

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

¹ahmadsyafiq@students.telkomuniversity.ac.id, ²kemasmuslim@telkomuniversity.ac.id

Abstrak

Seleksi fitur merupakan salah satu proses yang dilakukan untuk mengurangi dimensi data. Pengurangan dimensi bertujuan untuk meningkatkan performa model algoritma pembelajaran mesin. Turnover pegawai adalah suatu fenomena yang merujuk pada tingkat pegawai yang keluar dari suatu perusahaan. Penelitian mengenai implementasi algoritma pembelajaran mesin dalam memprediksi turnover pegawai sudah banyak dilakukan. Namun, performa model algoritma support vector machine (SVM) secara umum tidak menghasilkan performa yang baik. Dengan menggunakan metode seleksi fitur, hasil performa algoritma SVM diharapkan dapat menjadi lebih baik dalam memprediksi pegawai yang hendak melakukan turnover. Seleksi fitur digunakan pada dataset turnover pegawai sebelum dipelajari oleh model SVM yang dibangun. Metode seleksi fitur yang digunakan adalah filter methods, wrapper methods, dan embedded method. Penelitian ini menampilkan metode seleksi fitur mana yang paling baik dalam meningkatkan performa dari algoritma SVM. Matriks evaluasi seperti akurasi, recall, presisi, dan f1-score digunakan untuk menilai hasil akhir performan dari model SVM setelah dilakukan seleksi fitur. Hasil yang didapatkan adalah metode wrapper method meningkatkan performa dengan lebih baik dibandingkan metode lain. Nilai performa secara keseluruhan naik sebesar 4% dari performa sebelum dilakukan seleksi fitur.

Kata kunci : turnover pegawai, pembelajaran mesin, support vector machine, seleksi fitur

Abstract

Feature selection is one of the processes carried out to reduce the dimensions of data. Dimension reduction aims to improve the performance of machine learning algorithm models. Employee turnover is a phenomenon that refers to the level of employees leaving a company. Research on the implementation of machine learning algorithms in predicting employee turnover has been widely carried out. However, the performance of the support vector machine (SVM) algorithm model generally does not produce a good performance. By using the feature selection method, the results of the SVM algorithm's performance are expected to be better at predicting employees who want to make a turnover. Feature selection is used in the employee turnover dataset before being studied by the SVM model. The feature selection methods used are filter methods, wrapper methods, and embedded methods. This study shows which feature selection method is the best for improving the performance of the SVM algorithm. Evaluation matrices such as accuracy, recall, precision, and f1-score are used to assess the final performance of the SVM model after feature selection. The result obtained is that the wrapper method improves performance better than other methods. Overall performance value increased by 4% from performance before feature selection.

Keywords: employee turnover, machine learning, support vector machine, feature selection

1. Pendahuluan

Latar Belakang

Permasalahan yang dihadapi oleh perusahaan ketika berhadapan dengan karyawannya sendiri salah satunya adalah masalah tingkat perputaran (*turnover*) pegawai dalam perusahaan tersebut. Penting bagi perusahaan untuk dapat mengelola dan mengontrol tingkat perputaran pegawai dengan baik. Tingkat perputaran pegawai sangat berkaitan dengan tingkat rekrutasi suatu perusahaan. Semakin sering pegawai keluar dari perusahaan maka perusahaan juga harus semakin sering untuk melakukan rekrutasi pegawai baru [1]. Hal tersebut, selain dapat menyebabkan pengeluaran biaya yang tinggi, dapat menyebabkan suasana pekerjaan dalam perusahaan menjadi kurang menyenangkan.

Untuk mengatasi permasalahan tersebut, banyak penelitian yang telah membahas tentang bagaimana prediksi *turnover* pegawai menggunakan algoritma pembelajaran mesin. Penggunaan pembelajaran mesin ini dimaksudkan agar hasil analisis mengenai pegawai yang keluar dapat lebih akurat. Algoritma yang digunakan beraneka ragam contoh beberapa di antaranya adalah *decision tree*, *logistic regression*, *support vector machine*, *naïve bayes*, dan algoritma pembelajaran mesin modern lainnya. Dari beberapa hasil penelitian tersebut, beberapa algoritma memiliki hasil yang baik dan beberapa algoritma lain memiliki hasil yang kurang baik. Salah satu algoritma pembelajaran mesin yang memiliki hasil performa yang kurang baik adalah algoritma *support vector machine*.

Support vector machine (SVM) merupakan algoritma pembelajaran mesin yang menjadi fokus dalam penelitian ini. Berdasarkan hasil pada beberapa penelitian sebelumnya, SVM cenderung memiliki nilai performansi yang lebih buruk dari nilai performansi algoritma pembelajaran mesin lainnya [2]–[4]. Oleh karena itu, penelitian ini berfokus pada bagaimana memaksimalkan performa dari SVM. Salah satu cara yang digunakan adalah dengan melakukan seleksi fitur kepada dataset sebelum dipelajari oleh algoritma SVM.

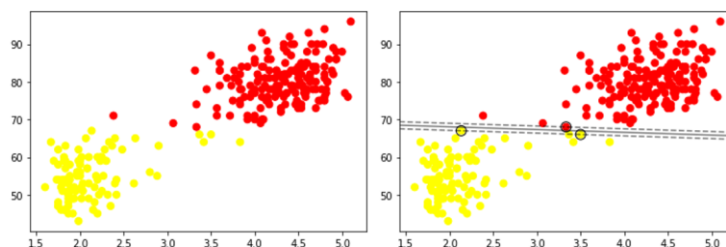
Feature selection atau seleksi fitur adalah suatu metode yang dapat digunakan untuk melakukan pengurangan dimensi terhadap dataset yang diberikan. Seleksi fitur dapat menyebabkan model yang dibangun lebih sederhana dan komprehensif, meningkatkan performa model, dan membantu untuk dapat memahami data [5]. Dalam kasus ini, seleksi fitur diharapkan dapat mengoptimasi hasil evaluasi dari model SVM yang telah dibangun. Sehingga, model SVM memiliki tingkat performansi yang lebih baik dari sebelumnya.

Topik dan Batasannya

Turnover pegawai, atau dalam bahasa Indonesia berarti perputaran pegawai, merujuk pada tingkat pergantian atau perputaran pegawai pada suatu perusahaan. Dalam perusahaan, pergantian pegawai merupakan hal yang lumrah terjadi dari waktu ke waktu. Namun, hal yang menjadi masalah adalah ketika pergantian pegawai terjadi secara terus menerus tanpa adanya evaluasi dari pihak perusahaan. Perusahaan tentu memiliki data-data terkait pegawai di perusahaan tersebut baik pegawai dengan status aktif atau pun pegawai dengan status sudah keluar dari perusahaan. Data-data pegawai dapat digunakan sebagai bahan evaluasi oleh pihak pengelola perusahaan, khususnya bagian pengelolaan sumber daya manusia, sehingga perusahaan dapat mengetahui indikasi yang membuat pegawai hendak keluar dari perusahaan.

Dalam teknologi pembelajaran mesin, terdapat berbagai macam metode yang dapat digunakan untuk mengklasifikasi suatu data ke dalam kelompok tertentu sesuai dengan topik data tersebut. Pada data terdapat satu fitur yang mengelompokkan data ke dalam klasifikasi yang diinginkan, fitur tersebut dapat disebut juga dengan variabel terikat. Variabel terikat tersebut dapat dijadikan rujukan dalam mengklasifikasi data ke dalam kelompok tertentu. Terdapat banyak metode pembelajaran mesin yang cukup terkenal untuk dapat mengatasi masalah klasifikasi data di antaranya adalah *naïve bayes*, *random forest*, *decision tree*, dan *support vector machine* (SVM). Untuk penelitian ini, metode pembelajaran mesin yang difokuskan adalah algoritma SVM.

Support vector machine (SVM) merupakan sebuah metode dalam pembelajaran mesin yang dapat digunakan untuk menyelesaikan permasalahan klasifikasi ataupun regresi. Algoritma ini termasuk dalam kategori *supervised learning*. *Supervised learning* berarti data yang akan dijadikan pembelajaran oleh algoritma ini telah memiliki label pada variabel terikat berdasarkan konteks data tersebut. SVM melakukan pendekatan klasifikasi dengan mencari *hyperplane* terbaik dengan jarak antar kelas yang paling optimal. *Hyperplane* adalah sebuah fungsi yang dapat dijadikan pemisah antar kelas dalam suatu data. Gambar 1.1 menunjukkan bagaimana garis *hyperplane* bekerja dalam mengklasifikasi dua jenis data yang berbeda.



Gambar 1. 1 Contoh implementasi SVM dengan hyperplane

Seleksi fitur digunakan sebagai metode dalam pengurangan dimensi dari data yang dipelajari oleh model SVM yang dibangun. Dataset dibagi ke dalam beberapa kategori yaitu dataset tanpa seleksi fitur dan dataset dengan seleksi fitur. Dataset dengan kategori seleksi fitur dibagi lagi ke dalam metode seleksi fitur yang digunakan. Hasil evaluasi dari tiap dataset dibandingkan untuk menilai seberapa efektif penggunaan metode seleksi fitur dalam mengoptimasi performa dari pembelajaran mesin.

Untuk menghindari terlalu banyak parameter yang digunakan, penerapan *hyperparameter tuning* tidak digunakan selama proses penelitian. Seleksi fitur dan model yang dibangun hanya menggunakan parameter bawaan yang sudah disediakan oleh *library*. Hal ini bertujuan agar fokus penelitian hanya pada perbandingan metode seleksi fitur mana yang paling baik dalam mengoptimasi model pembelajaran mesin.

Tujuan

Tujuan dari penelitian tugas akhir ini adalah untuk dapat membuat model sistem prediksi *turnover* pegawai dengan menggunakan algoritma *support vector machine*, melakukan implementasi metode seleksi fitur ke dalam dataset, dan menganalisis perbandingan hasil performa model algoritma *support vector machine* berdasarkan metode seleksi fitur yang digunakan.

Organisasi Tulisan

Penelitian ini disusun berdasarkan lima bab secara berturut-turut adalah Pendahuluan, Studi Terkait, Sistem yang Dibangun, Evaluasi, dan Kesimpulan. Studi Terkait menjelaskan tentang dasar-dasar pemahaman yang perlu diketahui mengenai penelitian yang dilakukan. Subbab Studi Terkait antara lain sebagai berikut: Penelitian Terkait, *Turnover Pegawai*, *Support Vector Machine*, dan *Feature Selection*. Sistem yang Dibangun menjelaskan tentang tahapan dalam membangun sistem dimulai dari dataset yang digunakan sampai implementasi pembelajaran mesin. Evaluasi menjelaskan tentang hasil dari sistem yang dibangun serta hasil analisis dari matrik evaluasi yang digunakan. Kesimpulan berisi rangkuman hasil dan kaitannya dengan tujuan penulisan serta saran untuk penelitian yang akan datang.

2. Studi Terkait

Penelitian Terkait

Penelitian terkait dengan prediksi turnover pegawai dapat dilihat pada Tabel 2.1. Hasil yang ditampilkan dalam Tabel 2.1 merupakan hasil dari penelitian untuk algoritma SVM.

Tabel 2. 1 Penelitian terkait prediksi turnover pegawai

No	Judul	Penulis/Tahun	Tujuan Penelitian	Hasil
1.	Employee Turnover Prediction with Machine Learning: A Reliable Approach [2]	Yue Zhao, Maciej K. Hryniewicki, Francesca Cheng, Boyang Fu, dan Xiaoyu Zhu/2019	Menyajikan deskripsi, demonstrasi, dan penilaian yang komprehensif terhadap pendekatan pembelajaran mesin untuk memprediksi <i>turnover</i> pegawai.	SVM termasuk dalam kategori metode pembelajaran mesin yang memiliki performa rendah. Nilai performa terendah yang diperoleh adalah 0,55.
2.	A Review On Employee Attrition Using Machine Learning [3]	Vishak Amin, JayantKumar A Rathod, Kshama, Mayuresh Kunder, Pathiksha Patkar/2021	Menyajikan struktur dalam memprediksi <i>turnover</i> pegawai menggunakan teknik klasifikasi	Dari tiga metode yang digunakan (<i>random forest</i> , <i>decision tree</i> , dan SVM), <i>random forest</i> memiliki hasil yang paling baik
3.	Prediction of Employee Turnover in Organizations using Machine Learning Algoritms [4]	Rohit Punnoose, Pankaj Ajit/2016	Mengeksplorasi pengaplikasian metode <i>extreme gradient boosting</i> sebagai peningkatan algoritma tradisional	SVM dengan menggunakan RBF kernel memiliki nilai akurasi 0,68, <i>run-time</i> 105 menit 30 detik, dan penggunaan kapasitas memori sebesar 12%

Penelitian terkait dengan implemenasi metode seleksi fitur dapat dilihat pada Tabel 2.2. Secara keseluruhan hasil evaluasi dengan menggunakan metode seleksi fitur mengalami peningkatan. Posisi penelitian ini dibandingkan dengan penelitian yang sudah ada adalah penelitian ini berfokus pada bagaimana meningkatkan performa untuk model SVM yang dibangun setelah dilakukan implementasi metode seleksi fitur.

Tabel 2. 2 Penelitian terkait implementasi metode feature selection

No	Judul	Penulis/Tahun	Metode	Hasil
1.	Implementation of the Naïve Bayes with Feature Selection using Genetic Algorithm for Sentiment Analysis of Fashion Online Companies [6]	Siti Ernawati, Eka Rini Yulia, Frieyadi, Samudi/2018	Genetic Algorithm	Hasil perbedaan untuk sebelum dan sesudah dilakukan seleksi fitur memiliki kenaikan yang cukup signifikan. Untuk akurasi memiliki nilai 68,5% menjadi 87,5% setelah seleksi fitur
2.	An Application of Machine Learning with Feature Selection to Improve Diagnosis and Classification of Neurodegenerative Disorders [7]	Josefa Diaz Alvarez, Jordi A. Matias-Guiu, Maria Nieves Cabrera-Martin, Jose L. Risco Martin dan Jose L.Ayala/2019	Cfs, Chi Squeard Attribute, Classifier Attribute, Wrapper Subset, Principal Component Analysis	Setelah fitur yang paling relevan diidentifikasi, jumlah data yang dapat diklasifikasi (benar) meningkat