

Studi perbandingan pemilihan fitur untuk support vector machine pada klasifikasi penilaian risiko kredit = Comparative study of feature selection for support vector machine in the classification of credit risk scoring

Silalahi, Desri Kristina, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20415003&lokasi=lokal>

Abstrak

[Penilaian kredit merupakan sistem atau cara yang digunakan oleh bank atau lembaga pembiayaan lainnya dalam menentukan calon debitur layak atau tidak mendapatkan pinjaman. Salah satu metode dalam penilaian kredit yang digunakan untuk mengklasifikasikan karakteristik calon debitur adalah Support Vector Machine (SVM). SVM mempunyai kemampuan generalisasi yang baik untuk menyelesaikan masalah klasifikasi dalam jumlah data yang besar dan dapat menghasilkan fungsi pemisah yang optimal untuk memisahkan dua kelompok data dari dua kelas yang berbeda. Salah satu keberhasilan menggunakan metode SVM adalah proses pemilihan fitur yang akan mempengaruhi tingkat akurasi klasifikasi. Berbagai metode dilakukan untuk pemilihan fitur, karena tidak semua fitur mampu memberikan hasil klasifikasi baik.

Pemilihan fitur yang digunakan dalam penelitian ini adalah Variance Threshold, Univariate Chi – Square, Recursive Feature Elimination (RFE) dan Extra Trees Classifier (ETC). Data dalam penelitian ini menggunakan data sekunder dari database dalam UCI machine learning repository. Berdasarkan simulasi untuk membandingkan nilai akurasi penggunaan metode pemilihan fitur pada SVM dalam klasifikasi penilaian risiko kredit, diperoleh bahwa metode Variance Threshold dan Univariate Chi – Square dapat mengurangi akurasi sedangkan metode RFE dan ETC dapat meningkatkan akurasi. Metode RFE memberikan akurasi yang lebih baik; Credit scoring is a system or method used by banks or other financial institutions to determine the debtor feasible or not get a loan. One of credit scoring method is used to classify the characteristics of debtor is Support Vector Machine (SVM). SVM has an excellent generalization ability to solve classification problems in a large amount of data and can generate an optimal separator function to separate two groups of data from two different classes. One of the success using SVM method is dependent on features selection process that will affect the level of classification accuracy.

Various methods have done to features selection, because not all the features are able to give best classification results. Features selection

that used this study is Variance Threshold, Univariate Chi - Square, Recursive Feature Elimination (RFE) and Extra Trees Classifier (ETC). Data in this study use secondary data from the database in UCI machine learning repository. Based on simulations to compare the accuracy of using feature selection method on SVM in classification of credit risk scoring, obtained that Variance Threshold and Univariate Chi – Square method can decrease accuracy while RFE and ETC method can increase accuracy. RFE method gives better accuracy., Credit scoring is a system or method used by banks or other financial institutions

to determine the debtor feasible or not get a loan. One of credit scoring method is used to classify the characteristics of debtor is Support Vector Machine (SVM). SVM has an excellent generalization ability to solve classification problems in a large amount of data and can generate an optimal separator function to separate two groups of data from two different classes. One of the success using SVM

method is dependent on features selection process that will affect the level of classification accuracy. Various methods have done to features selection, because not all the features are able to give best classification results. Features selection that used this study is Variance Threshold, Univariate Chi - Square, Recursive Feature Elimination (RFE) and Extra Trees Classifier (ETC). Data in this study use secondary data from the database in UCI machine learning repository. Based on simulations to compare the accuracy of using feature selection method on SVM in classification of credit risk scoring, obtained that Variance Threshold and Univariate Chi – Square method can decrease accuracy while RFE and ETC method can increase accuracy. RFE method gives better accuracy.]