

Analisis akurasi metode seleksi fitur gram-schmidt orthogonalization pada support vector machine untuk big data-studi kasus masalah prediksi klaim = Analyzing accuracy of feature selection method gram-schmidt orthogonalization for support vector machine on big data-a case study of claim prediction problem

Yuni Rosita Dewi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20493375&lokasi=lokal>

Abstrak

Prediksi klaim merupakan proses penting dalam industri asuransi karena perusahaan asuransi dapat menyiapkan jenis polis asuransi yang tepat untuk masing-masing pemegang polis potensial. Frekuensi prediksi klaim dewasa ini kian meningkat. Sehingga data prediksi klaim yang memiliki volume besar ini disebut *big data*, baik dari segi jumlah fitur maupun jumlah data pemegang polis. Salah satu alternatif solusi perusahaan asuransi untuk melihat pemegang polis melakukan klaim atau tidak, bisa menggunakan *machine learning* yang teruji dapat digunakan untuk klasifikasi dan prediksi. Salah satu metode *machine learning* untuk mengurangi jumlah fitur adalah dengan proses seleksi fitur, yaitu mencari urutan fitur berdasarkan tingkat pentingnya fitur. Metode seleksi fitur yang digunakan adalah *Gram-Schmidt Orthogonalization*. Metode ini sebelumnya digunakan untuk data tidak terstruktur namun pada penelitian ini diuji pada data terstruktur bervolume besar. Untuk menguji urutan fitur yang diperoleh dari proses seleksi fitur, digunakan *Support Vector Machine* karena termasuk metode *machine learning* yang populer untuk klasifikasi. Berdasarkan hasil simulasi, urutan yang diperoleh dari proses *Gram-Schmidt Orthogonalization* relatif konsisten. Selanjutnya, dapat diketahui fitur-fitur yang paling berpengaruh untuk menentukan pemegang polis klaim atau tidak. Simulasi juga menunjukkan bahwa hanya dengan menggunakan sekitar 26 % fitur, akurasi yang dihasilkan sebanding dengan menggunakan semua fitur.

.....Claim prediction is an important process in the insurance industry because insurance companies can prepare the right type of insurance policy for each potential policyholder. The frequency of today`s claim predictions is increasing. So that claim prediction data has a large volume called big data, both in terms of the number of features and the number of policyholders. One alternative solution for insurance companies to see whether policyholders claim or not, we can use machine learning that is proven to be used for classification and prediction. One of the machine learning methods to reduce the number of features is the feature selection process, which is to search for sequences of features based on their importance feature. The feature selection method used is Gram-Schmidt Orthogonalization. This method was previously used for unstructured data, but in this research is tested on large volume structured data. Support Vector Machine is used to test the ordered features obtained from the feature selection process because it is a popular machine learning method for classification. Based on a result, the ordered features obtained from the Gram-Schmidt Orthogonalization process is relatively stable. After that, it can also be seen the most important features to determine policyholders claim or not. The simulation also shows that using only about 26 % features, the resulting accuracy is comparable to using all features.