

INTISARI

PENGARUH PENUMPUKKAN *CONFIDENCE REGION* TERHADAP PERFORMA *MACHINE LEARNING* (SVM DAN ANN) PADA KLASIFIKASI TEH

Oleh

Fariz Harisuddin Dharmawan

20/455380/PA/19595

Dalam Statistika *confidence region* merupakan generalisasi multi-dimensi dari interval kepercayaan, ini adalah kumpulan titik dalam ruang n -dimensi. Pada penelitian sebelumnya dikatakan bahwa pada fitur PCA, terjadi penumpukkan *confidence region*. Pada proses klasifikasi jika *confidence region* bertumpuk antar kelasnya, maka dapat mempengaruhi performa sistem. Solusi dari permasalahan ini adalah melakukan analisis terhadap fitur dan *machine learning*. Dataset pada penelitian ini menggunakan raw data sensor teh dari pengambilan mesin enose. Tahap awal adalah melakukan variasi ekstraksi ciri untuk mereduksi informasi yang tidak diperlukan pada mesin pengenalan pola. Mengganti fitur menjadi LDA dengan tujuan mengelompokkan data ke dalam beberapa kelas. Selanjutnya penggunaan *machine learning* SVM, dengan tujuan sebagai metode pengklasifikasian dan ANN sebagai komparasi *machine learning* lainnya. Hasil dari penelitian ini adalah LDA mampu memisahkan *confidence region* secara visual score plot jika dibandingkan fitur PCA. Pengujian dilakukan dengan cara memvariasikan tiap ekstraksi ciri ke dalam fitur LDA dan SVM. Pada klasifikasi SVM akurasi terbaik mencapai nilai 92%, hal itu didapatkan dari ekstraksi ciri Integral dan Fractional. Namun saat dilakukan pengujian dengan gabungan ekstraksi ciri dan 2 ekstraksi ciri terbaik. Akurasi terbaik didapatkan oleh akurasi ANN dengan nilai 72.1% untuk gabungan ekstraksi ciri dan 64.9% untuk 2 ekstraksi ciri terbaik.

Kata kunci— fitur, ekstraksi ciri, *machine learning*

ABSTRACT

By

Fariz Harisuddin Dharmawan

20/455380/PA/19595

In statistics, a *confidence region* is a multi-dimensional generalization of a confidence interval, representing a collection of points in an n-dimensional space. In previous study, it is mentioned that there is an accumulation of *confidence regions* in PCA features. In the classification process, if *confidence regions* overlap among classes, it can influence the system's performance. The solution to this issue involves analyzing features and utilizing *machine learning*. The dataset in this research uses raw tea sensor data collected by an enose machine. The initial step involves various feature extraction methods to reduce unnecessary information in pattern recognition. The features are then transformed into LDA with the aim of grouping data into several classes. Subsequently, SVM is employed as a method for classification, and ANN is used as a comparison to other *machine learning* methods. The results of this study show that LDA can visually separate *confidence regions* in score plots when compared to PCA features. Testing is conducted by varying each feature extraction into LDA and SVM features. In SVM classification, the highest accuracy reaches 92%, obtained from the Integral and Fractional feature extractions. However, when testing with a combination of feature extractions and the two best feature extractions, the best accuracy is achieved by ANN with a value of 72.1% for combined feature extractions and 64.9% for the two best feature extractions.

Keywords: features, feature extraction, *machine learning*