

## INTISARI

### **Pengaruh Reduksi Fitur dengan *Principal Components Analysis* (PCA) terhadap performa *Automated Essay Scoring* (AES) Bahasa Indonesia**

Oleh

Andika Rahim Darusalam

16/394075/PA/17166

*Principal Components Analysis* (PCA) merupakan salah satu algoritma bertujuan untuk mengurangi banyak fitur pada dataset. Pada teks klasifikasi, fitur ter ekstraksi sangat banyak sehingga teks menjadi sulit di klasifikasikan. PCA dapat mengurangi fitur yang ada dengan cara mengdekomposisi ulang *Vector Space Model* (VSM) yang ada menjadi VSM dengan fitur yang lebih sedikit.

Penelitian ini mengusulkan PCA digunakan sebagai reduksi fitur dalam pembuatan model *classifier* pada pengklasifikasian teks jawaban berbahasa Indonesia dari sebuah pertanyaan. Penelitian ini melihat bagaimana performa dari model *classifier* yang dibangun ketika menggunakan PCA untuk mengurangi fitur pada datasetnya. *Classifier* yang digunakan dalam penelitian ini adalah *Decision Tree*, *Random Forest*, *Adaptive Boosting* (AdaBoost), *K-Nearest Neighbors*(KNN), dan *Support Vector Machine* (SVM). Performa dilihat dari sejauh mana PCA dapat meningkatkan *score* dari model *classifier* yang dibangun dibandingkan tanpa menggunakan PCA serta seberapa jauh peningkatan performa waktu eksekusinya.

Hasil Penelitian menunjukkan bahwa PCA tidak selalu mampu meningkatkan performa akurasi maupun *f1-score* secara signifikan dari model *classifier* yang dibangun walau secara waktu eksekusi terdapat peningkatan. Penurunan terjadi pada *classifier* yaitu *Decision Tree*. Penurunan paling signifikan terjadi pada dataset Jaket. Pada dataset Machu Pichu, AdaBoost dan *Random Forest* mengalami peningkatan tapi tidak signifikan. Peningkatan performa *f1-score* dengan PCA signifikan terjadi pada KNN. Peningkatan terjadi untuk semua dataset yaitu sekitar 3%. Performa waktu eksekusi juga meningkat signifikan untuk KNN.

**Kata kunci :** *PCA, Fitur Reduksi, Pembelejaran Mesin, Text Mining, Klasifikasi Teks*

## ABSTRACT

### The Effect of Feature Reduction with Principal Components Analysis (PCA) on the performance of Automated Essay Scoring (AES) Bahasa Indonesia

By

Andika Rahim Darusalam

Principal Components Analysis (PCA) is one algorithm that aims to reduce many features in the dataset. In text classification, features are extracted so much that the text becomes difficult to classify. PCA can reduce existing features by decomposing existing Vector Space Model (VSM) into VSM with fewer features.

This study proposes that PCA is used for feature reduction in making classifier models in classifying the answer text in Indonesian from a question. This study looks at how the performance of the classifier model that is built when using PCA to reduce features in the dataset. The classifiers used in this study are Decision Tree, Random Forest, Adaptive Boosting (AdaBoost), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Performance can be seen from the extent to which PCA can increase the score of the classifier model that is built compared without using PCA and how far the increase in execution time performance.

Research results show that PCA is not always able to significantly increase f1-score of the class model that was built even though the execution time has increased. The decrease occurred in the classifier namely Decision Tree. The most significant decrease occurred in the Jacket dataset. In the Machu Pichu dataset, AdaBoost and Random Forest have increased but not significantly. Accuracy and f1-scores increase significantly occurred in the KNN. The increase occurred for all datasets, around 3%. The execution time performance also improved for KNN.

**Keywords :** *PCA, Reduction Feature, Machine Learning, Text Mining, Text Classification*