

## **ABSTRACT**

### **INTEGRATION OF K-MEANS CLUSTERING AND NAIVE BAYES CLASSIFICATION ALGORITHM IN GROUPING THE NUMBER OF SCHOOLS, STUDENTS, AND TEACHERS IN WEST JAVA PROVINCE**

**By**

**SHAVIRA ZHALSABILLA**

Data Mining refers to the extraction process of implicit, previously unknown, and potentially useful information from data. However, as rapid development of technology, most of the data collection is hampered due to unidentified characteristics or label of the data. For having this sufficient labeled data need an enormous resources. The integration of clustering and classification techniques of this data mining implementation are proposed to overcome this problem. This clustering techniques provides advantages that helping to group the data that have similar characteristics, and also increase the accuracy of the classification model. In this study, the hybrid of K-means clustering and the Naive Bayes classification were implemented in grouping the number of schools, students, and teachers in West Java Province 2016-2020. This study aims to group each city/regency in West Java Province into clusters using the K-means clustering, then classify them using the Naive Bayes classification. Based on the results, the optimal number of clusters is 4 clusters with highest accuracy of 97.50% and an error rate of 0.025.

**Keywords:** Data mining, Clustering, K-means, Classification, Naive Bayes

## **ABSTRAK**

### **INTEGRASI ALGORITMA KLUSTERISASI K-MEANS DAN KLASIFIKASI NAIVE BAYES PADA PENGELOMPOKKAN JUMLAH SEKOLAH, MURID, DAN GURU DI PROVINSI JAWA BARAT**

**Oleh**

**SHAVIRA ZHALSABILLA**

*Data Mining* adalah proses mengekstraksi pola dan informasi yang berguna dari sebuah kumpulan data. Namun, seiring berkembangnya teknologi, sebagian besar pengumpulan data terhambat karena ketidaklengkapan label yang ada dalam data. Untuk mendapatkan label ini dibutuhkan sumber daya yang besar, sehingga untuk mengatasinya diterapkan penggabungan atau integrasi teknik klusterisasi dan klasifikasi pada *data mining*. Klusterisasi ini memberikan keuntungan atas teknik klasifikasi yaitu membantu mengidentifikasi kelompok data yang memiliki kesamaan karakteristik, dan juga meningkatkan akurasi pada model klasifikasi. Pada penelitian ini dilakukan penggabungan algoritma klusterisasi K-means dan klasifikasi Naive Bayes pada pengelompokan jumlah sekolah, murid, dan guru di Provinsi Jawa Barat tahun 2016-2020. Penelitian ini bertujuan untuk mengelompokkan kota/kabupaten di Provinsi Jawa Barat menjadi beberapa kluster menggunakan algoritma klusterisasi K-means dan mengklasifikasikannya menggunakan algoritma klasifikasi Naive Bayes. Berdasarkan hasil pengujian yang dilakukan, didapat jumlah kluster yang optimal sebanyak 4 kluster dan menghasilkan akurasi tertinggi sebesar 97.50% dan *error rate* 0.025.

**Kata Kunci:** *Data mining*, Klusterisasi, K-means, Klasifikasi, Naive Bayes