ABSTRACT

Indonesia is a country that has ethnic and cultural diversity. Each ethnic group in Indonesia has its own characteristics in terms of language, dialect or accent. Thus, the human voice has its own characteristics depending on the ethnicity of the area seen from the accent. In this study, a system that can identify a person's area of origin based on his accent has been made by using Linear Predictive Coding (LPC) feature extraction and K-Nearest Neighbor (K-NN) classification. This study used 160 training data and 80 replication data which were grouped into 4 classes of regional origin, namely West Sumatra, North Sumatra, West Java, and Central Java.

Voice data in the form of training data and test data are stored in .wav format which will then be analyzed by the LPC. After that, the K-NN classification method is used which will be tested with objects based on learning data that has the closest distance to the object. The choice of the K-NN classification method is because the K-NN method has good and efficient clustering capabilities.

The highest accuracy obtained in this study is 87.5% using the maximum matrix length (P) 12 and the cepstral window value (Q) 14 using the Euclidean and Minkowski distance rules on K-NN.

Keywords: Human Accent, Linear Predictive Coding, K-Nearest Neighbor.