

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan data kualitas air tambak menggunakan metode *Gaussian Mixture Model* (GMM) dengan algoritma *Expectation-Maximization* (EM). Data yang digunakan terdiri dari beberapa parameter penting, yaitu pH, suhu, kekeruhan, dan *total dissolved solids* (TDS), yang diperoleh dari hasil monitoring air tambak. Sebelum dilakukan proses klusterisasi, data terlebih dahulu diproses melalui tahap normalisasi menggunakan metode *Min-Max Scaling* untuk menyamakan skala antar fitur. GMM digunakan sebagai metode klusterisasi karena memiliki kemampuan memodelkan data dengan distribusi kompleks serta dapat memberikan nilai probabilitas keanggotaan tiap data terhadap suatu kluster. Evaluasi jumlah kluster dilakukan menggunakan *Bayesian Information Criterion* (BIC), sedangkan evaluasi kualitas hasil kluster dilakukan menggunakan *Silhouette Score* dan *Davies-Bouldin Index*. Hasil dari proses klusterisasi divisualisasikan menggunakan PCA (*Principal Component Analysis*). Sistem ini dibangun menggunakan Google Colab untuk eksperimen dan perhitungan, serta diimplementasikan dalam bentuk antarmuka web menggunakan Streamlit. Sistem mampu mengelompokkan data kualitas air tambak menjadi tiga kluster utama yang menggambarkan kategori kualitas air: baik (67 data), sedang (37 data), dan buruk (11 data). Hasil penelitian ini diharapkan dapat membantu pihak pengelola tambak dalam melakukan pengawasan dan pengambilan keputusan berbasis data.

Kata Kunci: *Gaussian Mixture Model* (GMM), kualitas air tambak, klusterisasi, *Expectation-Maximization*, BIC, *Silhouette Score*, Streamlit.

ABSTRACT

This study aims to cluster pond water quality data using the Gaussian Mixture Model (GMM) method with the Expectation-Maximization (EM) algorithm. The data used consists of several important parameters, namely pH, temperature, turbidity, and total dissolved solids (TDS), which are obtained from pond water monitoring results. Before the clustering process is carried out, the data is first processed through a normalization stage using the Min-Max Scaling method to equalize the scale between features. GMM is used as a clustering method because it has the ability to model data with complex distributions and can provide membership probability values for each data to a cluster. Evaluation of the number of clusters is carried out using the Bayesian Information Criterion (BIC), while evaluation of the quality of cluster results is carried out using the Silhouette Score and Davies-Bouldin Index. The results of the clustering process are visualized using PCA (Principal Component Analysis). This system is built using Google Colab for experiments and calculations, and implemented in the form of a web interface using Streamlit. The system was able to group pond water quality data into three main clusters, representing water quality categories: good (67 data points), moderate (37 data points), and poor (11 data points). The results of this study are expected to assist pond managers in monitoring and making data-driven decisions.

Keywords: *Gaussian Mixture Model (GMM), pond water quality, clustering, Expectation-Maximization, BIC, Silhouette Score, Streamlit.*