ABSTRAK

IMPLEMENTASI *CLUSTERING* K-MEANS++ PADA DATA PRODUK NG (*NO GOOD*) PT NSK BEARINGS MANUFACTURING INDONESIA

Naufal Althafi Handoyo
H1D021087

Permasalahan produk cacat masih menjadi tantangan signifikan dalam proses produksi di PT NSK Bearings Manufacturing Indonesia. Tingginya jumlah produk cacat berdampak pada peningkatan biaya produksi dan penurunan efisiensi operasional. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi pola kecacatan produk dengan menerapkan algoritma *K-Means++* sebagai metode *clustering* guna mengelompokkan data berdasarkan pola kecacatan yang serupa. Penelitian ini menggunakan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*). Data yang digunakan berupa 4.344 *dataset* hasil produksi dari Januari hingga Juni 2025, dengan variabel tanggal produksi, tipe *bearing*, *line* produksi, total produksi, jumlah kecacatan, jenis kecacatan, dan proses produksi. Proses *clustering* dilakukan menggunakan *Elbow Method* dan *Silhouette Score* untuk menentukan jumlah klaster optimal. Hasil terbaik diperoleh pada k = 3 dengan nilai *Silhouette Score* 0.3859. Klaster yang terbentuk dikategorikan sebagai Cacat Rendah, Cacat Sedang, dan Cacat Tinggi. Implementasi sistem menggunakan *framework* Streamlit memudahkan pihak perusahaan dalam melakukan analisis data, visualisasi hasil *clustering*, serta pengambilan keputusan berbasis data secara interaktif.

Kata kunci: *Clustering,* CRISP-DM, K-Means++, Produk Cacat, Streamlit.

ABSTRACT

IMPLEMENTATION OF K-MEANS++ CLUSTERING ON NG (NO GOOD) PRODUCT DATA AT PT NSK BEARINGS MANUFACTURING INDONESIA

Naufal Althafi Handoyo
H1D021087

Defective products remain a significant challenge in the production process at PT NSK Bearings Manufacturing Indonesia. The high number of defective products leads to increased production costs and decreased operational efficiency. Therefore, this study aims to identify patterns of product defects by applying the K-Means++ algorithm as a clustering method to group data based on similar defect patterns. The study adopts the CRISP-DM (Cross Industry Standard Process for Data Mining) approach. The dataset used consists of 4,344 production records from January to June 2025, including variables such as production date, bearing type, production line, total production, number of defects, defect type, and production process. The clustering process was carried out using the Elbow Method and Silhouette Score to determine the optimal number of clusters. The best result was obtained at k = 3 with a Silhouette Score value of 0.3859. The resulting clusters were categorized as Low Defect, Medium Defect, and High Defect. The system implementation using the Streamlit framework facilitates data analysis, visualization of clustering results, and interactive data-driven decision-making for the company.

Keywords: *Clustering, CRISP-DM, K-Means++, Defective Product, Streamlit.*