

ABSTRAK

Kecelakaan lalu lintas yang melibatkan kendaraan angkutan barang di Kabupaten Semarang masih menjadi permasalahan penting karena tingginya frekuensi kejadian dan dampaknya. Penelitian ini bertujuan memetakan pola kecelakaan kendaraan angkutan barang serta membangun model klasifikasi berdasarkan karakteristik kecelakaan yang terbentuk. Penelitian menggunakan 858 data kecelakaan kendaraan angkutan barang dari Satlantas Polres Semarang periode 2021–Agustus 2025. Analisis dilakukan menggunakan framework CRISP-DM. Proses clustering menerapkan metode Hierarchical Clustering, K-Means, dan K-Medoids dengan evaluasi Davies-Bouldin Index (DBI), sedangkan klasifikasi menggunakan algoritma Naïve Bayes, Decision Tree, dan K-Nearest Neighbor (KNN). Hasil penelitian menunjukkan bahwa Hierarchical Clustering merupakan metode terbaik dengan nilai DBI sebesar 0,111 dan menghasilkan dua kelompok kecelakaan. Cluster 0 merepresentasikan pola kecelakaan berisiko tinggi yang ditandai dominasi pengemudi tanpa SIM, kejadian pada jalan tanpa median, tingginya keterlibatan sepeda motor, serta jumlah fatalitas yang lebih tinggi dan banyak ditemukan di Kecamatan Tenganan, Ungaran Barat, dan Jambu. Sementara itu, Cluster 1 merepresentasikan pola kecelakaan dengan tingkat fatalitas yang relatif lebih rendah, didominasi pengemudi dengan SIM sesuai, lebih banyak terjadi pada jalan bermedian, serta banyak ditemukan di Kecamatan Bawen, Bergas, dan Pringapus. Pada tahap validasi, algoritma Naïve Bayes dengan rasio data 70:30 menghasilkan akurasi 87,25%, recall 86,63%, dan precision 87,07%.

Kata kunci: kecelakaan lalu lintas, angkutan barang, data mining, clustering, klasifikasi.

ABSTRACT

Traffic accidents involving trucks in Semarang Regency remain a significant issue due to their high frequency and severe impacts. This study aims to identify accident patterns involving trucks and develop a classification model based on the resulting accident characteristics. The study utilized 858 truck accident records obtained from the Semarang Regency Traffic Police covering the period from 2021 to August 2025. Data analysis was conducted using the CRISP-DM framework. The clustering stage applied Hierarchical Clustering, K-Means, and K-Medoids, which were evaluated using the Davies-Bouldin Index (DBI), while the classification stage employed Naïve Bayes, Decision Tree, and K-Nearest Neighbor (KNN) algorithms. The results indicate that Hierarchical Clustering achieved the best performance with a DBI value of 0.111 and produced two accident clusters. Cluster 0 represents a high-risk accident pattern characterized by the predominance of unlicensed drivers, occurrences on roads without medians, a high involvement of motorcycles, and a greater number of fatalities. This cluster was primarily found in Tenganan, West Ungaran, and Jambu districts. In contrast, Cluster 1 represents accidents with relatively lower fatality levels, dominated by licensed drivers, occurring more frequently on roads with medians, and primarily found in Bawen, Bergas, and Pringapus districts. In the validation stage, the Naïve Bayes algorithm with a 70:30 data split ratio achieved the best performance, yielding an accuracy of 87.25%, a recall of 86.63%, and a precision of 87.07%.

Keywords: traffic accidents, trucks, data mining, clustering, classification.