

**VEHICULAR ACCIDENTS DATA ANALYSIS
USING K-MEANS CLUSTERING ALGORITHM IN
CASTILLEJOS, ZAMBALES**

**A Thesis Presented to the Faculty of the
College of Communication and Information Technology
President Ramon Magsaysay State University
Castillejos, Zambales**

**In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science**

By:

**Cortez, Jhon Rexter V.
Lurzano, Jessica V.
Manila, Jhon Lloyd**

June 2023

Republic of the Philippines
PRESIDENT RAMON MAGSAYSAY STATE UNIVERSITY
 (Formerly Ramon Magsaysay Technological University)
 Castillejos Campus
 Castillejos, Zambales



APPROVAL SHEET

The thesis project entitled “**Vehicular Accidents Data Analysis using K- Means Clustering Algorithm in Castillejos, Zambales**” was prepared and submitted by **Cortez, Jhon Rexter V., Lurzano, Jessica and Manila, Jhon Lloyd** in partial fulfillment of the course requirements for the degree of **Bachelor of Science in Computer Science** has been examined and recommended for the oral examination.

[Signature]
IRATUS GLENN A. CRUZ, MSCS
 Thesis Adviser

Approved by the Panel of Examiners
 with a rating of 93 %

[Signature]
MICHAEL G. ALBINO, MIT
 Chair

[Signature]
MICHAEL N. FARIN, MSCS
 Member

[Signature]
MARIE CELIA R. AGLIBOT, MSCS
 Member

Accepted and approved as a requirement for the degree of
BACHELOR OF SCIENCE IN COMPUTER SCIENCE.

[Signature]
IVY H. CASUPANAN, Ed.D
 Campus Director

June 2023

ABSTRACT

This study focuses on the analysis of vehicular accidents in Castillejos, Zambales, using the K-means clustering algorithm. The objectives of the study were to gather accident data, utilize the K-means clustering algorithm for data analysis, identify common causes and high-risk locations, and improve road safety through data analytics.

Secondary data analysis or archival study was used as the research method. The proponents employed secondary data analysis and the k-means clustering algorithm allowing researchers to gain a deeper understanding of vehicular accident patterns, identify risk factors, and propose evidence-based strategies for prevention and intervention. However, it's important to acknowledge the limitations of the approach, such as the potential for biases in the original data collection process and the assumptions underlying the clustering algorithm. These limitations should be considered when interpreting and applying the results of the analysis.

The study also employed SPSS for data analysis, utilizing variables such as accident time, exact location, vehicle classes involved, weather condition, road condition, traffic status, type of accident, the influence of substance among individuals, indicated cause of accident, and areas in barangay.

The findings revealed that accidents were most likely to occur between 11:00 AM to 1:00 PM within each cluster. Specific areas, including along Olongapo – Bugallon National Road near KM5 Commercial Building, along Olongapo – Bugallon National Road near Precious Child Montessori of Zambales, Inc, and along Olongapo - Bugallon National Road near Camachile Bridge, were identified as high-risk locations requiring

targeted interventions. The algorithm was used to examine the parameters within each cluster to identify the factors that contribute to the occurrence of accidents. These are the remaining findings using the clustering algorithm. Certain vehicle combinations, such as tricycles and passenger buses, motorcycles, passenger vans, and passenger jeepneys, were more commonly involved in accidents, necessitating tailored safety measures. Accidents were more likely during light to moderate rain conditions with heavy winds, emphasizing the need for weather-responsive measures.

The study also identified factors contributing to accidents within each cluster, including light traffic, rear-end collisions, absence of substance influence, poor road conditions, and mechanical failure. Brgy. Magsaysay, Brgy. Nagbunga, and Brgy. San Jose were identified as high-risk areas requiring focused interventions. The study recommends implementing targeted measures, enhancing safety measures for specific vehicles, weather-responsive measures, road maintenance and infrastructure improvements, addressing specific causes of accidents, focusing on high-risk areas, fostering collaboration among stakeholders, and continuous monitoring and evaluation.

Keywords: Vehicular Accidents, K-means Clustering Algorithm, Road Safety

CHAPTER 2: FRAMEWORK OF THE STUDY

2.1 Introduction on Vehicular Accidents

2.2 Application and Usage of K-Means Clustering Algorithm in Analysis

2.3 K-Means Clustering Algorithm