

[Home](#) > [Smart Technologies, Systems and Applications](#) > Conference paper

Analysis of Hidden Patterns in Road Accident Dataset Using Clustering Techniques

| Conference paper | First Online: 03 September 2025

| pp 96–111 | [Cite this conference paper](#)



[Smart Technologies, Systems and Applications](#) (SmartTech-IC 2024)

[Henry O. Velesaca](#) , [Miguel Realpe](#), [Angel D. Sappa](#) & [Alice Gomez](#)

 Part of the book series: [Communications in Computer and Information Science](#)
((CCIS, volume 2393))

 Included in the following conference series:
[International Conference on Smart Technologies, Systems and Applications](#)

 73 Accesses

Abstract

This study presents the application of various clustering techniques for the analysis of a traffic accident dataset. Methods such as Gaussian Mixture Model, Agglomerative Clustering, MiniBatchK-Means and K-Means are used to identify hidden patterns in the

data, including features such as geographic coordinates, accident severity, cause and type of accident. The dataset is preprocessed by removing null values, encoding categorical variables, and robust scaling of features. Furthermore, PCA is applied to reduce the dimensionality of the dataset. The performance of the clustering techniques is evaluated using metrics such as the Silhouette Score, Davies–Bouldin Score and Calinski–Harabasz Score. The results indicate that K–Means, with 5 clusters, provides the best overall performance, according to the Elbow method and the evaluated metrics. Visualizations of 2D is included for a better interpretation of the clusters, highlighting the distribution and features of the groups formed. The code and dataset are available at Kaggle:

<https://www.kaggle.com/code/hvelesaca/smarttech-paper-id-42>, facilitating further research.

 This is a preview of subscription content, [log in via an institution](#)  to check access.

Access this chapter

Log in via an institution

Subscribe and save

✓ Springer+

from \$39.99 /Month

Starting from 10 chapters or articles per month

Access and download chapters and articles from more than 300k books and 2,500 journals

Cancel anytime

[View plans](#) →

Buy Now

^ Chapter

USD 29.95

Price excludes VAT (Ecuador)

Available as PDF

Read on any device

Instant download

Own it forever

Buy Chapter

▼ eBook

USD 109.00

▼ Softcover Book

USD 139.99

Tax calculation will be finalised at checkout

Purchases are for personal use only

[Institutional subscriptions](#) →

Similar content being viewed by others



**Extracting Hidden
Patterns Within Road
Accident Data Using
Machine Learning...**

Chapter | © 2018



**Unsupervised
Clustering of Asphalt
Pavement Conditions
Using Fuzzy C-Means...**

Chapter | © 2024



**The Real Sorting Hat –
Identifying Driving and
Scanning Strategies in
Urban Intersections...**

Chapter | © 2024

Explore related subjects

Discover the latest articles, books and news in related subjects, suggested using machine learning.

[Applied Statistics](#)

[Data Mining](#)

[Functional clustering](#)

[Multivariate Analysis](#)

[Sociological Methods](#)

[Programming Techniques](#)

Notes

1. <https://www.ant.gob.ec/visor-de-siniestralidad-estadisticas/>.
2. <https://www.ant.gob.ec/visor-de-siniestralidad-estadisticas/>.

References

1. Acosta Rodríguez, M.V.: Los peatones y los accidentes de tránsito, en el Cantón Ambato. B.S. thesis (2015)

[Google Scholar](#)

2. Amancha Moyolema, J.C.: El consumo de alcohol en los conductores y los accidentes de tránsito en la ciudad de Ambato. B.S. thesis (2015)

[Google Scholar](#)

3. Banerjee, K., et al.: Traffic accident risk prediction using machine learning. In: 2022 International Mobile and Embedded Technology Conference (MECON), pp. 76–82. IEEE (2022)

4. Bokaba, T., Doorsamy, W., Paul, B.S.: Comparative study of machine learning classifiers for modelling road traffic accidents. *Appl. Sci.* **12**(2), 828 (2022)

[Article](#) [Google Scholar](#)

5. Comi, A., Polimeni, A., Balsamo, C.: Road accident analysis with data mining approach: evidence from Rome. *Transp. Res. Procedia* **62**, 798–805 (2022)

[Article](#) [Google Scholar](#)

6. Doğru, N., SUBAŞI, A.: Comparison of clustering techniques for traffic accident detection. *Turk. J. Electr. Eng. Comput. Sci.* **23**(Sup 1), 2124–2137 (2015)

[Google Scholar](#)

7. Esenturk, E., Turley, D., Wallace, A., Khastgir, S., Jennings, P.: A data mining approach for traffic accidents, pattern extraction and test scenario generation for autonomous vehicles. *Int. J. Transp. Sci. Technol.* **12**(4), 955–972 (2023)

[Article](#) [Google Scholar](#)

8. Esenturk, E., Wallace, A.G., Khastgir, S., Jennings, P.: Identification of traffic accident patterns via cluster analysis and test scenario development for autonomous vehicles. *IEEE Access* **10**, 6660–6675 (2022)

[Article](#) [Google Scholar](#)

9. Franquesa Gómez, P.: Determining traffic accident patterns through clustering and direct count. B.S. thesis, Universitat Politècnica de Catalunya (2024)

[Google Scholar](#)

10. Fujii, Y., Shiobara, R.: The analysis of traffic accidents. *J. Navig.* **24**(4), 534–543 (1971)

[Article](#) [Google Scholar](#)

11. Gaibor Vásquez, J.E.: Los accidentes de tránsito y estado de salud de uno a tres días. B.S. thesis (2020)

[Google Scholar](#)

12. Islam, M.R., Jenny, I.J., Nayon, M., Islam, M.R., Amiruzzaman, M., Abdullah-Al-Wadud, M.: Clustering algorithms to analyze the road traffic crashes. In: 2021 International Conference on Science & Contemporary Technologies (ICSCT), pp. 1–6. IEEE (2021)

[Google Scholar](#)

13. Kumeda, B., Zhang, F., Zhou, F., Hussain, S., Almasri, A., Assefa, M.: Classification of road traffic accident data using machine learning algorithms. In: 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN), pp. 682–687. IEEE (2019)

[Google Scholar](#)

14. Mohammed, A.A., Ambak, K., Mosa, A.M., Syamsunur, D.: A review of traffic accidents and related practices worldwide. *Open Transp. J.* **13**(1) (2019)

[Google Scholar](#)

15. Nandurge, P.A., Dharwadkar, N.V.: Analyzing road accident data using machine learning paradigms. In: 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), pp. 604–610. IEEE (2017)

[Google Scholar](#)

16. Santos, D., Saias, J., Quaresma, P., Nogueira, V.B.: Machine learning approaches to traffic accident analysis and hotspot prediction. *Computers* **10**(12), 157 (2021)

[Article](#) [Google Scholar](#)

17. Taamneh, M., Taamneh, S., Alkheder, S.: Clustering-based classification of road traffic accidents using hierarchical clustering and artificial neural networks. *Int. J. Inj. Contr. Saf. Promot.* **24**(3), 388–395 (2017)

[Article](#) [Google Scholar](#)

18. Tippayanate, N., Impool, T., Sujayanont, P., Muttitanon, W., Chemin, Y., Som-Ard, J., et al.: Spatial distribution and cluster analysis of road traffic accidents in Khon Kaen municipality, Thailand. *Int. J. Geoinform.* **20**(4), 43–55 (2024)

[Google Scholar](#)

19. Velesaca, H.O., Araujo, S., Suárez, P.L., Sánchez, A., Sappa, A.D.: Off-the-shelf based system for urban environment video analytics. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 459–464. IEEE (2020)

[Google Scholar](#)

20. Yadav, J., Batra, K., Goel, A., et al.: A framework for analyzing road accidents using machine learning paradigms. In: *Journal of Physics: Conference Series*, vol. 1950, p. 012072. IOP Publishing (2021)

[Google Scholar](#)

Acknowledgments

This research has been partially supported by the ESPOL and University of Granada.

Author information

Authors and Affiliations

ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
Henry O. Velesaca, Miguel Realpe, Angel D. Sappa & Alice Gomez

Computer Vision Center, 08193, Bellaterra, Barcelona, Spain
Angel D. Sappa

Software Engineering Department, University of Granada, 18014, Granada, Spain
Henry O. Velesaca

Corresponding author

Correspondence to [Henry O. Velesaca](#).

Editor information

Editors and Affiliations

Universidad Politécnica Salesiana, Quito, Ecuador
Fabián R. Narváez

Universidad Politécnica Salesiana, Quito, Ecuador
Micaela N. Villa

Instituto Tecnológico Metropolitano, Medellín, Colombia
Gloria M. Díaz

Rights and permissions

[Reprints and permissions](#)

Copyright information

© 2026 The Author(s), under exclusive license to Springer Nature Switzerland AG

About this paper

Cite this paper

Velesaca, H.O., Realpe, M., Sappa, A.D., Gomez, A. (2026). Analysis of Hidden Patterns in Road Accident Dataset Using Clustering Techniques. In: Narváez, F.R., Villa, M.N., Díaz, G.M. (eds) Smart Technologies, Systems and Applications. SmartTech-IC 2024. Communications in Computer and Information Science, vol 2393. Springer, Cham. https://doi.org/10.1007/978-3-031-98290-3_8

[.RIS](#) [.ENW](#) [.BIB](#)

DOI	Published	Publisher Name
https://doi.org/10.1007/978-3-031-98290-3_8	03 September 2025	Springer, Cham
Print ISBN	Online ISBN	eBook Packages
978-3-031-98289-7	978-3-031-98290-3	Computer Science
		Computer Science (R0)

Keywords

[Machine learning](#) [Clustering](#) [MiniBatch K-Means](#) [Traffic accidents](#)

[Unsupervised technique](#)

Publish with us

[Policies and ethics](#) [🔗](#)