



PROCEEDINGS OF THE

3RD INTERNATIONAL CONFERENCE IN DATA SCIENCE

Autonomous AI Agents: Redefining Decision-Making in a Data-Driven World

25th and 26th of November 2025 at the University of Colombo, Sri Lanka

ICDS 2025



Organised By:



Department of Statistics
University of Colombo

Universiti Malaya
Malaysia

Silver Partner

Knowledge Partner





**Proceedings
of the
International Conference
in Data Science 2025
(ICDS 2025)**

25th - 26th November 2025
at the
**Department of Statistics, University of Colombo,
Colombo, Sri Lanka**

**ICDS 2025 is organised by the Center for Data Science jointly with
the
Department of Statistics, University of Colombo
and
Universiti Malaya, Malaysia**

Publisher: Center for Data Science, University of Colombo
ISSN: 3030-7163

Segmenting Graduate Unemployment: A clustering approach to understanding employment challenges in Sri Lanka

D G S P Thilakarathne¹, R R L U I Rajapaksha^{1*}

¹ Faculty of Computing and Technology, University of Kelaniya, Kelaniya, Sri Lanka
*rasikar@kln.ac.lk

This study investigates the demographic and socioeconomic factors influencing graduate unemployment in Sri Lanka through an unsupervised machine learning approach. Data from a structured national survey of 1,250 unemployed graduates collected between January 2022 and December 2023 were analyzed using Principal Component Analysis (PCA) and three clustering algorithms K-means, Hierarchical clustering (Ward's linkage), and Gaussian Mixture Models (GMM) to segment the graduate population into meaningful subgroups. After dimensionality reduction, GMM with six clusters was identified as the most suitable configuration, as indicated by strong validation metrics including the Silhouette Score, Davies Bouldin Index, and Calinski Harabasz Index. The six clusters were found to exhibit distinct demographic and educational profiles. Clusters 0, 2, and 3 consisted of younger graduates in internal study programs, with shorter unemployment durations (1.60–2.33 months) and residences primarily in urban or semi urban districts. Clusters 1, 4, and 5 comprised older, externally oriented graduates with longer unemployment durations (3.13–3.62 months), often located in rural districts. Gender based patterns were also observed: females in internal programs were associated with shorter unemployment durations (1.6–1.9 months) compared with males (2.3 months). Business and Management graduates were predominantly concentrated in urban districts such as Gampaha and Kandy, whereas Arts and Humanities graduates were largely situated in rural districts including Kurunegala and Batticaloa. The findings of this study provide a significant contribution by offering a data driven segmentation of unemployed graduates, enabling policymakers to design more precisely targeted interventions to reduce structural disparities in graduate employability.

Keywords: Clustering algorithms, Gaussian mixture models, Principal component analysis, Socio-economic segmentation, Unsupervised learning.