Unsupervised Machine Learning Approach Of COVID-19 Impact Analysis on Education

A. N. M. Sajedul Alam[1], Rimi Reza[1], Asir Abrar[1], Salsabil Ahmed[1], Tanvir Ahmed[1], Shihab Sharar[1], Annajiat Alim Rasel[1]

[1]*Computer Science and Engineering, BRAC University, Dhaka, Bangladesh*

{a.n.m.sajedul.alam, rimi.reza, asir.abrar, salsabil.ahmed, tanvir.ahmed8, shihab.sharar, annajiat}@g.bracu.ac.bd

Keywords: Clustering, Gaussian Mixture Model, Unsupervised learning, Machine Learning, COVID-19, Education, Students, Spectral Clustering, Agglomerative, Euclidean, K-means clustering. Hierarchical Clustering, Survey.

Abstract: The following paper presents a descriptive observation that is aimed at analysing the impact of COVID-19 on the education of students by implementing unsupervised machine learning techniques on survey datasets from different schools filled up by the students and their parents. The proposed approach analyses the COVID-19 impact on the education of students by the dataset named "COVID-19 Education Impact Survey" obtained from an open-source platform 'data.humdata.org'. The data pre-processing was done using tools such as OpenRefine and Python Pandas library. To implement the approach, different unsupervised learning algorithms such as K-means clustering, Spectral Clustering, Gaussian Mixture Model, Hierarchical Clustering and Principal Component Analysis are being used. The experimental results demonstrate that according to the silhouette score K-means algorithm Euclidean hyperparameter with four clusters outperforms among unsupervised learnings used on the dataset for the expected outcomes. On the other hand, focusing on the log-likelihood score, Gaussian Mixture Model performs better. In hierarchical clustering, the Agglomerative Complete hyperparameter gives the best results. Moreover, the radial basis function works better than the nearest neighbour in the Spectral Clustering.