

ABSTRAK

Sustainable Development adalah konsep yang muncul akibat tumbuhnya kesadaran akan krisis lingkungan yang akan segera terjadi. *United Nations* menetapkan *Sustainable Development Goals* (SDGs) sebagai rencana jangka panjang dalam menghadapi dunia yang terus berkembang. SDGs memiliki 17 target agenda yang perlu dicapai pada tahun 2030 dan dalam rangka pencapaian agenda tersebut, UN terus melakukan survei untuk melihat pencapaian setiap negara dalam memenuhi target SDGs. Dengan adanya kebutuhan untuk mengklasifikasikan pencapaian setiap negara UN, penelitian ini dilakukan untuk mengelompokkan negara-negara UN berdasarkan 2 target SDGs yaitu SDGs *Goal 1: No Poverty* dan SDGs *Goal 2: Zero Hunger* serta melihat indikator-indikator yang paling berpengaruh terhadap pencapaian progres SDGs menggunakan dua metode *clustering*. Metode *clustering* yang digunakan dalam penelitian ini adalah *K-Means Clustering* dan *Hierarchical Clustering Analysis* (HCA). Kemudian dilanjutkan dengan mencari variabel yang signifikan menggunakan metode *Feature Selection* yaitu *Random Forest Importance* (RFI). Metode *K-Means* menggunakan dua metode inisialisasi *centroid*, yaitu secara *random* dan menggunakan *K-Means++*. Kedua metode inisialisasi menghasilkan dua *cluster* yang diberikan label '*Perform Better*' dan '*Perform Worse*'. Kedua metode inisialisasi menunjukkan hasil optimal pada hasil evaluasi *cluster* menggunakan indeks DBI dan Calinski-Harabasz. Kemudian, hasil *clustering* dengan metode HCA pada empat metode *linkage* yaitu *single linkage*, *complete linkage*, *average linkage*, dan *ward linkage* menghasilkan jumlah *cluster* berbeda. Berdasarkan metrik evaluasi CPCC, metode *linkage* yang memiliki performa terbaik adalah *average linkage*. Jumlah *cluster* yang terbentuk menggunakan *average linkage* sejumlah dua *cluster* dan diberikan label '*Perform Better*' dan '*Perform Worse*'. Metode *Feature Selection* yaitu RFI mampu menunjukkan tingkat signifikansi masing-masing variabel dimana hasil RFI menunjukkan terdapat 6 variabel signifikan dengan variabel paling signifikan pada target SDGs 2.4.

Kata kunci: *Sustainable Development Goals* (SDGs), *K-Means Clustering*, HCA, *Feature Selection*

ABSTRACT

Sustainable Development is a concept that emerged as a result of the growing awareness of an imminent environmental crisis. United Nations has set Sustainable Development Goals (SDGs) as a long-term plan to face a constantly developing world. The SDGs have 17 target agendas that need to be achieved by 2030 and in order to achieve this agenda, the UN continues to conduct surveys to see the achievements of each country in meeting the SDGs targets. Given the need to classify the achievements of each UN country, this research was conducted to group UN countries based on the two SDGs which are SDGs Goal 1: No Poverty and SDGs Goal 2: Zero Hunger and also to find out which indicators that have the most influence on achieving SDGs progress using two clustering methods. Methods used are K-Means Clustering and Hierarchical Clustering Analysis (HCA). To find out the significant variables, this research used the Feature Selection method, namely Random Forest Importance (RFI). K-Means clustering was conducted using two centroid initialization methods, random and K-Means++. Both initialization methods resulted in two clusters that are being labeled 'Perform Better' and 'Perform Worse'. The two initialization methods also showed optimal results in the cluster evaluation results using the DBI and Calinski-Harabasz indexes. On the other hand, HCA clustering was conducted by using four linkage methods, namely single linkage, complete linkage, average linkage, and ward linkage which produced a different number of clusters. Based on the CPCC evaluation metrics, the linkage method that had the best performance was the average linkage. The number of clusters formed using the average linkage method was two clusters which are then being labeled 'Perform Better' and 'Perform Worse'. The Feature Selection method, namely RFI, is able to show the significance level of each variable where the RFI results show that there are 6 significant variables with the most significant variable being the SDG target 2.4.

Keywords: Sustainable Development Goals (SDGs), K-Means Clustering, HCA, Feature Selection