Subclustering of Type 2 Diabetes among Malaysian Population for Personalize Medication.

The evolution of healthcare is shifting from a standardized approach to a personalized one, driven by the unique needs of individual patients. Personalized healthcare relies on the integration of genetic, genomic, and other biological information to predict disease risk. Machine learning and data mining have emerged as rapidly advancing fields within healthcare, particularly in classifying patient cohorts from extensive datasets. Their application in diabetes subtyping holds the potential for groundbreaking insights.

Previous research has revealed that diabetes can be stratified into distinct subtypes based on clinical risk factors, genetic markers, and clinical characteristics, facilitating tailored treatment strategies. Notably, clustering algorithms, such as k-means clustering and hierarchical clustering, have played a pivotal role in identifying sub-clusters within the diabetes population.

To delve deeper into the realm of diabetes subtyping, it is essential to define specific objectives and methodologies for applying clustering algorithms to extensive datasets. This approach promises to yield novel insights and advancements in diabetes management.

In our recent study, we developed a sub-clustering model for Type 2 Diabetes (T2D) using k-means clustering, focusing on the Malaysian population. Our research journey encompassed several critical steps, including data preparation, exploratory data analysis, data transformation, and data reduction. The clustering model was meticulously crafted through k-means clustering, with optimization achieved through hyperparameter tuning. To ensure the robustness of our findings, we conducted validation using two distinct internal clustering validation techniques.

The results of our study unveiled the existence of three distinct T2D patient clusters, each characterized by specific clinical variables:

- Older Age-Related Cluster (48%): This cluster predominantly consists of patients whose diabetes is associated with advancing age.
- Obese-Related Cluster (31%): This cluster is primarily composed of patients whose diabetes is closely linked to obesity.
- Insulin Resistance-Related Cluster (21%): Patients in this cluster exhibit diabetes primarily driven by insulin resistance.

Our research underscores the potential for tailoring diabetes management strategies based on these distinct subtypes, thereby optimizing patient care and outcomes.