# Abstract

## Dalia A Kandeil

## An Integrated Clustering Model for the Application of CRM Customer Segmentation

In recent years the application of data mining techniques in customer relationship man- agement (CRM) has received much warranted attention, this is partially due to the abun- dance of customer data stored in company data warehouses. The pillar of CRM is customer identi cation where di erent customer types and their inherent characteristics are identi ed through the segmentation of customers using clustering data mining techniques, typically the standard k-means algorithm. For the customer segmentation task an enhanced clustering model is proposed, which groups customers based on their previous purchasing behavior. This is assessed by means of the Length, Recency, Frequency, Monetary (LRFM) behavioral segmentation model, which ranks customers according to their value and loyalty to the company based on, the customer's relationship length with the company (L), recency of latest transaction (R), purchasing frequency (F), and monetary value (M). The proposed clustering model proceeds to cluster customers according to their LRFM attributes using the k-means++ algorithm, a modi cation to the cluster center initialization of the standard k-means. Moreover, prior to performing the clustering procedure the number of clusters is Selected using an integration of the internal Calinski-Harabasz cluster validity index, as well as the external Rand cluster validity index with a bootstrapping technique. This is done to ensure the stability of the nal chosen clustering result. An interpretation of the clustering result follows in order to identify the various types of customers based on pro tability and longevity. Moreover, a comprehensive analysis of customer rmographics descriptors is performed. The proposed segmentation model clearly takes into account the stability and validity of the clustering result produced through the integration of a bootstrapping technique, o ering more reproducible clustering solutions. Moreover, randomness generated as a result of the clustering algorithm employed and the sample dataset are considered in the proposed work.