

ABSTRACT

Methods and systems for detecting fraudulent data points in a database of a computerized system include receiving, from a user interface, a request for detecting one or more fraudulent data points, choosing minimum and maximum values of k for use in clustering the data points in the database, k being a cluster number; and generating empty outlier scores corresponding to the data points. The methods and systems further include executing, starting with the minimum value of k , functions in an iterative or recursive manner, until the maximum value of k is reached. The functions include choosing k random points as centroids, performing k -means clustering on the chosen centroids, and computing a temporary outlier score for each of the data points in an iterative or recursive manner until a total number of data points is reached. The functions further include updating the outlier scores by adding the temporary outlier scores to the corresponding outlier scores and storing the updated outlier scores. When the maximum value of k is reached, the methods and systems further include normalizing the stored outlier scores and detecting a fraudulent data point based on the normalized outlier scores that indicate consistent degrees.