Today, money laundering (ML) poses a serious threat not only to financial institutions but also to the nation. This criminal activity is becoming more and more sophisticated and seems to have moved from the cliché of drug trafficking to financing terrorism and surely not forgetting personal gain. Most of the financial institutions internationally have been implementing anti-money laundering solutions (AML) to fight investment fraud activities. However, traditional investigative techniques consume numerous man-hours. Recently, data mining approaches have been developed and are considered as well-suited techniques for detecting ML activities. Within the scope of a collaboration project on developing a new data mining solution for AML Units in an international investment bank in Ireland, we survey recent data mining approaches for AML. In this paper, we present not only these approaches but also give an overview on the important factors in building data mining solutions for AML activities.

**INTRODUCTION:**
Money laundering (ML) is a process of disguising the illicit origin of “dirty” money and makes them appear legitimate. It has been defined by Genzman as an activity that “knowingly engage in a financial transaction with the proceeds of some unlawful activity with the intent of promoting or carrying on that unlawful activity or to conceal or disguise the nature location, source, ownership, or control of these proceeds”. Through money laundering, criminals try to convert monetary proceeds derived from illicit activities into “clean” funds using a legal medium such as large investment or pension funds hosted in
retail or investment banks. This type of criminal activity is getting more and more sophisticated and seems to have moved from the cliché of drug trafficking to financing terrorism and surely not forgetting personal gain. Today, ML is the third largest "Business" in the world after Currency Exchange and Auto Industry.

Today, financial institutions manage a huge banking data and more datasets are being recorded daily. The growth of financial data collected by far exceeds human capacities to manage and analyse them efficiently in a traditional way. Global competitions, dynamic markets, and rapidly increase in the technological innovation become important challenges for these organizations. They need to apply new business intelligent solutions, as the traditional statistical methods do not have the capacity to analyse large datasets.

In banking and finance, we can use DM to solve business problems in finding patterns, causalities and correlations in financial information that are not obviously apparent to managers because of the volume of data. DM can firstly be used to analyse huge datasets and build customer profiles of different groups from the existing data. It can generate rules and models that can be used for understanding business performance, making new marketing initiatives, market segmentation, risk analysis and revising company customer policies. DM methods used for customer profiles can be listed as: clustering, classification regression, association rule discovery and sequential pattern discovery. Another important contribution of DM in banking and financial is the prediction

**Current data mining approaches in anti money laundering**

**Clustering:**

Clustering is the process of grouping the data into classes so that objects within the same cluster have high similarity and objects within different clusters are very dissimilar. There are different clustering methods in the literature and they have been successfully exploited for scientific datasets, spatial datasets, business datasets, etc. In AML, clustering is normally used for grouping transactions/accounts into clusters based on their similarities. This technique helps in building patterns of suspicious sequence of transactions and detecting risk patterns of customers/account. One of the most challenges in clustering financial datasets is their size, this technique, for instance, should deal with millions of transactions during hundreds/thousands of time instances. applied a discretization process on their datasets to build clusters. They map their feature space "customer x time x transaction" to n+2 dimensional Euclidean space: n customer dimensions, 1 time dimension and 1 transaction dimension. They firstly discretize the whole timeline into difference time instances. Hence, each transaction is viewed as a node in one-dimensional
timeline space. They project all transactions of customers to the timeline axis by accumulating transactions and transaction frequency to form a histogram. They create clusters based on segments in the histogram. This approach improves firstly the complexity by reducing the clustering problem to a segmentation problem. Next, it avoids the iterative search existing in other clustering algorithms such as K-means. Furthermore, it is more or less appropriate for analysing individual behaviors or group behaviors by their transactions to detect suspicious behaviors related to "abnormal" hills in their histogram. However, as we have to analyse many customers with many transactions of variety amounts for a long period, it is difficult to detect suspicious cases, as there are very few or no "peak hills" in the histogram. Another global analysis is firstly needed and we can then apply this method for further analysis in this case.

**Data Mining frameworks for detecting money laundering**

In order to exploit DM techniques efficiently, they need to be integrated in a framework for detecting ML. A DM framework is normally consisted of four layers [11][13] corresponding to four levels of mining: transaction, account, institution and multi-institution. The most basic level is transactions. In this level, transaction records are extracted for an investigation. However, they provide a few analytical contexts because they do not constitute links to accounts or other data. In the second level, multiple transactions are associated with specific accounts. Aggregation of transaction with individual accounts gives a general view of these accounts on their financial activity. This view shows the degree of association between various accounts based on frequencies of their transactions. At the institution level, the same customer (business or individual) may have multiple accounts. A consolidation of these accounts may show that an institution maybe in ML suspicious and may involve multiple accounts related to different individuals. The last level investigates the ML involving multiple corporations, organizations and customers.