Market Manipulation and Fraudulent Activities Detection using Machine Learning

Harjit Singh
Neighbourhood Campus Dehla Seehan, Punjabi University, Punjab, India

Abstract- In financial markets, market manipulation and fraudulent activities can be detected by identifying unusual patterns in the financial market data. These unusual patterns and behaviour can be detected using machine learning techniques and models easily. In this paper we will discuss how machine learning technology can be applied to the financial market data for the detection of price distortions, market manipulation and fraudulent activities. The paper will discuss how supervised, unsupervised and semi supervised machine learning models can be used and what are the applications and challenges of this technique. Several machine learning models such as neural networks, support vector machines, clustering algorithms and decision trees can we applied to detect market manipulation and fraudulent activities. This paper discusses the use of real time data in market manipulation detection techniques and relates it with feature engineering.

Keywords – *Machine Learning, Market Manipulation, Fraudulent Activities.*

I. INTRODUCTION

Financial markets are much dynamic in nature and various factors such as financial fraud and market manipulation can happen anytime. By deeply analysing the trading volume, price fluctuations, order book irregularities, we can detect these activities in advance. Statistical analysis and other such tools have already been used for detecting such activities but machine learning models are much better than these traditional methods to handle large data sets and capture new data patterns to improve accuracy. The normal market behaviour and the behaviour of the market when some market manipulation or fraud is likely to occur is differentiated by detecting and usual data points. During the fraudulent activities and market manipulation the data points show different patterns than that were during the genuine market movements. Machine learning models can identify these usual patterns in data points to detect market manipulation and fraudulent activities which helps to ensure the integrity and stability of financial markets.

In this paper the applications of machine learning are discussed to detect market manipulation and fraudulent activities and how these techniques have benefits and challenges. This paper also discusses the detection of these activities in real time market conditions.

II. DIFFERENT TYPES OF MARKET MANIPULATION AND FRAUDULENT ACTIVITIES

There are various types of market manipulation and fraudulent activities done by using different ways. Some commonly occurring activities include the following:

Fraud: Many types of financial frauds such as ponzi schemes and insider trading results in irregular trading patterns.

Price: Unusual patterns appear when the price of stock is moved artificially in a direction using false or miss leading information.

Volatility: Many times market fundamentals do not support some sudden price movements or abnormal volatility spikes.

Market sentiment: Sometimes the investors behaviour suddenly changes from normal behaviour and the effects are shown in market movements.

Detection of these market manipulation and fraudulent activities is very important because these activities can impact the market stability and integrity which results investors losses and regulatory challenges. Timely detection of these fraudulent activities is very important to overcome the negative effects on financial markets.

III. MACHINE LEARNING APPLICATIONS TO DETECT MARKET MANIPULATION AND FRAUDULENT ACTIVITIES

Machine learning is a technique where the machine is trained to do a specific task using training data. By training the machine to detect market manipulation and fraudulent activities, it becomes so powerful that it can easily detect unusual complex data patterns and non linear relationships among them. The detection is automatically improved as the machine learns from the new data. The machine learning models to detect market manipulation and fraudulent activities can be categorised as:

3.1 Supervised Learning

In supervised learning, historical data is added to the machine to train it. This historical dataset is called training data which is a labeled data so that the machine can understand the patterns in the data. The historical data related to market manipulation and fraudulent activities is fed to the machine as training data. Using this data the machine becomes able to classify the normal and abnormal data points. Some supervised learning algorithms used for market manipulation and fraudulent activities are:

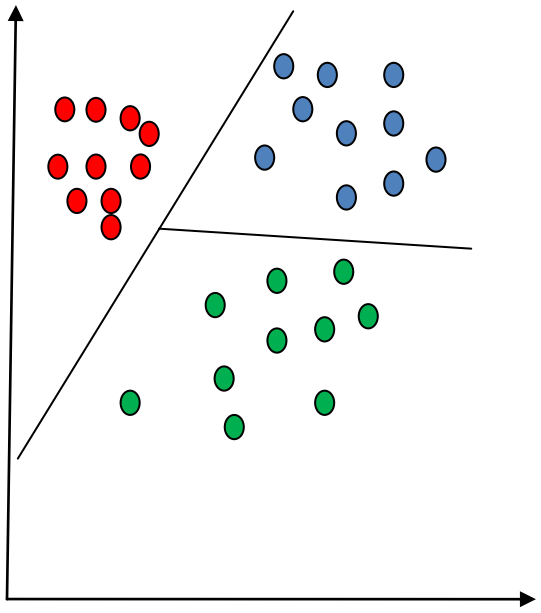

Figure 1. Support Vector Machine separates normal and abnormal patterns.

Support vector machines: Support vector machines are the best algorithms to differentiate between normal and abnormal data points. It is a popular method for the detection of abnormal patterns in datasets and can be best applied for detecting market manipulation and fraudulent activities.

Decision trees: A decision tree is a structure of nodes with branches. Node represent a feature and branches represent decision rules. In this way the machine learning model can identify those features that are related to market manipulation and fraudulent activities.

Neural networks: Autoencoders is a deep learning model that can be used for detecting market manipulation and fraudulent activities. Auto encoders work by taking the input data, compressing it and then reconstructing it. Reconstruction error is measured to find market manipulation or fraudulent activities. Higher the error means there is abnormality.

3.2 Unsupervised learning

In this type of learning the machine learning model do not use labeled data. In this type of learning the structure of the data is analysed to identify any abnormality. So it is used when labeled data is missing and is suitable for real time analysis of financial markets. Some unsupervised learning methods are:

Principal component analysis: The variance in the data is checked to identify principal components. The data points are analysed and the distance is measured from projected subspace to detect market manipulation and fraudulent activities.

Clustering algorithms: Some popular clustering algorithms are Density Based Special Clustering of Applications with Noise (DBSCAN) and K-means algorithm. These algorithms work by grouping the similar data points and then analysing the groups called clusters. Those points which do not come under any cluster represent market manipulation for fraudulent activity.

Isolation forest: The input data is randomly partitioned to isolate abnormal data from the normal data. This isolation method is very useful to detect market manipulation and fraudulent activities by using high dimensional datasets.

3.3 Semi-supervised learning

Supervised and unsupervised learning are combined together in semi-supervised learning. Both the labeled and unlabeled data is used in semi-supervised learning. The amount of labeled data is very small as compared to

unlabelled data. It is useful when we have limited labeled data but large amount of unlabeled data. Self organizing map(SOM) is a popular example of semi-supervised learning method and is suitable for detecting market manipulation and fraudulent activities.

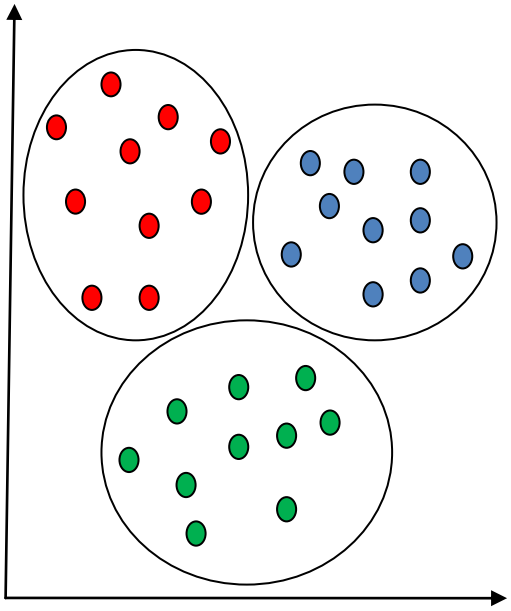

Figure 2: K-Means algorithm groups similar data points to make clusters

IV. MARKET MANIPULATION AND FRAUDULENT ACTIVITIES DETECTION USING FEATURE ENGINEERING

Future Engineering is used to convert raw financial data into meaningful features. These features are then used by machine learning models to detect market manipulation and fraudulent activities. Some important features in financial markets are:

Order book data: Order book data contains information about bids, asks and order depth. Buy analysing this information abnormal activities can be traced.

Price and volume trends: The normal market behaviour is learned from historical prices and volumes, which are then used to find any abnormal behaviour in the market.

Technical indicators: Many types of Technical indicators are analysed such as moving averages, bollinger bands, RSI to trace any abnormal market movements.

V. CASE STUDIES

Machine learning models have been used by many Financial Institutions and firms to detect market manipulation and fraudulent activities. Some examples are:

Risk management: Machine learning models are used by hedge funds to detect market manipulation and fraudulent activities so that they can reduce risk that can occur due to unexpected events for example flash crash.

Fraud detection: Machine learning models are used by banks to keep an eye on price movements and volumes so that they can detect abnormal patterns in this information to detect market manipulation and fraudulent activity.

Cryptocurrency: Cryptocurrency market dealers and analysts also use machine learning models. There is high volatility in cryptocurrency trading and market manipulation activities such as pump-and-dump schemes are used by manipulators.

VI. CHALLENGES

There are many advantages of machine learning to detect market manipulation and fraudulent activities as discussed above. But at the same time there are some challenges or limitations of using this technique:

Dynamic market conditions: Markets continuously goes on changing conditions which affects the machine learning models negatively. The changing market conditions affect the performance of trained models on historical data. So there is a need to continuously learn the new patterns for machine learning models to work well.

Quality of data: The lot of noise in financial market data reduces its quality. Also if labeled data is not available, it will come very difficult to train supervised learning models.

High dimensional data: High dimensional data creates challenges for machine learning models to detect market manipulation and fraudulent activities. Sometimes they cannot work well without over fitting.

Interpretable: In finance sector interpretability is required for decision making under regulatory compliance. Some

machine learning models such as deep learning are not transparent which is another challenge of using this approach.

VII. CONCLUSION

Machine learning is a very sophisticated tool for detecting market manipulation and fraudulent activities in the financial market. It is very flexible and scalable approach that can learn from historical financial data and can analyse complex patterns. Different machine learning techniques such as supervised, unsupervised and semi-supervised learning are available to detect market manipulation and fraudulent activities. But at the same time there are various challenges including dynamic market conditions, quality of data, high dimensional data and interpretablility in models. More research is needed to improve the machine learning models so that they can work well with unlabeled data and continue to work with changing market conditions. In this way machine learning technique will continue to provide a very important role to ensure market integrity and stability.

REFERENCES

[1] Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), 85-126.
[2] Poh, M. Z., & Wang, D. (2010). Predicting financial market anomalies using supervised learning. Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM), 991-996.
[3] Xia, Y., & Wang, S. (2012). Anomaly detection in financial time series using a hybrid model. Proceedings of the 2012 IEEE Symposium on Computational Intelligence for Financial Engineering & Economics (CIFEr), 64-71.
[4] Friedman, J., & Popescu, B. E. (2008). Predictive learning via rule ensembles. The Annals of Applied Statistics, 2(3), 916-954.
[5] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.
[6] Jain, A. K., & Kwak, S. (2006). Anomaly detection in time series data using machine learning algorithms. Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM), 589-598.
[7] Yu, P. S., & Liu, L. (2006). Anomaly detection in time series data using clustering techniques. Proceedings of the 2006 SIAM International Conference on Data Mining (SDM), 1-12.
[8] Chandola, V., & Kumar, V. (2007). A survey of outlier detection techniques in data mining. Proceedings of the 2007 International Conference on Data Mining (ICDM), 3-10.
[9] Iglewicz, B., & Hoaglin, D. C. (1993). How to detect and handle outliers. SAGE Publications.
[10] Markou, M., & Singh, S. (2003). Novelty detection: A review—Part 1: Statistical approaches. Signal Processing, 83(12), 2481-2497.