

# Workshop on Complex Networks in Banking and Finance

Fields Institute, University of Toronto, June 26th 2024

**Enhancing Financial Market Surveillance Processes through a  
Comprehensive Review of their Data Requirements and Machine  
Learning Algorithms**

Aceiton Campos P.

Ben-Abdallah R.

Elia E.

# Acknowledgments

- Financial Market Surveillance Intelligence Centre, École des Sciences de la Gestion, Université du Québec à Montréal.

# Table of content

Research context and question

Related work

Methodological approach

Results and contribution

Future work

# Research context

- Financial regulators have been exploring with machine learning to find new approaches for their market surveillance activities.
- Since 2016, the Financial Conduct Authority (FCA) of England has been organizing TechSprints (proof-of-concept initiatives).
- In 2022, the Bank of England along with the FCA presented the Artificial Intelligence Public-Private Forum (AIPPF) report.

# Research context

- The Monetary Authority of Singapore (MAS) has carried out proof-of-concept initiatives for exploring AI techniques (2018).
- In 2022, MAS along with the Bank for International Settlements (BIS), launched Project Ellipse, a proof-of-concept initiative to explore an integrated regulatory data and analytics platform.

# Research context

- In Québec, the Autorité des Marchés Financiers (AMF) tested in 2017 a machine learning algorithm for analyzing unstructured text in Over the Counter (OTC) derivative data.
- In 2018, The Regulatory Division of the Montréal Exchange signed a joint initiative with IVADO to explore innovative market data analysis and surveillance solutions to detect deceptive marketplace behaviors.
- In 2024, the AMF signed a [partnership](#) with Mila – the Québec Artificial Intelligence Institute (June 19<sup>th</sup>).

# Research question

- Financial regulators and institutions are exploring machine learning.
- Challenge: they require to define **the right data** according to the **case of study** and **machine learning** (“garbage-in / garbage-out”).
- Data set requirements allow to identify the right data.

*What are the data set requirements for machine learning to detect financial market manipulation practices?*

# Related work

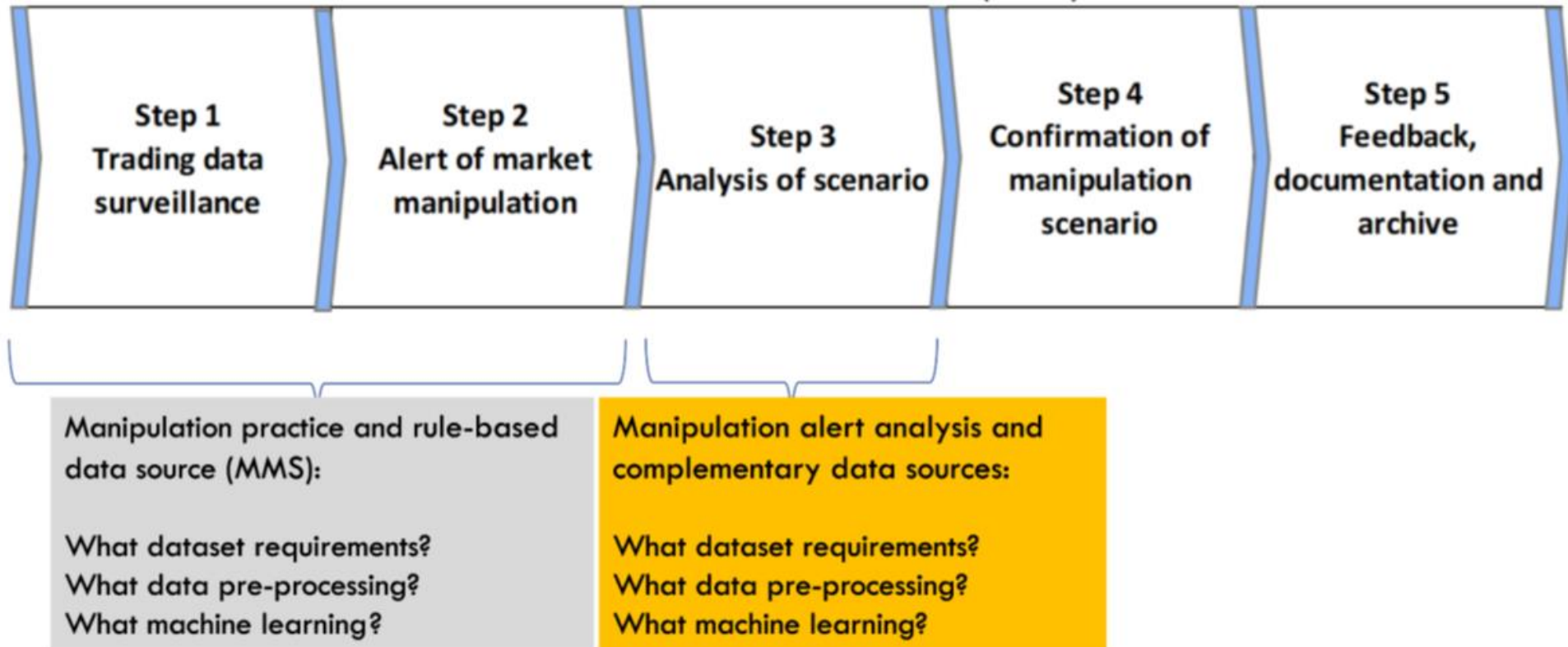
- Tiwari, S., Ramampiaro, H., & Langseth, H. (2021). Machine learning in financial market surveillance: A survey. *IEEE Access*, 9, 159734-159754.
- Zulkifley, M. A., Munir, A. F., Sukor, A., Edil, M., & Mohd Shafiai, M. H. (2023). A Survey on Stock Market Manipulation Detectors Using Artificial Intelligence. *Computers, Materials & Continua*, 75(2).



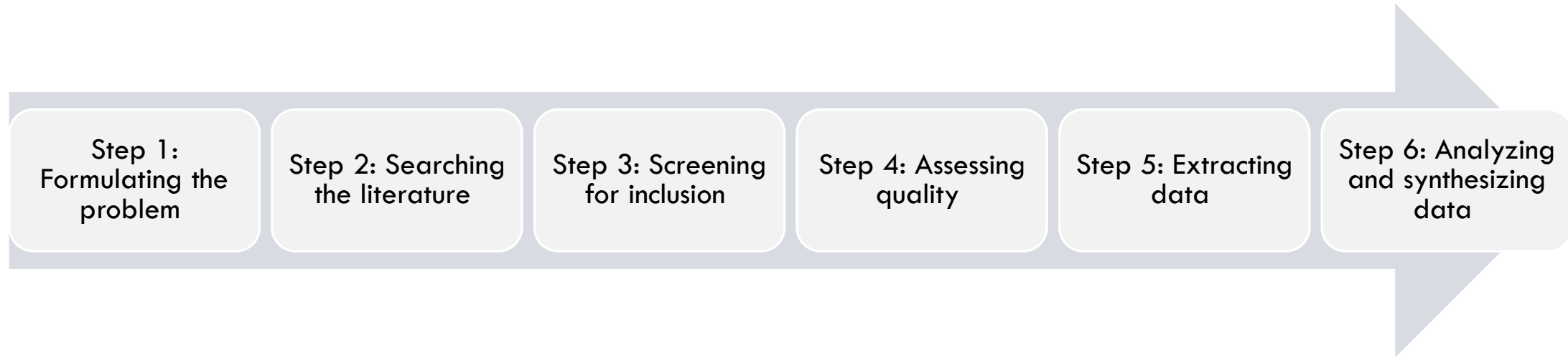
# Research approach

- Previous works show a **high-level description** of datasets, machine learning algorithms and metrics from academic documents related to financial market manipulation.
- Our approach supports financial regulators and institutions in the **seeking of** academic research related to their objectives.
- The research provides detailed findings about data sources, dataset requirements and data pre-process according to machine learning algorithms and the case study, considering a **Financial Market Surveillance Process (FMSP)**.

# Financial Market Surveillance Process (FMSP)



# Methodological approach: 6 review steps



This research follows the general steps and guidelines proposed by Templier et Paré (2015) for a narrative literature review in the domain of Information Systems (IS).

# Review: steps 1-4

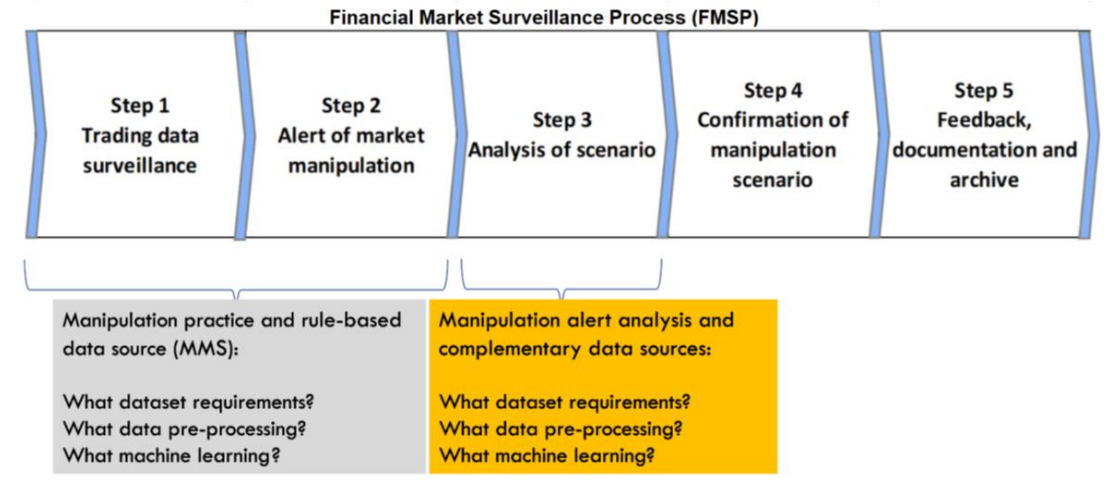


As a result of the first 4 steps, we identified 562 documents and **71 documents** were retained.

# Review: step 5



- We extracted data from documents retained (71), through Comprehensive Categories based on the FMSP configured in NVivo software.

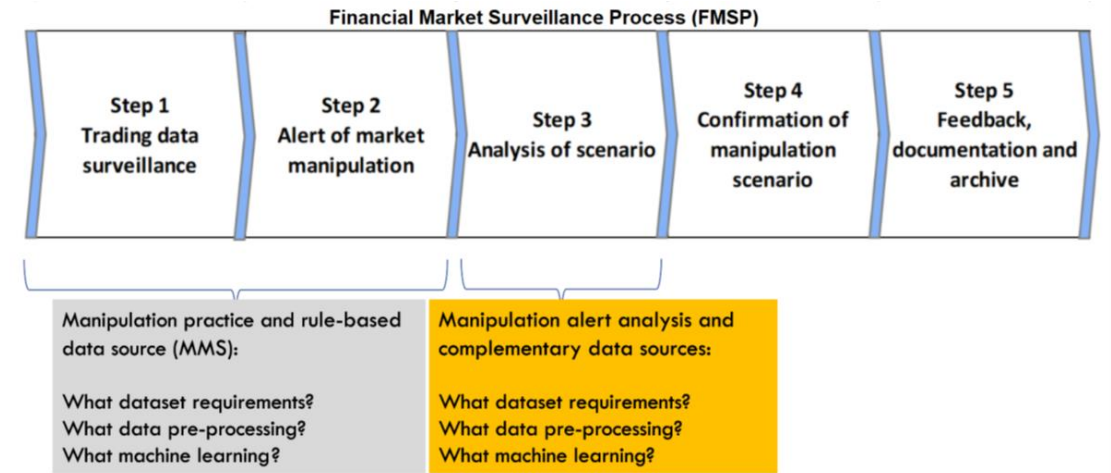


# Review: step 5



## Comprehensive Categories

1. Financial Market Surveillance Process (FMSP)
2. Financial Market Manipulation (specific and general market anomalies)
3. Data source
4. Dataset requirements
5. Data pre-processing
6. Machine Learning (supervised, unsupervised, ensemble)  
Hyperparameter definition  
Benchmark  
Performance Metrics



# About financial market manipulation and market anomalies



Comprehensive Categories
1. Financial Market Surveillance Process (FMSP)
2. Financial Market Manipulation (specific and general market anomalies)
3. Data source
4. Dataset requirements
5. Data pre-processing
6. Machine Learning (supervised, unsupervised, ensemble)
Hyperparameter definition
Benchmark
Performance Metrics



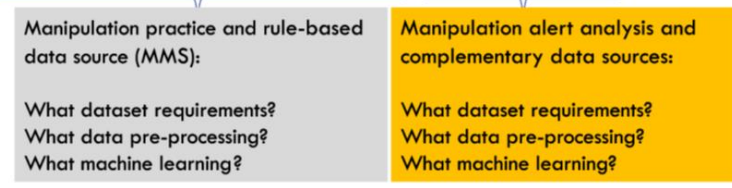
Financial market manipulation types considered in the review	
Insider Trading	Insider trading
Financial Market Manipulation Types	Pump and dump
	Short and distort
	Cornering
	Squeezing
	Advancing the bid
	Reducing the ask
	Matched orders
	Painting the tape
	Wash sales
	Capping
	Begging
	Marking the close
	Open the close
	Banging the close
	Front running
	Churning
	Scalping
	Spoofing
	Pinging
	Quote stuffing
Ramping	
Layering	
Order book fade	

# Review: step 6



- Findings were analyzed and synthesized based on the FMSP to keep the same research perspective.

Paper	FMSP
2021_Collusion_Set_Detection_within_the_Stock_Market_using_Graph_Clustering_amp_Anomaly_Detection -	Alert of Manipulation
2022_A Machine Learning Approach to Detection of Trade-Based Manipulations in Borsa Istanbul -	Alert of Manipulation
2022_A Novel Anomaly Detection Approach for Nifty Stocks using Machine Learning for -	Alert of Manipulation
2022_A Transformer Model for Stock Price Manipulation Detection in the Stock Exchange of Thailand -	Alert of Manipulation
2022_A Hybrid Prediction Model Integrating GARCH Models With a Distribution Manipulation Strategy Ba	Alert of Manipulation
2022_Detection and analysis of atypical stock transactions with possible misuse of insider information and mark	Alert of Manipulation
2022_ Identification of Stock Market Manipulation with Deep Learning -	Alert of Manipulation
2022_Stock market manipulation detection using feature -	Alert of Manipulation
2023_A machine learning attack on illegal trading -	Alert of Manipulation
2023_ Ensemble of supervised and unsupervised deep neural networks for stock -	Alert of Manipulation
2023_ Unsupervised Manipulation Detection Scheme for Insider Trading -	Alert of Manipulation
2022_Can Individual Human Financial Behaviour Be Mathematically Modelled -	Analysis of Scenario
2022_ Identification of Insider Trading in the Securities Market Based on Multi-task Deep Neural Network -	Analysis of Scenario
2022_Sentiment Analysis of Elon Musk's Twitter Data Using LSTM and ANFIS-SVM -	Analysis of Scenario
2023_LLD_A Low Latency Detection Solution to Thwart Cryptocurrency Pump amp Dumps -	Analysis of Scenario
2023_ The Doge of Wall Street Analysis and Detection of Pump and Dump Cryptocurrency Manipulations -	Analysis of Scenario





# Review: step 6 (example of results 1 of 4)



Paper	F MSP	Financial Market Ma	Market Anomalies	Dataset Source
2021_Collusion_Set_Detection_within_the_Stock_Market_using_Graph_Clustering_amp_Anomaly_Detection -	Alert of Manipulation		Anomaly defined by the researcher	Financial Market Data
2022_A Machine Learning Approach to Detection of Trade-Based Manipulations in Borsa Istanbul -	Alert of Manipulation		Regulator analysis	Financial Market Data
2022_A Novel Anomaly Detection Approach for Nifty Stocks using Machine Learning for -	Alert of Manipulation		Normal-abnormal behavior	Financial Market Data,Internet Data
2022_A Transformer Model for Stock Price Manipulation Detection in the Stock Exchange of Thailand -	Alert of Manipulation	Spoofing, Pump and dump		Financial Market Data
2022_A Hybrid Prediction Model Integrating GARCH Models With a Distribution Manipulation Strategy Be	Alert of Manipulation		Anomaly defined by the researcher	Financial Market Data
2022_Detection and analysis of atypical stock transactions with possible misuse of insider information and mark	Alert of Manipulation		Normal-abnormal behavior	Financial Market Data
2022_Identification of Stock Market Manipulation with Deep Learning -	Alert of Manipulation	Pump and dump, Squeezing		Financial Market Data,Internet Data,C
2022_Stock market manipulation detection using feature -	Alert of Manipulation		Regulator analysis	Financial Market Data
2023_A machine learning attack on illegal trading -	Alert of Manipulation	Insider trading		Financial Market Data
2023_Ensemble of supervised and unsupervised deep neural networks for stock -	Alert of Manipulation	Spoofing, Pump and dump	Normal-abnormal behavior	Financial Market Data
2023_Unsupervised Manipulation Detection Scheme for Insider Trading -	Alert of Manipulation	Insider trading		Financial Market Data
2022_Can Individual Human Financial Behaviour Be Mathematically Modelled -	Analysis of Scenario	Pump and dump	Anomaly defined by the researcher	Financial Market Data,Internet Data
2022_Identification of Insider Trading in the Securities Market Based on Multi-task Deep Neural Network -	Analysis of Scenario	Insider trading		Internet Data,Complementary Data
2022_Sentiment Analysis of Elon Musk's Twitter Data Using LSTM and ANFIS-SVM -	Analysis of Scenario		Suspicious transactions identified	Internet Data
2023_LLD_A Low Latency Detection Solution to Thwart Cryptocurrency Pump amp Dumps -	Analysis of Scenario	Pump and dump	Normal-abnormal behavior	Financial Market Data,Internet Data
2023_The Doge of Wall Street Analysis and Detection of Pump and Dump Cryptocurrency Manipulations -	Analysis of Scenario	Pump and dump		Financial Market Data,Internet Data



# Review: step 6 (example of results 2 of 4)



Paper	Dataset_Requireme	Data Preprocess	Synthetic Data
2021_Collusion_Set_Detection_within_the_Stock_Market_using_Graph_Clustering_amp_Anomaly_Detection -	Data Label, Data granular	Balanced dataset, Features detection and optimization, Time features,	Collusive transactions
2022_A Machine Learning Approach to Detection of Trade-Based Manipulations in Borsa Istanbul -	Data Label, Data granular	Balanced dataset, Features detection and optimization, Time features,	
2022_A Novel Anomaly Detection Approach for Nifty Stocks using Machine Learning for -	Data Label, Data cleaning	Time features, Indicators, Features and Control generation	
2022_A Transformer Model for Stock Price Manipulation Detection in the Stock Exchange of Thailand -	Data Label,	Balanced dataset, Indicators, Features and Control generation	Spoofing-Pump&Dump-WA
2022_A_Hybrid_Prediction_Model_Integrating_GARCH_Models_With_a_Distribution_Manipulation_Strategy_Ba	Data Label, Data granular	Supervised, Time features, Indicators, Features and Control generation	
2022_Detection and analysis of atypical stock transactions with possible misuse of insider information and mark	Data Label, Data granular	Unsupervised, Balanced dataset, Features detection and optimization,	
2022_Identification of Stock Market Manipulation with Deep Learning -	Data Label,	Time features,	
2022_Stock market manipulation detection using feature -	Data Label, Data granular	Indicators, Features and Control generation	
2023_A machine learning attack on illegal trading -	Data Label, Data granular	Time features, Indicators, Features and Control generation	
2023_Ensemble of supervised and unsupervised deep neural networks for stock -	Data Label, Data granular	Balanced dataset, Time features, Indicators, Features and Control gen	Spoofing-Pump&Dump-WA
2023_Unsupervised Manipulation Detection Scheme for Insider Trading -	Data granularity,	Features detection and optimization, Time features, Indicators, Feature	
2022_Can Individual Human Financial Behaviour Be Mathematically Modelled -	Data Label, Data granular	Time features,	
2022_Identification of Insider Trading in the Securities Market Based on Multi-task Deep Neural Network -	Data granularity,	Balanced dataset, Features detection and optimization, Indicators, Fea	
2022_Sentiment Analysis of Elon Musk's Twitter Data Using LSTM and ANFIS-SVM -	Data Label, Data granular	Supervised, Unsupervised, Time features, Sentiment and text analysis,	
2023_LLD_A_Low_Latency_Detection_Solution_to_Thwart_Cryptocurrency_Pump_amp_Dumps -	Data Label, Data granular	Features detection and optimization, Time features, Indicators, Feature	
2023_The Doge of Wall Street Analysis and Detection of Pump and Dump Cryptocurrency Manipulations -	Data Label, Data docume	Time features, Indicators, Features and Control generation	



# Review: step 6 (example of results 3 of 4)



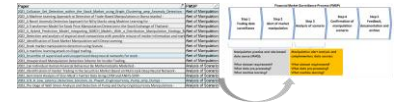
Paper	Supervised Machine Learning	Unsupervised Machine Learning	Ensemble Machine Learning
2021_Collusion_Set_Detection_within_the_Stock_Market_using_Graph_Clustering_amp_Anomaly_Detection -			Graph-Anomaly Detection clustering
2022_A Machine Learning Approach to Detection of Trade-Based Manipulations in Borsa Istanbul -	Decision Tree, Naive Bayes, SVM, Logistic Regression,		Random Forest
2022_A Novel Anomaly Detection Approach for Nifty Stocks using Machine Learning for -		OCSVM	
2022_A Transformer Model for Stock Price Manipulation Detection in the Stock Exchange of Thailand -	Transformation Model		
2022_A Hybrid Prediction Model Integrating GARCH Models With a Distribution Manipulation Strategy Based on LSTM Networks for Stock Market Volatility -			LSTM GARCH
2022_Detection and analysis of atypical stock transactions with possible misuse of insider information and market manipulation -	ANN Multilayer Perceptron,	Decision Tree, Naive Bayes,	AdaBoost-GradBoost- RandForest
2022_Identification of Stock Market Manipulation with Deep Learning -		LSTM-Dynamic Thresholding	
2022_Stock market manipulation detection using feature -	RNN-LSTM		
2023_A machine learning attack on illegal trading -		Nearest neighbor dynamic time warping-DTW	
2023_Ensemble of supervised and unsupervised deep neural networks for stock -	RNN-LSTM	LSTM-AutoEncoder	
2023_Unsupervised Manipulation Detection Scheme for Insider Trading -		Kernel Density Estimation-KDE and MKDE	
2022_Can Individual Human Financial Behaviour Be Mathematically Modelled -			
2022_Identification of Insider Trading in the Securities Market Based on Multi-task Deep Neural Network -	Deep Neural Network-Multitask		
2022_Sentiment Analysis of Elon Musk's Twitter Data Using LSTM and ANFIS-SVM -	RNN-LSTM		Fuzzy rule based model (ANFIS) -SVM
2023_LLD_A Low Latency Detection Solution to Thwart Cryptocurrency Pump amp Dumps -		LSTM-AutoEncoder	
2023_The Doge of Wall Street Analysis and Detection of Pump and Dump Cryptocurrency Manipulations -			Random Forest, AdaBoost-Decision Tree

# Review: step 6 (example of results 4 of 4)



Paper	Hyperparameter definitio	Benchmark	Performance Metric
2021_Collusion_Set_Detection_within_the_Stock_Market_using_Graph_Clustering_amp_Anomaly_Detection -	Hyperparameter definition		Performance Metrics
2022_A Machine Learning Approach to Detection of Trade-Based Manipulations in Borsa Istanbul -			Performance Metrics
2022_A Novel Anomaly Detection Approach for Nifty Stocks using Machine Learning for -	Hyperparameter definition		
2022_A Transformer Model for Stock Price Manipulation Detection in the Stock Exchange of Thailand -	Hyperparameter definition		Performance Metrics
2022_A_Hybrid_Prediction_Model_Integrating_GARCH_Models_With_a_Distribution_Manipulation_Strategy_Ba	Hyperparameter definition		Performance Metrics
2022_Detection and analysis of atypical stock transactions with possible misuse of insider information and mark			Performance Metrics
2022_Identification of Stock Market Manipulation with Deep Learning -	Hyperparameter definition	Statistical Models,	Performance Metrics
2022_Stock market manipulation detection using feature -	Hyperparameter definition	ML Supervised, Statistical Models,	Performance Metrics
2023_A machine learning attack on illegal trading -	Hyperparameter definition	ML Unsupervised, Statistical Mode	Performance Metrics
2023_Ensemble of supervised and unsupervised deep neural networks for stock -	Hyperparameter definition		Performance Metrics
2023_Unsupervised Manipulation Detection Scheme for Insider Trading -		ML Unsupervised,	Performance Metrics
2022_Can Individual Human Financial Behaviour Be Mathematically Modelled -			Performance Metrics
2022_Identification of Insider Trading in the Securities Market Based on Multi-task Deep Neural Network -	Hyperparameter definition	ML Supervised,	Performance Metrics
2022_Sentiment Analysis of Elon Musk's Twitter Data Using LSTM and ANFIS-SVM -	Hyperparameter definition		Performance Metrics
2023_LLD_A Low Latency Detection Solution to Thwart Cryptocurrency Pump amp Dumps -	Hyperparameter definition		Performance Metrics
2023_The Doge of Wall Street Analysis and Detection of Pump and Dump Cryptocurrency Manipulations -	Hyperparameter definition	ML Supervised,	Performance Metrics

# Research contribution



Our review approach:

- **Links** financial regulators/institutions requirements (FMSP) with academic research.
- Provides a good **starting point** for researchers interested in financial market manipulation and machine learning.
- Identifies **practices** for labelling data (from financial regulators documentation) and analysis of scenario (Reedit, X/Twitter, Telegram).
- Describes how previous research **prepared datasets** (data mapping, time features, data balance techniques, feature detection and synthetic data).
- Presents **machine learning algorithms, hyperparameters, benchmark and metrics**.

# Future work

- This review is a relevant component for proposing an **artifact** based on **Design Science** to support financial regulators and institutions in the development of machine learning in financial market manipulations.

# Workshop on Complex Networks in Banking and Finance

Fields Institute, University of Toronto, June 26th 2024

**Enhancing Financial Market Surveillance Processes through a  
Comprehensive Review of their Data Requirements and Machine  
Learning Algorithms**

Aceiton Campos P.

Ben-Abdallah R.

Elia E.

# NVIVO example

NVIVO Literature\_Review\_P1.nvp

Quick Access

IMPORT

- Data
  - Files
  - File Classifications
  - Externals

ORGANIZE

- Coding
  - Codes
  - Sentiment
  - Relationships
  - Relationship Types
- Cases
- Notes
- Sets

EXPLORE

- Queries
  - Query Criteria
  - Query Results
  - Coding Matrices

File Home Import Create Explore Share Modules

Clipboard Item Organize Query Visualize Code Autocode Range Code

Codes

Name	Files	References
2 Financial Market Manipulation	0	0
3. Dataset	0	0
4. Dataset_Requirements	0	0
5. Data Preprocess	0	0
6. Machine Learning	0	0
6.1 Supervised ML	43	76
6.1.1 ANN	5	5
6.1.1.1 ANN Genetic Algorithm	1	1
6.1.1.2 FFNN	1	1
6.1.1.3 RNN-LSTM	5	8
6.1.1.4 ELM	1	1
6.1.1.5 Transformation Model	1	1
6.1.1.6 ANN Multilayer Perceptron	1	1
6.1.1.7 Deep Neural Network and M	1	1
6.1.1.8 ANN	1	1
6.1.2 Decision Tree	11	11
6.1.3 Naive Bayes	6	6
6.1.4 SVM	10	11
6.1.5 Logistic Regression	6	6

PA 137 Items



## sample files.

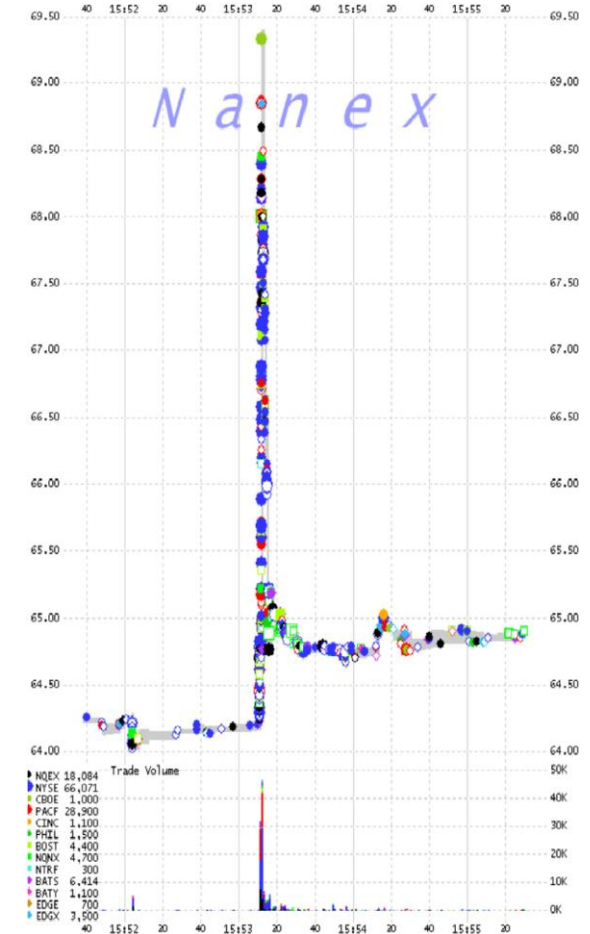
The sample files contain an 'orderbook' file, a 'message' file and a readme summarizing the data's properties. All sample files are based on the official NASDAQ Historical TotalView-ITCH sample.

## demo code.

We have prepared small demo codes for Matlab and R to help you get started with LOBSTER's data. The demo files for [Matlab](#) and [R](#) contain a small code sample, a sample file and a readme. The download links and further code is available in the [code help](#).

## download samples.

Name	Ticker	Level	Size (MB)	Download
Amazon	AMZN	1	0.7	<a href="#">download</a>
		5	2	<a href="#">download</a>
		10	4	<a href="#">download</a>
Apple	AAPL	1	1.3	<a href="#">download</a>
		5	4.2	<a href="#">download</a>
		10	6.5	<a href="#">download</a>
Google	GOOG	1	0.5	<a href="#">download</a>
		5	1.6	<a href="#">download</a>
		10	2.5	<a href="#">download</a>
Intel	INTC	1	3.3	<a href="#">download</a>
		5	5.8	<a href="#">download</a>
		10	6.7	<a href="#">download</a>



Snapshot of the short-lived period pump-and-dump scheme on the Westinghouse Air Brake (WAB) stock in the NYSE from December 14, 2011.

Source: (Photo Courtesy of Nanex).

From <https://doi.org/10.1016/j.eswa.2023.119698>.

# Review process

Research keywords were based on the taxonomy of financial market manipulations proposed by Siering et al. (2017) and the taxonomy of data mining tasks, methods and algorithms proposed by Sharda et al. (2018).

<b>Financial market manipulation types considered in the review</b>	
<b>Insider Trading</b>	Insider trading
<b>Financial Instrument Manipulation</b>	Pump and dump
	Short and distort
	Cornering
	Squeezing
	Advancing the bid
	Reducing the ask
	Matched orders
	Painting the tape
	Wash sales
	Capping
	Begging
	Marking the close
	Open the close
	Banging the close
	Front running
	Churning
	Scalping
Spoofing	
Pinging	
Quote stuffing	
Ramping	
Layering	
Order book fade	

# Review process

General Steps	Methodological Guidelines
Step 1: Formulating the problem	Specify the review's primary goal(s)
	Clearly define the review's key concept(s) and establish its boundaries
Step 2: Searching the literature	Use a combination of data sources and search approaches
	Avoid restrictions that are not based on the research question(s)
Step 3: Screening for inclusion	Use inclusion criteria that reflect the research question(s)
	Include studies from reputable sources
Step 4: Assessing quality	Consider the quality assessment in the selection of studies or the interpretation of the findings
Step 5: Extracting data	Use a structured procedure for data extraction
Step 6: Analyzing and synthesizing data	Summarize the included studies in detail

# Related work (example)

**TABLE 1.** State-of-the-art supervised, semi-supervised, and unsupervised machine learning techniques used for financial market surveillance.

Paper	Methods	Manipulation type	Evaluation measures	Data
Golmohammadi <i>et al.</i> [32]	k-NN, SVM, NN, RF, DT	Wash trade	Precision, Recall, F2 measure	Diaz <i>et al.</i> [7]
Frery <i>et al.</i> [48]	Learning to rank	–	Average, Precision	A highly unbalanced data
Öğüt <i>et al.</i> [6]	SVM, k-NN, ANN, LR	Trade-based manipulation	Recall, Precision	Istanbul Stock Market data
Ahmed <i>et al.</i> [49], Ahmed <i>et al.</i> [13]	Clustering algorithms	Trade-based manipulation	Recall, Precision, and F2 measure	ASX data
Das <i>et al.</i> [40], Das <i>et al.</i> [50]	AAD, iForest-AAD, Tree-based methods	Point anomaly	Quantitative measures	UCI datasets [51]
Cao <i>et al.</i> [3]	k-NN, OCSVM	Price manipulation	AUC/ROC curve	NASDAQ
Li <i>et al.</i> [8]	k-NN, DT, LR, SVM, ANN	Trade-based manipulation	AUC	China Security Regulatory Commission (CSRC)

•Tiwari, S., Ramampiaro, H., & Langseth, H. (2021). Machine learning in financial market surveillance: A survey. *IEEE Access*, 9, 159734-159754.

# Related work (example)

**TABLE 2.** State-of-the-art machine learning methods and statistical techniques used for anomaly detection in sequential/time-series data.

Paper	Methods	Manipulation type	Evaluation measures	Data
Mannila <i>et al.</i> [65]	Event sequencing using windowing technique	Trade-based manipulation	Sensitivity, F2 measure, Specificity	–
Atallah <i>et al.</i> [11]	Event sequencing using windowing technique	Collective, Contextual	Qualitative measures	Walmart data
Rossi <i>et al.</i> [66]	Frequent item set mining using windowing technique and categorical clustering	Collective, Contextual	Sensitivity, Clustering silhouette	Smart meter data
Golmohammadi and Zaiane [67]	CAD	Collective, Contextual	F-measure, Precision, Recall	S&P500 index
Laptev <i>et al.</i> [68]	ARIMA, Kalman filtering	Collective, Contextual	F1-Score	Time-series data
Shipmon <i>et al.</i> [31]	DNN, RNN, LSTM	Collective, Contextual	Confusion matrices, Recall, Precision	Google stream data
Zhu and Laptev [69]	Bayesian deep model	Collective, Contextual	–	Uber cab data
Munir <i>et al.</i> [70]	CNN	Point, Contextual	F-score	Yahoo Webscope
Zhang <i>et al.</i> [71]	Variational autoencoder	Collective, Contextual	AUC	UCR [72], UCI

•Tiwari, S., Ramampiaro, H., & Langseth, H. (2021). Machine learning in financial market surveillance: A survey. *IEEE Access*, 9, 159734-159754.

# Related work (example)

**Table 1:** Summary of the supervised conventional machine learning methods in stock market manipulation detection

Study	Data	Method	Strength	Weakness
Zare et al. [53]	Segmented tick data of Tehran stock exchange	Skewness analysis of Gaussian modelling of manipulated & non-manipulated stock prices	Local daily pattern using partitioned intraday data	Skewness information is not enough to capture complex manipulation schemes
Cao et al. [54]	Four stocks of Apple, Google, Intel & Microsoft (NASDAQ)	Log form of order price, volume & trading time input to SVM & kNN	Optimized for spoof trading & quote stuffing cases	Performance validation is suboptimal as testing is done on small number of cases
Golmohammadi et al. [55]	33 Manipulation cases from Dow Jones industrial companies	Six conventional classifiers: inference trees, RF, NB, SVM, ANN & kNN	Applied SMOTEBoost to equalize imbalanced dataset distribution during training	Low performance with the maximum F2 measure of 53%
Uslu et al. [56]	Daily data of Borsa Istanbul (2010–2015)	Benchmarked against six conventional classifiers: LR, kNN, decision tree, SVM, NB and RF	A set of comprehensive 20 input features of daily stock data	Manipulation cases selection is suboptimal as the duration span is too long in some cases & vice versa
Ougut et al. [57]	277 manipulation cases from Borsa Istanbul (1995–2004)	Four conventional classifiers: SVM, ANN, LR and discriminant analysis	Better dataset for the non-manipulated cases as it is collected from the same companies	Utilizes only three input features: average daily return, average change in trading volume and average volatility
Li et al. [58]	64 manipulation cases by China securities regulation commission	Opening, highest, lowest, closing prices & trading volume input to 7 classifiers: kNN, SVM, DT, LDA, QDA, LR, ANN	Quantified the performance using both tick and daily data input	Suboptimal tick data classifiers with around 50% detection rate only

•Zulkifley, M. A., Munir, A. F., Sukor, A., Edil, M., & Mohd Shafiai, M. H. (2023). A Survey on Stock Market Manipulation Detectors Using Artificial Intelligence. *Computers, Materials & Continua*, 75(2).