DATAMATICS

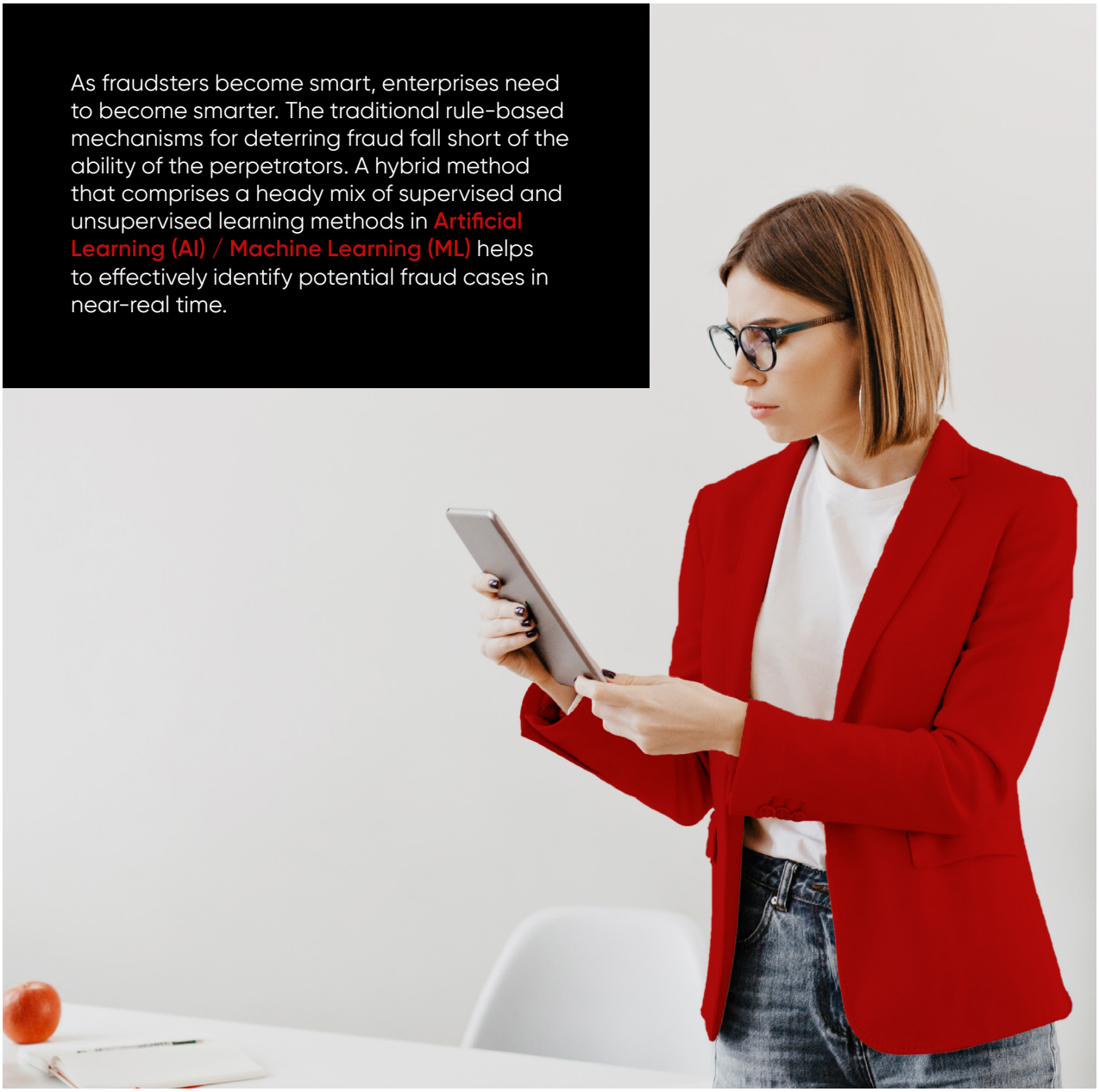
WHITEPAPER

HOW TO DISCERN PATTERNS FROM TRANSACTIONAL DATA FOR FRAUD DETECTION

TABLE OF **CONTENTS**

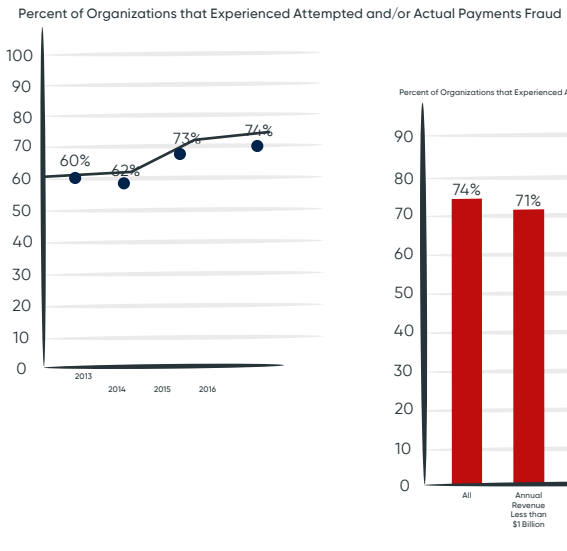
ABSTRACT

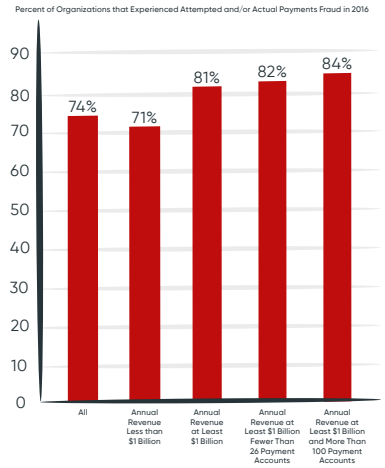
As fraudsters become smart, enterprises need to become smarter. The traditional rule-based mechanisms for deterring fraud fall short of the ability of the perpetrators. A hybrid method that comprises a heady mix of supervised and unsupervised learning methods in **Artificial Learning (AI) / Machine Learning (ML)** helps to effectively identify potential fraud cases in near-real time.

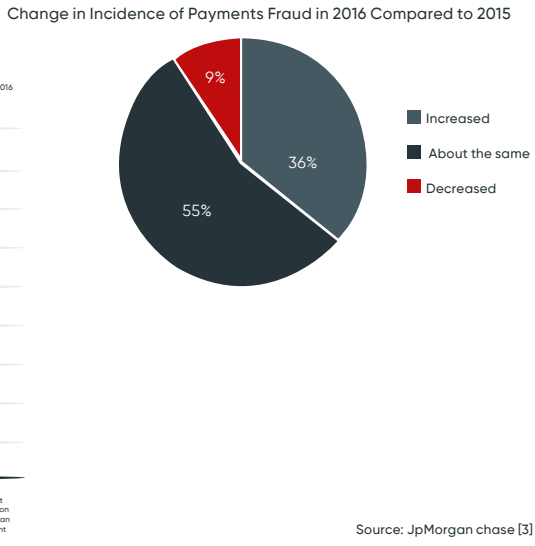


BILLIONS ARE LOST DUE TO FRAUDS

As enterprises become digital, fraudsters are also becoming technology savvy and are constantly on the lookout of loopholes to perpetrate fraud. Every year billions of dollars are lost due to fraud with minuscule possibilities of recovery. It is predicted that by 2020, the resulting revenue loss will reach a staggering $7 billion in USA alone.







Source: JpMorgan chase [3]

As enterprises become digital, fraudsters are also becoming technology savvy and are constantly on the lookout of loopholes to perpetrate fraud.

Even as enterprises strive to counter the bad actors and invest millions, the fraudsters are becoming increasingly sophisticated. They have the technology to effectively side-step the checks and measures that enterprises have implemented after detailed due-diligence. This not only affects the enterprise's revenue but also poses significant risk to its customers as well.

FRAUDSTERS ARE BECOMING SMARTER

Traditional fraud detection systems have limitations and are easy to circumvent. As enterprises become smart and embrace digital market and cashless transactions, fraudsters are becoming smarter. Businesses need fool-proof mechanism to detect potential fraud at the right time and save millions in terms of business revenue.

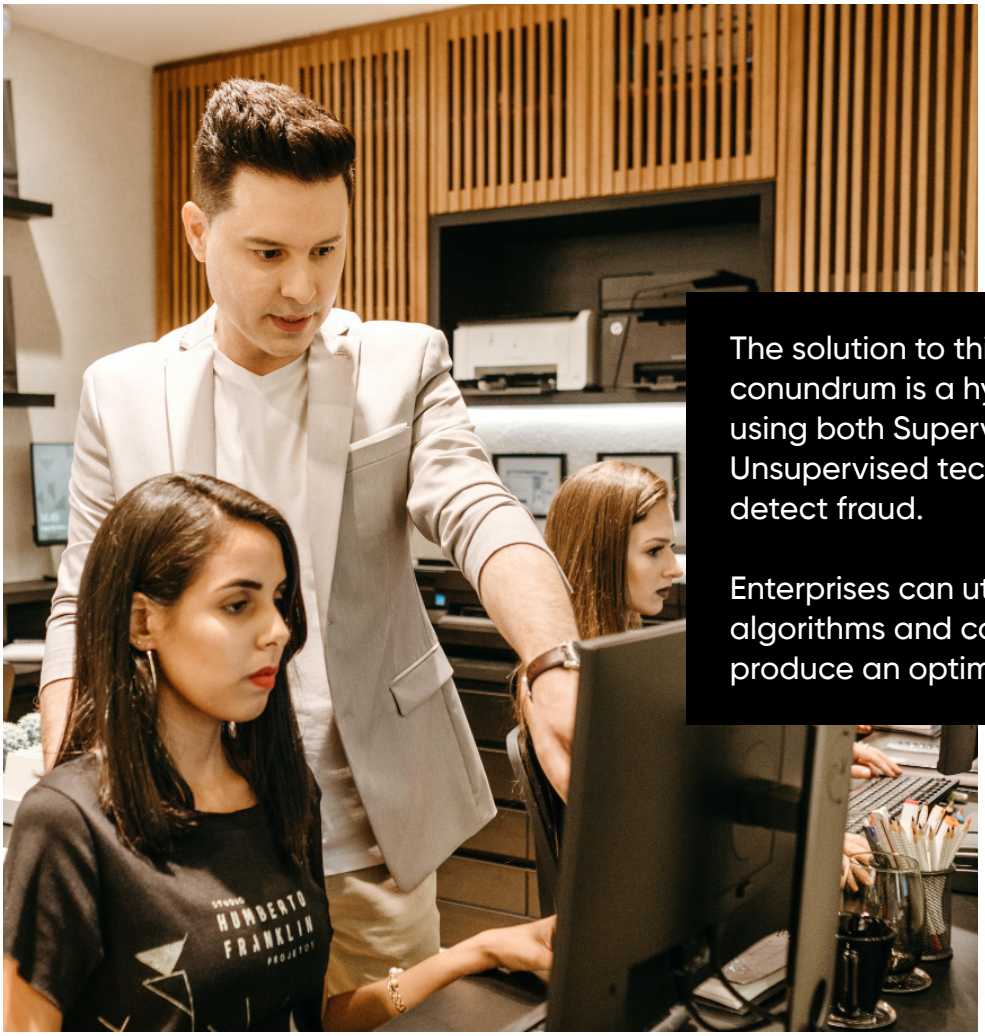
TRADITIONAL VS. NEWAGE MEASURES

The enterprises have traditionally used rule-based measures such as ID verification, blacklisting some IPs, setting a threshold on the maximum number of transactions allowed per day, etc. However, these solutions provide low coverage and are easy to beat making them obsolete sooner than later.

Enterprises are now compelled to resort to AI/ML that helps to overcome the limitations of a rule-based model. AI/ML based approach not only helps to identify patterns and detect anomalies but also sieve out potential frauds. In near-real-time, it can differentiate between a fraudster and a legitimate user without explicitly coding the rules and telling it what to look for, thereby preventing further fraud.

AI/ML is further segregated into Supervised and Unsupervised Learning. The drawback in Supervised models is that the model's accuracy is directly correlated to the amount of relevant training data available, that is, the number of labeled fraudulent cases. Unsupervised Learning model works on unlabeled dataset and thus scores over Supervised Learning models as often there are no known labeled examples of a fraud type. If a fraudster comes up with a new way of perpetrating fraud, then Supervised model will not be able to catch it as the model has not been trained for it thereby defeating the purpose it was meant to serve. Unsupervised models don't rely on any prior knowledge of fraud patterns and is able to discern patterns in datasets without being guided. They are able to detect anomalies in transaction data by identifying patterns that don't fit in with the known patterns.

A HYBRID APPROACH TO OUTSMART THE SMART



The solution to this business conundrum is a hybrid approach using both Supervised and Unsupervised techniques to detect fraud.

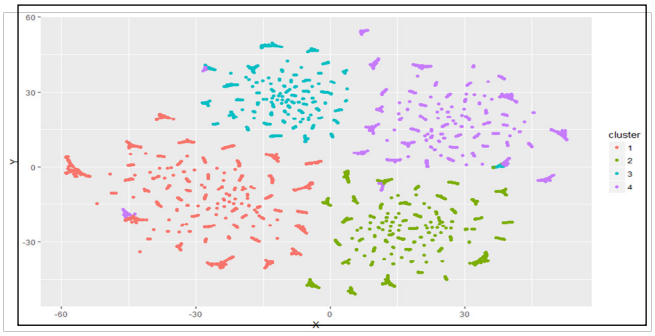
Enterprises can utilize different ML algorithms and combine these to produce an optimal model.

Few of the ML algorithms that are useful are:

01

Clustering

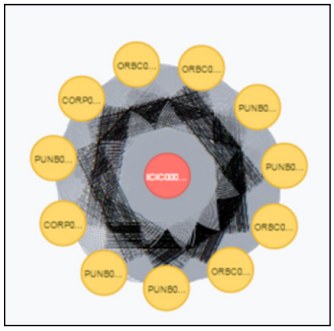
In this method, we first cluster the dataset using numerous parameters in a multivariate space to club together users or entities that have similar characteristics. Fraudsters that behave similarly and act differently from legitimate users either get grouped into one cluster or cannot be placed into any of the legitimate clusters with a high confidence. This group or set of transactions are then further evaluated and analyzed to check for the “abnormality” or “anomaly”. This way of segregating the data into components helps make the patterns within each cluster more visible and also aids in detecting the outliers present in the dataset.



02

Behaviour Patterns

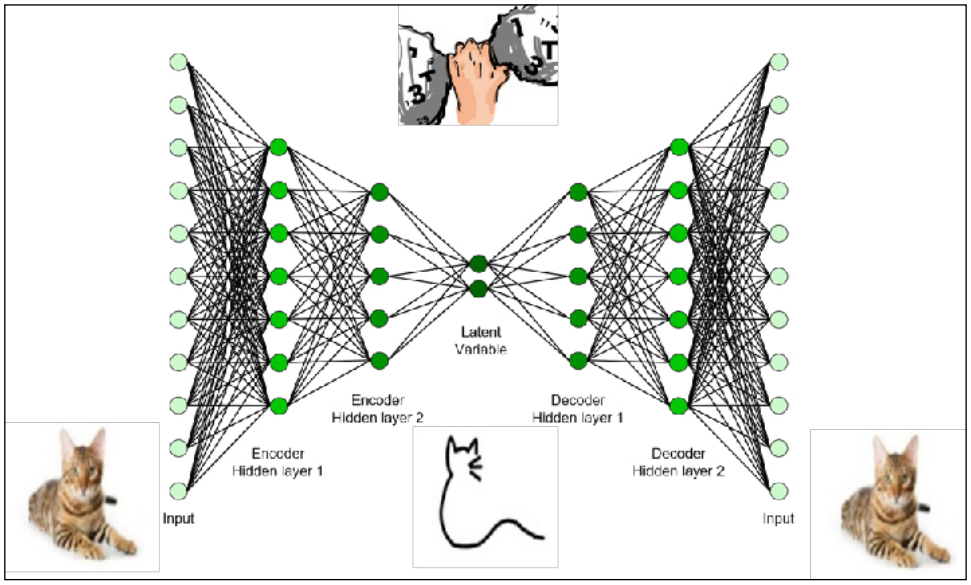
In this method, we use high-level ML algorithms and data analytics to understand and create a behavior profile of a user based on the past behaviour in the system. This typically includes information such as bands of money being transacted, frequency of transaction, time of transaction, with whom all does he/she transact, etc. This helps in determining the transaction patterns of an individual as well as helps in pegging repeated transactions that are unique to the individual. The classification of such behavior also reduces the number of false positives that occur. Further, if any deviation occurs from an observed behavior then these are flagged and further investigated to determine if it is a fraudulent one. If the transaction is found to be legitimate then it is fed back into the model to accommodate for the new behavioural pattern, so on and so forth. Sometimes individuals with suspicious behaviour may be put in a “watch-list” and their transactions closely monitored to assess their potential for perpetrating fraud.



03

Deep Learning

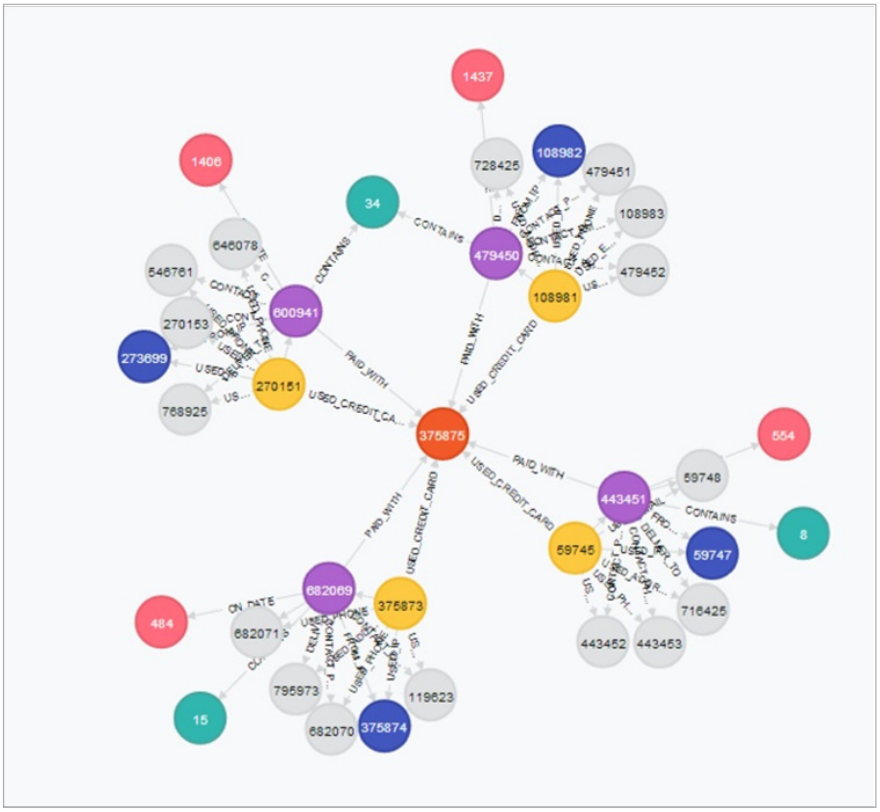
An autoencoder is a neural network that tries to reconstruct its input. In this process, it first creates a lower dimensional embedding of the input and then reconstructs this lower dimensional version back into something that closely resembles the original input. To detect fraud, we first train the autoencoder only on the dataset of transactions that contain no fraudulent cases. This way, after training, the model will be able to reconstruct legitimate transactions very well but will fail to reconstruct a fraudulent one. An anomaly detection autoencoder uses this reconstruction error to determine whether a transaction is fraudulent or not. Transactions with high reconstruction error are flagged as fraudulent.



04

Graph Analytics

In this method, we model the entire dataset as a property graph and use graph analytics to detect fraud. One of the ways fraud goes undetected is when fraudsters are able to bury their patterns in huge amount of data. Graphs provide a structure that naturally embraces relationships and are thus very useful to provide insight into connections that exist in the data. Using graph algorithms we are able to look beyond an individual entity to the relations that exist between entities, thereby, unearthing underlying links and connections among them that would otherwise have gone unnoticed. Based on the connections that are unearthed we are able to relate these entities and detect any collusive activities or fraud rings that may exist between them.



THE TAKEAWAY

Fraud is evolving rapidly and hitting harder across countries and industries than ever before. Though fraud comprises of less than 0.2% of the entire dataset it hits hard and results in huge losses. Supervised Learning method is insufficient to detect fraud in the face of insufficient precedents or training data. An optimal fraud model needs to be designed using traits and characteristics that are obtained from both Supervised and Unsupervised models in conjunction with the domain experts. A one size fits all solution fails when handling fraud as fraud differs by region and industry. As the world becomes increasingly digital and paperless and attacks occur from every corner, there is an ever increasing need to create accurate and adaptable fraud models. If applied correctly, AI/ML techniques are a highly effective weapon in the fight against fraud.

ABOUT
THE AUTHOR

DR. MINITA MATHEW

Lead Consultant - Artificial Intelligence and Cognitive Sciences

Dr. Minita Mathew has coordinated and developed multiple projects in the field of fraud detection and prevention.

Dr. Minita Mathew is a Lead Consultant in the Artificial Intelligence and Cognitive Sciences team of Datamatics. Her interests and responsibilities lie in the domain of Pattern detection and Graph analytics and she has coordinated and developed multiple projects in the field of fraud detection and prevention. She holds a doctorate degree from the Indian Institute of Science, Bangalore.

REFERENCES

https://www.aciworldwide.com/news-and-events/press-releases/2018/january/us-online-fraud-attempts-increase-22-percent-during-2017-holiday-shopping-season

https://www.rsa.com/en-us/blog/2018-08/rsa-report-rogue-mobile-apps-account-for-28-percent-of-fraud-attacks

https://commercial.jpmorganchase.com/jpmpdf/1320732417358.pdf

https://seda.college/seda-applies-artificial-intelligence-language-learning/

http://i-systems.github.io/HSE545/machine%20learning%20all/Workshop/CAE/06_CAE_Autoencoder.html

ABOUT **DATAMATICS**

Datamatics provides intelligent solutions for data-driven businesses to increase productivity and enhance the customer experience. With a complete digital approach, Datamatics portfolio spans across Information Technology Services, Business Process Management, Engineering Services and Big Data & Analytics all powered by Artificial Intelligence.

It has established products in Robotic Process Automation, Intelligent Document Processing, Business Intelligence and Automated Fare Collection.

Datamatics services global customers across Banking, Financial Services, Insurance, Healthcare, Manufacturing, International Organizations, and Media & Publishing.

The Company has presence across 4 continents with major delivery centers in the USA, India, and Philippines. To know more about Datamatics, visit www.datamatics.com

FOLLOW US ON

website: datamatics.com | email: business@datamatics.com

_012022

USA | UK | UAE | India | Philippines