

## Growing Importance of Machine Learning in Compliance and Regulatory Reporting

Dhrubajyoti Dey

### Abstract

Machine Learning is unquestionably one of the most exciting technology fields in the world today. The pace of discovery is almost unimaginably fast, and it seems each day we learn of new advancements, new technologies, new applications, and new opportunities. It's inspiring, but also overwhelming. In this paper we will take a examine the role of Machine learning in Financial sector

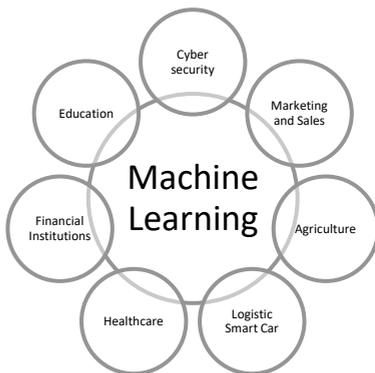
**Keywords:** Machine Learning, Regulatory Reporting , AI, Deep Learning

### Introduction:

Over the past few decades Machine Learning (ML) has become one of the backbones of information technology and with that, it has already revolutionized fields from image recognition to healthcare to financial markets.

Machine learning is magical, it teaches computers to do what comes naturally to creatures: learn from experience. Computational methods are used by ML to "learn" information directly from data without relying on a predetermined equation as a model. The algorithms improve their performance as the number of samples available for learning increases.

Few Real World Applications of ML-



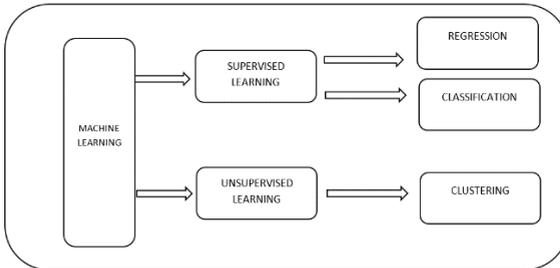
### Machine Learning Mechanism

Machine Learning (ML) is a highly interdisciplinary field which derives and builds upon ideas from statistics, computer science, engineering, cognitive science, optimisation theory and many other disciplines of science and mathematics. Broadly ML can be classified in 2 segments namely Supervised learning and Unsupervised learning .

Supervised learning is simply a formalization of the idea of learning from examples. In supervised learning, the learner (typically, a computer program) is learning provided with two sets of data, a training set and a test set. The idea is for the learner to "learn" from a set of labelled examples in the training set so that it can identify unlabelled examples in the test set with the highest possible accuracy

For example, All the recommendations at Amazon, Facebook, Google, Netflix, Youtube, etc. contains supervised learning techniques. All features or products of any companies such as face recognition, image classification, speech recognition would have used Supervised learning techniques

Supervised learning is extensively used in areas such as Bioinformatics , Database marketing , Handwriting recognition , Information retrieval , Object recognition in computer vision , Optical character recognition , Spam detection ,Pattern recognition



Unsupervised learning is more subjective than supervised learning, as there is no simple goal for the analysis, such as prediction of a response . The term “unsupervised learning” or “learning without a teacher” is generically associated with the idea of using a collection of observation  $X_1, \dots, X_n$  sampled from a distribution  $p(X)$  to describe properties of  $p(X)$ .

Clustering is basically grouping a set of data examples so that examples in one group (or one cluster) are more similar (according to some criteria) than those in other groups. This is often used to segment the whole dataset into several groups. Analysis can be performed in each group to help users to find intrinsic patterns. For instance, clustering individuals based on demographics might result in a clustering of the wealthy in one group and the poor in another

This is a data-driven method that can work well when there is adequate data; for instance, social information filtering algorithms, such as those that Amazon.com use to recommend books, are based on the principle of finding similar groups of people and then assigning new users to groups. In few cases, such as with social information filtering, the information about other members of a cluster (such as what books they read) can be sufficient for the algorithm to produce meaningful results. In other cases, it may be the case that the clusters are merely a useful tool for a human analyst. Unfortunately, even unsupervised learning suffers from the problem of overfitting the training data. There's no silver bullet to evade the problem because any algorithm that can learn from its inputs needs to be quite powerful.

Unsupervised learning has produced many successes, such as world-champion calibre backgammon programs and even machines capable of driving cars! It can be a powerful technique when there is an easy way to assign values to actions

### Machine learning to Deep Learning

Deep learning is making major advances in solving problems that have resisted the best attempts of the artificial intelligence community for many years .Deep learning can be based on both supervised and unsupervised methods . Deep learning allows computational models that are composed of several processing layers to learn representations of data with multiple levels of abstraction.

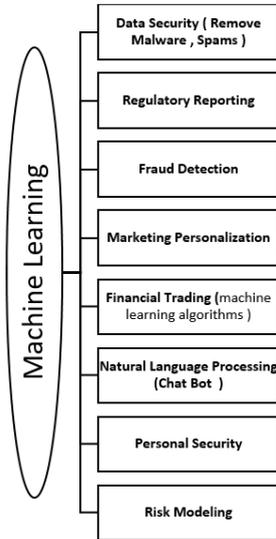
These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the **Backpropagation** algorithm to showcase how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer.

### Application within Financial Institutions ( FI)

With the advance of digital world , the amount of data gathered in Financial institutions (FI) have increased rapidly. The source of data for a Financial institutions could be many e.g. emails, internet banking, voice call, social media, websites etc . This information can be used to provide solution to meet the requirement of the customer which might be very unique and could be otherwise known and offered. Also After the financial crisis in 2008, many new regulations and supervisory measures were introduced that required FIs to report more detailed and more frequent data on more aspects of their business models and balance sheets

Data is critical not only for regulatory purposes but also for internal reporting, risk management purposes (i.e. in implementing the policies and practices set at the top of the organization), and identifying, assessing and adequately measuring risk areas. Additionally, Financial institutions need to have reliable data and manage it adequately in order to conduct predictive analytics on customer transactions, account activities, credit card balances and loan portfolios, and to better gauge their preferences, receptivity to marketing programs, and changes in the pattern of their activities.

From the perspective of the Financial institutions revamping the infrastructure represents their biggest challenge and Machine learning can be efficient tool to analyse the data and report the correct data .



### ML in Regulatory Reporting

With millions of transactions taking place every day, it will be impossible for humans to study what is going on and identify patterns. Therefore, the analytics tools themselves must be able to evolve, taking into account new information and comparing it to past data in order to make decisions about the validity of transactions. Rather than having associates spending weeks, potentially months glued to a computer screen, intelligent programs can improve efficiency and reduce cost by elevating human involvement to reviewing sample sets, interrogating search results and analysing the end product.

The industry's current focus has been improving regulatory reporting procedures for banks and financial services firms, such as increased automation and digitization of Know Your Customer (KYC) rules, Anti-Money Laundering (AML) rules, and tax reporting

Machine learning, robotics, artificial intelligence and other improvements in automated analysis and computer thinking create enormous possibilities when applied to compliance. Data mining algorithms based on machine learning can organize and analyze large sets of data, even if this data is unstructured and of a low quality, such as sets of e-mails, pdfs and spoken word.<sup>5</sup> It can also improve the interpretation of low-quality data outputs from payments systems. Machine learning can create self-improving and more accurate methods for data analysis, modeling and forecasting as needed for stress testing. In the future, artificial intelligence could even be applied in software automatically interpreting new regulations.

Machine learning identifies complex, nonlinear patterns in large data sets and makes more accurate risk models. By adjusting algorithms based on newly acquired information, their predictive power improves through use. This has several potential applications in compliance. In stress testing and risk management, it would benefit definition of models, the calculation and simulation of stress scenarios, and improve the accuracy and granularity of statistical analyses.

Big financial services firms are starting to use robotic process automation, natural language processing, and natural language generation to gather and organize structured and unstructured data for a wide range of business and compliance processes – from credit scoring, product pricing, limit management, and loan reviews to fraud management, vendor risk management, compliance testing, collateral management, report generation, and claims processing. This advanced automation improves efficiency and reduces compliance costs. Perhaps even more important, it enables a firm's valuable and scarce analysts and experts to do the jobs they were hired to do, instead of spending most of their time manually collecting data and updating spreadsheets.

Machine learning has a long way to go and it will be interesting to see how the world adopts to this new culture

## References

- [1] [1] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for Boltzmann machines.
- [2] Cognitive Science, 9:147–169, 1985.
- [3] [2] B. D. O. Anderson and J. B. Moore. Optimal Filtering. Prentice-Hall, Englewood Cliffs, NJ, 1979.
- [4] [3] C. E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems.
- [5] Annals of Statistics, 2(6):1152–1174, 1974.
- [6] [4] S. Arnborg, D. G. Corneil, and A. Proskurowski. Complexity of finding embeddings in a k-tree. SIAM
- [7] Journal of Algebraic and Discrete Methods, 8:277–284, 1987.
  
- [8] [5] D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent Dirichlet allocation. JMLR, 3:9931022, 2003.
- [9] [6] J. D. Lafferty, A. McCallum, and F. Pereira, Conditional random fields: Probabilistic modeling for segmenting and labeling sequence data, Proceedings of International Conference on Machine Learning (San Francisco, CA), vol. 18, Morgan Kaufmann, 2001, pp. 282–289.