VOL- VIII ISSUE- X OCTOBER 2021 PEER REVIEW e-JOURNAL IMPACT FACTOR 7.149 ISSN 2349-638x

Machine Learning and Artificial Intelligence in Finance

Dr. Akhilesh Saini
Associate Professor (Computer Science),
Ch. K.R.Godara Memorial College, Bashir,Tibbi, India

Abstract—

Artificial intelligence (AI) impacts numerous aspects of life in the form of smart devices and smart applications, designed to understand consumer behavior, needs, and preferences in order to deliver customized experiences. AI has been one of the primary drivers of innovation in marketing. Market-ers are already leveraging the advantages of AI to gain valuable insights into customers, competitors and markets. Besides, AI automate tasks, reduce costs, and improve workflows. This paper examines the current and potential applications of AI within marketing by providing comprehensive overview of existing academic research. Artificial intelligence and machine learning are now part of our lives. The expression "artificial intelligence" refers to the simulation of human intelligence by computers. Computers are trained to sense, reason, act, and adapt as humans do. This is made possible thanks to machine-learning algorithms that allow applying learning from data and experience towards future decisions or predictions. Current trends in artificial intelligence and machine learning include reinforcement learning, quantum computing, natural processing language, image analysis, recognition, biased data, neural networks, and deep learning.

I.Introduction

Artificial intelligence and machine learning have the potential to transform our world in a host of ways and will have an impact not just on the technology of how things are done but will have great social impact. Some, such as the futurist Ray Kurzweil, are even suggesting the dawn of a new epoch with concomitant revolutions in not just how humans act, but what humans are.

There is already a large and fast-growing literature that covers theoretical and computational developments; as well as a wide range of potential applications to finance. Several popular and prominent artificial intelligence and machine-learning capabilities are already in use by banks, insurance firms, and other financial institutions. The growing demand for artificial intelligence and machine learning technologies within the financial services industry is largely motivated by the huge amount of structured and unstructured data that can be used to predict and anticipate customer decisions; as well as to create strategies. Many researchers have investigated the realized or potential deployment of artificial intelligence and machine learning in financial contexts such as, credit decisions, risk management, quantitative trading, cyber security, financial forecasting, financial cycles, and fraud detection.

Artificial intelligence and Machine-learning capabilities are already in use by banks, insurance firms, and other financial institutions. The growing demand for artificial intelligence and machine learning technologies within the financial services industry is largely motivated by the huge amount of structured and unstructured data that can be used to predict and anticipate customer decisions; as well as to create strategies. Many researchers have investigated the realized or potential deployment of artificial intelligence and machine learning in financial contexts such as, credit decisions, risk management, quantitative trading, cyber security, financial forecasting, financial cycles, and fraud detection.

In order to promote our understanding, with respect to finance, of the uses and consequences of the development of artificial intelligence and machine-learning, Elsevier's Research in International Business and Finance is calling for papers for a special issue devoted to these subjects.

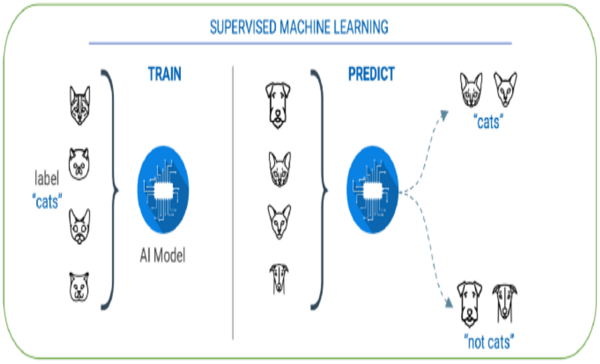
Topics of interest include, but are not limited to, the role and impact of artificial intelligence and machine learning in…

- Financial decision making · Risk assessment and management
- Financial forecasting · Performance and response of global
- Financial development financial markets to pandemics
- Trading algorithms · Capital budgeting
- Option pricing Corporate finance
- Portfolio analysis · Corporate governance
- Asset and liability management · Insurance
- Financial economics · Blockchain and cyber security
- Interest rate models · Financial planning
- Bank management · Financial engineering
- The impact of pandemics on financial · Facial payment technology market.

Machine Learning

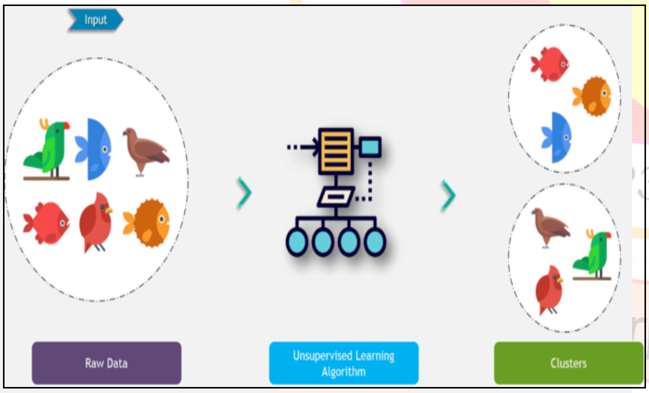
Machine learning, a subset of artificial intelligence, focuses on developing computer programs that autonomously learn and improve from experience without being explicitly programmed. The three broad types of machine learning are supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning



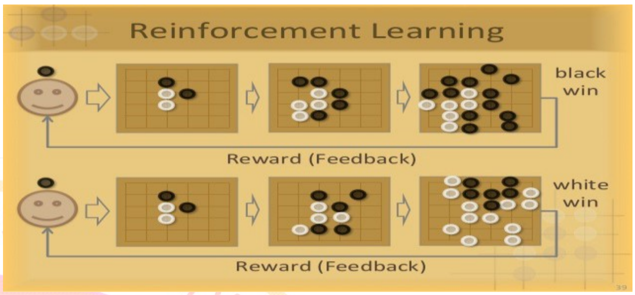
The goal of supervised learning is to create predictive models. Initially, a training data set with labeled input and output examples are fed to the algorithm (hence the name supervised). Then, the algorithm runs on the training set with its parameters adjusted until it reaches a satisfactory level of accuracy. From this analysis, the algorithm creates a function that can predict future outputs. In the image above, the AI model is given pictures of cats that are labeled as "cats". The model is then trained on the labeled data of cats until it can recognize the patterns in the images of cats. As a result, the model would be able to predict if later images are showing cats or not cats by responding to the previously recognized patterns.

Unsupervised Learning



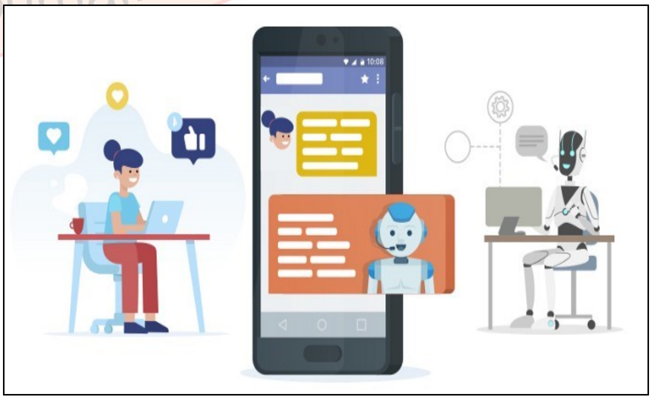
The goal of unsupervised learning is to find patterns in data. Contrary to supervised learning, an unsupervised algorithm is given a training set without classified or labeled examples (hence the name unsupervised). To discern patterns, the algorithm uses clustering. Each cluster is defined by the criteria needed to meet its requirements; that criteria are then matched with the processed data to form the clusters. The training set is then broken into clusters based on common features. In the image above, the input data has no class labels and comprises of fish and birds. An unsupervised model built using this input data will create one cluster of fish and another cluster of birds by grouping the data based on common features.

Reinforcement Learning



The goal of reinforcement learning is to train a model to make a sequence of decisions that will maximize the total reward. In reinforcement learning, a machine learning model faces a game-like situation where it uses trial and error to solve the problem it is facing. The programmer manipulates the model to act in a certain way by adding rewards and penalties. As a result, the model is incentivized to perform behaviors that have rewards and discouraged from performing behaviors that incur penalties (this feedback is the "reinforcement"). Once the model is left on its own to figure out the best approach to maximizing reward, it progresses from random trials to sophisticated tactics. For example, Google's Alpha Go computer program trained to play the game Go and ended up beating the world champion. This was a huge achievement because there are 10^{170} possible board configurations (more than the number of atoms in the known universe) and no computer program had previously beat a professional Go player.

Natural Language Processing



Natural language processing is another subset of artificial intelligence with uses in finance. The overarching goal of natural language processing is simple: decipher and understand human language. Speech recognition software (ex. Siri) isolates individual sounds from speech audio, analyzes these sounds, uses algorithms to find the best word fit, transcribes the sounds into text. After converting the natural language into a form a computer can understand, the computer employs algorithms to derive meaning and collect essential data from the text. Now that we understand machine learning and natural language processing, we can look at artificial intelligence in finance with a better understanding.

1. Big Data Is Boosting Intelligent Behavior in Machines

Machine learning (ML) and artificial intelligence (AI) are becoming dominant problem-solving techniques in many areas of research and industry, not least because of the recent successes of deep learning (DL). However, the equation AI=ML=DL, as recently suggested in the news, blogs, and media, falls too short. These fields share the same fundamental hypotheses: computation is a useful way to model intelligent behavior in machines. What kind of computation and how to program it? This is not the right question. Computation neither rules out search, logical, and probabilistic techniques, nor (deep) (un)supervised and reinforcement learning methods, among others, as computational models do include all of them. They complement each other, and the next breakthrough lies not only in pushing each of them but also in combining them.

Big Data is no fad. The world is growing at an exponential rate and so is the size of the data collected across the globe. Data is becoming more meaningful and contextually relevant, breaking new grounds for machine learning (ML), in particular for deep learning (DL) and artificial intelligence (AI), moving them out of research labs into production (Jordan and Mitchell, 2015). The problem has shifted from collecting massive amounts of data to understanding it—turning it into knowledge, conclusions, and actions. Multiple research disciplines, from cognitive sciences to biology, finance, physics, and social sciences, as well as many companies believe that data-driven and “intelligent” solutions are necessary to solve many of their key problems. High-throughput genomic and proteomic experiments can be used to enable personalized medicine. Large data sets of search queries can be used to improve information retrieval. Historical climate data can be used to understand global warming and to better predict weather. Large amounts of sensor readings and hyper spectral images of plants can be used to identify drought conditions and to gain insights into when and how stress impacts plant growth and development and in turn how to counterattack the problem of world hunger. Game data can turn pixels into actions within video games, while observational data can help enable robots to understand complex and unstructured environments and to learn manipulation skills.

However, is AI, ML, and DL really synonymous, as recently suggested in the news, blogs, and media? For example, when AlphaGo (Silver et al., 2016) defeated South Korean Master Lee Se-dol in the board game Go in 2016, the terms AI, ML, and DL were used by the media to describe how AlphaGo won. In addition to this, even Gartner's list (Panetta, 2017) of top 10 Strategic Trends for 2018 places (narrow) AI at the very top, specifying it as “consisting of highly scoped machine-learning solutions that target a specific task.”

2. Artificial Intelligence and Machine Learning

Artificial intelligence and ML are very much related. According to McCarthy (2007), one of the founders of the field,

AI is “the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.”

This is fairly generic and includes multiple tasks such as abstractly reasoning and generalizing about the world, solving puzzles, planning how to achieve goals, moving around in the world, recognizing objects and sounds, speaking, translating, performing social or business transactions, creative work (e.g., creating art or poetry), and controlling robots. Moreover, the behavior of a machine is not just the outcome of the program, it is also affected by its “body” and the environment it is physically embedded in. To keep it simple, however, if you can write a very clever program that has, say, human-like behavior, it can be AI. But unless it automatically learns from data, it is not ML:

ML is the science that is “concerned with the question of how to construct computer programs that automatically improve with experience,” (Mitchell, 1997).

So, AI and ML are both about constructing intelligent computer programs, and DL, being an instance of ML, is no exception. Deep learning (LeCun et al., 2015; Goodfellow et al., 2016), which has achieved remarkable gains in many domains spanning from object recognition, speech recognition, and control, can be viewed as constructing computer programs, namely programming layers of abstraction in a differentiable way using reusable structures such as convolution, pooling, auto encoders, variation inference networks, and so on. In other words, we replace the complexity of writing algorithms, that cover every eventuality, with the complexity of finding the right general outline of the algorithms—in the form of, for example, a deep neural network—and processing data. By virtue of the generality of neural networks—they are general function approximates—training them is data hungry and typically requires large labeled training sets. While benchmark training sets for object recognition, store hundreds or thousands of examples per class label, for many AI applications, creating labeled training data is the most time-consuming and expensive part of DL. Learning to play video games may require hundreds of hours of training experience and/or very expensive computing power. In contrast, writing an AI algorithm that covers every eventuality of a task to solve, say, reasoning about data and knowledge to label data automatically (Ratner et al., 2016; Roth, 2017) and, in turn, make, for example, DL less data-hungry–is a lot of manual work, but we know what the algorithm does by design and that it can study and that it can more easily understand the complexity of the problem it solves. When a machine has to interact with a human, this seems to be especially valuable.

This illustrates that ML and AI are indeed similar, but not quite the same. Artificial intelligence is about problem solving, reasoning, and learning in general. Machine learning is specifically about learning—learning from examples, from definitions, from being told, and from behavior. The easiest way to think of their relationship is to visualize them as concentric circles with AI first and ML sitting inside (with DL fitting inside both), since ML also requires writing algorithms that cover every eventuality, namely, of the learning process. The crucial point is that they share the idea of using computation as the language for intelligent behavior. What kind of computation is used and how should it be programed? This is not the right question. Computation neither rules out search, logical, probabilistic, and constraint programming techniques nor (deep) (un)supervised and reinforcement learning methods, among others, but does, as a computational model, contain all of these techniques.

Reconsidering AlphaGo: AlphaGo and its successor AlphaGo Zero (Silver et al., 2017) both combine DL and tree search—ML and AI. Alternatively, the "Allen AI Science Challenge" (Schoenick et al., 2017) should be considered. The task was to comprehend a paragraph that states a science problem, at the middle school level and then to answer a multiple-choice question. All winning models employed ML yet failed to pass the test at the level of a competent middle schooler. All winners argued that it was clear that applying a deeper, semantic level of reasoning with scientific knowledge to the question and answers, is the key to achieving true intelligence. In other words, AI has to cover knowledge, reasoning, and learning, using programmed and learning-based programmed models in a combined fashion.

3. The Joint Quest to Identify Intelligent Behavior in Machines

Using computation as the common language, we have come a long way, but the journey ahead is still long. None of today's intelligent machines come close to the breadth and depth of human intelligence. In many real-world applications, as illustrated by AlphaGo and the Allen AI Science Challenge, it is unclear whether problem formulation falls neatly into fully learning. The problem may well have a large component, which can be best modeled using an AI algorithm without the learning component, but there may be additional constraints or missing knowledge that take the problem outside its regime, and learning may help to fill the gap. Similarly, programmed knowledge and reasoning may help learners to fill their gaps. There is a symmetric difference between AI and ML, and intelligent behavior in machines is a joint quest, with many vast and fascinating open research problems:

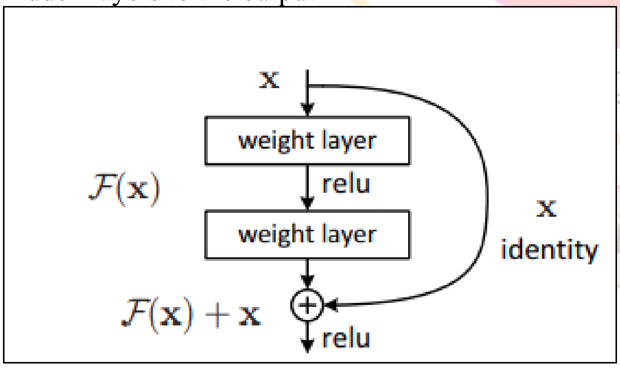
- How can computers reason about and learn with complex data such as multimodal data, graphs, and uncertain databases?
- How can preexisting knowledge be exploited?
- How can we ensure that learning machines fulfill given constraints and provide certain guarantees?
- How can computers autonomously decide the best representation for the data at hand?
- How do we orchestrate different algorithms, involving learned or not learned ones?
- How do we democratize ML and AI?
- Can learned results be physically plausible or easily understood by us?
- How do we make computers learn with us in the loop?
- How do we make computers learn with less help and data provided by us?

- Can they autonomously decide the best constraints and algorithms for a task at hand?
- How do we make computers learn as much about the world, in a rapid, flexible, and explainable manner, as humans?

Answering these and other similar questions will put the dream of intelligent and responsible machines into reach. Fully programmed computations, together with learning-based programmed computations, will help to better generalize, beyond the specific data that we have seen, whether a new pronunciation of a word or an image will significantly differ from those we have seen before. They allow us to go significantly beyond supervised learning, towards incidental and unsupervised learning, which does not depend so much on labeled training data. They provide a common ground for continuous, deep, and symbolic manipulations. They allow us to derive insights from cognitive science and other disciplines for ML and AI.

Learning the identity function is extremely difficult as the scope of all possible combination of weights and biases is enormous, thus the chance of learning the identity function is minuscule. As seen above, adding more layers to a neural network can actually do the opposite: more layers = lower accuracy (diminishing returns).

The paper identifies that there is one solution to this. That is by adding the inputs of the hidden layers to the output



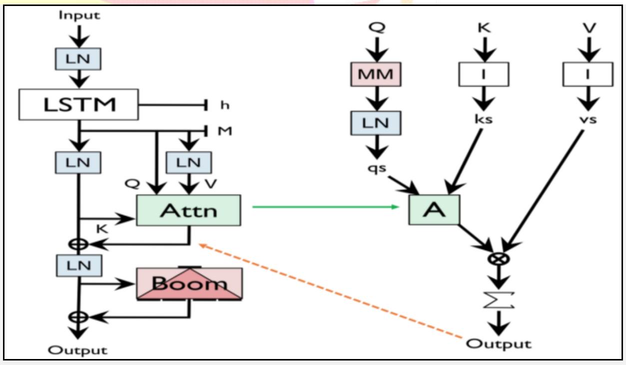
By implementing this idea on deeper networks, they are able to obtain a 28% relative improvement on the COCO object detection dataset. Deep residual nets are foundations of our submissions to ILSVRC & COCO 2015 competitions, where they also won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.

Single Headed Attention RNN: Stop Thinking With Your Head

The Harvard grad Steven Merity introduces a state-of-the-art NLP model called as Single Headed Attention RNN or SHA-RNN. Stephen Merity, an independent researcher that is primarily focused on Machine Learning, NLP and Deep Learning. The author demonstrates by taking a simple LSTM model with SHA to achieve a state-of-the-art byte-level language model results on *enwik8.*

The author's primary goal is to show that the entire field might have evolved in a different direction if we had instead been obsessed with a slightly different acronym and somewhat different results.

The central concept of the model architecture proposed by Steven consists of a **LSTM architecture with a SHA based network** with three variables (Q, K and V).



Source: Arvix

Each SHA-RNN layer contains only a single head of attention that helps with keeping the memory consumption of the model to the minimum by eliminating the need to update and maintain multiple matrices.

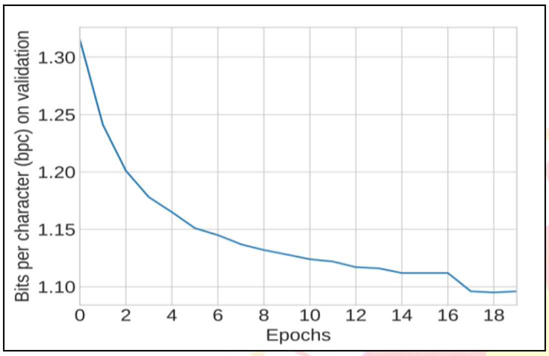
The Boom layer is related strongly to *the* **large feed-forward layer found in Transformers** and other architectures. This block reduces and removes an entire matrix of parameters compared to traditional down-projection layers by using Gaussian Error Linear Unit (GeLu) multiplication to break down the input to minimize computations.

Let's look at the actual comparison below. In 2016, The Surprisal-Driven Zoneout, a regularization method for RNN, achieved an outstanding compression score of 1.313bpc on the Hutter Prize

VOL- VIII | ISSUE- X | OCTOBER | 2021 | PEER REVIEW e-JOURNAL | IMPACT FACTOR 7.149 | ISSN 2349-638x

dataset, *enwiki8* which is a one-hundred-megabyte file of Wikipedia pages.

The SHA-RNN managed to achieve even lower (bpc) compared to the model in 2016. That is impressive. Bits per character is a model proposed by Alex Graves to approximate the probability distribution of the next character given past characters.



Further on, the Single Headed Attention RNN (SHA-RNN) managed to achieve strong state-of-the-art results with next to no hyper-parameter tuning and by using a single Titan V GPU workstation. And also, his work has undergone no intensive hyper-parameter tuning and lived entirely on a commodity desktop machine that made the author's small studio apartment a bit too warm to his liking. Now that's the passion for **Machine Learning**.

EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks:-

The authors systematically study model scaling and identify that carefully balancing network depth, width, and resolution can lead to better performance. A new scaling method that uniformly scales all dimensions of depth, width and resolution using a simple yet highly effective compound coefficient is demonstrated .

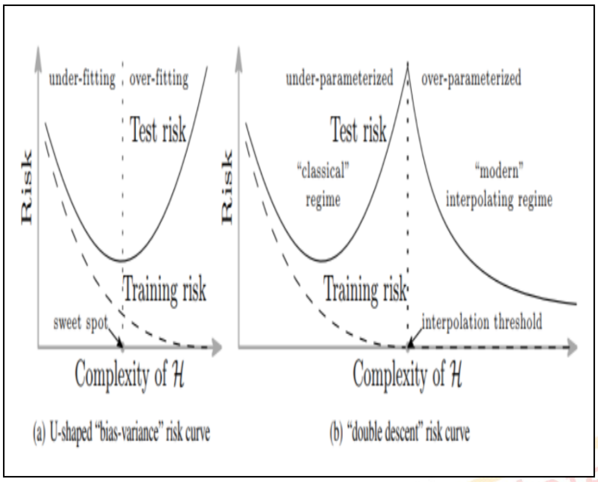
A network that goes through dimensional scaling (width, depth or resolution) improves accuracy. But the caveat is that the model accuracy drops with larger models. Hence, it is critical to balance all three dimensions of a network (width, depth, and resolution) during CNN scaling for getting improved accuracy and efficiency.

The ***compound scaling method*** as above consistently improves model accuracy and efficiency for scaling up existing models such as MobileNet (+1.4% Image Net accuracy), and ResNet (+0.7%), compared to conventional scaling methods.
Scaling doesn't change the layer operations; instead, they obtained their base network by doing a Neural Architecture Search (NAS) that optimizes for both accuracy and FLOPS. The scaled EfficientNet models consistently reduce parameters and FLOPS by an order of magnitude (up to 8.4x parameter reduction and up to 16x FLOPS reduction) than existing ConvNets such as ResNet-50 and DenseNet-169.

Efficient Nets also achieved state-of-the-art accuracy in 5 out of the eight datasets, such as CIFAR-100 (91.7%) and Flowers (98.8%), with an order of magnitude fewer parameters (up to 21x parameter reduction), suggesting that the Efficient Nets also transfers well.
The authors at OpenAI defines the effective model complexity (EMC) of a training procedure of a Neural Network as the maximum number of samples on which it can achieve close to zero training error. The experiments that were conducted suggests that there is a critical interval around the *interpolation threshold.*

Interpolation threshold means that the model is varied across the number of model parameters, the length of training, the amount of label noise in the distribution, and the number of train samples. The critical region is simply a small region between the under and over-parameterized risk domains.

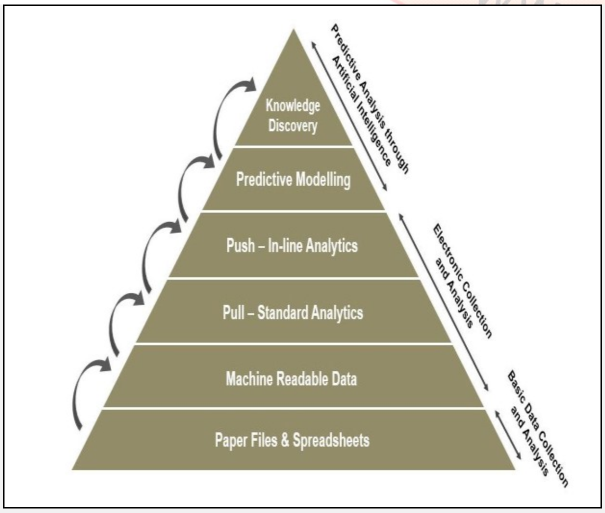
VOL- VIII | ISSUE- X | OCTOBER | 2021 | PEER REVIEW e-JOURNAL | IMPACT FACTOR 7.149 | ISSN 2349-638x

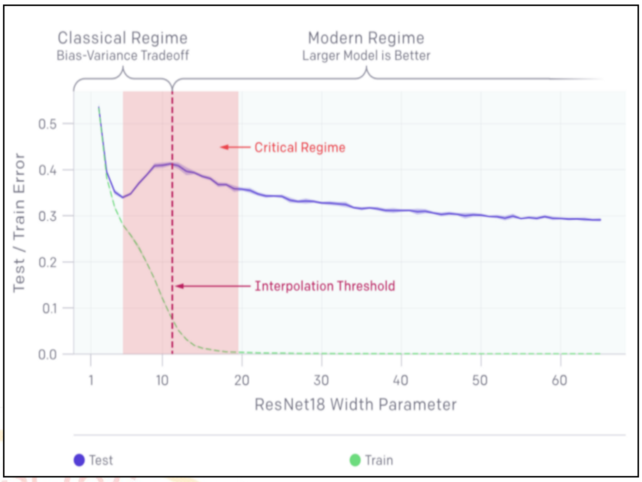


In most research, the *bias-variance trade-off* is a fundamental concept in classical statistical learning theory. The idea is that models of higher complexity have lower bias but higher variance.

Once a model complexity passes the critical interval, models overfit with the variance term dominating the test error, and hence from this point onward, increasing model complexity will only decrease performance called — double-descent phenomenon.

Financial Regulators make decisions about the health of financial institutions and their impact on the economy through compliance reports and financial market trends over a period of time. The speed and accuracy with which Regulators can sift through mountains of data, enables them to make timely and precise financial decisions. We are witnessing an explosion of data so vast that without the support of automation and artificial intelligence, efficient compliance will be a grave challenge.



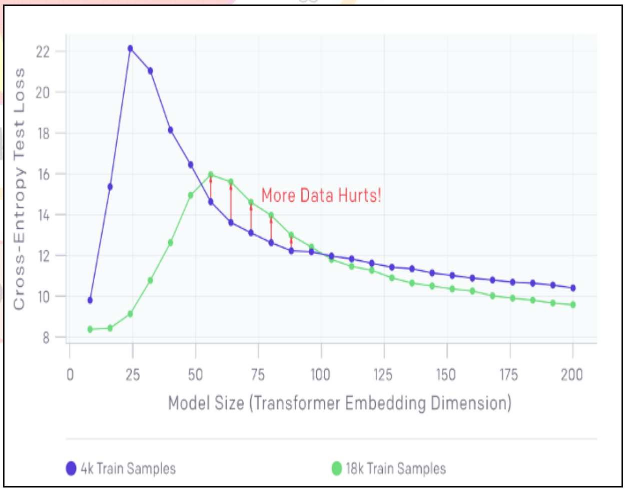


Model Regime

The demonstrate model-wise double descent occurrence across different architectures, datasets, optimizers, and training procedures.

The paper concludes that with the usual modifications that are performed on the dataset before training (e.g., adding label noise, using data augmentation, and increasing the number of train samples), there is a **shift in the peak in test error towards larger models**.

Also, in the chart above, the peak in test error occurs around the interpolation threshold, when the models are just barely large enough to fit the train set.



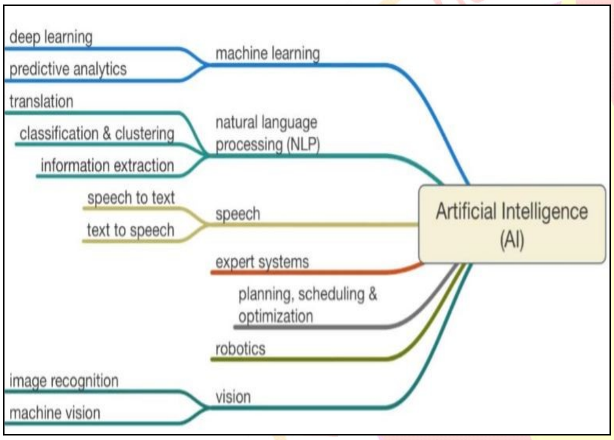
Sample Regime

In this section, the chart shows the effect of varying the number of training samples for a fixed model. Increasing the number of samples shifts the curve downwards towards lower test error but also shifts the peak error to the right.

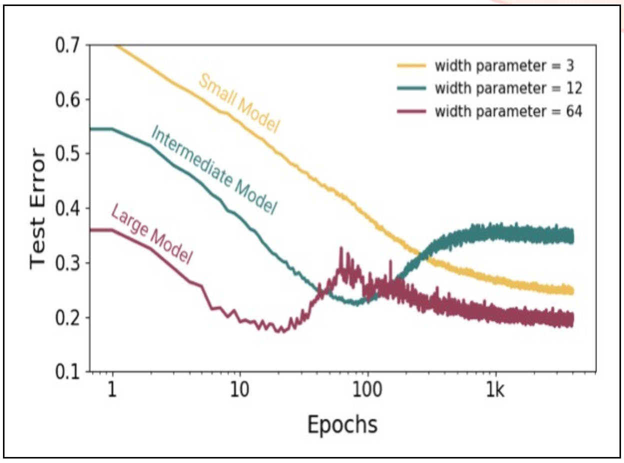
tool, the discovery of the uses of fire, forging metals, farming, the Industrial Revolution to the information age; we have been consistently inventing tools to make our lives easier.

Artificial Intelligence (AI) is one such tool. The term 'Artificial Intelligence' was first coined in 1956 by a Dartmouth University professor, John McCarthy. He defined Artificial Intelligence as:

"Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to stimulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves."



For a given number of optimization steps (fixed y-coordinate), test and train error exhibit model-size double descent. For a given model size as the training process proceeds, test and train error decreases, increases, and decreases again; we call this phenomenon epoch-wise double descent. Increasing the training time increases the EMC, thus a sufficiently large model transitions from under to over-parameterized throughout the full training phase.



For models at the interpolation threshold, there is effectively only one global model that fits the train data — and forcing it to fit even with small misspecified labels will destroy its global structure. The paper then concludes that there are *no good models* which both interpolate the train set and perform well on the test set. The characterization of these critical regimes, as stated above, provides a useful way of thinking for practitioners, **hopefully, to give a breakthrough in Machine Learning soon.**

4. Conclusions

Machine learning and AI complement each other, and the next breakthrough lies not only in pushing each of them but also in combining them. Our algorithms should support (re)trainable, (re) compostable models of computation and facilitate reasoning and interaction with respect to these models at the right level of abstraction. Multiple disciplines and research areas need to collaborate to drive these breakthroughs. Using computation as the common language has the potential for progressing learning concepts and inferring information that is both easy and difficult for humans to acquire.

To this end, the "Machine Learning and Artificial intelligence" section in Frontiers in Big Data welcomes foundational and applied papers as well as replication studies from a wide range of topics underpinning ML, AI, and their interplay. It will foster the scholarly discussion of the causes and effects of achievements providing a proper perspective on the obtained results. Using the common language of computation, we can fully understand how to achieve intelligent behavior in machines.

References:-

1. Blockchain is an example of a distributed ledger technology that allows electronic payments to be made without going through a financial intermediary.
   A new Fintech sector, combining the real estate market and technology
2. "Robots in Finance Bring New Risks to Stability, Regulators Warn. Silla Brush," Bloomberg Magazine. 1 Nov 2017. Available at: https://www.bloomberg.com/news/articles/2017-11-01/robotsin-finance-bring-new-risks-to-stability-regulators-warn

3. Asymptotic statistical properties stem from large sample theory and serve as a framework for statistical tests and estimated properties.
4. https://en.oxforddictionaries.com/definition/artificial_intelligence
5. Renaissance Technologies would eventually become famous for its financial signal processing techniques for use in pattern recognition.
6. Fair Isaac Corporation
7. Sponsored by The US Department of Treasury.
8. "17 minute trading glitch put Goldman's reputation on the line," Arash Massoudi and Tracy Alloway, Financial Times. 22 August 2013. Available at: https://www.ft.com/content/37fff9c6-0b36- 11e3-bffc-00144feabdc0
9. "Britain Urged to Take Ethical Advantage in Artificial Intelligence," John Thornhill, Financial Times. 16 April 2018.