IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 10, Issue 6, June 2020, ||Series -I|| PP 48-53

How Data Science Can Be Applied To Machine Learning - A Boom in Data Analytics

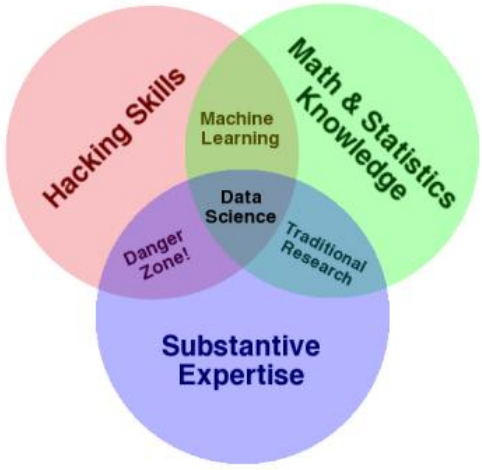
Isha Sharma Sharda
Research Analyst
Received 25 May 2020; Accepted 12 June 2020

I. INTRODUCTION

Intelligent computing technologies invented by human, have reached and jumped over the level of human accuracy level with time. Intelligence software on machine learning like Image recognition systems have reached the accuracy level from 72% to 96% from 2010 to 2015[1]. ML is a technology that works at the crossroads of data science, computer science and statistics as shown in figure 1. As a result , it is using essentials of all the three fields, collects, learn from data & patterns based on it for future predictions and action to be taken[2].

Almost every field in our daily life is occupied by machine learning applications. Healthcare, education, anomaly detection and pattern , voice recognition are some of the examples of Machine learning. According to Investors Guide to Artificial Intelligence published on 2017, by 2021, The Business Intelligence (BI) & analytics market within, Data Science objectives that support machine learning are envisioned to grow at a 13% CAGR.

Machine learning technology is evolved from Artificial Intelligence. Machine learning is a separate technology that basically mimics data into human learning. It helps in increasing the business values brands like Amazon , facebook and flipkart by using learning algorithms[3]. Elaborately, ML uses past huge data and deduce its pattern , applies algorithm to dig out best possible results. As a result, it enhances the business value by predicting the future outcome and forthcoming situation[4].



Machine learning within Data Science[5]

In terms of its role in business , machine learning helps in increasing digital business on the basis of data. ML covers mainly two types of analysis like exploratory (Unsupervised learning) in which algorithms are used to discover structure within data as shown in below figure 2.For example experience in years in industry and salaries are correlated. The other is predictive learning (supervised learning) in which algorithms works on predicting on the basis of past experiences. for example: A company wants to predict his future clients between the scenario. On the basis of few characteristics such as past clients and their size, field and customers , they can predict which companies can be possibly be their future clients[5].

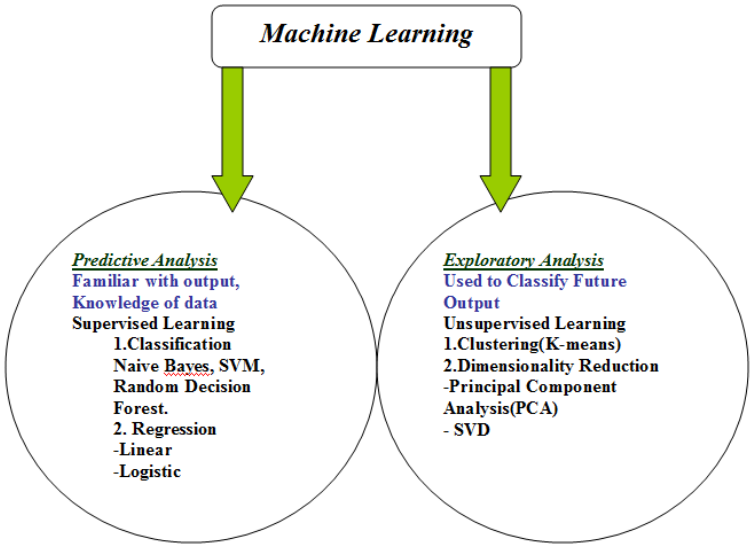


Figure 2. Machine Learning Life cycle.

Machine learning is successfully flourishing in areas of business like predicting financial results, fraud detection, spam filtering, risk assessment, threat detections systems, financial transactions, underwriting in insurance, bioinformatics, biometrics and identity management, manufacturing assembly[6].

The stages of ML covers a part of some stages of data science life cycle as shown in below figure 3.

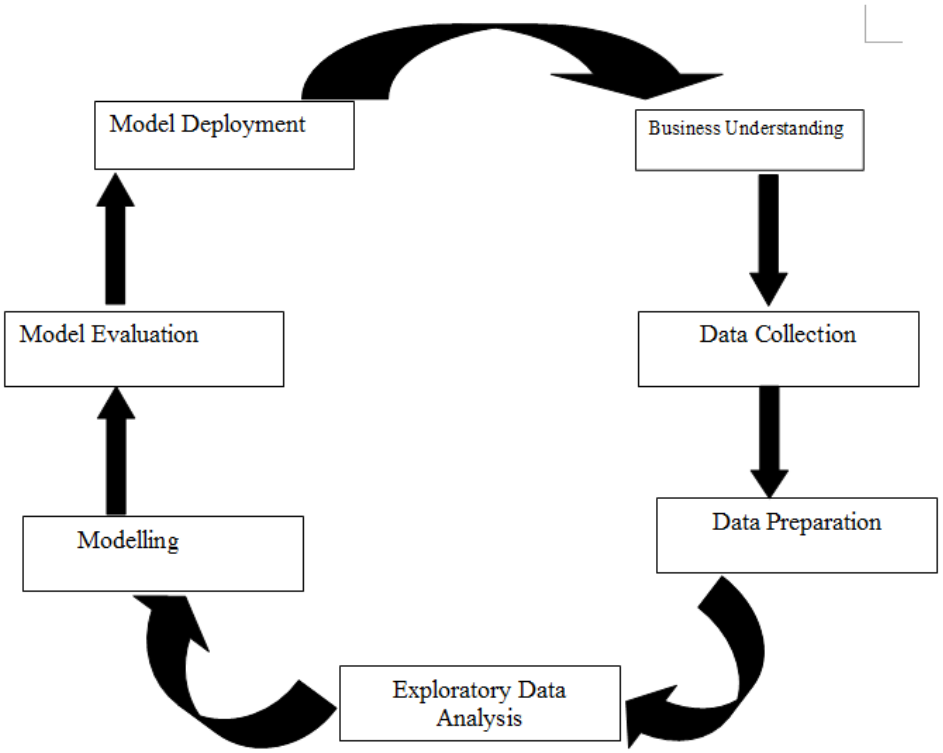


Figure 3. Data science life cycle.[7]

In other words, Machine Learning allows us to program how we make an outline instead programming explicitly. This provides numerous opportunities for data science and developer teams to develop product offerings, customer relationships, advertising, and marketing, process improvement, and much more[7].When starting out with ML, experience can be gained by obtaining few algorithms from the marketplace and

deploying them in the cloud with some data to perform experiments. These may uncover promising avenues for future business value, and eventually be expanded into formal ML deployments[8].

How Machine learning within Data Science is helping organization:

As shown in above figure 3, from business point of view , the project is executed in the same way. The above data science lifecycle is designed in way to complete a project that boats intelligence applications.
For its completion artificial intelligence as well as machine learning are deployed. Taking example of Facebook's machine learning algorithms gather behavioural information for every user on the social platform[9]. Lets discuss each step of data cycle:

1.Business Understanding: In this you need to understand the requirement of your project like you want to increase sales, or which product you need to sell etc. For this you need to uncover every single factors important for project completion[10].

2.Data Collection : By this we mean which factors are need to take into consideration for increasing business of sales like store location, its staff, products in store, store opening hours and many more.

3.Data Preparation : This stage means to clean data (removing noise and unwanted data) prepares it for the another stage of analysis[11].

4.Exploratory Data Analysis: In this phase , we look up for new insights into the data and explore which patterns and values the input data contains.

5. Modelling: . In this you will select the modelling technique. If finding any target value is qualitative apply classification under supervised learning and if quantitative than apply regression. If the analysis is non-target i.e. visualization for high dimensions applications then you can use PCA method under unsupervised learning. or if finding hidden structure with data set then use clustering[5]. In this you have to repeat the modelling assessment or a collection of models till you find a satisfactory result[5] .

6. Model Evaluation: In this stage the final model is evaluated against the test data and its accuracy is seen. This factor simply assess the performance of your model.

7. Model Deployment: After model evaluation , the results are taken and an approach is planned to deploy the model. After the end of project , the report is the only thing you need to update and monitor on regular rhythm[12].

Future of Data Science applying Machine Learning :
Platform Language-Python & Tensor Flow/ Kotlin

Python being beginner friendly and an easy scripting language when looking for optimizing performance and when you can interfere with low level libraries of C. It was a satisfactory language for building the end to end entire system. But in 2018 lattner announced Tensor flow would soon be supporting Swift. Swift is viable if you want to create national apps for iPhone or iPad. Therefore, you'll need java ,if you want to create apps for Android[13] .

Python proved less efficient where the challenge is to amend applications for slow, yet more efficient processors with less memory, for mobile devices where power &heat were main concerns.

Lattner and his team are glancing all of these concerns with Swift for TensorFlow. The syntax is almost same as Python. It has an fast interpreter for scripting and notebooks. Above all, they are providing the facility to run random Python code to help out migration and because by default , Swift is now the easiest choice for iOS app development, deploying to mobile.
Under the cover, Swift's open source compiler and static typing make it possible to aim particular AI chipsets at the build step[13].

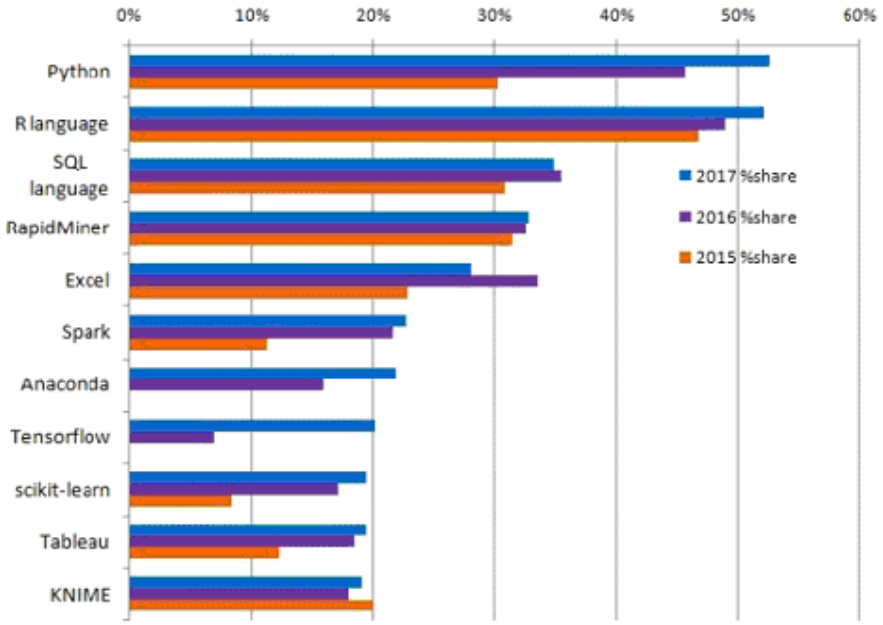


Figure 4. Analytics/Data Science 2017 Software Platforms polls [KDnuggets.com]

The above pie chart in figure 4. shows top 11 tools used in machine learning and data science projects ,which passed the threshold of 500 votes.

If we talk about multi-platform programming language that works on browser, mobile then kotlin is the best choice. Kotlin is short, secure, tool-friendly & interoperable. This programming language is statically-typed that runs on the Java virtual machine(JVM) and can also be compiled to the JavaScript source code or can use the LLVM compiler infrastructure.[14] But kotlin is also facing issues in data science like in this data structures have to be defined clearly which can add additional steps in working with data. Second drawback is numerical efficiency as boxing of number might create problems in libraries. In libraries and tooling Java libraries use java instead of kotlin in documentation. Width of data science libraries is upright but does not match python. Ad-hoc Analysis is another issue in which exploring data without any clear objective may be difficult without any data frame library called krangl. [*15]*

Future Prospects of Machine Learning for Data Science

Machine Learning(ML) and Data science are not unimaginable mechanisms instead when combined in a sequence of steps they provide influential solutions. In a latest article of August 2018, A machine learning model is applied to data science problem for knowing the reasons behind poverty in costa rica. After breaking the problem into component steps , ML is just easy to implement in simple steps .Researchers tried out experimenting techniques in the following steps as explained above in this article:

Our approach followed a sequence of processes:

1. Understand the problem and data
2. Perform data cleaning alongside exploratory data analysis
3. Engineer relevant features automatically and manually
4. Compare machine learning models
5. Optimize the best performing model
6. Interpret the model results and explore how it makes predictions

In model optimization is like searching for best settings for running machine learning model. As quoted by researchers that machine learning still lacks explainability gap which hinders its applicability as people wants explanation of its predictions.

Finally researchers , tried exploratory techniques and came out with surprising results. The techniques are Recursive feature elimination and uniform manifold approximation for dimension reduction (which uses many unsupervised machine learning algorithms) and visualization. Therefore At the end , researchers realized

that if you are not satisfied with the outcome, switch onto another exploratory techniques , practice them on your model as much to reach on particular decision[16].

Another field that will benefit is healthcare & pharmacy technology, Doctors can predict on the basis of past data like medical history to DNA and wearable fitness tracker, the evolution of certain diseases or health complications in patients. In Magnetic Resonance Imaging(MRI) , Data science and ML Algorithm are used to improve the image quality, extract information efficiently and providing accurate interpretations.[17].

Some of the future projects are Quantum computing, cognitive services, Collaborative learning, better unsupervised algorithms where the researchers are allowing learning algorithm to find on its own the hidden structure within input data and wont provide any labels.[18] . Many Small and flagship projects are going on as reported by stanford Data Science Inititative[19].

Machine learning was quite expensive to get to to project's budget, but today, public cloud providers offers machine learning services reasonable. Most cloud providers, such as Amazon, AWS, Google, provide support for three types of predictions. First is , Binary prediction bring into being in business processes such as credit check systems, order processing, and engines used to recommend music, videos, or other products to users based upon gathered data and learned responses, Second is ,Category prediction- This prediction is applied in Finance, manufacturing, and retail sector. Third one is Value Makes predictions in terms of future selling. This information helps to know, and save alot during manufacturing, planning and perhaps cut down on travel expenses as sales team follow up on leads. If you use a machine-learning system on cloud 1, then the data storage mechanism will normally be supported on cloud 1 as a native. However, your project database is not supported if you not provide data integration between your on-premises data storage system and those in the cloud. But if you're working with hybrid- or multi-cloud deployments then the separation of the data from the machine-learning engine will be challenging in terms of cost, performance, and usability.[20] Programmatic commerce is a new approach to bring advertising field to a new level using machine learning. It is collective use of automation, rich data and analytics to guess customers experience and company's ROI(return on Company's Investments). It is estimated that global programmatic market would grow from 14.2 billion to 36.8 billion from 2015 to 2019[21]. Sanzu, another standard for Data Science field, that will evaluate data. Till now there was only structure and data systems to check the work flow but under sanzu, various micro and macro level models will evaluate data sets on individual or real time cases[22]. Statistical Analysis of Soil fertility is inimitable application of Data science & ML. ANOVA and Principal Component Analysis(PCA) is applied on three thousand and eight hundred samples of soil using R software and came out with valuable results. It showed the concentration of different nutrients present in soil [23].A common ML assumption is that algorithms can learn better with more data and consequently provide more accurate results. Some work is done to protect , improve and increase the privacy using Map Reduce. Airavat is a method that is designed to executes trusted and untrusted Mapreduce computations on susceptible data. Still some challenges is needed to overcome in airavat provider reducer for working only in one domain[24]

II. SUMMARY

Data, nowadays is blowing up at an extraordinary rate from social media, mobile & medical devices, satellites and web technologies. Traditional approaches for data handling and extracting business values is not helping out these days. The prime reason for using techniques of ML and Data science is to discover large data, grab knowledge and making better decisions[25]. The McKinsey Global Institute has assured that Machine Learning will be one of the major technology of the Big Data revolution[26]. With the supply of Cloud-based IT services was a good start to make advanced Data Science a typical step, and now with Cloud and packaged algorithms, mid-sized and smaller businesses will have way in to Self-Service Business Intelligence and Analytics.

REFERENCES:

[1]. The Economist. 2016 From not working to neural networking. See http://www.economist.com/news/specialreport/21700756-artificial-intelligence-boom-based-old-idea-modern-twist-not (accessed 22 March 2017).
[2]. Shalev-Shwartz S, Ben-David S. 2014 Understanding machine learning: from theory to algorithms. Cambridge, UK: Cambridge University Press.
[3]. Ayon Dey / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (3) , 2016, 1174-1179.
[4]. Yogita , Anand Sharma, Monitoring network security from machine learning, International journal of computer & mathematical science, IJCMS, ISSN-2347-8527, Vol-7, Issue-2, Februrary-2018.
[5]. Edouard Duchesnay, Tommy Löfstedt, Statics & ML in python, Release 0.2, june22,2018
[6]. Machine learning: the power and promise of computers that learn by example Issued: April 2017 DES4702 ISBN: 978-1-78252-259-1

[7]. https://dataschool.com/data-science-life-cycle/
[8]. Preparing and Architecting for Machine Learning, Carlton E. Sap, 17 Janurary 2017, ID: G00317328, Technical professional Advice.
[9]. https://www.simplilearn.com/data-science-vs-data-analytics-vs-machine-learning-article
[10]. Rohan Joseph, Foundation in Data Analytics, https://dataschool.com/data-science-life-cycle/
[11]. https://www.sv-europe.com/crisp-dm-methodology/
[12]. http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html
[13]. https://heartbeat.fritz.ai/why-data-scientists-should-start-learning-swift-66c3643e0d0d.
[14]. https://kotlinlang.org/docs/books.html.
[15]. https://opensourceforu.com/2018/07/kotlin-a-language-for-modern-multi-platform -applications/]
[16]. [https://towardsdatascience.com/a-data-science-for-good-machine-learning-project-walk-through-in-python-part-two-2773bd52daf0.
[17]. Meghann chilcott, Rise of machines: future of machine learning & data science, forbes technology council, 30 july 2018
[18]. Stefan Franczuk, https://dzone.com/articles/5-predictions-about-the-future-of-machine-learning, dec 2017
[19]. https://sdsi.stanford.edu/sites/default/files/data_science_research_at_stanford_2017-18_0.pdf
[20]. https://techbeacon.com/machine-learning-cloud-how-it-can-help-you-right-now
[21]. Nico Neuman, The Power of Big Data and Algorithms for Advertising and Customer Communication, IWBIS 2016 978-1, DOI 978-1-5090-3477-2/16 IEEE, 2016
[22]. Alex Watson, Deepigha Shree Vittal Babu, and Suprio Ray, Sanzu: A Data Science Benchmark, 2017 IEEE International Conference on Big Data (BIGDATA), DOI 78-1-5386-2715-0/17, IEEE 2017
[23]. R. Ajith kumar1,2, M.K. Muhammed Aslam2 et. al, A Statistical Analysis of Soil Fertility of Thrissur District, Kerala, *2016 IEEE International Conference on Data Science and Engineering (ICDSE)*
[24]. K. Grolinger, M. Hayes, W. A. Higashino, A. L'Heureux, D. S. Allison, M. A. M. Capretz, "Challenges for Map Reduce in big data", *Proc. IEEE World Congr. Services (SERVICES)*, pp. 182-189, Jun. 2014
[25]. O. Y. Al-Jarrah, P. D. Yoo, S. Muhaidat, G. K. Karagiannidis, K. Taha, "Efficient machine learning for big data: A review", *Big Data Res.*, vol. 2, no. 3, pp. 87-93, Sep. 2015
[26]. [M. James, C. Michael, B. Brad, B. Jacques, Big Data: The Next Frontier for Innovation Competition and Productivity, New York, NY:McKinsey Global Institute, 2011

Isha Sharma. "How Data Science Can Be Applied To Machine Learning - A Boom In Data Analytics." *IOSR Journal of Engineering (IOSRJEN)*, 10(6), 2020, pp. 48-53.