ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.5

Review Paper on Importance of Data Science in 2020

Dhanashree Kulkarni[1], Ashish Kulkarni[2]
Assistant Professor[1], Business Analyst[2]
Department of E&TC[1]
DYPIET, Pimpri, Pune, India[1]
Xoriant Technologies, Pune, India[2]

Abstract:
Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it. Many researchers also think it is the best way to make progress towards human-level AI.

Keywords: Artificial Intelligence (AI)

I. INTRODUCTION

What is Machine Learning?

The term machine learning refers to the automated detection of meaningful patterns in data. In the past couple of decades it has become a common tool in almost any task that requires information extraction from large data sets. We are surrounded by a machine learning based technology: search engines learn how to bring us the best results (while placing profitable ads), anti-spam software learns to filter our email messages, and credit card transactions are secured by software that learns how to detect frauds. Digital cameras learn to detect faces and intelligent personal assistance applications on smart-phones learn to recognize voice commands. Cars are equipped with accident prevention systems that are built using machine learning algorithms. Machine learning is also widely used in scientific applications such as bioinformatics, medicine, and astronomy. One common feature of all of these applications is that, in contrast to more traditional uses of computers, in these cases, due to the complexity of the patterns that need to be detected, a human programmer cannot provide an explicit, fine detailed specification of how such tasks should be executed.

When do we need Machine Learning?

Two aspects of a given problem may call for the use of programs that learn and improve on the basis of their experience": the problem's complexity and the need for adaptivity. Tasks that are too complex to program.

- Tasks Performed by Animals/Humans: There are numerous tasks that we human beings perform routinely, yet our introspection concerning how we do them is not sufficiently elaborate to extract a well Introduction defined program. Examples of such tasks include driving, speech recognition, and image understanding. In all of these tasks, state of the art machine learning programs, programs that \learn from their experience," achieve quite satisfactory results, once exposed to sufficiently many training examples.
- Tasks beyond Human Capabilities: Another wide family of tasks that benefit from machine learning techniques are related to the analysis of very large and complex data sets: astronomical data, turning medical archives into medical knowledge, weather prediction, analysis of genomic data, Web search engines, and electronic commerce. With more and more available digitally recorded data, it becomes obvious that there are treasures of meaningful information buried in data archives that are way too large and too complex for humans to make sense of. Learning to detect meaningful patterns in large and complex data sets is a promising domain in which the combination of programs that learn with the almost unlimited memory capacity and ever increasing processing speed of computers opens up new horizons.
- Adaptivity- One limiting feature of programmed tools is their rigidity once the program has been written down and installed, it stays unchanged. However, many tasks change over time or from one user to another. Machine learning tools programs whose behavior adapts to their input data offer a solution to such issues; they are, by nature, adaptive to changes in the environment they interact with. Typical successful applications of machine learning to such problems include programs that decode handwritten text, where a fixed program can adapt to variations between the handwriting of different users; spam detection programs, adapting automatically to changes in the nature of spam e-mails; and speech recognition programs.

Applications of Data Science:

- **Internet search:**

Search engines make use of data science algorithms to deliver best results for search queries in fraction of seconds.

- **Digital Advertisements:**

The entire digital marketing spectrum uses the data science algorithms - from display banners to digital billboards. This is the mean reason for digital ads getting higher CTR than traditional advertisements.

- **Recommender Systems:**

The recommender systems not only make it easy to find relevant products from billions of products available but also add a lot to user experience. A lot of companies use this system to promote their products and suggestions in accordance to the user's demands and relevance of information. The recommendations are based on the user's previous search results.

Applications of Data Analysis:

- **Healthcare:** The main challenge for hospitals with cost pressures tightens is to treat as many patients as they can efficiently, keeping in mind the improvement of quality of care. Instrument and machine data is being used increasingly to track as well as optimize patient flow, treatment, and equipment use in the hospitals. It is estimated that there will be a 1% efficiency gain that could yield more than $63 billion in the global health care savings.

- **Travel:**

Data analytics is able to optimize the buying experience through the mobile/ web log and the social media data analysis. Travel sights can gain insights into the customer's desires and preferences. Products can be up-sold by correlating the current sales to the subsequent browsing increase browse-to-buy conversions via customized packages and offers. Personalized travel recommendations can also be delivered by data analytics based on social media data.

- **Gaming:**

Data Analytics helps in collecting data to optimize and spend within as well as across games. Game companies gain insight into the dislikes, the relationships, and the likes of the users.

- **Energy Management:**

Most firms are using data analytics for energy management, including smart-grid management, energy optimization, energy distribution, and building automation in utility companies. The application here is centered on the controlling and monitoring of network devices, dispatch crews, and manage service outrages. Utilities are given the ability to integrate millions of data points in the network performance and let the engineers to use the analytics to monitor the network.

To become a Data Scientist:

- **In-depth knowledge of SAS and/or R:** For Data Science, R is generally preferred.
- **Python coding:** Python is the most common coding language that is used in data science along with Java, Perl, C/C++.
- **Hadoop platform:** Although not always a requirement, knowing the Hadoop platform is still preferred for the field. Having a bit of experience in Hive or Pig is also a huge selling point.
- **SQL database/coding:** Though NoSQL and Hadoop have become a major part of the Data Science background, it is still preferred if you can write and execute complex queries in SQL.
- **Working with unstructured data:** It is most important that a Data Scientist is able to work with unstructured data be it on social media, video feeds, or audio

To become a Data Analyst:

- **Programming skills:**

Knowing programming languages are R and Python are extremely important for any data analyst.

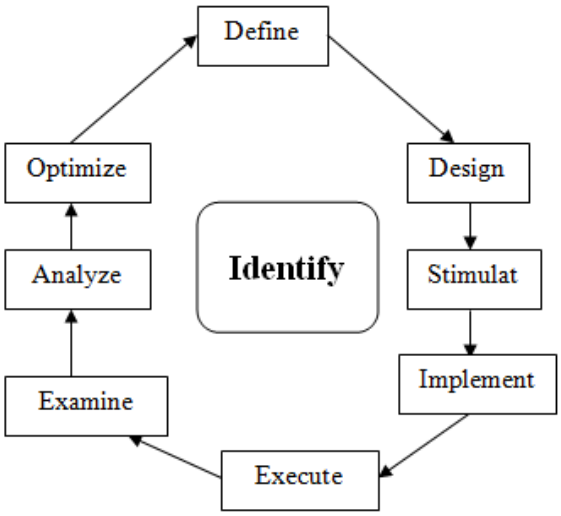
- **Statistical skills and mathematics:** Descriptive and inferential statistics and experimental designs are a must for data scientists.
- Machine learning skills
- **Data wrangling skills:**

The ability to map raw data and convert it into another format that allows for a more convenient consumption of the data.

- Communication and Data Visualization skills
- **Data Intuition:**

It is extremely important for professional to be able to think like a data analyst.

II. FLOWCHART:



III. REFERENCES

[1]. Fisher, D.H.(1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning, 2,* 139-172.

[2]. Quinlan, J.R. (1986). Induction of decision trees. *Machine Learning, 1,* 81-106.

[3].Rivest, R.L. (1987). Learning decision lists. *Machine Learning,2,* 229-246.

[4]. Sonal. A. Mishra1, Dhanashree S. Tijare2 and Dr. G. M. Asutkar "DESIGN OF ENERGY AWARE AIR POLLUTION MONITORING SYSTEM USING WSN".