

LITERATURE SURVEY

1. Data Science: Literature Review & State of Art Sanket Mantri(2016):

Since the evolution of internet the amount of data produced each minute is increasing substantially. In order to manage this data there has to be some mechanism. Thousands of terabytes produced each year need to be stored securely and should be accessible easily, this needs high amount of research and development. This new concern has given birth to field known as “data science” and people “data scientist”. Today data scientists are working hard to develop new solutions to process and store large amount of data using new techniques. Another major concern today is efficient use of available energy resources, data science helps immensely to predict the usage of resources and minimizing the wastage. This paper is about data science, current and future development in this field. The primary focus is on literature review of data science and various applications of data science.

2. Applied Data Science (lessons learned for data driven business) 2019:

“Data science” is a term that has entered public perception and imagination only since the first half of the decade. Even in the expert community, fundamental treatments such as “What Is Data Science?” (Loukides 2010) were first published as recently as 2010. Yet, the substance of what constitutes data science has been built up for much longer. An attempt to define the term “data science” can follow either a top-down or a bottom-up philosophy. On the one hand, looking “top-down,” data science is the research field that studies mechanisms and approaches necessary to generate value and insights from data, enabling the building of data products. Importantly, a “data product” is not just a product “dealing” with data, but it is a product deriving its value from the data and producing data itself (Loukides 2010). On the other hand, adopting the “bottom-up” view, data science is an interdisciplinary research field (Stockinger et al. 2015) that adopts a new, holistic way of exploiting data, looking beyond single aspects such as how to store data, or how to access it.

3. Ian Langmore Daniel Krasner - Columbia Applied Data Science:

One neat way we like to visualize the data science skill set is with Drew Conway's Venn Diagram[Con], see figure 1. Math and statistics is what allows us to properly quantify a phenomenon observed in data. For the sake of narrative let's take a complex deterministic situation, such as whether or not someone will make a loan payment, and attempt to answer this question with a limited number of variables and an imperfect understanding of those variables influence on the event we wish to predict. With the exception of your friendly real estate agent we generally acknowledge our lack of soothseer ability and make statements about the probability of this event. These statements take a mathematical form, for example

4. INTRODUCING DATA SCIENCE BIG DATA, MACHINE LEARNING, AND MORE, USING PYTHON TOOLS (DAVY CIELEN, ARNO D. B. MEYSMAN ,MOHAMED ALI) 2016:

It's in all of us. Data science is what makes us humans what we are today. No, not the computer-driven data science this book will introduce you to, but the ability of our brains to see connections, draw conclusions from facts, and learn from our past experiences. More so than any other species on the planet, we depend on our brains for survival; we went all-in on these features to earn our place in nature. That strategy has worked out for us so far, and we're unlikely to change it in the near future. But our brains can only take us so far when it comes to raw computing. Our biology can't keep up with the amounts of data we can capture now and with the extent of our curiosity. So we turn to machines to do part of the work for us: to recognize patterns, create connections, and supply us with answers to our numerous questions.

5. Hands-On Data Science and Python Machine Learning:

Let's take an example around unsupervised learning. Say I was clustering people instead of balls and dice. I'm writing a dating site and I want to see what sorts of people tend to cluster together. There are some attributes that people tend to cluster around, which decide whether they tend to like each other and date each other for example. Now you might find that the clusters that emerge don't conform to your predisposed stereotypes. Maybe it's not about college students versus middle-aged people, or people who are divorced and whatnot, or their religious beliefs. Maybe if you look at the clusters that actually emerged from that analysis, you'll learn something new about your users and actually figure out that there's something more important than any of those existing features of your people that really count toward, to decide whether they like each other. So that's an example of supervised learning providing useful results.