Topic Review: Data Science

Name: Deeksha Rao Gorige
Student Id: 1970751

Summary of the Paper:

This is a summary of the article "Data Science and Prediction" by Vasant Dhar. In this article, Vasant Dhar begins by addressing that use of "data science" is increasingly common, as is "big data." The article mainly focuses on-

- What exactly is Data Science and what is unique about it?
- What are the skills data scientists require to be productive in the world full of data?
- What are the implications for scientific inquiry?

All these questions are answered from the perspective of predictive modeling. He started his article with the basic understanding of what data science is. He later on explained about what makes an insight actionable and introduced a term "Predictive Power" and stated that requirement on predictive accuracy on observations that will occur in the future is a key consideration in Data Science.

He then explained about the implications. In the implications he started off with how data storage has become inexpensive and how large amounts of data are being used for the decision making. Later on, he explained how machine learning works on a large trove of data. This was explained in a very neat manner by taking an example of health-care database of individuals in which a group of people have been diagnosed by diabetes and some other subset of this group has developed complications. He elaborated this example and explained whether there are any patterns to the complications and whether the probability of complications can be predicted and acted upon. He later on made a pattern for getting the complication rate but the question which raised here is whether this is a pattern that is robust. Thus, introduced a new term robustness which is addressed extensively in Machine Learning and is a key point which the Data scientists should keep in mind.

In the future discussion of his paper he stated what are the skills that a data scientist should have in order to dealt with the large data sets. He mainly discussed about these skills-
- Statistics
- Skills that come from Computer Science
- Knowledge of correlation and causation
- The ability to solve problems in a very effective way

All these were explained in detailed and why a Data scientist requires all the above skills. He later on dived us into Knowledge discovery and what makes predictive models accurate i.e. what are the errors which we make in prediction. He discussed about three errors-
- Misspecification of Model
- Using a sample to estimate the model
- Randomness

He finally concluded by saying that Predictive modeling and machine learning are very important for any business organizations. But organizations and managers face significant challenges in adapting to the new world data. This requires a shift in organizational culture, one seen in organizations that have embraced the emerging world of data for decision making.

Strengths: A strong and interesting article on data science. The examples were clearly explained, and it was easy to understand the paper. The skills needed for a data scientist was the best part. The article was explained in a sequential manner. That is first he started off with what is data science and later the things which come under data science.

Weaknesses: Some of the topics where left incomplete and it was a sort of discontinuation.

Feedback: I felt it would be more interesting if the author has illustrated the concepts with more examples diagrammatically.