Maninderpreet Singh Puri- PSD Assignment 3

by Maninderpreet Singh Puri

In submitting your work, you are consenting that it may be copied and transmitted by the University for the detection of plagiarism.

"This is my own work. I have not copied any of it from anyone else."

Challenges of Big Data: Revisioning Staistical Eductors

"Implications of the Data Revolution for Statistical Education" by Jim Ridgway describes how the data revolution is transforming statistical education. He explains that technological advancements have opened new dimensions of data, resulting in huge volumes of structured and unstructured data. This brings new challenges for statistical educators as they have to mould students to acquire skills to pre-process data, model data mining techniques and interpret statistics. Although many factors are influencing modern statistics, Big Data and open data are at the heart of this revolution.

'Big Data' is the term used for the unstructured class of data acquired from devices like sensors and mobile phones, present in large volumes and is mostly spread over some time. Ridgway defines Big Data as "data that arise naturally from some system and are a by-product of the system"(p529). These data sets are collected by organizations for learning the behavior of their target audience or for learning patterns to predict associations. In contrast, 'open data' is more structured and is generated for purposes like surveys where parameters of data are defined and examined. For the most part, Ridgway elucidates on the problems posed by Big Data in terms of collection, storage and presentation.

Initially, a statistical problem starts with developing a model that fits the data according to its attributes. However, Ridgway warns that there can be negative consequences of using the same model for data with different 'distribution'(p530). Distribution here means the way data is spread across a scale of parameters, for instance, the number of occurrences of rain in a year. He supports this by giving an example of an event where George W. Bush introduced tax cuts and misclaimed the reduction in tax because of using an inappropriate model (p530-531). Especially with Big Data, this problem is even more severe. He states that this should be tackled using modern statistical analysis and suggests addressing it in the statistical curriculum.

Add to the references list.

Ridgway explains the problems with Statistics Curricula in terms of lack of awareness among students about the problem-solving approach rather than just performing calculations. He explains this idea by comparing the statistical curriculum for schools in England and Whales to New Zealand. Students in New Zealand learned how to use statistics which made them continue their studies and advance in statistics whereas the students in England and Whales had limited knowledge of statistics due to less emphasis on problem-solving skills (p535). Besides the intricacies in the statistical curriculum, he discusses the biases in statistical ethics.

The quality of sampling is another aspect which is the cause of faulty estimates of systems. Ridgway supports this with an example of the Netherlands where traffic lights which produced 80 million records each day and detected the number of vehicles particularly trucks was helpful in road planning (p538). However, many factors like location of sensors, small sample size and malfunction contributes to biases. External influences such as corruption can also alter the measurements for example websites selling 'likes'

and 'followers' on social networks (p538). This makes predictions hard based on such exaggerations in data. Moreover, with an increase in data dimensions, its association is another wide area of interest.

I am intrigued by the skill of finding associations between unrelated data and discovering new applications. To learn this skill I will have to learn data mining techniques and dig deep into pattern finding methods like Apriori, K-mean, classification using NN and so on. I see myself working as a Data Scientists in coming years to predict and analyse Big Data as data is exponentially growing and the utilization of data will be of high value.

To conclude, methods of teaching need to be remodelled so that the students understand the underpin laws of statistics. Care should be taken so that fewer biases are introduced and multiple data sources should be used. Awareness about data revolution can drive the ambitions of students to bring further developments in Data Science.

Refernces:

Ridgway, J. (2015, July 29). *Wiley Online Library*. Retrieved May 15, 2020, from Wiley Online Library: https://onlinelibrary.wiley.com/doi/full/10.1111/insr.12110

Word count (720)

Maninderpreet Singh Puri- PSD Assignment 3

Maninderpreet Singh Puri- PSD Assignment 3

FINAL GRADE

GENERAL COMMENTS

Instructor

92/100

PAGE 1

QM **Great work**

Text Comment. Add to the references list.

Text Comment. =====

Text Comment. =====

Text Comment. =======

PAGE 2