

HOMEWORK 1

01. Summary of articles.

- **Statistical Modeling: The Two Cultures**

The article highlights the problems of the two generalized approaches to data science analyses (i.e. The Data Modeling Culture and The Algorithmic Modeling Culture) and proposes a comprehensive road map towards a solution that intends to improve the applicability of data science concepts and also intends to diversify the tools used for such analyses. The article then highlights problems associated with both data modeling and algorithmic modeling approaches to statistical modeling. A major issue with data modeling is the uncritical nature of scientists who adopt and assume the model based entirely on the data they collect. Such an assumption becomes problematic as it leads to mischaracterization of data by in large and misleading conclusions. This issue is summed up using a popular idiom *"If all a man has is a hammer, then every problem looks like a nail."* The author then highlights the predictive accuracy of algorithmic models compared to data models. He furthers his point by implying that algorithmic models provide a better understanding of the underlying mechanisms. Algorithmic models are stated to have more realistic results as it involves describing the data against a theory and not patterns. He ends the article by optimistically predicting a cross discipline data sharing and migration of statisticians towards a more algorithmic based approach to problem solving.

- **50 Years of Data Science**

The article presents a compilation of the history of Data science and reviews important understanding regarding the subject, all while predicting what the field of data science will look like 50 years from now. The paper also presents potential of data science in virtually every field of science and study where data is being actively recorded and stored. Written nearly half a decade after the economic crisis of 2008 the author explores its potential applications and how data science helped companies and individuals recover from the economic devastation. The article also evaluates data analysis techniques and their evolution over time. The article predicts fifty years from now, empirical data analysis results will overrule the mathematical requisites. This is attributed primarily to the availability of big data and accuracy of data measurement techniques. Another intriguing prospect of the field is predicted as the development and application of a global data science methodology. Such a unified methodology will radically change how we collect and incorporate data in our analyses. The article ends in a reiteration of the definition

of data science as *“the science of learning from data; study of methods of analysis and processing data; improvement of said methods,”* and a prediction regarding how data science can help improve the data interpretation and concurrently improving the methodology of analysis is made. This self-improvement cycle is predicted to radically change the scientific approach to data with a predicted paradigm shift that will bring data science at the forefront of scientific explorations.

02. Misinformation on Covid-19

- **Articles on “COVID – 19 Related Infodemic and its impact on Public Health: A Global Social Media Analysis”**

The article highlights the correlation between misinformation on social media platform and impact of Covid-19 on public health. The article also attempts to evaluate different methodologies to collect and analyze social media data in relation with the spread of misinformation. The misinformation is then categorized into Rumor, Stigma and discrimination and conspiracy theories. Furthermore, these were specifically related to Covid-19 as misinformation regarding cause of the disease, illness, treatment, interventions, violence and miscellaneous. The Spread of misinformation was observed to spread more in countries where internet penetration and social media influence is deep and profound. Majority of these misinformation was characterized under rumors. Many examples were also highlighted regarding how spread of misinformation caused loss of life and property. The article concludes in highlighting the influence of social media platforms and the trend of misinformation spread with the spread of the actual disease itself.

There were several internet reports highlighted this and other articles regarding misinformation and its spread during the pandemic. One of the common highlights was the role of the social media in providing the infodemic with a platform that is accessible to billions. The role of content moderation and challenges in doing so is one of the key aspects in controlling the spread of misinformation. Failure of AI engines and data processing models in controlling the spread is one of the key takeaways from the infodemic that is still at large. The issue of an inherent bias in data modeling and difficulty to separate free speech from speculation. Data science professionals and government agencies are urged to develop control measures and regulations to control spread of infodemic especially during a healthcare crisis.

This infodemic, in fact, ties in with the article on statistical modeling: the two cultures as this highlights another issue with data modeling and implications of the

sheer ignorance to the underlying data generation mechanisms. As an example, Facebook is inundated with false information that is impractical to remove using manual intervention. The algorithms used by the social media site rely heavily on training data that comes from the user-generated database. This loop of misinformation being classified as popular and potentially innocuous is a big issue with automated content moderation. These articles will help governments in fabricating legislation to control the influence of social media and the spread of misinformation.