Big Data Analytics & Applications
4th of May 2019

A Symposium taking place
In Mathematical Research and Conference Center in Będlewo, Poland
Organized by Faculty of Mathematics, Computer Science and Econometrics
University of Zielona Góra, Poland

There are hosts of buzzwords in today's data-centric world, and especially in digital and print media. We encounter data in every walks of life, and for analytically and objectively-minded people, data is everything. However, making sense of the data and extracting meaningful information from it may not be an easy task. We come across buzzwords such as big data, high dimensional data, data visualization, data science, and open data without a proper definition of such words. The rapid growth in the size and scope of data sets in a host of disciplines has created a need for innovative statistical strategies analyzing such data. For example, many private and public agencies are using sophisticated data mining strategies and/or big data analytics to reveal patterns based on collected information. Some examples of big data that have prompted demand are digital marketing, customer service standards, gene expression arrays, social network modeling, clinical, genetics and phenotypic data.

The purpose of this workshop is bi-fold.

1. We focus on estimation of model parameters and prediction based on high dimensional data (HDD). In classical regression context, we define HDD where number of predictors (p) are larger than the sample size (n). A host of the classical techniques are available when p < n to reveal the data story. However, the existing classical strategies are not suited for providing solutions in the case of HDD. Over the past two decades, many penalized regularization approaches developed to perform variable selection and estimation simultaneously. We discuss model selection and post prediction strategies providing a better trade-off between model prediction and model complexity.

2. Most organizations have the ability to gauge their own individual performance through benchmarking and set KPI/Metrics. In the realm of ecommerce and digital marketing, the opportunity to incorporate consumer behavior will provide valuable insight on how the market demand could influence an organization's profitability. In this case, we will use an example of a retailer who has both an online and physical store. We will use an Application Programming Interface (API) where we will make a request and get information from Spotify streaming service to see what albums are in demand on their platform. This information will give us a method to assess inventory and consider what albums should be advertised both online and in the physical store. We will use Postman application to show the basics of an API call and then use Python to make a more detailed call. After we have ingested the data from Spotify, a medoid clustering approach is used to characterize album data. The medoid clustering is robust strategy than K-means approach; it is less sensitive to outliers. Based on the 3D clustering approach, we can identify which albums are in demand and assess our risk through Spotify's popularity metric. Finally, we extract the albums based on the cluster and start targeting to reflect the market trends, potentially shift our inventory to mitigate risk and create a strategy on niche albums.

   The suggested methodology is quite general in nature and has potential for implementation to a host of marketing scenarios.

PROGRAM:

Time	Session	Presenter
10:00 – 10-15	Welcome/Introduction	Roman Zmyślony
10:15 – 11:00	Big Data Analytics: challenges and opportunities	S. Ejaz Ahmed (Brock University, Canada)
11:00 – 11:30	*Healthy break*	
11:30 – 12:30	Identifying Advertisement Opportunities based on Integrating Public Data.	S. Jazib H. Ahmed (Fanxchange, Canada)
12:30 – 14:00	*Lunch break*	
14:00 – 15:00	High Dimensional Data Analysis: Model Selection and Post Prediction - I	S. Ejaz Ahmed (Brock University, Canada)
15:00 – 15:30	*Healthy break*	
15:30 – 16:15	High Dimensional Data Analysis: Model Selection and Post Prediction - II	S. Ejaz Ahmed (Brock University, Canada)
16:15 – 17:30	Panel Discussion – Where to from here?	Panelists are welcome
18:00	*Dinner*	

Speaker's information:

Dr. S. Ejaz Ahmed is Professor of Statistics and Dean of the Faculty of Math and Science at Brock University, Canada. Previously, he was Professor and Head of the Mathematics and Statistics Department at the University of Windsor, Canada and University of Regina, Canada as well as Assistant Professor at the University of the Western Ontario, Canada. He holds adjunct professorship positions at many Canadian and International universities. He has supervised numerous Ph.D. and Master Students, and organized several international workshops and conferences around the globe. He is a Fellow of the American Statistical Association. His areas of expertise include big data analysis, statistical learning, and shrinkage estimation strategy. Having authored several books, he edited and co-edited several volumes and special issues of scientific journals. He is Technometrics Review Editor for past ten years. Further, he is Editor and associate editor of many statistical journals. Overall, he published more than 175 articles in scientific journals and reviewed more than 100 books. Having been among the Board of Directors of the Statistical Society of Canada, he was also Chairman of its Education Committee. Moreover, he was Vice President of Communications for The International Society for Business and Industrial Statistics (ISBIS) as well as a member of the "Discovery Grants Evaluation Group" and the "Grant Selection Committee" of the Natural Sciences and Engineering Research Council of Canada.

For more information see: *https://brocku.ca/mathematics-science/mathematics/directory/syed-ejaz-ahmed/*

S. Jazib H. Ahmed is currently a Data Analyst at Fanxchange, an online Canadian secondary ticket exchange. He previously worked in the automotive and education industry. He has successfully implemented analytics in multiple organizations and has dealt with the stigma where decision-making is solely based on industry experience rather than incorporating a scientific data driven approach. He is an avid researcher on the industry's best practices where he incorporates those methods in its application in marketing.

PS: For any query/fee/further information about this Symposium see
www.impan.pl/en/activities/banach-center/conferences/19-multivariate

or please contact organizers:

Name	Augustyn Markiewicz	Roman Zmyślony
E-mail	*amark@up.poznan.pl*	*r.zmyslony@wmie.uz.zgora.pl*