Applied Data Science
Invited Talks Track at KDD-2017

The Applied Data Science (ADS) Invited Talks Track at KDD-2017 is a continuation of what has now become a "7-year tradition" at KDD conferences. This is the second year the track operates under the ADS name, an evolution from its origins at KDD-2011 as the "Industry Practice Expo".

The KDD Conference on Knowledge Discovery and Data Mining (KDD) is the world's first, largest and best conference on Data Science, Data Mining, and Knowledge Discovery. It brings together a healthy mix of academic researchers, industry and government researchers, and practitioners from a wide range of institutions and fields. The primary focus on KDD is on peer-reviewed research contributions and the academic advancement of the field. This is an important goal and in fact the KDD conference is now recognized as the most competitive and prestigious forum for presenting high quality research results. KDD, being fundamentally an applied field, needs the strong representation of applied work of big impact. Over the years of running the conference we observed that our initial speaker-selection approach needed to be re-thought because of the important contributions made to the field outside traditional academic, industrial and government research laboratories.

The result of this re-thinking was to create a forum that exposes important contributions to Data Science through Big Data Applications that address strategic problems. We wanted to effectively capture the rising importance of Data Science and Machine Learning especially in the Big Data environment where structured and unstructured data create special challenges, and of course present new opportunities. The goal of the Invited Talks Track is to curate contributions from leaders in our field who have made important contributions through the development of a system, the creation of a new and important business, or the development and market introduction of a product,. Some of these important contributions may never see an academic paper or detailed peer-reviewed paper written about them, yet they are of critical importance to our very applied field.

To give you an idea of how rapidly growing this area is, and how this sector of our industry and promises to be highly disruptive across many industries, we cite a couple of articles out of a plethora of such coverage: According to IDC, the global revenues from Big Data and business will grow from $130.1 billion in 2016 to more than $203 billion in 2020, at a compound annual growth rate (CAGR) of 11.7% [1]. Furthermore, to quote from a Forbes article: ""Data monetization" will become a major source of revenues, as the world will create 180 zettabytes of data (or 180 trillion gigabytes) in 2025, up from less than 10 zettabytes in 2015." [2]

Over the past nine months we have worked hard to create a strong program of invited speakers. The focus is first and foremost on applications: deployed, real-world applications addressing strategic problems in industry, in private or public sectors, with quantifiable value delivered. This year we particularly focused on two areas that, in our opinion, are not getting sufficient attention in the Data Science, and Machine Learning communities: the establishment of actual repeatable benchmarks for the field and the creation of tools to track, manage, and bring order and repeatability to the (often messy) work environment of Data Scientists.

The Invited Talks are clustered around the following themes:

Data Science in Sensor Data: David Potere, CEO of Tellus Labs will speak on how common the analysis of spaceborne data has become in his talk: "Spaceborne data enters the mainstream." A related topic is analysis of climate data and will be covered by Professor Vipin Kumar of the

University of Minnesota. Professor Jonathan How of MIT will cover the issues of uncertainty in learning and planning when it comes to dealing with this type of data.

Benchmarks & Process Management in Data Science: Eduardo Ariño de la Rubia, Chief Data Scientist of Domino Data Lab will speak on experiences with large enterprises in deploying process management tools for data scientists. Szilard Pafka, Chief Scientist of Epoch will share experiences and learnings from attempting to uild large scale benchmark metrics and data sets in the field.

Understanding Behavior with Data Science: Professor Andy Berglund, of the University of Florida will speak on "Mining Big Data in NeuroGenetics to Understand Muscular Dystrophy." Mainak Mazumdar, EVP and Chief Research Officer of Nielsen will speak on challenges in media measurement with BigData and Paritosh Desai, SVP Enterprise Data and Analytics of Target addresses the real considerations of successfully delivering results at a large retailer with his talk: "It Takes More than Math and Science to Hit the Bullseye with Data."

Applied Machine Learning: Joshua Bloom, VP of Data and Analytics of GE speaks on machine learning is practical industrial/manufacturing settings.

Rajesh Parekh, Director of Analytics at of Facebook delivers a talk on "Designing AI at Scale to Power Everyday Life."   Professor Longbing Cao of UTS speaks on the dramatic results achieved with applying data science for tax enforcement in working with the Australian Taxation Office.

These talks present a rare opportunity to hear from the very best in the field about the most exciting topics when it comes to building highly scalable platforms for Data Science and deploying practical, real-world applications. Our invited speakers will share key insights from their experiences and present valuable lessons learned.

Invited Panel Session: The ADS Invited talks track concludes with a Invited Panel Session titled: "**Benchmarks and Process Management in Data Science: Will We Ever Get Over this Mess?**" which addresses the issues of benchmarks and management of the work environments of Data Scientists to achieve consistency, objectivity, and repeatability.This panel will be moderated by Usama Fayyad of Open Insights. The panelists are: Arno Candel, CTO of H20.ai; Eduardo Ariño de la Rubia, Chief Data Scientist of Domino Data Lab; Szliard Pafka, Chief Scientist of Epoch, Anthony Chong, Founder and CEO of IKASI; and Jeong Yoon Lee, Technical Evangelist at Microsoft.

REFERENCES

[1] See IDC report: *Worldwide Semiannual Big Data and Analytics Spending Guide* - https://www.idc.com/getdoc.jsp?containerId=prUS41826116 - October 2016.

[2] See Forbes Magazine article: "6 Predictions For The $203 Billion Big Data Analytics Market" - https://www.forbes.com/sites/gilpress/2017/01/20/6-predictions-for-the-203-billion-big-data-analytics-market/#61f37fe82083 - January 2017.

Usama M. Fayyad *(Open Insights)*
Evangelos Simoudis *(Synapse Partners)*
Ashok Srivastava *(Verizon)*