Open Data Science

In partnership with مركز الشباب YOUTH HUB

Sponsored by IBM R data iku ASPIRE DISCOVER TALENT

Time	ML track	Business track
10:00 - 11:00	Registration and Breakfast	
11:00 - 11:10	Greetings	Greetings
11:10 - 12:00	**Towards Fairness in Machine Learning Models** - Jack Fitzsimons, Principal Machine Learning Scientist @ Qera *Fairness, through its many forms and definitions, has become an important issue facing the machine learning community. In this talk, we consider how to incorporate group fairness constraints in kernel regression methods. More specifically, we focus on examining the incorporation of these constraints in decision tree regression when cast as a form of kernel regression, with direct applications to random forests and boosted trees amongst other widespread popular inference techniques. We show that order of complexity of memory and computation is preserved for such models and*	**When machine learning facing real Machines: selected ML cases in manufacturing** - Roman Chebotarev, ML Architect, Head of Delivery @ Zyfra *The talk will be focused on ML/AI applications in the manufacturing industry - its specific issues and common ML approaches to solve it. The talk will come with the following real-world examples:* • *Increasing oil flow extraction in artificial pumping,* • *Reducing costs for polymer-modified bitumen production,* • *Product flow rate increase in dichloromethane (CH2Cl2) production.*

	bound the expected perturbation to the model in terms of the number of leaves of the trees. Importantly, the approach works.	
12:05 - 12:40	## Image Restoration using Deep Learning: Latest Trends - Orest Kupyn, Senior Research Engineer @ SoftServe *In this talk, we'll discuss the general approach for the Image Restoration task as well the latest research trends in this area. We'll see how conditional GANs can be used to improve the results of Image Deblurring, Dehazing, Denoising, and Enhancement in general. The audience must be familiar with the basic concepts of Convolutional Neural Networks and Deep Learning in general.*	## Towards high-quality Machine Learning models for business needs - Justin Bewsher, Senior Machine Learning Researcher @ Qera *Artificial Intelligence (AI) is seeing ever increasing integration into commercial and business applications. The benefits are potentially enormous, with many heralding AI as the new electricity. Despite this, practical knowledge and expertise is limited. Especially when it comes to quality assurance and verification of an integrated Machine Learning model. With a new and unfamiliar technology, how can we ensure that it both fulfills our business need and performs? In this talk we discuss some of the issues and considerations to move towards ensuring high-quality ML business solutions.*
12:45 - 13:20	## Model Explainability - Aoun Lutfi, AI Solutions Engineer @ IBM *In this talk, we will explore how to ensure the explainability of a model and how to explain the results obtained by a model. This is important in a various business where a decision is based on the results of the data. We will explore various approaches to explain a model's results and how to handle situations where the explained results do not match the human intuition.*	## Your Path to Enterprise AI - Hylke Visser, Director of Sales and Business Development @ Dataiku *In this talk, we will discuss why an Enterprise AI strategy is needed for many companies, but as well how to get there. Based on the experience, ranging from Fortune 500 companies to startups, a maturity model will be discussed.*
13:25 - 14:00	## Time Series Forecasting automation - Andrew Mackay, Director of Digital Transformation @ Teambase *To get an insight on how to leverage the power of Dataiku machine learning automaton to get your*	## How good of an employee are you? My journey from a university lab to leading an HR Analytics team - Hamza Bendemra, Analytics Manager @ Royal Group UAE

	Time Series Forecasting initiative off the ground. He will share valuable lessons learned from the field including getting business buy-in and scaling the solution for large enterprises with many business streams.	*In this talk, I'll be discussing my experience transitioning from academia to a corporate environment; and how Data Analytics with Dashboarding and ML predictive analysis can lead to quick results in organizations. Indeed, companies wanting to leverage ML techniques first need strong foundations to build a data-driven culture and I'll be discussing my work at Royal Group's HR Analytics team as an example.*
14:00 - 14:55	Lunch break	
15:00 - 15:35	## Face anti-spoofing in biometric systems - Evgenii Makarov, R&D Architect @ Emirates NBD *We will discuss techniques to detect 2D and 3D face spoofing attacks. Learn how neural networks cope with this task and will talk about the specifics of anti-spoofing detection for mobile devices.*	## Hidden engineering debt in AI systems - Dmitriy Selivanov, machine learning engineer & co-founder @ stealth mode startup *We will consider an example from Dmitriy's recent project - real-time personalization system for one of the largest retail players in the region. Dmitriy will describe different components of the architecture and demonstrate how they interoperate. Hopefully by the of the talk audience will have a high-level understanding about efforts, resources, and skills required for building such real-world systems.*
15:40 - 16:10	## Non-standard problems in feature engineering - Pavankumar Gurazada, Sales Manager @ KIMMCO *Engineering hand crafted features is fun. It is an inherently creative process that demands undivided attention to the data set. In this talk, we take a deep dive into the feature engineering process using a selection of data sets. We focus on three 'non-standard problems' (for those who haven't had the joy of reading old Russian mathematics books, we refer to problems that do not admit a canned solution). Problem 1: Distributions and Transformations. Look beyond histograms. Feature distributions can inform modeling choices. Problem 2: Missingness. What to do when we our data set misses a beat?*	## ML-driven Risk Management System: How to use clustering algorithms to calculate events impact on revenue - Pavel Golubev, Data Scientist @ Reaktor *A talk on how PCA helps to manage risks for airlines tickets sales: impacting events clustering, confidence intervals analysis, solution architecture, production deployment.*

	Problem 3: Peeking into predictive features. Before training a model, take a pit stop and run feature - target correlations; get a feel for what to expect from fitted models.	
16:15 - 16:45	## Why do stack models win data science competitions? - Nader Hamadeh, Data scientist @ dubizzle *Ensemble models are often used in boosting the accuracy of predictive models. These type of models are very popular in random forest and gradient boosting decision trees. The idea behind it is to combine weak learners and use them as a collective to predict the specified target. A new modern approach nowadays is taking shape, by rather than combining week learners it combines strong but diverse learning. The idea behind this approach is to combine strong machine learning models and use them as a collective power. Building a powerful ensemble has many resembles business units in a company. The more diverse, yet specialized in a specific area, the team is the more the business will benefit. A simple kind of ensemble is to take the unweighted average of the model's prediction. For example, if three models (Random Forest, SVM, and linear logistic regression) were used together, the final prediction will be the average vote of the three. But this is so simplistic and does not consider that some models must be more powerful than the others. An even better approach might be to estimate these weights more intelligently by using another layer of the learning algorithm. This approach is called model stacking. Stacking models have been in the top of Kaggle competitions for 3 years now and I will be talking about some techniques to do stacking with some example of the most popular winning ones.*	## Finding the Holy Grail: Addressing the Challenges of a Governed Data Science Pipeline - Zeeshan Mehdi, IBM Big Data & AI Technical Leader *Zeeshan will walk you through the key challenges of implementing an end to end governed Data Science Pipeline in a local government & enterprises.* *The majority of organizations locally face challenges in the region can be placed into three categories: Storing data in their right repositories, organizing & governing that data and then analyzing that data. We will see how to addresses these challenges through an agile, containerized governed data science platform.*
16:45 - 17:30	Networking	