



## Data Science Community (DSC)

Študentský klub pôsobiaci pri Ekonomickej Univerzite v Bratislave

Harmonogram výučby:	Letný Semester 2019/2020
Stretnutia:	MAREC: utorok 17:00-18:30
Miesto:	EUBA (učebňa D117)
Čo potrebujete:	Vlastný notebook s JUPYTER-om

Prihlášky a otázky, prosím, smerujte na formulár:

[https://docs.google.com/forms/d/1DrhvuzFlyaS0nad7m2pUXUTyCp3YHTjZE8mdBy\\_GGSE](https://docs.google.com/forms/d/1DrhvuzFlyaS0nad7m2pUXUTyCp3YHTjZE8mdBy_GGSE)

Kontakt: [dsceuba@gmail.com](mailto:dsceuba@gmail.com)

### Letný semester: Predikcia cien na trhu s nehnuteľnosťami.

Nový semester, nový harmonogram DSC aktivít. V letnej edícii nadviažeme na znalosti, ktoré sme spolu naberali od septembra do decembra minulého roka.

Informácia pre tých, ktorí neboli súčasťou minulo-semesterálnych aktivít klubu. Znalosť Python-u nie je nutná. Lekcie budú pokrývať iba základné príkazy a nebude náročné ich zvládať. Komplikované by mohlo byť to, že sa budeme venovať iba veľmi úzkej a špecifickej časti programovania bez širšieho kontextu, ktorý bol predstavený v zime.

Na programe budú tri prednášky, počas ktorých vypracujeme pod vedením odborníkov z praxe kompletný end-to-end machine learning projekt. Cieľom týchto troch lekcií je naprogramovať model umelej inteligencie, ktorý bude predikovať ceny nehnuteľnosti v Kalifornii. Počas lekcií si zopakujeme ako pracovať so základnou python knižnicou – pandas, osviežime si vizualizáciu dát v tomto jazyku a naučíme sa ako sa tvoria modely umelej inteligencie v knižnici – scikit-learn.

Kurzy budú prebiehať iba v priebehu MARCA v klasickom čase v utorok medzi 17.00 až 18.30 v miestnosti D117 (stará budova). Konkrétne dátumy sú uvedené nižšie.

Tešíme sa na vašu účasť.

#### 1. Exploratívna analýza vstupných dát o cenách nehnuteľnosti (10. Mar. 2020)

- Lektor: [Juraj Krivda](#), Big Data Science Lab Team, ERSTE GROUP IT.

#### Dáta:

- California housing prices je jeden z najznámejších datasetov, na ktorých sa ukazuje tvorba machine learning modelu.

#### Obsah hodiny:

- Počas prvej hodiny si zopakujeme prácu s pandas a vizualizáciu. Všetko aplikujeme pri práci s reálnymi dátami – rôzne informácie o nehnuteľnostiach. Budeme hľadať vzťahy v dátach, ktoré môžu byť kľúčové pre navrhnutie dobrého modelu.



## 2. Úprava dát na vstup do modelu (17. Mar. 2020)

---

- Lektor: **Filip Dzuroška**, Big Data Science Lab Team, ERSTE GROUP IT.



### Obsah hodiny:

- Vysvetlíme si ako funguje supervised learning (učenie pod dohľadom).
- Upravíme dáta do podoby, ktorá vyhovuje zvolenému modelu. Každý model má svoje špecifiká a vyžaduje si úpravu vstupných dát do určitej podoby.
- Pri úprave dát budeme opäť pracovať v knižnici pandas.

## 3. Tvorba AI modelu a interpretácia výsledkov (24. Mar. 2019)

---

- Lektori: **Filip Dzuroška & Juraj Krivda**, Big Data Science Lab Team, ERSTE GROUP IT.



### Obsah hodiny:

- Machine Learning: Vytvorenie modelu, ktorý sa učí z dát bez explicitného definovania pravidiel. Model bude predikovať ceny nehnuteľnosti na základe ich špecifických charakteristikách ako je rozloha, lokalita, vek a pod.
- Zhodnotenie presnosti modelu: Definujeme si metriky na hodnotenie presnosti predikcií modelu a povieme si, či by sme s týmto modelom uspeli na trhu s nehnuteľnosťami. Vysvetlíme si, čo spôsobilo nepresnosti a načrtneme si ako by vyzerali výsledky ak by sme použili iný model z množiny supervised learning.

### Po kurze:

---

Študentská spätná väzba: [Feedback na DSC aktivity](#)

- Zhodnotenie semestra, návrhy na zlepšenie a ďalšie kroky klubu.



### Odporúčaná literatúra

---

1. Python Data Science Handbook: Essential Tools for Working with Data, Author: Jake VanderPlas, 2016 (available On-line)
2. Hands-On Machine Learning with Scikit-Learn and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems, Author: Aurélien Géron, 2017

