

Strojové učení nad velkými daty

Kód kurzu: MLC_BDATA

Cílem tohoto kurzu je představit různé nástroje a koncepty ze strojového učení nad velkými daty. Po dokončení tohoto kurzu by měli účastníci být schopni říct jaký nástroj použít pro daný problém, zjistit jestli neexistuje jednodušší řešení a znát časté chyby a umět se jim vyhnout. Speciální pozornost věnujeme Sparku jakožto univerzálnímu nástroji, který lze použít jak pro zpracování velkých dat, tak pro ML nad velkými daty.

Pobočka	Dnů	Cena kurzu	ITB
Praha	1	4 990 Kč	0

Uvedené ceny jsou bez DPH.

Termíny kurzu

Datum	Dnů	Cena kurzu	Typ výuky	Jazyk výuky	Lokalita
-------	-----	------------	-----------	-------------	----------

Uvedené ceny jsou bez DPH.

Požadované vstupní znalosti

- Základy práce v Pythonu a v nástroji Google Colab
- Znalosti strojového učení na úrovni kurzu Úvod do strojového učení.

Studijní materiály

Studijní materiál společnosti Machine Learning College.

Osnova kurzu

- Přehled konceptů a nástrojů ve zpracování velkých dat
- Od malých k velkým datům a odhad jejich hodnoty
- Řádkové a sloupcové databáze
- HDFS (Hadoop Distributed File System)
- Formáty dat – Parquet, ORC, Avro
- Komprese – gzip, snappy, zstd
- SQL databáze – BigQuery, Redshift, Clickhouse, Snowflake, Vertica
- Praktický příklad na srovnání malých a velkých dat
- Úvod do Sparku
- MapReduce
- Spark Computing Engine a RDDs (Resilient Distributed Datasets)
- DataFrames
- Spark ekosystém
- Nejčastější chyby
- Kde pustit Spark
- Alternativy – Apache Beam (Dataflow), Dask, lambdas
- Praktický příklad se Sparkem
- ML strategie pro velká data
- Inkrementální učení
- Dávkové učení pro neuronové sítě
- Distribuované trénování
- Federated learning
- Alternativní strategie
- Náhodné vzorkování
- Podmodely

GOPAS Praha

Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk



Copyright © 2020 GOPAS, a.s.,
All rights reserved

Strojové učení nad velkými daty

- Větší výpočetní kapacity
- Frameworky
- Scikit-learn a partial_fit
- MLlib
- Dask-ML
- Praktické příklady s frameworky
- Nejčastější chyby

GOPAS Praha

Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk



Copyright © 2020 GOPAS, a.s.,
All rights reserved