



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



Název projektu	Rozvoj vzdělávání na Slezské univerzitě v Opavě
Registrační číslo projektu	CZ.02.2.69/0.0./0.0/16_015/0002400

Dolování dat

Úvodní informace a požadavky na absolvování

Jan Górecki



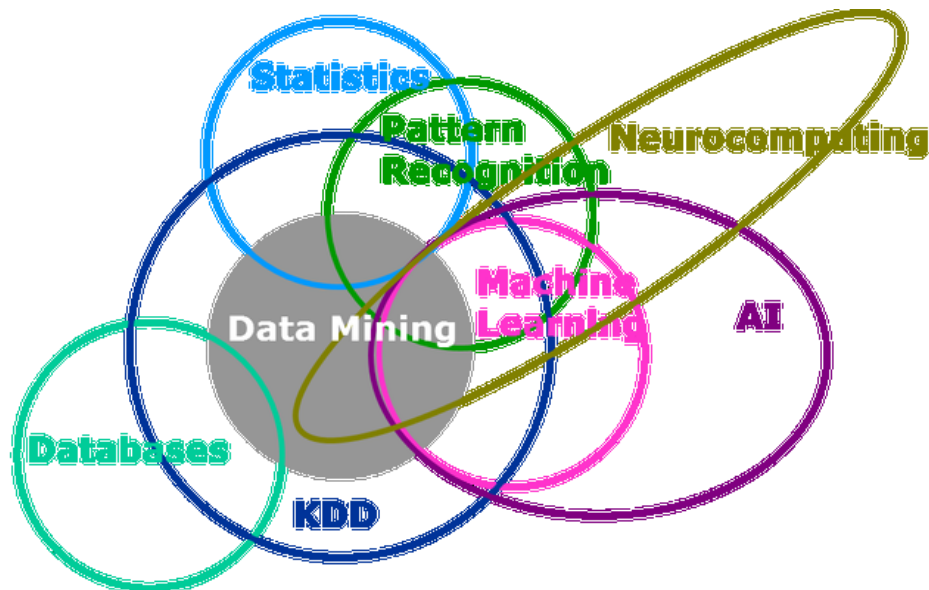
**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Dolování dat (Data mining)



SLEZSKÁ
UNIVERZITA
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVÍNĚ

Non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns from data (Fayyad a kol., 1996)



The Rise of Deep Learning

'Deep Voice' Software Can Clone Anyone's Voice With Just 3.7 Seconds of Audio

Using snippets of voices, Baidu's 'Deep Voice' can generate new speech, accents, and tones.



with DEEPMIND STARCRRAFT TRIUMPH FOR

Let There Be Sight: How Deep Learning Is Helping the Blind 'See'



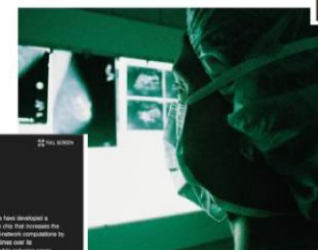
Technology outpacing security measures

Facial Recognition | Features and Interviews



AI beats docs in cancer spotting

A new study provides a fresh example of machine learning as an important diagnostic tool. Paul Binger reports.



AI Can Help In Predicting Cryptocurrency Value



'Creative' AlphaZero leads way for chess computers and, maybe, science

Former chess world champion Garry Kasparov likes what he has to say about a computer that could be used to find cures for diseases



How an A.I. 'Cat-and-Mouse Game' Generates Believable Fake Photos

By CADE METZ and KEITH HOLENBERG - JUNE 2, 2018



AI: Faked Data



AI-generated faces show how far AI image generation has come in just four years

AI-generated faces are on the right aren't real; they're the product of machine learning



Stock Predictions Based On AI: Is the Market Truly Predictable?



Neural networks everywhere

New chip reduces neural networks' power consumption by up to 95 percent, making them practical for battery-powered devices.

DeepMind | Wed, 01/10/2018 - Boston | Comment by Kerry Walker - Digital Reporter - @RandPhugase

After Millions of Trials, These Simulated Humans Learned to Do Perfect Backflips and Cartwheels



Researchers introduce a deep learning method that converts mono audio recordings into 3D sounds using video scenes



Automation And Algorithms: De-Risking Manufacturing With Artificial Intelligence

Sarah Goehke Contributor Manufacturing 1 item in the Industrialization of additive manufacturing

TWEET THIS The two key applications of AI in manufacturing are pricing and manufacturability feedback

Complex of bacteria-infecting viral proteins modeled in CASP-13. The complex consists of 12 proteins that were modeled individually. PROTEIN DATA BANK

Google's DeepMind acs protein folding

By Robert F. Service | Dec. 6, 2018, 12:05 PM



Dolování dat:

- **Prezenční forma: 13 přednášek a 12 seminářů,**
 - **Kombinovaná forma: 3 přednášky**
 - **zakončena zkouškou**
-

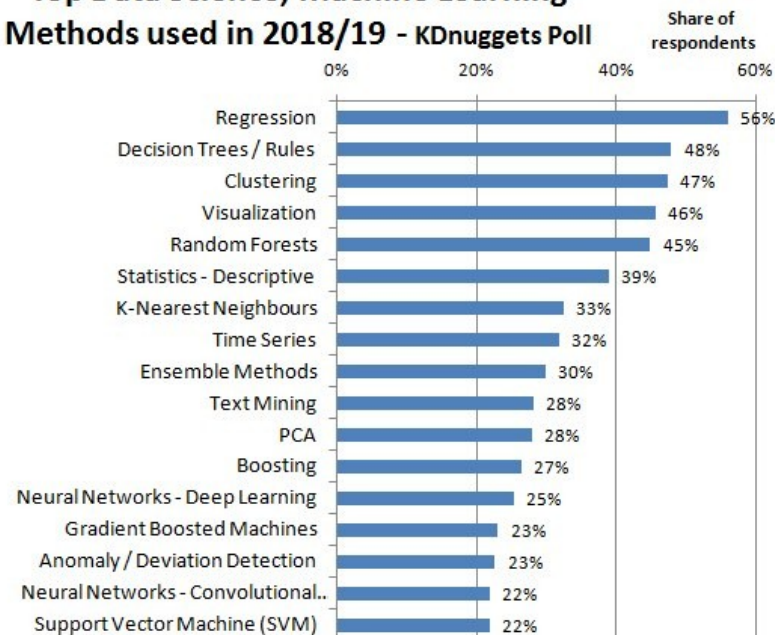
Stručná anotace předmětu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- **Proces dolování dat**
Dolování dat, úlohy dolování dat, metodiky pro dolování dat
- **Statistika v kontextu dolování dat**
Kontingenční tabulky, regresní analýza, diskriminační analýza, shluková analýza
- **Strojové učení**
Základní pojmy, principy strojového učení, typy strojového učení, formy strojového učení, trénovací data, atributy, chybová funkce
- **Metody dolování dat**
Rozhodovací stromy, Rozhodovací pravidla, Neuronové sítě, Genetické algoritmy, bayesovské metody, metody založené na analogii
- **Evaluace modelů**
kritéria, deskriptivní úlohy, klasifikační úlohy, vizualizace modelů, vizualizace klasifikací, porovnávání modelů, volba nejvhodnějšího algoritmu, kombinování modelů
- **Předzpracování dat**
Příprava dat, strukturovaná data, více vzájemně propojených tabulek, odvozené atributy, příliš mnoho objektů, příliš mnoho atributů, numerické atributy, kategoriální atributy, chybějící hodnoty

**Top Data Science, Machine Learning
Methods used in 2018/19 - KDnuggets Poll**



Požadavky na absolvování předmětu



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

- **docházka na semináře min. 60% (10 % hodnocení),**
- **zpracování seminární práce (30% hodnocení),**
 - Analýza vybraných dat dle metodiky CRISP-DM pomocí metod dolování dat (alespoň 5 metod celkově, z nichž alespoň 2 statistické a alespoň 3 ze strojového učení)
- **zkouška (60% hodnocení)**

Celkem maximum: 100

Požadované minimum: 60



- **Veškeré elektronické materiály je možné nalézt na školní síti: L:\gorecki\public\NPDOD-NKDOD \ (přes <https://raimundo.opf.slu.cz/NetStorage/> popř. files.opf.slu.cz)**
-

Povinná:

- BERKA, P. a GÓRECKI, J., 2017. *Dolování dat*. Skripta SU OPF, Karviná.
- BERKA, P., 2003. *Dobývání znalostí z databází*. Praha: Academia. ISBN 80-200-1062-9.

Doporučená:

- CLARK, B., E. FOKOUE a H. H. ZHANG, 2009. *Principles and theory for data mining and machine learning*. New York: Springer. ISBN 978-0-387-98134-5.
 - MURPHY, K. P., 2012. *Machine learning: A probabilistic perspective*. London, England: The MIT Press. ISBN 978-0-262-01802-9.
-

Software



- **MATLAB**

- Statistics and Machine Learning Toolbox
- <https://www.mathworks.com/solutions/data-science.html>
- trial verze z mathworks.com
- Octave – free verze MATLABu

- **Python**

- **R**

- **RapidMiner**





- **Nejlépe vlastní**
 - **UC Irvine Machine Learning Repository**
<https://archive.ics.uci.edu/ml/index.php>
 - **Kaggle: Your Home for Data Science**
<https://www.kaggle.com/>
 - **KEEL - dataset repository**
<http://www.keel.es/datasets.php>
-

Kontakty



**SLEZSKÁ
UNIVERZITA**
OBCHODNĚ PODNIKATELSKÁ
FAKULTA V KARVINĚ

Ing. Jan Górecki, Ph.D.

gorecki@opf.slu.cz

A407

- **konzultace po domluvě emailem**

Sekretariát Katedry informatiky a matematiky

A402
