

CENTRO UNIVERSITÁRIO UNINORTE
CURSO DE PÓS-GRADUAÇÃO EM: Pós
Graduação em Gerência de Banco de Dados.
DISCIPLINA: Mineração de Dados



Apresentação

Prof.º: Manoel Limeira
juniorlimeiras@gmail.com

Apresentação

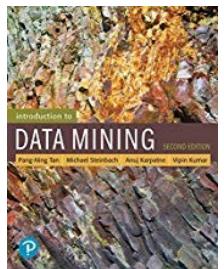
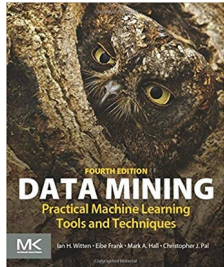
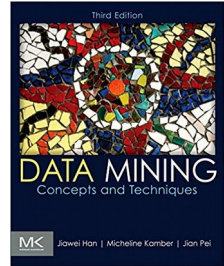
- Professor
 - Quem é o docente?
 - Expectativa com a turma
- Alunos
 - Experiência na área de mineração de dados
 - Expectativa com a disciplina

Ementa

- Processo de KDD (*Knowledge Discovery in Databases*)
- Tarefas de Mineração de Dados
 - Preditivas
 - Classificação e Regressão
 - Descritivas
 - Clusterização
 - Regras de Associação
- Ferramentas de Mineração de Dados

Bibliografia

- HAN, J; KAMBER, M. **Data Mining: Concepts and Techniques**. 3 ed. Morgan Kaufmann Publishers, 2011.
- WITTEN, Ian H.; FRANK, Eibe; HALL, Mark A. **Data Mining Practical Machine Learning Tools and Techniques**. 4 ed. Morgan Kaufmann, 2016.
- TAN, Pang-Ning; STEINBACH, Michael; KARPATNE, Anuj; et al. **Introduction to Data Mining**. 2 ed. Pearson, 2018.
- FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. **From Data Mining to Knowledge Discovery in Databases**. AI Magazine, v. 17, n. 3, p. 37–37, 1996.



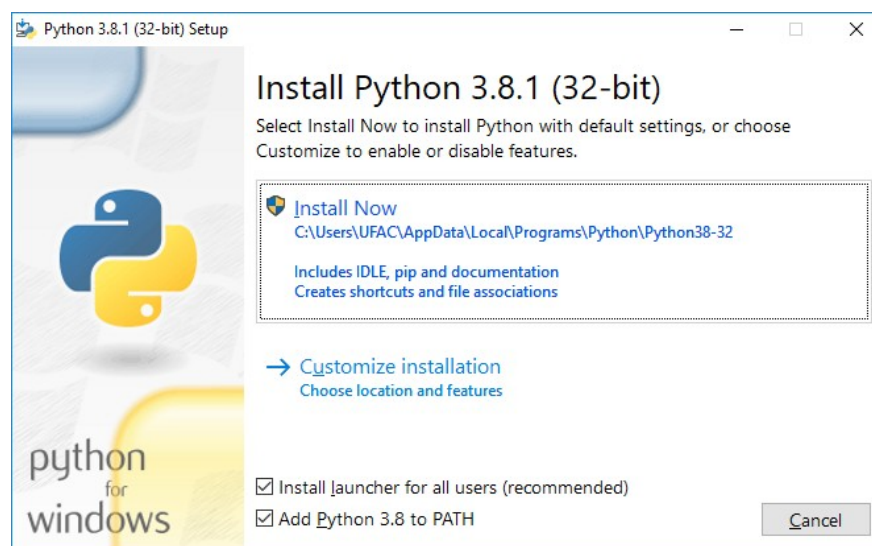
[illegible]

Instalação Python

- www.python.org/downloads



- Windows
- MacOS
- Linux



- Para verificar a versão
 - `python -V`

Instalação scikit-learn



- <https://scikit-learn.org/stable/>
- Para atualizar o PIP
 - `python -m pip install --upgrade pip`
- Para instalar o scikit-learn
 - `pip install scikit-learn`
- Para fazer o upgrade do scikit-learn
 - `pip install -U scikit-learn`
- Para desinstalar
 - `pip uninstall scikit-learn`

Instalação jupyter

- <https://jupyter.org/install>



- Para instalar o jupyter
 - `pip install notebook`
- Para instanciar o jupyter no terminal ou prompt de comando
 - `jupyter notebook`

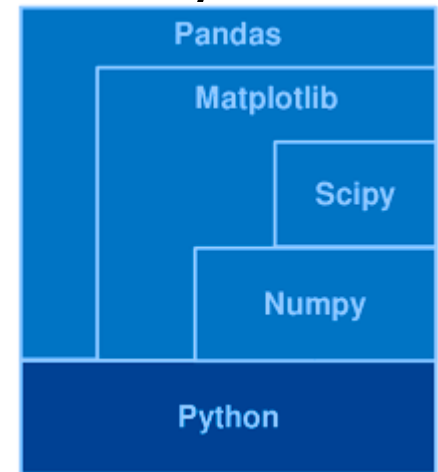
Google Colab

- <https://colab.research.google.com>
- Para acessar
 - Precisa de uma conta google
- Vantagens
 - Configuração mínima
 - Compartilhamento
 - Bibliotecas
 - Machine Learning



Bibliotecas

- Pandas
 - `pip install pandas`
 - <https://pandas.pydata.org/pandas-docs/stable/index.html>
- NumPy
- SciPy
- Matplotlib
 - `pip install -U matplotlib`
 - <https://matplotlib.org/index.html>

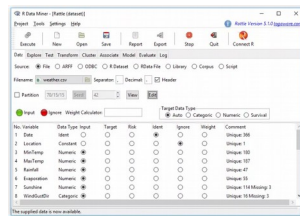


Outras Ferramentas



Spyder disponível em:

<https://spyder-ide.org>



Rattle disponível em:

<https://rattle.togaware.com/>



Weka disponível em:

<https://www.cs.waikato.ac.nz/ml/weka/download.html>



Orange disponível em:

<https://orange.biolab.si/download/>

Repositórios de dados

Sites para encontrar conjuntos de dados!

1. **Portal Kaggle**

[Kaggle.com](https://kaggle.com)

kaggle

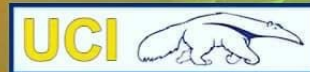
2. **Portal Brasileiro de Dados Abertos**

[Dados.gov.br](https://dados.gov.br)



3. **UCI Machine Learning Repository**

archive.ics.uci.edu



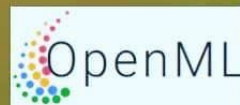
4. **Google Dataset Search**

toolbox.google.com/datasetsearch

Google

5. **OpenML**

openml.org



CENTRO UNIVERSITÁRIO UNINORTE
CURSO DE PÓS-GRADUAÇÃO EM: Pós
Graduação em Gerência de Banco de Dados.
DISCIPLINA: Mineração de Dados



Apresentação

Prof.º: Manoel Limeira
juniorlimeiras@gmail.com