Course Number: 286.4112

Spring Semester (B), 2022, Wednesdays, 8:30-11:45

Course Syllabus: Applied ML for Healthcare

Lecturer:

Ortal Dayan, ort.dayan@gmail.com

Prerequisites:

Basic Python

Target Audience:

Biostatistics MPH program

Course Objectives & Learning Outcomes:

The course provides a detailed guide on how to complete Machine Learning Projects End to End using tabular data. Students will gain a good understanding of the theory behind ML algorithms and the methodologies for completing a data science research project using Python.

Course Structure:

Each session will comprise 4 hours of lectures in which we will look under the hood of ML algorithms, learn about the steps in the workflow of ML projects and how to implement these in Python by studying Python programming concepts for ML and reviewing code implementations.

Methodology:

Revision of the topics covered in class and gaining practical experience by completing an end to end data science research project.

Course Assessment:

Compulsory attendance in 80% of meetings

Passing Grade: 60

90% of the course grade will be based on completing an end-to-end data science project in a team and 10% will be based on participation in class.

Dates Class is on Zoom:

8th and 29th March and 24th May

Course Content:

Lesson	Topics	Subtopics
1	Intro to ML with Emphasis on Applications in Healthcare	- What is ML and why to use it - Types of ML - Main Challenges of Training
2 - 3	Review of Python for ML	- Python Libraries: NumPy, Pandas, re (Regex), Scikit-Learn (Pieplines), Imbalanced-Learn, Matplotlib and Seaborn - OOP in Python for ML - File formats incl. Pickle, Jason and HDF5 - Overview of Git and Githab
4 - 6	ML Project Workflow	- Tidy data requirements - Splitting into train and test - EDA (Explatory Data Analysis) - Data preparation (incl. how to treat missing data and transformation pipelines) - Training and compering performance between models - Hyperparameter Tunning (Grid & Random Search, Optuna) - Model Explainability - Feature importance (Gini, Shap and Lime) - Error Analysis - Evaluation on Test Set
7	Classification	- Binary, multiclass, multilabel and multipotput - Performance measures - Changing the decision threshold - Model Calibration - Upsampling and undersampling (SMOTE & GANs)
8	Linear Regression	- Linear and Polynomial Regression - Regularized Linear Models - Logistic and Softmax Regression
9	Support Vector Machines	Linear and Non-Linear kernel SVMs for classification and regression
10	Decision Trees	- Classification and Regression Decision Trees - Pre-Prunning vs. Post Prunning
11	Ensemble Methods and Fine Tunning	- Voting, Bagging and Pasting Classifiers - Out of Bag Evaluation - Random Forest and Extra Trees - Boosting: AdaBoost and Gradient Boosting - Stacking - Differences between XGBoost, CatBoost and LightGBM and how to Fine tune them for optimum performance incl. early stopping
12	Unsupervised Learning	The K-means Algorithm and its applications

13	Intro to NLP and LLMs	- Text preprocessing and transformers for entity recognition - Large Language Models including GPT and ChatGPT for entity recognition and tabular predictions

Compulsory Literature:

Aurelien Geron. (2019). Hands-On Machine Learning with Scikit-Learn, Keras and Tensorflow, O'reilly Media, Inc.

*סטודנט יקר,

אם יש לך לקות למידה או מוגבלות/בעיה רפואית שעשויה להשפיע על לימודיך, ושבגינה את/ה זקוק/ה להתאמות אנא פנה/י לדיקנאט הסטודנטים למדור נגישות ולקויות למידה:

דואל LDA@univ.haifa.ac.il טל: 04-8249265

לאבחון והתאמות בגין לקות למידה ו/או הפרעת קשב יש לפנות ליה"ל:

דואל mhait@univ.haifa.ac.il טל: 04-8249022

אם יש ברשותך מכתב התאמות מהאוניברסיטה ואת/ה זקוק/ה להנגשה בקורס אנא פנה/י אליי בשעות הקבלה או במייל בסמוך לתחילת הקורס.

If you have a disability that may affect your studies and for which you may require accommodations, please contact the Accessibility and Learning Disabilities Department at the Dean of Students office

e-mail: LDA@univ.haifa.ac.il Phone number: 04-98249265

Students that receive accommodation letters, and need academic adjustments, please meet with me to discuss the provisions of those accommodations as early in the semester as possible.