UNIVERSIDADE FEDERAL DE SANTA CATARINA – UFSC
CENTRO TECNOLÓGICO – CTC
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA - INE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO - PPGCC

INE410XXX– Syllabus (2021.2)

1. IDENTIFICATION

Course: INE410XXX - Applied Machine Learning
Credits: 60 hours/class – 4 credits
Professors: Jônata Tyska
Mateus Grellert
Vania Bogorny
Pre-requisites: None
Course schedule: XXXX
Location: Online during the Covid-19 pandemic

2. SYLLABUS

Machine learning workflow. Basic machine learning techniques. Data preparation and pre-processing. Practical and technical aspects for applying machine learning on scientific and commercial applications.

3. GOALS

3.1 Main goal

Provide students the required knowledge to solve practical problems with learning-based techniques, and also how to prepare and improve the data used in these approaches to build more efficient models.

3.2 Specific goals:

- Present common learning-based techniques that can be used to solve many types of problems
- Explain the main steps involved in data preparation and which techniques can be used to improve a given data set
- Introduce a general machine learning pipeline, including techniques to avoid overfitting
- Discuss how machine learning is being employed in different research fields to solve important problems
- Provide practical experience on applying machine learning techniques to scientific and/or commercial applications

4. COURSE PROGRAM

4.1) Basic Learning Techniques and Concepts

- Types of learning (Supervised, Unsupervised, Reinforcement Learning)
- Clustering
- Linear and Logistic Regression
- Gradient Descent
- Introduction to Neural Networks (Perceptron, MLP and backpropagation)
- Decision Trees
- Random Forests
- Ensemble methods (bagging, boosting, gradient boosting)

4.2) Machine Learning Pipeline

- Data split
- Evaluation metrics
- Model Evaluation and Selection
- Bias-variance
- Avoiding overfitting

4.3) Data preparation

- Data Cleaning
- Data Reduction and Transformation
- Feature Extraction and Portability
- Feature Engineering
- Dimensionality Reduction
- Feature Selection

4.4) Applied machine learning seminars

4.5) Final project

5. METHODOLOGY

- All classes will be given online, using synchronous and asynchronous activities.
- The topics will be discussed in theory, but students will also be challenged to implement some techniques in practical machine learning problems. These exercises will be given throughout the semester and will be part of the final grade.
- Aside from the synchronous lectures, this course will use different means for delivering content, such as (not limited to): recorded video lectures, book chapters, and scientific papers. The discussions will be encouraged during the synchronous meetings, as well as through the course discussion forum.

6. GRADING

According to the normative resolution nº 95/CUn/2017 from 04/04/2017:

- The minimum class attendance for approval is 75%.
  - Obs.: The attendance will be computed as following:
    - For synchronous activities students have to join and participate in the video web calls.
    - For asynchronous activities, attendance will be computed based on the handing in of the assignments (observing the deadlines) proposed on the virtual learning environment (Moodle).
  - Half of the total attendance will be synchronous and half of it asynchronous.
- The grading will be given in the range 0 (zero) to 10,0 (ten), being 7,0 (seven) the minimum grade for approval.

Graded Activities:

- **SE** = Seminar on Applied Machine Learning, in which students will discuss a research paper that applies Machine Learning in a field of their interest
- **HW** = Homework exercises that will be given throughout the semester and that must be handed in using the Moodle platform.
- **FP** = Final course project based on a Machine Learning practical implementation, including the production of an article reporting the results obtained;

- The final grade (FG) will be computed using the equation below:

$$\mathbf{FG = 0.3*HW + 0.4*FP + 0.3*SE}$$

7. BIBLIOGRAPHY

7.1 Main bibliography

AGGARWAL, Charu C. **Data mining: the textbook**. Springer, 2015.

FRIEDMAN, Jerome et al. The elements of statistical learning. New York: Springer series in statistics, 2001.

HAN, Jiawei; PEI, Jian; KAMBER, Micheline. **Data mining: concepts and techniques**. Elsevier, 2011.

7.2 Complementary Bibliography

DUBOUE, Pablo. **The Art of Feature Engineering: Essentials for Machine Learning**. Cambridge University Press, 2020.

KAZIL, Jacqueline; JARMUL, Katharine. **Data wrangling with Python: tips and tools to make your life easier**. O'Reilly Media, Inc.", 2016.

OSBORNE, Jason W. **Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data**. Sage, 2013.

SKIENA, Steven S. **The data science design manual**. Springer, 2017.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to data mining.** Pearson Education India, 2016.