1. IDENTIFICATION

Course: INE410154 - Applied Machine Learning
Credits: 45 hours/class – 3 credits
Professor: Jônata Tyska Carvalho
   - all contact information and a tool for scheduling online meetings with me can be found at my website: https://tyska.prof.ufsc.br

Pre-requisites: English proficiency Level (B1) - very important
Course schedule: Every Thursday 14h to 16h30
Location: Informatics Laboratory (possibly INE 313)

Description: This course seeks to provide students with the required knowledge to solve practical problems with learning-based techniques. Its focus is on the general pipeline usually employed in machine learning applications, rather than on the specifics of each machine learning algorithm. So, it will show steps and techniques for data preprocessing and improving the data quality to build more efficient machine learning models and applications.

2. SYLLABUS

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

3. GOALS

After the course, the students will be able to:

3.1 Main learning goal

- solve practical problems with learning-based techniques, and also be able to prepare and improve the data used in these approaches in order to build more efficient machine learning models.

3.2 Specific learning goals:

- describe 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
• define a general machine learning pipeline, including techniques to avoid overfitting
• discussing how machine learning is being employed in different research fields to solve relevant problems
• applying machine learning techniques to scientific and/or commercial domains and applications
• read and critically analyze research papers applying machine learning methods to different fields
• design and implement experiments applying machine learning to their research areas
• write a research paper reporting their scientific findings regarding the application of machine learning methods
• present their scientific findings regarding the application of machine learning methods to an audience

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 AND COURSE POLICIES

- All classes will be in-person in the lab.
- The topics will be discussed in theory, with mini-lectures, and 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.
- The course will use different means for delivering content, such as (not limited to): recorded video lectures, book chapters and scientific papers readings. The discussions will be encouraged during the classes.

**Technologies used in this course:** Moodle Learning Management System, data show, computers and Internet access. Students will see examples of code using a tool called Jupyter Notebook, and practice building their own codes. GitHub will also be used as the repository system for instructors and students.

**Plagiarism policy:** Plagiarism, including the uncredited use of others' ideas or work, is strictly prohibited in this course, even when unintentional. The first instance will result in a formal notice, a failed assignment, and the opportunity to redo the assignment for a maximum of 50% of the original grade. In the second instance, a formal notice and a failed assignment will be issued. A third instance of plagiarism will result in automatic failure of the course. It is mandatory to cite all sources properly and adhere to academic integrity standards.

6. ATTENDANCE AND GRADING

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

- The minimum class attendance for approval is 75%.
- The grading will be given in the range of 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 research papers that apply Machine Learning on a field of their interest
- **IC** = Four Inclass exercises that will be given throughout the semester and that must be handed in using the Moodle platform. The exercises will be related to the student's research topic whenever possible.
- **FP** = Final course project based on a Machine Learning practical implementation, including the production of an article reporting the results obtained and its presentation to the class;

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

\[
FG = 0.3 \times IC + 0.4 \times FP + 0.3 \times SE
\]
7. BIBLIOGRAPHY

7.1 Main bibliography


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

7.2 Complementary Bibliography


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.
