Introduction to Data Mining and Machine Learning

Background

The Faculty of Computer Science at Ostfalia has maintained a partnership with the U.S. University of Wisconsin - Parkside (UWP) for several years. Within the framework of this partnership, we organize joint courses, excursions and semesters abroad for our students. In the autumn term of 2023, we offer you the opportunity to take a course on Data Mining and Machine Learning at UWP as an elective. This course be held as an <u>online lecture</u> by Prof. Vijayalakshmi Ramasamy, who is an Associate Professor at the University of Wisconsin - Parkside.

Course Description:

This course will help students to understand the conceptual underpinnings of the process, methods, computing tools, and visualization fundamental to data mining (DM) and machine learning (ML). Students will be required to work on projects using large real-world data sets to become acquainted with the strengths and limitations of various techniques like Association and Cluster analysis, Classification (identifying similar groups), and ensemble methods. Students will complete a term project emphasizing the different steps of the machine learning process, from data cleaning to the extraction of accurate models and the visualization of results. It also includes extensive hands-on coding practice in Google Colaboratory.

Course Goals & Objectives:

In this course, you will learn to:

- Describe how to create accurate and generalizable models from large and messy datasets.
- Implement Python code to clean data and derive a model using an appropriate machine-learning algorithm.
- Present solutions to stakeholders using visualizations and professional machine-learning workflows.

Tentative Topics:

- Data: Types, preprocessing (DM: Ch. 2), statistical modeling, data visualization, intro to Google Colaboratory, Python Libraries
- Python Programming: Basics
- DM/ML process: basic concepts
- Data preprocessing: Missing Data and Data Imputation Techniques
- Classification models: Decision Tree, model-overfitting, selection, evaluation, comparison
- Types of Classifiers: rule-based, KNN, Naive Bayes, SVM, ensemble methods, class imbalance problem
- Regression Analysis: simple and multiple linear regression, visualizing features, training and testing model

- Association analysis: Frequent itemset generation, rule generation, evaluation of association patterns
- Cluster analysis: overview, K-Means, hierarchical clustering, cluster evaluation
- Introduction to Text mining, NLP, and sentiment analysis

Credits

2.5 Credits (Wahlpflichtfach)

Shedule

- **Sep 6, 2023*** - Dec 15, 2023 **(*Achtung: Start vor regulärem Vorlesungsbeginn)**
- Tuesdays and Thursdays
- 11:00 AM – 12:22 PM (**Chicago Time**)
- Please note that the time change between summer and winter time is not synchronized between Germany and the USA. This will result in changes of the lecture times in KW44 for you as a German student.

Participation requirements

- Open for students from the department of computer science
- Knowledge in Java Programming and Algorithms and Data Structures

Exams

- 25 % written exam (Exam 1 in week 6 and Exam 2 in week 12)
- 25 % participation (quizzes, worksheets)
- 25 % project presentation at the end of the lecture
- 25 % final exam at the end of the lecture

Enrollment and further information

Please join the Moodle Course „WPF Introduction to Data Mining and Machine Learning (WS 23/24)" (https://moodle.ostfalia.de/course/view.php?id=9246) for further information on the enrollment process and additional Q&As.

This course serves as an information platform to reach out to all interested students. Registration for this Moodle course is not a binding registration for the lecture. Feel free to enroll this moodle course even if you are interested but not sure yet if you want to attend the lecture.