Big Data Analysis in Energy Systems

Course Code: 46392
Course Type: Theoretical
Credits: 3
Course Status: Elective Specialized Ph.D. Course
Prerequisite: Supervisor Approval

Aim/Scope/Objectives: This course intends to give the required knowledge to graduate students for analyzing the big data sets in fields of energy systems using different data mining methods. This course covers all required theoretical bases plus training over the useful codes and software in the data mining area such as RapidMiner and Hadoop platform.

Course Outline:

1- Getting to know different data
- 1-1- Introduction
- 1-2- Different types of data sets
- 1-3- Important characteristics of structured data
- 1-4- Data objectives and different attributes.
- 1-5- Statistical descriptions of data
- 1-6- Measuring data similarity and dissimilarity
- 1-7- Data visualization methods

2- Data preprocessing
- 2-1- An overview on different big data and their analysis methods
- 2-2- Redundancy and correlation analysis
- 2-3- Different data reduction methods and Principal Components Analysis (PCA)
- 2-4- Data normalization

3- Big data regression models
- 3-1- Basic concepts
- 3-2- Linear regression
- 3-3- Variance analysis (ANOVA)
- 3-4- Logistic regression

4- Basic concepts and methods in mining frequent patterns
- 4-1- Basic concepts
- 4-2- Frequent itemsets mining methods
  - 4-2-1- Apriori mining algorithm
  - 4-2-2- Eclat mining algorithm by vertical data format
  - 4-2-3- FP-Growth approach for mining frequent itemsets
- 4-3- Generating association rules from frequent itemsets
- 4-4- Pattern evaluation methods

5- Big data classification methods
- 5-1- Basic concepts
- 5-2- Decision tree induction
- 5-3- Bayes classification methods
- 5-4- Rule-based classification
- 5-5- Model evaluation and selection
- 5-6- Techniques to improve classification accuracy

6- Big Data clustering methods
- 6-1- Basic concepts
- 6-2- Expected performances in clustering
- 6-3- Partitioning approach for clustering
- 6-4- Hierarchical approach for clustering

7- Temporal Data Mining
- 7-1- Temporal data representation
- 7-2- Similarity measures between sequences
- 7-3- Mining sequences

Grading: 35% Midterm exam, 35% Final exam, 10% Homework, 20% Research Project

References:

1- J. Han, M. Kamber, J. Pei, “Data Mining Concepts and Techniques”, Third edition, 2012.
2- Mehmed Kantardzic, “Data Mining, Concepts, Models, Methods, and Algorithms”, Second edition, 2020.
3- Ian H. Witten, Eibe Frank, “Data Mining, Practical Machine Learning Tools and Techniques”, Fourth edition, 2017.
4- Vijay Kotu, Bala Deshpande, “Predictive Analytics and Data Mining Concepts and Practice with RapidMiner”, Elsevier, 2015.
5- Chapman and Hall, “Temporal Data Mining”, CRC press, 2010.
6- K.P. Murphy, “Machine Learning: A Probabilistic Prospective”, MIT press, 2010.
7- T. Agami Reddy, “Applied Data Analysis and Modelling for Energy Engineers and Scientists”, Springer, 2011.
8- F. Magoules, H X Zaho, “Data Mining and Machine Learning in Building Energy Analysis”, Wiley, 2016.
9- V. Bahatnagar, “Data Mining and Analysis in the Engineering Field”, IGI global, 2014.
10- Chapman and Hall, “Computational Intelligence Data Analysis for Sustainable Development”, CRC Press, 2013.
11- Carol L. Stimmel, “Big Data Analytics Strategies for the Smart Grid”, CRC press, 2015.