DATA MINING AND MACHINE LEARNING

ENGE 413

Lecture	**: 4**	**Year : IV**
Tutorial	**: 0**	**Part : I**
Practical	**: 2**	

Course Objectives:

The objective of this course is to introduce data mining and machine learning with a special focus on applications in geomatics engineering. Students will learn theoretical concepts as well as practical techniques for extracting meaningful patterns and knowledge from large geospatial datasets.

1 Introduction (6 hours)

- 1.1 Overview of data mining and machine learning
- 1.2 Data mining views: Data, application, knowledge and technique
- 1.3 Data mining: Major issues, ethics, resources
- 1.4 Importance and applications in geospatial domain

2 Data Preprocessing and Cleaning (8 hours)

- 2.1 Data quality and missing data
- 2.2 Data integration and transformation
- 2.3 Data reduction techniques
- 2.4 Data warehouse: Architecture, Data Cube and OLAP

3 Exploratory Data Analysis (8 hours)

- 3.1 Descriptive statistics
- 3.2 Data segregation and classification
- 3.3 Geospatial data visualization

4 Supervised Classification Techniques (10 hours)

- 4.1 Decision trees
- 4.2 Bayesian classification
- 4.3 k-Nearest Neighbors (k-NN)
- 4.4 Support vector machines (SVM)
- 4.5 Random forest (RF)
- 4.6 Geospatial classification applications

5 Unsupervised Classification Techniques (10 hours)

- 5.1 Introduction to clustering

5.2 k-Means clustering
5.3 Hierarchical clustering and grid based clustering
5.4 Density-based spatial clustering of applications with noise (DBSCAN)
5.5 Applications of clustering in geospatial domain

6 Association Rule Mining (10 hours)

6.1 Basic concepts of frequent, closed and max-patterns, and association rules
6.2 Introduction to Apriori Algorithm
6.3 FP-Growth Algorithm
6.4 Spatial association rules
6.5 Spatial co-location pattern
6.6 Mining Semantics-rich movement patterns and periodic movement patterns

7 Geospatial Applications and Case Studies (8 hours)

7.1 Case studies in land use and land cover classification
7.2 Predictive modeling in disaster management
7.3 Spatial data mining for urban planning
7.4 Integrating GIS and machine learning for geospatial solutions

Practical (30 hours)

1. Geospatial data collection from different sources, cleaning and preparation using python and GIS tools
2. Visualization of geospatial data using GIS software and exploratory data analysis with python libraries
3. Implementation of classification algorithms on geospatial data
4. Clustering geospatial data, geospatial data for market basket analysis and finding spatial pattern using geospatial data
5. Developing and training neural networks for geospatial applications
6. Implementing time series analysis for geospatial data

Final Exam

The questions will cover all the chapters in the syllabus. The evaluation scheme will be as indicated in the table below:

Chapter	Hours	Mark distribution*
1	6	6
2	8	8
3	8	8
4	10	10
5	10	10
6	10	10
7	8	8
Total	**60**	**60**

* There may be minor deviation in marks distribution.

References

1. Tan, P.N., Steinbach, M., Kumar, V. (2019). Introduction to data mining. Pearson.
2. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
3. de Smith, M. J., Longley, P. A., Goodchild, M. F. (2020). Geospatial analysis: A comprehensive guide. Winchelsea Press.
4. Yang, X.-S. (2019). Introduction to algorithms for data mining and machine learning. Academic Press.
5. Kumar, V., Tan, P.-N., Steinbach, M. (2019). Introduction to data mining. Pearson.