ECSEVAC02 DATA ANALYSIS WITH PYTHON

UNIT – I

Handling Raw Data using Numpy and Pandas: Arrays and operations on Arrays using NumPy – Data structure of Pandas – Essential functionality – Summarizing and Computing Descriptive Statistics – Handling Missing Data - Inserting and Exporting data from CSV, XLS, JSON, database files – Data cleansing – Data Operations – Aggregation and Join operations– Data transformation: Removing duplicates – Transformation of data using function or mapping – Discretization and Binning – Renaming Axis indexes - detecting and filtering outliers – Permutation and Random Sampling.

UNIT – II

Data Visualization: Data mining – Presenting an analysis – Studying the Titanic dataset - Data Visualization - Charts – Multiple plots – Playing with text – Styling plots – Box plots – Heatmaps – Scatter plots with histograms – Area Plots – Bubble charts.

UNIT – III

Machine Learning: Types of machine learning – Decision trees – Linear regression – Logistic regression – Naïve Bayes Classifier – K means Clustering – Hierarchical clustering.

UNIT – IV

Case Studies: Performing predictions with a linear regression – Estimating the likelihood of events with Logistic regression.

UNIT – V

Case Studies: Generating recommendations with Collaborative Filtering – Applying Segmentation with k-means Clustering.

TEXT BOOKS

Samir Madhavan, “Mastering Python for Data Science”, PACKT Publishing, 2015. (ISBN 978-1-78439-015-0)

Wes McKinney, “Python for Data Analysis”, O’Reilly, 2013.(ISBN: 978-1-449-31979-3)

REFERENCES

Alberto Boschetti, Luca Massaron, “Python Data Science Essentials”, PACKT Publishing, Third Edition, 2018.

Gopi Subramanian, “Python Data Science Cookbook”, PACKT Publishing, 2015.

Jake VanderPlas, “Python Data Science Handbook”, O’Reilly, 2017. (ISBN: 978-1-491-91205-8).

Joel Grus, “Data Science from Scratch”, O’Reilly, 2015.

COURSE OUTCOMES

At the end of this course, the students will be able to:

- Apply data cleansing, transformation techniques and obtain descriptive statistics on data.
- Analyse datasets and Create simple visualization plots of data.
- Compare the machine learning techniques.
- Develop machine learning models for chosen problems of classification and prediction.
- Create machine learning models for generating recommendations and clustering data.