Format No. QSP/7.1/01.F01 (B) Issue No. 05, Rev. No. 5, Dated: Jan 1, 2017

UNIVERSITY OF PETROLEUM & ENERGY STUDIES

School of Computer Science

Dehradun

COURSE PLAN

Programme	:	B. Tech- CSE with Specialization in BAO
Course	:	Data Mining & Prediction Modeling Lab
Subject Code	:	CSIB 338
No. of credits	:	1
Semester	:	V
Session	:	July 2019 - Dec2019
Batch	:	2017-2021
Prepared by	:	Mr. Sudhanshu Srivastava
Email	:	s.srivastava@ddn.upes.ac.in

Approved By

_____ _____

Faculty HOD

UPES Campus
"Energy Acres"
P.O. Bidholi, Via Prem Nagar, Dehradun

Tel: +91-135-2770137
Fax: +91 135- 27760904
Website: www.upes.ac.in

COURSE PLAN

Subject: Data Mining & Prediction Modeling lab

Course: B.Tech (CS+BAO)

Duration: 120 min

Subject code: CSIB 338

One Session: 120 Minutes

A. OBJECTIVES:

This course is intended to provide students understand how to develop models to predict categorical and continuous outcomes, using such techniques as neural networks, decision trees, logistic regression, support vector machines and Bayesian network models. Know the use of the binary classifier and numeric predictor nodes to automate model selection. Understand when and how to use each model. Also learn how to combine two or more models to improve prediction.

PREREQUISITE:

a. Basic Knowledge of statistics
b. Basic Knowledge of programming

B. PROGRAM OUTCOMES (POs) and PROGRAM SPECIFIC OUTCOMES (PSOs) for Business Analytics and Optimization:

B1. PROGRAM OUTCOMES (POs)

PO1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.
PO5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.
PO6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.
PO7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.
PO8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
PO9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
PO10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
PO11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
PO12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

B2. Program Specific Outcomes (PSOs)
PSO1. Perform system and application programming using computer system concepts, concepts of Data Structures, algorithm development, problem solving and optimizing techniques
PSO2. Apply software development and project management methodologies using concepts of front-end and back-end development and emerging technologies and platforms.
PSO3. Ability to develop the understanding of quantitative modeling and data analysis techniques and to apply these to real world business problems, communicate findings, and effectively present results for improved decision-making.

COURSE OUTCOMES FOR Data Mining & Prediction Modeling Lab: At the end of this course student should be able to

CO1 Understand basic concepts of Data Mining, understand the KDD Process, learn various phases of CRIP- DM, and learn the challenges of Data Mining.

CO2 Know the steps involved in data understanding and preparation, read data from various sources, visualization of the data, understand how to partition and aggregate the data.

CO3 Know the various models and algorithms used in data mining (e.g. Clustering,

classification, decision trees, logistic regression, association rule, etc.)

CO4 Know how to deploy the model, understand the rule induction using CHAID, know how to validate the model, compare and combined the model, and how to access the model performance

Table: Correlation of POs and PSOs v/s COs

PO/CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	1														2
CO2				1	1			1	1						2
CO3		1	1			1				2		1			2
CO4			1				1				1				2

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

C. COURSE OUTLINE

Sr.No	Experiment	Contents
1	Experiment No 1	Introduction to Weka Tool
2	Experiment No 2	Perform basic algorithm with inbuilt data on Weka Tool
3	Experiment No 3	Introduction to Python with supporting libraries
4	Experiment No 4	Linear and Multiple Linear Regression
5	Experiment No 5	Logistics Regression Analysis
6	Experiment No 6	Decision Tree
7	Experiment No 7	Naïve Bayes
8	Experiment No 8	KNN (k-Nearest Neighbor) Algorithm
9	Experiment No 9	SVM(Support Vector Machine)
10	Experiment No 10	k-Means Clustering
11	Experiment No 11	DBSCAN (Density Based Spatial Clustering of Applications with Noise)
12	Experiment No 12	Outliers and Anomalous Data Analysis

D. PEDAGOGY

1. Lab performance on Weka/Python
2. Assignments
3. Digital and analog Presentations
4. Lab File

E. COURSE COMPLETION PLAN

Total no of lab session plan are 12. Six experiments will be cover before mid-semester examination, remaining part will be completed after mid semester examination.

F. EVALUATION & GRADING

Description	Weight age	Schedule
1. Internal Examination	50%	Detailed Below
2. End-term Exam	50%	Academic Calendar

Internal Assessment: WEIGHTAGE- 50%

Internal Assessment shall be based on the following:

Sl. No.	Description	% of Weightage out of 50%
1	Continuous Assessment	Sole Discretion of the Teacher
2	Viva-Voce/Quiz	
3	General Discipline	

Continuous Assessment: Based on the weekly evaluation of the experiments actually performed by the students in the Laboratory and submitted on the same day or on the very next turn.

MANDATARY: Min. 12 experiments to be performed by each student.

VIVA-VOCE/QUIZ: Two Viva-Voce examinations will be held; one after completion of at least three experiments by each student/three turns of Laboratory Class and second after completion of at least six experiments by each student/six turns of Laboratory Class. In addition to the Viva-Voce Examination, a Quiz examination based on objective type questions will be held after completion of at least nine experiments by each student/nine turns of laboratory class. Those who do not appear in Viva-Voce and quiz examinations shall lose their marks.

GENERAL DISCIPLINE: Based on student's regularity, punctuality, sincerity and behaviour in the class.

End -Sem Examination: 50% Weightage

End Term Examination shall be Three Hours duration and shall be conducted by actually performing the experiment.

G. Lab Experiment

Experiment No 1- Introduction to Weka Tool

Explore the tools with different options like

- Preprocess
- Classify
- Cluster
- Associate
- Select Attributes
- Visualize

Experiment No 2- Implement the different Machine Learning on Weka Tool

- Analyzing data in different views of relation contact-lenses, iris data
- Implement Classification, Association Rule, Clustering, Visualize

Experiment No 3- Introduction to Python with supporting libraries

- Data formats read by python
- Reading free-field text data files
- Reading statistics data files
- Reading Excel files
- Viewing the data

Experiment No 4- Logistics Regression Analysis

- Perform a classification rule method on given data set using logistic regression.
- Draw the residual plots for the logistic regression

Experiment No 5- Multiple Linear Regression

- Demonstrate the implementation of Multiple Linear Regression on Boston house pricing dataset using Scikit-learn

Experiment No 6- Decision Trees Algorithm

- Write a program to demonstrate ID3 Algorithm to classify dataset

Experiment No 7- Naïve Bayes

- Write a program to demonstrate Naïve Bayes Algorithm to classify dataset

Experiment No 8- KNN (k-Nearest Neighbor) Algorithm

- Perform a classification rule method on given data set using neural network.
- Perform a classification rule method using support vector machine

Experiment No 9- SVM (Support Vector Machine)

- Write a program to demonstrate SVM Classification algorithm on Iris dataset.

Experiment No 10- k-Means Clustering

- Write a program to demonstrate k-means clustering algorithm on numerical as well as image dataset.

Experiment No 11- DBSCAN (Density Based Spatial Clustering of Applications with Noise)

- Write a program to demonstrate DBSCAN clustering algorithm on given dataset

Experiment No 12- Outliers and Anomalous Data Analysis

- For the given dataset produce the summary of the variables including the details on skewness and kurtosis.
- Explore the minimum and maximum value of the variables

Experiment No 12- Looking for relationships in data

Find the correlation between the numerical variable for the given dataset.

H. SUGGESTED READINGS:

H1: TEXT BOOKS:

T1. Book provided by IBM- Data Mining & Prediction Modeling (Course code BAO2SG01 V1.0), Introduction to Machine Learning (Course Code AIML03G01 V1.0)

H2: REFERENCE BOOKS

R1. Book provided by IBM

R2.Research methodology – Methods & Techniques by C.R. Kothari

H3: VIDEO RESOURCES (URL LINK) AND NPTEL LECTURES

H4: From Wikipedia: http:// en.wikipedia.org/ wiki/ Powerball

GUIDELINES

***Cell Phones and other Electronic Communication Devices*:** Cell phones and other electronic communication devices (such as Blackberries/Laptops) are not permitted in classes during Tests or the Mid/Final Examination. Such devices MUST be turned off in the class room.

E-Mail and LMS: Each student in the class should have an e-mail id and a pass word to access the LMS system regularly. Regularly, important information – Date of conducting class tests, guest lectures, via LMS. The best way to arrange meetings with us or ask specific questions is by email and prior appointment. All the assignments preferably should be uploaded on LMS. Various research papers/reference material will be mailed/uploaded on LMS time to time.

Attendance: Students are required to have **minimum attendance of 75%** in each subject. Students with less than said percentage shall **NOT** be allowed to appear in the end semester examination.

Passing criterion: Student has to secure minimum 40% marks of the "highest marks in the class scored by a student in that subject (in that class/group class)" individually in both the 'End-Semester examination' and 'Total Marks' in order to pass in that paper.

- Passing Criterion for B. Tech: minimum 40% of the highest marks in the class
- Passing Criterion for M. Tech: minimum 40% of the highest marks in the class