COURSE TITLE	**MACHINE LEARNING**
COURSE CODE	**05MD0203**
COURSE CREDITS	**5**

Course Outcomes: After completion of this course, student will be able to:

1 Understand applications of ML in real life.
2 Implement preprocessing steps on data to make it ready for analysis.
3 Apply classification algorithms for supervised learning
4 Apply regression algorithms for predictive analysis
5 Apply clustering algorithms and association rule mining algorithms for real life problems

Pre-requisite of course: Knowledge of Python

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
4	0	2	50	30	20	25	25

Contents : Unit	Topics	Contact Hours
1	**Introduction to machine learning and preparing a model** Introduction to subject and syllabus, Human learning and types of human learning, What is machine learning?, Types of machine learning, Applications of machine learning, Tools in machine learning, Machine learning activities and basic data types, Exploring data structure – 1, Exploring data structure – 2, Data quality and remediation, data pre-processing	10
2	**Model selection and feature engineering** Machine Learning Activities, Selecting a model , Training a model, Model representation and interpretability, Features – understand your data better, Feature extraction and engineering, Feature engineering on – numerical data & categorical data & text data, Feature scaling, Feature selection, Dimensionality reduction	10
3	**Supervised learning – Classification** Intro to supervised learning and example, classification model, Introduction to Bayesian learning, importance of Bayesian methods, Bayes theorem and concept learning – 1, Bayes theorem and concept learning – 2, Bayes theorem and concept learning – 3, Bayesian belief network, Class test – 1, Common classification algorithm – K nearest neighbour, Decision tree, Random forest, Support vector machine, Evaluating performance of classification models	10

Contents : Unit	Topics	Contact Hours
4	**Supervised learning – Regression** Intro to regression and example - simple linear regression, Multiple linear regression, Assumption and problem in regression analysis, Improving accuracy, polynomial regression model, Logistic regression and maximum likelihood estimation, Evaluating performance of regression models	10
5	**Unsupervised learning** Intro to unsupervised learning, comparison , Applications of unsupervised learning, Clustering as machine learning task, techniques, Partitioning methods – K Means, K-medoids, Hierarchical clustering , DBSCAN, Evaluating performance of clustering, Introduction of Association Mining, Apriori Algorithm, Examples, Class test - 2	10
Total Hours		**50**

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**Machine Learning Practicals** Download Auto-MPG data set (https://www.kaggle.com/uciml/autompg-dataset) and perform following operations i. Write program to read dataset (Text,CSV,JSON,XML) ii. Which of the attributes are numeric and which are categorical? iii. Performing Data Cleaning • Handling Missing Data • Removing Null data iv. Rescaling Data v. Encoding Data vi. Feature Selection and Dimensionality Reduction • Implement Principle Component Analysis,, Download Airline data set (https://www.kaggle.com/open-flights/airline-database) and perform all the above operations., Write a python code to apply Naive Bayesian algorithm to classify that whether a person can buy computer or not based on given test data, Write a python code to implement decision tree for below given dataset. Identify Job offered or not., Write a python code to implement K-nearest neighbourhood program for the given dataset. (for above both the data sets), Implement supervised machine learning algorithm (Classification – K Nearest Neighbourhood) in python to classify breast tumour data into malignant breast tumour or benign breast tumour (use breast tumour dataset) and obtain its accuracy level., Implement supervised machine learning algorithm (Classification – K Nearest Neighbourhood) in python to classify iris data into setosa, virginica, versicolor using iris dataset and obtain its accuracy level., Build a classification model in python that classifies if a student gets admission in a course or not given his last two examination scores for the dataset available at https://docs.google.com/spreadsheets/d/1g0mjTUZ9Ado5prXA1UnAvNjmdzTrV0TzkFkIoU-Lpbk/edit?usp=sharing, Implement supervised machine learning algorithm (Classification – Support Vector Machine) in python to classify breast tumour data into	30

	malignant breast tumour or benign breast tumour (use breast tumour dataset) and obtain its accuracy level., Write a python program to build an email spam classifier using support vector machines for the Spam base dataset from UCI machine learning repository., Implement supervised machine learning algorithm (Classification - Naïve Bayes algorithm) in python on Pima Indians Diabetes dataset and obtain its accuracy level., Predict the CPU time if the DiskIO = 40 on the basis of following data :, Write a python code to predict profit of hotel chain given the population of the area (city) using the data at https://docs.google.com/spreadsheets/d/1Ks20skBgEefHFU36sFqVzozoFtz2EZE2rxBIgXOrUg/edit?usp=sharing., Write a python code to predict salary on the basis of experience in years using the data at https://github.com/tarunlnmiit/machine_learning/blob/master/SimpleLinearRegression.csv, Write a python code to predict the price of house given square feet and number of bed rooms in the house for the dataset available at https://docs.google.com/spreadsheets/d/1DHVK7gKo4TSyj7mFLwofHamj1Sl4SOZma2q51w1ZvyE/edit?usp=sharing, Build a logistic regression model to classify flower type based on the dataset of iris flower., Build a multivariate logistic regression model to classify glass type of glass given different glass mixture features using the Glass Identification Dataset from UCI Machine Learning Repository., Implement unsupervised machine learning algorithm (Clustering – K Means) in python on Titanic dataset to cluster data (use Titanic dataset) by removing the class label., Implement unsupervised machine learning algorithm (Clustering – K Means) in python on Breast Tumour dataset to cluster data (use Breast Tumour dataset) by removing the class label., Implement unsupervised machine learning algorithm (Clustering – Hierarchical) in python on Titanic dataset to cluster data (use Titanic dataset)., Implement unsupervised machine learning algorithm (Clustering – Hierarchical) in python on Breast Tumour dataset to cluster data (use Breast Tumour dataset) by removing the class label., Implement Apriori algorithm in python to find rules which explain association between different products for given transactions at a retail store. (The data is available at https://drive.google.com/file/d/1NUXoptUlHY8z4KcFKpFA6sQN5KnWzk3p/view?usp=sharing), Generating Association rule mining for following data set. of mobile cases., Generating Association rule mining for following data set. - Apple, bear purchase order, practical test	
Total Hours		**30**

Textbook :

1 Machine Learning, Saikat Dutt et al, Pearson, 2019

2 Practical Machine Learning with Python, Dipanjan Sarkar et al., APress, 2018

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery and evaluation					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking**

Supplementary Resources:

1 www.edx.org
2 www. coursera.org
3 www.kaggle.com

Digitally signed by (Name of HOD) **Digitally signed by (Name of Dean/ Principal)**