COURSE MODULE: MACHINE LEARNING AND DATA ANALYTICS USING PYTHON

Course Coordinator: Prof. Divyashree B K				Academic Year:**2024-25**	
Department: Master of Computer Application					
Course Code	**Course Title**	**Core/Elective**	**Prerequisite**	**Contact Hours** L:T: P:S	**Total Hrs/ Sessions**
MMC201	MACHINE LEARNING AND DATA ANALYTICS USING PYTHON	IPCC		2:2:0:0	40(TH) + 10(LAB)

Course Learning Objective:

The course will enable the students to:

1. Understand foundational concepts in machine learning and data analytics.
2. Gain proficiency in Python for data analysis and machine learning tasks.
3. Learn and apply various machine learning algorithms and techniques.
4. Develop skills in data preprocessing, visualization, and model evaluation.
5. Prepare students for industry roles involving data-driven decision making and predictive modeling.

Teaching-Learning Process (General Instruction):

1. Adopt different types of teaching methods to develop the outcomes through PowerPoint presentations and Video demonstrations.
2. Adopt collaborative (Group Learning) Learning in the class.
3. Adopt Problem Based Learning (PBL), which fosters students' Analytical skills and develops thinking skills such as evaluating, generalizing, and analyzing information.

Module-1

INTRODUCTION TO MACHINE LEARNING AND PYTHON:

INTRODUCTION TO MACHINE LEARNING: Definition and importance of machine learning, Types of machine learning: Supervised, unsupervised, and reinforcement learning, Applications of machine learning in various domains.

PYTHON FOR DATA ANALYSIS: Introduction to Python programming, Python libraries for data analysis: NumPy, Pandas, Matplotlib, Data manipulation and visualization using Pandas and Matplotlib.

DATA PREPROCESSING: Data cleaning and transformation, Handling missing values and outliers, Feature scaling and normalization.

TLP: Lectures with PowerPoint presentations, Hands-on coding exercises using Jupyter notebooks, Interactive discussions and problem-solving sessions, Assignments and quizzes for assessment.

Module-2

SUPERVISED LEARNING:

REGRESSION: Linear regression, Polynomial regression, Model evaluation metrics: MAE, MSE, RMSE. Classification: Logistic regression, K-Nearest Neighbors (KNN), Decision Trees and **Random Forests, Model evaluation metrics:** Accuracy, precision, recall, F1-score, ROC-AUC.

MODEL TRAINING AND EVALUATION: Train-test split and cross-validation, Hyper parameter tuning using GridSearchCV, Overfitting and underfitting.

TLP:Lab exercises on regression and classification models, Practical coding sessions with real- time problem-solving, Group projects on developing and evaluating supervised learning models, Continuous assessment through quizzes and coding challenges.

Module-3

UNSUPERVISED LEARNING:

CLUSTERING: K-Means clustering, Hierarchical clustering, Evaluation of clustering results. **DIMENSIONALITY REDUCTION:** Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) t-Distributed Stochastic Neighbor Embedding (t-SNE).
ASSOCIATION RULE LEARNING: Apriori algorithm, Market Basket Analysis, Evaluation metrics for association rules

TLP: Lab exercises on clustering and dimensionality reduction, Practical coding sessions with unsupervised learning techniques, Group projects on applying unsupervised learning to real-world data, Continuous assessment through quizzes and practical tests.

Module-4

ADVANCED MACHINE LEARNING TECHNIQUES:
ENSEMBLE METHODS: Bagging and Boosting, Gradient Boosting Machines (GBM), Extreme Gradient Boosting (XGBoost).
SUPPORT VECTOR MACHINES (SVM): Linear and non-linear SVM, Kernel trick, Model evaluation and tuning.
NEURAL NETWORKS AND DEEP LEARNING: Introduction to neural networks, Building and training neural networks using TensorFlow and Keras, Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).

TLP: Practical sessions on advanced machine learning techniques, Interactive coding exercises to implement neural networks, Group projects on applying advanced techniques to complex data problems, Continuous assessment through quizzes and practical tests.

Module-5

DATA ANALYTICS AND REAL-WORLD APPLICATIONS:
EXPLORATORY DATA ANALYSIS (EDA): Data visualization techniques, Statistical analysis and hypothesis testing, Identifying patterns and insights from data.
TIME SERIES ANALYSIS: Introduction to time series data, Time series forecasting using ARIMA and Prophet, Evaluating time series models.
INTEGRATING MACHINE LEARNING MODELS: Deployment of machine learning models, Building web applications with Flask and Django, Case studies on real-world applications of machine learning.

TLP: Lab exercises on EDA and time series analysis, Practical sessions on deploying machine learning models, Group discussions on real-world case studies, Final project presentation and assessment.

Course Outcomes:

At the end of the course, the student will be able to:

Sl. No.	Description	Blooms Level
CO1	Demonstrate an understanding of machine learning concepts and techniques.	L3
CO2	Perform data preprocessing and exploratory data analysis using Python.	L4
CO3	Develop and evaluate machine learning models using Python libraries.	L5
CO4	Apply machine learning algorithms to real-world data problems.	L4
CO5	Integrate data analytics and machine learning models into practical applications.	L5

Assessment Details (both CIE and SEE)

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing mark for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements and earned the credits allotted to each subject/ course if the student secures not less than 50% (50 marks out of 100) in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

CIE for the theory component of IPCC

1. Two Tests each of 20 Marks
2. Two assignments each of 10 Marks/One Skill Development Activity of 20 marks
3. Total Marks of two tests and two assignments/one Skill Development Activity added will be CIE for 60 marks, marks scored will be proportionally scaled down to 30 marks.

CIE for the practical component of IPCC

- On completion of every experiment/program in the laboratory, the students shall be evaluated and marks shall be awarded on the same day. The 15 marks are for conducting the experiment and preparation of the laboratory record, the other 05 marks shall be for the test conducted at the end of the semester.
- The CIE marks awarded in the case of the Practical component shall be based on the continuous evaluation of the laboratory report. Each experiment report can be evaluated for 10 marks. Marks of all experiments' write-ups are added and scaled down to 15 marks.
- The laboratory test at the end /after completion of all the experiments shall be conducted for 50 marks and scaled down to 05 marks. Scaled-down marks of write-up evaluations and tests added will be CIE marks for the laboratory component of IPCC for 20 marks.

SEE for IPCC

Theory SEE will be conducted by University as per the scheduled timetable, with common question papers for the course (duration 03 hours)

1. The question paper will be set for 100 marks and marks scored will be scaled down proportionately to 50 marks.
2. The question paper will have ten questions. Each question is set for 20 marks.
3. There will be 2 questions from each module. Each of the two questions under a module (with a maximum of 3 sub-questions), should have a mix of topics under that module.
4. The students have to answer 5 full questions, selecting one full question from each module.

The theory portion of the IPCC shall be for both CIE and SEE, whereas the practical portion will have a CIE component only. Questions mentioned in the SEE paper shall include questions from the practical component).

- The minimum marks to be secured in CIE to appear for SEE shall be the 15 (50% of maximum marks-30) in the theory component and 10 (50% of maximum marks -20) in the practical component. The laboratory component of the IPCC shall be for CIE only. However, in SEE, the questions from the laboratory component shall be included. The maximum of 04/05 questions to be set from the practical component of IPCC, the total marks of all questions should not be more than the 20 marks.

SEE will be conducted for 100 marks and students shall secure 40% of the maximum marks to qualify in the SEE. Marks secured will be scaled down to 50. (Student has to secure an aggregate of 50% of maximum marks of the course(CIE+SEE)

List of Text Books

Text Books:

1. Practical Machine Learning for Data Analysis Using Python Abdulhamit Subasi.
2. Advance Machine Learning with Python by John Hearty.

Reference Textbooks:

1. "Python Machine Learning: Machine Learning and Deep Learning with Python, scikit- learn, and Tensor Flow " by Sebastian Raschka and Vahid Mirjalili.
2. "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron. "Introduction to Machine Learning with Python".

The Correlation of Course Outcomes (CO's) and Program Outcomes (PO's)

Subject Code: MMC101	TITLE: MACHINE LEARNING AND DATA ANALYTICS USING PYTHON	Faculty: DIVYASHREE B K

List of Course Outcomes	Program Outcomes											
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO-1	3	2	2		2							
CO-2	3	3	2	2	3							
CO-3	3	3	3	2	3							
CO-4	3	3	3	2	3				2			
CO-5	3	3	3	3	3				3	2		
Total												

Note: 3 = Strong Contribution 2 = Average Contribution 1 = Weak Contribution - = No Contribution