MODULE FOR SUMMER INTERNSHIP PROGRAMME 2025
(BY **MANTRA ASSOCIATES & E&ICT ACADEMY IIT GUWAHATI**)
on
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING TECHNIQUES

OBJECTIVE:

1. To provide a comprehensive introduction to artificial intelligence (AI) and machine learning (ML) concepts, focusing on their practical applications.

2. To develop foundational skills in data analysis and exploratory data analysis (EDA) for understanding and preparing datasets.

3. To gain hands-on experience with core machine learning algorithms such as Support Vector Machines (SVM), Decision Trees, and Random Forests.

4. To explore advanced ML topics like convolutional neural networks (CNNs) for image and audio classification and transfer learning with pre-trained models.

5. To understand and apply natural language processing (NLP) techniques for text classification and sentiment analysis tasks.

6. To leverage Python libraries (e.g., NumPy, pandas, scikit-learn, TensorFlow, PyTorch) for developing and deploying machine learning models.

OUTCOME:

1. **Foundational Knowledge**: Demonstrate a solid understanding of AI and ML concepts, including supervised and unsupervised learning.
2. **Data Preparation Skills**: Apply techniques for data cleaning, visualization, and exploratory data analysis to prepare datasets for ML tasks.
3. **Algorithm Proficiency**: Implement and evaluate core ML algorithms for both classification and regression problems, understanding their strengths and limitations.
4. **Deep Learning Applications**: Develop and train convolutional neural networks (CNNs) for tasks such as image and audio classification, leveraging transfer learning for enhanced performance.
5. **NLP Expertise**: Utilize natural language processing methods and tools for text classification, sentiment analysis, and other linguistic tasks.
6. **Model Evaluation**: Analyze and optimize models using appropriate evaluation metrics to ensure robust performance and reliability.

DURATION: ONE MONTH **(120 HOURS)**

PREREQUISITES:

1. **Basic Programming Knowledge**: Familiarity with Python programming, including loops, functions, and basic data structures (e.g., lists, dictionaries).
2. **Mathematical Foundations**: Understanding of linear algebra, probability, and statistics at a basic level.

INTERNSHIP STRUCTURE BREAKDOWN

DAY NO. & DATE	TOPICS TO BE COVERED	TIME DURATION
DAY 1 (TUESDAY) 01-07-2025	Introduction to Machine Learning, Types of Machine Learning, Key concepts: Features, Labels, Training, and Testing	2.5 HRS
DAY 2 (WEDNESDAY) 02-07-2025	Python Libraries for Machine Learning, Introduction to NumPy and Pandas, Overview of Scikit-learn library	2.5 HRS
DAY 3 (THURSDAY) 03-07-2025	Exploratory Data Analysis (EDA), Descriptive statistics, Handling missing values, Introduction to Matplotlib, Visualizing distributions and relationships in the data	2.5 HRS
DAY 4 **(FRIDAY)** 04-07-2025	Data Pre-processing Techniques, Strategies for handling missing values, Using Pandas for missing data handling, Methods of feature scaling (Normalization, Standardization), Encoding categorical variables (One-hot encoding, Label encoding)	3.5 HRS **(MCQ TEST 1)**
DAY 5 **(SATURDAY)** 05-07-2025	PROJECT WORK (9:30 AM to 5:00 PM)	7.5 HRS (ONLINE)
DAY 6 **(SUNDAY)** 06-07-2025	PROJECT WORK (9:30 AM to 5:00 PM)	7.5 HRS (ONLINE)
DAY 7 (MONDAY) 07-07-2025	Building and Evaluating Models, Training a model using Scikit-learn, Evaluation metrics for regression (MAE, MSE, RMSE, R^2), Evaluation metrics for classification (Accuracy, Precision, Recall, F1 Score)	2.5 HRS
DAY 8 (TUESDAY) 08-07-2025	Advanced Topics and Real-World Applications, Overview of advanced algorithms (SVM, Random Forest, K-means Clustering), Brief introduction to TensorFlow and Keras	2.5 HRS
DAY 9 (WEDNESDAY) 09-07-2025	Introduction to Supervised Learning, Types of supervised learning problems: Classification vs. Regression, Basics of neural networks, Training neural networks: Forward propagation and backpropagation	2.5 HRS
DAY 10 (THURSDAY) 10-07-2025	Decision Trees and Ensemble Methods, Implementing a decision tree using Scikit-learn, Overview of ensemble methods, Random Forest, Gradient Boosting	2.5 HRS
DAY 11 **(FRIDAY)** 11-07-2025	Support Vector Machines (SVM), Kernel trick and different types of kernels, SVM for classification problem	3.5 HRS **(MCQ TEST 2)**
DAY 12 **(SATURDAY)** 12-07-2025	PROJECT WORK (9:30 AM to 5:00 PM)	7.5 HRS (ONLINE)
DAY 13 **(SUNDAY)** 13-07-2025	PROJECT WORK (9:30 AM to 5:00 PM)	7.5 HRS (ONLINE)

Day	Topic	Hours
DAY 14 (MONDAY) 14-07-2025	Model Selection and Hyper parameter Tuning, Cross-validation techniques, Model comparison and selection criteria, Grid search vs. random search	2.5 HRS
DAY 15 (TUESDAY) 15-07-2025	Building a Feedforward Neural Network from Scratch, Loss functions and optimization algorithms (SGD, Adam), Training the neural network on a simple dataset, Evaluating the performance of the neural network	2.5 HRS
DAY 16 (WEDNESDAY) 16-07-2025	Implementation of Neural Networks using TensorFlow/Keras, Training neural networks: Forward propagation and backpropagation	2.5 HRS
DAY 17 (THURSDAY) 17-07-2025	Convolutional Neural Networks (CNNs) for Image Classification, Key concepts: filters, feature maps, and receptive fields, Building a CNN using TensorFlow/Keras	2.5 HRS
DAY 18 (FRIDAY) 18-07-2025	Visualizing Classification Results, Confusion matrix, ROC curve and AUC, Precision-recall curves	3.5 HRS (MCQ TEST 3)
DAY 19 (SATURDAY) 19-07-2025	PROJECT WORK (9:30 AM to 5:00 PM)	7.5 HRS (ONLINE)
DAY 20 (SUNDAY) 20-07-2025	PROJECT WORK (9:30 AM to 5:00 PM)	7.5 HRS (ONLINE)
DAY 21 (MONDAY) 21-07-2025	Transfer Learning and Fine-Tuning Pre-trained Models, Overview of popular pre-trained models (e.g., VGG, ResNet, Inception), Implementing transfer learning with TensorFlow/Keras	2.5 HRS
DAY 22 (TUESDAY) 22-07-2025	Audio Data Processing with Spectrogram Analysis, Understanding audio data, Introduction to spectrograms, Generating and visualizing spectrograms using Python libraries (e.g., Librosa, Matplotlib)	2.5 HRS
DAY 23 (WEDNESDAY) 23-07-2025	Building an Audio Classification Model using CNNs, Training the CNN on an audio dataset (e.g., UrbanSound8K), Evaluating the performance of the audio classification model	2.5 HRS
DAY 24 (THURSDAY) 24-07-2025	Introduction to NLP and NLTK, Key concepts and challenges in NLP, Overview of NLTK library, Basic NLTK functionalities (corpora, tokenization, stemming, lemmatization)	2.5 HRS
DAY 25 (FRIDAY) 25-07-2025	Introduction to Text Classification, Preprocessing Text Data, Understanding tokenization, Stemming, Lemmatization, Cleaning text data (removing stop words, punctuation, and special characters)	3.5 HRS (MCQ TEST4)
DAY 26 (SATURDAY) 26-07-2025	PROJECT WORK (9:30 AM to 5:00 PM)	7.5 HRS (ONLINE)
DAY 27 (SUNDAY) 27-07-2025	PROJECT WORK (9:30 AM to 5:00 PM)	7.5 HRS (ONLINE)

DAY 28 (MONDAY) 28-07-2025	Types of text classification problems (binary, multi-class, multi-label), Overview of common text classification algorithms (Naive Bayes, SVM, Logistic Regression), Building a text classification pipeline using NLTK and Scikit-learn	2.5 HRS
DAY 29 (TUESDAY) 29-07-2025	Vectorizing text data (Bag-of-Words, TF-IDF), Visualizing Text Data, Using bar charts to visualize word frequency, Creating word frequency bar charts using Matplotlib, Creating word clouds using Word Cloud library	2.5 HRS
DAY 30 (WEDNESDAY) 30-07-2025	Understanding Named Entity Recognition (NER), Using NLTK for NER, Introduction to sentiment analysis, Building a sentiment analysis model with NLTK	2.5 HRS
DAY 31 (THURSDAY) 31-07-2025	DOUBT CLEARING SESSION	1 HR & VALEDICTORY SESSION

PROJECTS TO BE ASSIGNED TO THE INTERNS (MIN. 10):

1. **Iris Flower Classification:** Classify flower species using EDA and ML algorithms with the Iris dataset.

2. **Titanic Survival Prediction**: Predict passenger survival on the Titanic using logistic regression.

3. **Handwritten Digit Recognition**: Recognize digits from the MNIST dataset using logistic regression or CNNs.

4. **Email Spam Classification**: Classify emails as spam or ham using the SMS Spam Collection dataset.

5. **House Price Prediction**: Predict house prices using features from the Boston or Ames Housing dataset.

6. **Energy Efficiency Prediction**: Estimate heating and cooling loads using the Energy Efficiency dataset.

7. **Image Classification with Cats and Dogs**: Build a CNN to classify images of cats and dogs.

8. **Audio Classification**: Classify environmental sounds using the UrbanSound8K dataset.

9. **Sentiment Analysis**: Perform sentiment classification of IMDB movie reviews as positive or negative.

10. **Text Classification**: Categorize news articles into topics using the 20 Newsgroups dataset.

11. **Transfer Learning with CIFAR-10**: Use pre-trained CNNs to classify CIFAR-10 images into categories.

12. **Time Series Forecasting**: Forecast air passenger numbers using the Air Passenger dataset and LSTMs.