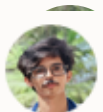
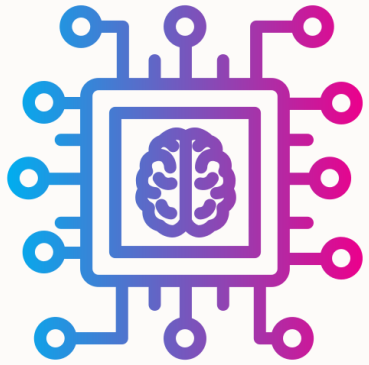




# Machine learning and Data analytics using Python



# What is Programming?

## What is Programming Language?



PRAMOD NAIK

**What is Programming? : Writing a Program**

**What is Programming Language? : Is a Collection of **Tokens**.**



**PRAMOD NAIK**

## **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.



## • **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.



## 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



## • **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).



- **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.





## Experiments or Lab Programs:

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file.
2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.
3. Write a program to demonstrate the working of the decision tree based ID3 algorithm. Use an appropriate data set for building the decision tree and apply this knowledge to classify a new sample.
4. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
5. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print both correct and wrong predictions.
6. Build an Artificial Neural Network by implementing the Backpropagation algorithm and test the same using appropriate data sets.
7. Write a program to demonstrate Regression analysis with residual plots on a given data set.
8. Write a program to compute summary statistics such as mean, median, mode, standard deviation and variance of the given different types of data.
9. Write a program to implement k-Means clustering algorithm to cluster the set of data stored in .CSV file.



## • Suggested Learning Resources:

### 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.
3. "Introduction to Machine Learning with Python".



## Web links and Video Lectures (e-Resources):

1. <https://youtu.be/7eh4d6sabA0?si=JWHNJRVs6NhQOPYe>
2. <https://youtu.be/kqtD5dpn9C8?si=LBep4HWaMFRrAPsD>
3. <https://youtu.be/4SJ7bEILPjk?si=5LurvjzUOuCew1W9>

