



Rocket Data Analytics operates as an affiliate of Data Minds Analytics pvt ltd

DATA SCIENCE INTERNSHIP

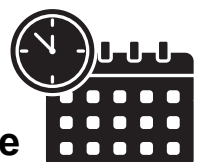


LIVE



**CLASSROOM
ONLINE
HYBRID MODEL**

2 MONTHS- Internship Course



MODULE 06

MACHINE LEARNING



Foundations of Machine Learning

Introduction to Machine Learning

- Overview of machine learning
- Setting up development environments (Python, Jupyter, sklearn libraries)

Supervised Learning

- Definition of supervised learning
- Explanation of the difference between supervised and unsupervised learning.

Regression

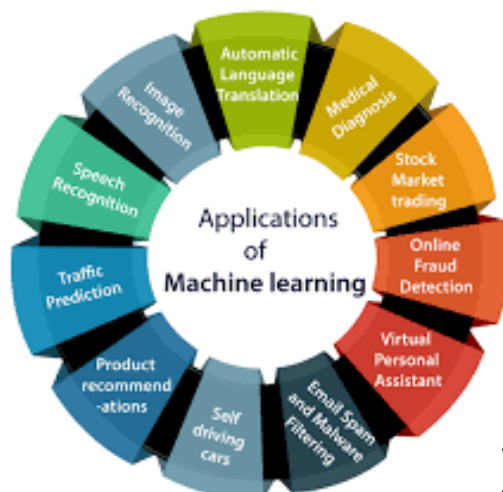
- Correlation
- Scatter Diagram
- Correlation coefficient
- Correlation analysis
- Correlation coefficient
- Regression

Linear Regression

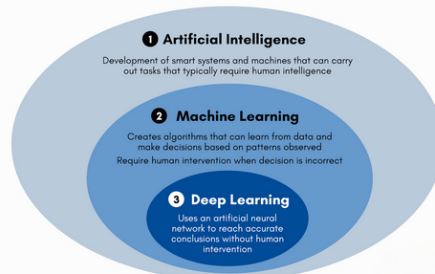
- Simple Regression
- Linear Equation - coefficients, intercept
- Residuals, Least Squares Method:
- Assumptions of Linear Regression
- Homoscedasticity. Heteroscedasticity
- Multicollinearity
- Polynomial Regression

Naive Bayes

- Probability
- Bayes Rule
- Naïve Bayes Classifier



ARTIFICIAL INTELLIGENCE VS
MACHINE LEARNING VS DEEP LEARNING



Math Fundamentals

- Python programming basics
- Linear algebra and calculus essentials

Model Selection

LIBRARIES

Classification

- Definition of classification.
- Understanding the concept of class labels.
- Binary and multiclass classification.

Logistic regression

- Types of Logistic regression
- Logit and Log-Likelihood
- Sigmoid function
- Analysis of logistic regression results
- Multiple Logistic regression
- Evaluation metrics
- Confusion matrix
- AUC / ROC for binary classifier

k-Nearest Neighbors (k-NN)

- Deciding the K value
- Thumb rule in choosing the K value.
- Building a KNN model by
- splitting the data

MODEL EVALUATION METRICS

- Mean Squared Error
- Root Mean Squared Error
- RSquare
- Adjusted RSquare
- Mean Absolute Error
- Mean Absolute Percentage Error
- Confusion matrix
- Accuracy, Precision,
- Recall, F1 Score
- Sensitivity, Specificity
- False Positive, False Negative
- True Positive, True Negative
- AUC - Area Under Curve
- ROC - (Receiver operating characteristics)

MODULE 06



MACHINE LEARNING



Decision Tree

- Elements of classification tree
- Root node, Child Node, Leaf Node, etc.
- Greedy algorithm
- Measure of Entropy
- Information gain
- Gini Index
- Decision Tree C5.0 and understanding
- Various arguments
- Checking for Underfitting and
- Overfitting in Decision Tree
- Pruning – Pre and Post Prune techniques
- Generalization and Regulation Techniques to avoid overfitting in Decision Tree

Kernel Method - SVM

- Hyperplanes
- Best Fit "boundary"
- Linear Support Vector Machine using
- Maximum Margin
- SVM for Noisy Data
- Non- Linear Space Classification
- Non-Linear Kernel Tricks
- Linear Kernel
- Polynomial, Sigmoid, Gaussian RBF
- SVM for Multi-Class Classification
- One vs. All, One vs. One
- Directed Acyclic Graph (DAG) SV

Advanced Topics in Machine Learning

Ensemble Techniques

- Ensemble Learning Introduction:
- Types of Ensemble Methods
- Bagging (Bootstrap Aggregating)
- Boosting
- Stacking

Cross-validation Techniques

- Need for Cross-Validation
- K-Fold Cross-Validation
- Leave One Out Cross Validation
- Stratified cross validation method
- Cross-Validation for Hyperparameter Tuning
- Grid search and randomized search
- For optimal hyperparameters.
- Model Selection and Comparison
- Extension to Monte Carlo Cross-Validation

Bagging - Random Forest

- Introduction to Random Forest
- Random Forest Works - Majority Voting
- Advantages of Random Forest - oob
- Random Forest Variations
- Comparison with Other Algorithms
- Evaluation and Interpretability

Boosting- AdaBoost

- Motivation for Boosting
- Feature Importance in Boosting
- Weak Learners and Strong Learners
- Decision Stump
- Assigning weights to misclassified instances
- Over fitting & Under fitting Challenges

Optimization Tehniques

- Loss Functions
- Gradient Descent
- Learning Rate and Hyperparameter Tuning
- Regularization Techniques:
- L1 (Lasso) and L2 (Ridge)regularization
- Early Stopping
- Hyperparameter Search Strategies
- Bias-Variance Tradeoff

Gradient Boosting

- Gradient Boosting
- Components of Gradient Boosting
- loss function, weak learners
- Loss Functions and Error Minimization
- Strategies for dealing with noisy data

Applications and Use Cases Live Projects

XGBoost, LightGBM and CatBoost:

MODULE 06

MACHINE LEARNING



*"predicting the future is'nt magit,
its Artificial Intelligence !"*

Introduction to Unsupervised Learning:

- Definition of unsupervised learning and its contrast with supervised learning.
- Explanation of learning from unlabeled data
- and discovering inherent patterns.

Clustering

- Distance Metrics
- k-Means clustering
- Hierarchical Clustering
- Non-Hierarchical Clustering DBSCAN
- Clustering Evaluation metrics

K-Means Clustering:

- In-depth coverage of the K-means algorithm, its initialization methods, and convergence
- Practical implementation and examples.

Association Rules

- Association rules mining
- Market Basket Analysis
- Apriori Algorithm, Fp Growth
- Metrics - Support, Confidence, Lift

Recommender Systems

- User Based Collaborative Filtering
- Similarity Metrics
- Item Based Collaborative Filtering
- Search Based Methods
- SVD Method

Dimensionality Reduction:

Principal Component Analysis (PCA):

- In-depth coverage of PCA, including eigenvalue decomposition and feature extraction.
- Applications in reducing dimensionality.

t-Distributed Stochastic Neighbor Embedding (t-SNE):

- Explanation of t-SNE and its use for visualizing high-dimensional data.
- Comparison with other dimensionality reduction techniques.

Anomaly Detection

- Discussion of anomaly detection as a task in unsupervised learning.
- Techniques for identifying unusual patterns.

Natural Language Processing (NLP)

Tokenization and text processing

Introduction to language models

- Text Mining and Natural Language Processing (NLP)
- Sources of data
- Bag of words
- Pre-processing, corpus Document
- Term Matrix (DTM) & TDM
- Word Clouds
- Corpus-level word clouds
- Sentiment Analysis
- Positive Word clouds
- Negative word clouds
- Unigram, Bigram, Trigram
- Semantic network
- Extract, user reviews of the product/services from Amazon and tweets from Twitter
- Install Libraries from Shell
- Extraction and text analytics in Python
- LDA / Latent Dirichlet Allocation
- Topic Modelling
- Sentiment Extraction
- Lexicons & Emotion Mining

Applications and Use Cases

Live Projects



ABOUT US



Rocket Data Analytics operates as an affiliate of Data Minds Analytics pvt ltd

Our professional courses are instructed by industry experts actively engaged in real-time practices, utilizing the latest teaching tools and techniques. The combination of our Learning Management System (LMS) and dedicated support mentors constitutes key elements that facilitate easy and simplified learning

- 3000 + Successfully trained students
- 1056 + Facilitated career transitions
- 10+ Industry 4.0 diverse range of Digital Transformation courses
- Flexible training options, including Classroom, Online, E-Learning, and Corporate Trainings.

DATA MINDS ANALYTICS PVT LTD

PROFESSIONAL COURSES

- | | | |
|---------------------|-------------------|---------------------------|
| ✓ Data Science | ✓ Data Analytics | ✓ Data Engineering |
| ✓ Gen - AI | ✓ Block Chain | ✓ Artificial Intelligence |
| ✓ Python Full Stack | ✓ Java Full Stack | ✓ R Language |
| ✓ Power BI / Apps | ✓ Aws / Devops | ✓ Azure Cloud |
| ✓ Cyber Security | ✓ ERP - SAP | ✓ Digital Marketing |

Our Collaborations & Placements

We foster meaningful collaborations with industry partners to enhance opportunities for our students. Our commitment to facilitating placements ensures that our graduates seamlessly transition into promising career paths, establishing a strong foundation for their professional journey.

accenture



amazon

cognizant

Deloitte.



Google



Infosys®



Internstime

