

**ZYTRIX TRAINING**

**Build your skills for Brighter Future**

**DIPLOMA IN DATA SCENCE AND MACHINE  
LEARNING CERTIFICATION COURSE SYLLABUS**

## **Module 1: Introduction to Data Science and Machine Learning**

- Overview of Data Science and its applications in various industries
- Introduction to Machine Learning concepts and algorithms
- Importance of Python programming language for Data Science and Machine Learning
- Installing Python and required libraries (NumPy, pandas, Matplotlib, scikit-learn)

## **Module 2: Data Acquisition and Cleaning**

- Understanding different types of data (structured, unstructured, semi-structured)
- Collecting data from various sources (APIs, databases, web scraping)
- Data preprocessing techniques:
  - Handling missing values
  - Removing duplicates
  - Dealing with outliers
- Data normalization and standardization

## **Module 3: Exploratory Data Analysis (EDA)**

- Statistical analysis of data:
  - Measures of central tendency and dispersion
  - Correlation analysis
- Data visualization using Matplotlib and Seaborn:
  - Histograms, box plots, scatter plots
  - Pair plots, heatmaps
- Feature engineering and selection:
  - Encoding categorical variables
  - Dimensionality reduction techniques (PCA, t-SNE)

## **Module 4: Supervised Learning Algorithms**

- Introduction to supervised learning:
  - Classification vs. regression
- Linear models:
  - Linear regression
  - Logistic regression

- Non-linear models:
- Decision trees
- Random forests
- Support Vector Machines (SVM)
- Model evaluation metrics:
- Accuracy, precision, recall, F1-score
- ROC curve, AUC-ROC

### **Module 5: Model Evaluation and Validation**

- Cross-validation techniques:
- K-fold cross-validation
- Stratified cross-validation
- Hyperparameter tuning using Grid Search and Random Search
- Bias-variance tradeoff:
- Underfitting vs. overfitting
- Model interpretation and feature importance

### **Module 6: Unsupervised Learning Algorithms**

- Introduction to unsupervised learning:
- Clustering
- Dimensionality reduction
- Clustering algorithms:
- K-means clustering
- Hierarchical clustering
- DBSCAN
- Dimensionality reduction techniques:
- Principal Component Analysis (PCA)
- Singular Value Decomposition (SVD)

### **Module 7: Time Series Analysis and Forecasting**

- Introduction to time series data:
- Trends, seasonality, noise
- Time series decomposition:
- Trend extraction, seasonal decomposition
- Forecasting methods:
- Moving averages
- Exponential smoothing
- ARIMA models

- Evaluation metrics for time series forecasting

### **Module 8: Natural Language Processing (NLP)**

- Introduction to NLP:
- Tokenization, stemming, lemmatization
- Text preprocessing techniques:
- Removing stopwords, punctuation
- TF-IDF vectorization
- Text classification:
- Naive Bayes classifier
- Support Vector Machine (SVM)
- Named Entity Recognition (NER) and sentiment analysis

### **Module 9: Neural Networks and Deep Learning**

- Basics of artificial neural networks (ANNs):
- Perceptron, activation functions
- Introduction to deep learning frameworks (TensorFlow, Keras)
- Convolutional Neural Networks (CNNs) for image classification
- Recurrent Neural Networks (RNNs) for sequence data
- Transfer learning and pre-trained models

### **Module 10: Model Deployment and Productionization**

- Converting trained models to deployable formats (PMML, ONNX)
- Deploying models using cloud platforms (AWS, Azure)
- Building RESTful APIs for model inference
- Model monitoring, logging, and versioning

### **Module 11: Ethics and Responsible AI**

- Ethical considerations in Data Science and Machine Learning
- Fairness, transparency, and accountability in AI systems

### **SOFTWARE TOOLS**

Data Science Topics		
Parts	Topics	Classes(Hours)
Python	Data Structure	4
	Functions	1
	OOPS	3
	Web and Image Scrapping, Mongo DB	3
ML visualization	Numpy	2
	Matplotlib	1
	Seaborn	1
	Plotly	1
ML	pandas	3
Statistics	CLM	1
	covariance & corelation	1
	Dispersion	1
	Probability	1
Machine Learning	Theorems	1
	ML basic	2
	EDA	2
Machine Learning(algorithms)	Linear Regression	1
	Lasso Regression	1
	Ridge Regression	1
	Logistic Regression	1
	Elastic Net Regression	1
	Cross Validation	1
	Hyper parameter Tuning	1
	Decision Tree	2
	SVC, SVM	2
	Naive Bayes	2
	Random Forest	2
	Adaboost, Xgboost, Gradientboost	2
Deep Learning(Neural Networks)	KNN, PCA	2
	Optimizers, Tensorflow, Pytorch	3
	CNN, Alex Net, LeNet, ResNet	4

	RCNN, Yolo	3
Deep Learning(GANs)	Siamese, DCGAN, WGAN	3
NLP	Tokenization, Stemming, Lemmatization	1
	RNN	2
	LSTM	2
	Transformers	2
	BERT	2
BigData	intro to BigData	1
	Hadoop	1
	Spark	1
	RDD	1
ML Projects	3-4 Projects	12
Time Series	EDA	1
	Model Building	1
Gen AI	intro	1
	Basic understanding	1
R Language	intro	2

Data Analyst Topics		
Parts	Topics	Classes
SQL	Databases	2
	Tables	2
SQL Statements	Select	1
	Update	1
	Where	1
	Delete/ Insert	1
	Order BY	1
	AND / OR / NOT	1
Aggregate functions	Min/ Max	1
	Count / Avg / Like	1
Joins	Inner /Full	1
	Left /Right /Self	1

SQL	Union /Group by/ Having	1
	Case	1
	Null Functions	1
	Stored	1
	Comments /Operators	1
Power BI	Installation	1
	Cleaning and Processing of Data	2
	DAX	2
	Power Query	2
	Power Pivot	2
	Data Visualization various charts	4
	Dashboards	2
	2-3 Projects	6
Tableau	Installation	1
	Basics	1
	Data Visualization various charts	2
	Joins & Unions	1
	Dashboards	2
	Tableau Server	1
	Sets / Parameters /Table Calculations	2