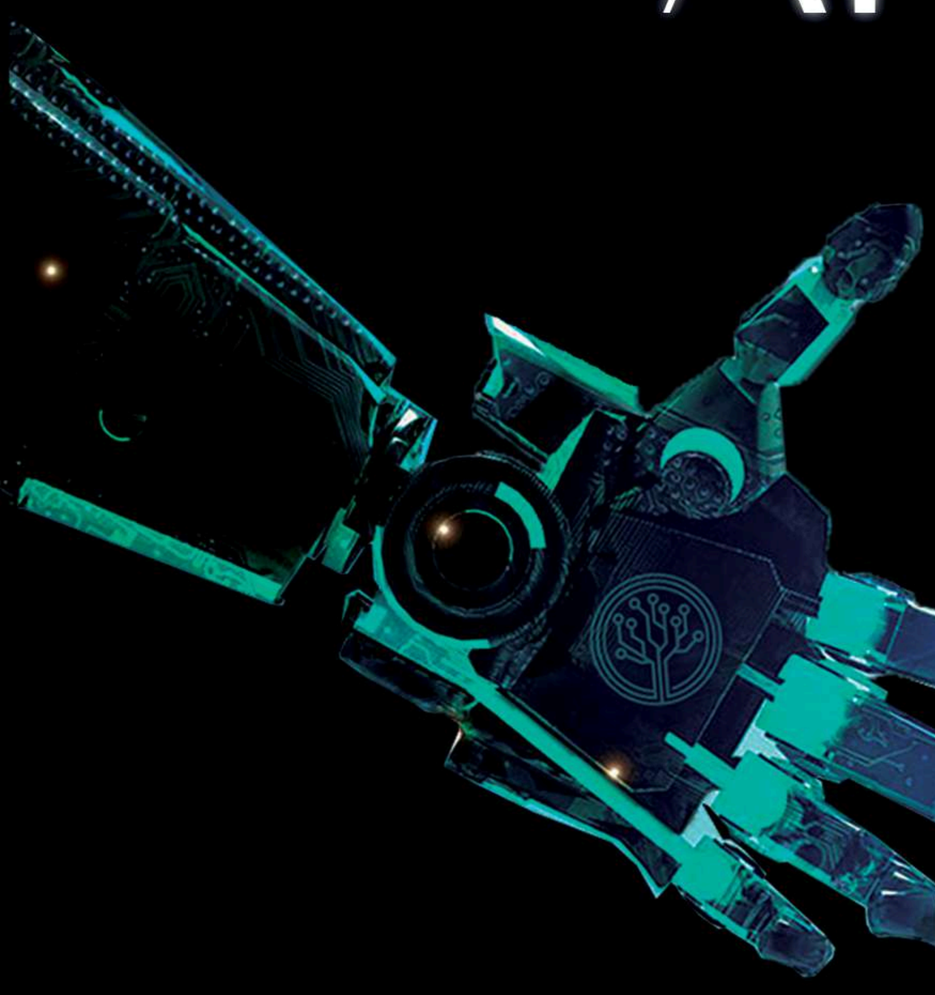
WHITE SCHOLARS

CERTIFIED DATA SCIENCE & AI COURSE

Classroom | Online



in association with

Microsoft

Why Trust Us?

- 100+ Batches
- Dedicated placement support
- 25+ hiring partners
- Senior Data scientists as faculty
- 2000+ passed out students
- Guaranteed internship opportunity
- Placement assistance

Companies that hire for Data Science

FAQS

Duration

Weekday: 4 Months

Weekends: 5 Months

Laptop Configuration

OS: Windows X,

Ram: 6GB, HDD: 1TB

Eligibility

BE/MBA/BSc/B.Tech.MSC

M.Tech - IT/Statistics/CSE

Avg. Packages

3.2-5 LPA (Fresher)

5-15 LPA (Experienced)

Certifications

White Scholars and Microsoft

Pre Requisite

No pre-requisites to learn Data Science

Roles

Data Scientist, Data Engineer, Machine Learning, Business Analytics Specialist, Data Visualization Developer, BI Engineer, BI Solution Architect, BI Specialist, Analytics Manager, Machine Learning Engineer, Statistician, Data Mining

Training Flow

1. Course Enrollment
2. Live Classes
3. Weekly Assignments
4. Project Work
5. Completion of Modules
6. Job Applications
7. Mock Interviews
8. Resume Improvements
9. Calls from HRs
10. Interviews and Tests
11. Offer Letter

Month on month journey

1

Basics of Mathematics & Statistics

- Descriptive Statistics
- Introduction to Probability
- Linear Algebra
- Inferential Integrity
- Exploratory Data Analysis & Data Visualization

2

Python Programming

- Python Language
- Intro to Python for Data Science team feature selection
- Learn exploratory data analysis
- Learn about Numpy, Pandas,Torchvision

3

Basic & Advanced ML Tools

- Decision Trees and Random Forest Test
- Clustering
- Support Vector Machines
- Dimensionally Reduction

4

Building your Profile

- Github Profile Building
- Practice via competitions like - Analytics Vidhya. Kaggle Datahack
- Discussion Forum - Kaggle Discussion

5

Apply for Job and Internship

- Identify the right jobs and apply on Naukri | Linkedin
- Analytics Vidhya | Datajobs
- Kaggle Job Portal
- Internshala
- indeed.co

Curriculum

Modules

Maths & Statistics

Python

Power BI

Machine Learning

SQL

Tableau

Concepts

1. Introduction to Programming and Python

- What is programming?
- History of Python
- Setting up the development environment
- IDLE
- Jupyter Notebook
- VS Code
- Writing and running your first Python program

2. Basic Syntax

- Comments
- Indentation
- Printing to the console

3. Variables and Data Types

- Numbers (integers, floats)
- Strings
- Booleans

4. Basic Operators

- Numbers (integers, floats)
- Strings
- Booleans

5. Control Flow

- Conditional statements (if, elif, else)
  Loops (for loops, while loops)
- break, continue, and pass statements

6. Functions

- Defining and calling functions
  Function arguments and return values
- Scope and lifetime of variables
  Lambda functions

7. Data Structures: Lists

- Creating and accessing lists
- List operations (indexing, slicing, adding, removing elements)
- List methods
- append
- extend
- insert
- remove
- pop
- clear
- index
- count
- sort
- reverse

8. Data Structures: Tuples and Sets

- Creating and accessing tuples
- Tuple methods
- Creating and accessing sets
- Set operations
- union
- intersection
- Difference

9. Data Structures: Dictionaries

- Creating, accessing, and modifying dictionaries Dictionary methods
- keys
- values
- items
- get
- pop
- update

10. Working with Strings

- String operations
- concatenation
- slicing
- formatting
- String methods
- find
- replace
- split
- join
- lower
- upper
- strip

Intermediate Concepts

11. File Handling

- Reading from files
- Writing to files
- Working with CSV files using
- the csv module

12. Modules and Packages

- Importing modules
- Standard library overview
- math
- datetime
- random
- Installing and using third-party packages (pip)

13. Error Handling

- Understanding exceptions
- try, except, else, finally blocks
- Raising exceptions

14. Object-Oriented Programming (OOP)

- Classes and objects
- Attributes and methods
- Inheritance
- Polymorphism

15. Comprehensions

- List comprehensions
- Dictionary comprehensions
- Set comprehensions

16. Decorators and Generators

- Understanding and using decorators
- Creating and using generators

17. Regular Expressions

- Introduction to regular expressions
- Using the re module for
- pattern matching

18. Advanced Data Handling with Pandas

- Introduction to Pandas
- DataFrames and Series
- Reading from and writing to different file formats
- CSV
- Excel
- JSON
- Data cleaning and manipulation

19. Data Visualization

- Introduction to Matplotlib and Seaborn
- Plotting graphs and charts
- Customizing plots

20. Web Scraping

- Introduction to web scraping
- Using BeautifulSoup and requests
- Handling web scraping challenges
- pagination
- dynamic content

21. APIs and Web Services

- Understanding APIs
- Making HTTP requests using requests
- Parsing JSON data

22. Working with Databases

- Introduction to SQL and databases
- Using SQLite with Python
- Performing CRUD operation

23. Concurrency and Parallelism

- Understanding concurrency vs. parallelism
- Using threading and multiprocessing modules

24. Testing and Debugging

- Writing unit tests with unittest and pytest
- Debugging techniques and tools

25. Advanced OOP Concepts

- Abstract classes and interfaces
- Design patterns
- Metaclasses

26. Advanced Python Libraries

- Introduction to NumPy for numerical computing
- SciPy for scientific computing
- Exploring additional libraries as per interest
- TensorFlow for machine learning

Data Visualization

AdvancedModule 1: Introduction to Data Visualization OOP Concepts

- Fundamentals of Data Visualization
- Importance of data visualization
- Key principles of effective data visualization
- Overview of common visualization types
- bar charts
- line charts
- scatter plots
- Introduction to Tableau and PowerBI
- Introduction to Tableau: installation, interface overview
- Introduction to PowerBI: installation, interface overview

Module 2: Data Connection and Preparation

Tableau

- Connecting to data sources (Excel, CSV, databases)
- Data cleaning and preparation
- Understanding Tableau data types and relationships

PowerBI

- Connecting to data sources (Excel, CSV, databases)
- Data cleaning and preparation
- Understanding PowerBI data types and relationships

Module 3: Basic Visualization Techniques

Tableau

- Creating basic charts (bar, line, pie)
- Customizing charts (colors, labels, tooltips)
- Using filters and sorting data

PowerBI

- Creating basic charts (bar, line, pie)
- Customizing charts (colors, labels, tooltips)
- Using filters and sorting data

Module 4: Advanced Visualization Techniques

Tableau

- Creating advanced charts (heat maps, tree maps, bullet charts)
- Using calculated fields
- Parameters and input controls

PowerBI

- Creating advanced charts (heat maps, tree maps, bullet charts)
- Using DAX (Data Analysis Expressions) for calculations
- Parameters and input controls

Module 5: Dashboards and Interactive Reports

Tableau

- Designing interactive dashboards
- Adding interactivity with actions (filters, highlights)
- Creating stories for data presentation

PowerBI

- Designing interactive dashboards
- Adding interactivity with slicers and filters
- Creating reports for data presentation

Module 6: Data Preparation and Cleaning Techniques

Tableau

- Data blending
- Data joins and unions
- Handling null values and outliers

PowerBI

- Data transformation using Power Query
- Data joins and merges
- Handling null values and outliers

Module 7: Advanced Data Analysis

Tableau

- Time series analysis
- Forecasting and trend analysis
- Cohort analysis

PowerBI

- Time series analysis
- Forecasting and trend analysis
- Cohort analysis

Module 8: Geographic and Geospatial Analysis

Tableau

- Creating maps
- Spatial joins and distance calculations
- Advanced map visualizations

PowerBI

- Creating maps
- Spatial joins and distance calculations
- Advanced map visualizations

Module 9: Integration with Other Tools

Tableau

- Exporting Tableau visualizations
- Embedding Tableau in websites and applications
- Connecting Tableau to R and Python for advanced analytics

PowerBI

- Exporting PowerBI visualizations
- Embedding PowerBI in websites and applications
- Connecting PowerBI to R and Python for advanced analyticss

Module 10: Best Practices and Case Studies

- Best Practices for Effective Data Visualization
- Design principles
- Avoiding common pitfalls
- Case Studies and Real-world Applications
- Reviewing industry-specific use cases
- Hands-on case study analysis

Statistics

1. Introduction to Statistics

- Descriptive Statistics
- Measures of central tendency: mean, median, mode
- Measures of dispersion: range, variance, standard deviation, interquartile range
- Data Visualization
- Histograms, bar charts, pie charts
- Box plots, scatter plots
- Probability Basics
- Probability theory and rules
- Conditional probability and Bayes' theorem
- Probability distributions: discrete and continuous

2. Probability Distributions

- Discrete Distributions
- Bernoulli
- Binomial
- Poisson distributions
- Continuous Distributions
- UniformNormal (Gaussian)
- Exponential]]\ distributions
- Central Limit Theorem
- Multivariate Distributions
- Multinomial distribution
- Multivariate Normal distribution

3. Statistical Inference

- Sampling and Sampling Distributions
- Point Estimation and Properties of Estimators
- Bias
- Variance
- Mean Squared Error (MSE)
- Interval Estimation
- Confidence intervals for means and proportions
- Hypothesis Testing
- Null and alternative hypotheses
- Type I and Type II errors
- p-values and significance levels
- t-tests
- chi-square tests
- ANOVA

4. Bayesian Statistics

- Bayesian vs Frequentist methods
- Prior, likelihood, and posterior distributions
- Bayesian inference and applications

Machine Learning

1. Introduction to Machine Learning

- Overview and history of machine learning
- Types of machine learning: supervised, unsupervised, reinforcement learning
- Key terminology
- features
- labels
- training set
- test set
- validation set

2. Data Preprocessing

- Data cleaning
- handling missing values
- outliers
- Feature scaling
- normalization
- standardization
- Encoding categorical variables
- one-hot encoding
- label encoding
- Feature engineering and selection techniques

3. Supervised Learning

- Linear Models
- Linear regression
- Logistic regression
- Decision Trees
- Construction and interpretation of decision trees
- Pruning techniques to avoid overfitting
- Ensemble Methods
- Bagging (Bootstrap Aggregating)
- Random Forests
- Boosting (AdaBoost, Gradient Boosting)
- Support Vector Machines (SVM)
- Concept of hyperplane and support vectors
-
- Kernel tricks: linear, polynomial, radial basis function (RBF)
- K-Nearest Neighbors (KNN)
- Distance metrics: Euclidean, Manhattan
- Choosing the value of K
- Gradient Boosting Machines (GBMs)
- XGBoost
- LightGBM
- CatBoost
- Advanced SVM techniques
- Advanced kernel techniques
- Support Vector Regression (SVR)

4. Unsupervised Learning

- Clustering
- K-means clustering
- Hierarchical clustering: agglomerative and divisive
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- Dimensionality Reduction
- Principal Component Analysis (PCA)
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Linear Discriminant Analysis (LDA)
- Association Rule Learning
- Apriori algorithm

5. Model Evaluation and Validation

- Train-test split
- Cross-validation
- k-fold
- stratified k-fold
- Evaluation metrics for regression
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- R^2
- Evaluation metrics for classification
- accuracy
- precision
- recall
- F1 score
- ROC-AUC

6. Introduction to Neural Networks

- Perceptron and multi-layer perceptron
- Activation functions
- sigmoid
- tanh
- ReLU
- Backpropagation and gradient descent

7. Introduction to Time Series Analysis

- Definition and examples of time series data
- Components of time series
- trend
- seasonality
- cyclic patterns
- irregular components
- Time series vs. cross-sectional data

8. Descriptive Analysis

- Plotting time series data
- Moving averages and smoothing techniques
- Seasonal decomposition of time series
- STL (Seasonal and Trend decomposition using Loess)

9. Stationarity and Differencing

- Definition of stationarity
- Augmented Dickey-Fuller (ADF) test for stationarity
- Differencing to achieve stationarity
- Seasonality adjustment

10. Autocorrelation and Partial Autocorrelation

- Autocorrelation Function (ACF)
- Partial Autocorrelation Function (PACF)
- Identifying patterns using ACF and PACF plots

11. Time Series Models

- Autoregressive (AR) models
- AR(p) process and Yule-Walker equations
- Moving Average (MA) models
- MA(q) process and invertibility
- Autoregressive Moving Average (ARMA) models
- ARMA(p,q) process and parameter estimation
- Autoregressive Integrated Moving Average (ARIMA) models
- ARIMA(p,d,q) process and model selection
- Seasonal ARIMA (SARIMA) models

12. Model Evaluation and Validation

- Splitting time series data
- training and testing sets
- Walk-forward validation
- Forecasting accuracy metrics
- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)

Advanced Machine Learning Concepts

1. Deep Learning

- Convolutional Neural Networks (CNNs)
- Convolution and pooling layers
- Famous architectures: LeNet, AlexNet, VGG, ResNet
- Recurrent Neural Networks (RNNs)
- Vanishing gradient problem
- Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)

2. Natural Language Processing (NLP)

- Text Preprocessing
- Tokenization
- Stemming
- Lemmatization
- Stopwords removal
- Word Embeddings
- Word2Vec
- GloVe
- Sequence Models
- RNNs
- LSTMs
- GRUs
- Transformers
- Attention mechanism
- BERT
- GPT models

3. Model Deployment and Productionization

- Model serving and APIs
- Containerization with Docker
- Model monitoring and management
- Scaling machine learning models in production

Deep Learning Foundations

Module 1: Deep Learning Basics

- Overview of deep learning and neural networks
- Key concepts
- neurons
- layers
- activation functions
- loss functions
- Introduction to deep learning frameworks
- TensorFlow
- PyTorch

Module 2: Optimization and Training Techniques

- Gradient descent and variants
- SGD
- Adam
- RMSprop
- Learning rate schedules and techniques to avoid overfitting
- dropout
- regularization
- Batch normalization and its impact on training

Module 3: Advanced Neural Network Architectures

- Convolutional Neural Networks (CNNs)
- Review of CNNs
- Advanced CNN architectures (ResNet, DenseNet, EfficientNet)
- Techniques to improve CNN performance
- data augmentation
- transfer learning
- Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)
- Overview of RNNs and LSTMs
- Advanced architectures: GRU, Bidirectional RNNs, and Attention mechanisms
- Applications of RNNs in time series and sequence modeling

Module 4: Transformers and Attention Mechanisms

- Introduction to Attention Mechanisms
- The need for attention in deep learning
- Types of attention mechanisms
- self-attention
- cross-attention
- Applications of attention in various domains
- Transformers: Theory and Architecture
- The transformer architecture
- encoder
- decoder
- multi-head attention
- Positional encoding and its importance
- Understanding the transformer model through the "Attention is All You Need" pape