

Data Science Course Syllabus at CSDTCentre Patna: A Comprehensive Guide

Introduction to Data Science

In the broad subject of data science, knowledge and insights are extracted from both organised and unstructured data using scientific procedures, systems, algorithms, and methods. The goal of the CSDTCentre Patna's Data Science course is to equip students with a solid foundation in data science theories and methods, enabling them to take on a variety of positions in this quickly expanding sector. The course syllabus for Data Science is described in full in this extensive handbook.

For an IT or Non-IT background student pursuing a data science course, the syllabus might include foundational IT concepts along with a comprehensive understanding of data science principles. Here is a complete syllabus:

Core Modules in Data Science

1. Introduction to Data Science

Overview of Data Science

- **Definition and Importance:** Understanding what data science is and its significance in today's data-driven world.
- **Data Science Lifecycle:** Key stages in the data science process including data collection, cleaning, analysis, and interpretation.
- **Applications of Data Science:** Real-world applications and use cases in various industries.

Setting Up the Environment

- **Installing Anaconda:** Setting up the Anaconda distribution.
- **Introduction to Jupyter Notebooks:** Using Jupyter Notebooks for data analysis and visualization.

2. Python for Data Science

Python Basics

- **Data Types and Variables:** Introduction to variables, strings, lists, tuples, sets, and dictionaries.
- **Control Structures:** Using loops and conditional statements in Python.
- **Functions and Modules:** Writing reusable code with functions and modules.

Advanced Python

- **Object-Oriented Programming:** Classes, objects, inheritance, and polymorphism.
- **Error and Exception Handling:** Writing robust code with error and exception handling.

3. Data Manipulation and Analysis

NumPy

- **Arrays and Matrices:** Creating and manipulating arrays and matrices.
- **Mathematical Operations:** Performing mathematical and statistical operations with NumPy.

Pandas

- **DataFrames and Series:** Introduction to Pandas DataFrames and Series.
- **Data Manipulation:** Merging, grouping, reshaping, and cleaning data with Pandas.
- **Handling Missing Data:** Techniques for managing missing data.

4. Data Visualization

Matplotlib

- **Basic Plotting:** Creating basic plots (line, bar, scatter) with Matplotlib.
- **Advanced Visualizations:** Customizing plots, adding annotations, and creating subplots.

Seaborn

- **Statistical Visualizations:** Creating advanced visualizations (heatmaps, pair plots, box plots) with Seaborn.
- **Customization:** Customizing and styling Seaborn plots.

5. Statistics and Probability

Descriptive Statistics

- **Measures of Central Tendency:** Mean, median, mode.
- **Measures of Dispersion:** Range, variance, standard deviation.

Inferential Statistics

- **Probability Distributions:** Understanding normal, binomial, and Poisson distributions.
- **Hypothesis Testing:** Performing t-tests, chi-square tests, and ANOVA.

6. Machine Learning

Supervised Learning

- **Regression Analysis:** Simple and multiple linear regression, polynomial regression.
- **Classification Algorithms:** Logistic regression, k-nearest neighbors, support vector machines, decision trees, random forests.

Unsupervised Learning

- **Clustering:** K-means clustering, hierarchical clustering.
- **Dimensionality Reduction:** Principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE).

Model Evaluation

- **Train-Test Split:** Splitting data into training and testing sets.
- **Cross-Validation:** Implementing cross-validation techniques.
- **Metrics:** Evaluating model performance with metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

7. Deep Learning

Introduction to Neural Networks

- **Perceptron and Multilayer Perceptron:** Basics of neural networks.
- **Activation Functions:** Understanding different activation functions (ReLU, Sigmoid, Tanh).

Deep Learning Frameworks

- **TensorFlow:** Introduction to TensorFlow and its components.
- **Keras:** Building and training neural networks using Keras.

Convolutional Neural Networks (CNNs)

- **CNN Architecture:** Understanding the architecture of CNNs.
- **Applications:** Implementing CNNs for image classification tasks.

Recurrent Neural Networks (RNNs)

- **RNN Architecture:** Understanding the architecture of RNNs.
- **Applications:** Implementing RNNs for time series forecasting and natural language processing tasks.

8. Natural Language Processing (NLP)

Text Processing

- **Tokenization:** Splitting text into tokens.
- **Stemming and Lemmatization:** Normalizing text data.

Text Analysis

- **Bag of Words:** Creating a bag of words model.
- **TF-IDF:** Calculating term frequency-inverse document frequency.

NLP Applications

- **Sentiment Analysis:** Analyzing the sentiment of text data.
- **Topic Modeling:** Identifying topics in text data using LDA.

9. Big Data Technologies

Hadoop Ecosystem

- **Introduction to Hadoop:** Understanding the Hadoop ecosystem.
- **HDFS and MapReduce:** Basics of Hadoop Distributed File System and MapReduce programming model.

Apache Spark

- **Introduction to Spark:** Understanding the Spark ecosystem.
- **Spark SQL and DataFrames:** Working with structured data using Spark SQL and DataFrames.
- **Spark MLlib:** Implementing machine learning algorithms using Spark MLlib.

10. Data Science Projects

Project Work

- **Capstone Projects:** Working on real-world data science projects.
- **End-to-End Data Science Pipeline:** Implementing the complete data science pipeline from data collection to model deployment.

Career Path and Opportunities in Data Science

Key Roles for Data Scientists

1. **Data Scientist:** Analyzing complex data to extract actionable insights.
2. **Data Analyst:** Interpreting data and generating reports.
3. **Machine Learning Engineer:** Developing and deploying machine learning models.
4. **Data Engineer:** Building and maintaining data infrastructure.
5. **Business Analyst:** Using data to drive business decisions.

Skills Required

- **Technical Skills:** Proficiency in programming (Python, R), data manipulation (Pandas, SQL), and machine learning.
- **Analytical Skills:** Strong ability to analyze and interpret complex data.
- **Communication Skills:** Effective communication of insights and findings.

Career Growth and Opportunities

Professionals with expertise in data science are in high demand in a variety of sectors, including technology, finance, healthcare, and retail. Professionals can progress to senior positions like Chief Data Officer (CDO), Lead Data Scientist, and Data Science Manager with extensive data science abilities.

Certifications to Boost Your Career

- **Certified Data Scientist (CDS):** Validates expertise in data science concepts and tools.
- **Google Data Analytics Professional Certificate:** Recognizes proficiency in data analysis using Google tools.
- **IBM Data Science Professional Certificate:** Demonstrates skills in data science and machine learning.

Salary Expectations

Data science professionals enjoy competitive salaries. According to industry reports, the average salary for a Data Scientist ranges from \$90,000 to \$150,000 per year, with higher salaries for specialized roles and senior positions.

Conclusion

Gaining expertise in data science can lead to a variety of job options in the ever-evolving fields of machine learning and data analysis. This extensive curriculum offers a strong basis for advancing data science abilities and succeeding in a variety of tech-related careers.

Csdtdo Centre