* **Data Science Mastery Certification Syllabus**

**(K.M.IQBAL , Duration : 24weeks – 250 Hours, lessons - 27 )** Here’s a **general employability-focused Data Science syllabus** aimed at building skills needed to land a job as a data scientist or related roles. This syllabus covers practical and industry-relevant topics to ensure job readiness.

**Topics Covered:**

**1. Introduction to Data Science**

* What is Data Science?
* Applications in real life (healthcare, finance, biology, etc.)
* Data Science workflow/lifecycle

**2. Programming and Data Manipulation**

**Core Programming Skills**

* **Python:** Data Types, Data Structures,Function and OOP Concepts
* **R Programming:** Data manipulation and statistical analysis
* **SQL:** Writing complex queries, joins, aggregations

**3. Python for Data Science**

* Installing Python and Jupyter Notebook
* Data types: int, float, string, boolean
* Lists, Tuples, Dictionaries, Sets
* Conditional statements and loops
* Functions and basic error handling

**4. Working with Libraries**

* NumPy: Arrays, vectorized operations
* Pandas:
* Series and DataFrames
* Reading CSV/Excel files
* Filtering, sorting, indexing
* Aggregations and groupby

**5. Data Cleaning (Preprocessing)**

* Handling missing values
* Handling duplicates
* Data type conversions
* String operations and date-time handling
* Dealing with outliers
* Data normalization and transformation

**6. Exploratory Data Analysis (EDA)**

* Descriptive statistics (mean, median, mode, std)
* Data visualization for understanding:
* Matplotlib: Line, Bar, Scatter plots
* Seaborn: Histograms, Boxplots, Heatmaps
* Detecting outliers and skewness

**7. Statistical Foundations**

* Types of data: Nominal, Ordinal, Interval, Ratio
* Measures of central tendency and spread
* Basics of probability
* Probability Theory and Distributions
* Normal, Binomial, Poisson, Exponential
* Hypothesis Testing:
* t-test, ANOVA, Chi-Square Test
* Confidence Intervals and p-values
* A/B Testing and Experimental Design

**8. Data Visualization and Storytelling**

**Visualization Techniques**

* **Python:** Matplotlib, Seaborn, Plotly
* **R:** ggplot2
* **Interactive** **Dash boarding:** Tableau, Power BI

**Storytelling with Data**

* Creating clear, impactful visualizations
* Crafting narratives from data insights
* Presenting analysis to non-technical stakeholders

**9. Real-World Mini Projects**

* Example datasets: Titanic, Iris, Student performance, Sales data
* Practice data loading, cleaning, analysis, and visualization
* Summary writing and basic insights

**Machine Learning**

**10. Introduction to Machine Learning**

* What is Machine Learning?
* Types of Machine Learning: Supervised, Unsupervised, Reinforcement (overview)
* ML Workflow: Data → Model → Evaluation → Deployment
* Train-Test Split and Cross-validation

**11. Supervised Learning Algorithms**

**Regression Techniques**:

* Linear Regression and polynomial Regression
* Assumptions, implementation, metrics
* Logistic Regression for classification
* Binary classification, sigmoid function
* Lasso

**Classification Algorithms:**

* Decision Trees
* Gini index, entropy, pruning
* Random Forest
* Bagging and feature importance
* k-Nearest Neighbors (k-NN)
* Distance metrics, choosing k
* Support Vector Machines (SVM)
* Naïve Bayas

**12. Model Evaluation and Metrics**

* Classification metrics:
* Accuracy, Precision, Recall, F1-score
* Confusion Matrix
* ROC Curve and AUC
* Regression metrics:
* MAE, MSE, RMSE, R²
* Cross-validation (k-Fold, Stratified)

**13. Feature Engineering**

* Feature creation: Interaction terms, domain knowledge features
* Feature selection:
* Correlation, Chi-square, Mutual Information
* Recursive Feature Elimination (RFE)
* Encoding:
* One-Hot Encoding
* Label Encoding
* Scaling:
* MinMaxScaler, StandardScaler, RobustScaler

**14. Unsupervised Learning Algorithms**

* k-Means Clustering
* Elbow method, silhouette score
* Hierarchical Clustering
* Dendrograms, linkage methods
* DBSCAN
* Density-based clustering
* Dimensionality Reduction:
* PCA (Principal Component Analysis)
* t-SNE (for visualization)

**15. Time Series Analysis (Introduction)**

* Components: Trend, Seasonality, Residuals
* Time-based features (day, month, hour, etc.)
* Simple Moving Average (SMA)
* ARIMA (intro concept)

**16. Handling Real-World Datasets**

* Working with:
* Imbalanced datasets
* Large datasets (chunking)
* Noisy or dirty data
* Pipelines and workflow organization using scikit-learn

**17. Intermediate-Level Mini Projects**

* Titanic survival prediction (with modeling)
* Customer segmentation (clustering)
* Sales forecasting (time series)
* Heart disease classification (health dataset)

**Deep Learning**

**18. Deep Learning Basics**

* Introduction to Neural Networks
* Neurons, activation functions (ReLU, Sigmoid)
* Forward and backward propagation
* Building simple neural networks with TensorFlow/Keras
* Overfitting and regularization (Dropout, L2)
* Optimizers: SGD, Adam, RMSProp

**19. Convolutional Neural Networks (CNNs)**

* Concepts: filters, pooling, strides, padding
* Image classification tasks
* Transfer learning using pretrained models (e.g., VGG, ResNet)

**20. Recurrent Neural Networks (RNNs) and LSTM**

* Understanding sequences and time dependency
* RNNs vs LSTM vs GRU
* Applications: time series forecasting, sequence modeling

**21. Natural Language Processing (NLP)**

* Text Preprocessing:
* Tokenization, Lemmatization, Stemming
* Feature Extraction:
* TF-IDF, Word Embeddings (Word2Vec, GloVe)
* NLP Models:
* Sentiment Analysis
* Text Classification
* Named Entity Recognition (NER)
* Advanced Techniques:
* Transformers (BERT, GPT)
* Language Modeling

**22. Time Series Forecasting (Advanced)**

* Stationarity, ACF, PACF
* ARIMA, SARIMA, SARIMAX
* Facebook Prophet
* Deep learning for time series: LSTM/GRU models

**23. Big Data for Data Science**

* Introduction to Hadoop, HDFS
* Spark basics (PySpark for distributed computing)
* MapReduce concepts
* Working with big datasets in chunks (Dask, Vaex)

**24. Data Engineering for Data Scientists**

* Data Pipelines: Building ETL processes
* Data Warehousing:
* SQL and NoSQL Databases (MySQL, MongoDB)
* Big Data Technologies:
* Apache Hadoop and Spark

**25. Deployment and Model Serving**

* Building APIs using Flask or FastAPI
* Containerization with Docker
* Deploying on Cloud Platforms (AWS, Azure)
* Monitoring and Maintenance of Deployed Models

**26. Responsible AI & Ethics**

* Fairness and bias in models
* Privacy and data anonymization
* Explainable AI (XAI)
* Ethics in automated decision-making

**27. Capstone Projects (Advanced Level)**

 **Choose something a project.**

* **Project 1:** Predicting Loan Defaults using Logistic Regression
* **Project 2:** Building a Recommender System for E-commerce
* **Project 3:** Image Classification using CNN
* **Project 4:** Real-Time-Sales-Data-Analysis-Application
* **Project 5:** Industry-Specific Case Studies:
* Finance: Fraud Detection
* Healthcare: Disease Prediction
* Biology : Predicting heart disease using machine learning
* Retail: Sales Forecasting

**28. Soft Skills and Professional Development**

* Resume Building and Interview Preparation:
* Mock Interviews
* Behavioral Interview Tips

**Educational Background**

* **Bachelor’s Degree** in:
* BE / B.Tech (From any Branch)
* BBA / B.Com / BSc (in any Branch)
* **Master’s Degree**
* M.Tech / MBA / MCA / M.Sc. / M.Com (in any Branch)

**Data Science tools**

**Python** / **R / SQL** / **Pandas** / **NumPy** / **Matplotlib** / **Seaborn** / **Scikit-learn**

**XGBoost** / **TensorFlow** / **PyTorch** / **Jupyter Notebook** / **Google Colab**

**Kaggle** / **Docker** / **Flask** / **Apache Spark** / **Tableau / Power BI**