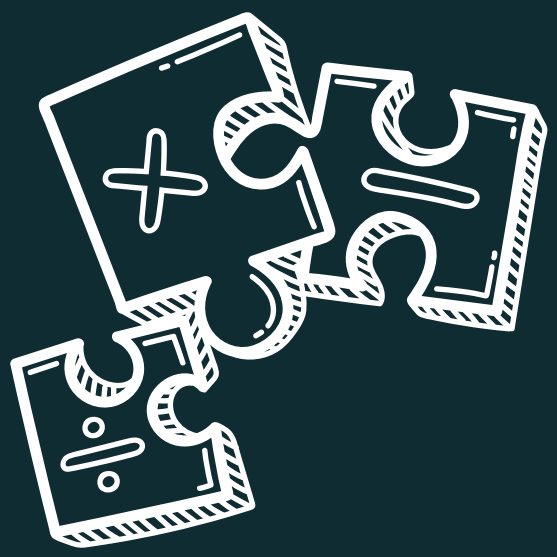
Data Scientist RoadMap



Foundational Knowledge:

Mathematics:

Linear Algebra, Calculus, Probability and Statistics

Data Manipulation and Visualization:

Programming:

Python:

Syntax and Basic Concepts, Data Structures, Control Structures, Functions , Object-Oriented Programming

R (optional, based on preference)

Swipe >>>

Data Manipulation and Visualization:

Data Manipulation:

Numpy (Python),
Pandas (Python),
Dplyr (R)

Data Visualization:

Matplotlib (Python),
Seaborn (Python),
ggplot2 (R),
Interactive Visualization Tools

Exploratory Data Analysis (EDA) and Preprocessing:

Swipe >>>

Machine Learning:

Supervised Learning:

Regression:

Linear Regression , Polynomial Regression, Regularization Techniques,

Classification:

Logistic Regression, k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Decision Trees, Random Forest, Gradient Boosting

Unsupervised Learning:

Clustering:

K-means, DBSCAN, Hierarchical Clustering, Dimensionality

Reduction:

Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (tSNE), Linear Discriminant Analysis (LDA), ⇔ Association Rule Learning

Deep Learning:

Reinforcement Learning,
Model Evaluation and Validation:

Swipe >>>

Deep Learning:

Neural Networks:
Perceptron, Multi-Layer Perceptron (MLP),

Convolutional Neural Networks (CNNs):
Image Classification, Object Detection, Image Segmentation,

Recurrent Neural Networks (RNNs):
Sequence-to-Sequence Models , Text Classification, Sentiment Analysis

Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU):
Time Series Forecasting, Language Modeling

Generative Adversarial Networks (GANs):
Image Synthesis, Style Transfer, Data Augmentation

Advanced Topics:

Swipe >>>

Big Data Technologies:

Natural Language Processing (NLP):
Text Preprocessing, Word Embeddings (e.g., Word2Vec, GloVe), Recurrent Neural Networks for NLP, Transformer Models (e.g., BERT, GPT)

Time Series Analysis:
Time Series Decomposition, Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), Exponential Smoothing Methods ,Prophet

Recommender Systems:
Collaborative Filtering, Content-Based Filtering, Matrix Factorization, Hybrid Methods

Advanced Topics:

Causal Inference:
Experimental Design, Observational Studies, Propensity Score Matching, Instrumental Variable Analysis,

Advanced Deep Learning:
Advanced Architectures (e.g., Transformers, GPT models), Generative Models (e.g., VAEs, flow-based models), Advanced Techniques for NLP and Computer Vision

Bayesian Statistics and Probabilistic Programming:
Bayesian Inference, Markov Chain Monte Carlo (MCMC), Probabilistic Graphical Models, Stan, PyMC3, or Edward for Probabilistic Programming

Swipe >>>

Big Data Technologies:

Hadoop:
HDFS, MapReduce
Spark:
RDDs, DataFrames, MLlib
NoSQL Databases:
MongoDB, Cassandra, HBase, Couchbase
Stream Processing Frameworks: Apache Kafka, Apache Flink,Apache Storm

Domain Knowledge and Soft Skills:

Industry-specific Knowledge, Problem-solving, Communication Skills, Time Management, Teamwork

Data Visualization and Reporting:

Dashboarding Tools:
Tableau, Power BI, Dash (Python), Shiny (R), Storytelling with Data, Effective Communication

Ethical Considerations and Bias in Data Science:

Swipe >>>

Recommended Resources:

Deployment and Productionization:

Model Deployment Techniques, Containerization (e.g., Docker), Model Serving and APIs, Scalability and Performance Optimization

Continuous Learning and Staying Updated

Online Courses and Tutorials, Books and Research Papers, Blogs and Podcasts, Conferences and Workshops, Networking and Community Engagement

Ethical Considerations and Bias in Data Science:

Fairness in Machine Learning, Bias Detection and Mitigation, Privacy and Data Security

Swipe >>>

Recommended Resources

Online Courses:

- Coursera - Data Science Specialization,
- edX - Data Science MicroMasters Program Kaggle Courses

Books:

- "Python for Data Analysis" by Wes McKinney,
- "Hands-On Machine Learning with Scikit-Learn and TensorFlow" by Aurélien Géron,
- "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville.

YouTube Channels:

- Sentdex,
- Data School,
- 3Blue1Brown,
- PyData ,
- StatQuest with Josh Starmer

Swipe >>>

DID YOU FIND THIS POST HELPFUL ?

Follow us on

@engineer_bhaiya_yt

@engineeringwalabhaiya

topmate.io/shivan_kumar1

For Mentorship & Free Text Queries,
WhatsApp at: +91 6280 963 446

Like, Save, and Share with your friends !!!

Shivan Kumar

Data Scientist & Kaggle Master