DATA SCIENCE ROADMAP

THE RAVIT SHOW

INTRODUCTION

DATA SCIENCE IS A FIELD THAT USES SCIENTIFIC METHODS, PROCESSES, ALGORITHMS, AND SYSTEMS TO EXTRACT KNOWLEDGE AND INSIGHTS FROM STRUCTURED AND UNSTRUCTURED DATA.

DATA SCIENCE IS AN IMPORTANT FIELD IN VARIOUS INDUSTRIES, INCLUDING HEALTHCARE, FINANCE, MARKETING, AND MORE.

FOUNDATION IN MATHEMATICS AND STATISTICS

- **LINEAR ALGEBRA:** VECTORS, MATRICES, MATRIX MULTIPLICATION, EIGENVECTORS, AND EIGENVALUES.

- **CALCULUS:** LIMITS, DERIVATIVES, AND INTEGRALS.

- **PROBABILITY AND STATISTICS:** PROBABILITY THEORY, STATISTICAL INFERENCE, BAYESIAN STATISTICS, AND STATISTICAL MODELLING.

- **HYPOTHESIS TESTING:** HYPOTHESIS FORMULATION, NULL HYPOTHESIS, ALTERNATIVE HYPOTHESIS, STATISTICAL SIGNIFICANCE, AND P-VALUES.

PROGRAMMING LANGUAGES

- **PYTHON:** WIDELY USED IN DATA SCIENCE FOR ITS SIMPLICITY, VERSATILITY, AND LARGE COMMUNITY OF USERS. PYTHON LIBRARIES FOR DATA SCIENCE INCLUDE NUMPY, PANDAS, MATPLOTLIB, SCIKIT-LEARN, AND TENSORFLOW.

- **R:** SPECIFICALLY DESIGNED FOR DATA ANALYSIS AND HAS A WIDE RANGE OF STATISTICAL TOOLS BUILT-IN. R PACKAGES FOR DATA SCIENCE INCLUDE DPLYR, GGPLOT2, TIDYR, AND CARET.

- **SQL:** USED FOR QUERYING AND MANIPULATING DATA IN RELATIONAL DATABASES. ESSENTIAL FOR WORKING WITH LARGE DATASETS.

DATA WRANGLING

- **DATA CLEANING:** REMOVING IRRELEVANT, INCORRECT, OR INCOMPLETE DATA.

- **DATA PREPROCESSING:** SCALING, NORMALIZING, STANDARDIZING, AND ENCODING DATA.

- **DATA TRANSFORMATION:** FEATURE ENGINEERING, AGGREGATION, DISCRETIZATION, AND BINNING.

DATA VISUALIZATION

- **TYPES OF VISUALIZATIONS:** BAR GRAPHS, HISTOGRAMS, SCATTERPLOTS, HEATMAPS, BOXPLOTS, AND MORE.

- **TOOLS FOR VISUALIZATION:** MATPLOTLIB, SEABORN, PLOTLY, TABLEAU, AND MORE.

- **GOOD PRACTICES FOR VISUALIZATION:** CHOOSING THE RIGHT CHART, USING COLORS EFFECTIVELY, REMOVING CHART JUNK, ADDING CONTEXT, AND TELLING A STORY.

MACHINE LEARNING

- **SUPERVISED LEARNING:** REGRESSION, CLASSIFICATION, DECISION TREES, RANDOM FORESTS, AND GRADIENT BOOSTING MACHINES.

- **UNSUPERVISED LEARNING:** CLUSTERING, DIMENSIONALITY REDUCTION, PRINCIPAL COMPONENT ANALYSIS, AND SINGULAR VALUE DECOMPOSITION.

- **REINFORCEMENT LEARNING:** MARKOV DECISION PROCESSES, Q-LEARNING, AND POLICY GRADIENT METHODS.

DEEP LEARNING

- **NEURAL NETWORKS:** FEEDFORWARD NETWORKS, CONVOLUTIONAL NEURAL NETWORKS, AND RECURRENT NEURAL NETWORKS.

- **CONVOLUTIONAL NEURAL NETWORKS:** IMAGE RECOGNITION, OBJECT DETECTION, AND IMAGE SEGMENTATION.

- **RECURRENT NEURAL NETWORKS:** NATURAL LANGUAGE PROCESSING, TEXT GENERATION, AND SPEECH RECOGNITION.

NATURAL LANGUAGE PROCESSING

- **PREPROCESSING TEXT:** TOKENIZATION, STEMMING, LEMMATIZATION, AND STOP WORD REMOVAL.

- **FEATURE EXTRACTION FROM TEXT:** BAG OF WORDS, TF-IDF, WORD EMBEDDINGS, AND SENTIMENT LEXICONS.

- **SENTIMENT ANALYSIS:** BINARY CLASSIFICATION, MULTI-CLASS CLASSIFICATION, AND ASPECT-BASED SENTIMENT ANALYSIS.

- **TEXT CLASSIFICATION:** SPAM FILTERING, TOPIC MODELING, AND NAMED ENTITY RECOGNITION.

BIG DATA

- **HADOOP:** DISTRIBUTED STORAGE AND PROCESSING OF LARGE DATA SETS.

- **SPARK:** FAST AND GENERAL ENGINE FOR LARGE-SCALE DATA PROCESSING.

- **NOSQL:** NON-RELATIONAL DATABASES FOR UNSTRUCTURED AND SEMI-STRUCTURED DATA.

REAL-WORLD APPLICATIONS

- **INDUSTRY-SPECIFIC APPLICATIONS:** HEALTHCARE, FINANCE, MARKETING, AND MORE.

- **CASE STUDIES:** REAL-WORLD EXAMPLES OF HOW DATA SCIENCE IS USED IN PRACTICE.

- **FUTURE TRENDS:** EMERGING TECHNOLOGIES AND TRENDS IN DATA SCIENCE,INCLUDING EXPLAINABLE AI, AUTOMATED MACHINE LEARNING, AND QUANTUM COMPUTING.

THE RAVIT SHOW