Complete Data Science Roadmap

Beginners guide

- Programming
  - Python
    - Data structures: https://www.youtube.com/watch?v=D_ZG3N80ziA
    - Pandas: https://youtube.com/playlist?list=PL31XenDVPq_n5H4lJzOZUtnPzyZUie51t
    - Numpy: https://youtube.com/playlist?list=PL31XenDVPq_1_DVMqdAvbL1R8R5anWwPh
    - Matplotlib: https://youtu.be/8qAeUe5oY7k
    - Seaborn
  - R
- Statistics
  - Mean, Median, Mode
  - Confidence Interval
  - Null hypothesis, Alternate hypothesis
  - Statistical tests: https://youtu.be/52UlOiLgBas
- Probability
  - Distributions: https://www.youtube.com/watch?v=XUZcqx1rOSI&t=94s
  - pdf, pmf, cdf
  - QQ-plot
  - KL-divergence
  - Conditional probability: https://www.youtube.com/watch?v=254xt1VJLLo
- Machine learning
  - Exploratory Data Analysis
    - Data cleaning
    - Preprocessing
    - Handling missing and null values
    - Outliers: https://www.youtube.com/watch?v=fed0ApMSfSE&t=1s
    - Class imbalance: https://www.youtube.com/watch?v=my2NQkBCyDc
  - Supervised

        - Regression
            - Linear Regression: https://www.youtube.com/watch?v=-uC1ZP61EYg&t=504s
        - Classification
            - Logistic Regression: https://www.youtube.com/watch?v=N6l46rYSCpM
            - K-nn: https://www.youtube.com/watch?v=cY6NFyLghzM&t=310s
            - Decision Tree: https://www.youtube.com/watch?v=wsH55R5dJCY&t=1s
            - SVM: https://www.youtube.com/watch?v=KlrDOh2WobU&t=332s
            - Naive Bayes: https://www.youtube.com/watch?v=T5x2haAR4rE
            - Random Forest
            - Gradient Boosting

    - Overfitting and Underfitting: https://www.youtube.com/watch?v=zQB7-gFjHgk&t=1s
    - Regularization
    - Dimensionality Reduction: https://www.youtube.com/watch?v=NmGXU-4QajE&t=331s
    - Cross validation: https://www.youtube.com/watch?v=SdjMU2iqTG4&t=41s
    - Loss & metrics: https://www.youtube.com/watch?v=AgHXr2CDjNo&t=2s
    - Unsupervised
        - Clustering
            - K-means

    - Reinforcement

- **Interview Preparation:**
    - Soft skills: https://youtu.be/Hp9zvuV2-o8
    - Multiple choice questions
    - Scenario based questions
    - Industrial exposure: https://www.youtube.com/watch?v=YbshjBO0DGU&t=576s

- **Solving use cases**
  - Binary Classification
    - Text based
    - Tabular
  - Multiclass classification
    - Tabular
  - Regression

Advanced

- Linear Algebra
  - Vectors and matrices
  - Eigen vectors & eigen values
  - SVD: https://www.youtube.com/watch?v=iWw3QxhgDoo&t=37s
  - NMF
- Deep learning
  - Perceptron
  - Neurons and activations
  - CNN
  - RNN
  - Popular Architectures
    - Alexnet
    - VGGNet
    - UNet
  - Capsule Net
  - Understanding popular applications
    - ChatGPT
    - Dall e-2
  - Transformers
    - Bert
    - GPT
- Interview Prep
- Use Cases
  - Text based
  - Image based

- Productionisation
  - Data leakage: https://www.youtube.com/watch?v=vtKTTtj_668
  - Docker
  - Airflow
  - Flask and Rest API