Data Science Roadmap 2025-2026 [MLE Version]

Stastics

Descriptive Stats

1. Central tendency(Mean,Median,Mode)
2. Dispersion(Rage,variance,std dev,iqr)
3. Distributions (e.g., normal distribution)
4. Data visualization: histograms, box plots, bar charts

Inferential Stats

1. **Probability theory**
2. **Sampling distributions**
3. **Hypothesis testing** (t-tests, z-tests, chi-square, ANOVA)
4. **Confidence intervals**
5. **Regression models**
6. **Bayesian inference**
7. **A/B testing**
8. **Central limit Theorem**

Machine Learning

Basics

1. What are Features and labels(X,y)
2. Numerical vs Categorical/Structured vs unstructured
3. Handling Missing values
4. Feature Scaling/mormalization
5. Overfitting /underfitting
6. Bias vs Variance tradeoff
7. Train Validation blind sets
8. Cross validation
9. Training process
10. Metrices

Classification

1. Binary/multiclass
2. Distance based algos(logistic,KNN,SVM)
3. Boosting vs Bagging
4. Tree Based(Decission tree,RF,XG Betc..)
5. Precission ,recall,f1 score,accuracy
6. AUC ROC curv
7. Precission recall curv
8. Class imbalance
9. Feature selection(RFE,Permutation improtance)

Regression

MLOps

1. ML Flow
2. basic CI CD pipelines
3. AWS EC2/Lambda etc..
4. REST API building understanding
5. Docker/Kubernates

1. Assumptions of linear regression
2. loss/cost functions
3. optimizers
4. Eval metrics(mse,rmse,mae,r2)
5. L1 and l2 regularization
6. multicollinearity
7. feature engineering
8. k fold cross validation
9.