ML INTERVIEW PREPARATION – FULL MOCK SESSION

Candidate Profile:

18+ years overall experience | 1 year ML/DS

Target Role: ML Engineer / Analyst

LINEAR REGRESSION

• Definition, assumptions, cost function

• Bias–Variance tradeoff

• Overfitting vs Underfitting

• Regularization (L1 vs L2)

• Multicollinearity, Outliers

• Evaluation metrics

BIAS – VARIANCE (INTERVIEW DEPTH)

• Bias: error from overly simple models

• Variance: error from overly complex models

• Overfitting = high variance

• Underfitting = high bias

• Prefer slight bias over high variance in production

• Regularization increases bias, reduces variance

CROSS-VALIDATION & METRICS

• K-Fold Cross Validation

• Avoid CV for time-series data

• Metrics: Precision, Recall, F1, ROC-AUC

• Accuracy not suitable for imbalanced data

• Evaluate mean + std across folds

PROJECT: WRONG / DOUBLE PAYMENT PREDICTION

Business Problem:

• Duplicate and incorrect invoice payments

• Manual verification costly and slow

Solution:

• Kafka-based real-time pipeline

• ML model classifies invoices as valid/suspicious/wrong

• Only suspicious invoices manually verified

Model:

• Supervised classification

• Random Forest chosen for robustness and non-linearity

• Handled class imbalance using oversampling + class weights

Evaluation:

• Precision, Recall, F1, ROC-AUC

• Focus on recall for wrong payments

Production:

• No data leakage (time-based split, offline training)

• Kafka used only for transport

• Threshold tuning for false positives

• Periodic retraining and monitoring

INTERVIEW GOLD LINES

• Prefer stability over aggressive accuracy

• Metrics should reflect business risk

• Kafka moves data; learning happens offline

• Slight bias is better than high variance in production