

## **Claim Denial Risk Prediction – Interview Q&A; (1 Page)**

### **Project Overview**

Q: What problem were you solving?

A: Predict insurance claim denial risk early to reduce rework, delays, and operational cost.

Q: What was your approach?

A: Built an ML pipeline with feature engineering, model training, batch scoring, and real-time API deployment.

### **Data & Feature Engineering**

Q: What features did you engineer?

A: Billing patterns, length of stay flags, age buckets, insurance risk flags, provider experience indicators, and diagnosis presence.

Q: How did you handle missing data?

A: Used median imputation, default values, and rule-based flags to keep logic simple and consistent.

### **Modeling & Evaluation**

Q: Which model did you choose and why?

A: Logistic Regression for better recall, interpretability, and stability on imbalanced data.

Q: How did you handle class imbalance?

A: Used class\_weight='balanced' and threshold tuning to improve recall for denied claims.

### **Case Study Questions**

Q: What if billing amount is negative but model still predicts low risk?

A: Input validation should be handled before the model. The model focuses on risk patterns, not rule enforcement.

Q: How would you reduce false positives?

A: Tune thresholds, add more claim lifecycle features, and retrain with updated data.

Q: How would you improve performance if recall is low?

A: Adjust decision threshold, rebalance training data, and add denial-specific features.

### **Production & Deployment**

Q: How did you ensure training and inference consistency?

A: Used a single sklearn pipeline for feature engineering, preprocessing, and prediction.

Q: How is batch scoring handled?

A: New data placed in input folder, scheduled job runs daily, outputs scored claims CSV.

Q: How is the model deployed?

A: Exposed via FastAPI, containerized using Docker, deployed on Render.

### **Model Limitations**

Q: What are the limitations?

A: Depends on historical data quality, no hard business rule validation, requires periodic retraining.

### **Production Experience**

Q: Do you have real production ML experience?

A: Worked with production healthcare data and batch pipelines. Designed this project to mirror real-world ML workflows.