PROBLEM:

Predict which companies have high layoff risk based on their layoff history, funding information, and industry metrics. To help investors and stakeholders identify unstable companies early.

DATA:

- Dataset: Layoffs Dataset
- Features: 9
  (industry, stage, funds_raised, region, recency,
  events_deviation, recency_deviation, layoff_events_category, funds_raised_binned)
- Target: high_risk (1 = high risk if risk_score >= 65th percentile, 0 = low risk)
- Train/Test split: 80/20 (968 train, 242 test)

MODEL:

- Best Model: XGBoost with hyperparameter tuning
- Optimal Parameters:
  max_depth = 3
  learning_rate = 0.01
  n_estimators = 300
  subsample = 0.7
  colsample_bytree = 0.7
- Decision Threshold: 0.5 (default)

RESULTS:

- Accuracy: 78.51%
- Precision: 67.37%
- Recall: 75.29%
- F1-Score: 0.7111
- AUC: 0.8375

The model successfully identifies 75.29% of high-risk companies.
Out of positive predictions, 67.37% are correct.
Overall, the model provides useful risk assessment with room for improvement.

LIMITATIONS:

1. Missing features: company growth rate, CEO changes, market sentiment, etc.
2. Limited dataset size (1,209 companies), which may restrict the model's ability to generalize.
3. Risk threshold (65th percentile) is arbitrary; optimal threshold selection could improve model performance.

NEXT STEPS:

1. Collect additional features (revenue growth, CEO changes, hiring trends, news sentiment)
2. Combine models trained on different feature groups (layoff history, financials, sentiment) to improve stability and reduce variance.
3. Add prediction confidence intervals instead of binary classification
4. Conduct fairness analysis to ensure the model behaves consistently across industries and regions