

A dark blue background featuring abstract white and light blue geometric shapes. These include a large triangle at the top, several curved lines forming loops and swirls, and a globe-like sphere with a grid pattern. The overall aesthetic is futuristic and technological.

Bank Marketing Campaign Analysis

Predictive modeling for term deposit subscription using machine learning

By Matt LeRoi

11/12/2025

Business Understanding

Campaign Context

Portuguese bank direct marketing via phone calls

Goal: predict term deposit subscriptions

Success Metrics

ROC-AUC score optimization

Balance true positive vs. false positive rate

Identify key features



Dataset Overview

45K

17

11.7%

Total Records

Complete dataset
entries

Features

Customer and
campaign attributes

Conversion Rate

Clients who
subscribed



Key Feature Categories



Demographics

Age, job, marital status, education level



Financial Profile

Balance, credit default, housing loan, personal loan



Campaign Contact

Contact type, day, month, call duration



Previous Campaigns

Contact history, prior outcomes, campaign count

Data Limitations and Considerations

Cultural Context

Data originates from Portuguese customers; observed behavioral patterns and marketing responses may not universally apply to other markets, such as the US, due to cultural differences.

Missing Values

Key features like 'previous outcome' and 'contact type' contain a notable percentage of unknown entries. Despite this, these features still demonstrate predictive power, indicating potential hidden information or biases.

Correlation vs. Causation

The predictive model highlights strong correlations between features and term deposit subscription. However, these statistical relationships do not necessarily imply direct causal links, requiring deeper investigation for actionable insights.

Data Preparation Strategy

01

Feature Encoding

One-hot encoding for categorical variables

02

Data Splitting

75% training, 10% validation, 15% test

03

Feature Scaling

StandardScaler fitted on training data

04

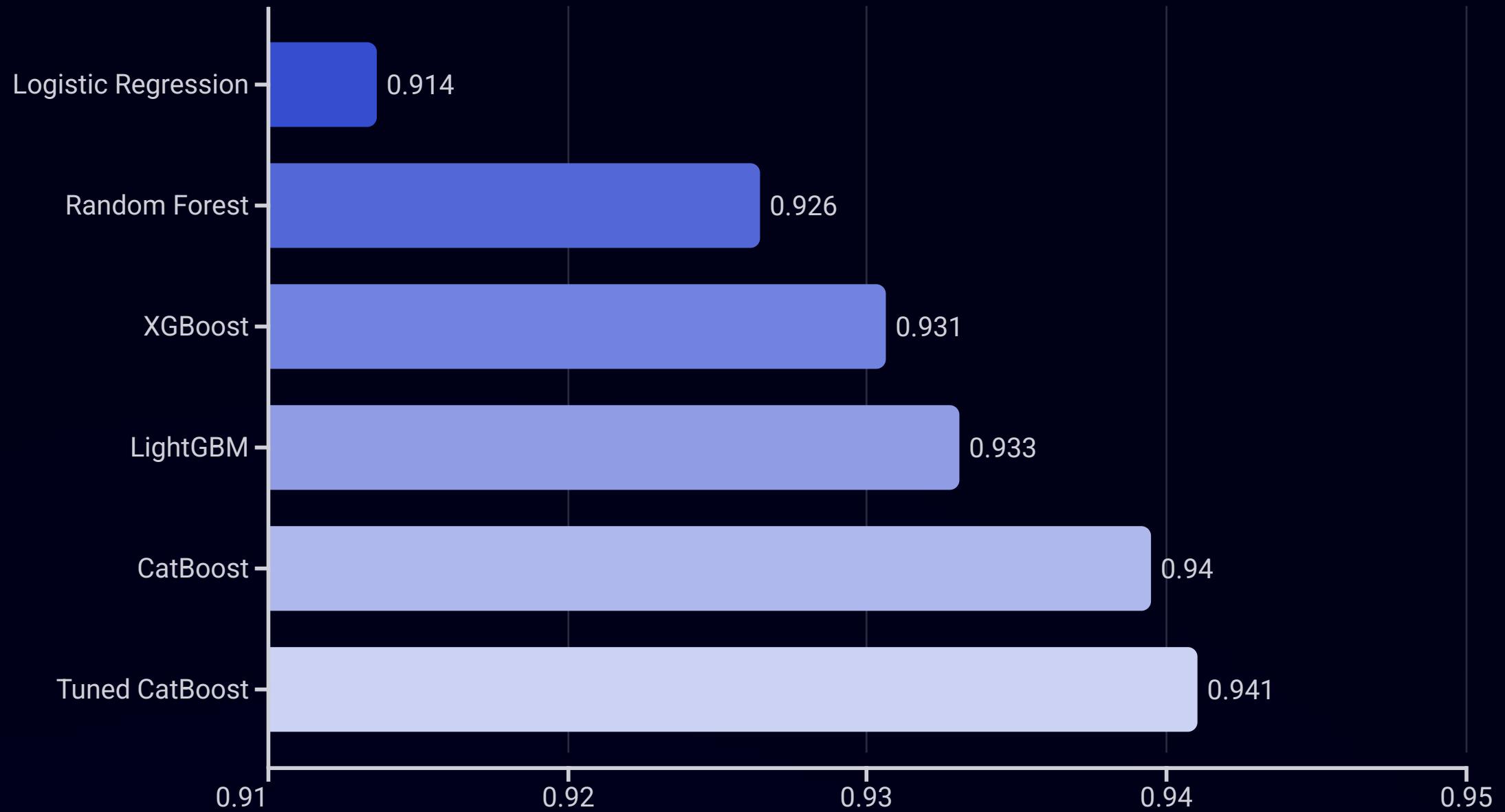
Class Balancing

SMOTE oversampling at 0.35 ratio

Data Cleaning Preprocessing Workflow



Model Performance Comparison (ROC-AUC score)



CatBoost emerged as the top performer after hyperparameter optimization

Top Predictive Features

Duration

Call duration most critical predictor across all models

Previous Outcome

Success in prior campaigns strongly indicates conversion

Housing Loan

Customers without housing loans more likely to subscribe

Contact Method

Cellular contact shows higher conversion rates

Campaign Count

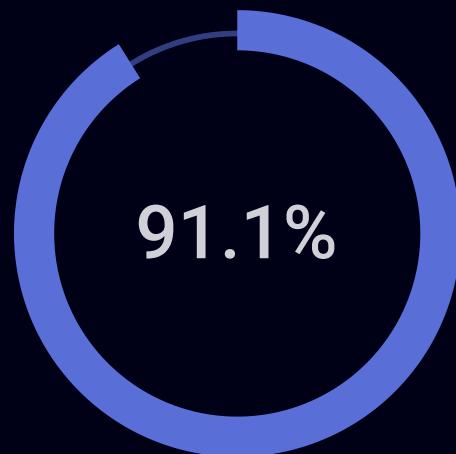
Number of contacts correlates with subscription likelihood

Final Model Performance

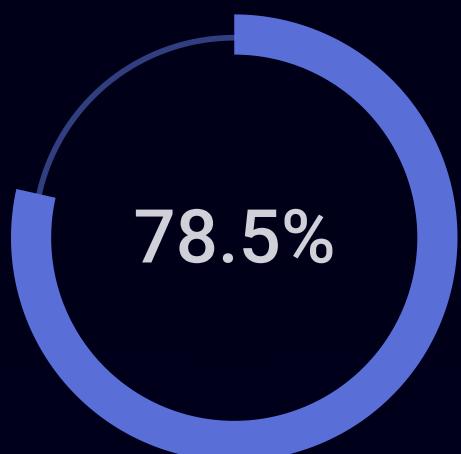
Test Set Results



Overall model score



Correctly identified non-subscribers



Successfully predicted subscribers

Optimization Details

- Depth: 8 layers
- Learning rate: 0.07
- L2 regularization: 1
- Class weights: 1:3 ratio
- Early stopping: 50 rounds



Strategic Recommendations



Run a causality experiment

Test whether longer call durations or more contacts causally drive conversions versus correlation

Collect more complete data

Filling in the missing data for previous outcomes and contact type may lead to further insights.

Experiment for seasonality

There may be differences in signing rates based on the month, but the existing data is not conclusive.

Optimize Contact Strategy

Train agents based on the learnings from this model and these experiments



Next Steps

Implementation Phase

1. Deploy model to production environment
2. Integrate with CRM system for real-time scoring
3. Train sales team on insights
4. Set up monitoring dashboards

Continuous Improvement

1. Collect feedback from controlled experiments
2. Retrain model with new campaign data
3. Refine feature engineering based on results
4. Iterate on targeting strategies

Thank You!

Email: mcleroi@gmail.com

GitHub: @mattleroi

LinkedIn: linkedin.com/in/mcleroi/