

Marketing Conversion Prediction

Keisha Walker | Case study | October 2026

This project evaluates whether seven customer and prior-campaign features can rank term-deposit subscriptions in historical bank marketing data. The selected random forest showed weak discrimination on the final holdout. The main finding is the large change in subscription rates over time, which limits confidence in both ranking and probability estimates.

Question and data

The research question is whether a model can prioritize likely subscribers before a current campaign call. The study uses 41,188 records from UCI's bank-additional-full.csv, ordered by date from May 2008 through November 2010. It is an educational marketing analysis, not a financial eligibility or lending system. [1]

The original seven inputs are job, education, default, housing, loan, previous, and poutcome. Call duration is excluded because it is only known after the call. Six categorical inputs use one-hot encoding; previous is standardized. All preprocessing is fitted on training data only.

Evaluation design

Ordered split	Records	Subscribers	Subscription rate
Train	24,712	1,188	4.81%
Validation	8,238	912	11.07%
Test	8,238	2,540	30.83%

The first 60% of rows train the models, the next 20% select the model, and the last 20% provide the final test. Row order is preserved rather than shuffled. These are ordered blocks, not separately verified calendar periods.

A constant training-rate baseline is compared with logistic regression and a random forest. The forest uses 300 trees, maximum depth 8, minimum leaf size 30, and seed 42. These settings are fixed rather than searched. Validation average precision selects the model; neither test scores nor labels select it. The selected training-fitted model is not refitted on validation data.

Contribution and provenance

The original Steps 1-6 script is preserved unchanged. The completed package adds an AI-assisted continuation for model comparison, frozen holdout evaluation, ranking summaries, data checks, and documentation. These additions should be reviewed and understood before presenting them as independent work.

Results and outcome drift

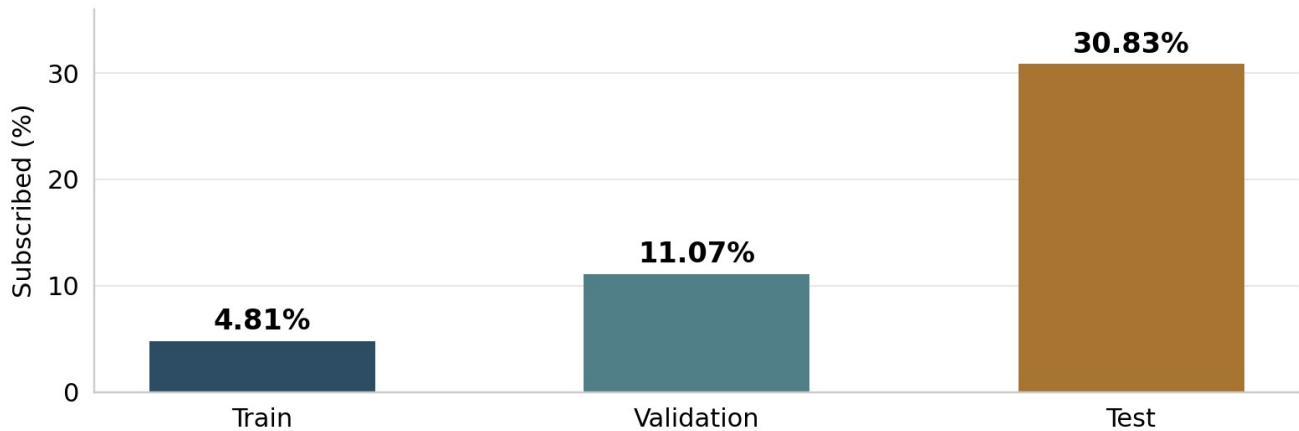
The random forest was selected by validation average precision, 0.1478 versus 0.1457 for logistic regression. This small observed gap does not establish meaningful superiority. Logistic regression had the higher validation ROC AUC and slightly lower Brier score.

Split	Model	AP ↑	ROC AUC ↑	Brier ↓
Validation	Training-rate baseline	0.1107	0.5000	0.1024
Validation	Logistic regression	0.1457	0.5718	0.1020
Validation	Random forest	0.1478	0.5610	0.1020
Test	Random forest	0.3393	0.5376	0.2796
Test	Training-rate baseline	0.3083	0.5000	0.2810

AP is average precision, a ranking summary across recall levels. Its constant-score reference equals the split's subscription rate. ROC AUC measures ranking discrimination, with 0.5 as the chance reference. Brier score measures squared probability error; lower is better. Metrics are rounded to four decimals.

On the previously examined test block, ROC AUC was 0.5376 and AP was 0.3393, compared with a prevalence reference of 0.3083. These results show limited ranking separation. The constant baseline still predicts the 4.81% training rate, far below the 30.83% test rate, so its probability estimates are badly mismatched to the holdout.

Subscription rates rise across the ordered blocks



Source: reports/split_summary.csv. Outcome-rate drift is observed; these aggregates alone do not identify its cause or prove a specific type of concept drift.

Illustrative top 20 percent ranking

The 1,648 highest-scored test records contain 572 subscribers: 34.71% precision, 22.52% recall, and 1.126 times the full-test subscription rate. The 20% capacity is an illustration, not a profit-optimized threshold. This is a retrospective concentration measure, not evidence that contacting these people causes additional subscriptions.

Limitations and next steps

- The public data are historical and come from one banking context. Results cannot be assumed to transfer to a current organization, customer population, or campaign.
- There are no null cells, but unknown is an explicit category, including 8,597 default values. Twelve exact duplicate rows are retained to preserve the original ordered split. No exact full rows overlap across splits, but absent customer IDs prevent a customer-level independence check.
- The same ordered holdout was examined in earlier project work. This is a repeated evaluation, not a new independent test. No uncertainty interval or significance test supports the small model differences. Do not tune on this holdout.
- The analysis does not establish reliable probability calibration, campaign revenue, incremental conversions, or fairness across groups. Job, education, and financial-status fields require careful use review. The model is not production ready.

A stronger next experiment

Use rolling temporal validation and a new final holdout to assess stability. Evaluate calibration on recent development data and compare simpler baselines. Investigate how the available features and outcomes change over time, and verify which inputs are available at the actual scoring moment. Before operational use, review consent and fairness, define contact costs and capacity, then use an appropriately designed experiment to measure incremental campaign effects.

Plain language walkthrough

The code reads the data, checks labels and data quality, and keeps the rows in their original order. It learns category encodings and model patterns from the first block. The middle block chooses between two fixed model candidates. That decision is written to a file before the chosen model scores the final block. CSV reports and figures record what happened, including the changing subscription rate and the limited test performance.

Suggested interview explanation

“My project asks whether information known before a marketing call can help rank likely subscribers. I kept the original ordered split and excluded call duration to avoid using information learned after the call. An AI-assisted continuation compared a fixed random forest with logistic regression and evaluated the selected model on a previously examined holdout. The forest’s test ROC AUC was only 0.538. The most important lesson was that average precision must be interpreted against the outcome rate, which changed sharply between blocks. I would validate stability and calibration before recommending any operational use.”

Modest resume wording

Marketing conversion prediction portfolio using Python and scikit-learn: chronological validation, baseline comparison, and outcome-drift analysis on 41,188 public records, with AI-assisted extensions.

Sources and reproducibility

[1] Moro, S., Rita, P., and Cortez, P. (2014). Bank Marketing. UCI Machine Learning Repository. <https://doi.org/10.24432/C5K306>. Dataset licensed CC BY 4.0: <https://creativecommons.org/licenses/by/4.0/>. This analysis selects seven inputs and creates ordered splits; source records are retained.

Results: reports/validation_metrics.csv, test_metrics.csv, split_summary.csv, ranking_metrics.json, data_quality.json, and frozen_model_decision.json. Implementation: marketing_conversion.py. Original source: original/marketing_conversion.py. Recorded run: Python 3.12.14, pandas 2.2.3, scikit-learn 1.8.0.