

The background of the slide is a collage of financial and business-related items. In the top left, a portion of a black calculator is visible, showing keys for '3', '6', '+', and '='. Below the calculator, there are several charts: a bar chart at the top with months from May to December on the x-axis, a pie chart in the center, and a line graph at the bottom left with data points connected by lines. To the right of the pie chart, there is a legend with colored squares corresponding to months from January to July. In the top right corner, a stack of Euro banknotes is fanned out. Below the banknotes, a silver compass is shown with its needle pointing towards the top right. At the bottom of the slide, a black pen lies diagonally across a table with numerical data. The table has several columns of numbers, including 125,058, 125,487, 124,000, 154,568, 56,845, 110,000, 150,000, 35,000, 95,054, 97,511, 99,011, 99,216, 101,090, 101,684, 101,962, 124,500, 125,000, 154,000, 95,000, 154,200, 110,000, 89,000, 50,000, and 700.

Bank Marketing Success Classification problem Rui Yuan


Clients and problem

Bank wants to know about what their target clients are (that are likely to subscribe bank term deposit or other financial products).

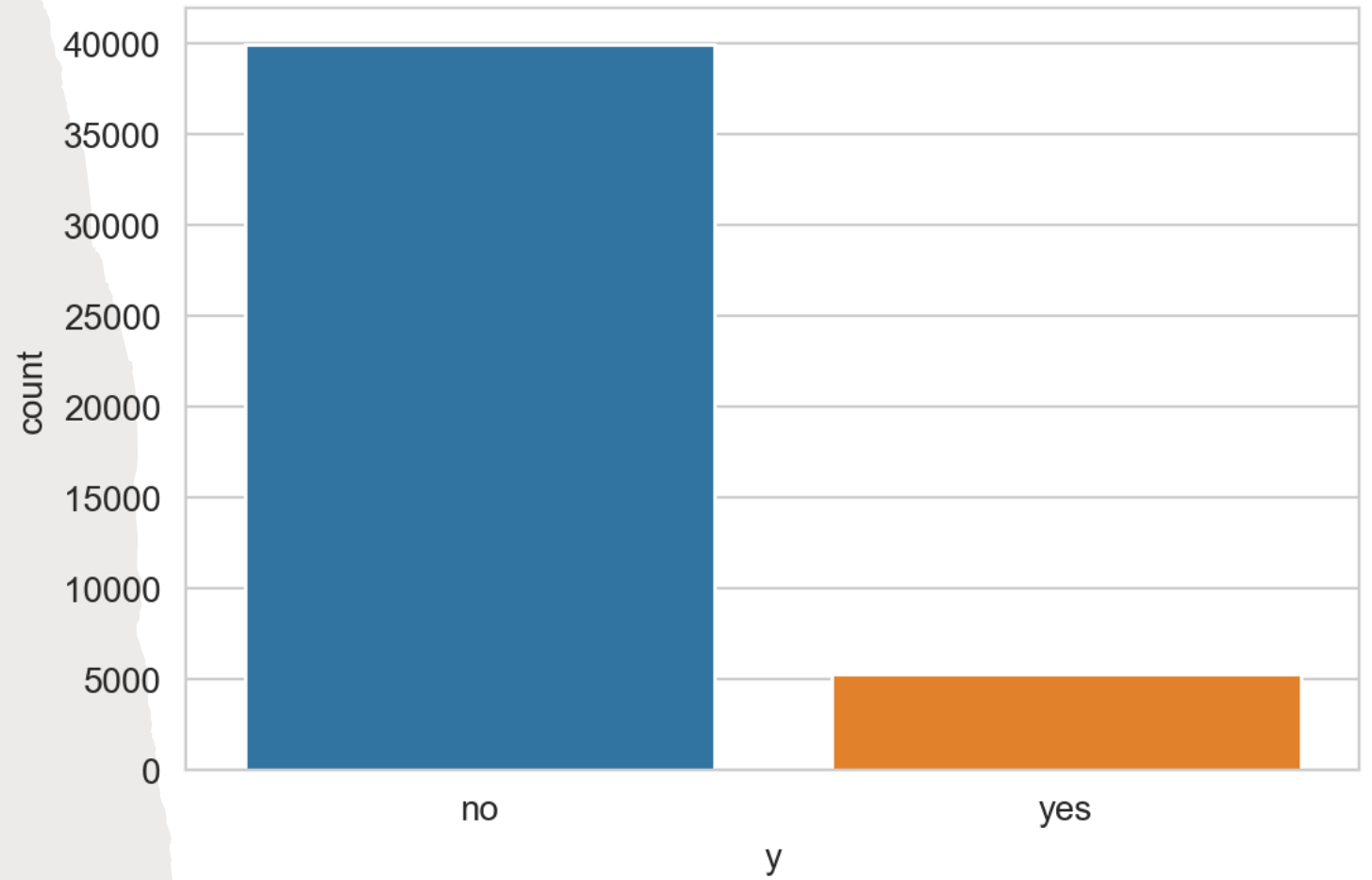
Data

- The data is about bank telemarketing which consists of clients' information, communication frequency and marketing outcome.
- 45211 rows
- Target: Whether client subscribe a term deposit (1 = yes, 0 = no)
- 15 Features: clients: age, bank balance, loan....

Classification Modeling Goal

- Optimization Goal:  precision score on positive class (subscribe) & number of successful deals
- In Business Sense: Adjustable model depending on the business capability. I.e., how many clients bank can reach to or how many phone calls bank can do in a given period (month).

Class Distribution




Modeling & Performance Metric Report

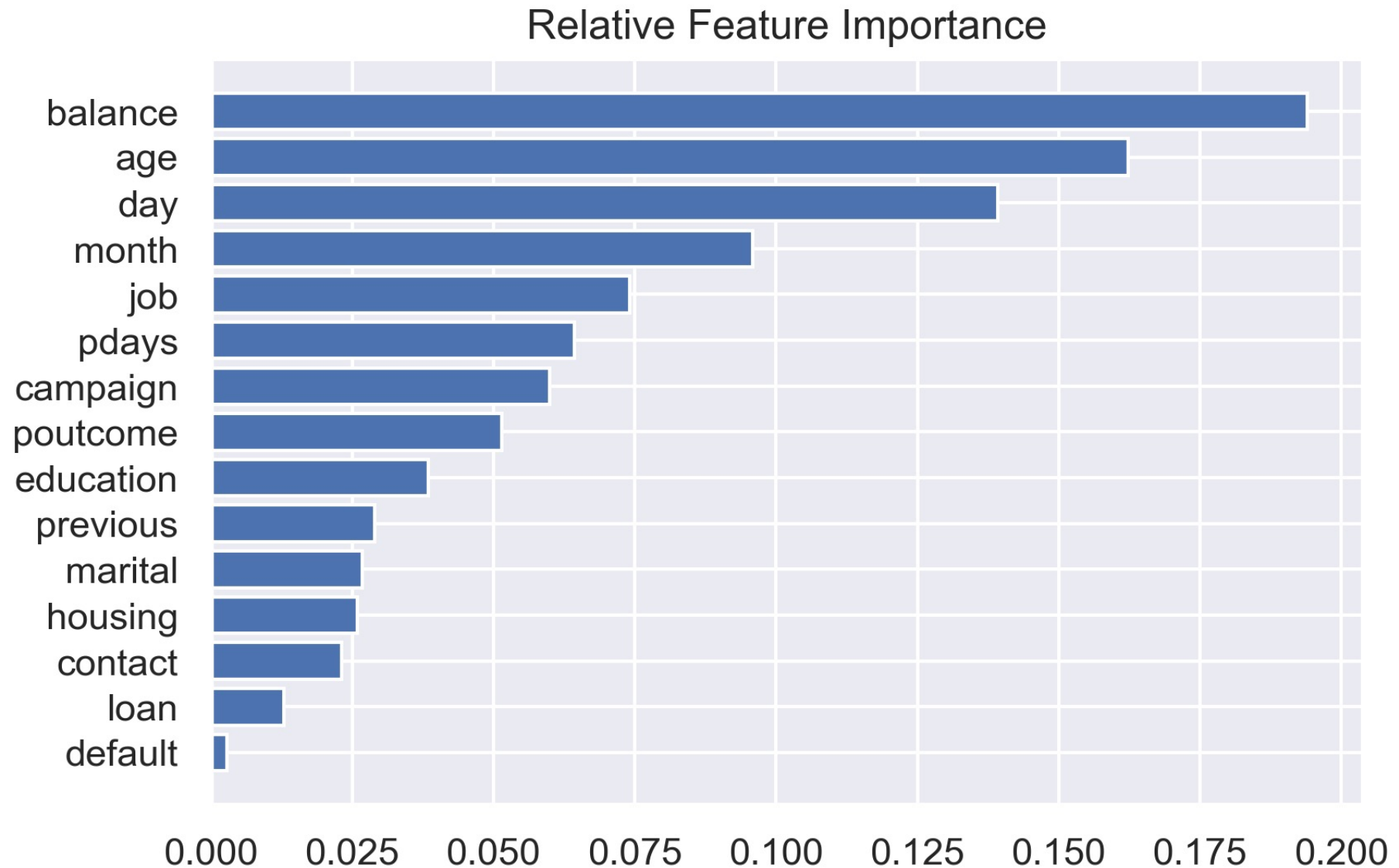
Phase 1: Model Testing

Model	Precision	Recall	F-1	Accuracy
1. K-Nearest Neighbor Baseline	0.43	0.12	0.18	0.88
2. K-Nearest Neighbor Optimized with Grid Search	0.48	0.04	0.07	0.88
3. Logistic Regression Baseline	0.50	0.00	0.00	0.88
4. Logistic Regression Regularized	0.48	0.04	0.07	0.88
5. Random Forest Baseline	0.69	0.21	0.32	0.89
6. Random Forest Optimized with Random Search	0.71	0.19	0.30	0.89

Phase 2: Handle Class Imbalance

Model	Precision	Recall	F-1	Accuracy
7. Random Forest with Sampling method	0.55	0.29	0.38	0.89
8. Random Forest with Adjusted Class Weight	0.69	0.19	0.30	0.89
9. Random Forest with Probability Threshold Adjustment 	Adjustable	Adjustable	Adjustable	Adjustable

Feature Importance from RF model

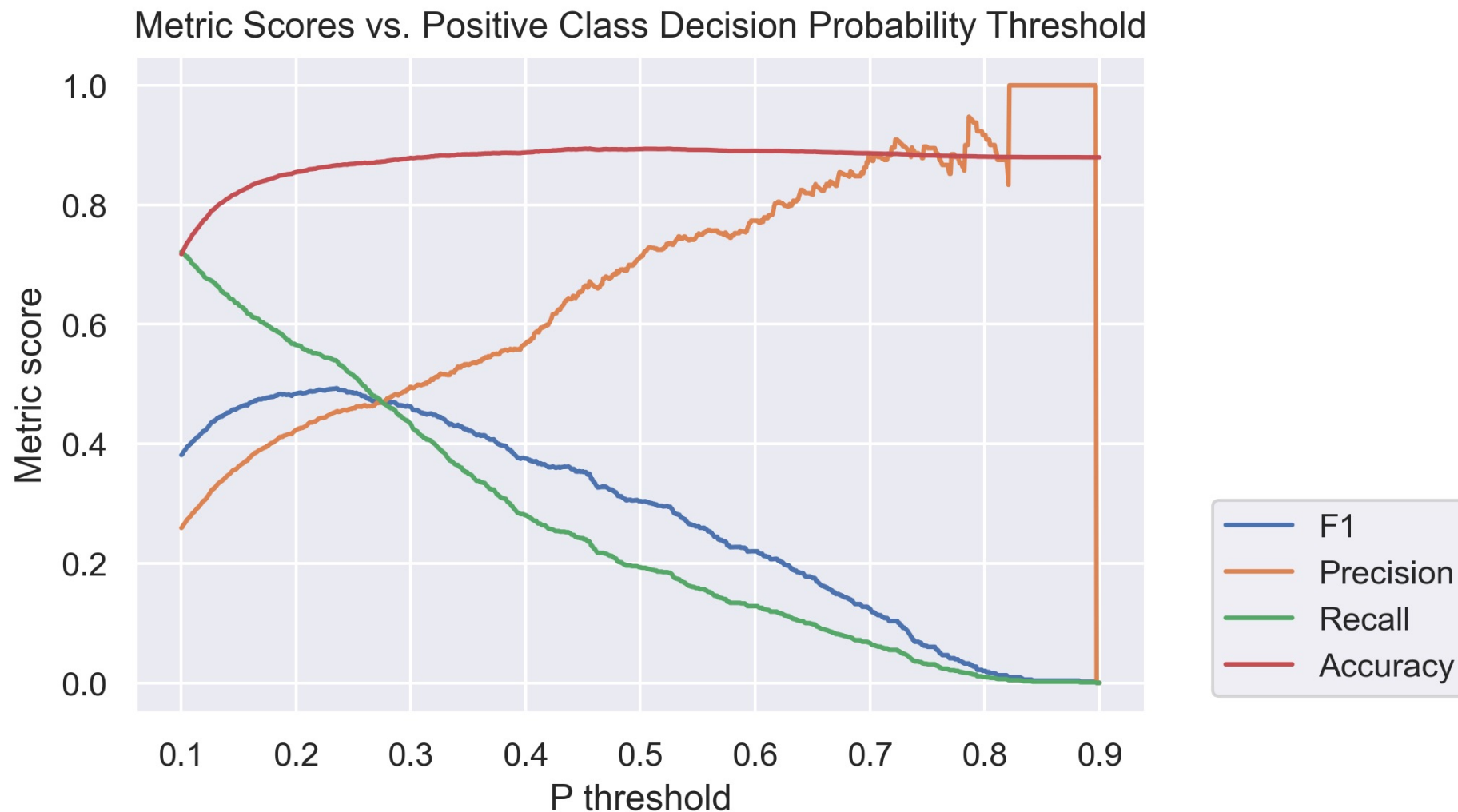


Closer look at important features

Graph idea

- Balance mean for negative and positive class ---- matplotlib bargrah
- Age (further classified to 20s, 30s, 40s--
- Job ---sns.scountplot
- Campaign(number of contacts during the campaign) ----

Solution: RF model w/ Probability Threshold Controlling



Solution: RF model w/ Probability Threshold Controlling

