

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 fan of Euro banknotes. In the bottom right corner, a silver compass is shown. A black pen lies diagonally across the bottom left of the slide, resting on a table with numerical data.

Bank Telemarketing Success Classification problem Rui Yuan

125,058	154,568	95,054	124,500
125,487	56,845	97,511	125,000
124,000	110,000	99,011	154,000
1450	150,000	99,216	95,000
	35,000	101,090	154,200
		101,684	110,000
		101,962	89,000
			50,000
			10,700

Clients and Problem

- Bank wants to implement a telemarketing campaign.
- And they want to know about the campaign performance: success rate; what their target clients are (that are likely to subscribe bank term deposit or other financial products).

Data

- The data is download from UCI ML Repository, related with direct marketing campaigns of a Portuguese banking institution dated from May 2008 to November 2010.
- The marketing campaigns were based on phone calls.
- 45,211 rows and 16 columns.
- Target: Whether client subscribe bank term deposit (1 = yes, 0 = no)

15 Features

bank clients' data:

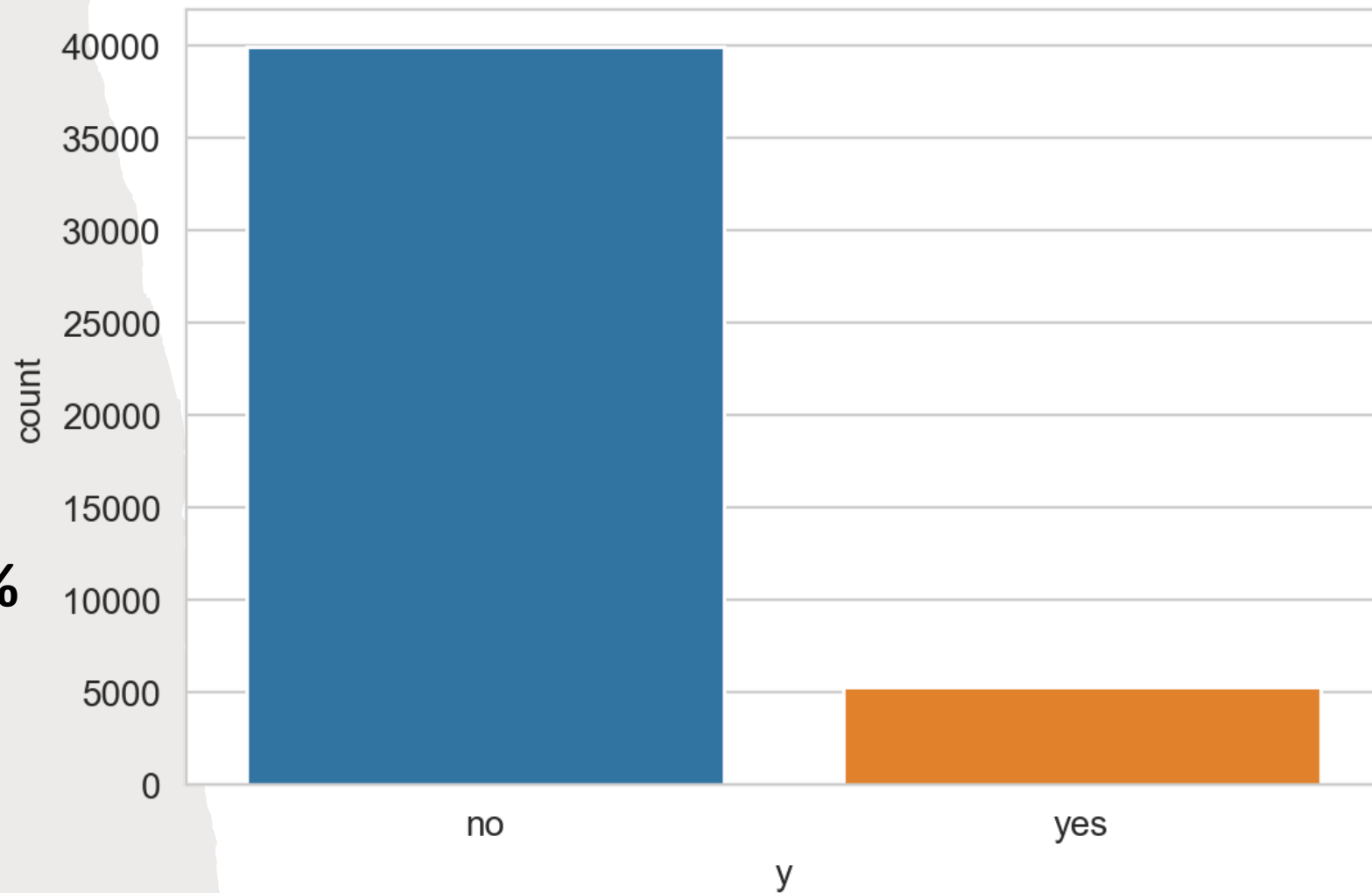
- 1 - **age**: numeric, 18 to 95.
- 2 - **job** : type of job (categorical: "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed", "retired", "technician", "services")
- 3 - **marital status** (categorical: "married", "divorced", "single"; note: "divorced" means divorced or widowed)
- 4 - **education** (categorical: "unknown", "secondary", "primary", "tertiary")
- 5 - **default**: has credit in default? (binary: "yes", "no")
- 6 - **balance**: average yearly balance, in euros (numeric)
- 7 - **housing**: has housing loan? (binary: "yes", "no")
- 8 - **loan**: has personal loan? (binary: "yes", "no")

campaign data and other attributes:


- 9 - **contact**: contact communication type (categorical: "cellular", "telephone", "unknown")
- 10 - **day**: last contact day of the month
- 11 - **month**: last contact month of year (categorical: "jan", "feb", "mar", ..., "nov", "dec")
- 12 - **campaign**: number of contacts performed during this campaign and for this client (numeric, includes last contact)
- 13 - **pdays**: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
- 14 - **previous**: number of contacts performed before this campaign and for this client (numeric)
- 15 - **poutcome**: outcome of the previous marketing campaign (categorical: "failure", "nonexistent", "success")

Class Distribution

Success rate= $\text{yes/all} = 12\%$



Classification Modeling Goal


- Goal:  precision score on positive class & number of true positives (subscribe)
- In Business Sense: Model being able to target clients upon changing needs.
 - Business capability can vary depending on the amount of sources, such as number of employees, phone plan fees, etc.
 - So how many clients bank can reach to or how many phone calls bank can do in a given period (week/month/year) may vary, and we want the model being able to target potential clients upon business changing capability.

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