

Erdős Institute Data Science bootcamp 2023: Predicting subscription to a term deposit.



Team 11: Drew D. Ash, Laura Brade, Jamie Kimble

Framing the problem and data

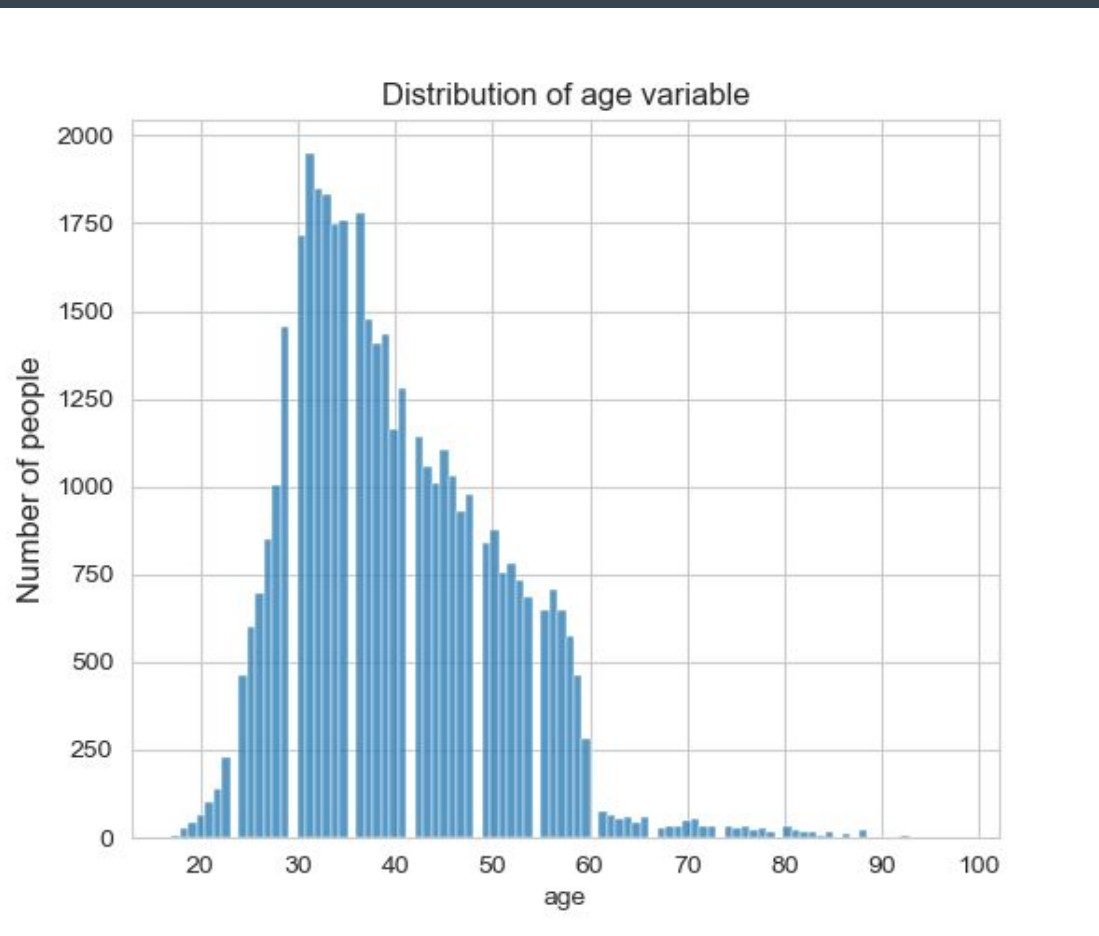
Statement of problem: Given specific data about a person, how likely are they to subscribe to a term deposit with this Portuguese bank?

Key stakeholders: This specific bank is the primary stakeholder. The model, however, could be adapted to different datasets and different questions revolving around any consumer or customer base.

Data: <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

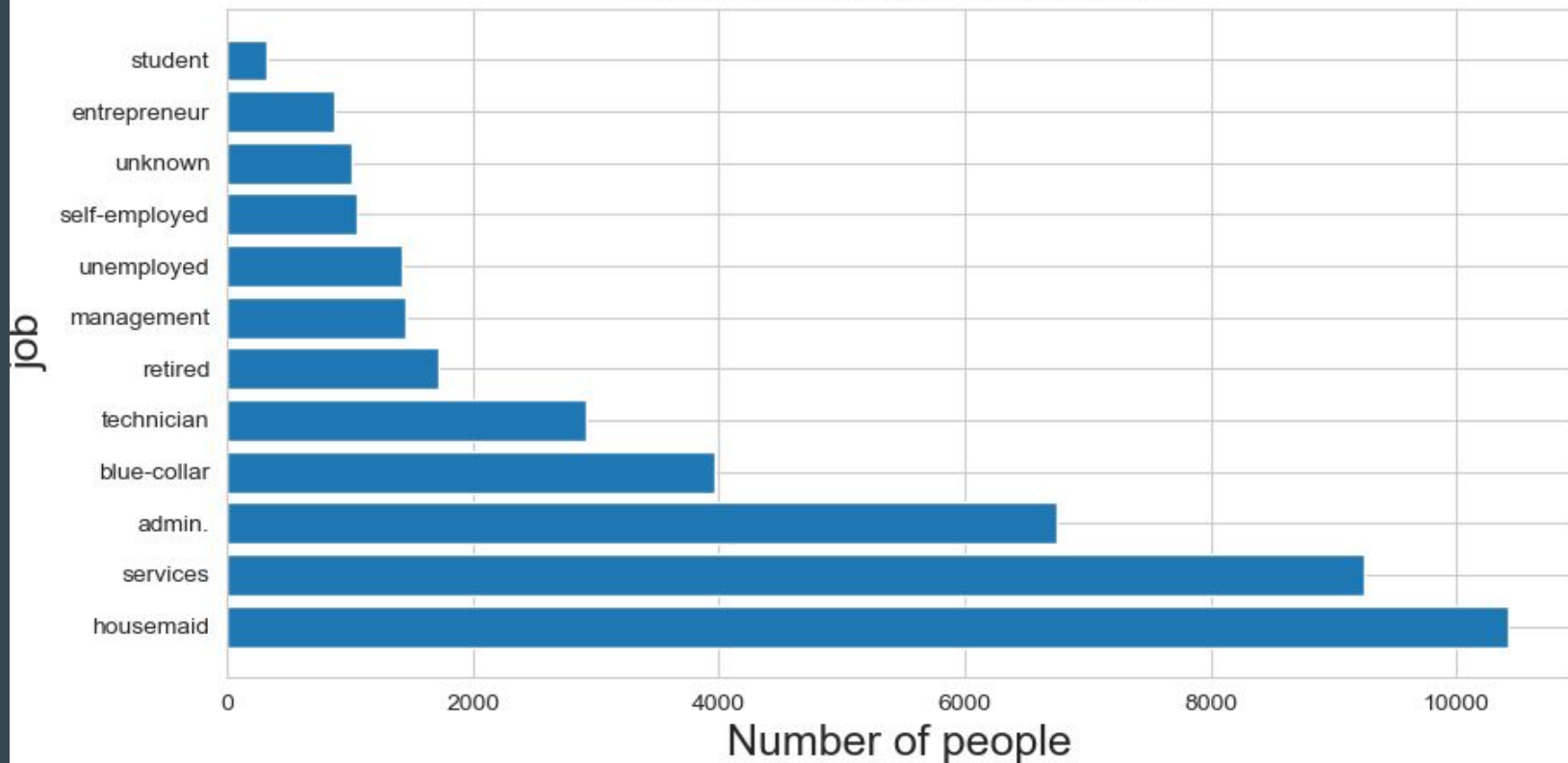
Code and materials: https://github.com/DrewAsh13/Erdos-2023_Project

Exploring the data and preprocessing



EDA continued

Bank Data Jobs Distribution



Model Choices, metrics and outcomes

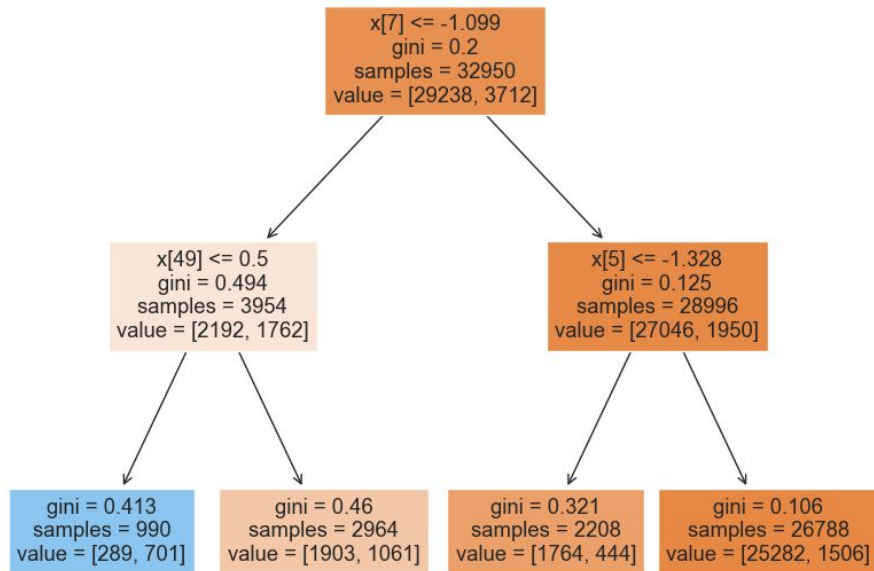
Model	Precision	Recall	True Positive Rate (TPR)	False Negative Rate (FNR)
Logistic Reg.	0.6442	0.2166	0.2166	0.7834
KNN	0.5274	0.2802	0.2802	0.7198
Decision Tree	0.6745	0.1853	0.6745	0.3255
Random Forest	1	0.0011	1	0

Final choice and recommendations

Recommended Model: Decision Tree

- Highest TPR and Lowest FNR
 - Given that roughly 11% of contacted people deposited money, having high TPR and low FNR is essential to this problem.
- Transparency
 - The decision tree model is transparent in that it outputs how it made choices to classify will deposit or will not deposit. We can use these insights to better inform the next direct marketing campaign.

Decision Tree



Future Directions

1. Tuning hyperparameters
 - a. Revisiting the random forest model to correct overfitting.
 - b. Experiment with various other KNN models
 - c. Try different depths on the decision tree
2. Feature selection
3. Try other classification models
 - a. Neural Networks
 - b. SVM

Data: <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

Code and materials: https://github.com/DrewAsh13/Erdos-2023_Project