


Machine Learning Project 4 Portuguese Bank Direct Marketing Campaign Data



Group 6
Clayton Knight, Erik Malrait,
Jacque Meyer, Saudia Epps



Introduction to the Dataset

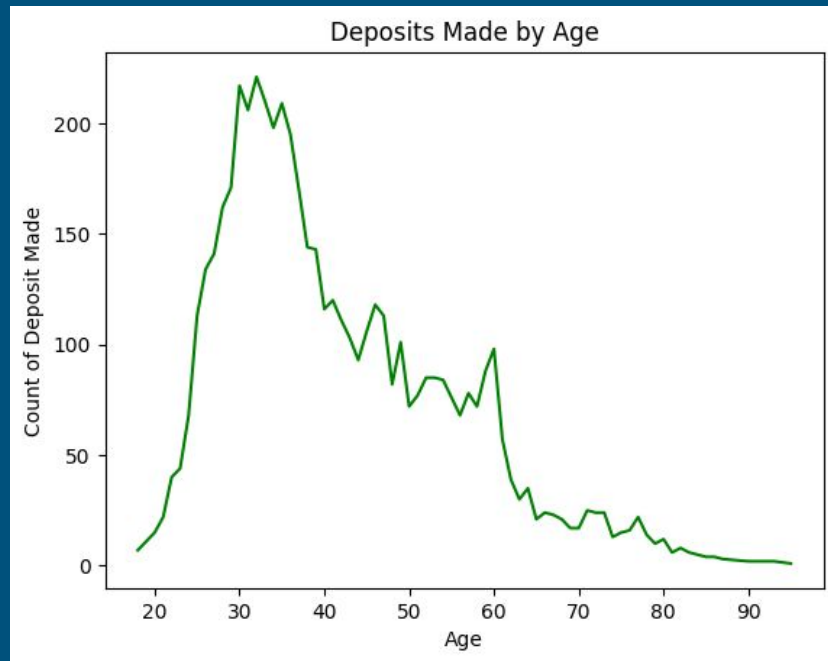
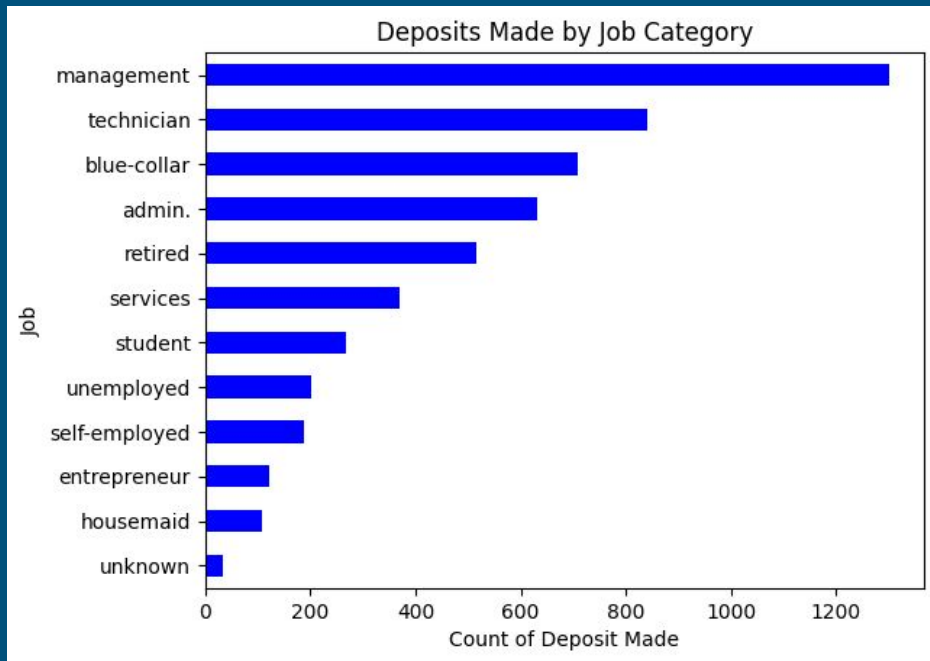
The Bank Marketing Dataset is derived from a direct marketing campaign (via phone calls) conducted by a Portuguese banking institution.

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Business	Classification
Feature Type	# Instances	# Features
Categorical, Integer	45211	16

Objective

- Employ machine learning
 - Supervised Learning
 - Binary Classification
- Forecast an individual's likelihood of making a term deposit
- Factors such as:
 - Age
 - Occupation
 - Education level
 - Marital status
 - Avg Yearly Balance
 - Home Loan

Data Exploration

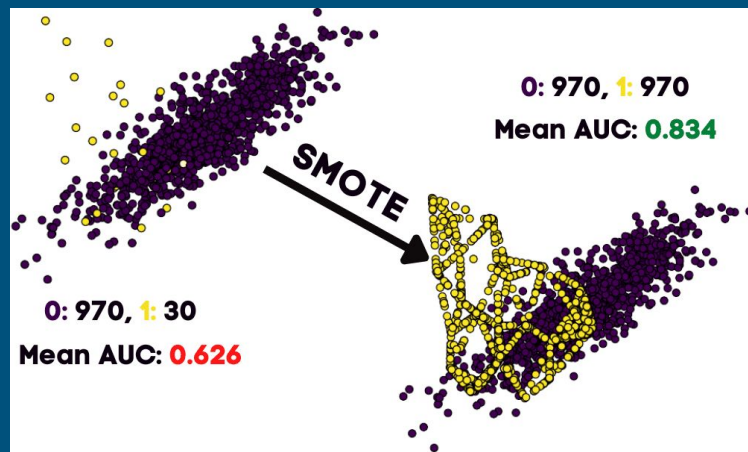


Applied Techniques

- Synthetic Minority Oversampling Technique (SMOTE)
- Keras Tuner
- Early Stopping

Train on 84 samples, validate on 28 samples

```
Epoch 1/200
84/84 - 0s - loss: 0.0181 - accuracy: 1.0000 - val_loss: 0.2464 - val_accuracy: 0.9286
Epoch 2/200
84/84 - 0s - loss: 0.0253 - accuracy: 0.9881 - val_loss: 0.2413 - val_accuracy: 0.9286
Epoch 3/200
84/84 - 0s - loss: 0.0184 - accuracy: 1.0000 - val_loss: 0.2083 - val_accuracy: 0.9286
Epoch 4/200
84/84 - 0s - loss: 0.0146 - accuracy: 1.0000 - val_loss: 0.2026 - val_accuracy: 0.9286
Epoch 5/200
84/84 - 0s - loss: 0.0166 - accuracy: 1.0000 - val_loss: 0.2169 - val_accuracy: 0.9286
Epoch 6/200
84/84 - 0s - loss: 0.0124 - accuracy: 1.0000 - val_loss: 0.2900 - val_accuracy: 0.9286
Epoch 7/200
84/84 - 0s - loss: 0.0337 - accuracy: 0.9881 - val_loss: 0.2882 - val_accuracy: 0.9286
Epoch 8/200
84/84 - 0s - loss: 0.0376 - accuracy: 0.9762 - val_loss: 0.4433 - val_accuracy: 0.8571
Epoch 9/200
84/84 - 0s - loss: 0.1473 - accuracy: 0.9524 - val_loss: 0.3199 - val_accuracy: 0.8929
```



```
Trial 5 Complete [00h 01m 01s]
val_accuracy: 0.7699999809265137
```

```
Best val_accuracy So Far: 0.9350000023841858
Total elapsed time: 00h 05m 07s
```

Search: Running Trial #6

Hyperparameter	Value	Best Value So Far
conv1	32	96
conv2	64	64
fc	512	768
learning_rate	0.001	0.1
tuner/epochs	12	12
tuner/initial_e...	10	10
tuner/bracket	13	13
tuner/round	10	10

```
Epoch 1/2
25/25 [=====] - 29s 1s/step - loss: 1.0732 - accuracy: 0.7027 - val_loss: 0.5772 - val_accuracy: 0.7600
Epoch 2/2
6/25 [=====>.....] - ETA: 16s - loss: 0.3675 - accuracy: 0.9095
```

Logistic Regression Results

	Predicted No	Predicted Yes
Actual No	9718	263
Actual Yes	863	459

	precision	recall	f1-score	support
0	0.92	0.97	0.95	9981
1	0.64	0.35	0.45	1322
accuracy			0.90	11303
macro avg	0.78	0.66	0.70	11303
weighted avg	0.89	0.90	0.89	11303

Random Forest Results

	Predicted No	Predicted Yes
Actual No	9699	282
Actual Yes	804	518

	precision	recall	f1-score	support
0	0.92	0.97	0.95	9981
1	0.65	0.39	0.49	1322
accuracy			0.90	11303
macro avg	0.79	0.68	0.72	11303
weighted avg	0.89	0.90	0.89	11303

Important Features

duration	0.266084
balance	0.095127
age	0.092993
day	0.082115
poutcome_success	0.045386

* We realized AFTER that this is for multi-class classification

[illegible]

ML Test		Precision	Recall	F1 Score
Logistic Regression	0	92%	97%	95%
	1	64%	35%	45%
Random Forest	0	92%	97%	95%
	1	65%	39%	49%

Optimized Results

```
➡ Trial 60 Complete [00h 00m 52s]  
val_accuracy: 0.9043616652488708  
  
Best val_accuracy So Far: 0.9050694704055786  
Total elapsed time: 00h 28m 37s
```

```
➡ 354/354 - 1s - loss: 0.2168 - accuracy: 0.9051 - 1s/epoch - 4ms/step  
Loss: 0.21683882176876068, Accuracy: 0.9050694704055786
```

References

Data source: <https://archive.ics.uci.edu/dataset/222/bank+marketing>

Early Stop callback:

<https://www.geeksforgeeks.org/choose-optimal-number-of-epochs-to-train-a-neural-network-in-keras/>

SMOTE documentation:

https://imbalanced-learn.org/stable/references/generated/imblearn.over_sampling.SMOTE.html