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.

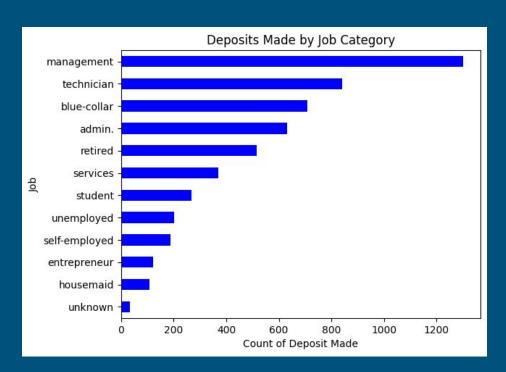
Dataset Characteristics	Subject Area	Associated Tasks		
Multivariate	Business	Classification		
Feature Type	# Instances	# Features		
Categorical, Integer	45211	16		

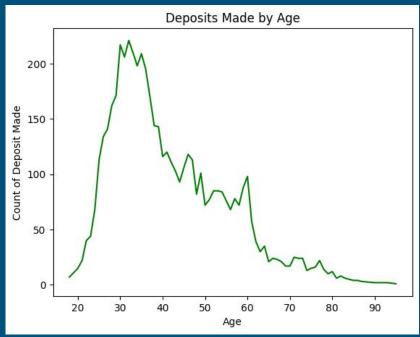
Objective

- 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

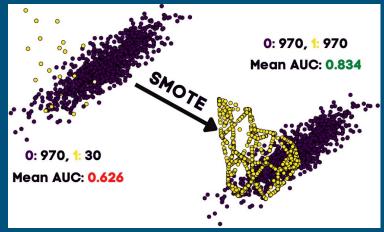




Applied Techniques

- 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
                 132
conv2
                 164
                                   1768
learning rate
                 10.001
                                   10.1
tuner/epochs
tuner/initial_e...|0
tuner/bracket
tuner/round
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 1	0.92 0.64	0.97 0.35	0.95 0.45	9981 1322
accuracy macro avg weighted avg	0.78 0.89	0.66 0.90	0.90 0.70 0.89	11303 11303 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

Model Iterations

10.00	Testi	ng	Train	ing	Total	Laye	r 1:	Lay	er 2:	Lay	er 3:	Laye	er 4:	Laye	r 5:	5: Layer 6:		CE CESTIONE		100
Model Iteration	Accuracy	Loss	Accuracy	Loss	Layers	Neurons	Func	Neurons	Func	Neurons	Func	Neurons	Func	Neurons	Func	Neurons	Func	Epochs	SMOTE	Early Stop
NN #1	90.3%	21.3%	91.0%	19.4%	2	6	relu	1	sigmoid									100		
NN #2	90.1%	22.7%	92.5%	17.2%	2	30	relu	1	sigmoid									100		
NN #3 w/SMOTE	86.6%	34.4%	91.9%	21.2%	2	30	relu	1	sigmoid									100	1.0	
NN #4 w/SMOTE	88.0%	29.0%	90.9%	22.4%	2	30	relu	1	sigmoid									100	0.5	
NN #5 w/SMOTE	87.3%	38.6%	93.8%	15.4%	3	30	relu	30	tanh	1	sigmoid							100	0.5	
NN #6 w/SMOTE	11.0%	15.0%	33.0%	39.0%	3	30	relu	30	tanh	1	softmax*							100	0.5	
NN #7 w/SMOTE	87.0%	34.8%	91.0%	21.1%	3	30	relu	30	leaky_relu	1	sigmoid							100	0.5	
NN #8 w/SMOTE	88.2%	34.0%	92.7%	17.9%	3	30	relu	30	leaky_relu	1	sigmoid							100	0.5	7
NN #9 w/SMOTE	88.0%	26.1%	90.1%	23.5%	3	30	relu	30	relu	1	sigmoid							100	0.5	5
NN #10 w/SMOTE	87.7%	25.8%	91.8%	19.8%	4	30	relu	30	relu	30	tanh	1	sigmoid					100	0.5	6
NN #11 w/SMOTE	88.1%	26.0%	87.5%	27.3%	4	30	relu	30	relu	30	relu	1	sigmoid					100	0.5	4
NN #12 w/SMOTE	86.9%	26.5%	88.1%	27.8%	4	144	relu	96	relu	48	tanh	1	sigmoid					100	0.5	1
															* We reali	zed AFTER t	hat this is	for multi-cl	ass classi	fication
After Running Keras Tuner:	c 0.01 to 10 to	101611	46		-	65		50								S		101 1 20		

Arter Running Keras Tuner:																			
KT-Best fit (no synthetic data)	90.5%	21.7%		6	3	tanh	1	tanh	7	tanh	5	tanh	1	tanh	1	sigmoid	20		
KT-Best fit w/SMOTE	89.5%	37.2%	1	3	5	sigmoid	9	sigmoid	1	sigmoid							3	0.5	

ML Test		Precision	Recall	F1 Score
Logistic Regression	0	92%	97%	95%
	1	64%	35%	45%
Random Forest	0	92%	97%	95%
personal to transport your state of	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