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

Deposits Made by Job Category

Deposits Made by Age
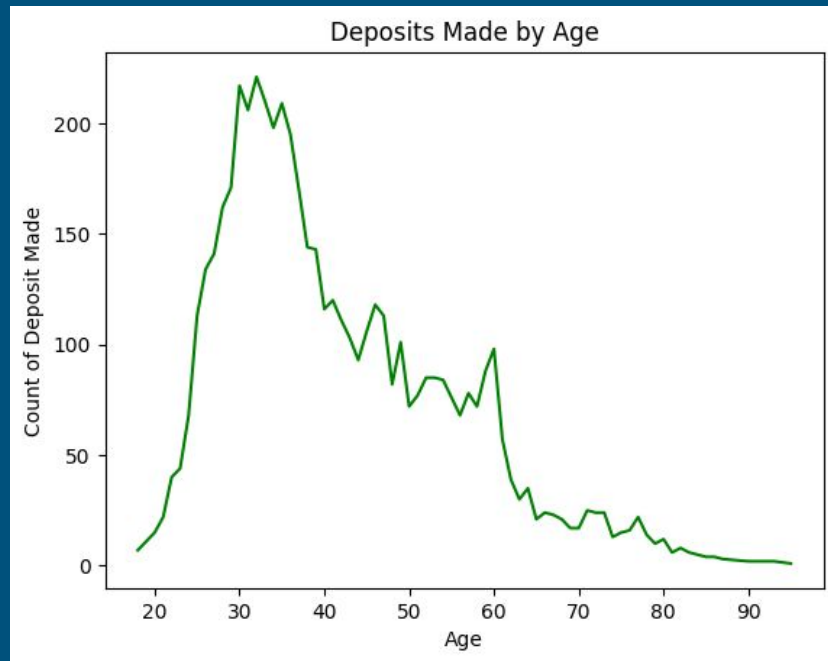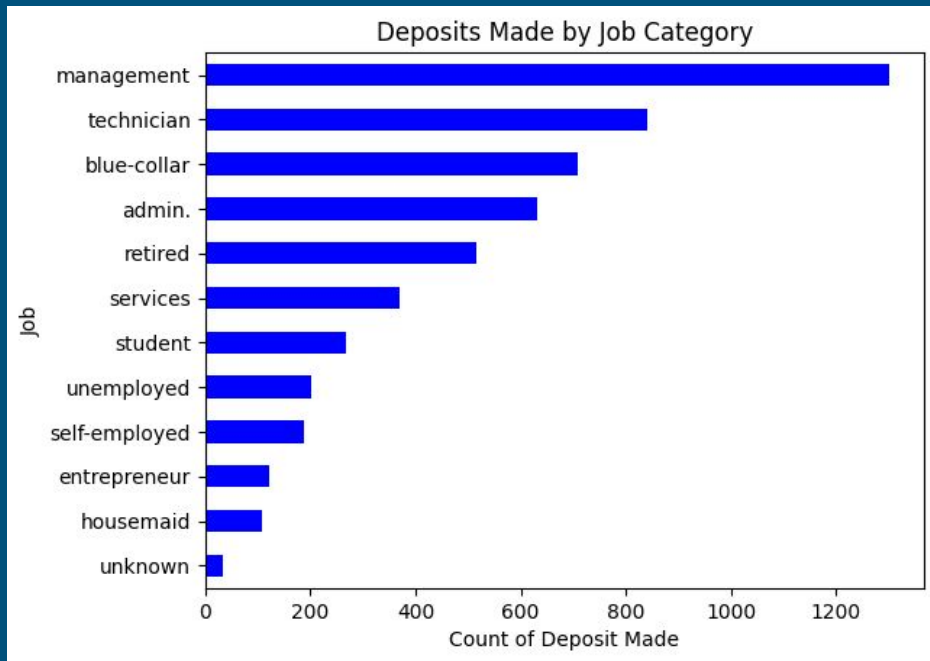
Applied Techniques

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

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
```

Model Iterations

Model Iteration	Testing Accuracy	Loss	Training Accuracy	Loss	Total Layers	Layer 1: Neurons	Func	Layer 2: Neurons	Func	Layer 3: Neurons	Func	Layer 4: Neurons	Func	Layer 5: Neurons	Func	Layer 6: 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 realized AFTER that this is for multi-class classification				
After 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%			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%
	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