

# **EXPLORING BANK DATA**

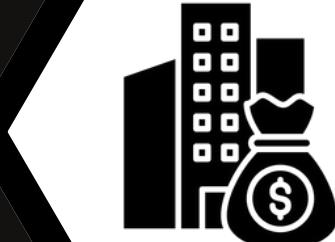
**TO PREDICT AND IDENTIFY  
POTENTIAL CUSTOMERS**

**OPERATIONS ANALYTICS**

**GROUP 4: RUBEN CUESTA, KRISHAN KANT, SARAH RAUBENHEIMER,  
JOSEPHINE LEITZ, LUIS GONZALEZ**

# CURRENT CHALLENGES

High marketing expenditures  
with moderate conversion rate



Increased risk exposure due  
to wrong target audience



Marketing mismatch  
harming brand reputation



# ADDRESSING THE CHALLENGES WITH DATA

01

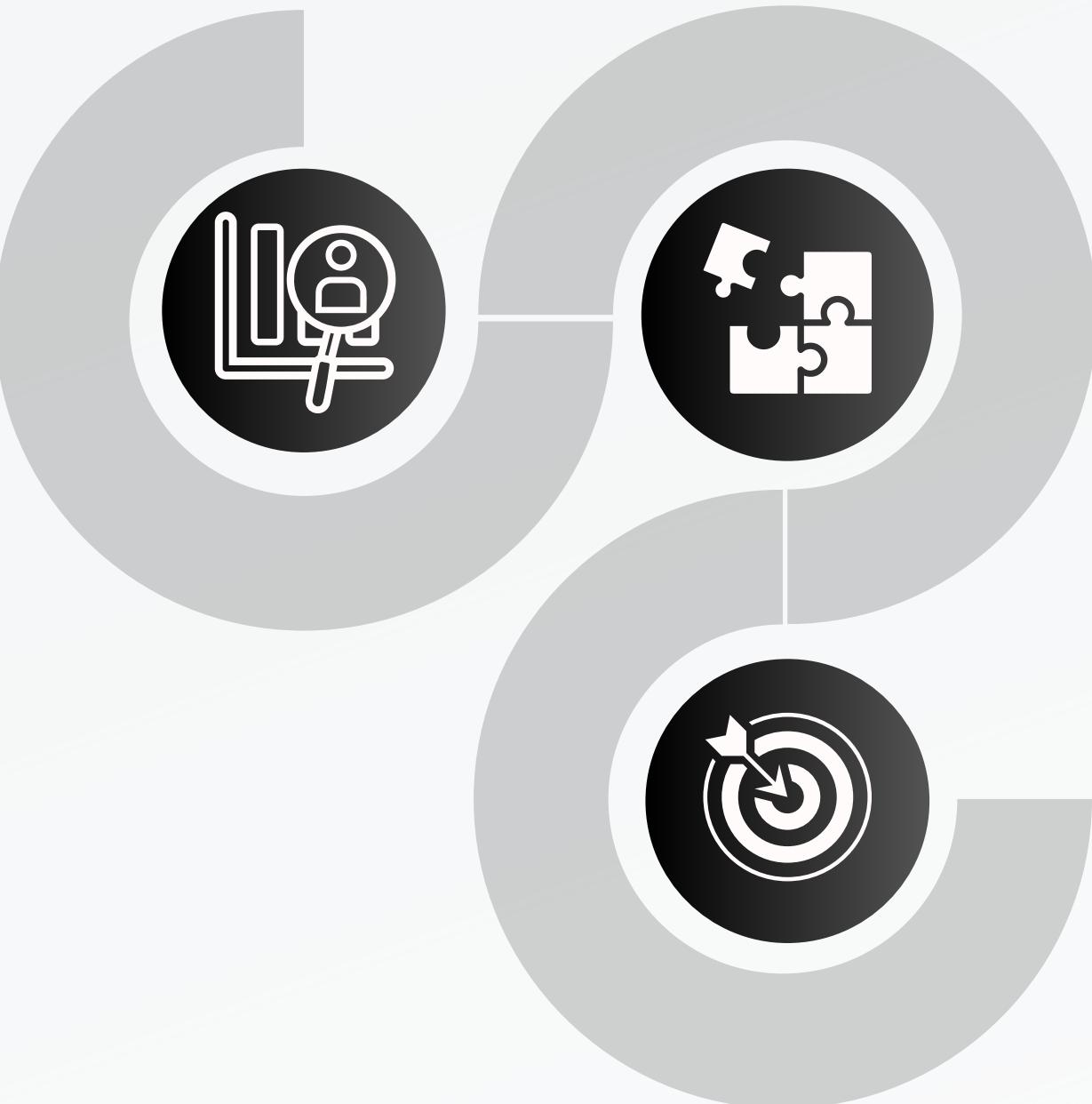
Preparing and exploring an internal bank marketing dataset

02

Creating models that predict if a customer will subscribe to a loan (Logistic Regression, Neural Nets & Support Vector Machine)

03

Fine-tuning models and deriving business insights from predictions (e.g., feature importance and accuracy)



# MODELS AND FINDINGS

Model	Balanced Accuracy	Insights
1. Logistic Regression (all variables*)	77.73%	Most significant variables (campaign timing, customer persona)
<b>2. Logistic Regression (only significant variables)</b>	<b>77.34%</b>	<b>Business insights (best model to predict a customer**)</b>
3. Support Vector Machine	75.97%	Increasingly complex model did not yield better results
4. Neural Network	74.49%	Black Box model + high computational power needed (not explainable nor scalable)

\* Models including all variables are prone to overfitting

\*\* Reducing the amount of variables will reduce computing power (increasing efficiency) and less data gathered reduces costs

# WHAT ARE THE MOST IMPORTANT PREDICTORS?

MODEL: LOGISTIC REGRESSION WITH ACCURACY OF 77.34% & IMPORTANT VARIABLES

## MARKETING STRATEGY

Predictor	Effect on customer purchasing decision
# of Contact Points	Negative
Contact Call Duration	Positive

## CAMPAIGN TIMING

Predictor	Effect on customer purchasing decision
Day of the Week	Positive
Month	Positive

## CUSTOMER-CENTRIC

Predictor	Effect on customer purchasing decision
Job: retired	Positive
Have a personal loan	Negative
Previously successfully targeted	Positive

# TRANSLATING FINDINGS INTO BUSINESS RECOMMENDATIONS

## Who?

Customers that are more likely to subscribe to a term deposit:

- Retired customer
- Does not have personal loan
- Prev. campaign sensitive

**Customized Marketing for Target Customer Traits**

## When?

Time of the year when most customers will subscribe to a term deposit:

- June
- October
- End of week

**Implement seasonal & weekly campaigns**

## How?

Marketing strategy to ensure customers subscribe to the new loan:

- Longer call duration
- Less # of contacts

**Call customers less but keep them engaged on the phone**



# APPENDIX STRUCTURE

- Stratified Sampling
- Model findings - Detailed
- Logistic Regression
- SVM
- Neural Network
- Link to GitHub Code

# APPENDIX: STRATIFIED SAMPLING

Bar Chart: Distribution of Target Variable (Term Deposit Subscription)

More observations for no term deposit: Data is Skewed

Solution: Stratified Sampling

80/20 Training and Testing set



# APPENDIX: MODEL FINDINGS - DETAILED

Model	Training Accuracy	Testing Accuracy	Balanced Accuracy	Precision	Recall
1. Logistic Regression (all variables)*	90.16%	90.28%	77.73%	39.05%	63.08%
<b>2. Logistic Regression (only significant variables)</b>	<b>90.35%</b>	<b>90.06%</b>	<b>77.34%</b>	<b>35.24%</b>	<b>62.71%</b>
3. Support Vector Machine	89.63%	89.83%	75.97%	38.10%	59.70%
4. Neural Network	90.17&	89.50%	74.49%	42.86%	56.25%

\* Model including all variables is prone to overfitting

# APPENDIX: LOGISTIC REGRESSION

Logistic Regression 1:

- All features
- Accuracy: 90.16%

		Actual	
		No	Yes
Predicted	No	776	64
	Yes	24	41

Logistic Regression 2:

- 16 features
- Accuracy: 90.35%

		Actual	
		No	Yes
Predicted	No	778	68
	Yes	22	37

# APPENDIX: LOGISTIC REGRESSION

## Model 1 Summary Statistics:

- Logistic Regression 1
- Coefficient Estimates
- Significance of Variables

	Coefficients:	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.875701	0.090724	-31.697	< 2e-16	***
age	-0.052003	0.086065	-0.604	0.54569	
`jobblue-collar`	-0.109582	0.111735	-0.981	0.32672	
jobentrepreneur	-0.025878	0.078872	-0.328	0.74284	
jobhousemaid	-0.002923	0.072675	-0.040	0.96791	
jobmanagement	-0.010042	0.111722	-0.090	0.92838	
jobretired	0.165856	0.078143	2.122	0.03380	*
`jobsself-employed`	-0.057123	0.081021	-0.705	0.48078	
jobservices	-0.022099	0.090043	-0.245	0.80612	
jobstudent	0.013687	0.059391	0.230	0.81774	
jobtechnician	-0.086708	0.099989	-0.867	0.38585	
jobunemployed	-0.097658	0.078424	-1.245	0.21304	
jobunknowm	0.066276	0.061823	1.072	0.28371	
maritalmarried	-0.186520	0.096957	-1.924	0.05439	.
maritalsingle	-0.120252	0.103580	-1.161	0.24566	
educationsecondary	0.068479	0.114377	0.599	0.54936	
educationtertiary	0.204434	0.120929	1.691	0.09093	.
educationunknown	-0.095494	0.082934	-1.151	0.24955	
defaultyes	0.088972	0.060626	1.468	0.14222	
balance	-0.018846	0.057684	-0.327	0.74389	
housingyes	-0.118694	0.077602	-1.530	0.12613	
loanyes	-0.177463	0.079857	-2.222	0.02627	*
contacttelephone	-0.066785	0.067687	-0.987	0.32380	
contactunknown	-0.726836	0.116498	-6.239	4.40e-10	***
day	0.167743	0.076547	2.191	0.02843	*
monthaug	-0.079916	0.099508	-0.803	0.42191	
monthdec	0.047732	0.047953	0.995	0.31954	
monthfeb	0.100522	0.071765	1.401	0.16130	
monthjan	-0.242029	0.082420	-2.937	0.00332	**
monthjul	-0.270879	0.104948	-2.581	0.00985	**
monthjun	0.286438	0.107121	2.674	0.00750	**
monthmar	0.199419	0.044220	4.510	6.49e-06	***
monthmay	-0.129843	0.122848	-1.057	0.29054	
monthnov	-0.207752	0.088102	-2.358	0.01837	*
monthoct	0.251645	0.049767	5.056	4.27e-07	***
monthsep	0.108150	0.049972	2.164	0.03045	*
duration	1.113656	0.058945	18.893	< 2e-16	***
campaign	-0.208277	0.098602	-2.112	0.03466	*
pdays	0.066676	0.110390	0.604	0.54584	
previous	-0.030853	0.073844	-0.418	0.67608	
poutcomeother	0.131694	0.062163	2.119	0.03413	*
poutcomesuccess	0.464793	0.053381	8.707	< 2e-16	***
poutcomeunknown	0.059630	0.140183	0.425	0.67057	
---					
Signif. codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2581.3 on 3615 degrees of freedom  
 Residual deviance: 1703.8 on 3573 degrees of freedom  
 AIC: 1789.8

Number of Fisher Scoring iterations: 6

# APPENDIX: LOGISTIC REGRESSION

Most Accurate Model

Summary Statistics:

- Logistic Regression 2
- Coefficient Estimates
- Significance of Variables

	Coefficients:	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-2.82173	0.08780	-32.139	< 2e-16 ***	
contact_unknown	-0.78405	0.11385	-6.887	5.71e-12 ***	
day	0.17736	0.07411	2.393	0.016693 *	
month_feb	0.29511	0.06292	4.690	2.73e-06 ***	
month_jun	0.57881	0.09523	6.078	1.22e-09 ***	
month_mar	0.29189	0.04050	7.207	5.72e-13 ***	
month_oct	0.36998	0.04348	8.510	< 2e-16 ***	
month_sep	0.20995	0.04648	4.517	6.28e-06 ***	
month_may	0.25165	0.10268	2.451	0.014254 *	
month_aug	0.25843	0.07499	3.446	0.000568 ***	
month_apr	0.19901	0.06226	3.197	0.001390 **	
month_dec	0.11298	0.04485	2.519	0.011773 *	
duration	1.09548	0.05746	19.064	< 2e-16 ***	
loanyes	-0.15372	0.07705	-1.995	0.046049 *	
campaign	-0.20454	0.09496	-2.154	0.031249 *	
job_retired	0.13956	0.05161	2.704	0.006848 **	
poutcome_other	-0.41145	0.06774	-6.074	1.25e-09 ***	
poutcome_failure	-0.82035	0.09290	-8.830	< 2e-16 ***	
poutcome_unknown	-0.97977	0.09592	-10.215	< 2e-16 ***	
---					

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2581.3 on 3615 degrees of freedom

Residual deviance: 1735.5 on 3597 degrees of freedom

AIC: 1773.5

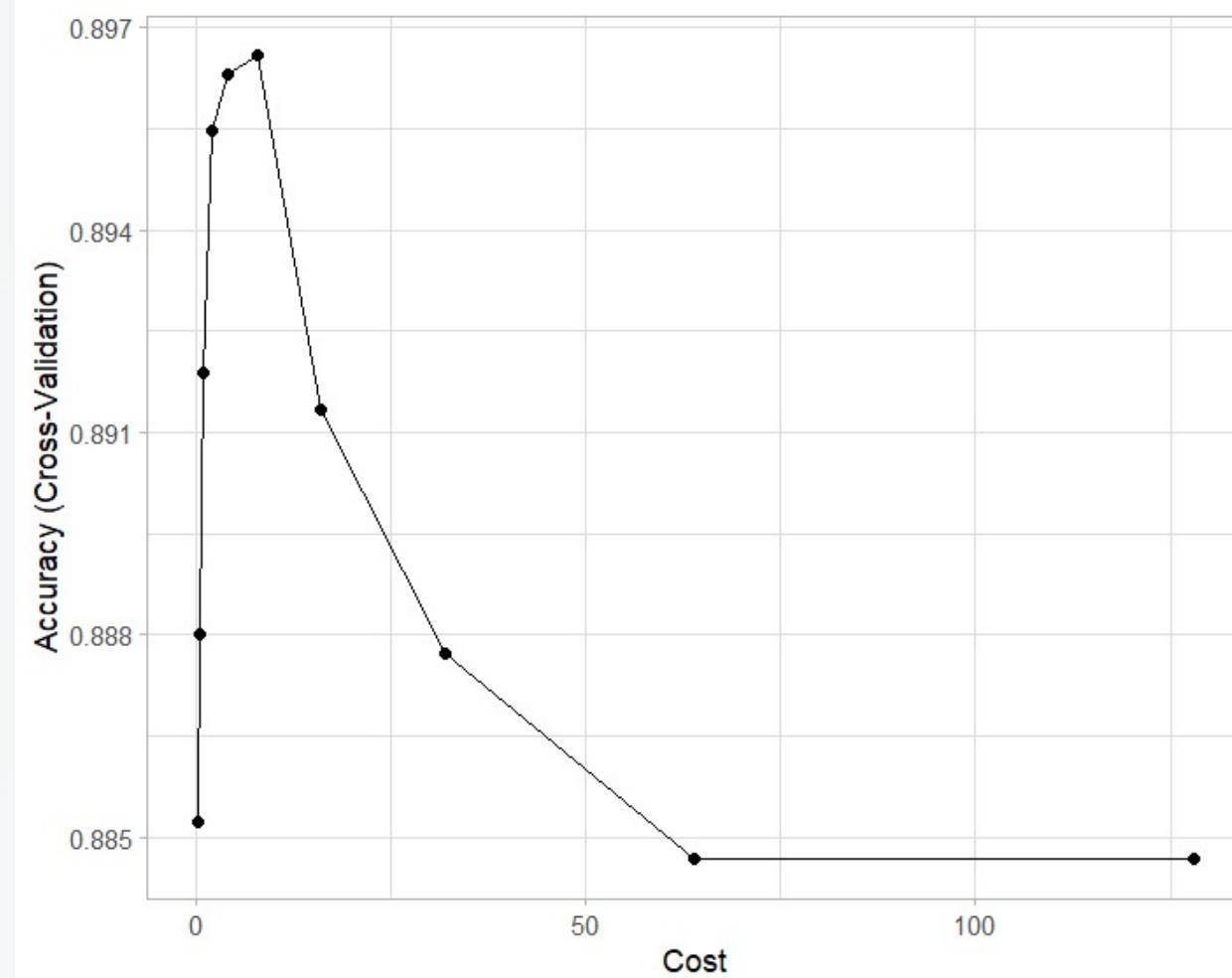
Number of Fisher Scoring iterations: 6

# APPENDIX: SVM

## Model 1

- All features
- Cost = 8
- 16 features
- Accuracy: 89.66%

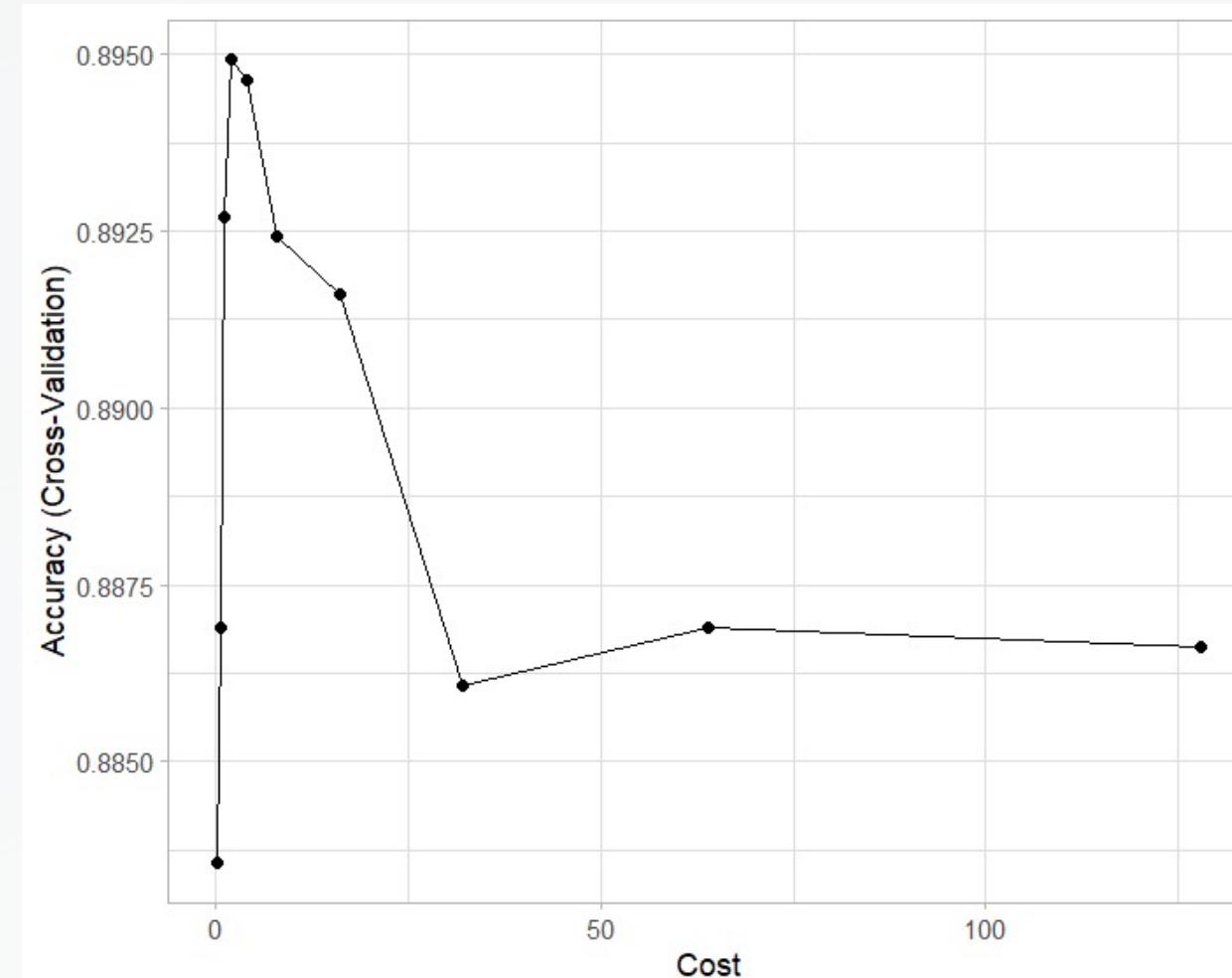
		Actual	
		Predicted	No
		Yes	
No		773	65
Yes		27	40



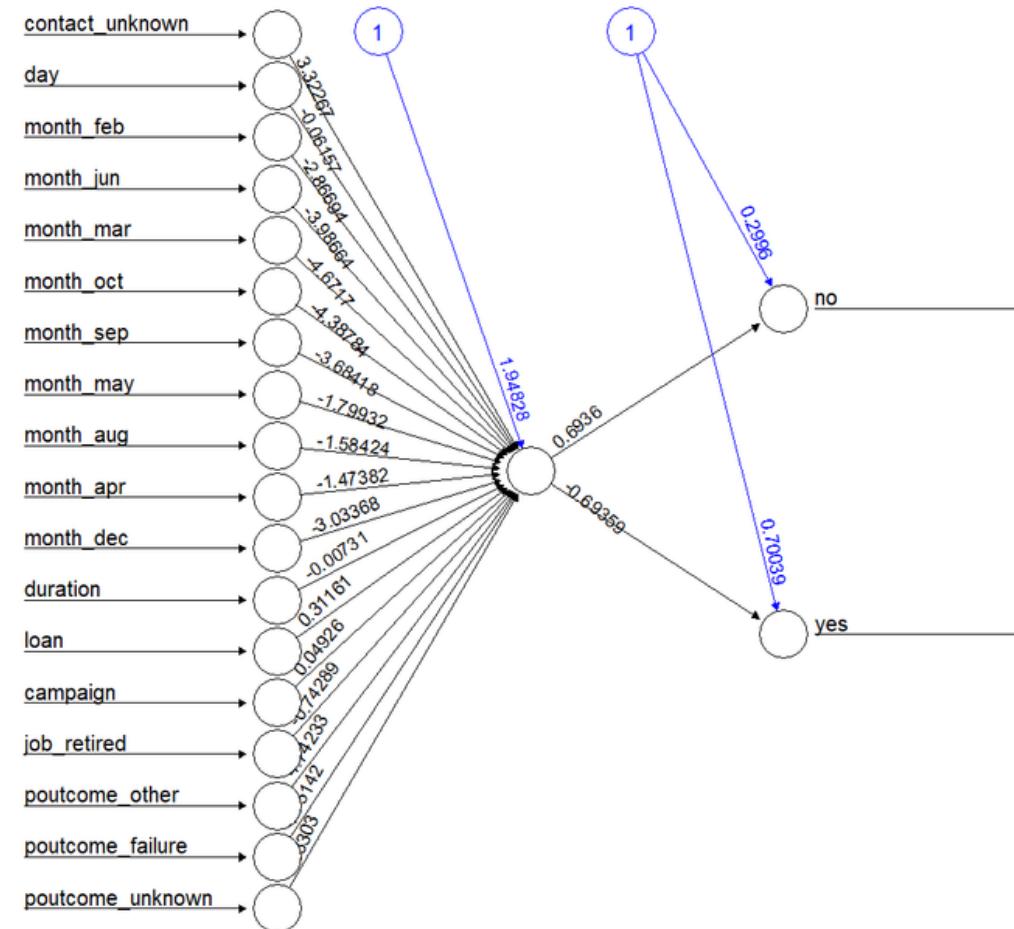
## Model 2

- All features
- Cost = 2
- 16 features
- Accuracy: 89.49%

		Actual	
		Predicted	No
		Yes	
No		776	73
Yes		24	32



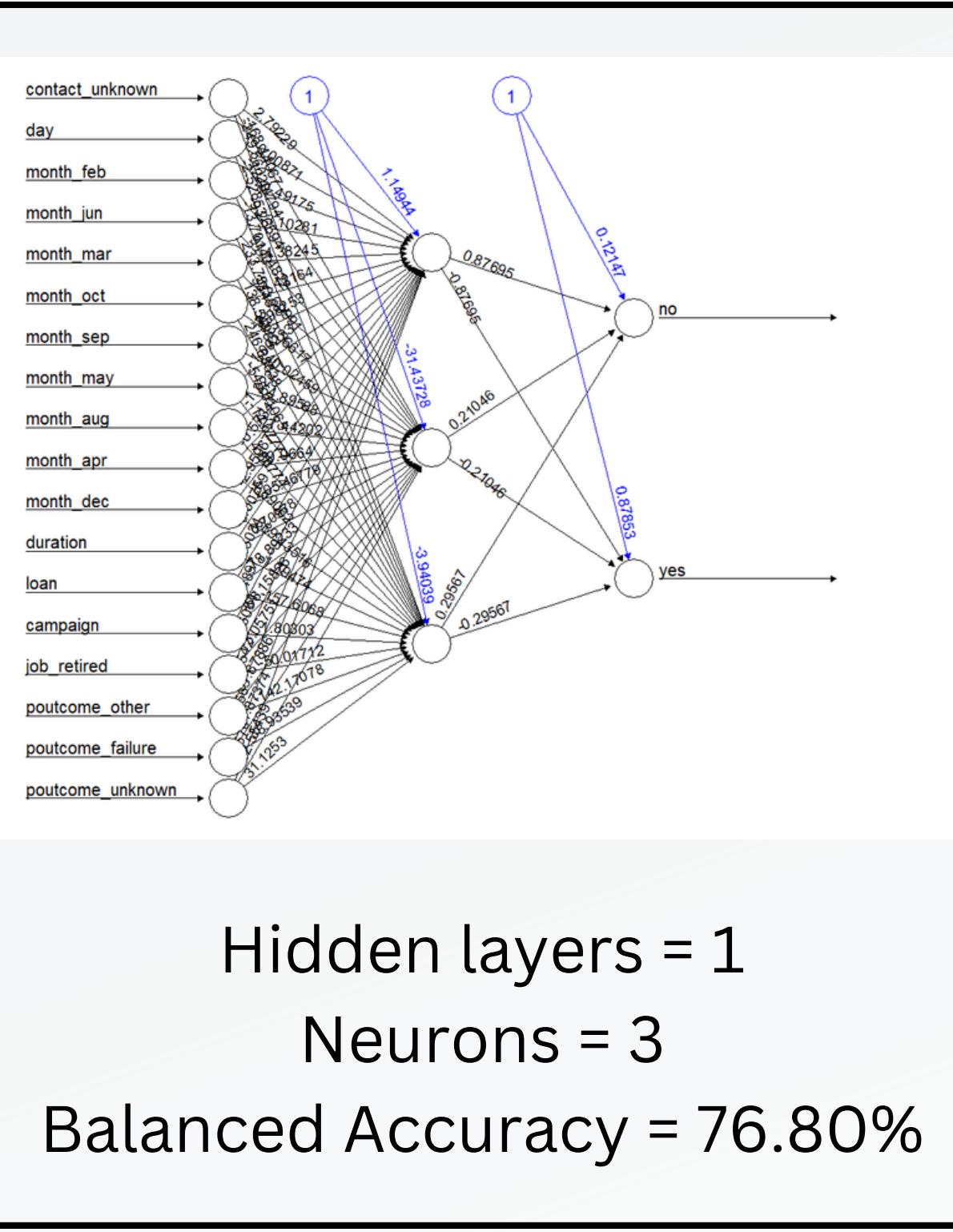
# APPENDIX: NEURAL NETWORK



Hidden layers = 1

Neurons = 1

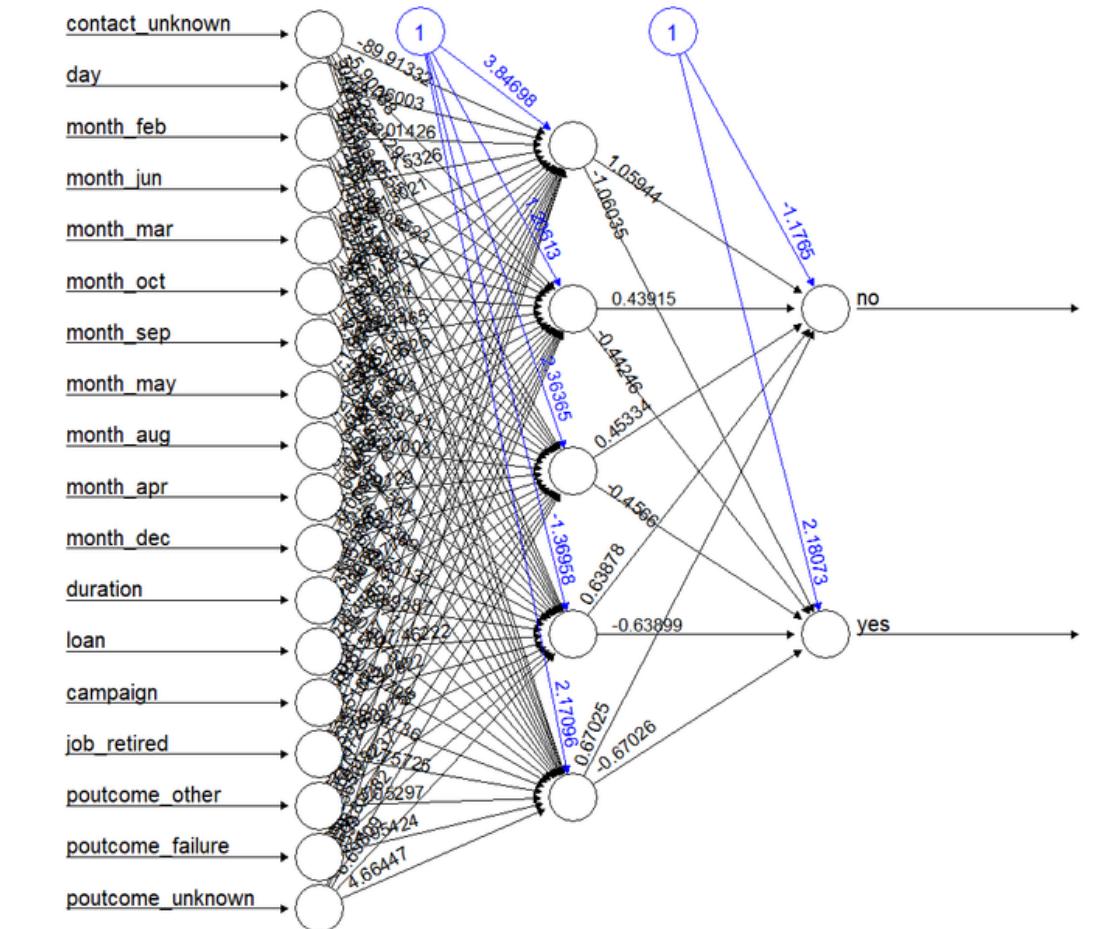
Balanced Accuracy = 75.92%



Hidden layers = 1

Neurons = 3

Balanced Accuracy = 76.80%



Hidden layers = 1

Neurons = 5

Balanced Accuracy = 74.49%

# **APPENDIX: GITHUB LINK**

[https://github.com/KrishanTCD/TCDBA\\_Operations\\_Group\\_4](https://github.com/KrishanTCD/TCDBA_Operations_Group_4)