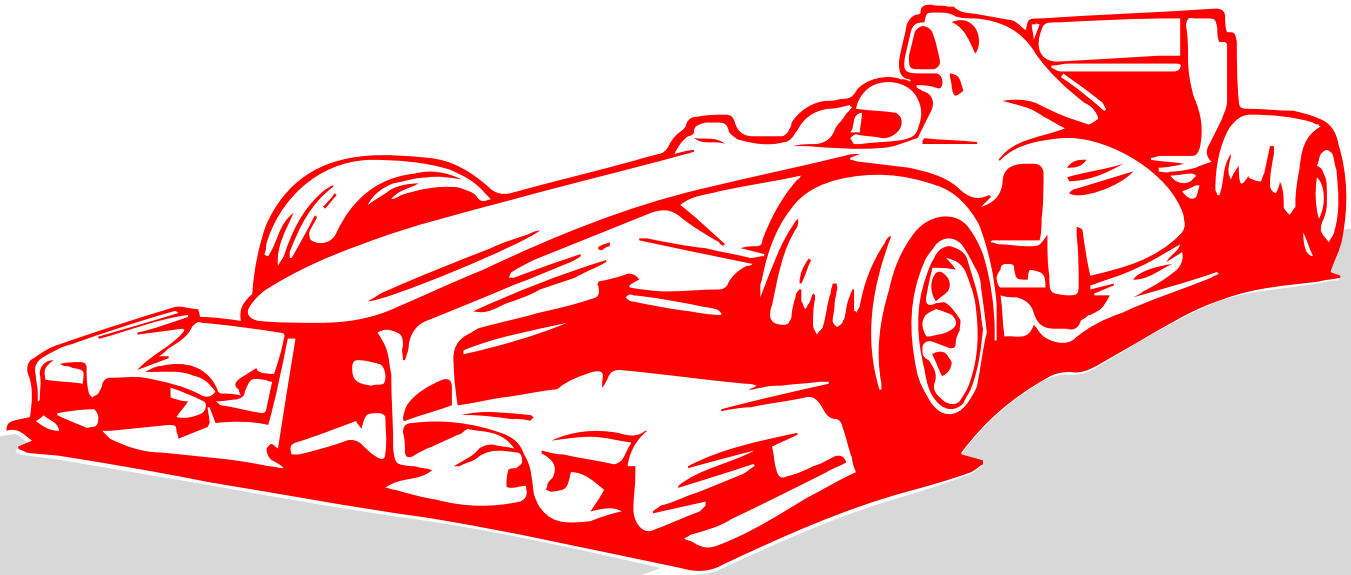
Charles

Betting and Formula 1

Rishyak Panchal and Juan Castano



THE GRID

#AUSGP

STARTING GRID

1ST	VERSTAPPEN
2ND	RUSSELL
3RD	HAMILTON
4TH	ALONSO
5TH	SAINZ
6TH	STROLL
7TH	LECLERC
8TH	ALBON
9TH	GASLY
10TH	HULKENBERG
11TH	OCON
12TH	TSUNODA
13TH	NORRIS
14TH	MAGNUSSEN
15TH	DE VRIES
16TH	PIASTRI
17TH	ZHOU
18TH	SARGEANT
19TH	BOTTAS
20TH	PEREZ

Introduction

Dataset

Predicting results of F1 races

Dataset

Data obtained from Kaggle Formula 1 World Championship (1950-2023)

Data size

dataset had 23,699 rows we potentially could use

Data chosen

We selected only data after 2014(3,707 rows)

Training data

Training data was 80% (2,965 rows)

Testing data

Testing data was 20% (742 rows)

Two models

Multiclass and Binary

Features

01 Grid Position
The starting position of a driver for a race

02 Driver Standing
Driver’s position in the championship, indicative of success

03 Constructor Standing
Constructor’s position in the championship, indicative of success

04 Driver Wins
Total number of wins for the driver, indicative of experience

05 Constructor Wins
Total number of wins for the constructor, indicative of experience

Features

06 Driver Points

Points won by a driver during a championship, indicative of success

07 Constructor Points

Points won by a constructor during a championship, indicative of success

08 Circuit Latitude

North-South position of a circuit, affects weather and temperatures

09 Circuit Longitude

East-West position of a circuit, affects sunlight, times, and tyre behaviour

10 Circuit Altitude

Circuit height, affects oxygen density, aerodynamics, cooling, combustion

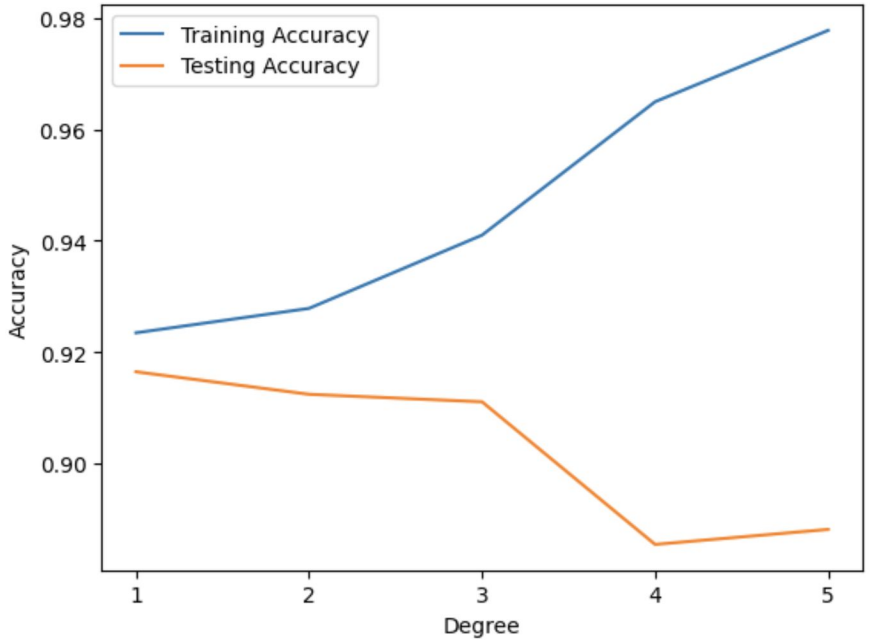
Logistic Regression

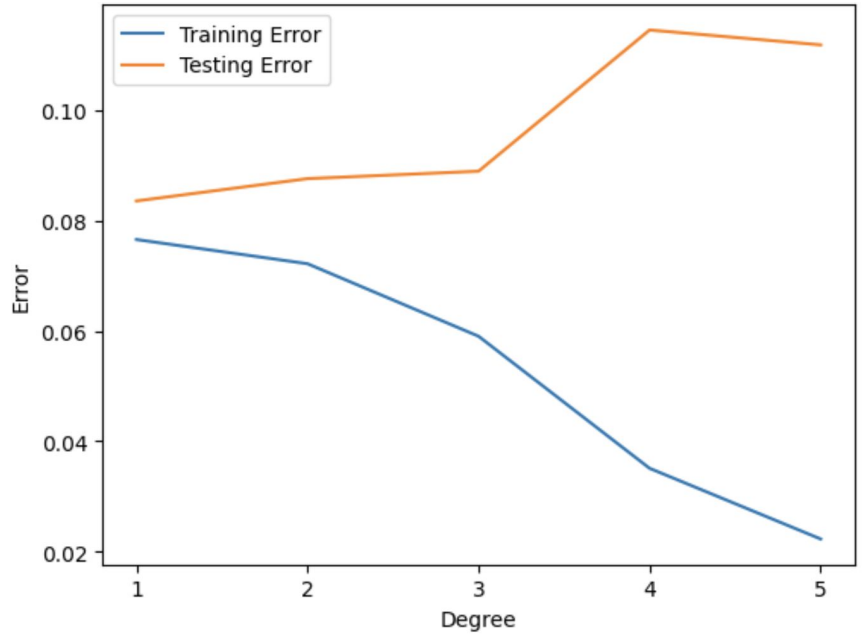
Logistic Regression (Binary)

	Binary Model							
	Lambda=0		Lambda=0.01		Lambda=0.1		Lambda=1	
Degree	Test accuracy	Training accuracy	Test accuracy	Training accuracy	Test accuracy	Training accuracy	Test accuracy	Training accuracy
1	91.51%	92.38%	91.51%	92.38%	91.51%	92.41%	91.64%	92.34%
2	91.51	92.72%	91.51%	92.72%	91.37%	92.75%	91.24%	92.78%
3	90.43%	95.99%	90.43%	95.21%	90.84%	94.50%	91.11%	94.10%
4	87.33%	100%	87.20%	98.92%	88.68%	97.74%	88.54%	96.49%
5	87.33%	100%	87.06%	99.63%	88.01%	98.92%	88.81%	99.77%

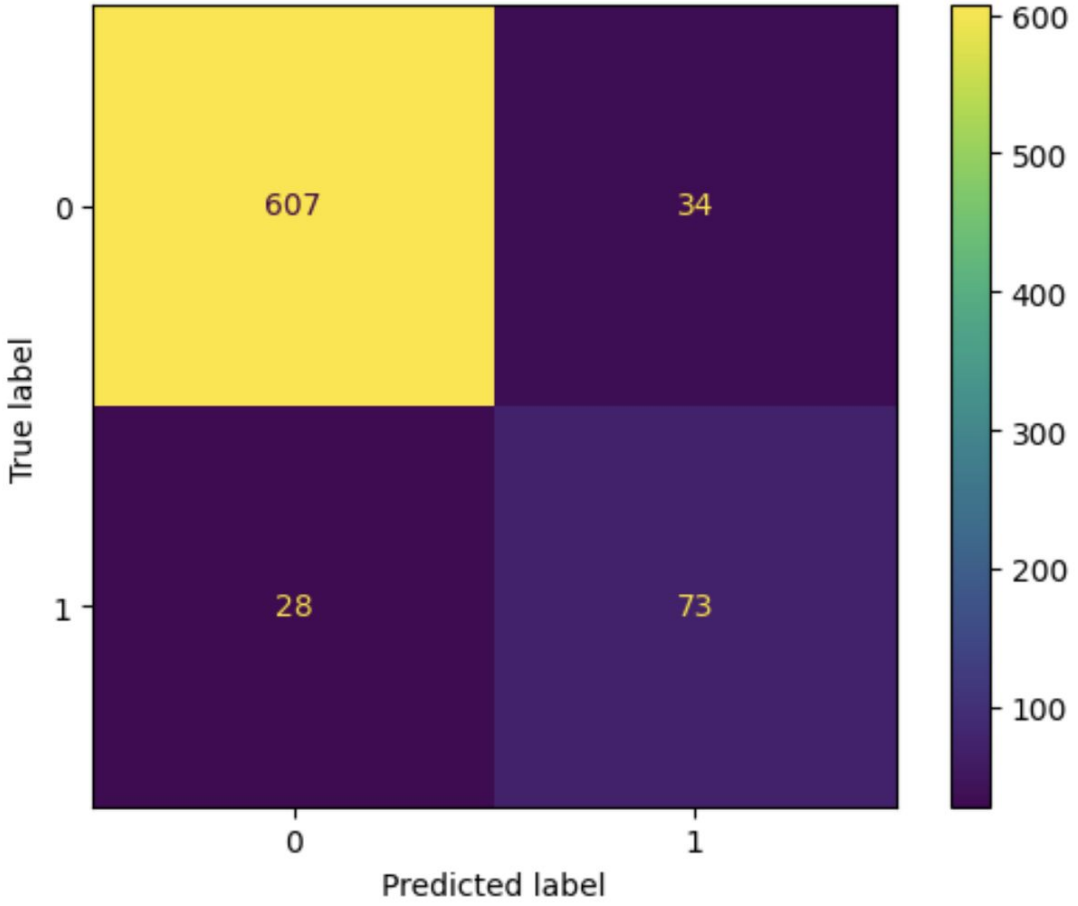
Logistic Regression (Binary)

Lambda = 1





Logistic Regression (Binary)

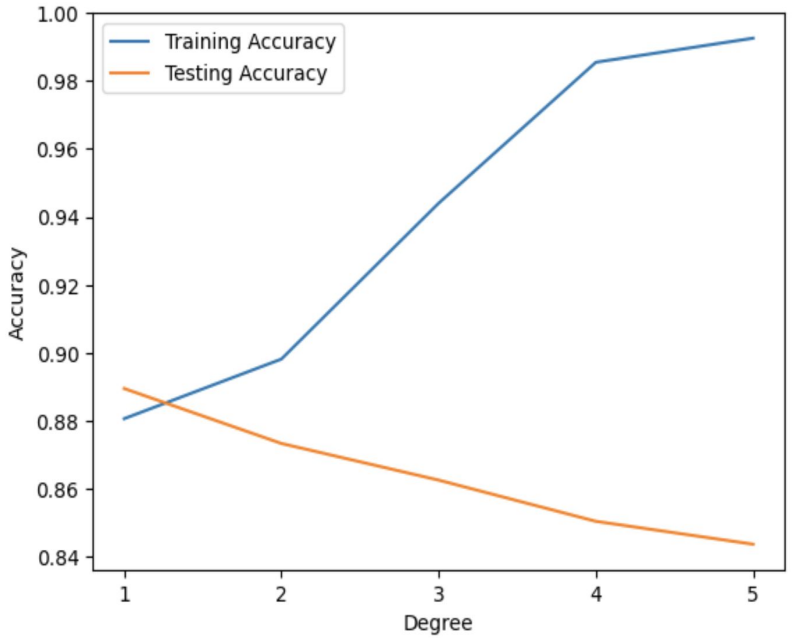


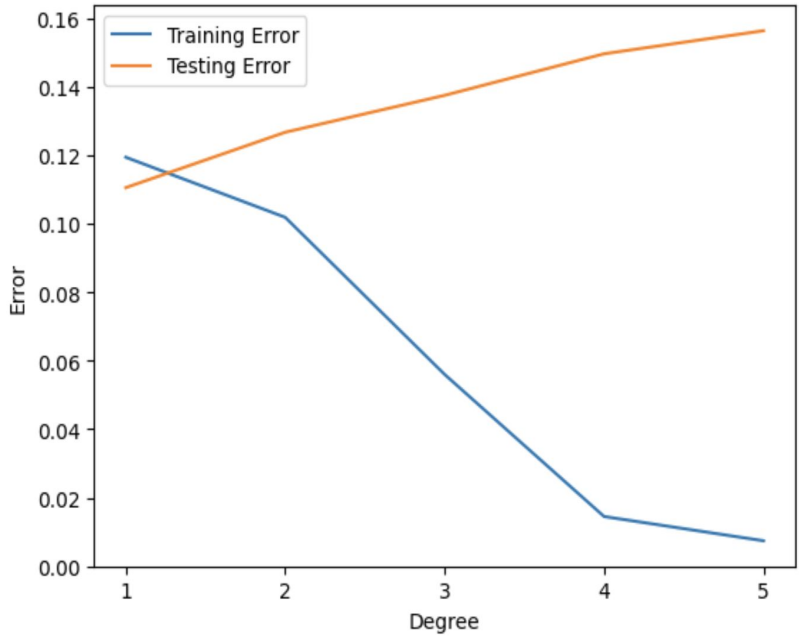
Logistic Regression (Multi-class)

	Multinomial Model							
	Lambda=0		Lambda=0.01		Lambda=0.1		Lambda=1	
Degree	Test accuracy	Training accuracy	Test accuracy	Training accuracy	Test accuracy	Training accuracy	Test accuracy	Training accuracy
1	88.95%	88.06%	88.95%	88.06%	88.95%	88.06%	88.81%	88.09%
2	87.60%	89.98%	87.47%	89.98%	87.33%	89.81%	87.06%	90.05%
3	83.56%	96.96%	85.31%	95.85%	86.25%	94.40%	87.06%	92.75%
4	84.23%	100.00%	83.02%	99.46%	85.04%	98.55%	86.52%	96.36%
5	83.69%	100.00%	83.29%	99.90%	84.37%	99.26%	85.44%	98.28%

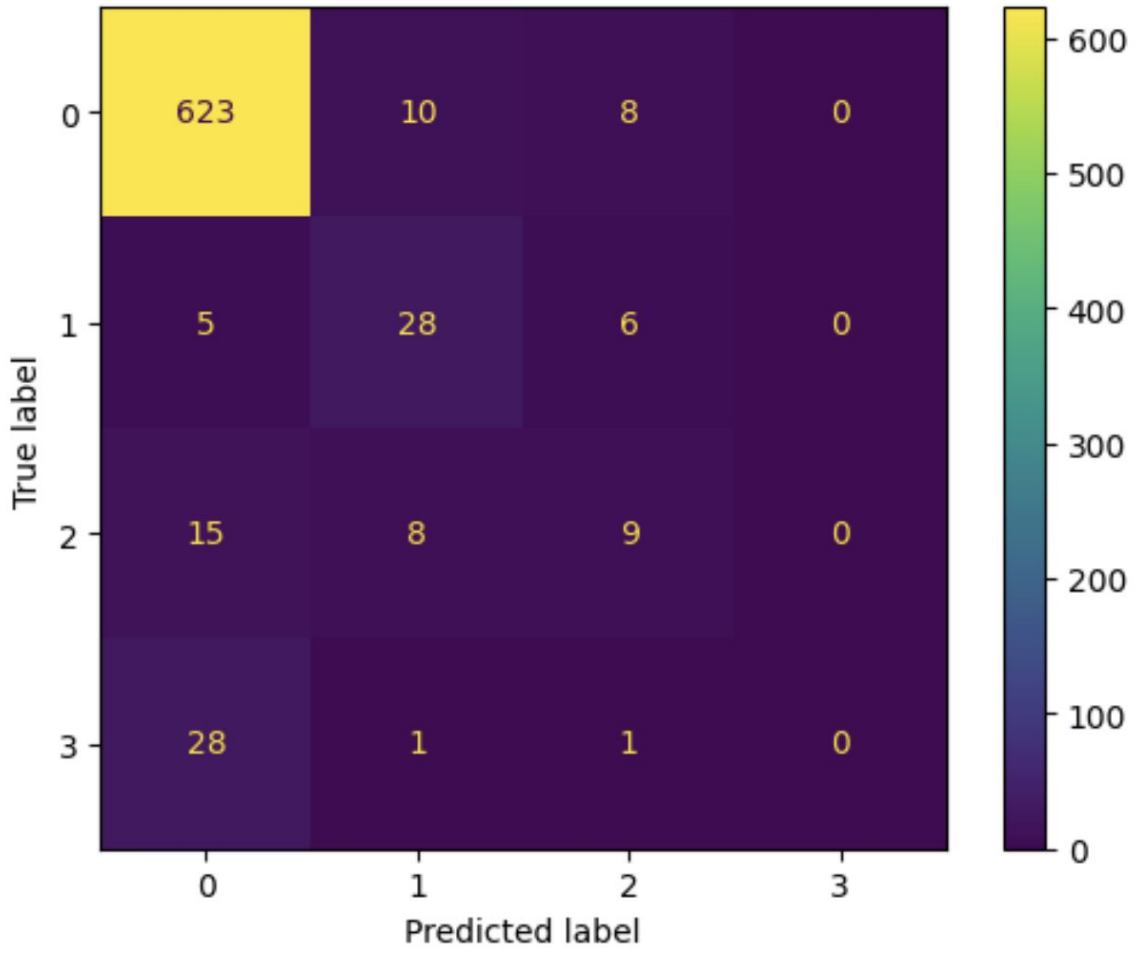
Logistic Regression (Multi-class)

Lambda = 0.1



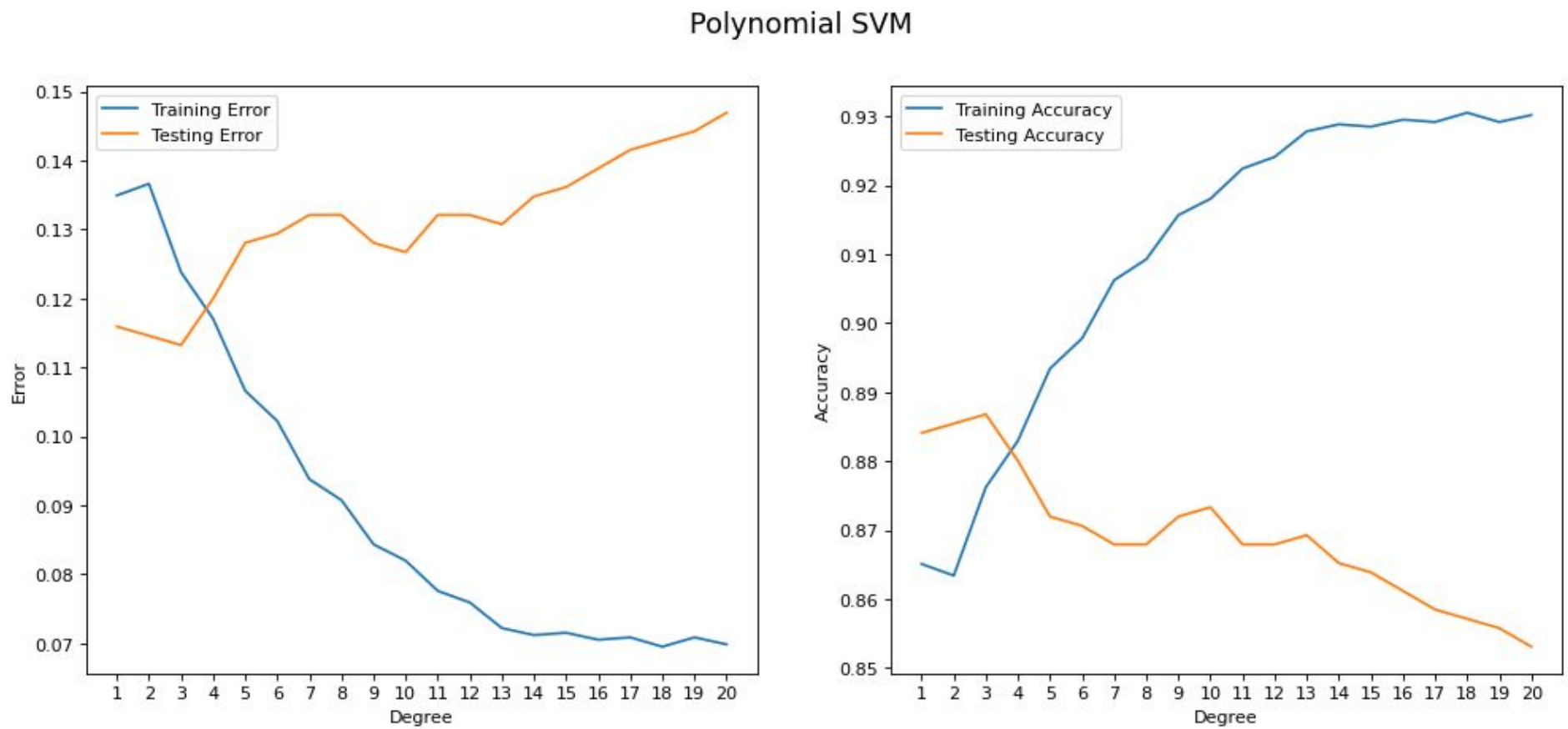


Logistic Regression (Multi-class)



Support Vector Machines

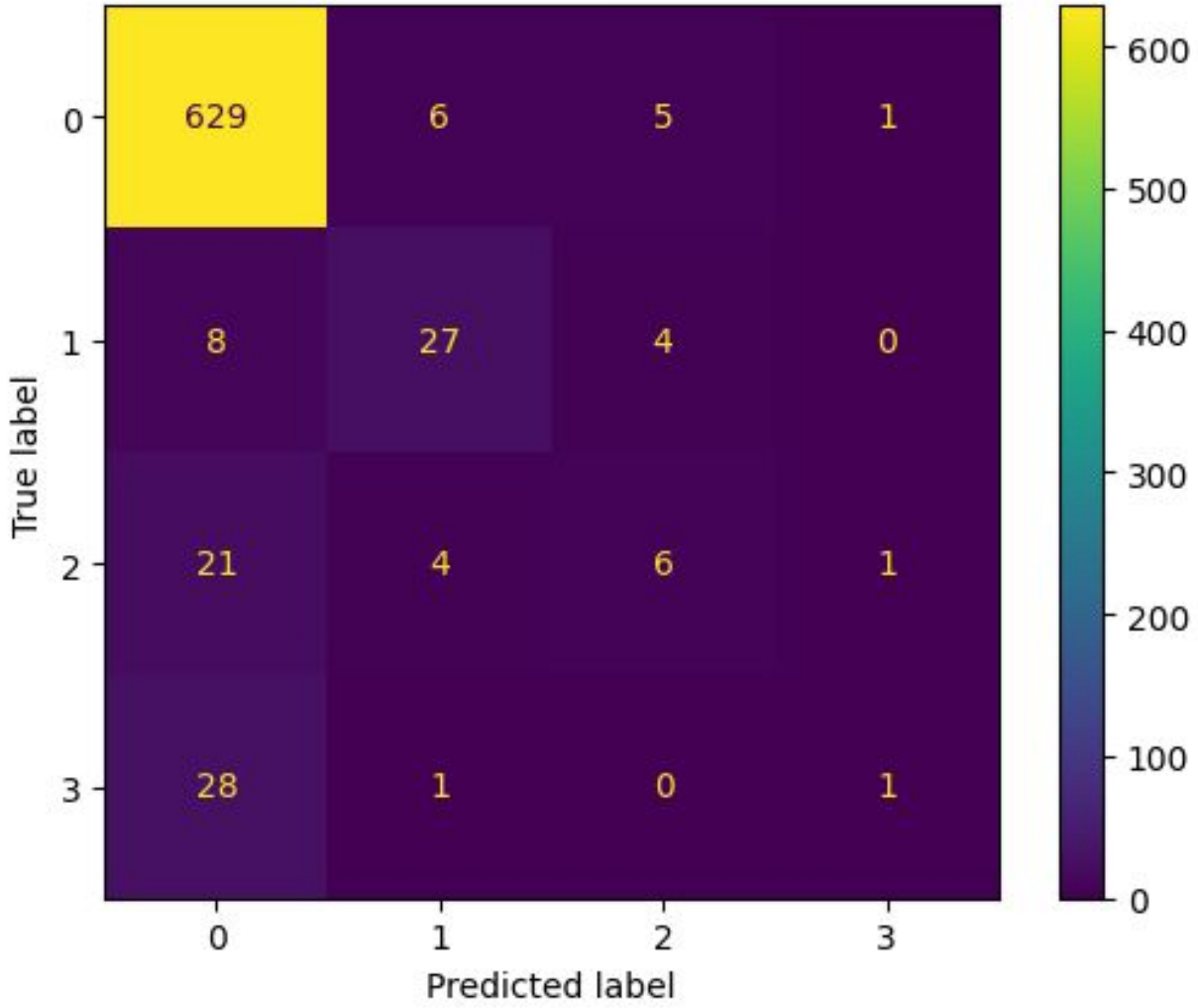
SVM Polynomial



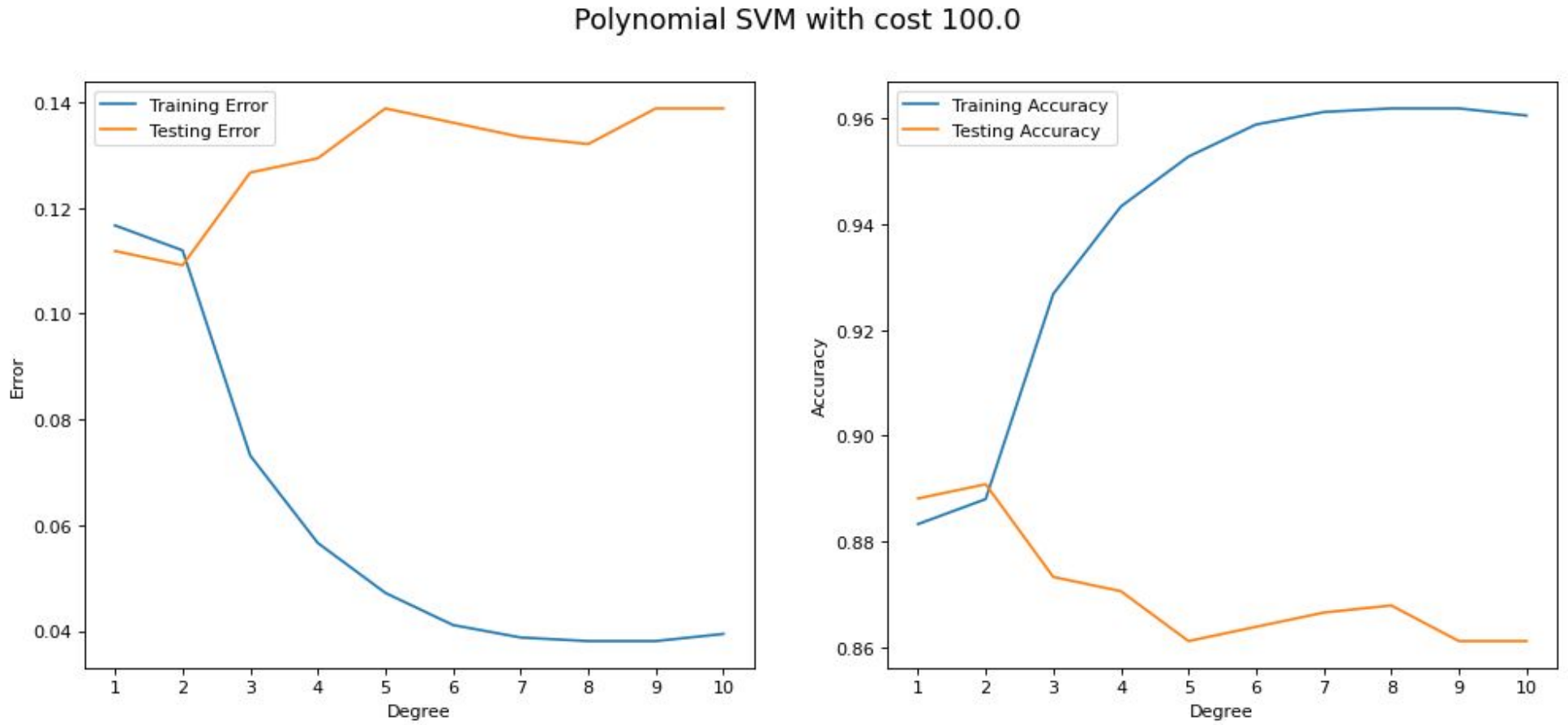
SVM Polynomial

Cost	0.0001		0.001		0.01		0.1	
Degree	**Test accuracy**	**Training accuracy**	**Test accuracy**	**Training accuracy**	**Test accuracy**	**Training accuracy**	**Test accuracy**	**Training accuracy**
1	86.39%	84.99%	86.39%	84.99%	87.20%	85.53%	88.41%	86.51%
2	86.39%	84.99%	87.06%	85.50%	88.14%	86.14%	88.54%	86.34%
3	86.79%	85.40%	87.47%	85.94%	88.27%	86.48%	88.68%	87.62%
4	87.20%	85.70%	87.87%	86.14%	88.27%	86.75%	88.01%	88.30%
5	87.47%	85.94%	88.27%	86.68%	87.60%	87.49%	87.20%	89.34%
6	88.27%	86.61%	88.14%	87.32%	87.20%	88.36%	87.06%	89.78%
7	87.87%	87.22%	87.60%	87.82%	86.79%	89.14%	86.79%	90.62%
8	87.47%	87.55%	87.06%	88.26%	86.93%	89.51%	86.79%	90.93%
9	86.39%	88.13%	86.66%	88.84%	86.66%	90.32%	87.20%	91.57%
10	86.25%	88.47%	86.93%	89.54%	86.66%	90.59%	87.33%	91.80%
Cost	**1**		***10***		**100**		**1,000**	
Degree	**Test accuracy**	**Training accuracy**	**Test accuracy**	**Training accuracy**	**Test accuracy**	**Training accuracy**	**Test accuracy**	**Training accuracy**
1	88.95%	87.28%	88.81%	88.09%	88.81%	88.33%	89.08%	88.26%
2	88.68%	87.25%	***89.35%***	88.33%	89.08%	88.80%	89.22%	88.97%
3	88.27%	89.61%	88.14%	90.86%	87.33%	92.68%	86.12%	94.44%
4	87.74%	89.98%	87.20%	91.74%	87.06%	94.33%	85.18%	96.73%
5	87.20%	91.13%	86.12%	93.42%	86.12%	95.28%	85.18%	97.34%
6	86.93%	91.74%	85.98%	93.96%	86.39%	95.89%	84.91%	97.74%
7	87.06%	92.82%	86.25%	94.47%	86.66%	96.12%	86.25%	97.84%
8	86.52%	93.12%	87.06%	94.64%	86.79%	96.19%	85.98%	97.71%
9	86.39%	93.56%	87.47%	94.74%	86.12%	96.19%	86.52%	97.47%
10	86.12%	93.59%	87.47%	94.74%	86.12%	96.05%	85.58%	97.37%

SVM Polynomial



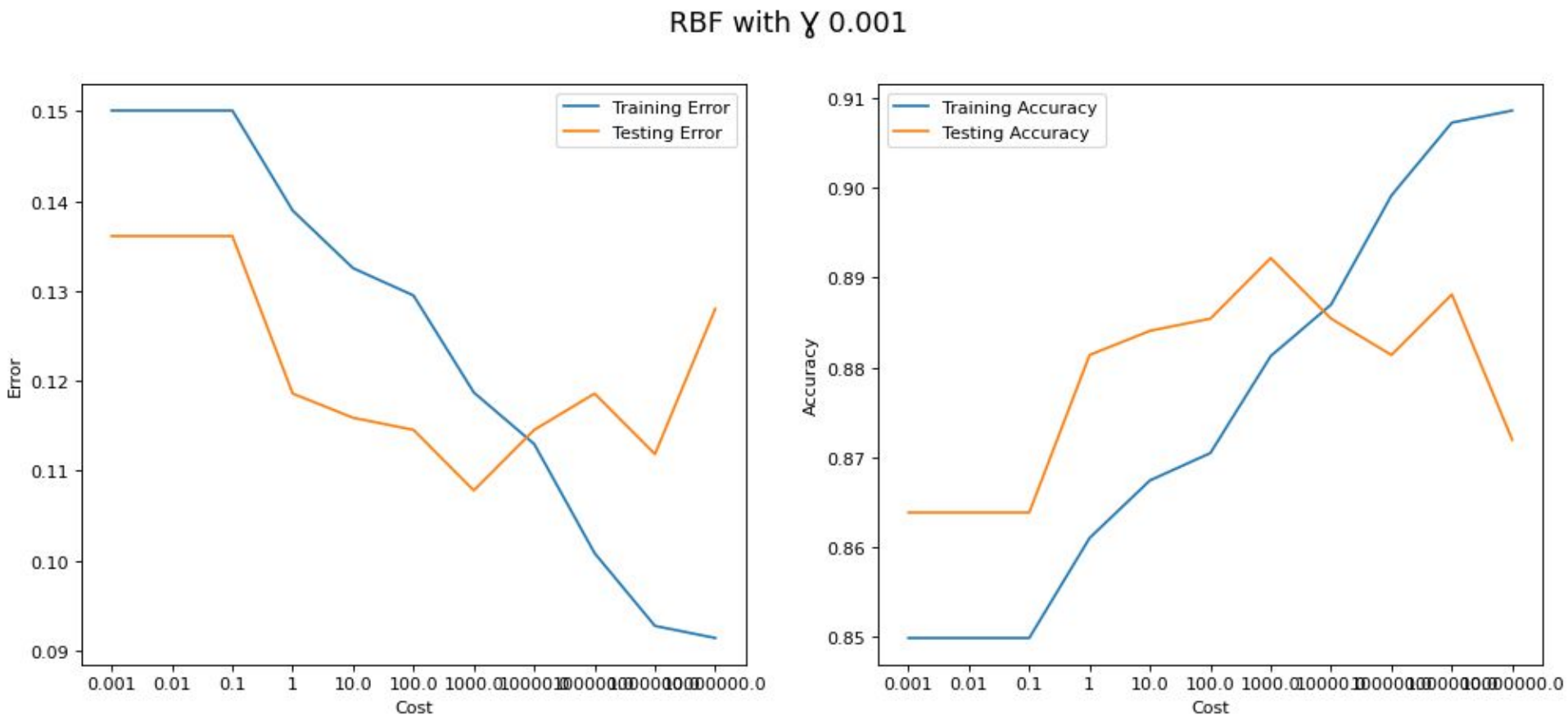
SVM Polynomial



SVM RBF

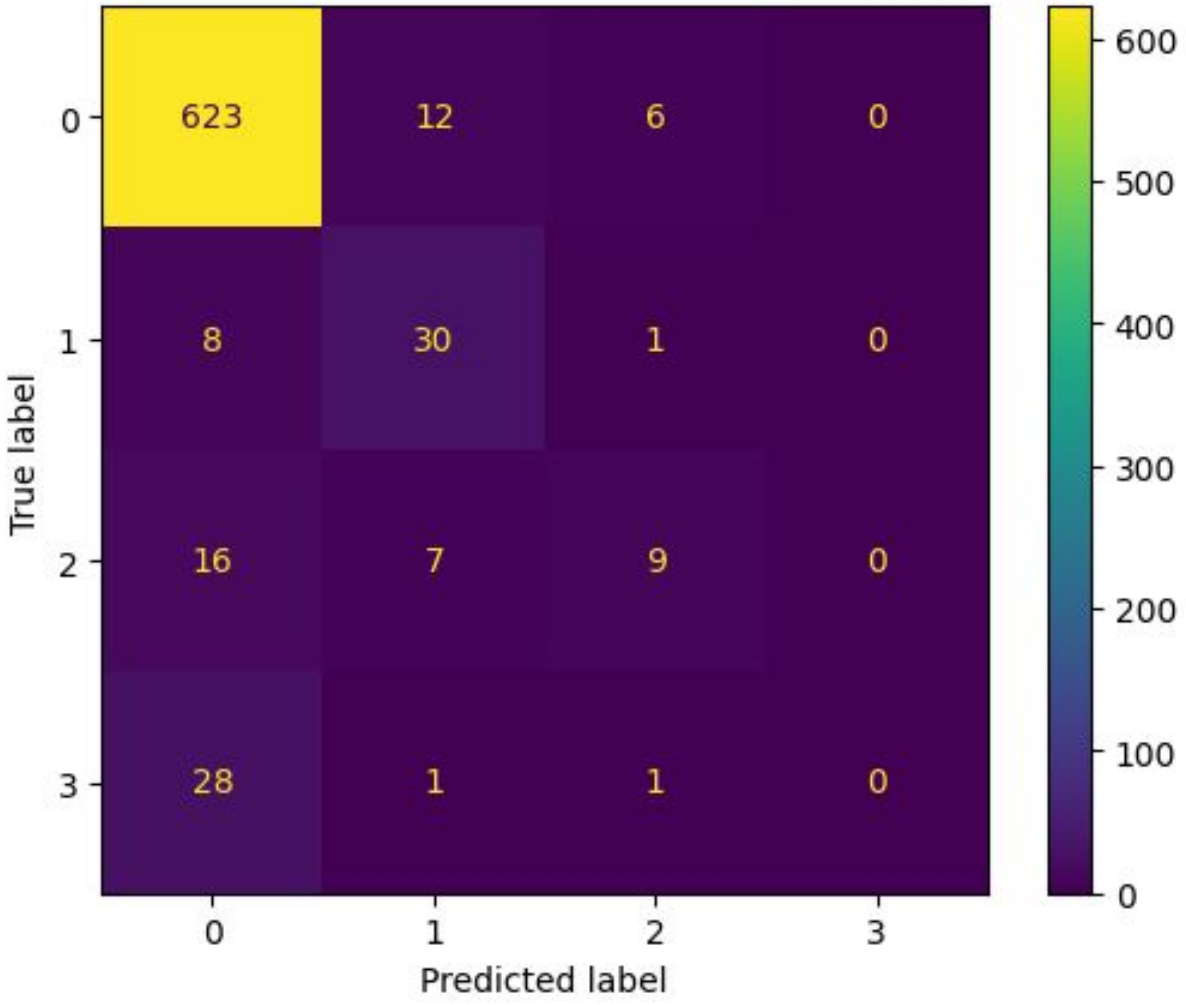
Gamma	Best Cost	Test accuracy	Training accuracy
0.000001	1000	88.14%	86.07%
0.00001	1000	88.54%	86.81%
0.0001	1000	89.08%	87.35%
0.001	***1000***	***89.22%***	***88.13%***
0.01	100	88.81%	88.74%
0.1	0.1	88.54%	86.37%
1.000	1	87.33%	92.28%
10.00	1	86.39%	98.75%
100.0	1	86.39%	10000.00%

SVM RBF



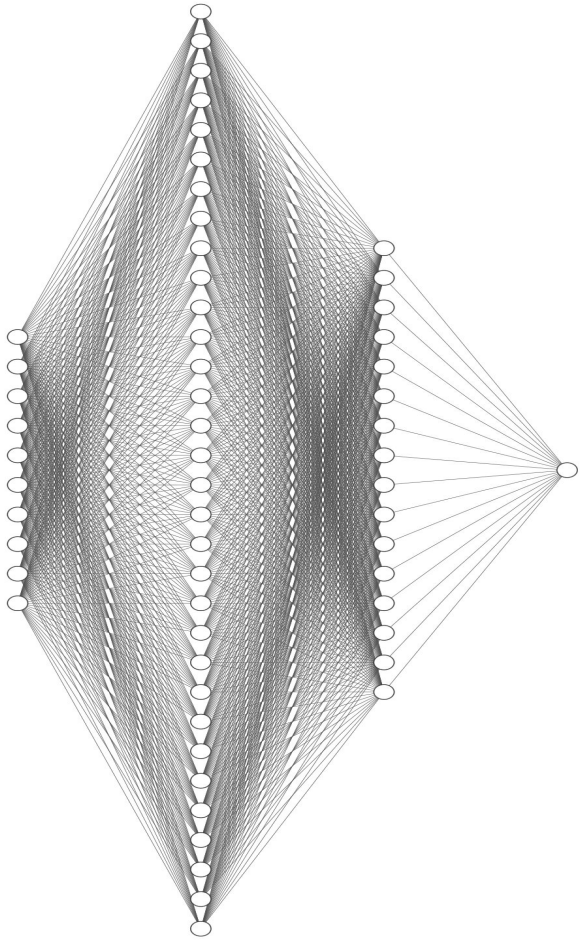
we apologise for the x-axis

SVM RBF


623
12
6
0
8
30
1
0
16
7
9
0
28
1
1
0
0
1
2
3
True label
Predicted label
600
500
400
300
200
100
0

Neural Network

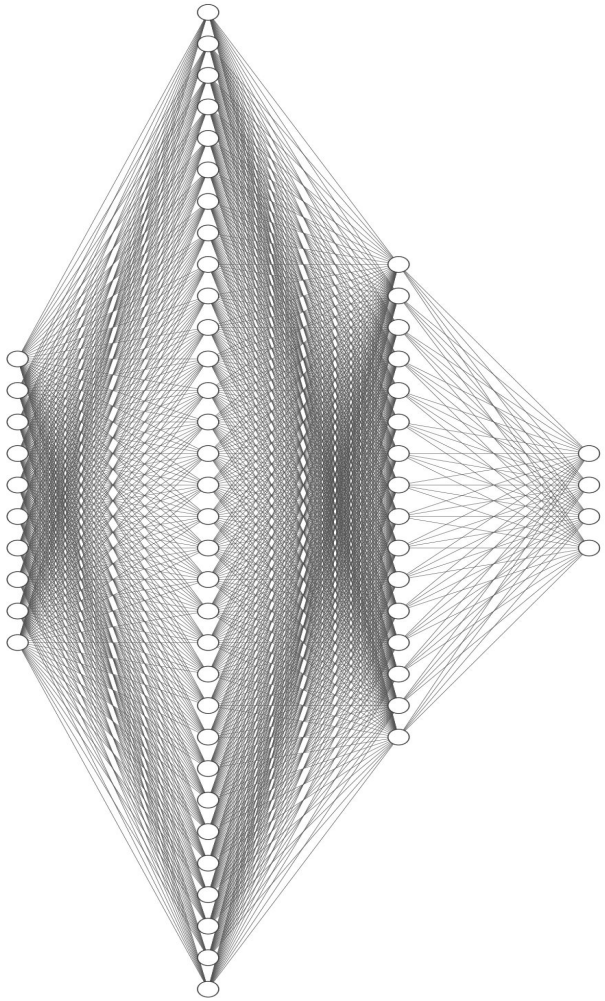
Neural Network (Binary)



Neural Network (Binary)

Binary Neural Network Accuracies			
Hidden Layer 1	**Hidden Layer 2**	**Output Layer**	**Accuracy**
ReLU	ReLU	ReLU	90.30%
ReLU	ReLU	Sigmoid	90.30%
Sigmoid	Sigmoid	Sigmoid	91.11%
tanh	tanh	Sigmoid	90.97%
ELU	ELU	Sigmoid	91.24%
SELU	SELU	Sigmoid	90.97%

Neural Network (Multi-class)

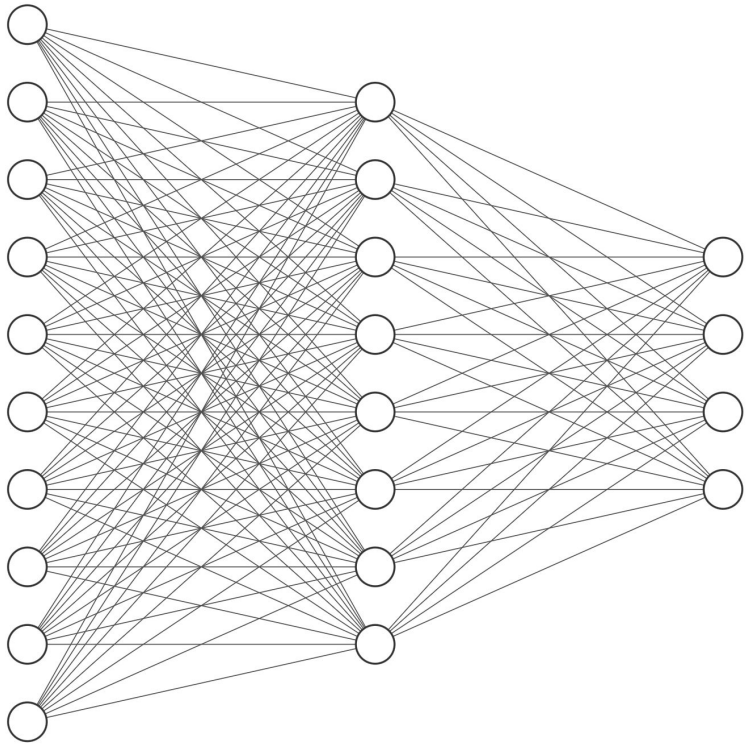


Neural Network (Multi-class)

Multinomial Neural Network Accuracies			
Hidden Layer 1	**Hidden Layer 2**	**Output Layer**	**Accuracy**
ReLU	ReLU	ReLU	69.54%
ReLU	ReLU	Sigmoid	9.84%
ReLU	ReLU	Softmax	10.38%
Sigmoid	Sigmoid	Softmax	9.70%
Softmax	tanh	Softmax	9.97%

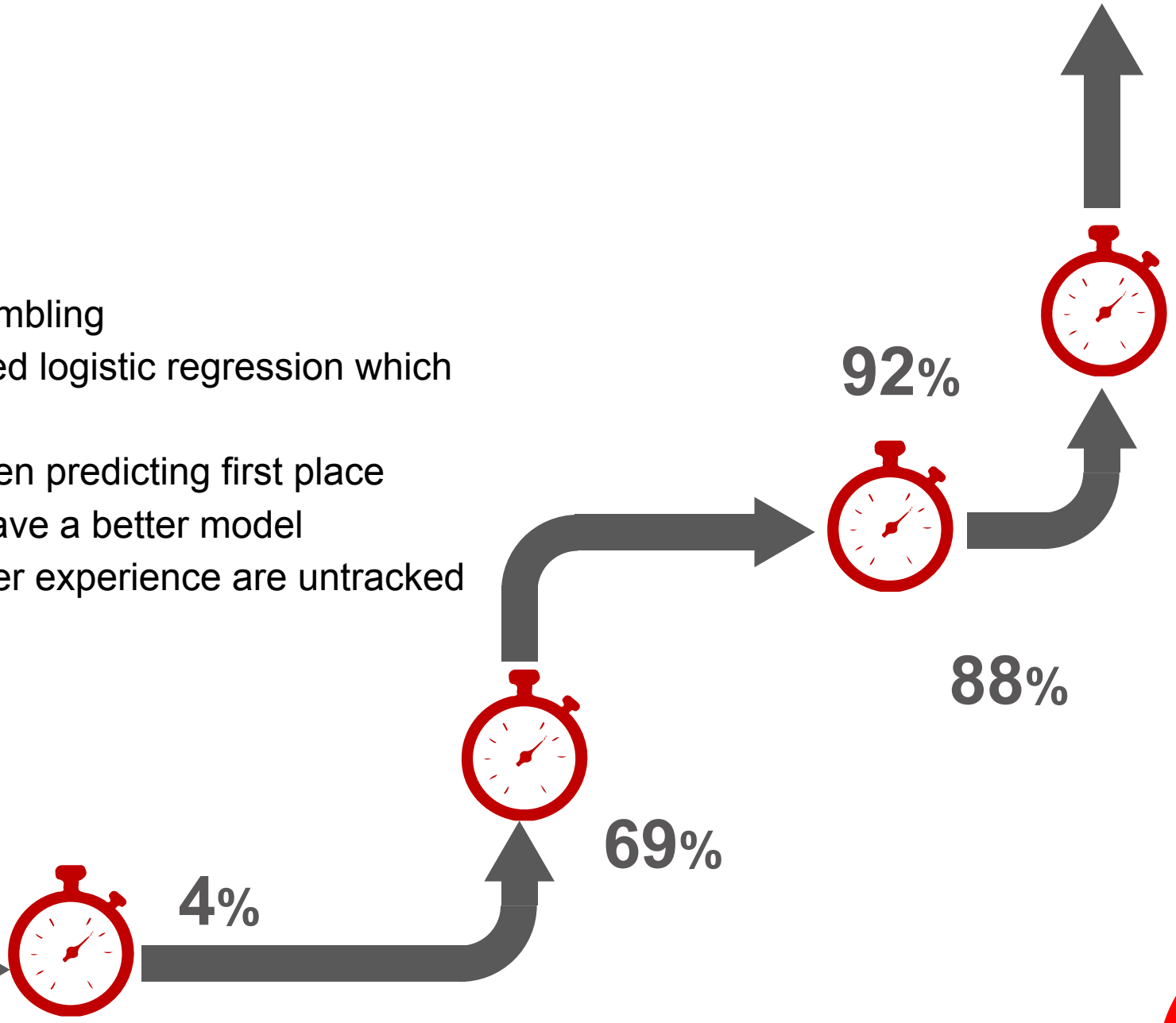
Neural Network (Simplified Multi-class)

91.64% Accuracy yayyyyy!



Conclusion

- Do not use this model for gambling
- Neural networks outperformed logistic regression which outperformed SVMs
- Surprisingly good results when predicting first place
- We need more features to have a better model
- Things like weather and driver experience are untracked



Thank You

Feel free to use our model but we are not responsible for any lost bets