



Website Classification

by Neurotic Networkers

Table of contents

01

Problem Overview

02

Model Introduction

03

Model Development

04

Experiment Results

05

Discussion

01

Problem Overview

A quick summary of the problem being addressed





Closed-World Multi-Class

(classification of monitored websites)

Open-World Binary

(classification between monitored and unmonitored websites)

Open-World Multi-Class

(classification of monitored websites and accounting for unmonitored websites)

02

Model Introduction

Introduction to our choice of model

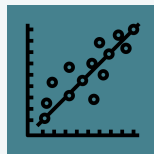


Empirical Results for Comparison: Closed-World Multi-Class



SVM

Model Accuracy: 35.95%

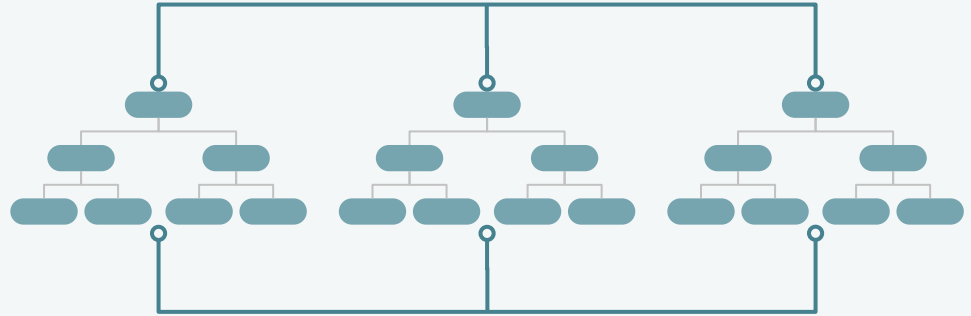


**Linear
Regression**

Model Accuracy: 2.97%

Random Forest

- Suited to large datasets
- Able to handle complex, non-linear relationships
- Parallelization
- Able to classify monitored websites & identify unmonitored websites



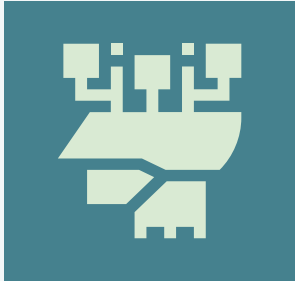
03

Model Development

Feature engineering and parameter selection



Experimental Situation



Processor Used: AMD Ryzen 9 5900HS, 3301 Mhz, 8 cores, 16 logical processors

Physical Memory: 16 GB

Sampling Method: Random (random state = 42)

Size of Data: 19,000 instances of monitored websites, 10,000 instances of unmonitored websites

Tuning

Techniques Used:

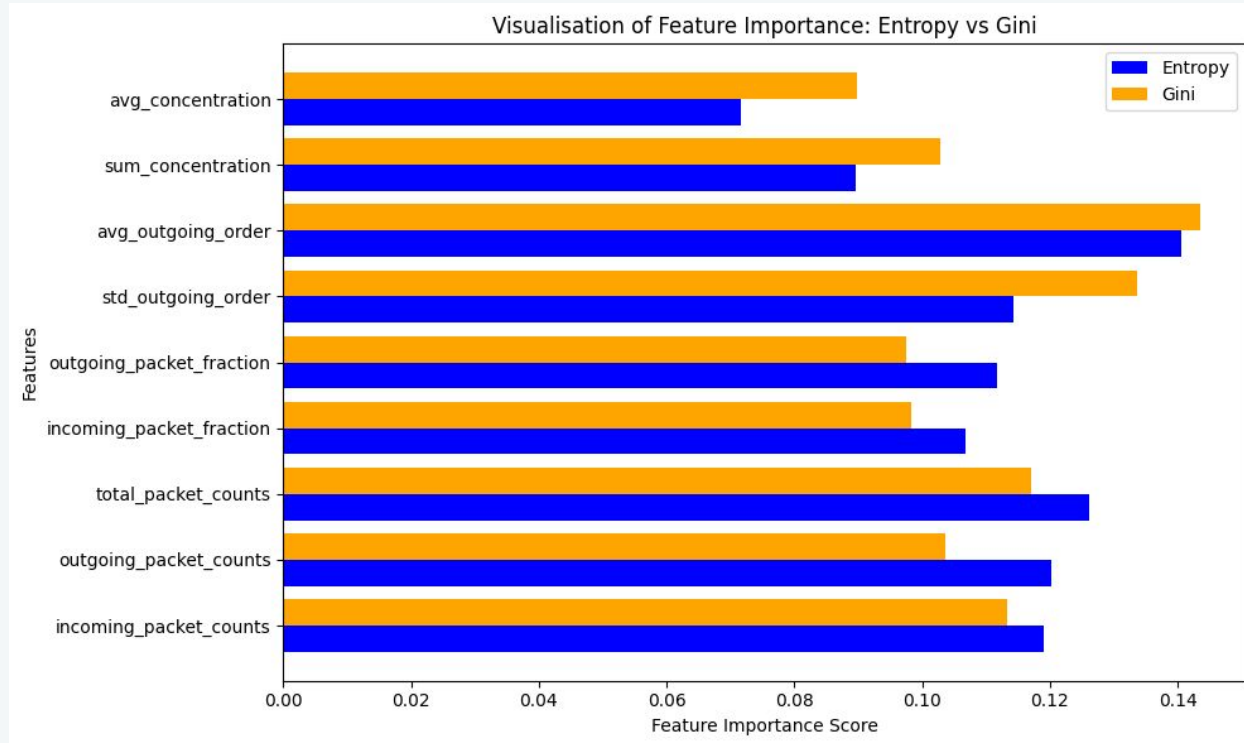


Parameter Grid



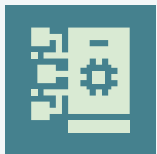
**Grid Search
(Cross Validation)**

Analyzed Feature Importance



Trial Feature Selection

Best features found:



**Number of
Incoming
Packets**



**Total Number of
Packets**



**Average of the
Outgoing
Packets
Ordering List**



**Sum of All Items
in Alternative
Concentration
Feature List**

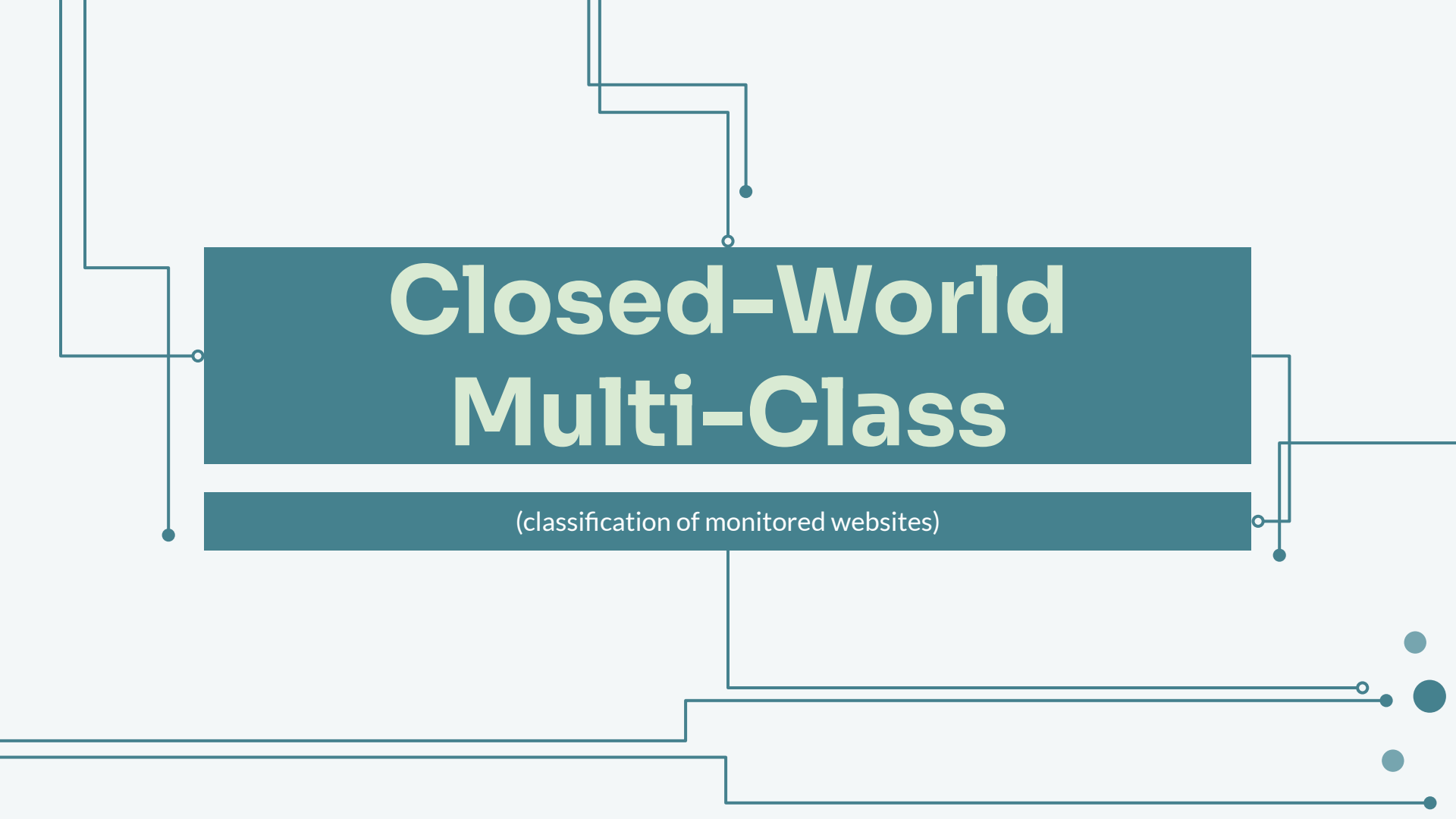
Accuracy went down by 3% and time only went down by 0.003 seconds (open-world multi-class), so **reducing features is unnecessary**

04

Experiment Results

Reporting results for each model





Closed-World Multi-Class

(classification of monitored websites)

Closed-World Multi-Class: RF Initial Metrics



Training

Memory used: 208.852 MB

Time taken to predict: 0.629 seconds

Model Accuracy: 83.7%

Metric	Precision	Recall	F1-Score
Accuracy			0.84
Macro Avg	0.85	0.84	0.84
Weighted Avg	0.85	0.84	0.84



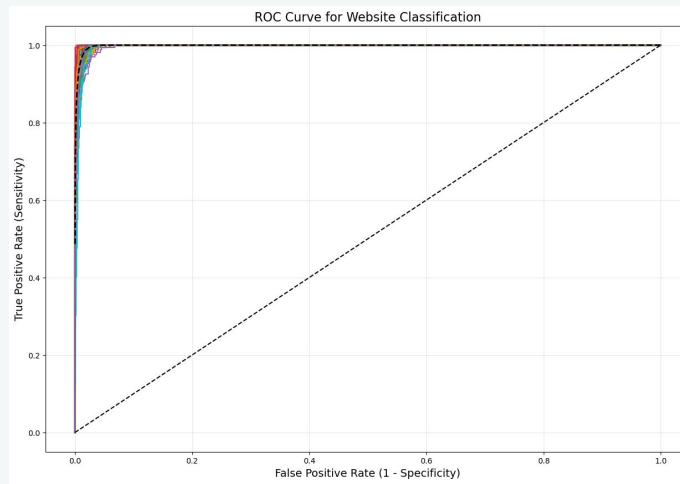
Testing

Memory used: 209.195 MB

Time taken to predict: 0.143 seconds

Model Accuracy: 67.2%

Metric	Precision	Recall	F1-Score
Accuracy			0.67
Macro Avg	0.68	0.68	0.67
Weighted Avg	0.68	0.67	0.67



Closed-World Multi-Class: RF Final Metrics

Using best parameters found: {'bootstrap': True, 'criterion': 'entropy', 'max_depth': 20, 'min_samples_leaf': 5, 'min_samples_split': 10, 'n_estimators': 200}



Training

Memory used: 1034.355 MB

Time taken to predict: 1.181 seconds

Model Accuracy: 87.1%

Metric	Precision	Recall	F1-Score
Accuracy			0.87
Macro Avg	0.87	0.87	0.87
Weighted Avg	0.87	0.87	0.87



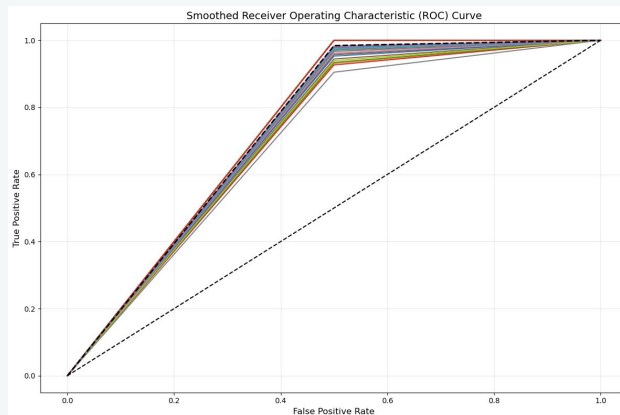
Testing

Memory used: 1034.434 MB

Time taken to predict: 0.269 seconds

Model Accuracy: 70.2%

Metric	Precision	Recall	F1-Score
Accuracy			0.70
Macro Avg	0.70	0.70	0.70
Weighted Avg	0.71	0.70	0.70





Open-World Binary

(classification between monitored and unmonitored websites)

Open-World Binary: RF Initial Metrics



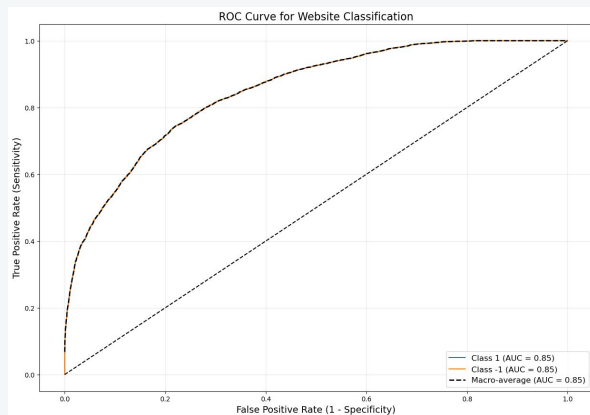
Training

Memory used: 225.375 MB

Time taken to predict: 0.145 seconds

Model Accuracy: 77.3%

Metric	Precision	Recall	F1-Score
Accuracy			0.77
Macro Avg	0.79	0.69	0.71
Weighted Avg	0.78	0.77	0.75



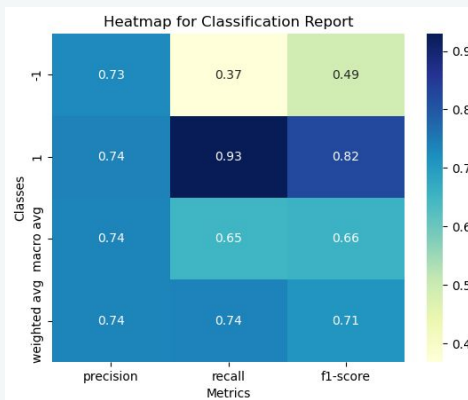
Testing

Memory used: 225.383 MB

Time taken to predict: 0.043 seconds

Model Accuracy: 73.8%

Metric	Precision	Recall	F1-Score
Accuracy			0.74
Macro Avg	0.74	0.65	0.66
Weighted Avg	0.74	0.74	0.71



Open-World Binary: RF Final Metrics

Using best parameters found: {'bootstrap': True, 'criterion': 'gini', 'max_depth': 30, 'min_samples_leaf': 5, 'min_samples_split': 10, 'n_estimators': 200}



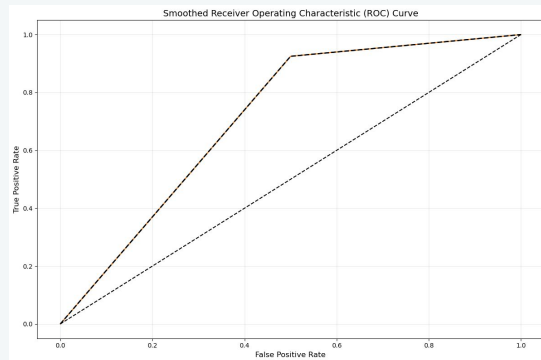
Training

Memory used: 288.023 MB

Time taken to predict: 0.638 seconds

Model Accuracy: 94.0%

Metric	Precision	Recall	F1-Score
Accuracy			0.94
Macro Avg	0.95	0.92	0.93
Weighted Avg	0.94	0.94	0.94



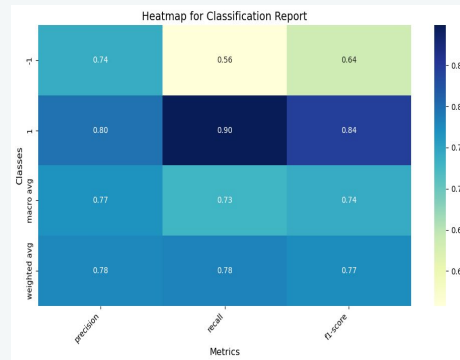
Testing

Memory used: 288.188 MB

Time taken to predict: 0.183 seconds

Model Accuracy: 78.2%

Metric	Precision	Recall	F1-Score
Accuracy			0.78
Macro Avg	0.77	0.73	0.74
Weighted Avg	0.78	0.78	0.77



The background features a light gray color with a network of thin, dark teal lines. These lines are mostly vertical and horizontal, with some diagonal segments, creating a circuit-like or architectural feel. Small teal dots are placed at various points where the lines intersect or terminate. The main title is centered within a large teal rectangle, and a subtitle is centered within a smaller teal rectangle below it.

Open-World Multi-Class

(classification of monitored websites and accounting for unmonitored websites)

Open-World Multi-Class: RF Initial Metrics



Training

Memory used: 618.020 MB

Time taken to predict: 0.938 seconds

Model Accuracy: 70.3%

Metric	Precision	Recall	F1-Score
Accuracy			0.70
Macro Avg	0.88	0.57	0.66
Weighted Avg	0.78	0.70	0.68



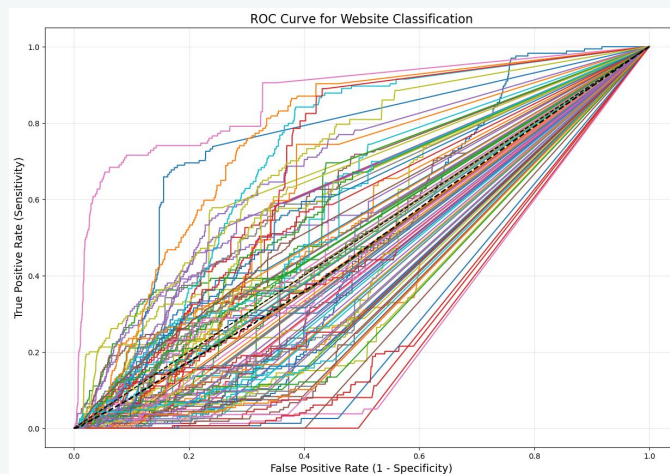
Testing

Memory used: 618.043 MB

Time taken to predict: 0.245 seconds

Model Accuracy: 60.2%

Metric	Precision	Recall	F1-Score
Accuracy			0.60
Macro Avg	0.77	0.46	0.53
Weighted Avg	0.68	0.60	0.57



Open-World Multi-Class: RF Final Metrics

Using best parameters found: {'bootstrap': True, 'criterion': 'entropy', 'max_depth': 20, 'min_samples_leaf': 5, 'min_samples_split': 10, 'n_estimators': 200}



Training

Memory used: 715.512 MB

Time taken to predict: 2.914 seconds

Model Accuracy: 82.9%

Metric	Precision	Recall	F1-Score
Accuracy			0.83
Macro Avg	0.93	0.76	0.82
Weighted Avg	0.86	0.83	0.82



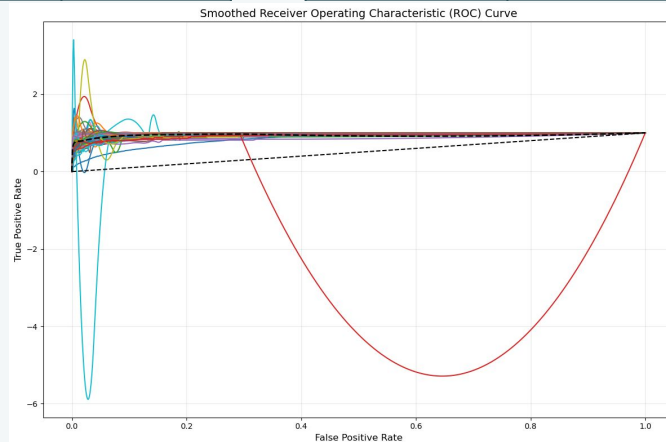
Testing

Memory used: 717.316 MB

Time taken to predict: 0.818 seconds

Model Accuracy: 66.8%

Metric	Precision	Recall	F1-Score
Accuracy			0.67
Macro Avg	0.78	0.57	0.63
Weighted Avg	0.71	0.67	0.65




05

Discussion

Interpretations and learnings from this experiment





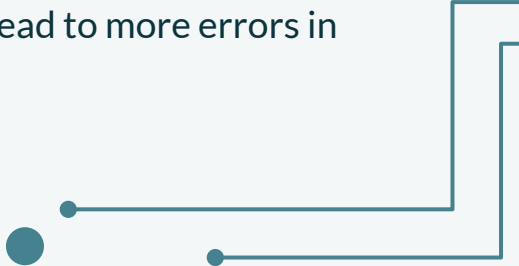
Final test accuracy		
Closed world multi-class	Open world binary	Open world multi-class
70.2%	78.2%	66.8%

Open world binary classification > Closed world multi-class > Open world multi-class

This ranking is logical as:

- Fewer classes makes decision boundary easier
- Open world multiclass adds further difficulty by requiring the model to recognize an additional class

The binary model also had a higher recall and precision than the other two models

- Multi-class models must distinguish among several classes, which can lead to more errors in both precision and recall.
- 

The slide features a light blue background with a network of dark blue lines and dots. These lines, resembling circuit traces, enter from the top and left, and exit towards the bottom and right. Some lines terminate in small solid dots, while others end in open circles. In the bottom right corner, there are three solid dark blue circles of varying sizes. The main content is presented in three horizontal dark teal bars.

Thanks!

The Neurotic Networkers

[Project Github Link](#)

Thanks!

Do you have any questions?

youremail@freepik.com

+34 654 321 432

yourwebsite.com



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

