A COMPARISON OF PREDICTION METHODS FOR LAPSES AND SURRENDERS, USING INSURANCE AND ANNUITY DATA

by Joseph Laquis

Introduction

Goal

- Attain an understanding of policy lapses and surrenders
- Significant financial impacts and benefits of predicting lapses and surrenders
- Process involved in the collection and analysis of the data
- Examination of the predictive methods for optimal results

Definition

Lapses

- A lapse happens when policyholders cease premium payments, resulting in the termination of the policy.
- **Process:** It refers to the discontinuation of a policy initiated by the policyholder, either through action or inaction, beyond the grace period.

Surrenders

- A surrender occurs when policyholders terminate their contract prematurely, and this can happen at any point.
- **Process:** Policy surrender involves the policyholder deciding to end their insurance plan. During a surrender, the policyholder receives any accumulated cash value from the policy.

Why Predict Lapses and Surrenders?

- Predicting lapses and surrenders enables insurers to reduce risks, lower costs, and retain policyholders.

- Proactive measures can lead to financial stability, reduced losses, and optimized investment strategies.

- Predicting these events helps insurance companies build stronger relationships with their clients, fostering trust and long-term loyalty.

Financial Impacts of Lapses

- Lapses result in lost premium income, negatively impacting an insurer's profitability.
- Predicting and managing lapses is vital for maintaining financial stability and ensuring policyholder satisfaction.
- Insurance companies may also face policy reissuance costs when lapses occur.

Advantages of Predicting Lapses

- Predicting lapses allows insurers to implement proactive retention strategies, reduce financial losses, and maintain policyholder satisfaction

- It leads to improved profitability and long-term financial stability.

- It also aids in avoiding reissuance costs and policy reinstatement efforts.

Financial Impacts of Surrenders

- Surrenders lead to the loss of policy reserves and can disrupt an insurer's investment strategy.

- Predicting surrenders is crucial for maintaining financial stability and ensuring a sustainable business model.

- Insurance companies may also face early surrender charges and administrative costs when surrenders happen.

Advantages of Predicting Surrenders

- Predicting surrenders helps insurers retain policyholders, maintain a stable customer base, and optimize investment strategies.

- It contributes to financial stability, enhanced customer relationships, and long-term growth.

- Predicting surrenders can also save insurers from incurring early surrender charges and administrative expenses.

Methodology and Data Analysis Overview

Data and Variables

- This study represents the first empirical investigation of Trinidad and Tobago's life insurance market, utilizing genuine policyholder data from a local insurance company spanning a five-year period (2015 to 2019).

- The study encompasses the following product types:

1) Traditional Life (Whole of Life, Term Life and Endowment)
2) Deferred Type of Accumulation Annuity
3) Universal-Life

Continued…

- The analysis is based on a dataset containing policy-related variables that directly influence policyholder behavior.

- In a binary grouping, a confusion matrix displays the expected outcomes, allowing a closer examination of the results, and is as follows:

		Predicted Group	
		+	-
Actual Group	+	True Positive (TP)	False Negative (FN)
	-	False Positive (FP)	True Negative (TN)

Continued…

- Key performance metrics used were:

Accuracy: How often are our predictions correct

Precision: Gauges exactness in our positive predictions

Recall: Captures the effectiveness of catching all actual positive instances

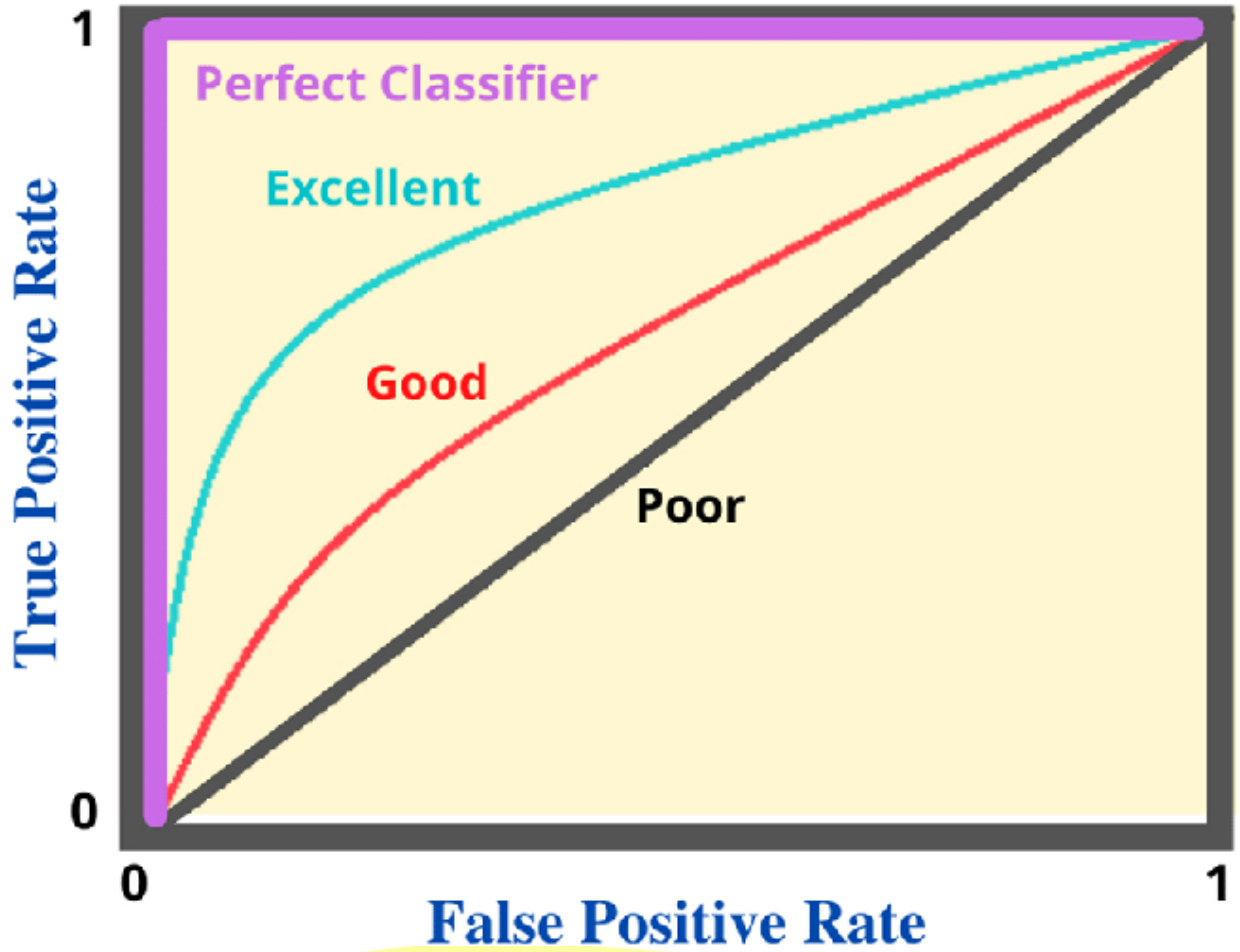
F1 Score: Strikes a balance between Precision and Recall

AUC (Area under the ROC curve): Evaluates the model's ability to distinguish between two outcomes

Cohen's Kappa Statistic: Measures how well predictions match actual values

RMSE (Root Mean Square Error): How far, on average, are we from the target?

Receiver Operating Characteristic Curve



Predictive Analytics

Predictive analytics uses advanced data analysis for forecasting lapses and surrenders, by analyzing policyholder behavior and various data sources to create accurate predictive methods

Predictive Methods can be categorized as follows:

Supervised Method:

Explores the relationship between independent variables (X) and a dependent variable (Y).

Semi-Supervised Method:

1. Particularly useful for managing large datasets.
2. Doesn't necessitate names or labels for every element in the data.

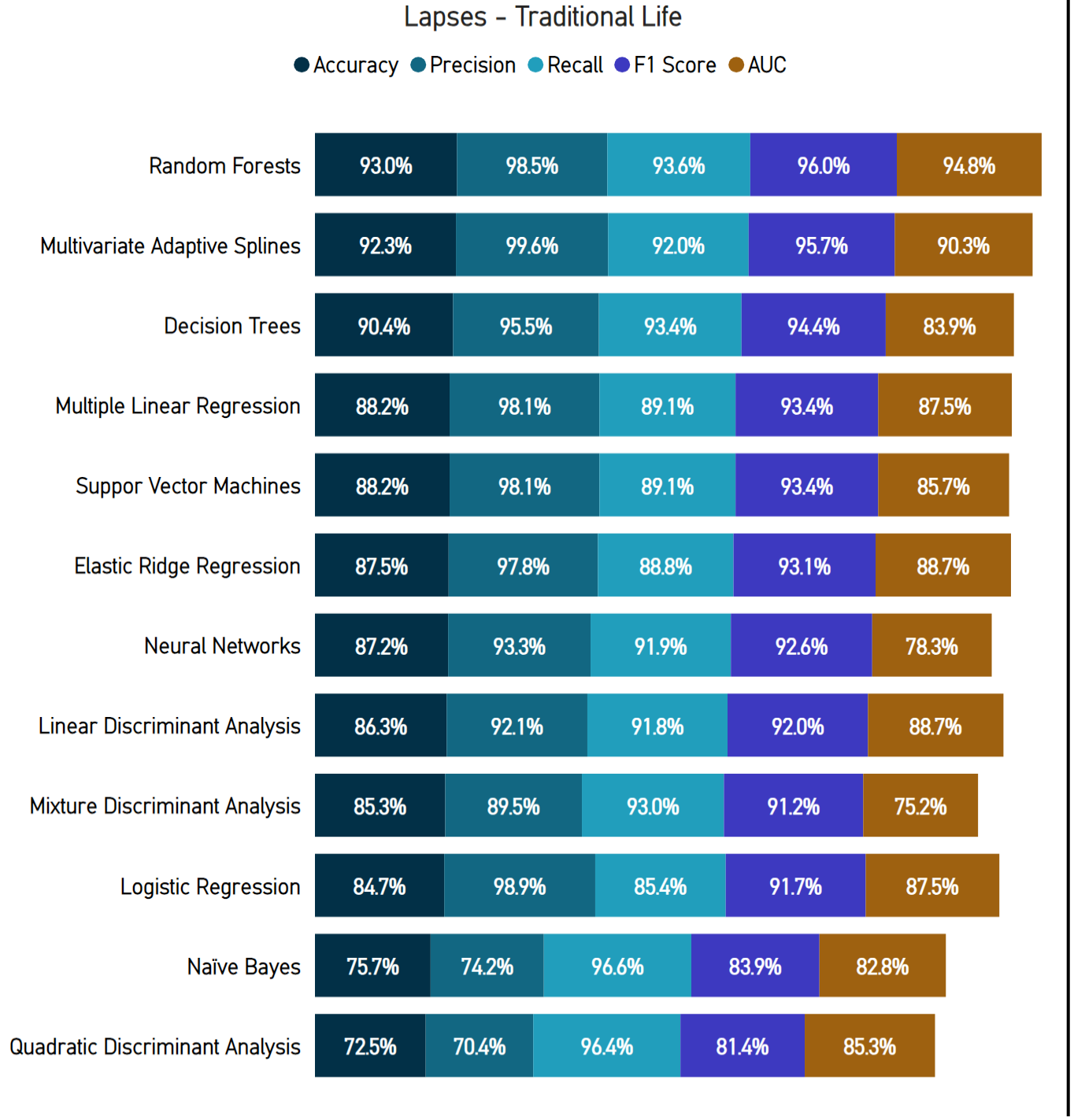
Unsupervised Method:

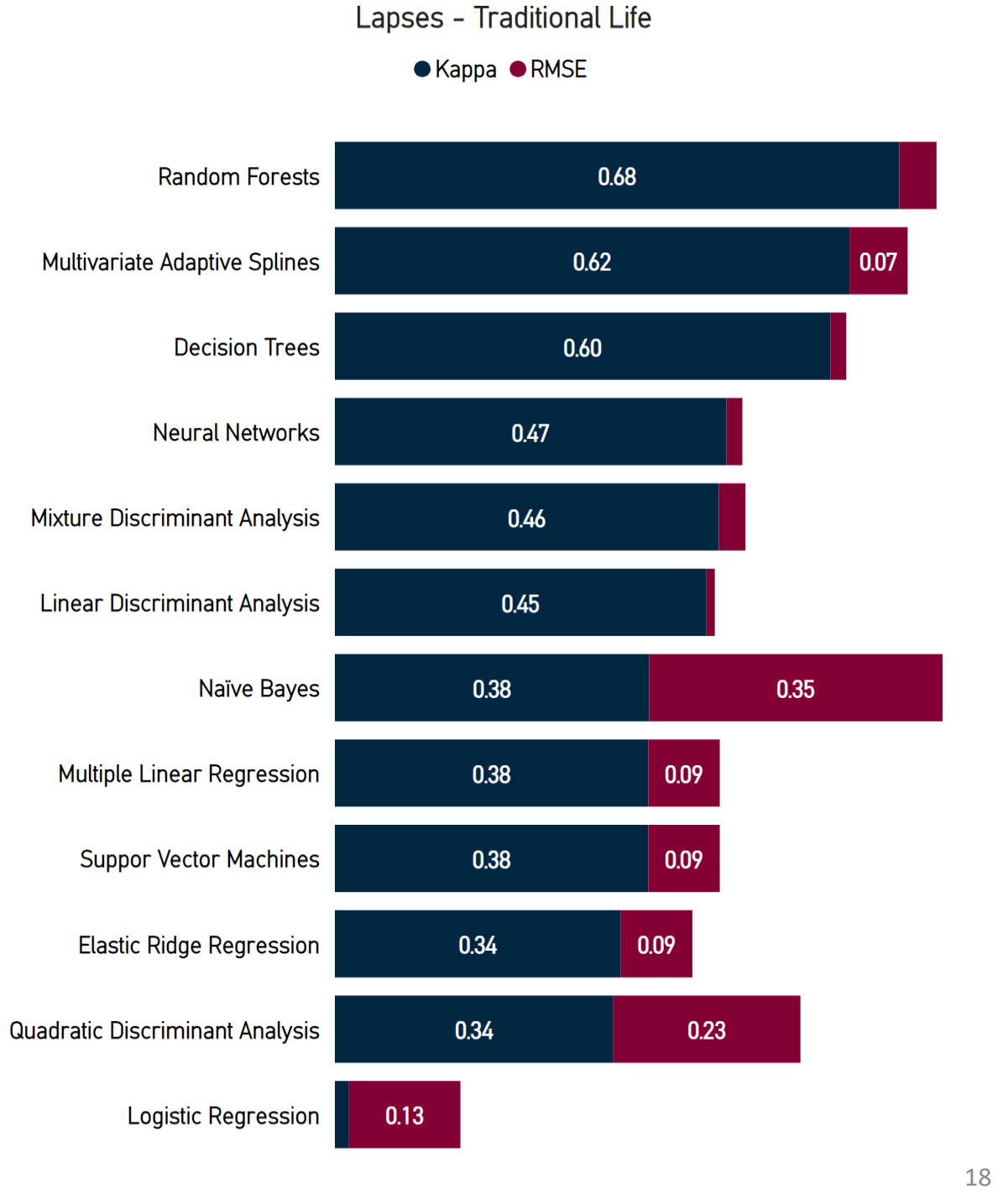
1.Identifies patterns within datasets without prior knowledge or labels.
2.Reveals underlying structures or relationships within the data.

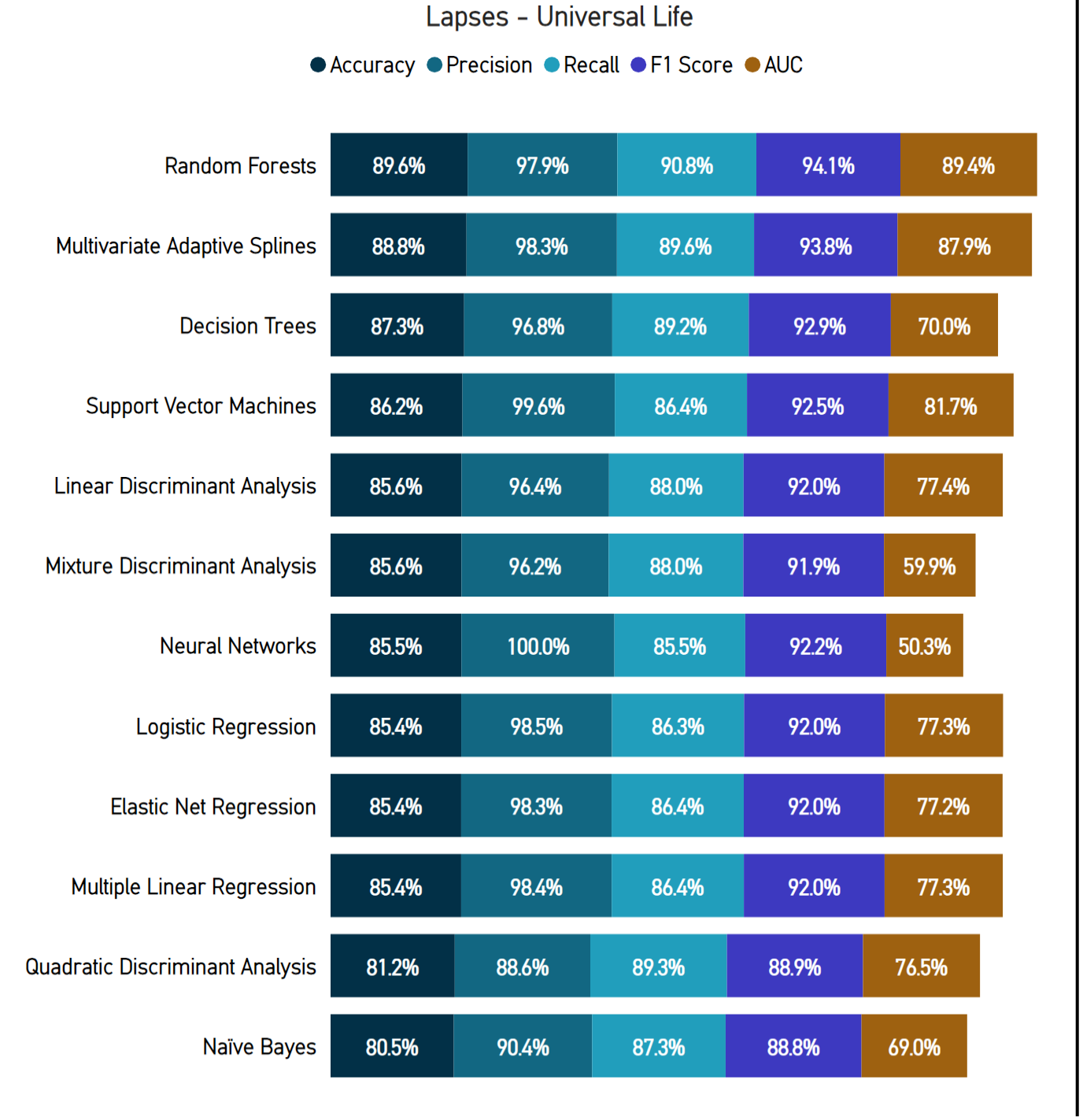
Predictive Methods

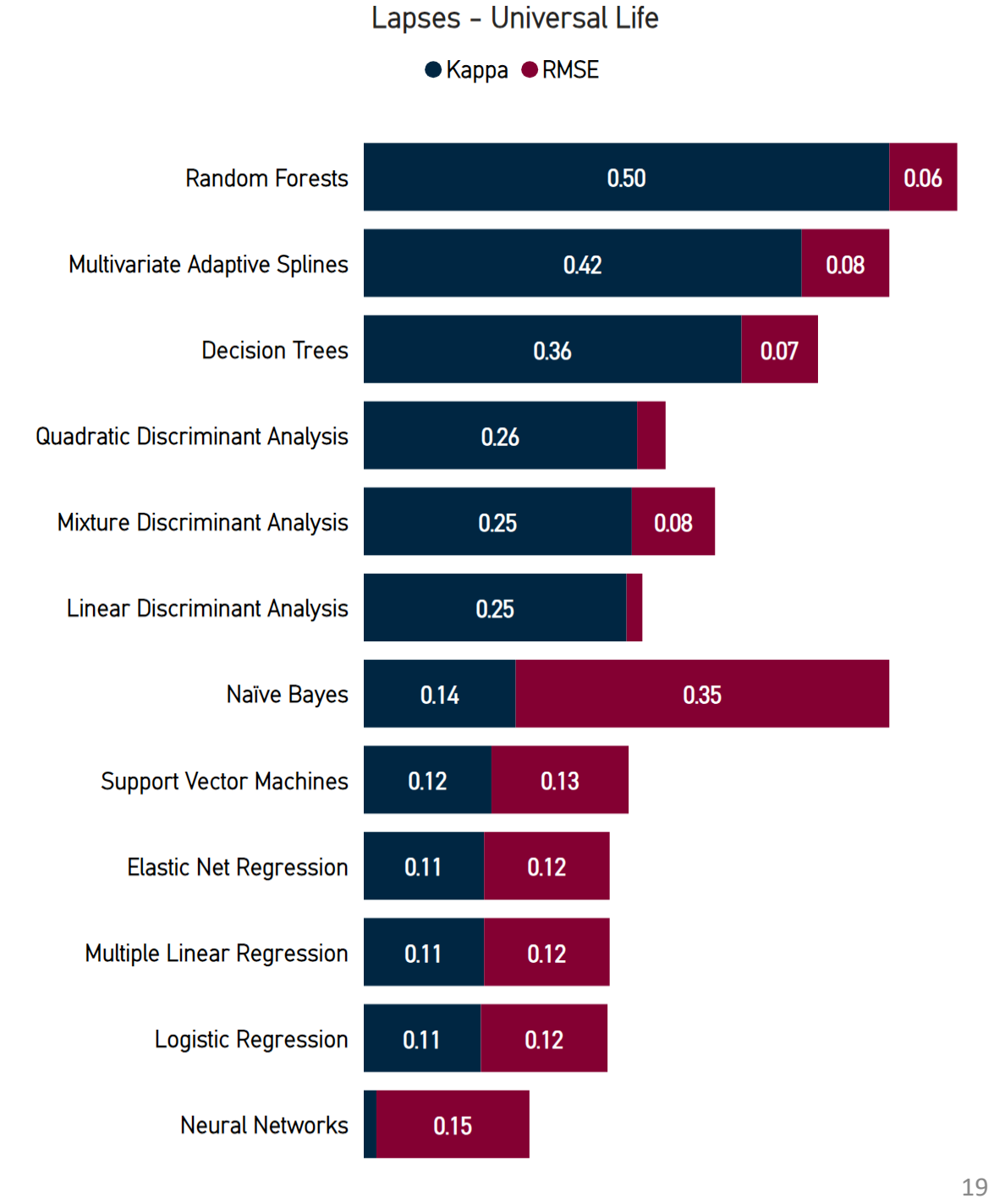
Method	Description
Naïve Bayes' Classifier	Probability-based classifier
Multiple Linear Regression	Linear relationship between multiple independent variables and a dependent variable
Logistic Regression	Probability-based classification for binary outcomes
Elastic Net Regression	Selects and weighs variables for accurate predictions in complex datasets
Linear Discriminant Analysis	Maximizes class separability in linearly transformed feature space
Quadratic Discriminant Analysis	Extension of LDA, considers quadratic decision boundaries
Mixture Discriminant Analysis	Combine both linear and quadratic components for a more adaptable decision boundary
Decision Trees	Hierarchical tree structure for decision-making
Random Forests	Ensemble of decision trees for improved accuracy
Multivariate Adaptive Regression Splines (MARS)	Flexible non-linear regression technique
Artificial Neural Networks (ANN)	Mimics human brain structure for complex tasks
Support Vector Machines (SVMs)	Finds optimal hyperplane for classification tasks

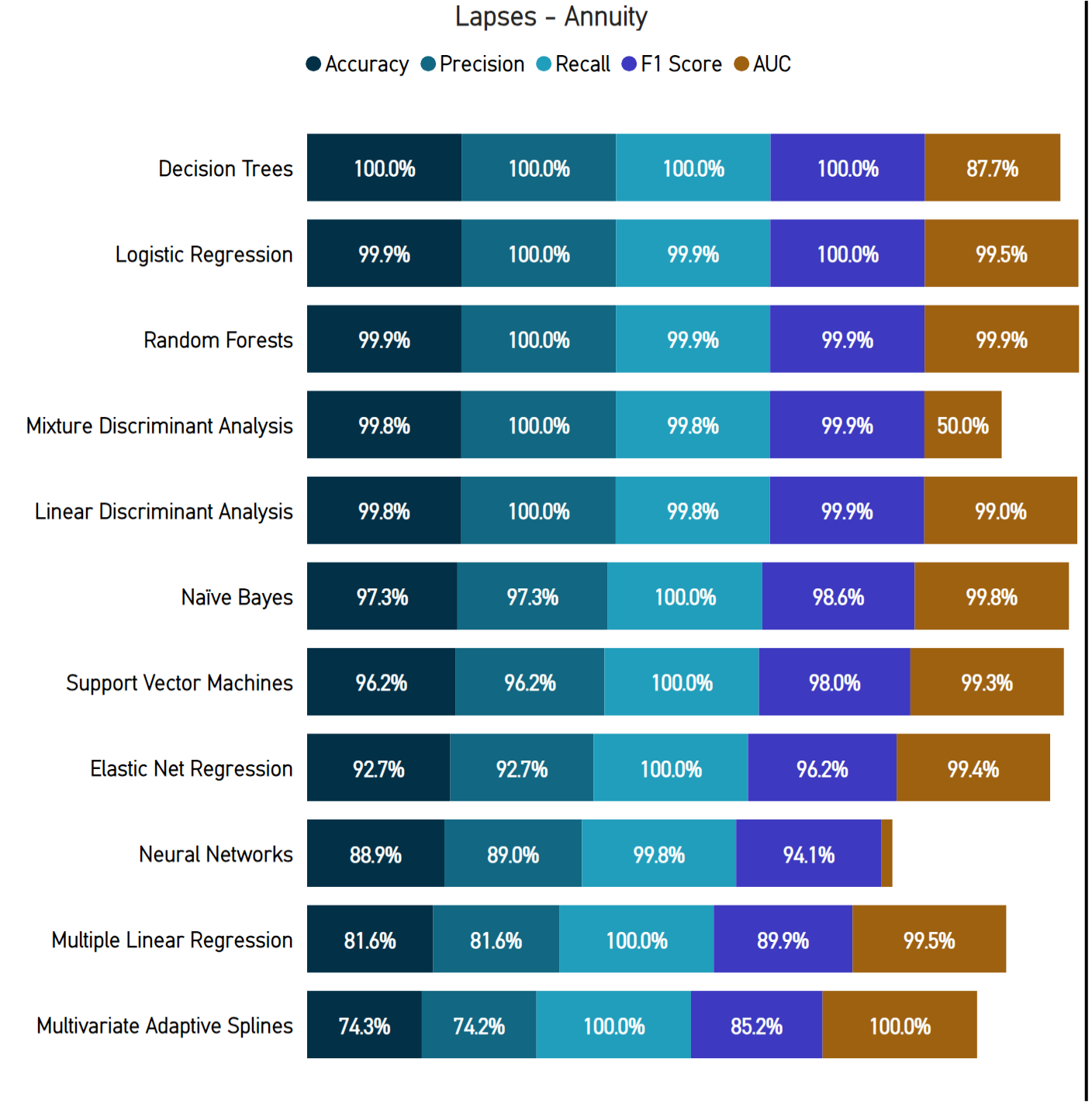
Results

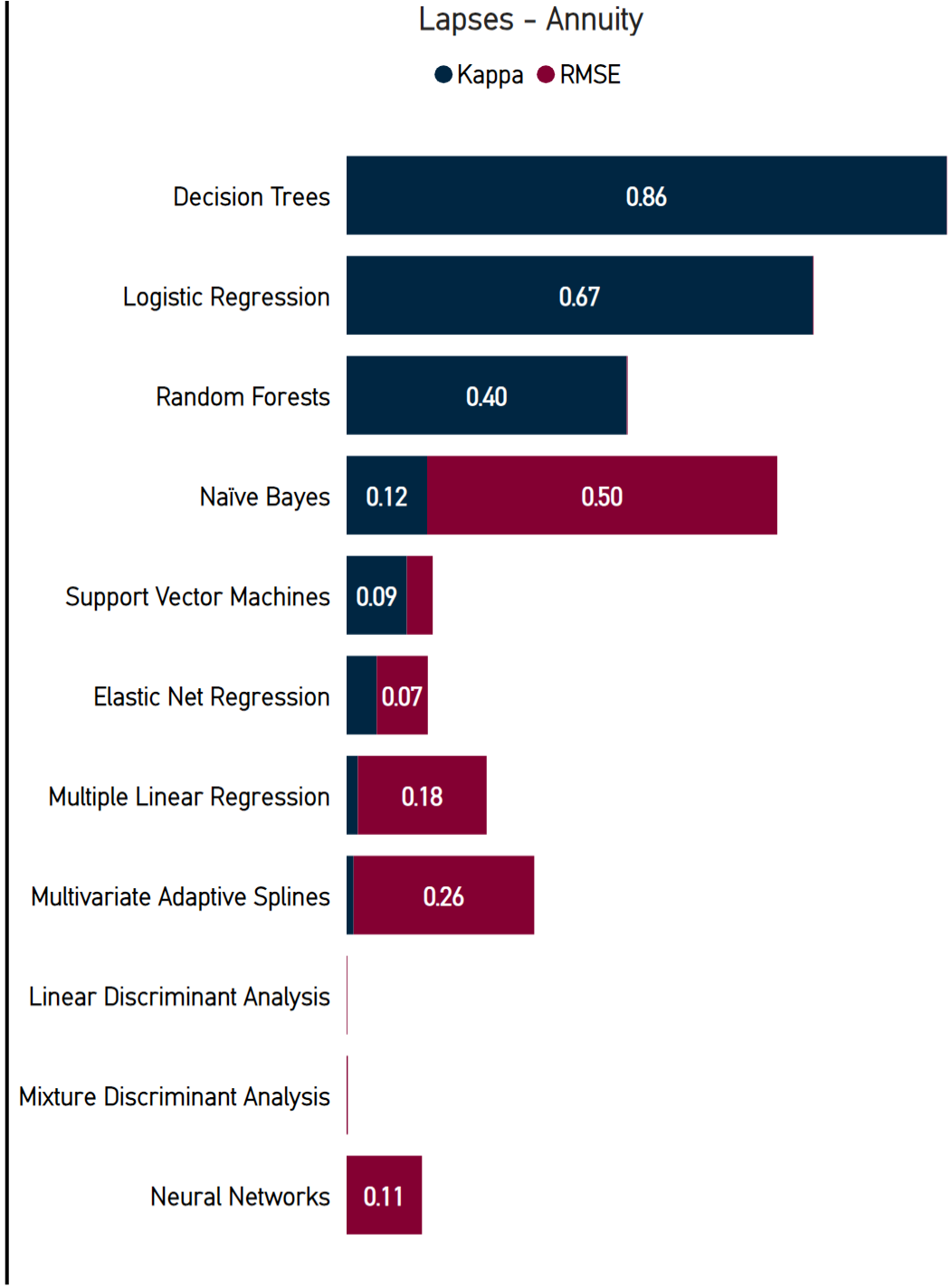

Lapses - Traditional Life
Accuracy
Precision
Recall
F1 Score
AUC
Random Forests
93.0%
98.5%
93.6%
96.0%
94.8%
Multivariate Adaptive Splines
92.3%
99.6%
92.0%
95.7%
90.3%
Decision Trees
90.4%
95.5%
93.4%
94.4%
83.9%
Multiple Linear Regression
88.2%
98.1%
89.1%
93.4%
87.5%
Suppor Vector Machines
88.2%
98.1%
89.1%
93.4%
85.7%
Elastic Ridge Regression
87.5%
97.8%
88.8%
93.1%
88.7%
Neural Networks
87.2%
93.3%
91.9%
92.6%
78.3%
Linear Discriminant Analysis
86.3%
92.1%
91.8%
92.0%
88.7%
Mixture Discriminant Analysis
85.3%
89.5%
93.0%
91.2%
75.2%
Logistic Regression
84.7%
98.9%
85.4%
91.7%
87.5%
Naïve Bayes
75.7%
74.2%
96.6%
83.9%
82.8%
Quadratic Discriminant Analysis
72.5%
70.4%
96.4%
81.4%
85.3%

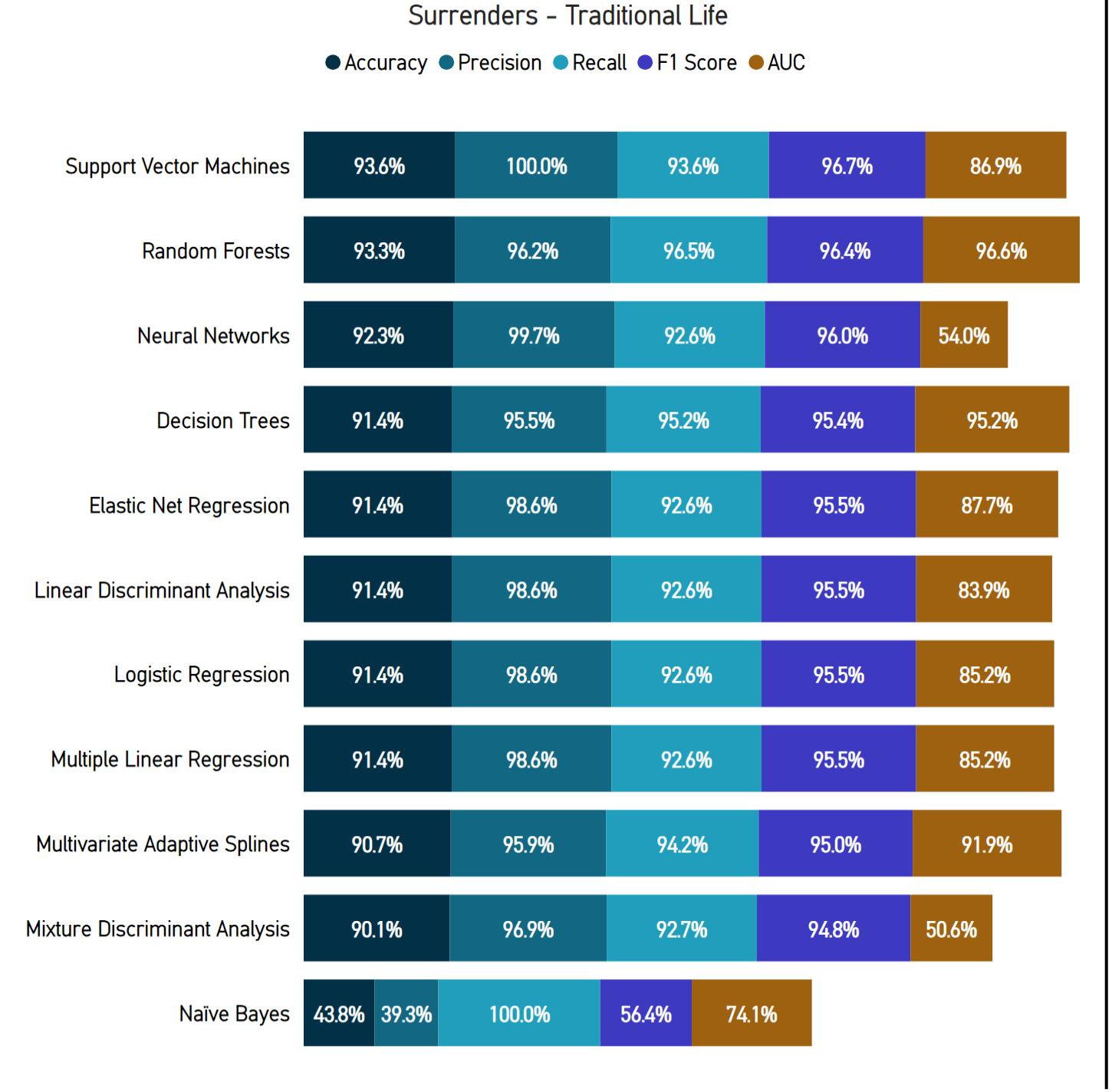

Lapses - Traditional Life
Kappa
RMSE
Random Forests
0.68
Multivariate Adaptive Splines
0.62
0.07
Decision Trees
0.60
Neural Networks
0.47
Mixture Discriminant Analysis
0.46
Linear Discriminant Analysis
0.45
Naïve Bayes
0.38
0.35
Multiple Linear Regression
0.38
0.09
Suppor Vector Machines
0.38
0.09
Elastic Ridge Regression
0.34
0.09
Quadratic Discriminant Analysis
0.34
0.23
Logistic Regression
0.13

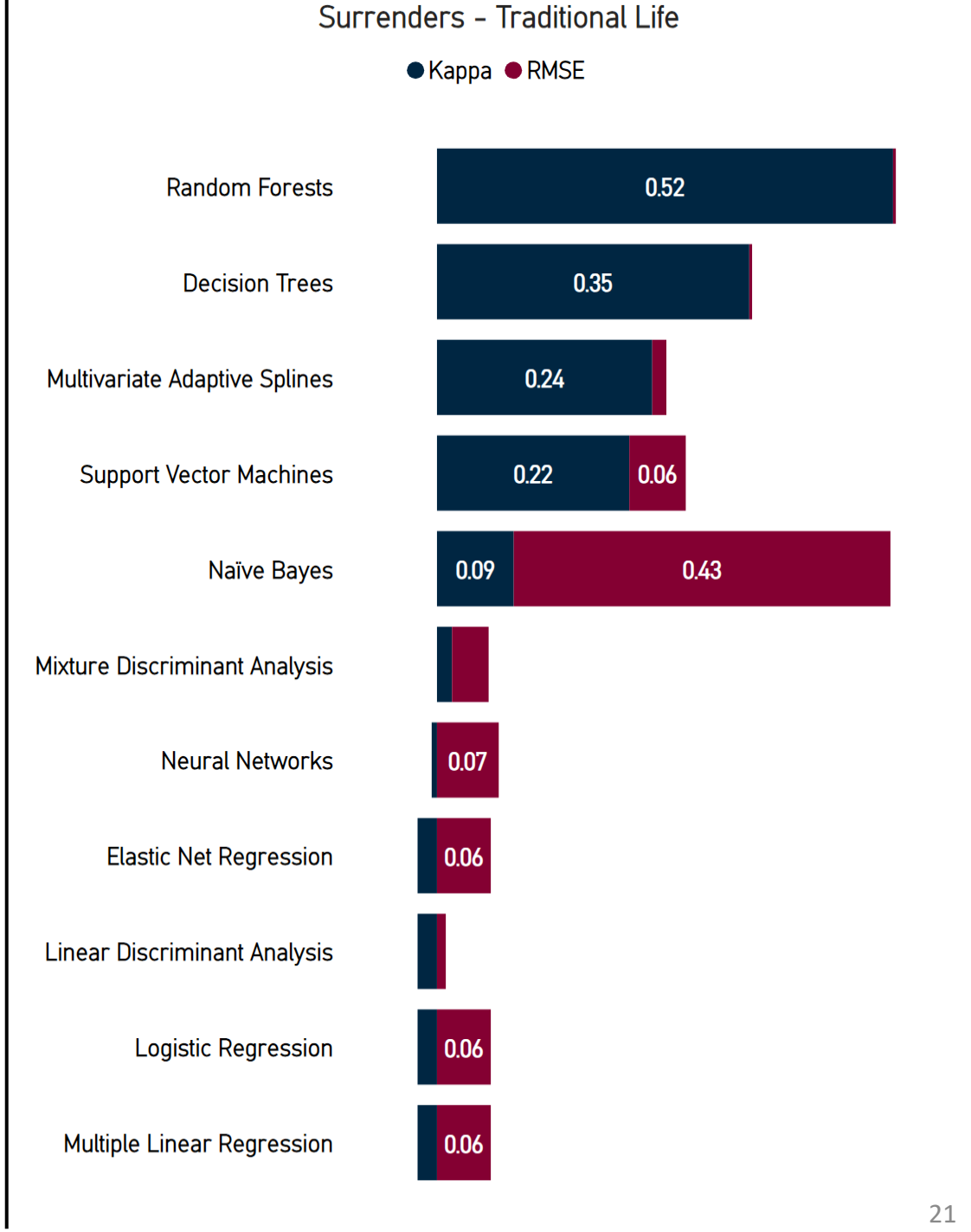

Lapses - Universal Life
Accuracy
Precision
Recall
F1 Score
AUC
Random Forests
89.6%
97.9%
90.8%
94.1%
89.4%
Multivariate Adaptive Splines
88.8%
98.3%
89.6%
93.8%
87.9%
Decision Trees
87.3%
96.8%
89.2%
92.9%
70.0%
Support Vector Machines
86.2%
99.6%
86.4%
92.5%
81.7%
Linear Discriminant Analysis
85.6%
96.4%
88.0%
92.0%
77.4%
Mixture Discriminant Analysis
85.6%
96.2%
88.0%
91.9%
59.9%
Neural Networks
85.5%
100.0%
85.5%
92.2%
50.3%
Logistic Regression
85.4%
98.5%
86.3%
92.0%
77.3%
Elastic Net Regression
85.4%
98.3%
86.4%
92.0%
77.2%
Multiple Linear Regression
85.4%
98.4%
86.4%
92.0%
77.3%
Quadratic Discriminant Analysis
81.2%
88.6%
89.3%
88.9%
76.5%
Naïve Bayes
80.5%
90.4%
87.3%
88.8%
69.0%

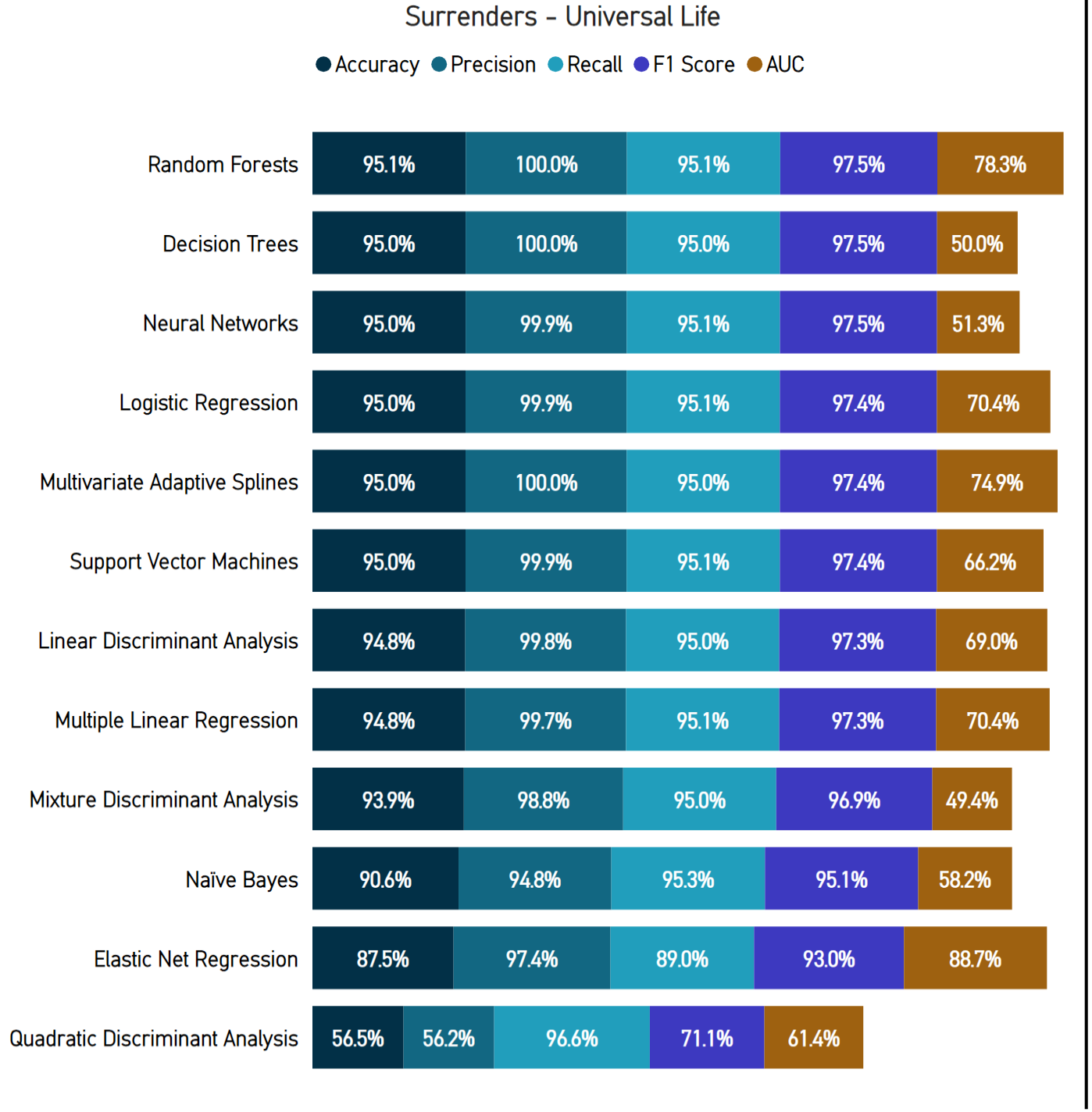

Lapses - Universal Life
Kappa
RMSE
Random Forests
0.50
0.06
Multivariate Adaptive Splines
0.42
0.08
Decision Trees
0.36
0.07
Quadratic Discriminant Analysis
0.26
Mixture Discriminant Analysis
0.25
0.08
Linear Discriminant Analysis
0.25
Naïve Bayes
0.14
0.35
Support Vector Machines
0.12
0.13
Elastic Net Regression
0.11
0.12
Multiple Linear Regression
0.11
0.12
Logistic Regression
0.11
0.12
Neural Networks
0.15

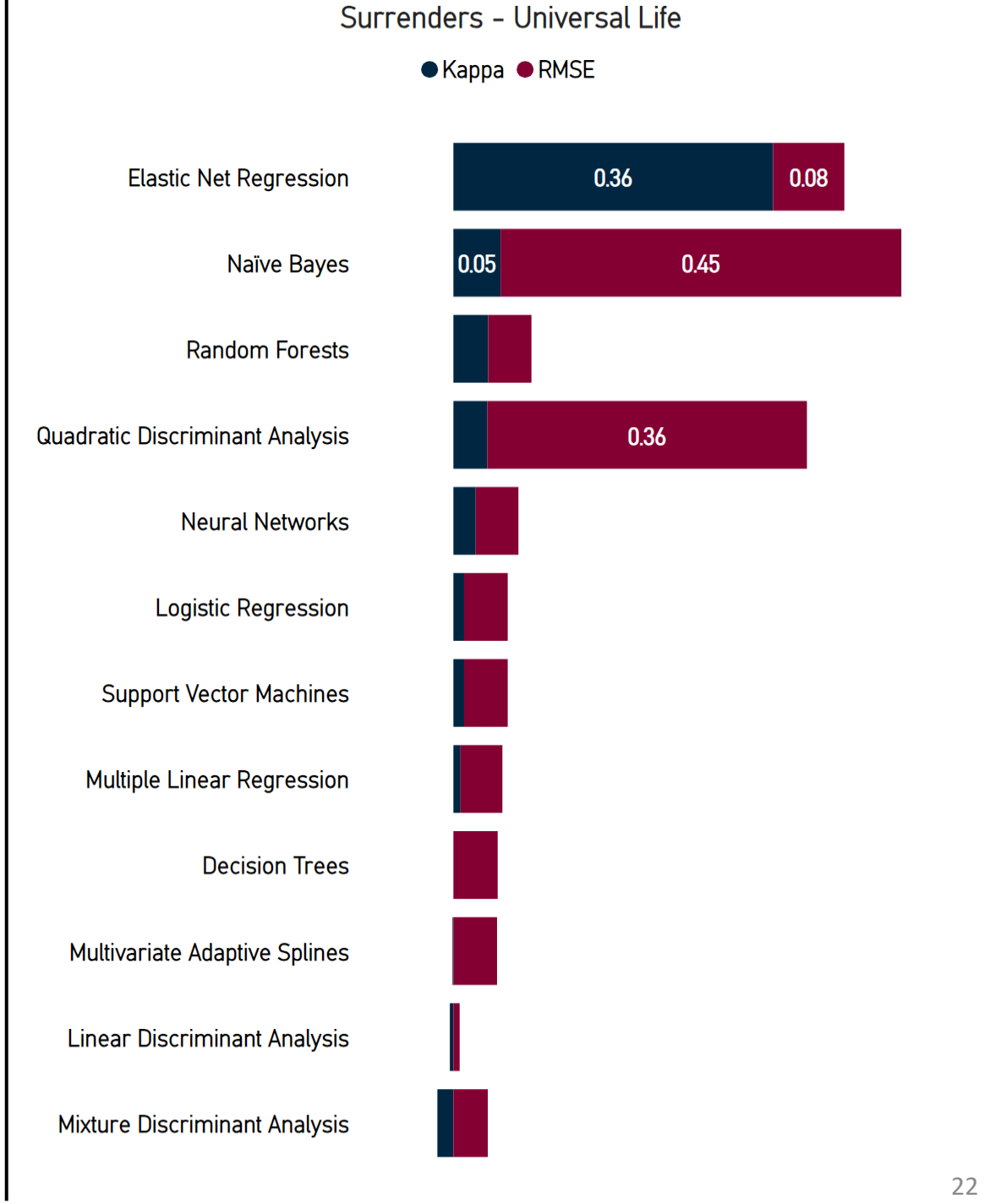

Lapses - Annuity
Accuracy
Precision
Recall
F1 Score
AUC
Decision Trees
100.0%
100.0%
100.0%
100.0%
87.7%
Logistic Regression
99.9%
100.0%
99.9%
100.0%
99.5%
Random Forests
99.9%
100.0%
99.9%
99.9%
99.9%
Mixture Discriminant Analysis
99.8%
100.0%
99.8%
99.9%
50.0%
Linear Discriminant Analysis
99.8%
100.0%
99.8%
99.9%
99.0%
Naïve Bayes
97.3%
97.3%
100.0%
98.6%
99.8%
Support Vector Machines
96.2%
96.2%
100.0%
98.0%
99.3%
Elastic Net Regression
92.7%
92.7%
100.0%
96.2%
99.4%
Neural Networks
88.9%
89.0%
99.8%
94.1%
Multiple Linear Regression
81.6%
81.6%
100.0%
89.9%
99.5%
Multivariate Adaptive Splines
74.3%
74.2%
100.0%
85.2%
100.0%

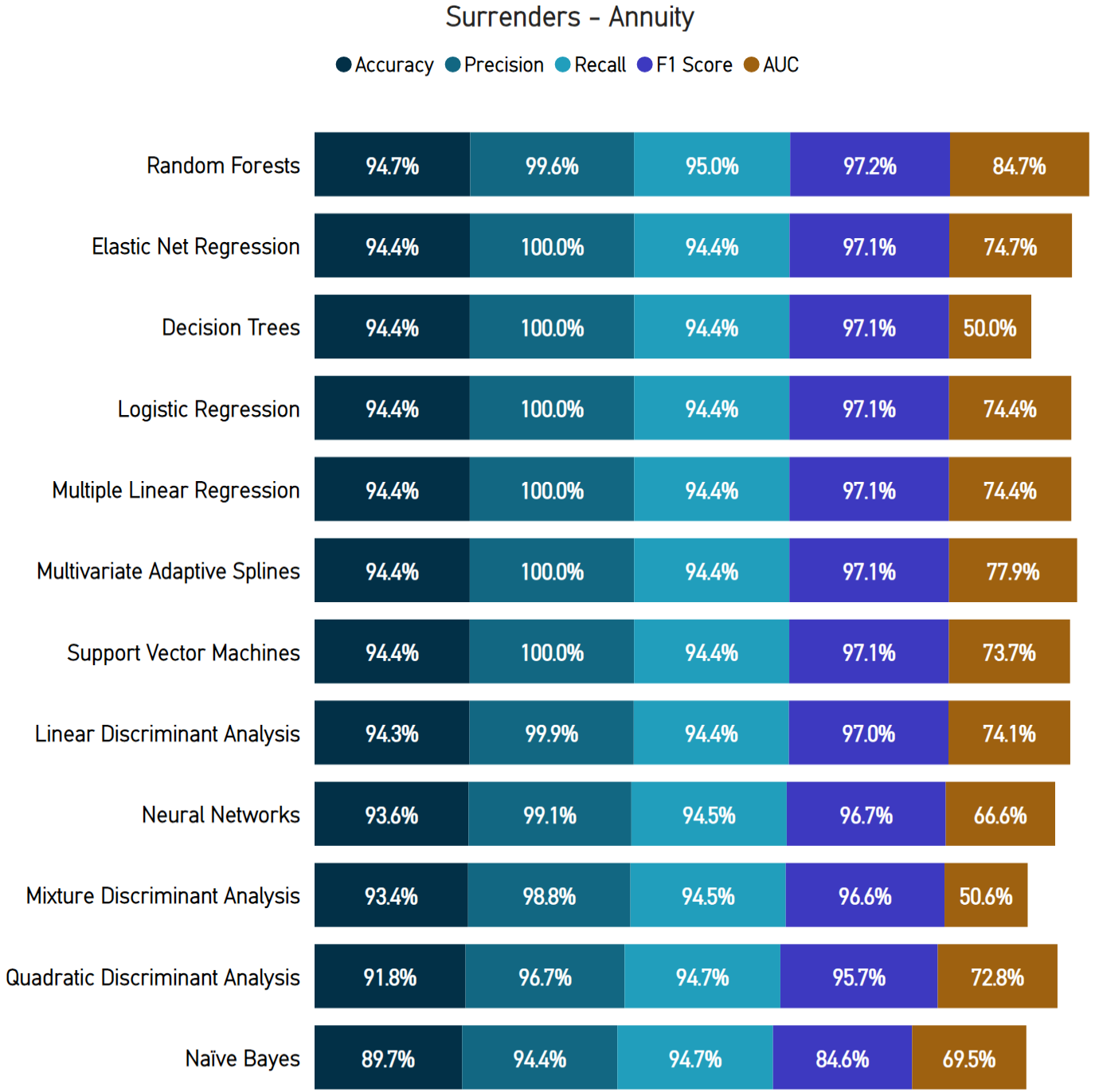

Lapses - Annuity
Kappa
RMSE
Decision Trees
0.86
Logistic Regression
0.67
Random Forests
0.40
Naïve Bayes
0.12
0.50
Support Vector Machines
0.09
Elastic Net Regression
0.07
Multiple Linear Regression
0.18
Multivariate Adaptive Splines
0.26
Linear Discriminant Analysis
Mixture Discriminant Analysis
Neural Networks
0.11

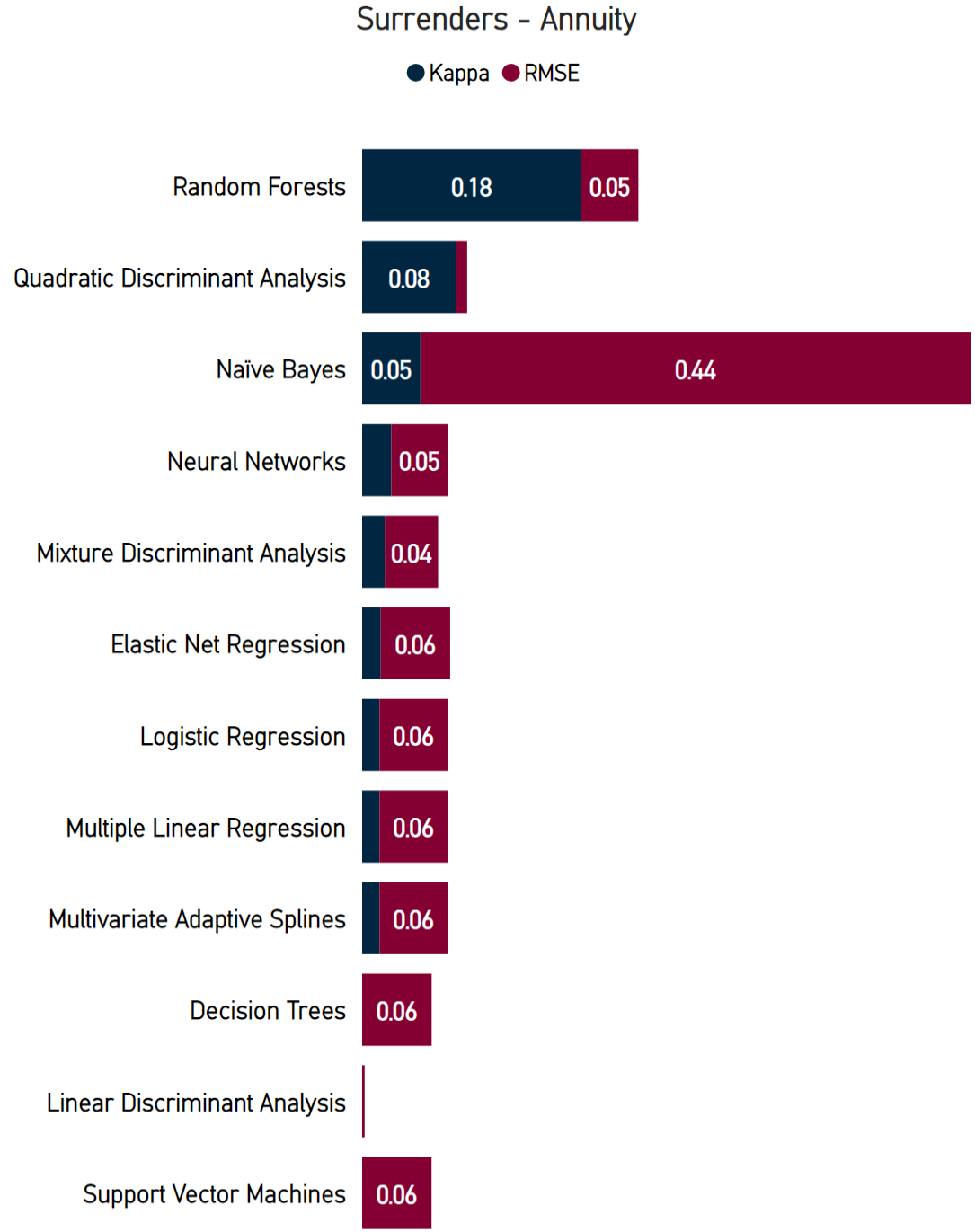

Surrenders - Traditional Life
Accuracy
Precision
Recall
F1 Score
AUC
Support Vector Machines
93.6%
100.0%
93.6%
96.7%
86.9%
Random Forests
93.3%
96.2%
96.5%
96.4%
96.6%
Neural Networks
92.3%
99.7%
92.6%
96.0%
54.0%
Decision Trees
91.4%
95.5%
95.2%
95.4%
95.2%
Elastic Net Regression
91.4%
98.6%
92.6%
95.5%
87.7%
Linear Discriminant Analysis
91.4%
98.6%
92.6%
95.5%
83.9%
Logistic Regression
91.4%
98.6%
92.6%
95.5%
85.2%
Multiple Linear Regression
91.4%
98.6%
92.6%
95.5%
85.2%
Multivariate Adaptive Splines
90.7%
95.9%
94.2%
95.0%
91.9%
Mixture Discriminant Analysis
90.1%
96.9%
92.7%
94.8%
50.6%
Naïve Bayes
43.8%
39.3%
100.0%
56.4%
74.1%


Surrenders - Traditional Life
Kappa
RMSE
Random Forests
0.52
Decision Trees
0.35
Multivariate Adaptive Splines
0.24
Support Vector Machines
0.22
0.06
Naïve Bayes
0.09
0.43
Mixture Discriminant Analysis
Neural Networks
0.07
Elastic Net Regression
0.06
Linear Discriminant Analysis
Logistic Regression
0.06
Multiple Linear Regression
0.06


Surrenders - Universal Life
Accuracy
Precision
Recall
F1 Score
AUC
Random Forests
95.1%
100.0%
95.1%
97.5%
78.3%
Decision Trees
95.0%
100.0%
95.0%
97.5%
50.0%
Neural Networks
95.0%
99.9%
95.1%
97.5%
51.3%
Logistic Regression
95.0%
99.9%
95.1%
97.4%
70.4%
Multivariate Adaptive Splines
95.0%
100.0%
95.0%
97.4%
74.9%
Support Vector Machines
95.0%
99.9%
95.1%
97.4%
66.2%
Linear Discriminant Analysis
94.8%
99.8%
95.0%
97.3%
69.0%
Multiple Linear Regression
94.8%
99.7%
95.1%
97.3%
70.4%
Mixture Discriminant Analysis
93.9%
98.8%
95.0%
96.9%
49.4%
Naïve Bayes
90.6%
94.8%
95.3%
95.1%
58.2%
Elastic Net Regression
87.5%
97.4%
89.0%
93.0%
88.7%
Quadratic Discriminant Analysis
56.5%
56.2%
96.6%
71.1%
61.4%


Surrenders - Universal Life
Kappa
RMSE
Elastic Net Regression
0.36
0.08
Naïve Bayes
0.05
0.45
Random Forests
Quadratic Discriminant Analysis
0.36
Neural Networks
Logistic Regression
Support Vector Machines
Multiple Linear Regression
Decision Trees
Multivariate Adaptive Splines
Linear Discriminant Analysis
Mixture Discriminant Analysis


Surrenders - Annuity
Accuracy
Precision
Recall
F1 Score
AUC
Random Forests
94.7%
99.6%
95.0%
97.2%
84.7%
Elastic Net Regression
94.4%
100.0%
94.4%
97.1%
74.7%
Decision Trees
94.4%
100.0%
94.4%
97.1%
50.0%
Logistic Regression
94.4%
100.0%
94.4%
97.1%
74.4%
Multiple Linear Regression
94.4%
100.0%
94.4%
97.1%
74.4%
Multivariate Adaptive Splines
94.4%
100.0%
94.4%
97.1%
77.9%
Support Vector Machines
94.4%
100.0%
94.4%
97.1%
73.7%
Linear Discriminant Analysis
94.3%
99.9%
94.4%
97.0%
74.1%
Neural Networks
93.6%
99.1%
94.5%
96.7%
66.6%
Mixture Discriminant Analysis
93.4%
98.8%
94.5%
96.6%
50.6%
Quadratic Discriminant Analysis
91.8%
96.7%
94.7%
95.7%
72.8%
Naïve Bayes
89.7%
94.4%
94.7%
84.6%
69.5%


Surrenders - Annuity
Kappa
RMSE
Random Forests
0.18
0.05
Quadratic Discriminant Analysis
0.08
Naïve Bayes
0.05
0.44
Neural Networks
0.05
Mixture Discriminant Analysis
0.04
Elastic Net Regression
0.06
Logistic Regression
0.06
Multiple Linear Regression
0.06
Multivariate Adaptive Splines
0.06
Decision Trees
0.06
Linear Discriminant Analysis
Support Vector Machines
0.06

Discussion

Model's Performance

Though many models performed well, the following models were the best in each category:

Lapses

Place	Traditional Life	Universal Life	Annuity
1st Place	Random Forests	Random Forests	Logistic Regression
2nd Place	Multivariate Adaptive Regression Splines	Multivariate Adaptive Regression Splines	Random Forests
3rd Place	Decision Trees	Support Vector Machines	Decision Trees

Continued…

Surrenders

Place	Traditional Life	Universal Life	Annuity
1st Place	Random Forests	Random Forests	Random Forests
2nd Place	Decision Trees	Elastic Net Regression	Neural Networks
3rd Place	Support Vector Machines	Multivariate Adaptive Regression Splines	Quadratic Discriminant Analysis

Conclusion and Future Research

Conclusion

- On average, the Random Forests model performed the best

Future Research

- Investigate the use of macroeconomic variables, in addition to those used in this research

- Explore other modeling methods

Thank You!!