# Predicting Caravan Insurance Policy Ownership

Erik Carrion, Juna Iafelice, John Makhijani

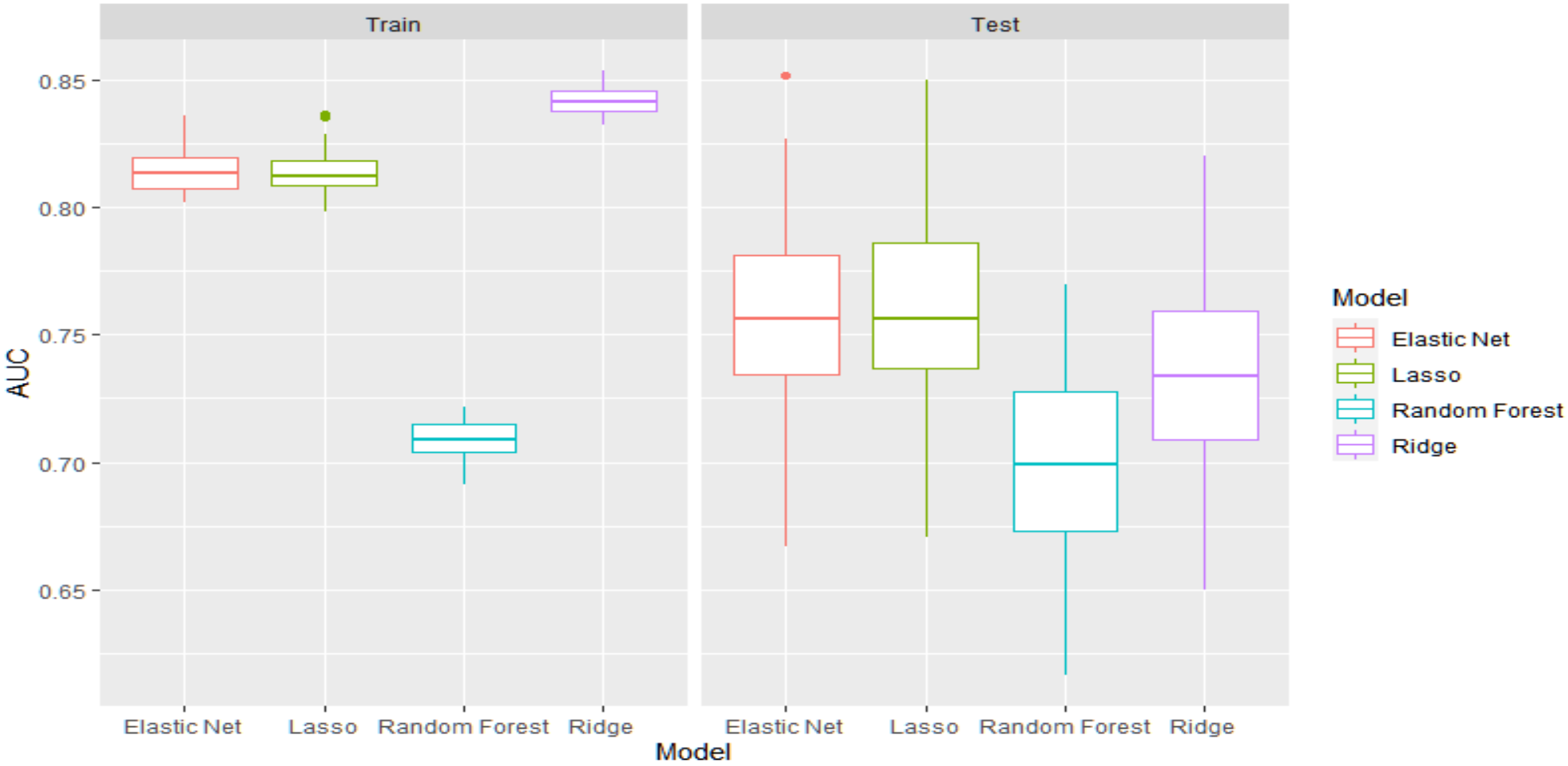
December 13, 2021

# Caravan Dataset Description

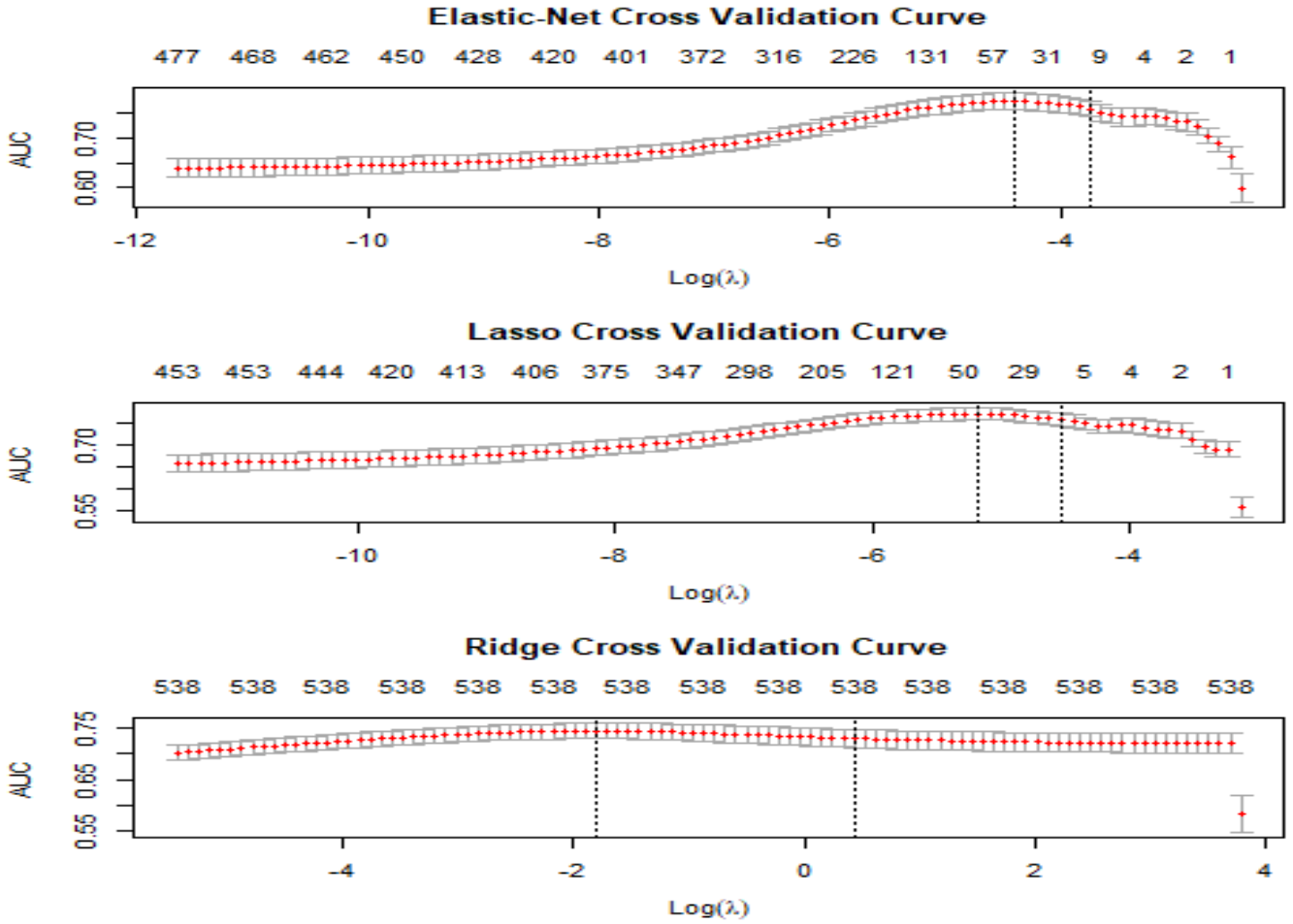
**Goal** - Predict who would be interested in buying a caravan insurance policy

- Data Breakdown
  - Sample Size (n) – 5822 | Predictors (p) – 86
  - Predictors include product usage data and socio-demographic data derived from zip area codes
  - Response Variable (Binary) – Caravan insurance policy owner
  - Predictors – Multinomial with a varying number levels 2-40
  - Data Source – http://kdd.ics.uci.edu/databases/tic/tic.html

# $AUC_{test}$ and $AUC_{train}$ by Method



# Cross Validation Curves




| CV TIME | |
|---|---|
| ELASTIC NET | 112 secs |
| LASSO | 134 secs |
| RIDGE | 64 secs |

# Model Performance / Time Tradeoff

| | Test AUC (median) | Time |
|---|---|---|
| ELASTIC NET | 0.76 | 129 secs |
| LASSO | 0.76 | 146 secs |
| RIDGE | 0.73 | 57 secs |
| RANDOM FOREST | 0.70 | 155 secs |


Standardized Elastic Net Coefficients
Coefficient
0.4
0.2
0.0
Standardized Lasso Coefficients
Coefficient
0.6
0.4
0.2
0.0
Standardized Ridge Coefficients
Coefficient
0.10
0.05
0.00
-0.05
Random Forrest Variable Importance
Importance
12.5
10.0
7.5
5.0
2.5
0.0

# Conclusion

- **Important Features:**

| | |
|---|---|
| PPERSAUT6 | Contribution Car Policies, L6 |
| PBRAND4 | Contribution Fire Policies, 3 policies |
| APLEZIER1 | No. of Boat Policies, 1 policy |
| PBRAND3 | Contribution Fire Policies, 2 policies |
| MKOOPKLA7 | Purchasing Power Class, level 7 |

- **Time vs Performance:** Elastic Net or Lasso are both parsimonious and give us approximately the same median AUC.

- **Best Model:** Elastic Net