MODELS FOR CLASSIFYING POLICYHOLDER LAPSE

KARTHIK TUMULURU

Actuarial Associate, MetLife

AGENDA

- Introduction
- Research and data
- Preliminary analysis
- Model testing
- Results and use
- Conclusion

THE ISSUE

- Higher lapses in life insurance may result in significant loss of premiums
- A lapse can be voluntary or involuntary, wildly varying reasons possible
- Finding key factors in common however helps act proactively on people who might lapse

RESULTS - A QUICK SNAPSHOT

- Achieved an 80% accuracy for classifying first-year lapse, and a 90% accuracy for lapse by 3rd year
- Certain models performed better in short term, others better in longer term
- Similarly, certain variables were more important in shorter-term, others in longer-term.

ACTUARIAL LITERATURE REVIEW

- Economic status of household affects policyholder behaviour
- Products with better rates might result in more policyholders wanting to replace their existing policy
- Other variables such as commission structure or policyholders' education play a role as well

DATA COLLECTION AND PREPARATION

- Updated datasets taken several months apart for analysis and model testing
- In-force and lapsed policyholders for Ordinary Life taken, ages 18-65
- Variables like modal premium, age, cover amount, agent code present
- Variables extrapolated include no. of payments, no. of riders, no. of dependents, years left to maturity

PRELIMINARY DATA ANALYSIS

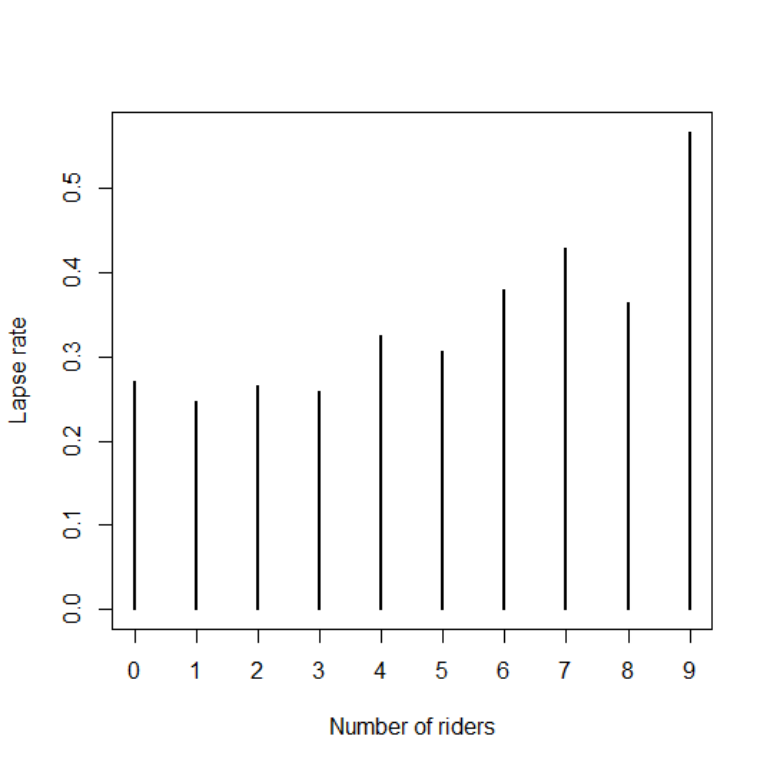
- Lapse defined as paying the last premium within 3 years from date of issue
- No significant differences/trends found between lapse rates from datasets taken at different months

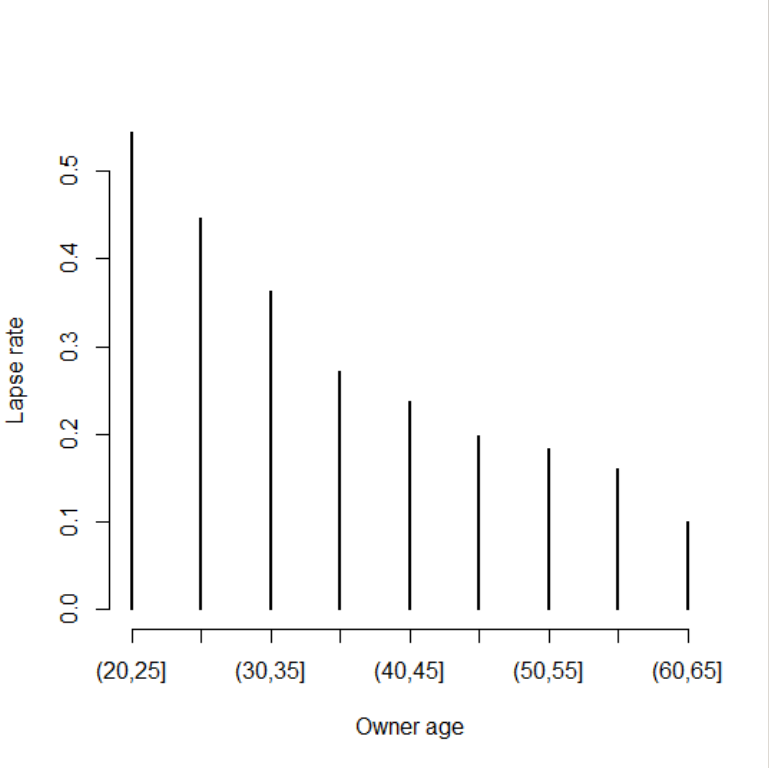
PRELIMINARY DATA ANALYSIS

- Relations between lapse rate and different variables including:
  - Cohort
  - Product type
  - Age
  - Agency selling the policy,

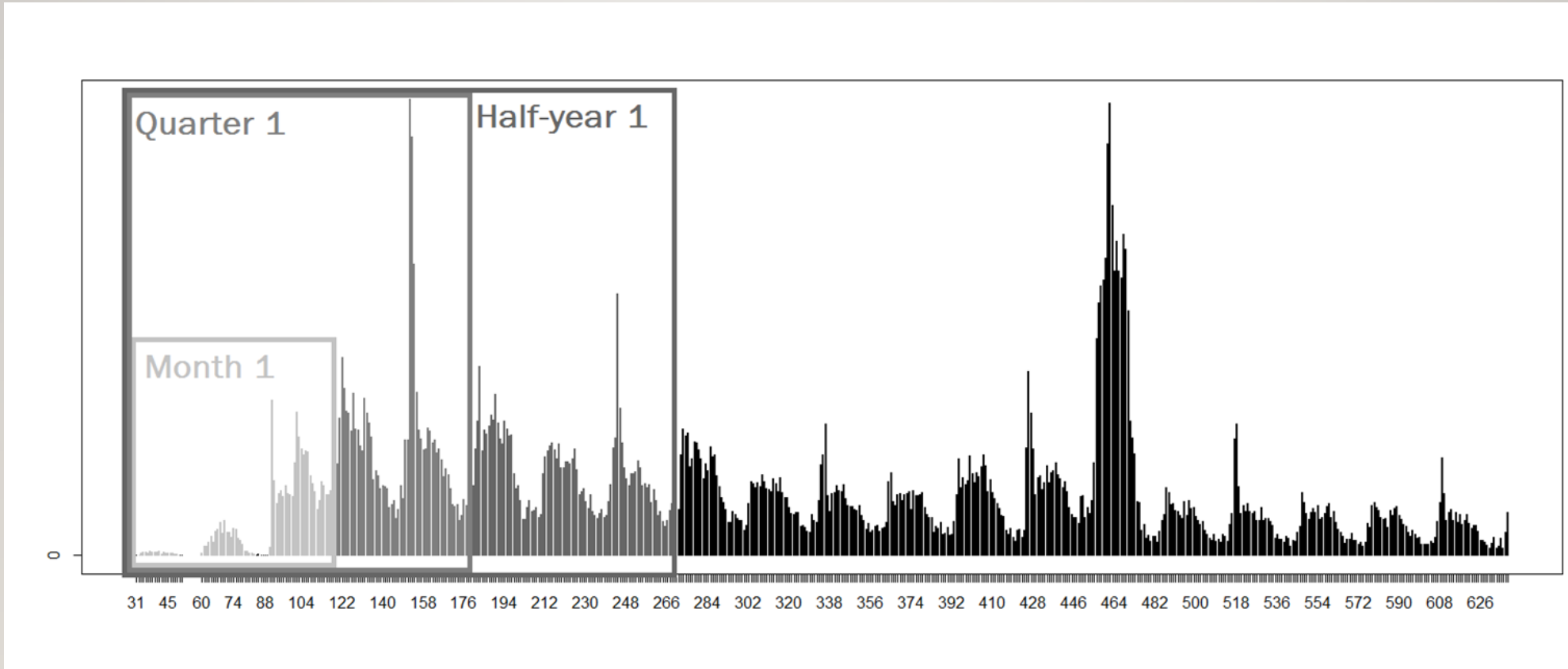
and distributions of certain variables like number of payments were studied, as was seasonality in lapses

PRELIMINARY DATA ANALYSIS





PRELIMINARY DATA ANALYSIS



PICKING THE RIGHT MODEL

- Right model for the right task
- Complexity
- Interpretability
- Variables used

PICKING THE RIGHT MODEL

GLM

- Multivariate linear regression
- (+) Efficient to train
- (-) Only good with linearity

Random Forest

- A 'forest' of decision trees
- (+) Robust
- (-) Black box

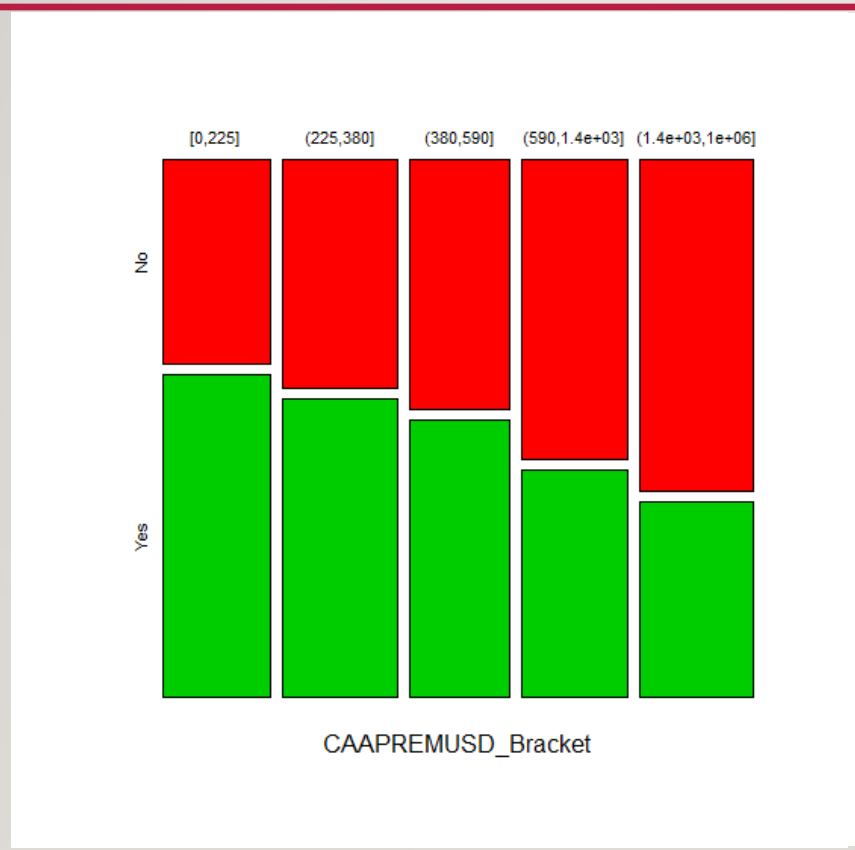
Naïve Bayes

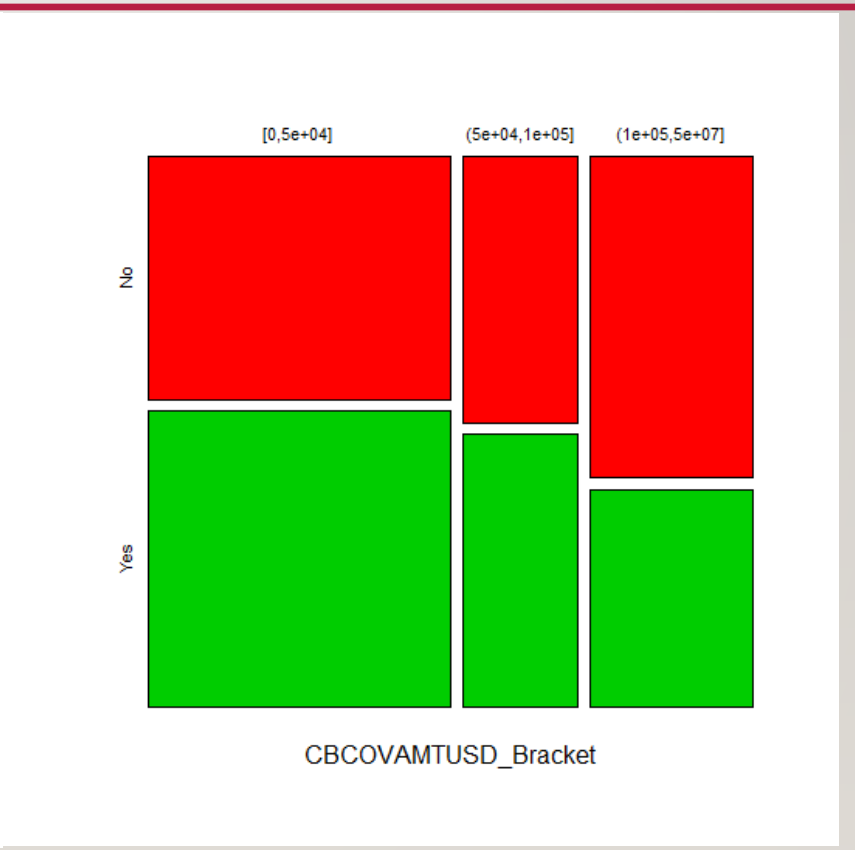
- Based on Bayes Theorem
- (+) Easy to understand
- (-) Assumes independence

C5.0

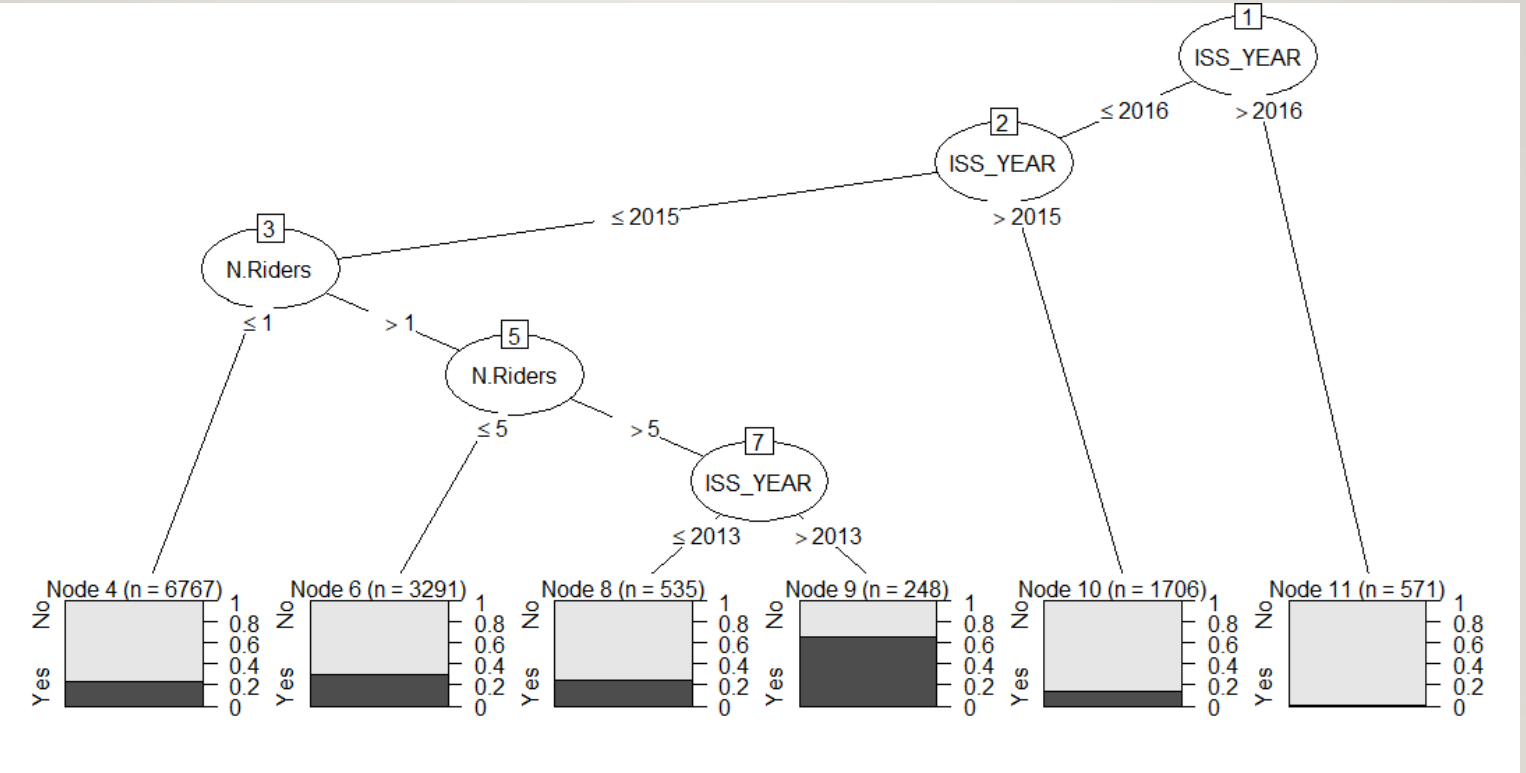
- (+) Rules are simple to interpret
- (-) Small data variation -> different tree

Naive Bayes - sample graphs





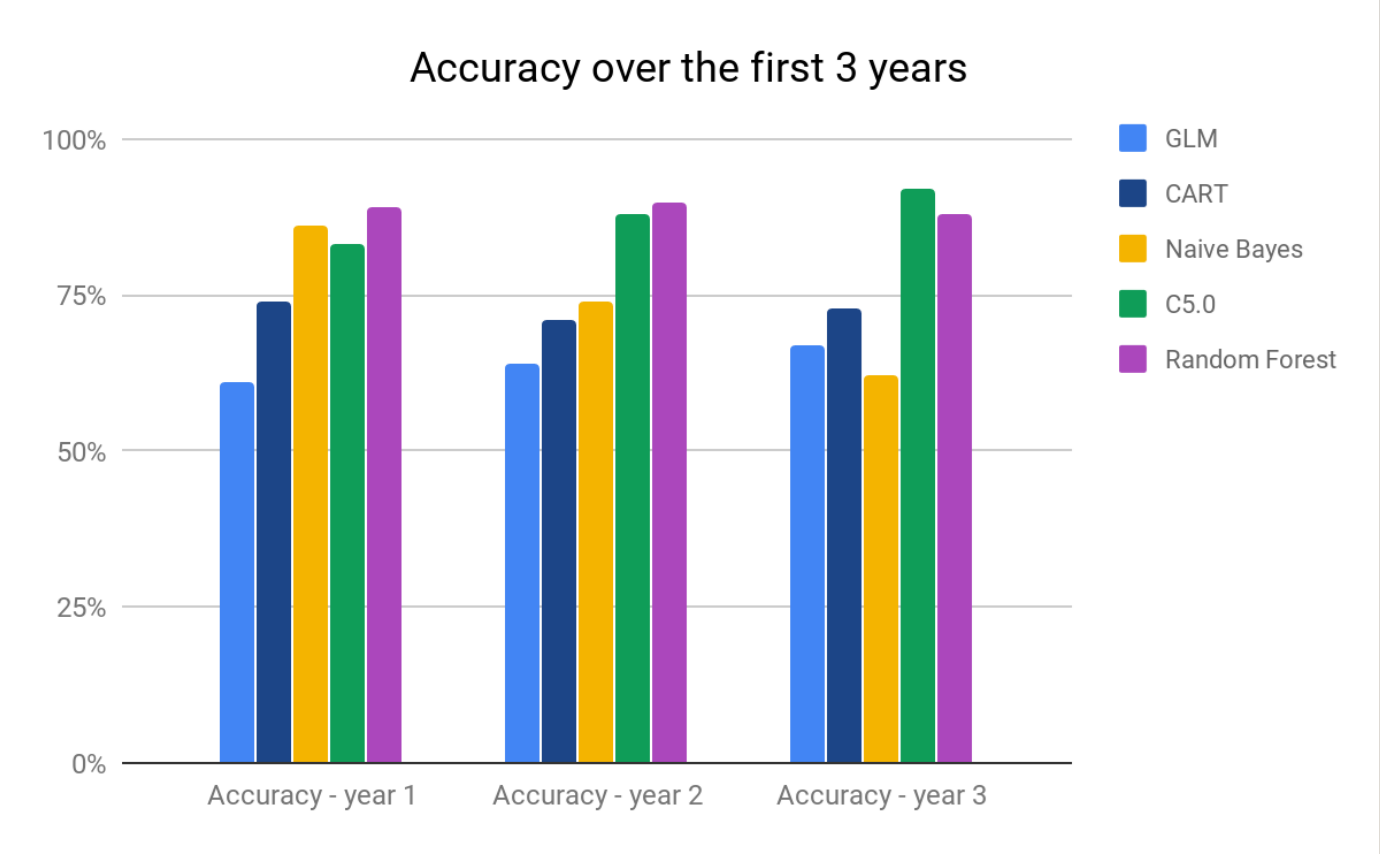
C5.0 - a sample graph



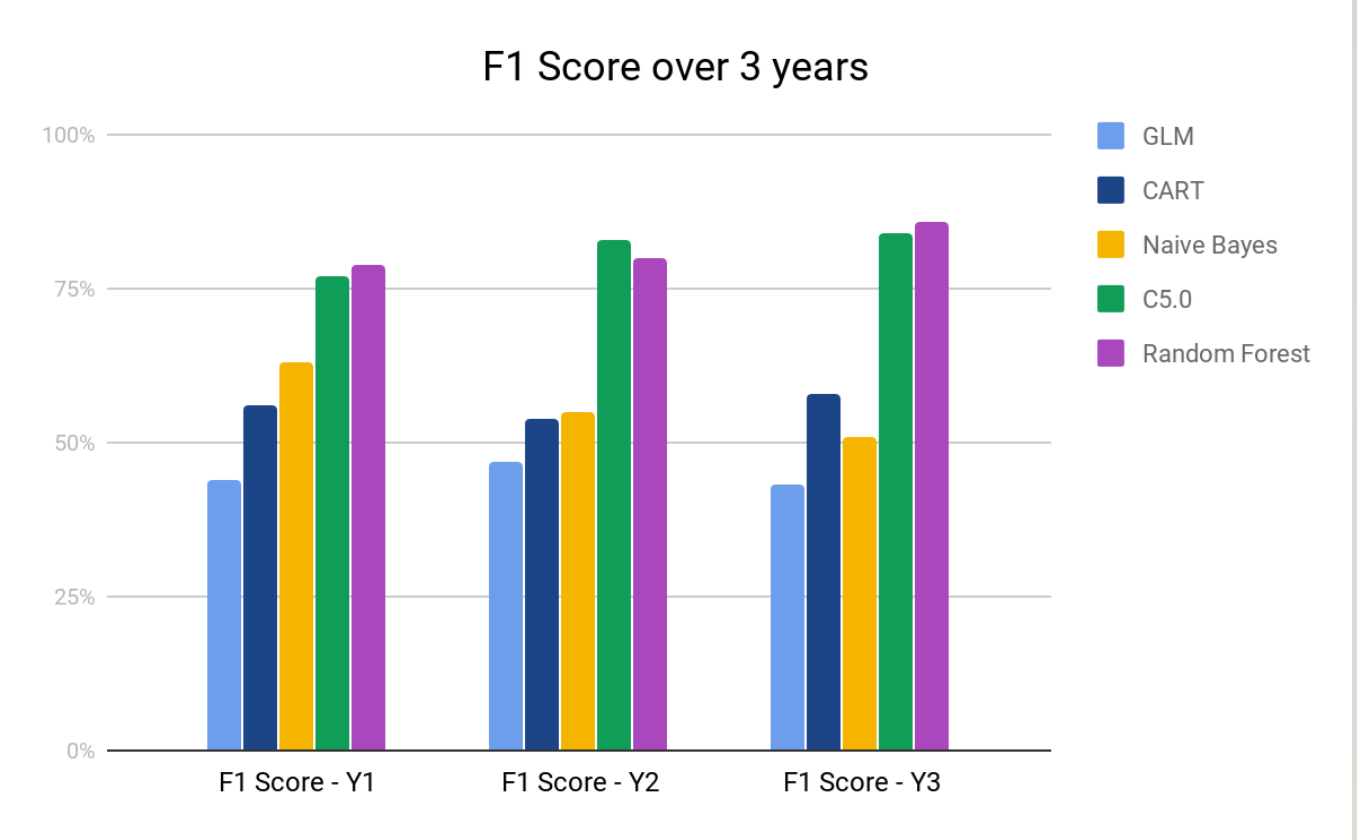
TRAINING AND TESTING

- GLM used as benchmark, other models like CART, C5.0, Naïve Bayes tested
- Models tested on 2 datasets in 2018 and 2019
- Accuracy and F1-Score used as 2 performance metrics

RESULTS


Accuracy over the first 3 years
100%
75%
50%
25%
0%
Accuracy - year 1
Accuracy - year 2
Accuracy - year 3
GLM
CART
Naive Bayes
C5.0
Random Forest

RESULTS


F1 Score over 3 years
100%
75%
50%
25%
0%
F1 Score - Y1
F1 Score - Y2
F1 Score - Y3
GLM
CART
Naive Bayes
C5.0
Random Forest

USING THE RESULTS

Name	Policy #	Age Bracket	Gender	Premium Size	Lapse Risk	Priority	Cross-Selling Opportunities
John Doe	271828182	30-40	M	20000-50000	Medium	High	Savings Plan 5
Jane Doe	987654321	50-60	F	50000-100000	Low	Medium	Pension Annuity
Bob Ross	123456789	20-30	M	2000-5000	High	Medium	5-Year Term Life
Fred Rogers	314159265	20-30	M	10000-20000	Low	Low	None

USING THE RESULTS

Agent	Product	Prob_Y1	Prob_Y2	Prob_Y3
Agent 1	Product 1	0.89	0.34	0.23
Agent 1	Product 2	0.74	0.34	0.11
Agent 1	Product 3	0.82	0.22	0.05
Agent 2	Product 4	0.93	0.86	0.23
Agent 3	Product 4	0.88	0.78	0.13
Agent 4	Product 4	0.91	0.85	0.31
Agent 5	Product 1	0.73	0.86	0.43
Agent 5	Product 2	0.46	0.44	0.13
Agent 6	Product 2	0.74	0.85	0.88

CONCLUSION

- Certain classification models can be used to predict whether a policyholder is likely to lapse in a given period of time
- The results of these models can be used to proactively sort customers with missed payments, in order of priority. Ideally any products matching the customer profile (using a product recommendation system for example) could be used to cross-sell.
- These results can also be used to retrospectively analyze which agents/products perform well or poorly in terms of persistency

Q&A