



AUTO INSURANCE ENTERPRISE

MAGUETTE PAYE

NUPUR MITTAL

PASCAL NGUYEN TANG

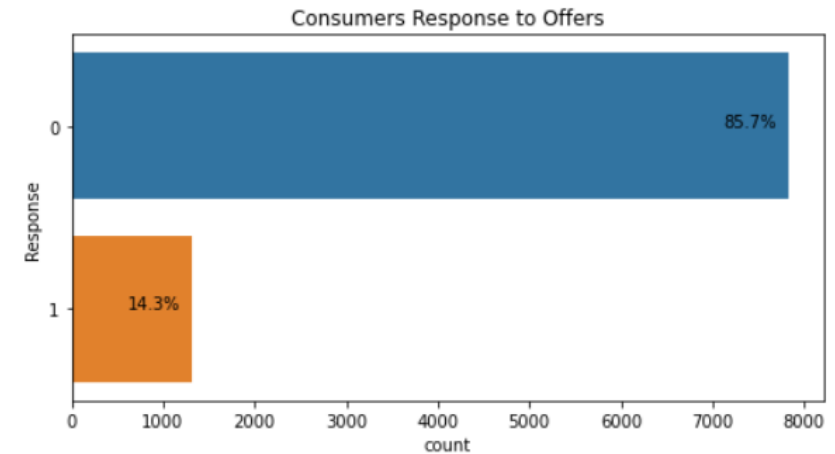
SHIVANGI SONI

VIVEK SAAHIL

REPOSITORY_AUTO_INSURANCE_ENTERPRISE

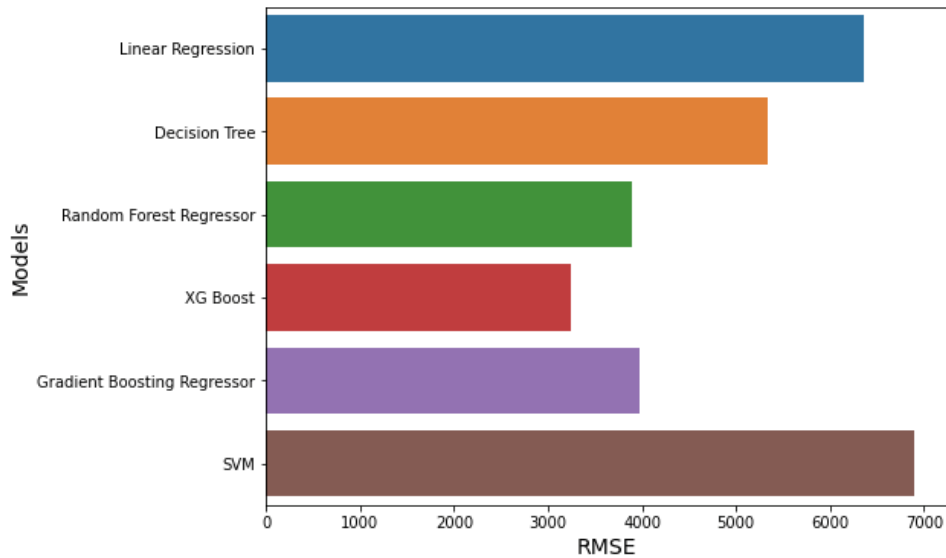
CONTEXT

- Automotive Insurance Dataset, Synthetic Data – IBM Watson
- Policies of customers expiring within next two months
- All customers were offered a renewal offer
- **Major Pain Point**
 - 14% of the customers accepted the renewal offer
 - Understanding the lower acceptance rate
- Our solution
 - Regression – Predicting the Customer Lifetime Value
 - Take targeted actions to increase profitable customer response, retention, and growth
 - $CLV > \text{Cost of acquisition and retention}$
 - Classification – Determining whether the customer would respond to the offer
 - Understanding which customers would respond negatively and target those

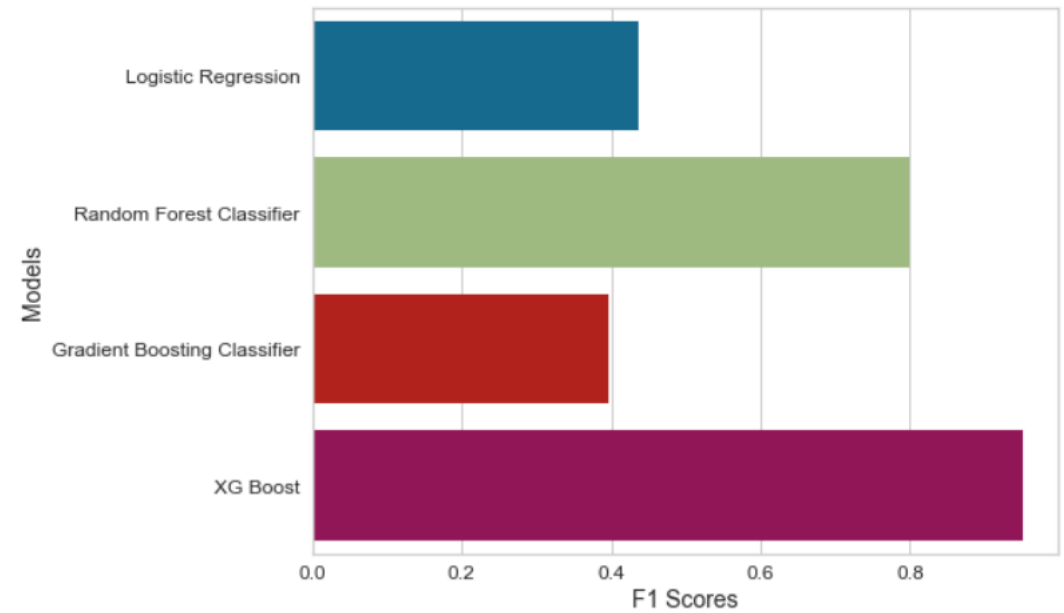


RESULTS

RMSE- Regression



F1 Scores - Classification



XGBOOST performed best for both Regression and Classification tasks, as shown by the RMSE and F1 Scores.

F1 Scores were used as performance metrics as opposed to precision, accuracy and recall because it is better suited for imbalanced data sets.

XGBOOST had the lowest **RMSE of \$3,250** and the highest **F1 score of 0.9508**.

RESULTS AND CAUSAL INFERENCE

Causal Inference



Treatment Variable	Outcome	P-value	Causal Estimate
<u>Employment Status</u>	<u>Response</u>	<u>2.06e-185</u>	<u>0.30</u>
Employment Status	Customer Lifetime Value	0.55	-231.57
<u>Renewed Offer 2¹</u>	<u>Response</u>	<u>1.91e-72</u>	<u>0.14</u>
<u>Renewed Offer 2¹</u>	<u>Customer Lifetime Value</u>	<u>2.79e-05</u>	<u>-600.89</u>
Coverage ²	Response	0.37	0.012
Coverage ²	Customer Lifetime Value	0.10	-406.53
Income ³	Response	0.536	-0.0086
Income ³	Customer Lifetime Value	0.36	-240.85

¹ Offer 2 was found to be the most popular offer

² Coverage was either listed as Premium or not

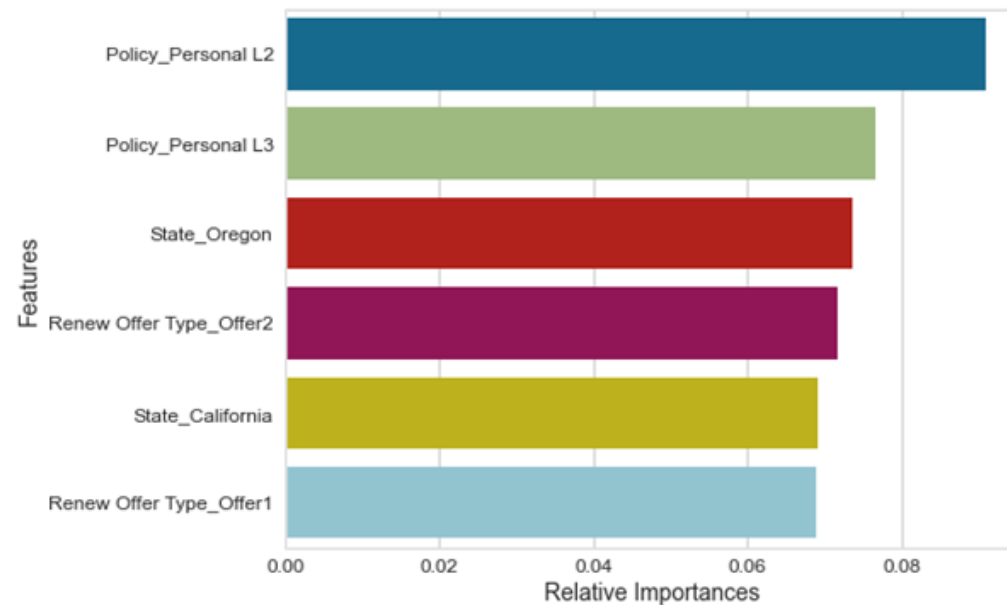
³ Income is listed as true if above average

THREATS TO VALIDITY

- **Threats due to assumptions** !
 - “No” in *Response* means *Customer Lifetime Value* is stopped.
- **Uncertainties and Risks** 
 - There may be a relationship between low amounts of *Customer Lifetime Value* with the *Response* value “No”, which can skew our predictions.
- **Data Quality Issue Risks** 
 - The lack of data dictionaries made it difficult to make business interpretations.

EXPLAINABILITY AND FEATURE IMPORTANCE

Feature Importance



Policies were divided into multiple categories, which were themselves divided into 3 partitions: L1, L2 and L3. The Offer type for the renewal of a given policy was divided into 4 partitions.

The graph shows the most impactful features on the classification model.

BUSINESS IMPLICATIONS

- Targeting consumers with high CLV who were predicted to respond negatively to offers
 - Providing other offer options to gauge their interest
- Offering Offer 2 to more consumers
- Causal Inference Recommendation
 - Focus on developing similar offers to Offer 2
 - Focusing on volume

LESSONS LEARNT AND NEXT STEPS

Lessons learnt

- Data is not enough to answer the 'why' of the problem
- Data overlay maybe required to determine a strategy
- Need to SMEs to understand the features/data better

Next Steps (from company's perspective)

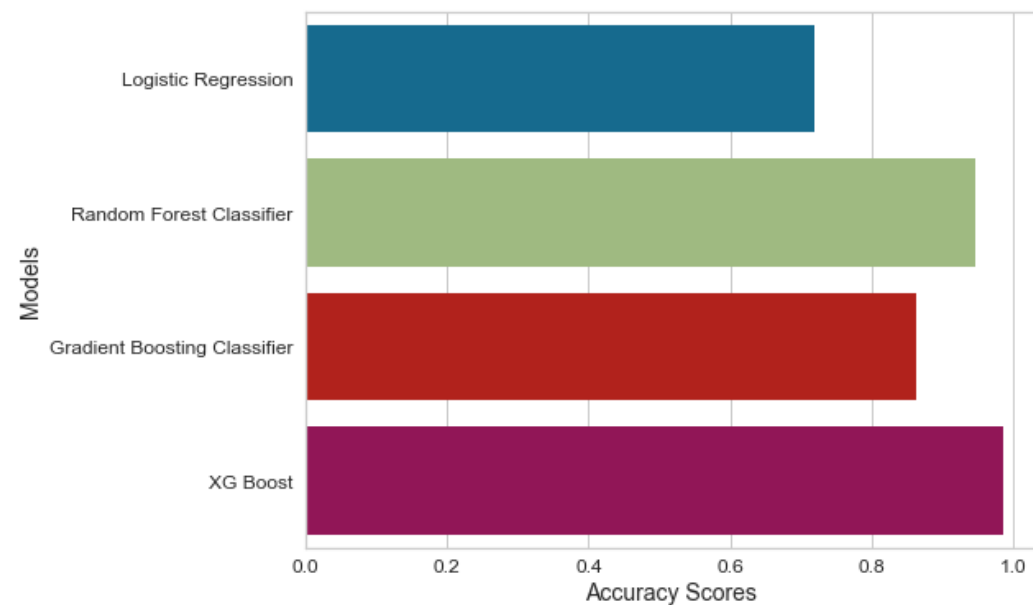
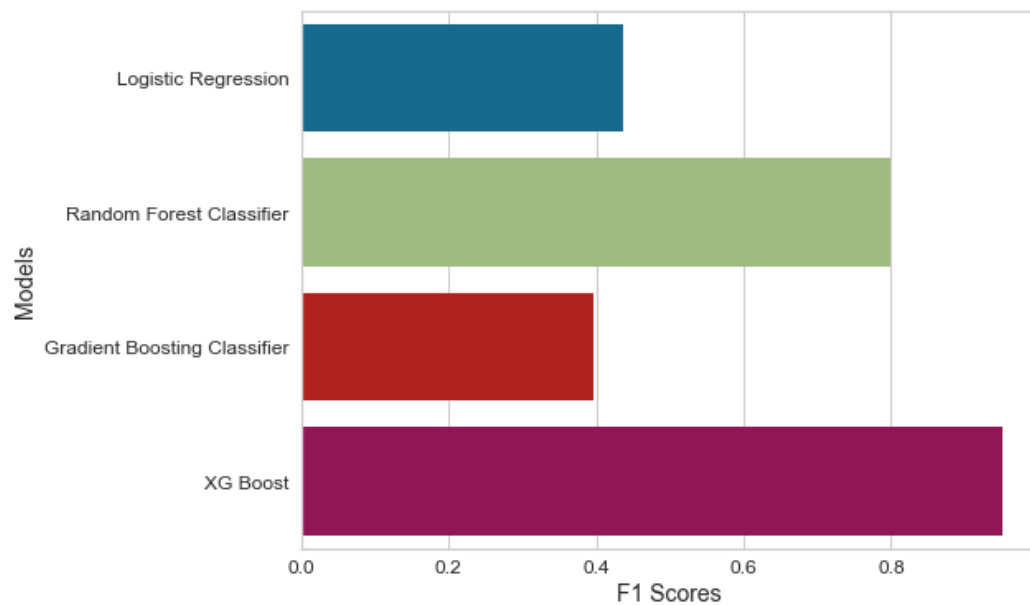
- Understanding who your customers are and what's driving them
 - Conducting surveys to determine what kind of offers people would be willing to accept
 - Understanding the reason of people rejecting the offer
 - Understanding if giving multiple offers would improve the acceptance rate
- Renewal offers should be distributed after understanding different customer segments
 - Personalizing the type of offers being provided to customers



APPENDIX

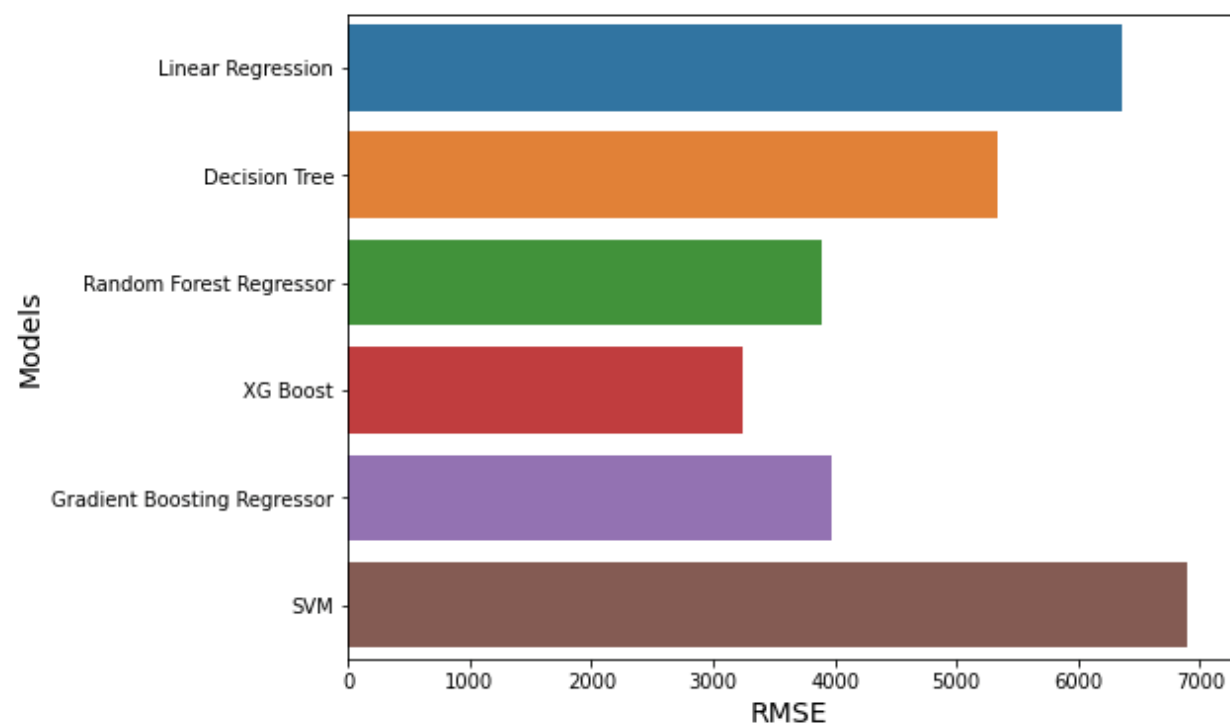


F1 SCORES AND ACCURACY SCORES – CLASSIFICATION



	Models	Accuracy Scores	Precision Score	Recall Score	F1 Score
0	Logistic Regression	0.7190	0.3147	0.7153	0.4371
1	Random Forest Classifier	0.9470	0.9504	0.6889	0.7988
2	Gradient Boosting Classifier	0.8628	0.6029	0.2942	0.3954
3	XG Boost	0.9850	0.8138	0.9497	0.9508

RMSE



	Models	RMSE
0	Linear Regression	6353
1	Decision Tree	5342
2	Random Forest Regressor	3886
3	XG Boost	3250
4	Gradient Boosting Regressor	3970
5	SVM	6902

FEATURE IMPORTANCE

