1

Tim DS, Astro Boys - 2021

HEALTH INSURANCE CROSS SELL PREDICTION

Data Science Team of

ASTRO BOYS INSURANCE. INC

BINTANG ADI KUSUMA

ROBERTSEN PUTRA SUGIANTO

TOSSY ADHAHIR RUKMANA RAUF

The team

Data Science Team, Astro Boys - 2021

Presentation Highlights

- Background
- Business Insight
- Modeling
- Potential Impact
- Conclusion

BACKGROUND

What happened in India?

contribute

6%

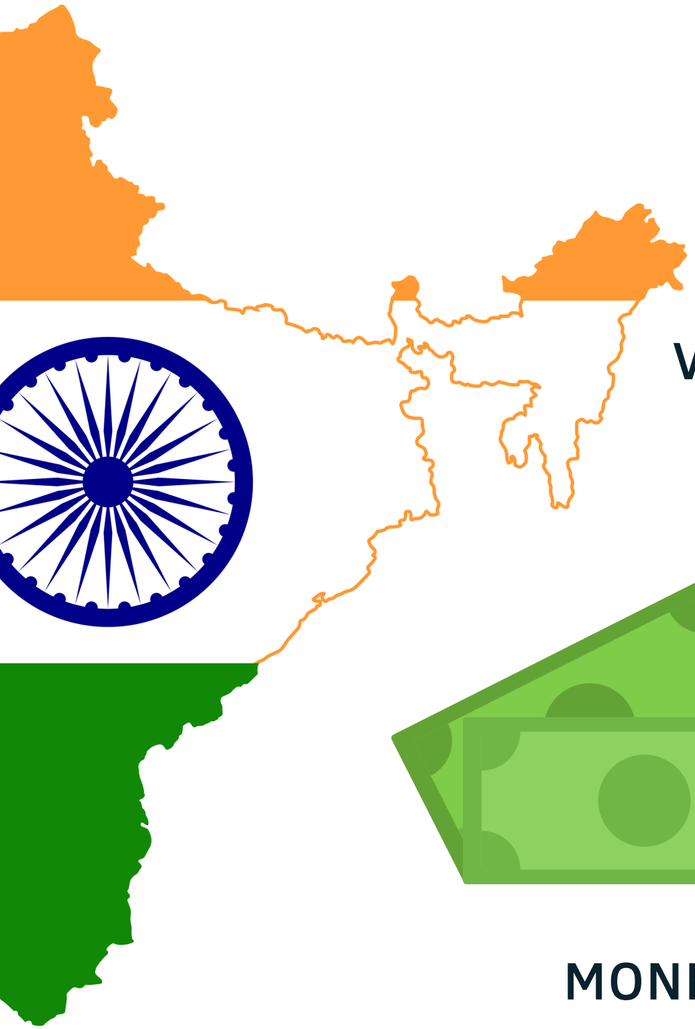
OF TOTAL GLOBAL ACCIDENT

57 %

UNINSURED

HTTPS://WWW.STATISTA.COM/TOPICS/5982/ROAD-ACCIDENTS-IN-INDIA/

What happened in India?



Road accident

LOSS

VEHICLE DAMAGE

LOSS

LOSS

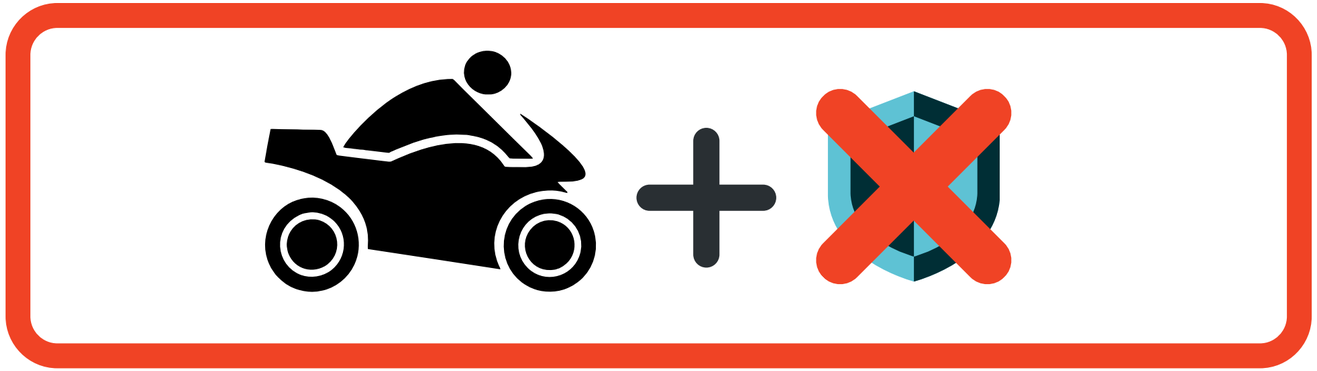
MONEY

LOSS

HEALTH ISSUE (HOSPITALIZED)

HTTPS://WWW.FORBES.COM/ADVISOR/IN/CAR-INSURANCE/MOTOR-INSURANCE-HAS-A-NEW-HOPE-IN-DIGITAL/

What happened in India?



MOTOR VEHICLE ACT, 1988

Astro Boys Insurance

Health

Vehicle

What is the problem?

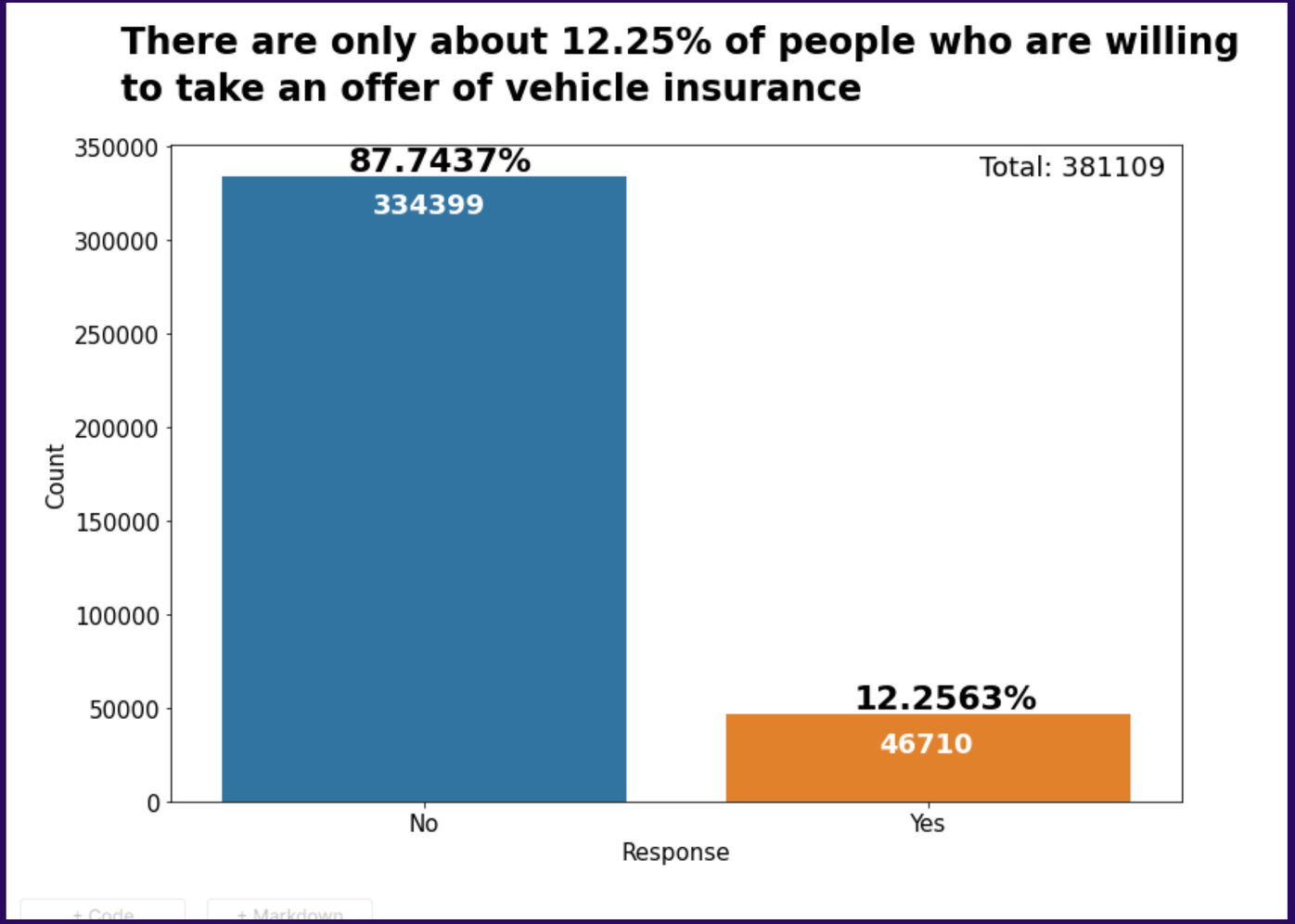
Win Rate

Sales Team just converted

12.2563%

or

46.710 people from **381.109**



Data Science Team, Astro Boys - 2021

Goals

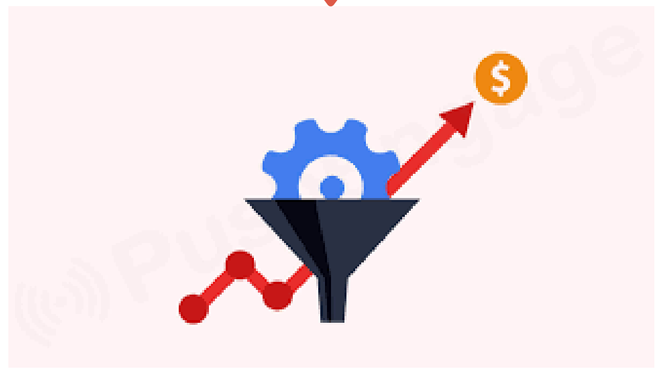
Revenue

Win Rate

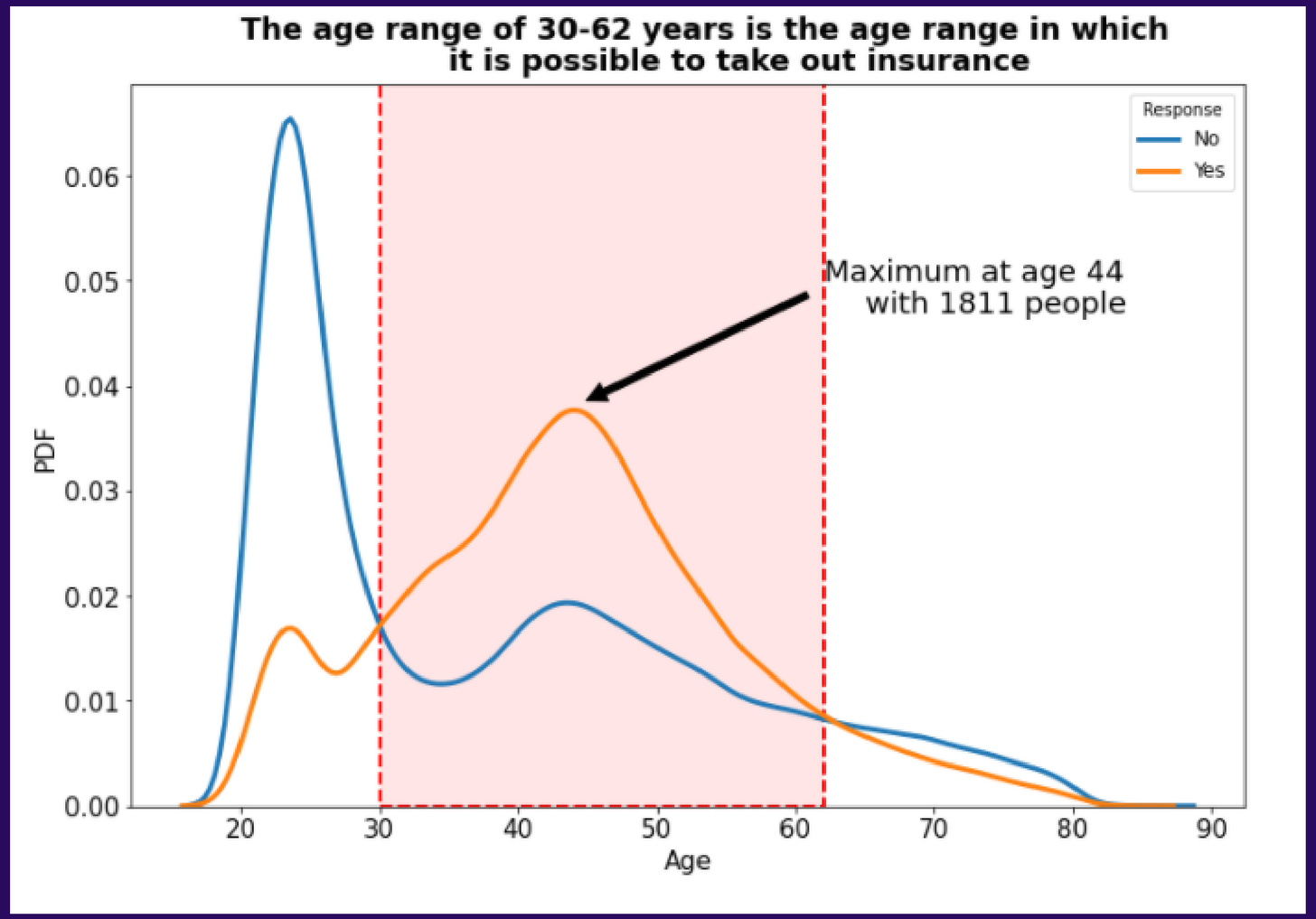
Customer

Objective

Model



INSIGHT



We can reach this range of age by divided it by two groups:

Millenials

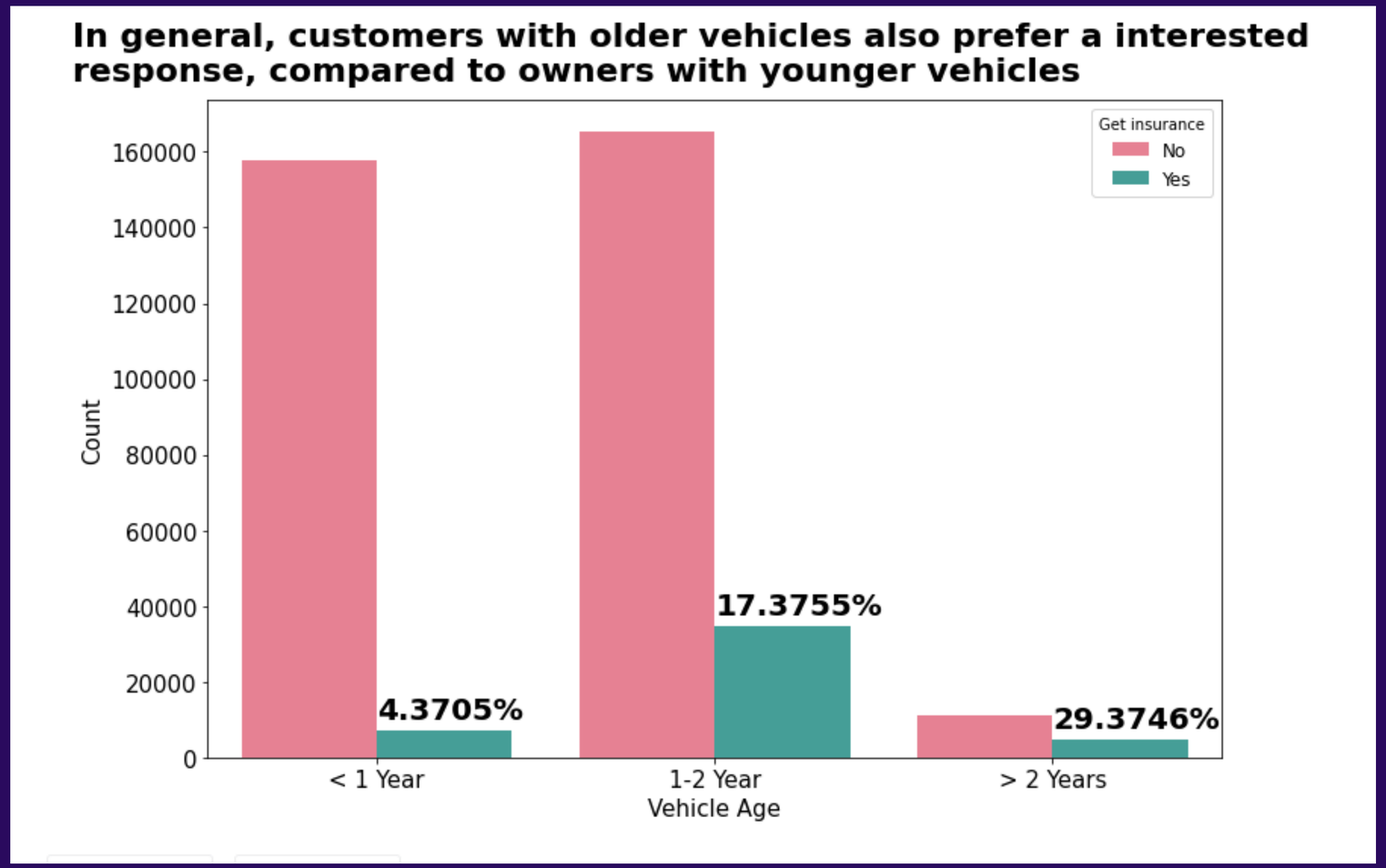
(<=39 y.o.)

Give Promotion/ Discount

X

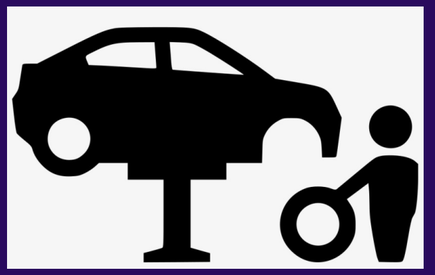
(>39 y.o.)

- Easy apply
- Voucher



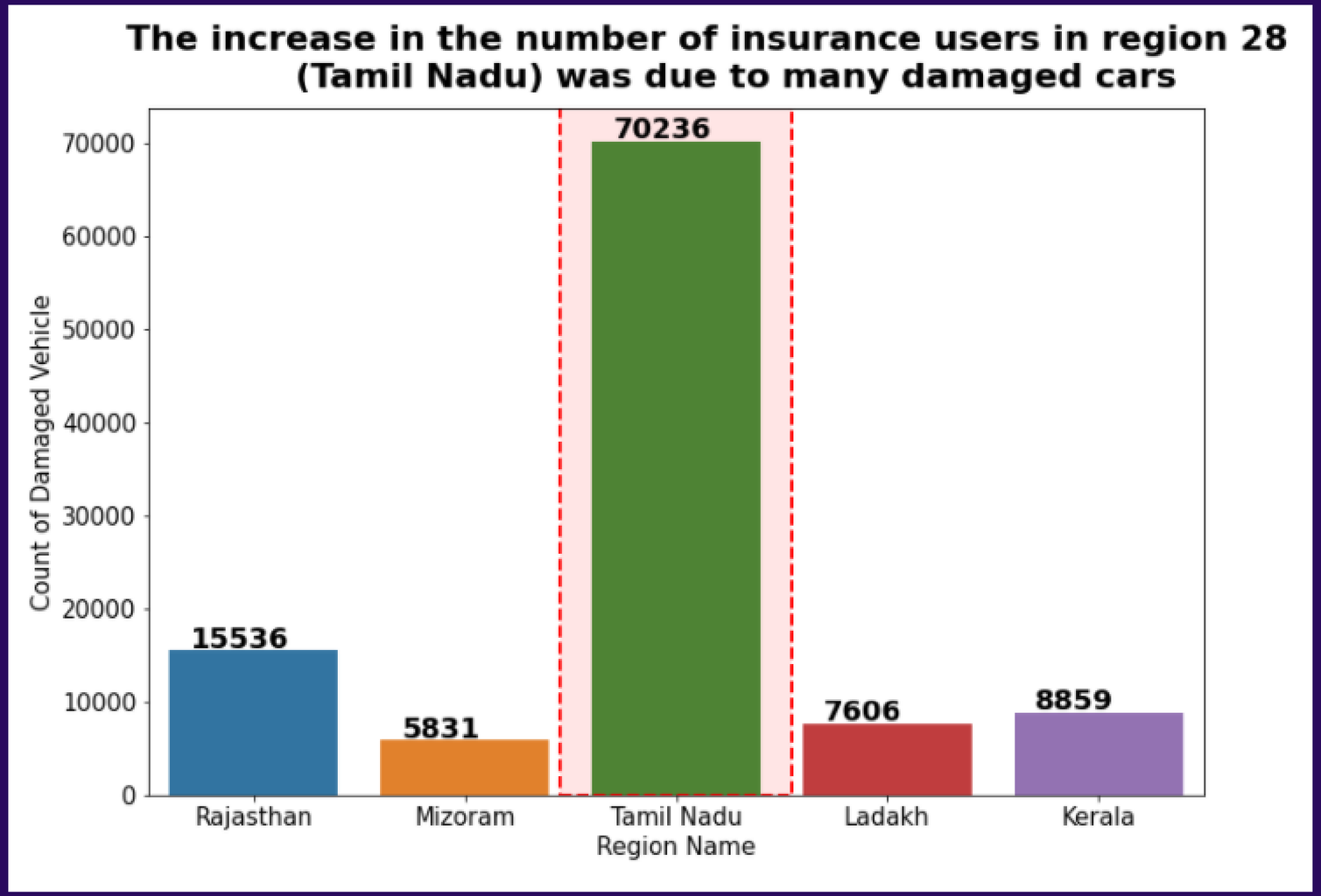
New Vehicle owner typically **not interested** to Vehicle Insurance

We can reach people with new vehicle, by:



Give the new customer **a voucher for maintenance** if take vehicle insurance

14



We can reach the people with damaged cars, especially in Tamil Nadu, by:

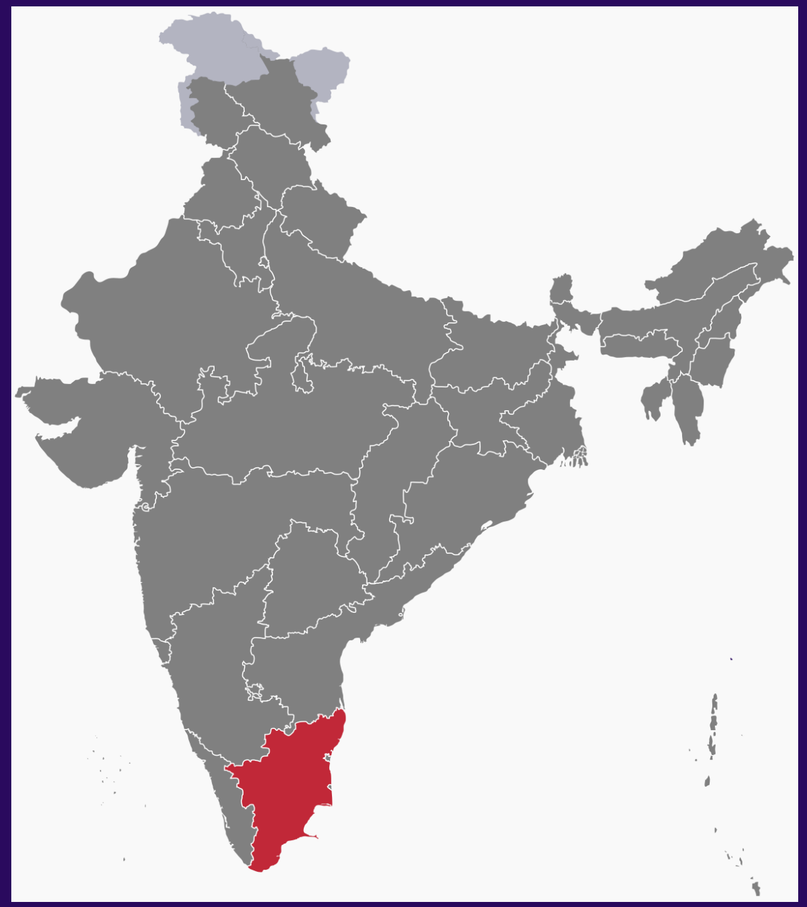
1

Car Repair Shop (**profit sharing**):
- Subcription promo
- Discount

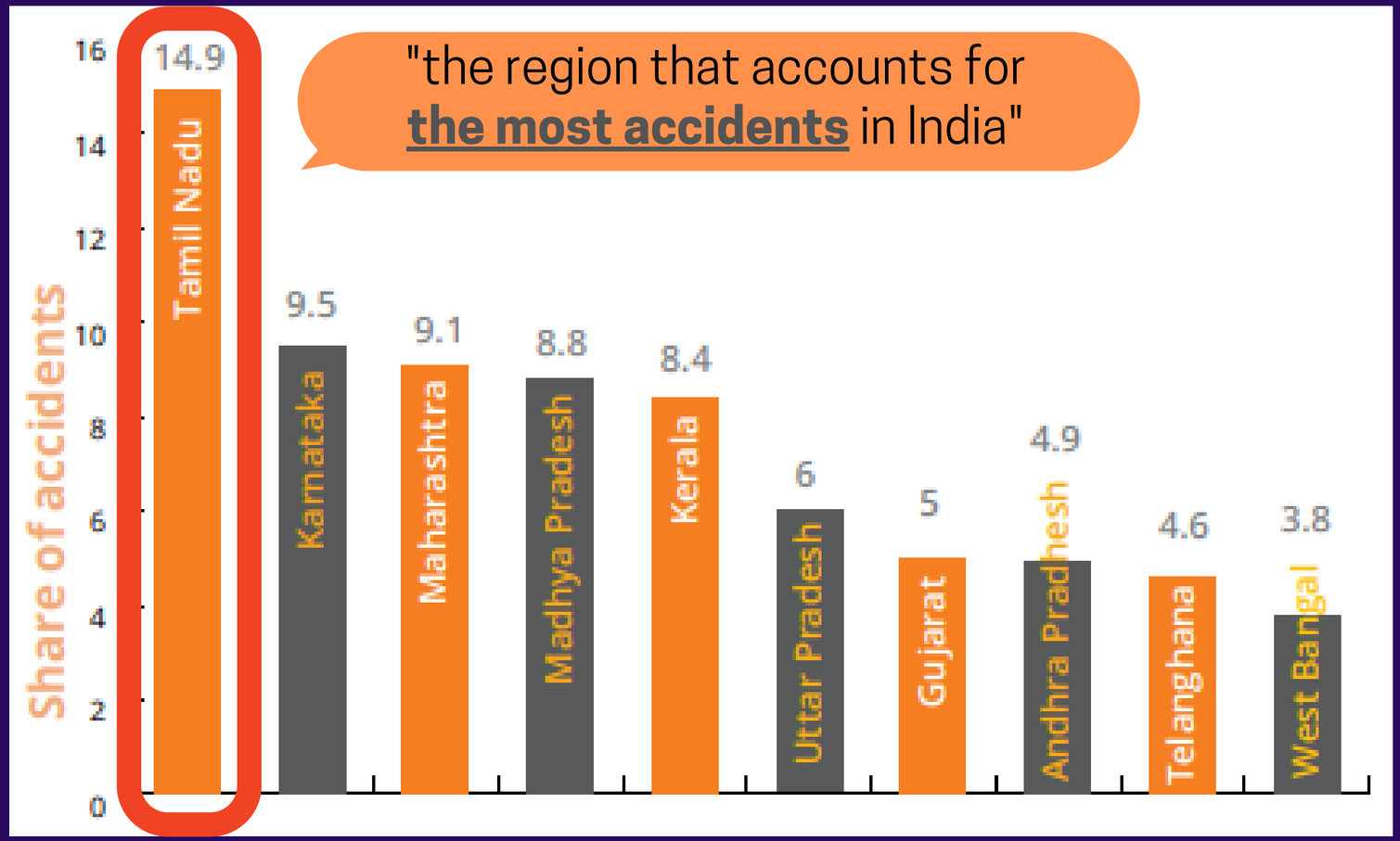
2

Used Car Dealer (**profit sharing**):
- Subcription promo
- Discount

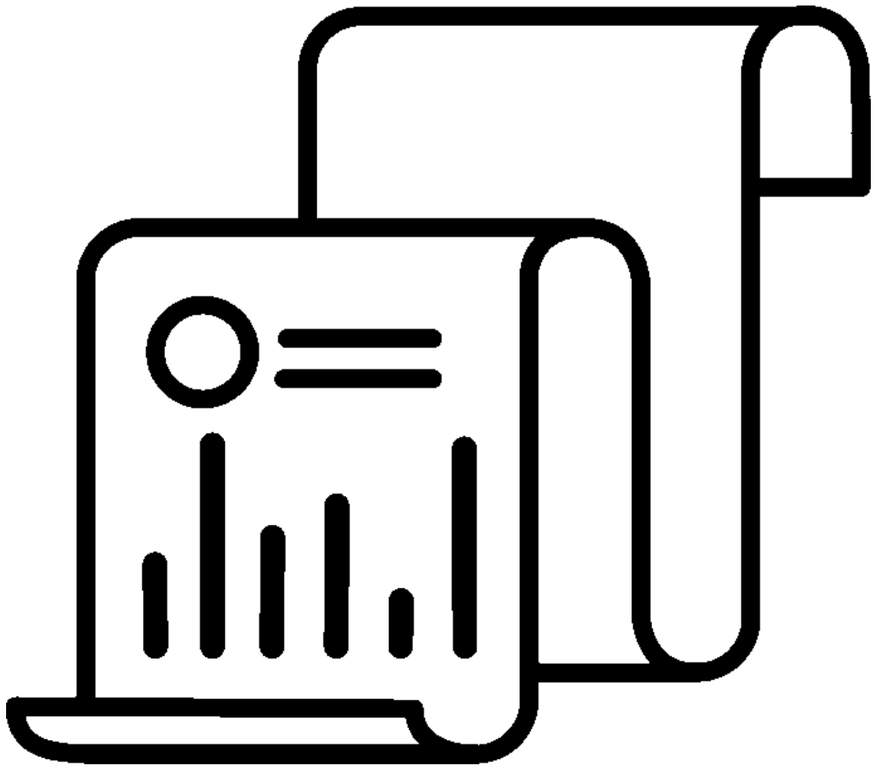
Tamil Nadu (Region 28)



Chennai one of the most crowded city on tamil nadu



SUMBER: 2017. NATIONAL INSTITUTE OF MENTAL HEALTH AND NEURO SCIENCES. ADVANCING ROAD SAFETY IN INDIA IMPLEMENTATION IS THE KEY



DATA PREPROCESSING

381.109

entry

12 KOLOM

- id
- gender
- age
- driving license
- region code
- previously insured
- vehicle age
- vehicle damage
- annual premium
- policy sales channel
- vintage
- response

Outlier Handling

Standardization

Feature Engineering

90.310 rows

Encoding Categorical Column

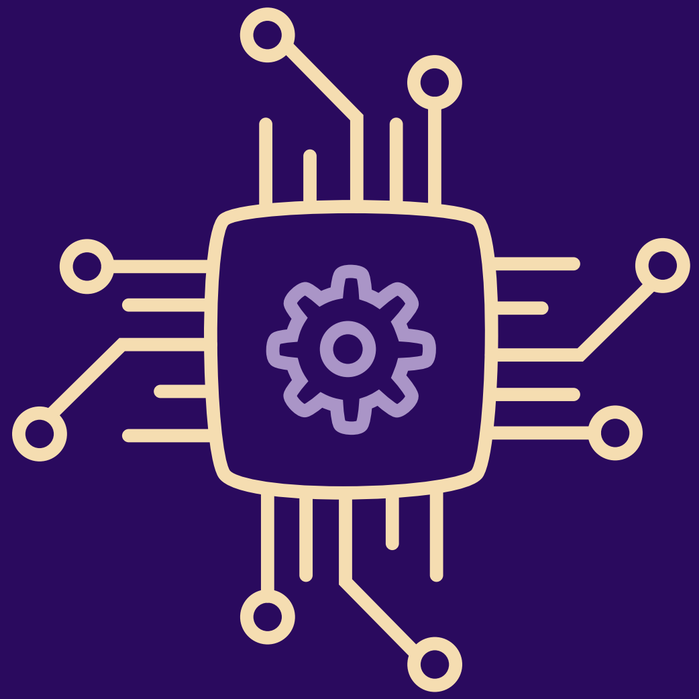
One Hot Encoding
Binary Encoding

Class Imbalance (Undersampling)

90.310 data is enough
45.155 each class

Previous Iteration:

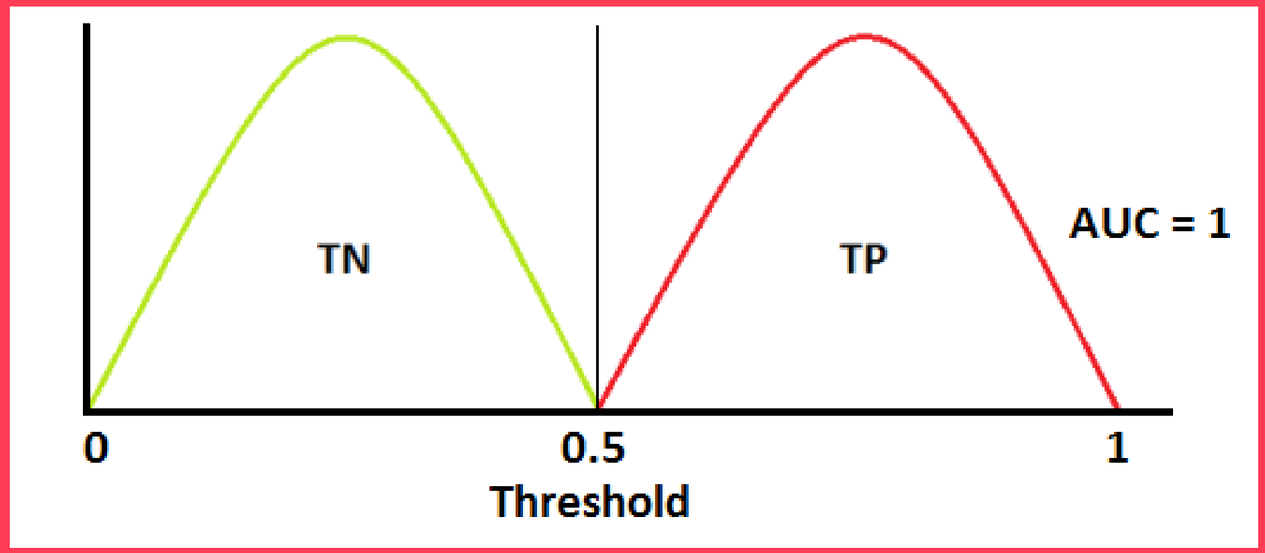
Just One Hot Encoding
One Hot Encoding + Outliers Handling
One Hot Encoding + Outliers Handling + Standardization



MODELING

ROC-AUC

~measure of data separabiity



Ataman, Kaan. *Learning to Learning to rank by maximizing the A y maximizing the AUC with linear pr UC with linear programming for problems with binary output*. 2007. University of Iowa.

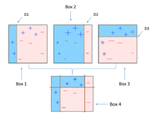
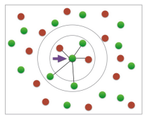
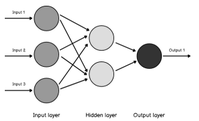
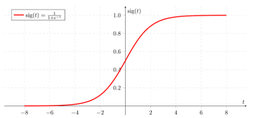
Precision

~measure the ratio between true positive and all predicted positive

$$\text{Precision} = \frac{\text{people who will } \textit{actually} \text{ take out the insurance}}{\text{everyone who } \textit{predicted would take} \text{ vehicle insurance}}$$

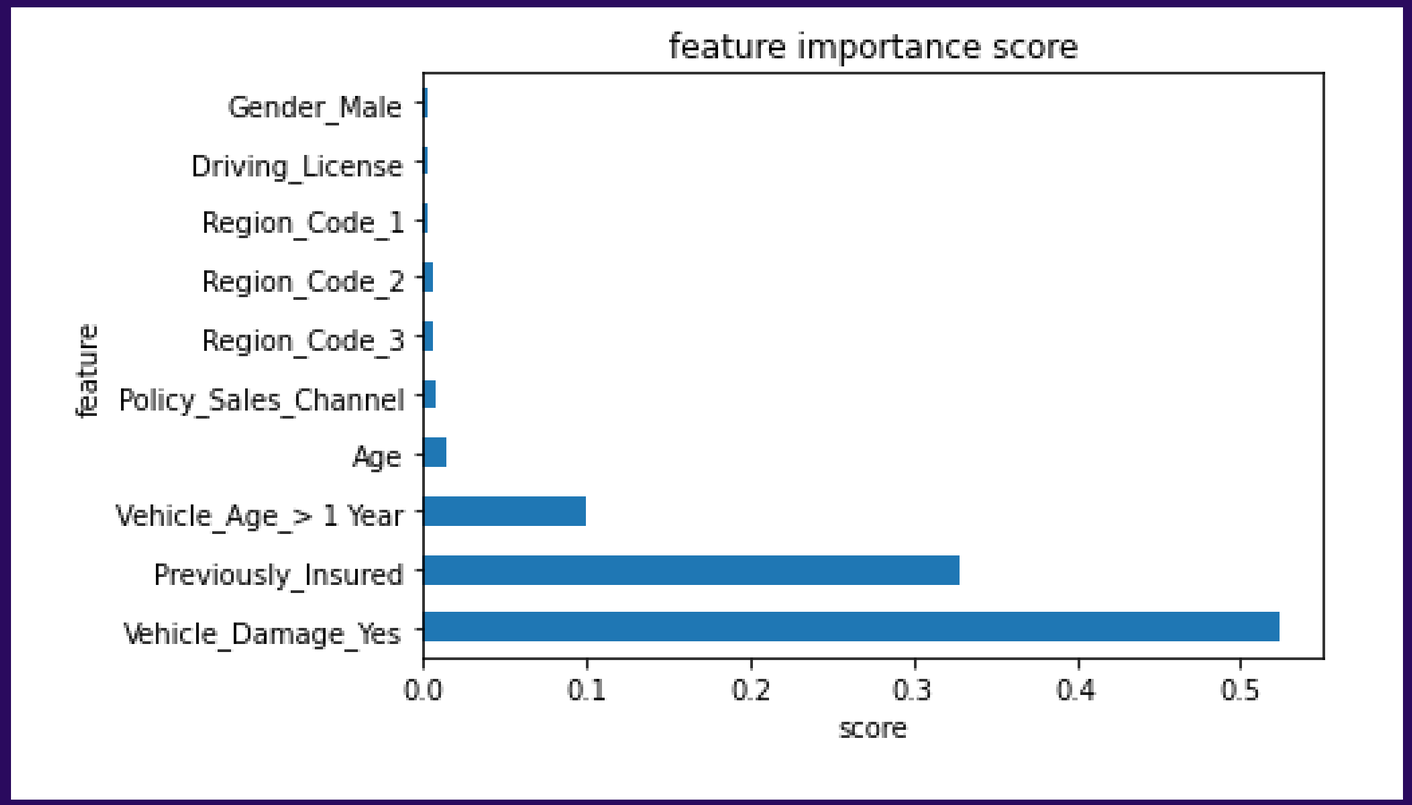
We care about **win rate** to save **offering cost** and **increase the revenue**

Researched Model

Rank based Precision			Rank based AUC	
73.617 %	(eXtremeGradientBoosting) (XGBoost)	dmlc XGBoost	**85.69 %**	**1**
73.563 %	(Decision Tree Classifier)		**83.84 %**	**6**
73.459 %	(Random Forest Classifier)	Random Forests®	**85.56 %**	**2**
73.380 %	(Adaptive Boosting) (AdaBoost)		**85.44 %**	**3**
72.549 %	(K-Nearest Neighbors)		**82.98 %**	**7**
72.549 %	(Gaussian Naive Bayes)	$P(A\|B) = \frac{P(B\|A)P(A)}{P(B)}$	**82.24 %**	**8**
72.354 %	(Artificial Neural Network)		**84.87 %**	**4**
70.794 %	(Logistic Regression)		**84.15 %**	**5**

*Hyperparameter testing included

Feature Importance



From the extraction of feature importance from the model, there are several parameters for the best probability leaf:

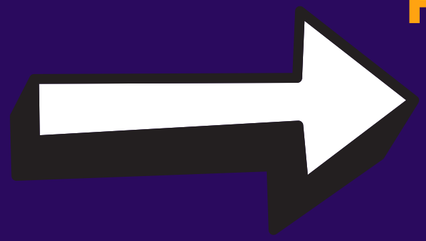
1. Customer got his/her vehicle damaged in the past
2. Customer doesn't have Insurance
3. Vehicle Age > 1 year

Potential Customer

Random sampling

the probability that the customer will be interested is:

12,26 %



Filtering based on feature importance

the probability that the customer will be interested is:

28,10 %

POTENTIAL IMPACT

Win Rate

Before using model

12,26%

repondents are interested

Filtering based on feature importance

28,10%

repondents are interested

Using dmlc XGBoost

73,62% (precision)

repondents are interested

Win Rate ~ Number of Customer

Revenue

*simulated

100.000

prospective customer reached by marketing team

with mean annual premium of

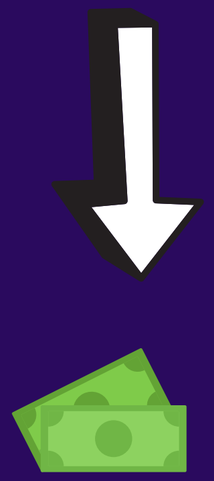
Rp. 5.000.000

Revenue

Before using model

12.256

customer

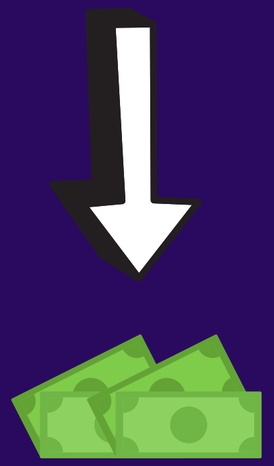


Rp. 61,280 billion /year

Filtering based on feature importance

28.098

customer

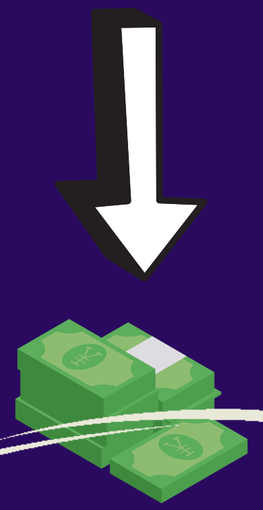


Rp. 140,490 billion /year

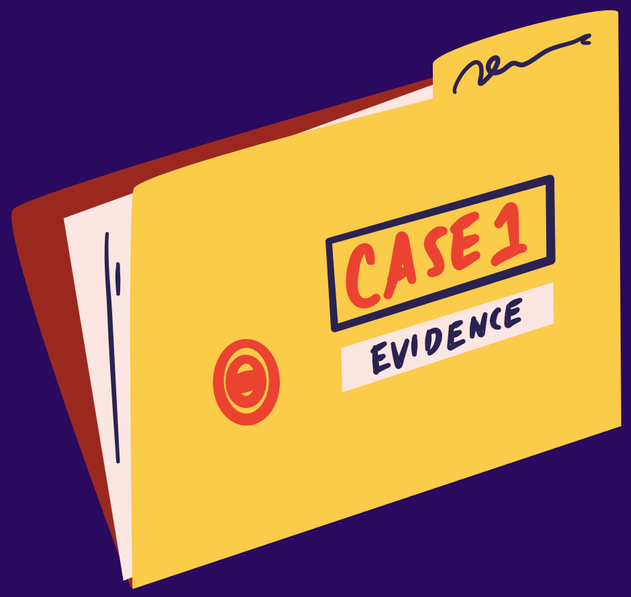
Using dmlc XGBoost

73.617

customer



Rp. 368,085 billion /year



CONCLUSION & RECOMMENDATION

Conclusion

Before using model

12.256
customer

Rp. 61,280
billion / year

12,26%

Filtering based on feature importance

28.098
customer

Rp. 140,490
billion / year

28,10%

Using dmlc XGBoost

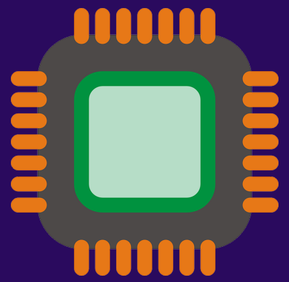
73.617
customer

Rp. 368,085
billion / year

73,62%

30

Recommendation



Use the ***proposed model*** to predict potential customer

Focusing marketing in ***Region Code 28*** (Tamil Nadu) by adding more store or join an exhibition

Reach people who are ***30-62 years of age***

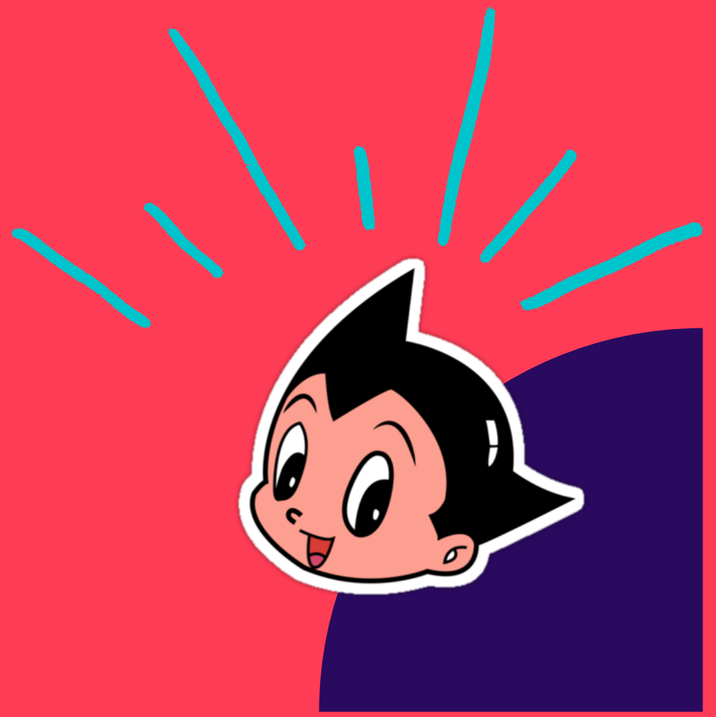
(give any **promotion/dicsount** for millenials, and **easy apply** to x generation)

Reach out to people ***who have damaged their car*** and ***old car age***

(by **collaborating** with second hand dealer and garage)

Any questions?

Enquiries and concerns are welcomed :)



Data Science Team, Astro Boys - 2021

Thank you for your attention

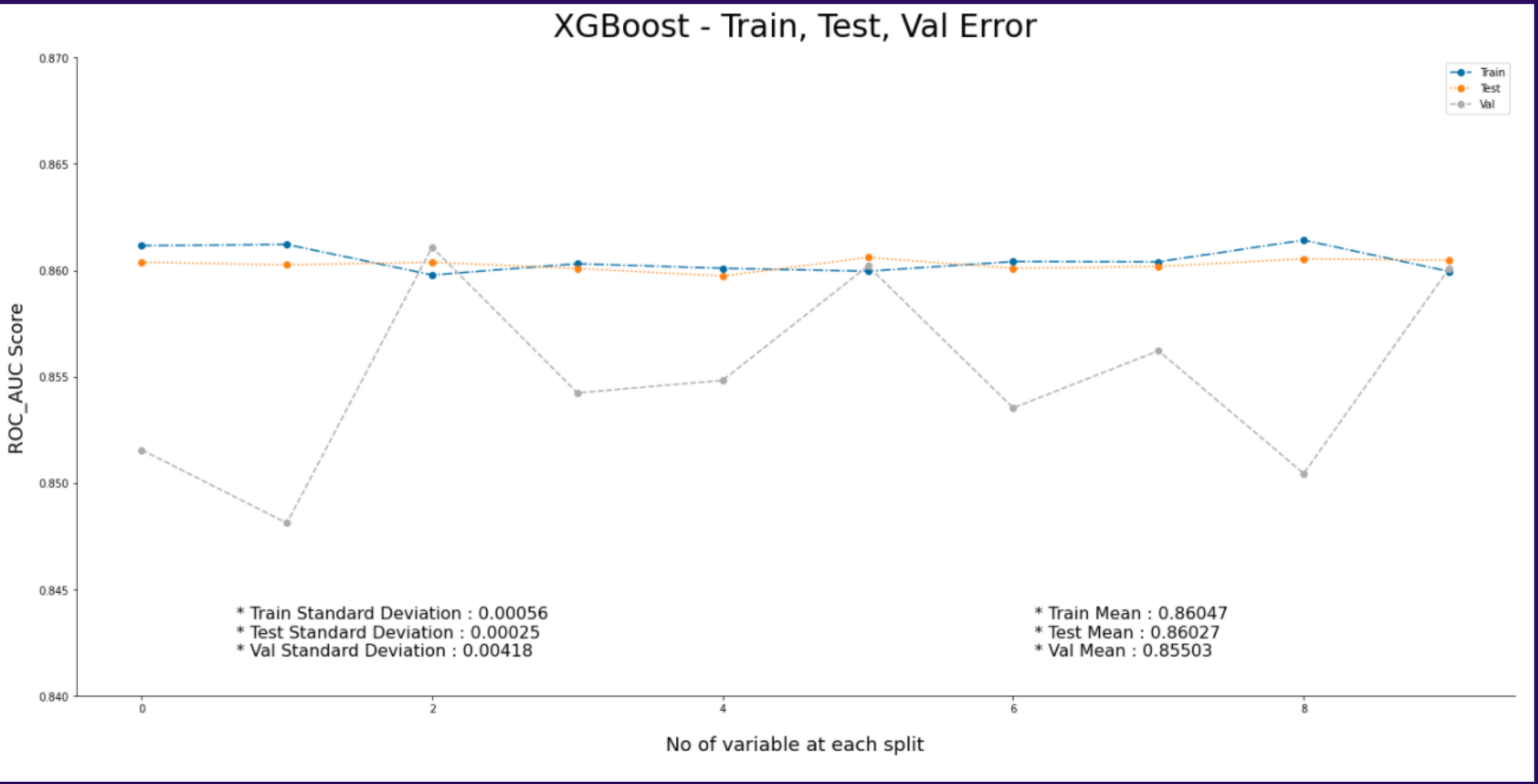
Tim DS Astro Boys Insurance Inc.

Lampiran

Tim DS Astro Boys Insurance Inc.

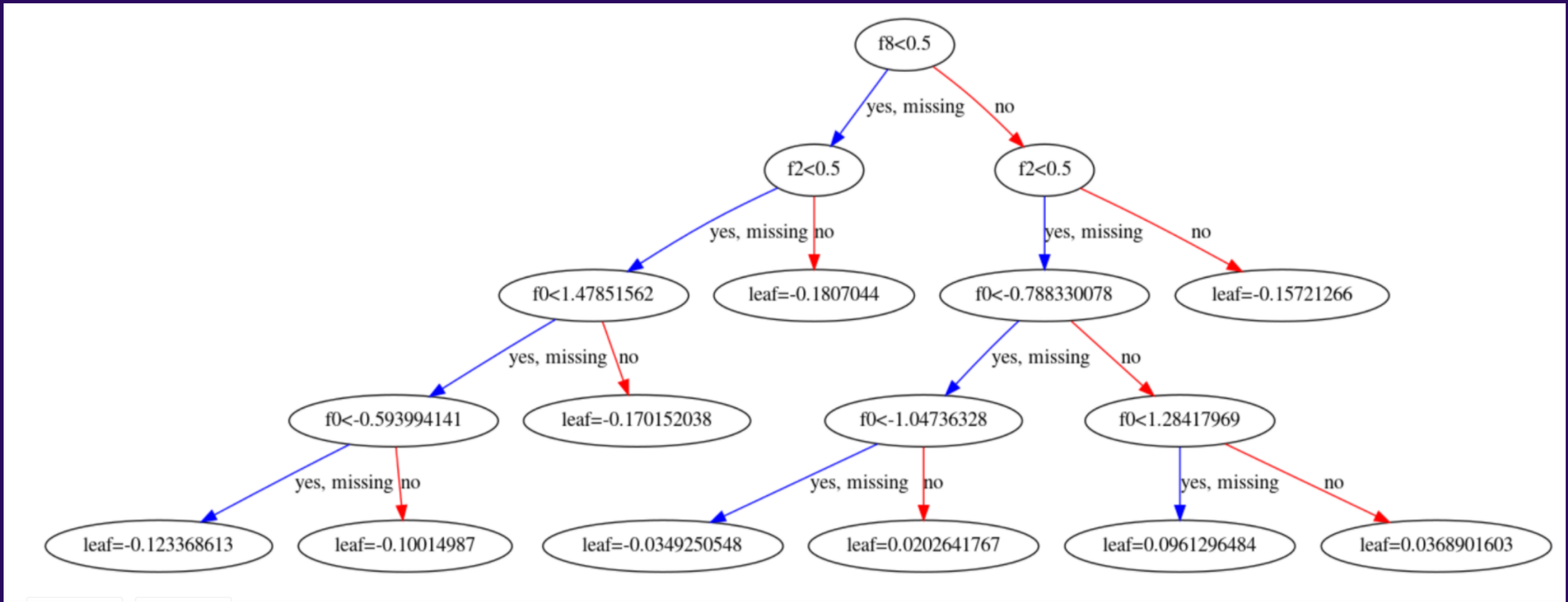
Cross Validation

Tim DS, Astro Boys - 2021



08

TREE NUMBER 1 XGBOOST



Tim DS, Astro Boys - 2021

MODEL RESEARCH

	Method	Precision	Recall	AUC
0	Logistic Regression	70.7937	96.9439	84.1565
1	KNN	72.5487	89.007	82.9802
2	Naive Bayes	72.5487	89.007	82.9802
3	Decision Tree	73.5635	90.6103	83.8452
4	Random Forest	73.4593	93.6575	85.564
5	ANN	72.3544	95.3317	84.879
6	XGB	73.6174	93.507	85.6998
7	AdaBoost	73.38	93.07	85.4421

TUNED XGBOOST


```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=0.7, eta=0.005, gamma=0.0,
              gpu_id=-1, importance_type='gain', interaction_constraints='',
              learning_rate=0.1, max_delta_step=0, max_depth=4,
              min_child_weight=5, missing=nan, monotone_constraints='()',
              n_estimators=100, n_jobs=4, num_parallel_tree=1,
              objective='binary:logistic', random_state=42, reg_alpha=0,
              reg_lambda=1, scale_pos_weight=1, subsample=1,
              tree_method='exact', use_label_encoder=True,
              validate_parameters=1, verbosity=None)
```

Manual Filter Using Fitur Importance


```
df_manual_filter_D=df_manual_filter[df_manual_filter['Vehicle_Damage'].str.contains('Yes')]
df_manual_filter_D['Response'].sum()/len(df_manual_filter_D['Response'])*100
```

```
23.765545987017507
```

```
df_manual_filter_D_P=df_manual_filter_D[df_manual_filter_D['Previously_Insured']==0]
df_manual_filter_D_P['Response'].sum()/len(df_manual_filter_D_P['Response'])*100
```

```
25.010548465403776
```

```
df_manual_filter_D_P_A=df_manual_filter_D_P[~df_manual_filter_D_P['Vehicle_Age'].str.contains('< 1 Year')]
df_manual_filter_D_P_A['Response'].sum()/len(df_manual_filter_D_P_A['Response'])*100
```

```
28.09858187314237
```