

[Problem Statement](#)

[Data Preprocessing](#)

[Handling Missing Values](#)

[Data Visualization](#)

[Univariate Analysis](#)

[Bivariate Analysis](#)

[Preparing data for modelling](#)

[Building Model](#)

[Finetune Cutoff probability](#)

[AUC Score](#)

## Problem Statement

Predicting whether a given customer will become a lead for availing credit card

## Data Preprocessing

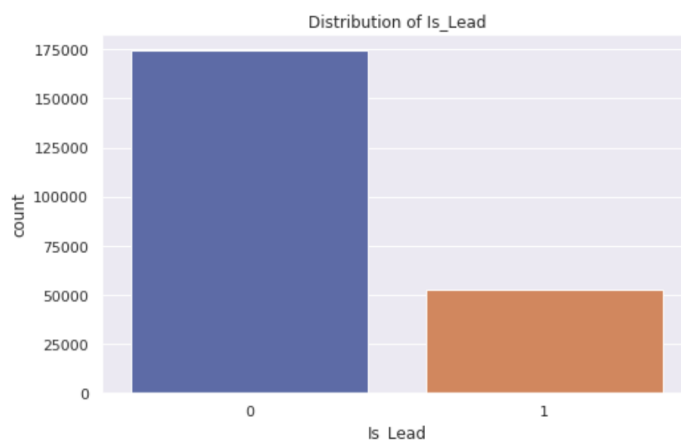
### Handling Missing Values

- Credit\_Product has 11 % missing values, It was replaced with mode
- Non impacting column ID was removed

## Data Visualization

### Univariate Analysis

- Target distribution



- Similarly all the features distribution was checked using seaborn's distplot and countplot

### **Inferences from univariate analysis ¶**

- Count of Lead is 3.2 times that of Not Lead
- Median age is 43
- Median Vintage is 32
- Median of average account balance is 894,601
- Top 3 region code are RG268 (35k), RG283 (29k), RG254 (26k)
- Occupation Self employed is highest while entrepreneur is least
- Channel code X1 is highest while X4 is least
- Count of customers not active in last 3 months is 1.5 times that of active
- Count of customers not using credit product is 2.4 times that of using
- Count of male customers are 1.2 times that of females

## **Bivariate Analysis**

- Relationship between all the features and is\_lead was analysed using seaborn's boxplot, countplot and barplot
- Outlier treatment
  - Rows with average account balance > 2510200 and vintage > 120 were cropped

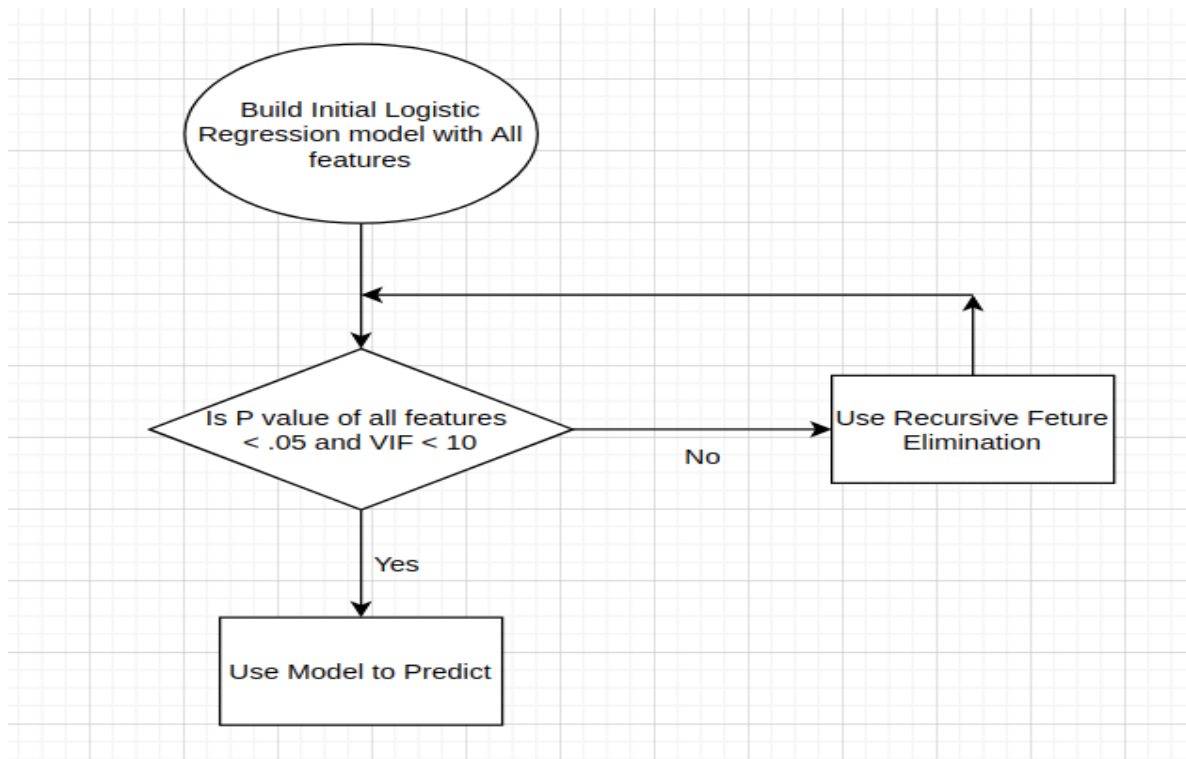
### **Inferences from Bivariate Analysis ¶**

- With increase in Avg\_Account\_Balance lead conversion chances increases
- Customers in the age group 45 to 65 and 75 to 85 have higher chances of becoming lead
- Customers with vintage 70 to 110 higher chances of becoming lead
- Customers active in last 3 months have 8% higher chance of becoming lead
- Customers using credit product have 11% higher chance of converting to leads
- Channel\_Code X3 has highest lead conversions 36 % while channel X1 has lowest- 9%
- Occupation Entrepreneur has highest lead conversions 65 %
- Region code 283, 284, 263 have higher conversions around 30 %
- male customers have 6 % higher chances to become lead

## **Preparing data for modelling**

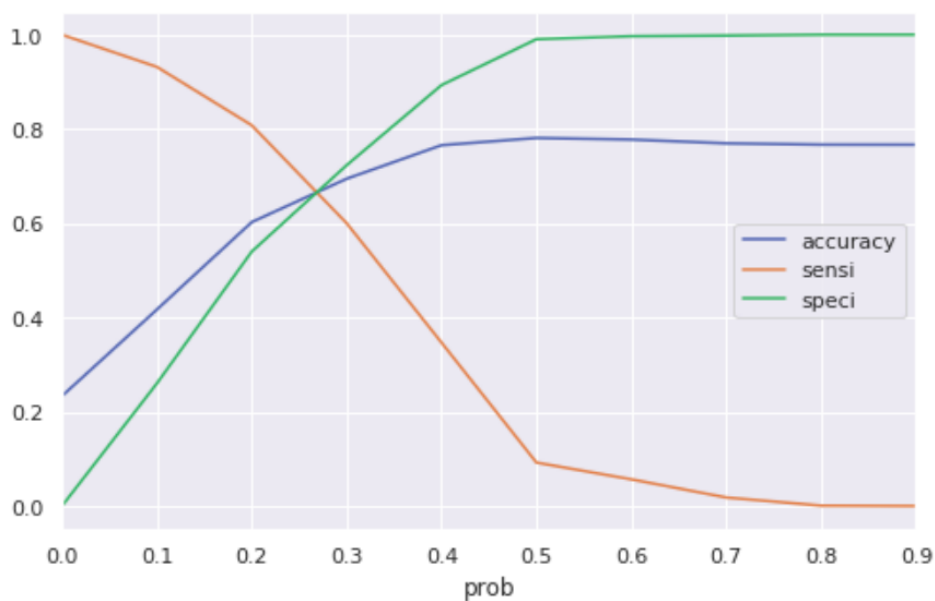
- Categorical columns were encoded to numerical
- Data was scaled using MinMax Scaler
- Correlation of all the features with is\_lead was analysed

## Building Model



## Finetune Cutoff probability

- The initial prediction was made by choosing probability as 0.5
- This cutoff was further fine tuned to 0.27 analysing, sensitivity, specificity and accuracy



## AUC Score

