

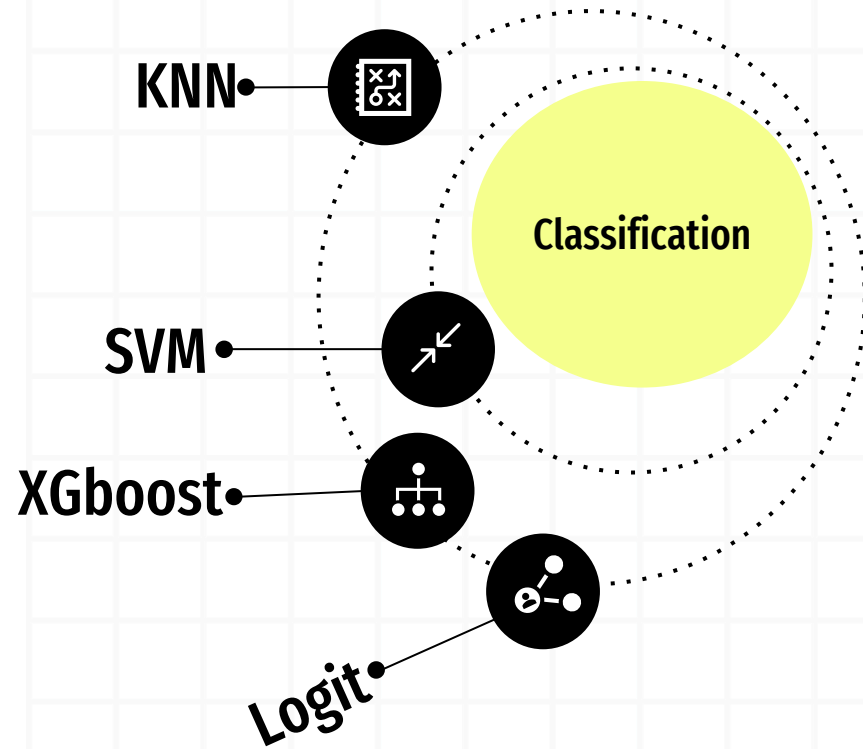
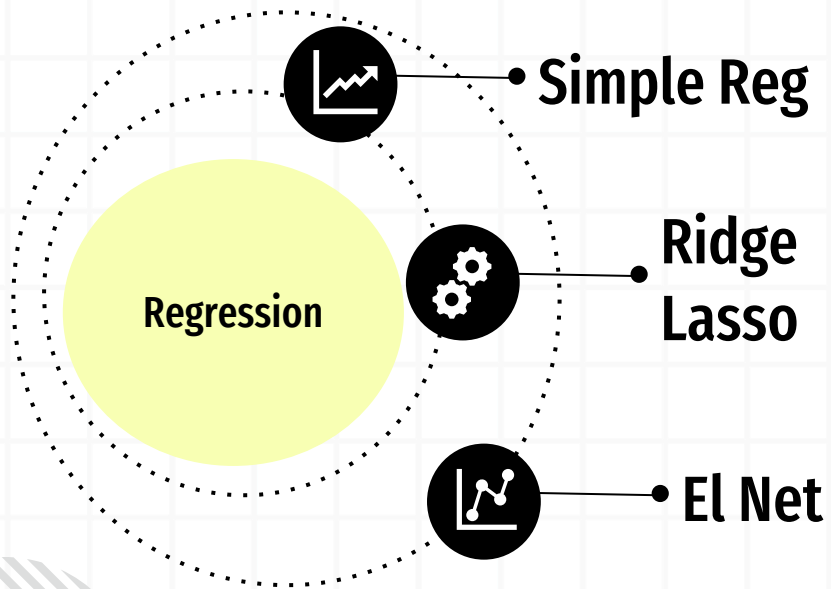
Machine Learning



Pearly Tantra 455722

Project Presentation

Models



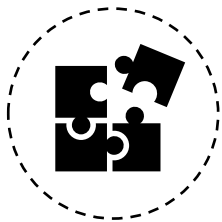
Regression RoadMap



Explanatory Data

- NA Values
- Converting Values
- Outliers

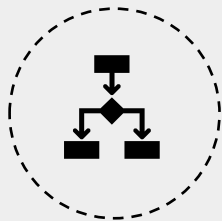
• • • • •



Feature Selection

Correlation Matrix

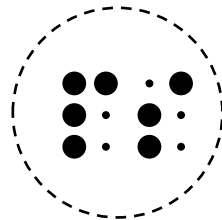
• • • • •



Sample Split

Proportion:
90 - 10

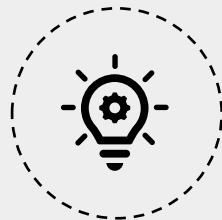
• • • • •



Feature Scalling

Numerical Variables

• • • • •



Modelling

- Linear Reg
- Ridge/Lasso
- Elastic Net

• • • • •

Regression

Prior to Modelling

01 Explanatory

01. NA Values : categorical (omit) & numerical (mean/median)

02. Convert : character -> factors : ordered / unordered

03. Outliers : Numerical Vars w/ Z-score

02 Feature Selection

Correlation Analysis : `mother_delivery_weight` ✗
`mother_body_mass_index`

03 Sample Split

90/10 == 1,7 mio/198k

04 Feature Scalling

Numerical Vars



Regression

Modelling



01

Simple Regression

- General Model
- GETS : omit father_age

02

Ridge and Lasso

- Regularization
- Predictors converted to Matrix
- Optimal Lambda

03

Elastic Net

- Set up Train Control
- Model

Regression

Performance - MAPE

Simple Regression

16%

Ridge and Lasso

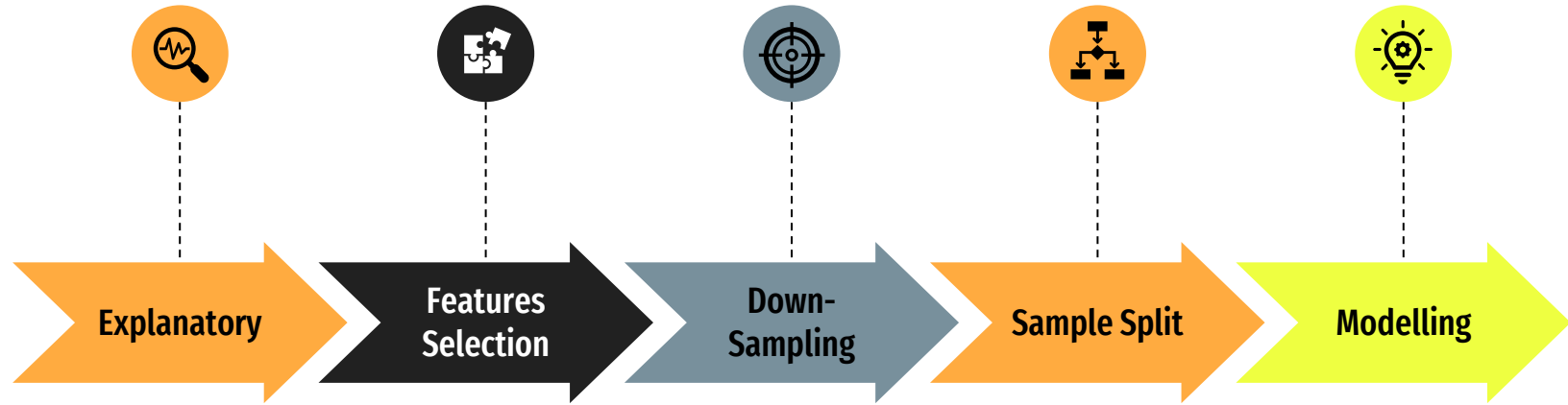
12,71 %



Elastic Net

12,41%

Classification RoadMap



Classification

Prior to Modelling

01

Explanatory



- NA Values : categorical (omit) & numerical (mean/median)
- Convert : character -> factors : ordered / unordered

02

Features Selection



Correlation Analysis

- NumVar: omitting 'remaining_credit_limit' & 'customer_available_credit_limit'
- CatVar : omitting 'customer_sex'

03

Down-Sampling



To balance the data between 'open' and 'closed'

04

Sample-Split



80/20 : 290 ~ 290

Classification

Modelling

Simple Logit

- Train General Model
- GETS method :omit customer_education

XGBOOST

- Convert X test and train data into model matrix
- Train the model

SVM

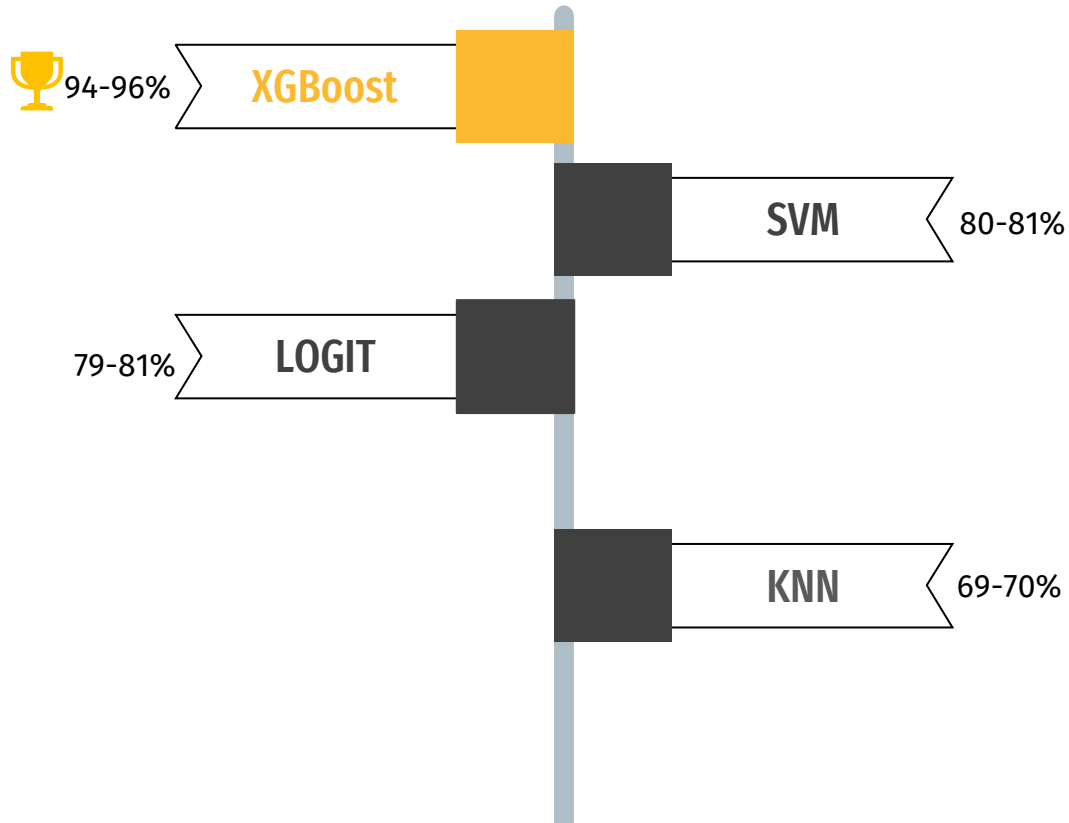
- Train the model w/ additional parameters. Type = c-classification, Kernel = linear

KNN

- Set a train control
- Train the optimum K-Value (k= 9)
- Train the model

Classification

Performance – balanced accuracy





Conclusion