Advanced Analytics in Fundraising

How knowing donors
helps in growing donors

Group 03
10th December 2020

THE TEAM

Paul Cazilhac

Prineet Kaur Bhurji

Venkat Jayanarasimhan

Kiran TL Rao

PROJECT DEFINITION

DSC	1985	22	20	1.8%
Your Partner in Fundraising	*Foundation Year Country: Belgium*	*Humanitarian Organizations*	*Enthusiastic People*	*Average Campaign Response Rate*

TO PREDICT BEST POSSIBLE DONORS

Direct Social Communications
your partner in fundraising

Python Predictions
The art of data science

IESEG SCHOOL OF MANAGEMENT
EMPOWERING CHANGEMAKERS FOR A BETTER SOCIETY

METHODOLOGY

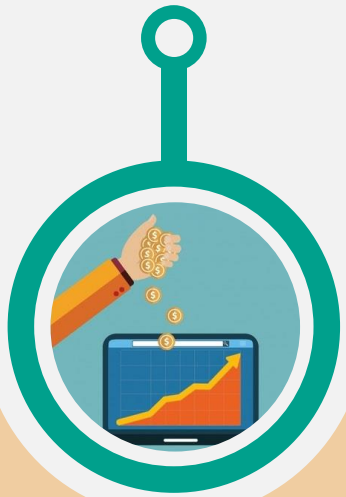
- Data Preparation
- Model Building
- Model Validation
- Model Usage

Data Preparation

Cleanse & Qualify Data

DATASET OVERVIEW

Donor Data



Train Data

Campaign Data

Gift Data

Test Data

Direct Social Communications
your partner in fundraising

Python Predictions
The art of data science

IÉSEG
SCHOOL OF MANAGEMENT
EMPOWERING CHANGEMAKERS FOR A BETTER SOCIETY

CAMPAIGN

367 X 4	*Rows X Columns*
0 NaN	*Missing Values*
1 Outlier	*LetterSent & CostUnit*
High SD	*LetterSent*
Different	*Formats (Int & Object)*

367 X 7	*Rows X Columns (Date → DD,MM & YYYY)*
0 NaN	*Missing Values*
1 Outlier	*LetterSent & CostUnit*
High SD	*LetterSent*
Standard	*Formats (Date, Int & Float)*

DONORS

44K X 7	*Rows X Columns*
2 NaN	*Zipcode*
3	*Categorical Variables (Gender, Lang. & Province)*
2	*Insignificant Variables (DoB & Region)*
2 Outlier	**Age Feature*
Different	*Formats (Int & Object)*

44K X 24	*Rows X Columns*
0 NaN	*Subset Lang. & ZipCode (Replace with Mode)*
0	*Categorical Variables Perform Dummy Encoding*
0	*New Feature – *Age Drop 2 insig. variables*
0 Outlier	*Age Reassignment*
Standard	*Formats (Date & Int)*

GIFT

216K X 4	*Rows X Columns*
15K NaN	*Campaign ID*
*4K NaN	*Still Missing in Camp. ID*
2 Outlier	*Gift Amount (Huge Diff.)*
Different	*Formats (Date)*

212K X 9	*Rows X Columns (Date → DD,MM & YYYY)*
*4K NaN	*NaN b/w 2 consecutive ID (Replace it with that ID)*
0 NaN	*Dropping the NaN Cam. ID*
0 Outlier	*Dropping Outliers*
Standard	*Formats (Date to New Var.)*

Model Building

Predict Outcomes

VARIABLES

Predictors

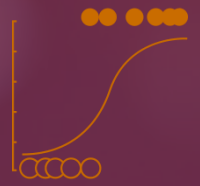
01	Gender_M
02	Gender_F
03	EN
04	FR
05	age
06	Zipcode

Target

##	Campaign Responders 1 – Donor Responded 0 – Donor Not Responded

MODEL APPLICATION

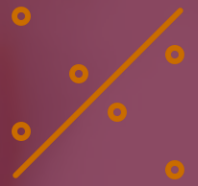
Decision Tree



Logistic Regression

Random Forest

Gradient Boosting



SVM

Neural Network

Nearest Neighbors

Model Validation

Evaluate Outcomes

AUC METRIC

Boosted Tree ML & Neural Network Model provides the Best Performance

All other remaining ML models doesn't perform well for the given datasets

But we can achieve only 61% Model Performance with the current predictors!

Is it possible to increase the Model Performance?

FEATURE ENGINEERING

Donor

01	Age @ Train Campaign Year
02	Age @ Test Campaign Year

Gift

01	Number Of Donations
02	Donations Per Year

VARIABLES

Predictors

01	Gender_M
02	Gender_F
03	EN
04	FR
05	age_6169 / age_7244
06	Number_Of_Donation
07	Donation_Per_Year
08	Flemish Brabant
09	Antwerp
10	Brussels
11	Walloon Brabant
12	West Flanders
13	East Flanders
14	Liege
15	Luxembourg
16	Namur
17	Hainaut
18	Limburg

Target

	Campaign Responders
##	1 – Donor Responded
	0 – Donor Not Responded

Train

Test


0
20
40
60
80
100
Neural Net
63
56
Boosted Tree
72
61
SVM
Logistic
Random Forest
Neighbors
Decision Tree

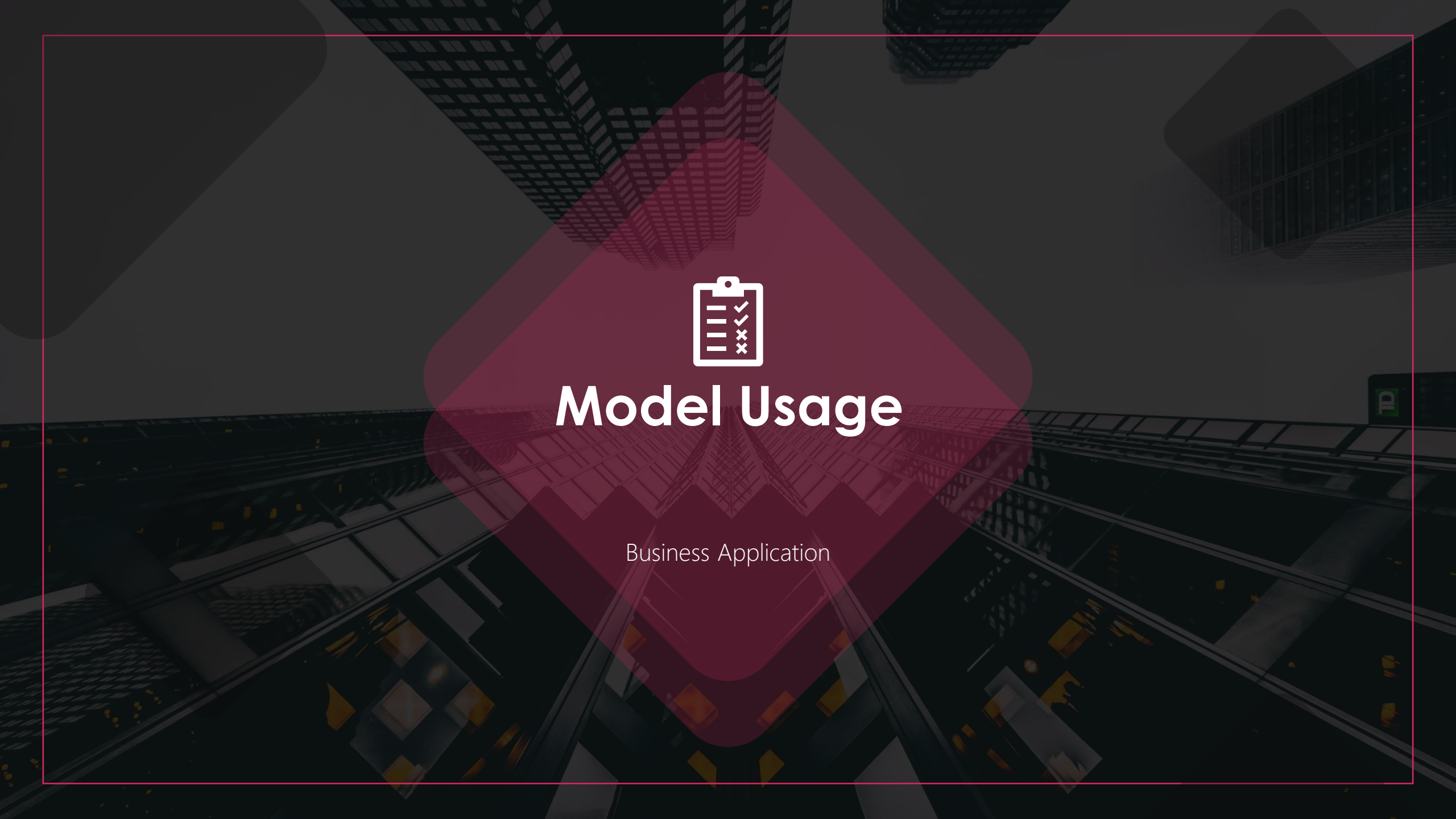
INCREASED AUC METRIC

Neural Network ML Model provides the Best Performance both in Train & Test Datasets

Boosted Tree ML Model also provides the Best Performance, but there's a drop in performance between Train & Test Datasets

All other remaining ML models doesn't perform well for the given datasets

CONFUSION MATRIX



Model Usage

Business Application

IDEAL DONOR PROFILE

Female

Dutch

Flemish Brabant

49 Years Old

Avg. € 110


Thank You!
Any Questions?