Credit Default Group 5

Diwei Zhu   **Gabriela Caballero**   **Kunyang Que**   **Ullas Srivastava**   **Yangxing Liu**

Agenda

Business Objective

1

Data overview

2

Data
Cleaning

3

Exploratory data
Analysis

4

Models
comparison

5

Chosen
model

6

Business
Implications

8
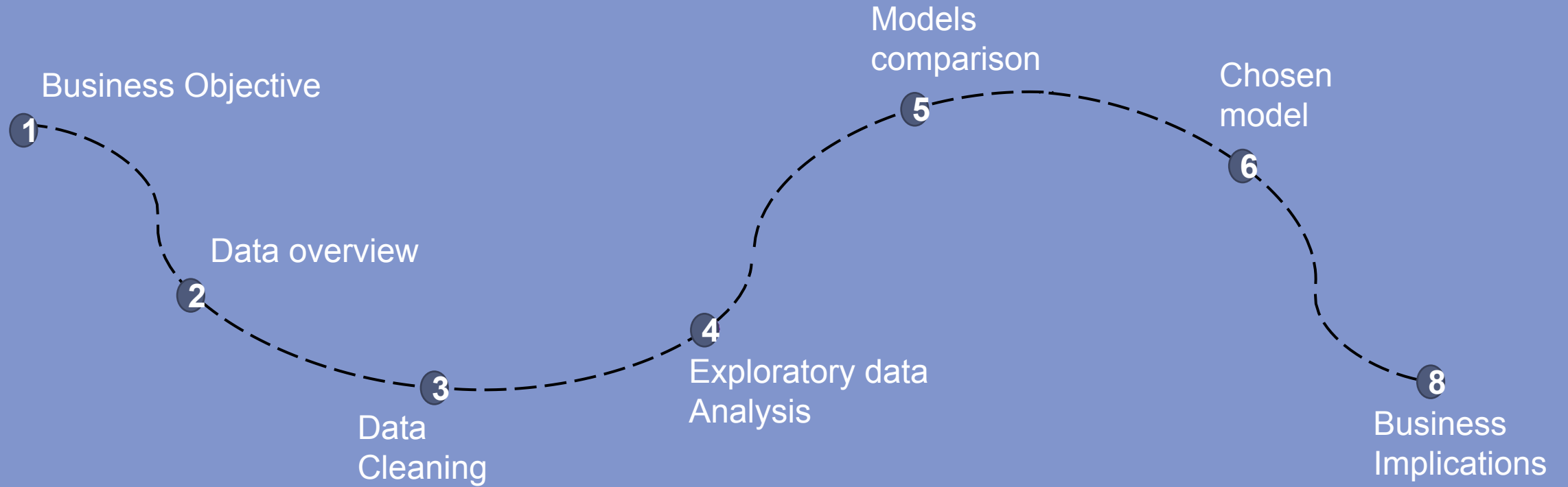
Business Objective

> "
> Nearly **1/3** of personal loan seekers leave the bank unhappy, and as many as **21%** fail to pay back their loans on time
>
> Deloitte Insights 2017

Problem:

How do our clients meet the **loan requirements**?

Solution:

Find what it takes to get your loan accepted

Using our predictive modelling and the first hand banking data related to the **consumer loans**, **revolving loan** and **cash loans**

Data Overview

292,132 rows

304 columns

4 files

Cleaning Technic

- Removing outlier

- Filling missing data

- Creating bins for trees

- One hot encoding

- Dropping duplicate columns

Data Review

application_train

- SK_ID_CURR (unique loan id)
- TARGET (binary, prediction variable)
- info about loan and loan applicant
- history record count
- approved rate
- amount of loans to pay
- usage of different type of loans

SK_ID_CURR

previous_application

- all previous loans history

Credit_card_balance

-monthly balance of credit card
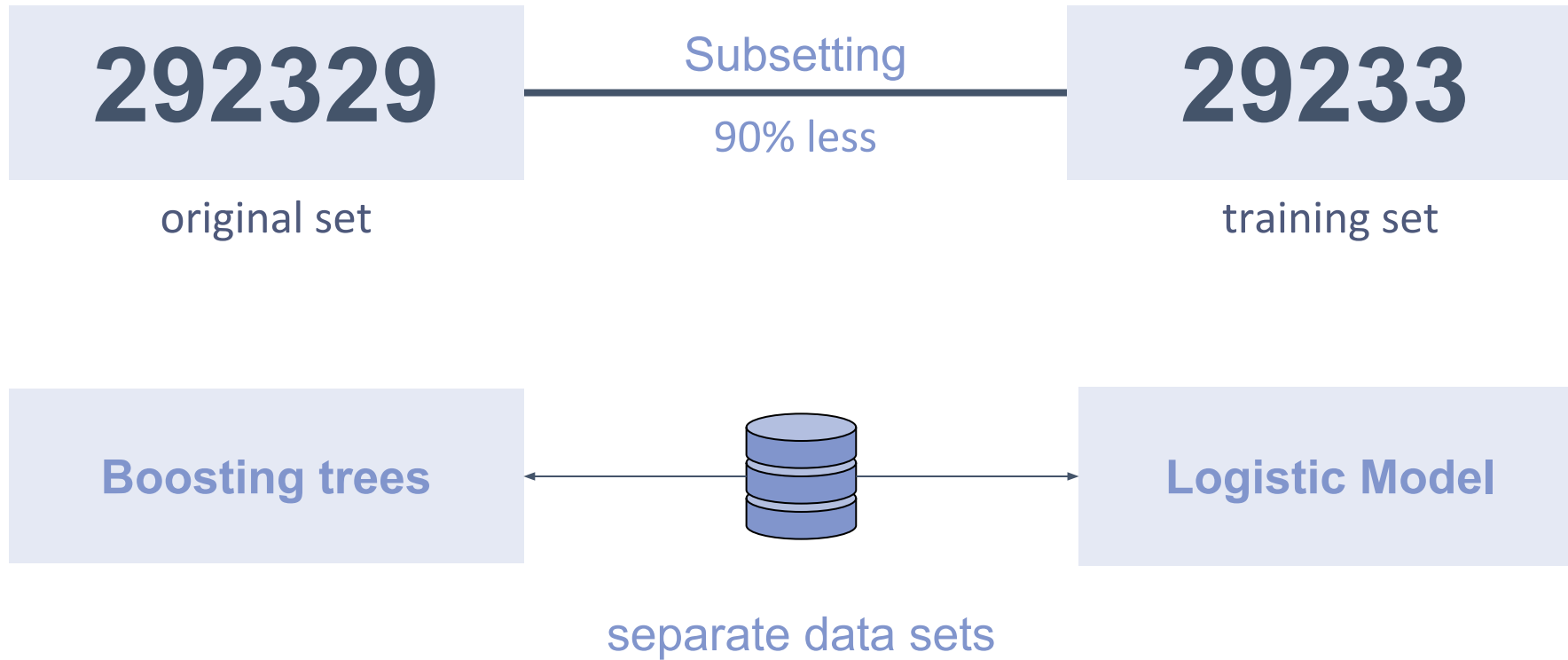-behavioral data

SK_ID_CURR

Models Run without joined external data

	Logistic Unbalanced	Logistic Balanced	Boosting Tree	Boosting Tree with cross validation
Accuracy	91%	78%	92%	71%
Sensitivity	94%	79%	99%	72%
Specificity	19%	50%	1.5%	64%

Models Run with joined external data

	Logistic Unbalanced	Logistic Balanced	Boosting Tree	Boosting Tree with cross validation	Gradient Boosting Tree
Accuracy	92%	65%	92%	71%	83%
Sensitivity	99%	65%	100%	72%	81%
Specificity	0.2%	59%	0%	64%	85%

Data Sets

292329

Subsetting

90% less

29233

original set

training set

Boosting trees

separate data sets

Logistic Model

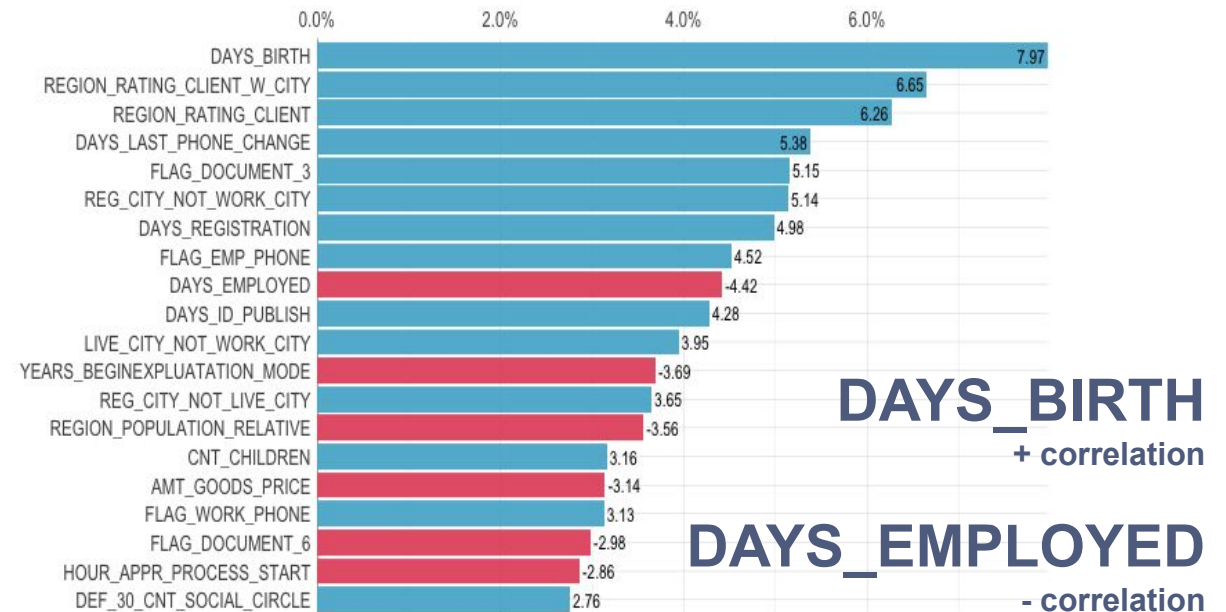# Exploratory Data Analysis

260000+

20000+

Able   Unable

Highly unbalanced data

Our solution:
OVERSAMPLING

Correlations of TARGET [%]
Top 20 out of 46 variables (original & dummy)

Variable	Value
DAYS_BIRTH	7.97
REGION_RATING_CLIENT_W_CITY	6.65
REGION_RATING_CLIENT	6.26
DAYS_LAST_PHONE_CHANGE	5.38
FLAG_DOCUMENT_3	5.15
REG_CITY_NOT_WORK_CITY	5.14
DAYS_REGISTRATION	4.98
FLAG_EMP_PHONE	4.52
DAYS_EMPLOYED	-4.42
DAYS_ID_PUBLISH	4.28
LIVE_CITY_NOT_WORK_CITY	3.95
YEARS_BEGINEXPLUATATION_MODE	-3.69
REG_CITY_NOT_LIVE_CITY	3.65
REGION_POPULATION_RELATIVE	-3.56
CNT_CHILDREN	3.16
AMT_GOODS_PRICE	-3.14
FLAG_WORK_PHONE	3.13
FLAG_DOCUMENT_6	-2.98
HOUR_APPR_PROCESS_START	-2.86
DEF_30_CNT_SOCIAL_CIRCLE	2.76

DAYS_BIRTH
+ correlation

DAYS_EMPLOYED
- correlation
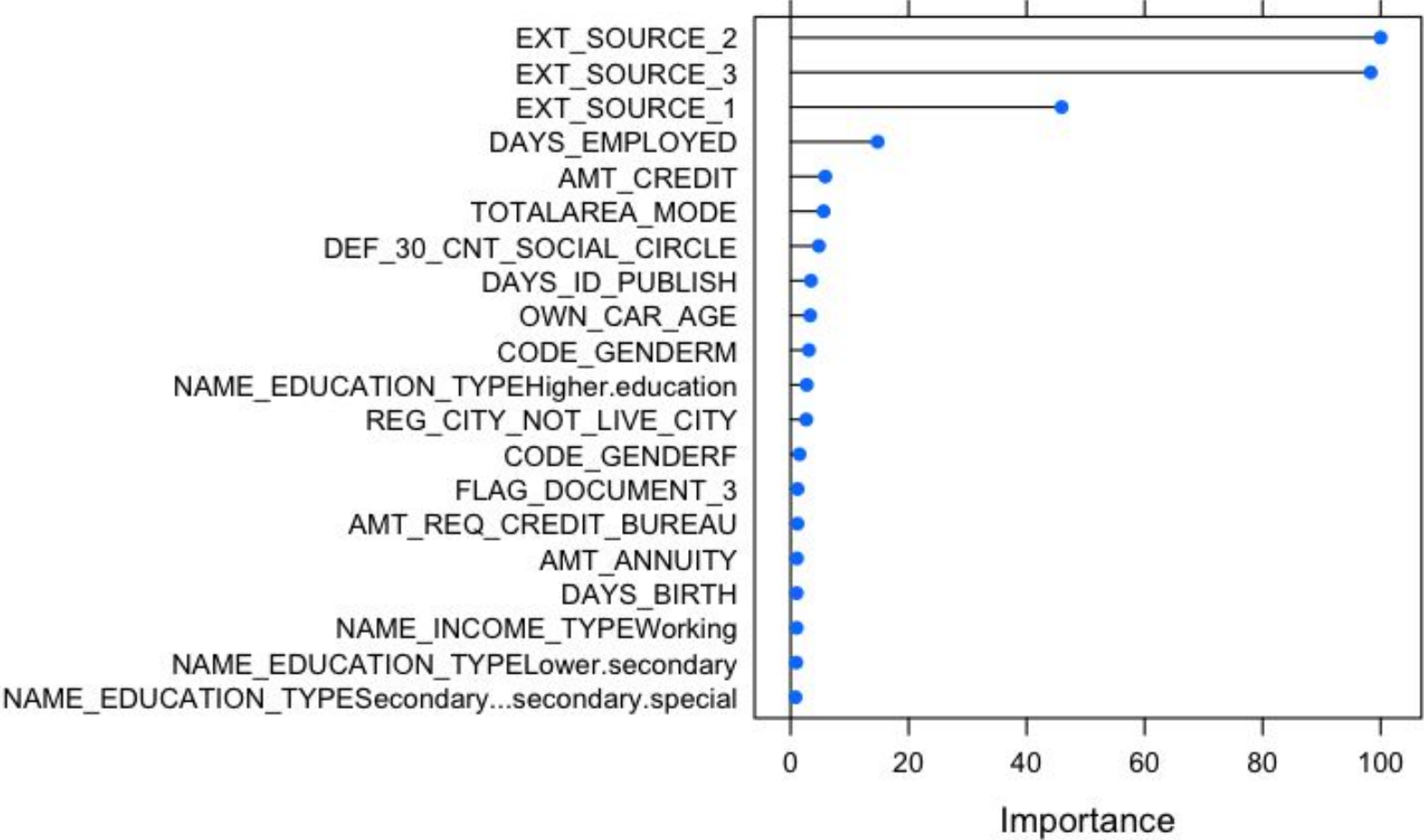
Models Used

Unbalanced Logistic Regression
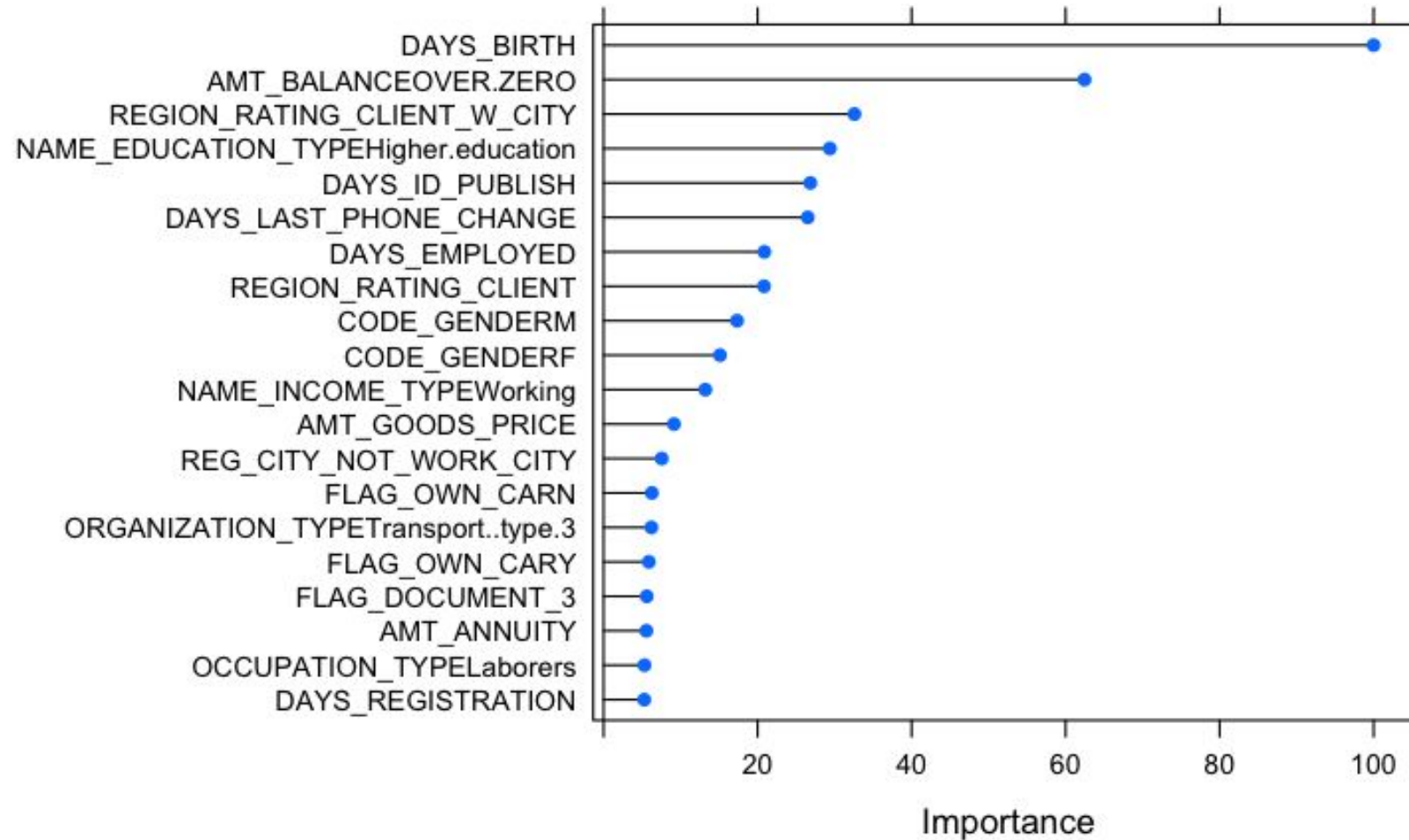
Balanced Logistic Regression

Gradient Boosting Tree

Boosting
Tree

Boosting Tree with cross validation

Features of importance of Boosting Tree without joined external data

Features of importance of Boosting Tree with joined external data

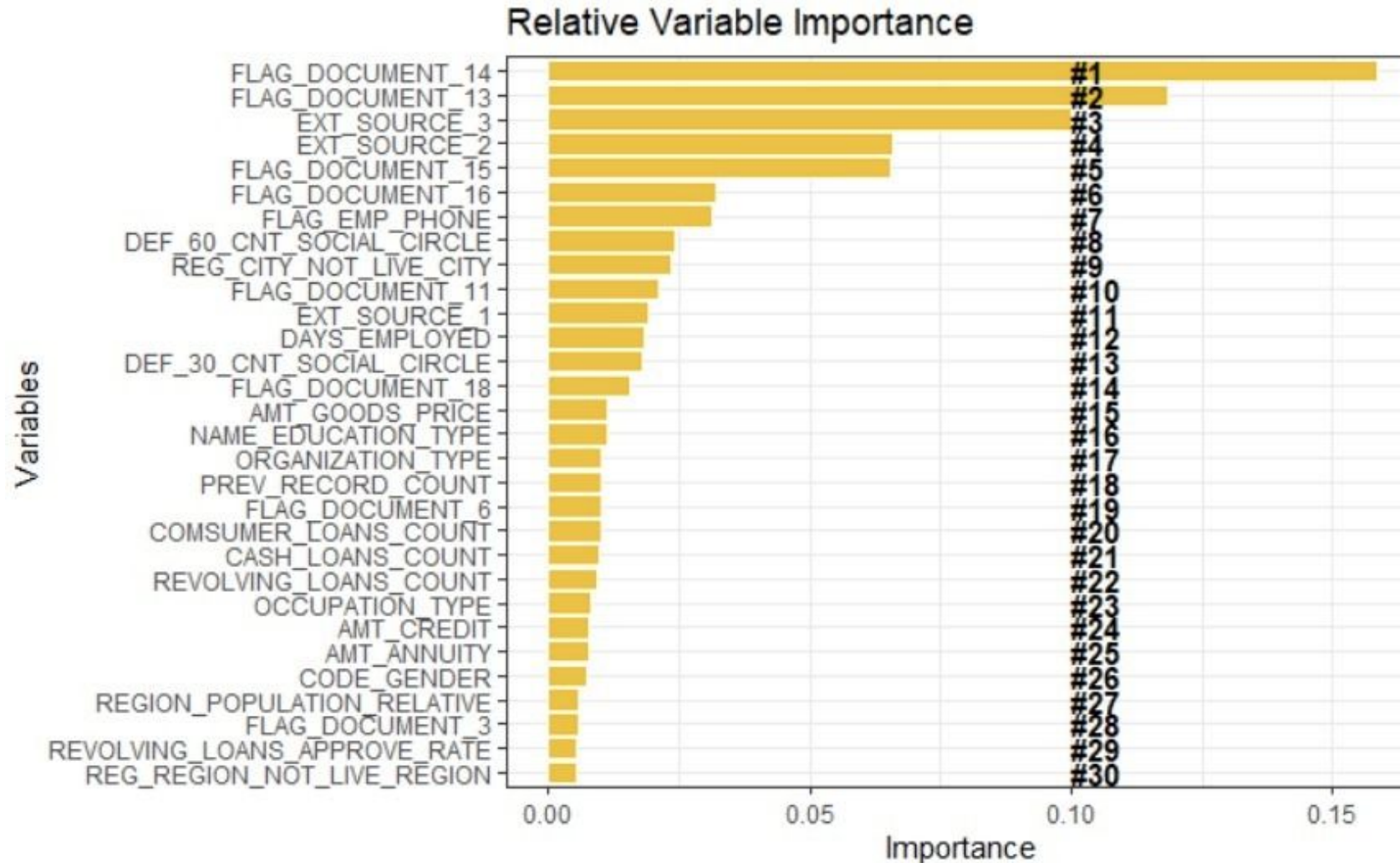# Model Chosen: Gradient Boosting Tree

83%
Accuracy

81%
Sensitivity

85%
Specificity

Features of importance of Gradient Boosting Tree

Relative Variable Importance

LoanProof
Get your Loan Approved
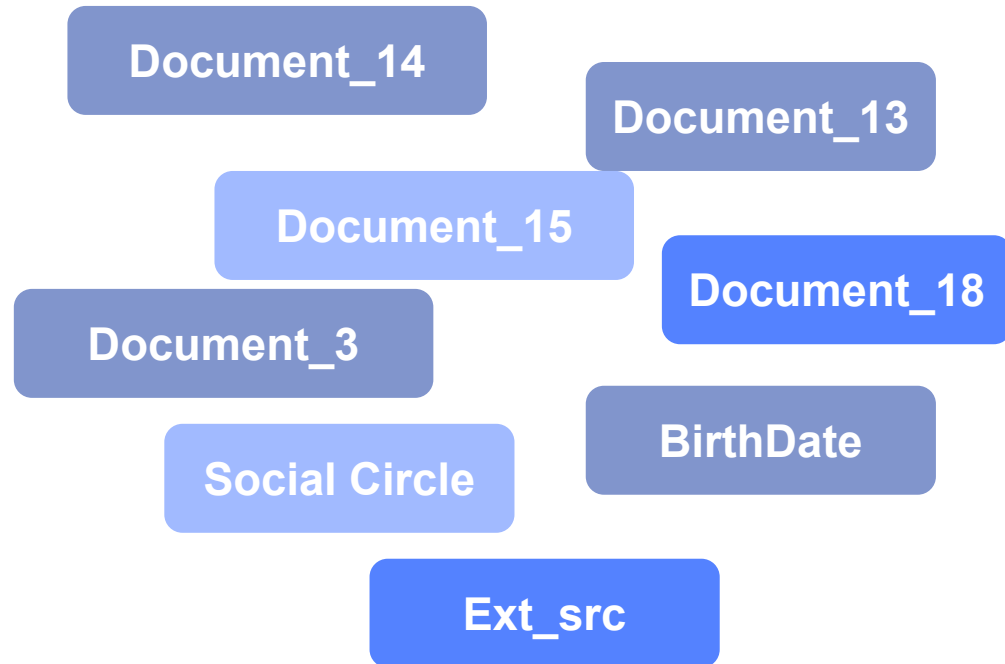
Three things to focus:

External sources

Documentations

background

Things we can help you!

Service : Document preparation

Predictive analysis shows the following documents are a priority:

Document_14

Document_13

Document_15

Document_18

Document_3

BirthDate

Social Circle

Ext_src

Our assistance include:

- Important forms curations

- Forms identification

- Top-level writers

- Experienced consultants

Service: Background Evaluation
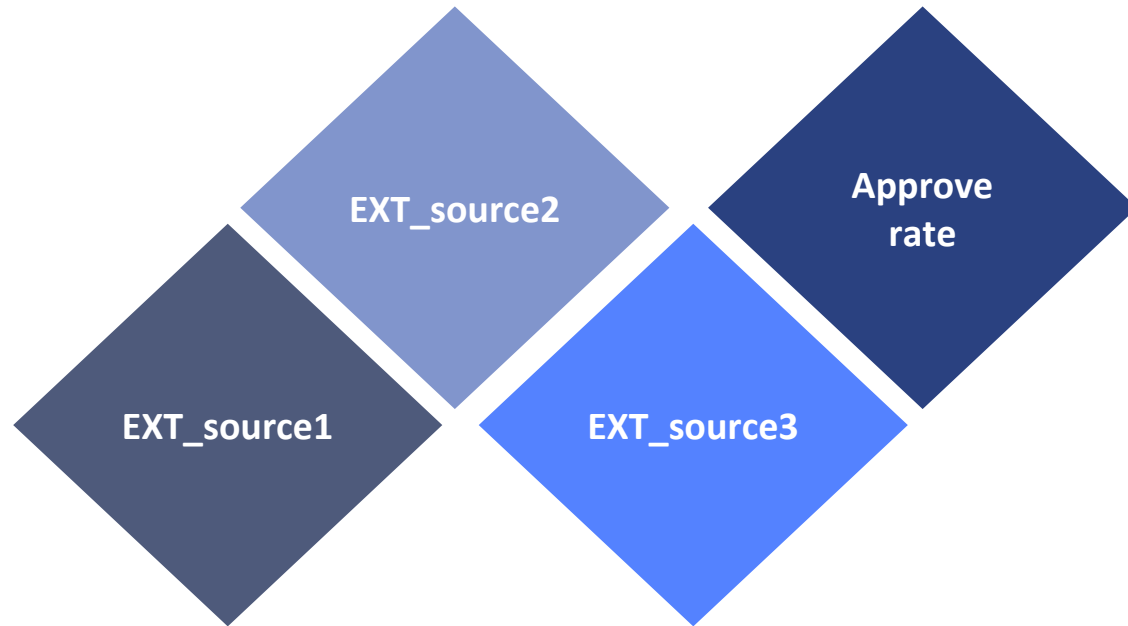
Your background matters to get a loan

How we can help:

Living Place	
	Employment
Credit Report	
	Property

- Finance expert

- Property consultants

- Career counsellors

- Secure handling of sensitive data
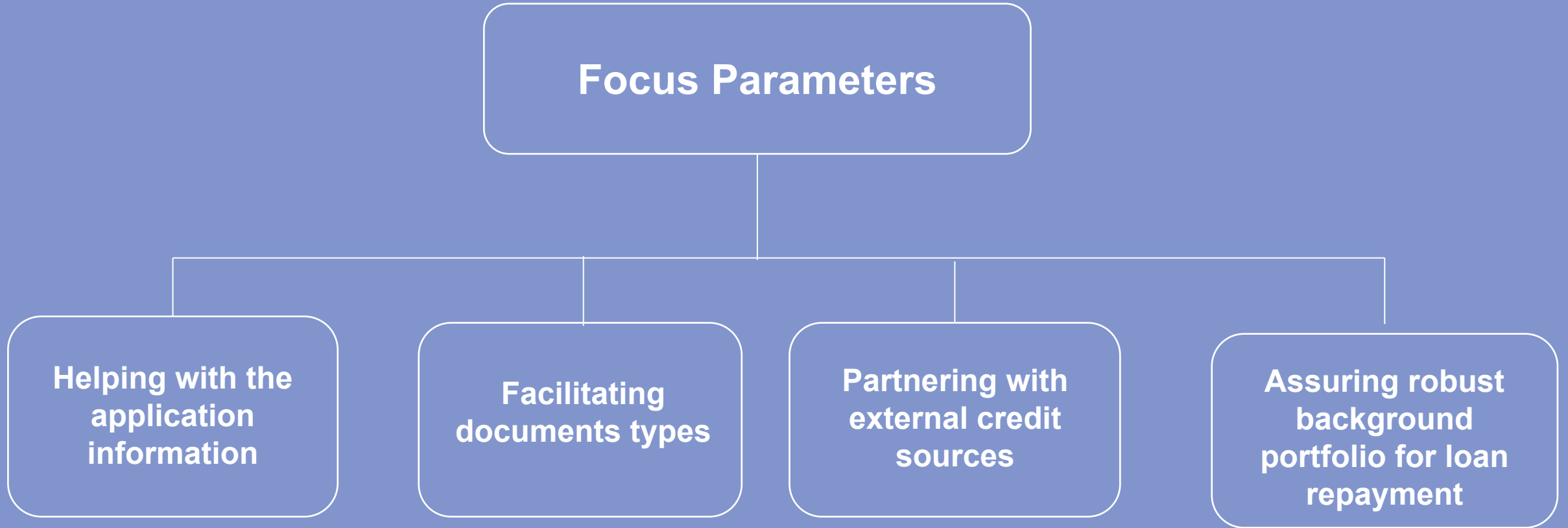
Service: Credit Evaluation

Banks are highly depends on external source to estimate your pay back ability.

How we can help:

- Good relationship with source institution

- Techniques to nail credit evaluation

- Portfolio boosting

Our Success Relies on

Focus Parameters

Helping with the application information

Facilitating documents types

Partnering with external credit sources

Assuring robust background portfolio for loan repayment

Thank you!