KickStarter Increases Profit

By Jakolien den Hollander & Michael Leder
November, 5th 2020

KICKSTARTER

Agenda

Successful Projects

Failed Projects

01

Introduction

Stakeholder & Business Case

Stakeholder:

Stakeholder of this data science project is the **Kickstarter C-Level Management**.

Kickstarter's Business Model:

The **more successful** a project is, the **more money** can be raised from backers (amount pledged), the **more profit** can be gained for Kickstarter (KS charges **5% fee** on the total pledged amount).
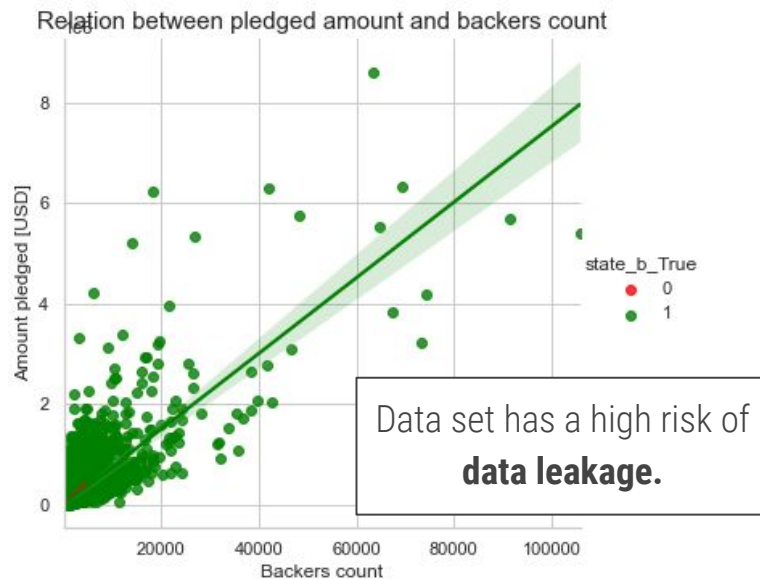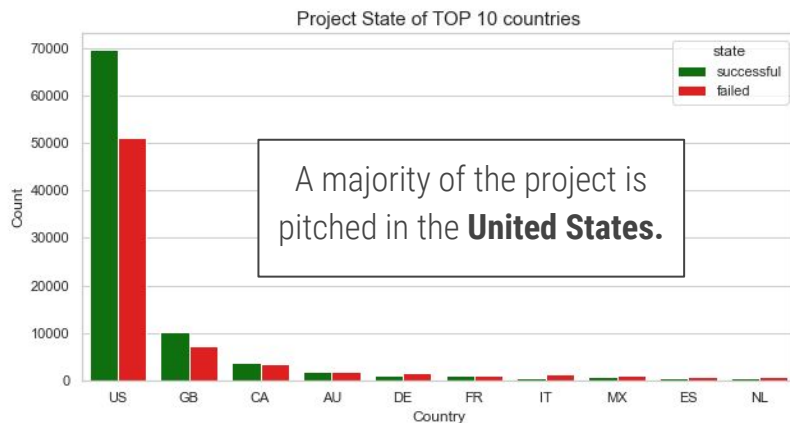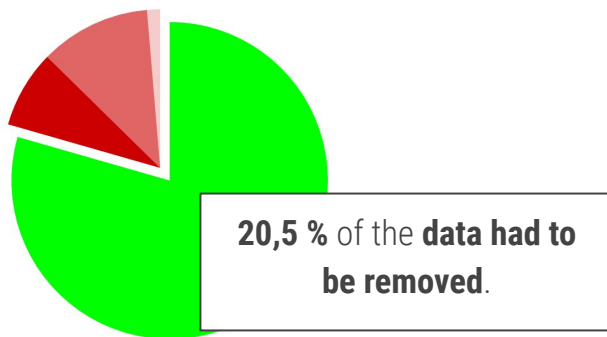
The underlying question of the business case:

A. **Main question:** How likely is it that a project is successful on Kickstarter based on an assessment of factors that contribute to success?
B. **Additional question:** How much money can be pledged based on an assessment of factors that contribute to success?

A Brief Data Overview

Data Cleaning

- Trimmed Set
- Wrong State
  [Live, Suspended, Canceled]
- Duplicates
- Goal
  [> 2 Mio. $; < 100 $]

20,5 % of the **data had to be removed**.

Relation between pledged amount and backers count

Data set has a high risk of **data leakage.**

Project State of TOP 10 countries

A majority of the project is pitched in the **United States.**

02
Models Used

Models Used to Predict

For Classification*

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. Adaboost

Target → State (success or failure)

For Regression**

1. Multi Linear Regression
2. Random Forest

Target → Pledged amount

* Main question | Predict the success or fail of a project.
** Additional question | Predict the amount of pledged money.

KICKSTARTER

03

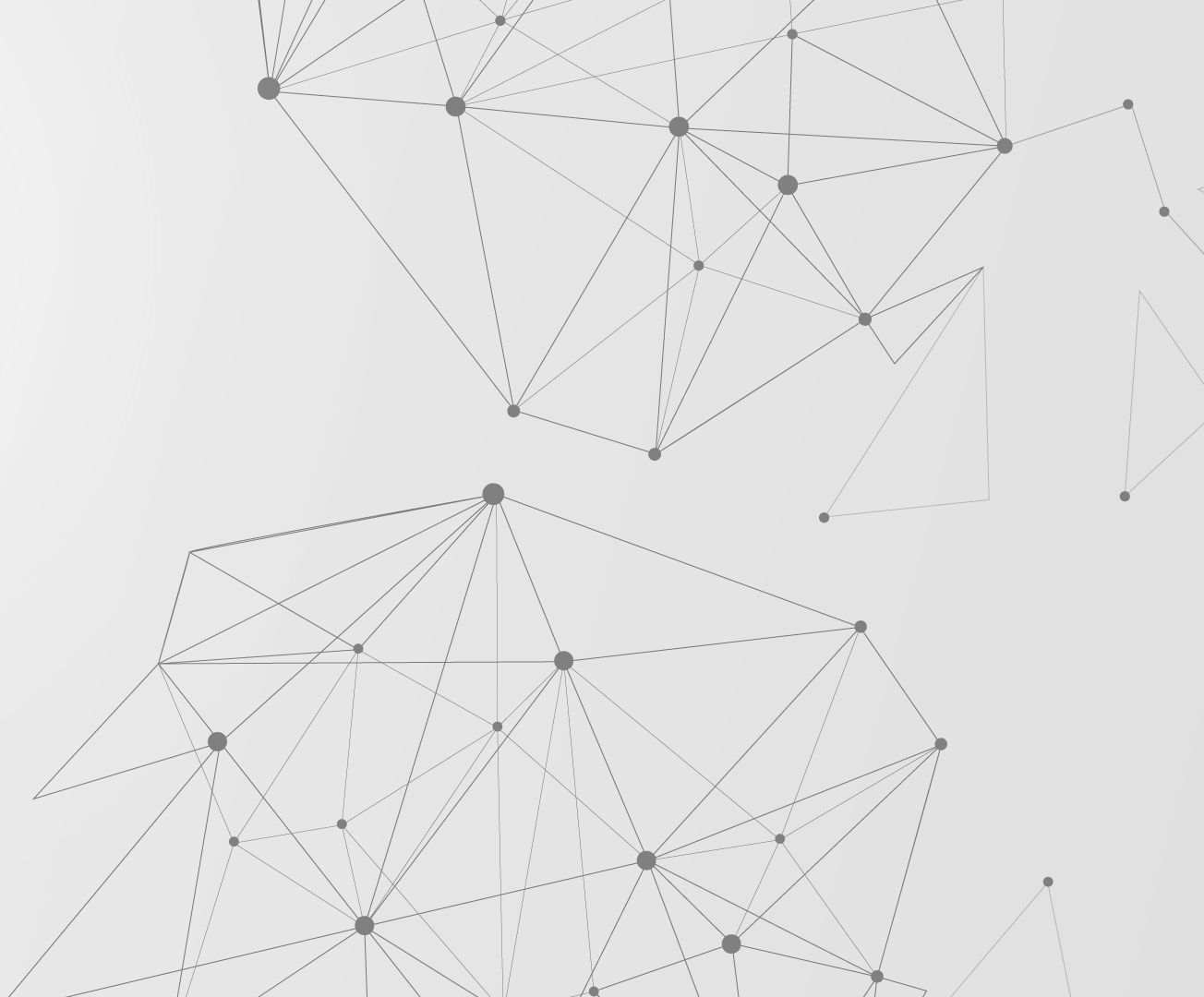# Model Evaluation

Performance metrics

For Classification

1. F1-Score
2. Accuracy

For Regression
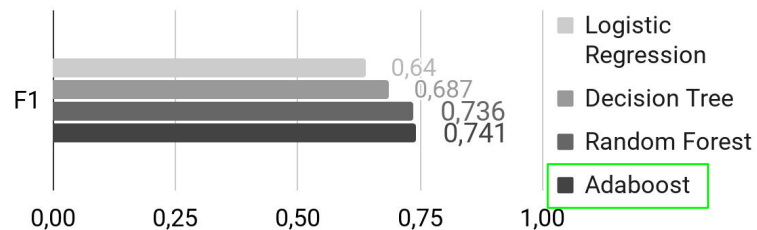
1. R^2
2. Root Mean Squared Error (RMSE)
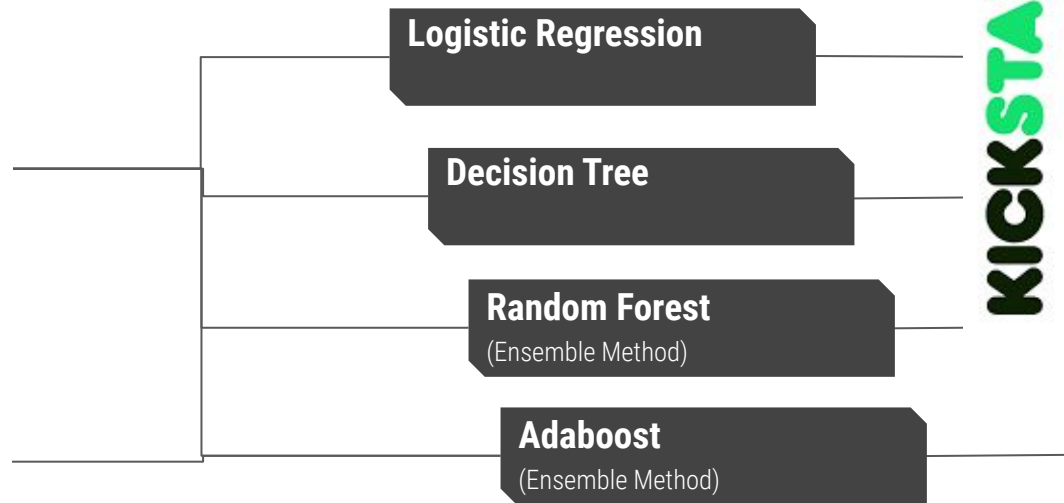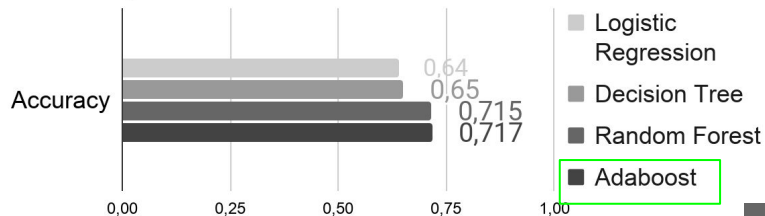
KICKSTARTER

04
Results

Our Numbers - Classification

F1-Score

F1
- Logistic Regression
- Decision Tree
- Random Forest
- Adaboost

0,64
0,687
0,736
0,741

0,00   0,25   0,50   0,75   1,00

Accuracy

Accuracy
- Logistic Regression
- Decision Tree
- Random Forest
- Adaboost

0,64
0,65
0,715
0,717

0,00   0,25   0,50   0,75   1,00

Logistic Regression

Decision Tree

Random Forest
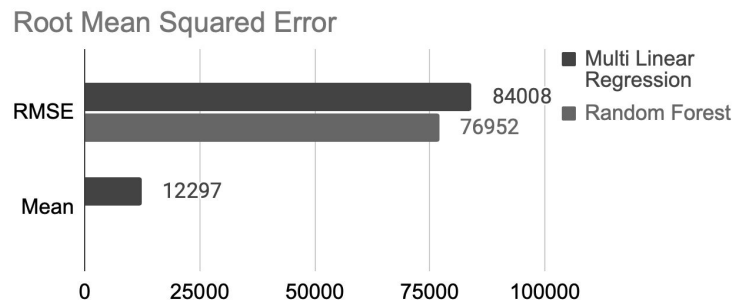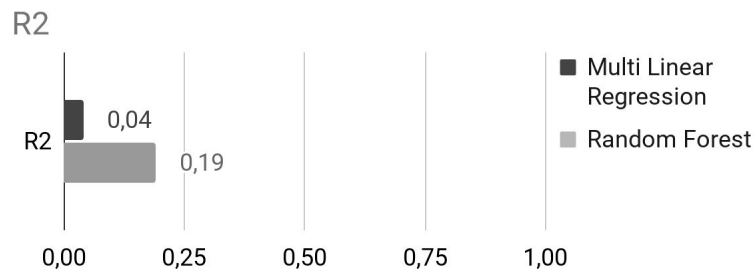(Ensemble Method)

Adaboost
(Ensemble Method)

Performance:
adaboost > rand. for. > dec. tree > log. reg.
Succes: rand. for.; **Failure:** adaboost
Features: category, goal, duration

Conclusion

11

Our Numbers - Regression

R2

- Multi Linear Regression
- Random Forest

0,04
0,19

R2

0,00   0,25   0,50   0,75   1,00

Root Mean Squared Error

- Multi Linear Regression
- Random Forest

RMSE   84008
        76952

Mean   12297

0   25000   50000   75000   100000

Multi Linear Regression

Random Forest
(Ensemble Method)

With given data and the applied models **prediction of the pledged amount** is **not possible** respectively **not accurate** enough.

Conclusion

05
Future Work

Future Work

A. **Main question:** How likely is it that a project is successful on Kickstarter project based on an assessment of factors that contribute to success?

Increase model performance

1

B. **Additional question:** How much money can be pledged based on an assessment of factors that contribute to success?

Try **different models** suitable for regression
1. Decision Tree
2. RIDGE Regression
3. LASSO Regression

2

Engineer **additional features**

Engineer feature to see if creators already created a project before **(multi creators)**

3

THANKS

Any questions?

Sources (Data Set)