PREDICTING KICKSTARTER SUCCESS

Empowering Investors and Founders
with Data-Driven *Insights*

THE TEAM

FRANZISKA RAUSCHGOLD

Data Scientist

ALEXANDER SIMAKOV

Data Scientist

WHAT IS KICKSTARTER

Kickstarter is a crowdfunding platform where creators **raise funds to bring their projects** to life. Backers pledge money to support projects, which only receive funding if they meet their financial goal.

THE PROBLEM

Founders and investors often **struggle to predict** which Kickstarter campaigns will succeed, leading to **uncertainty and financial risk**. Making informed decisions is crucial but challenging without reliable data-driven insights.

OUR PRODUCT

Our machine learning model can **analyze patterns** from past Kickstarter campaigns to accurately **predict the likelihood of success.** This helps founders and investors make more **informed decisions**, **reducing risk** and increasing the chances of **funding successful projects.**

MODELING PHASE

DATASET

Our Dataset consists of **200k instances** from successfully launched, failed, canceled, suspended and at that point live campaigns from **2009-2018**. While **3,4%** were live data, **40%** is from failed projects (incl. suspended and canceled) and **56%** is from successful projects.

DATA CLEANING

When it comes to classes of our target variable, our **dataset is rather balanced** but we are facing some **dependent variables** that we need to exclude for our model. Additionally, also **dummies** were needed for categorical data.

TRAINING & TUNING

We tune the model to improve **accuracy** as the cost of false positives, e.g. predicting a project as successful but it fails is high.

FINDINGS

01

SUCCESSFUL COUNTRIES

While most data comes from US, we see the highest successful-rate in Honkong (70%) and Luxemburg (66%). In general, most successful projects have been realised in the US and Spain.

02

PROMISING CATEGORIES

Categories with most successfully implemented projects are: (1) Product Design, (2) Tabletop Games and (3) Software.

03

TIME DOES NOT MATTER

We don't see months where in general projects are more successful, the only thing we can see from the data is that in december less projects are created.

OUR MODEL

We evaluated the performance of 3 algorithms:
(1) logistic regression
(2) knn
(3) random forest

For the prediction we used the following features:
- Creation date
- Fundraising goal
- Country
- Category
- Communcation
- Starrable status
- Staff pick marketing

After tuning we see the best performance on a **knn-model** with a **neighborhood size of 14** using **distance weight** and **manhattan distance** and the random forest.
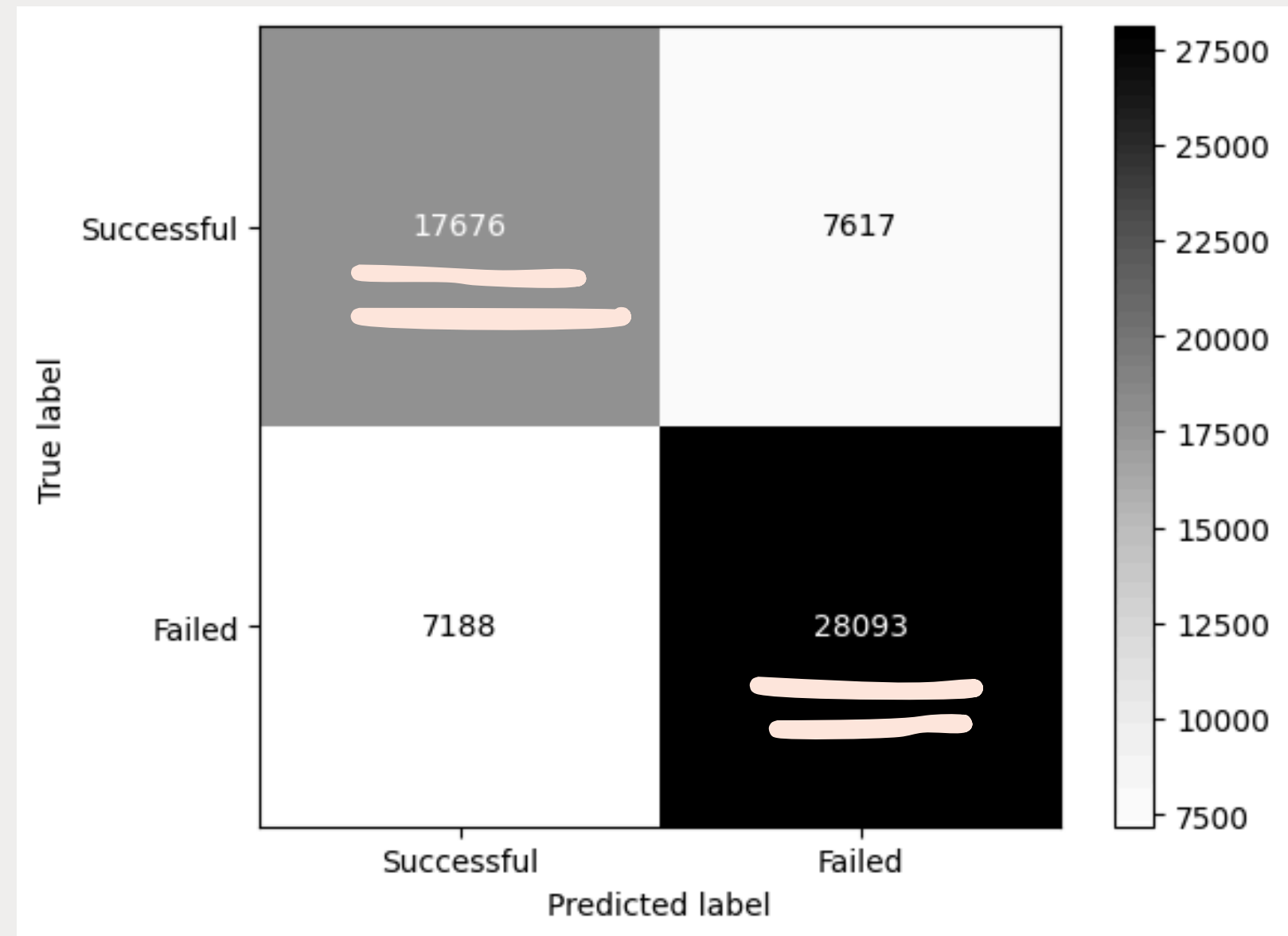
RECOMMENDATIONS

If you find your project to be predicted as failed, check if you can **adjust your starting date, invest in staff pick marketing** or least, adapt your product to a more promising **category**. Also evaluate your **cost planning** again to see if you can decrease your campaign goal.

Feature Importance for KNN Model
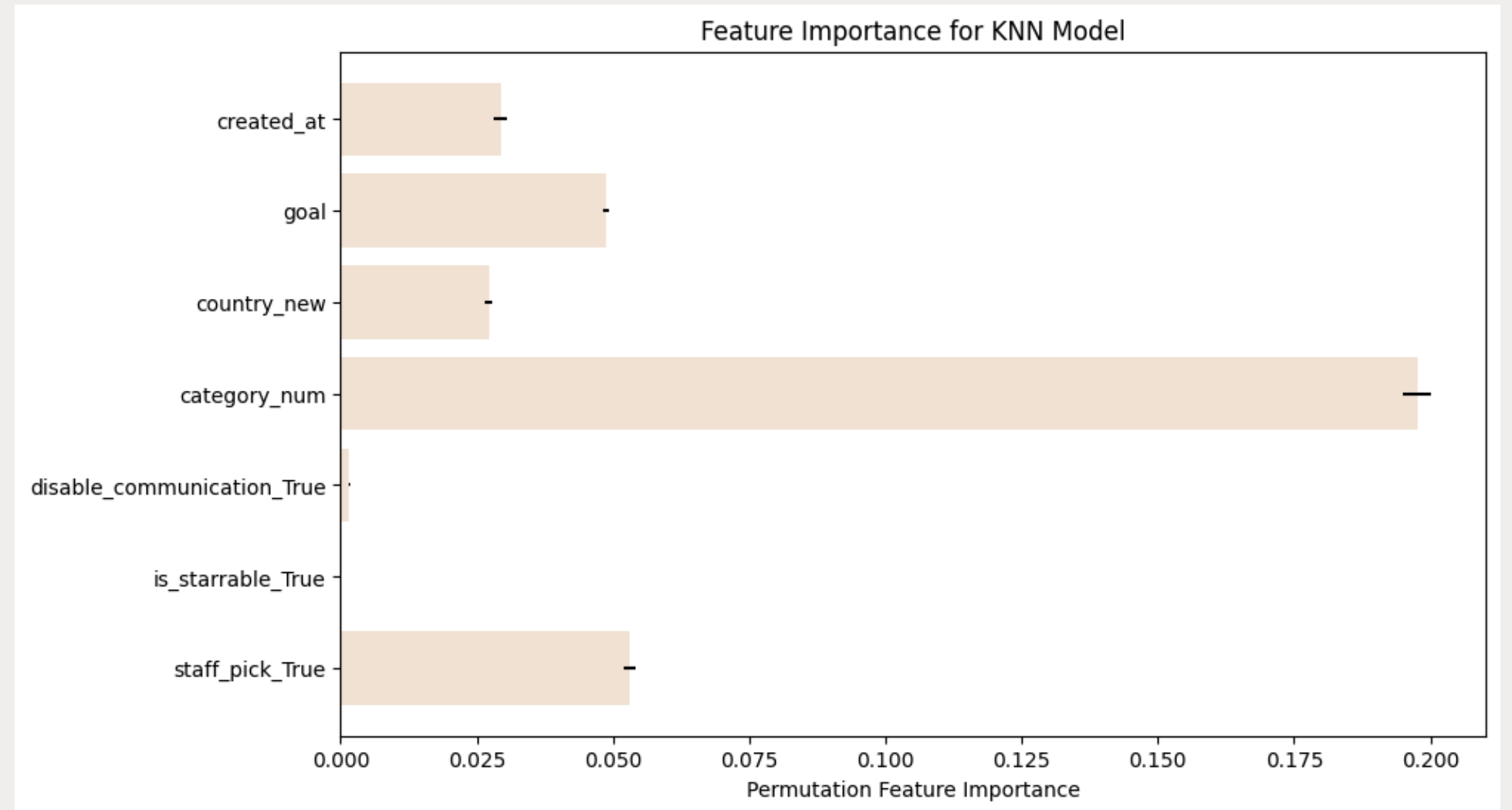
FUTURE WORK

01
IMPROVE MODEL FURTHER
As our dataset also has data from 'live' projects, these can be used as unseen training data to predict, if the projects will be successful or not and ensemble methods can be used to improve accuracy further.

02
SEASONALITY
Suggest additional services or products for founders with seasonal products to add to their portfolio to ensure constant revenue throughout the year.

03
PREDICT ROI IN MODEL
For investors it might be specifically interesting to also predict the precise return on invest.

THANK YOU