# Loan Approval Process Automation

Arash Aghagol

# Table of Contents

# 01

**Objectives**

# Objectives

- ❖ Reliable models

- ❖ Influential factors

- ❖ Actionable advice

- ❖ Future works

# 02

## Overview

# General Information

❖ A Bank in Czech Republic

❖ Scope of Data

    ❖ Transaction

        ❖ Purpose

        ❖ Amount

        ❖ Category

  ❖ District Info

  ❖ Loan

  ❖ Credit Card

  ❖ Customers Personal Info

# Challenges

❖ Dataset in Czech language
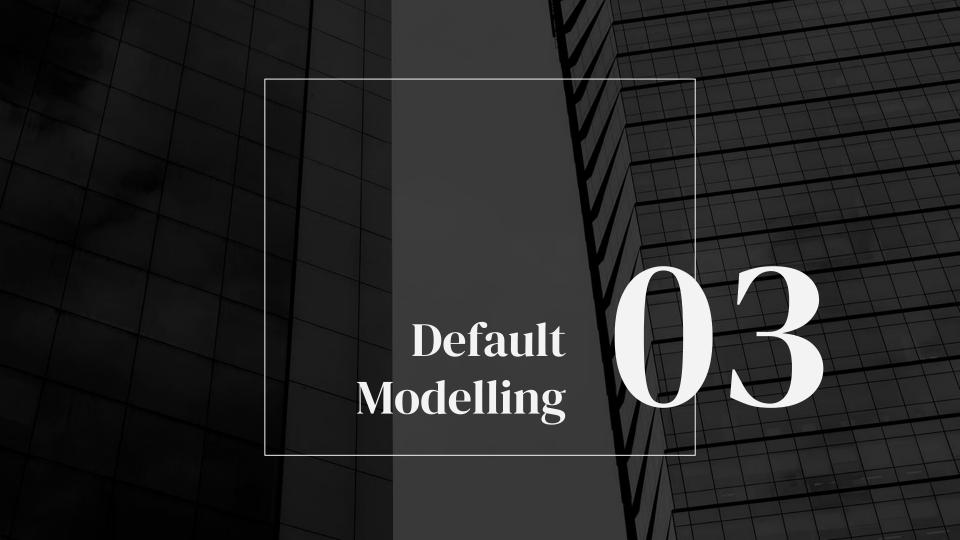❖ Coded columns names
❖ Transaction level
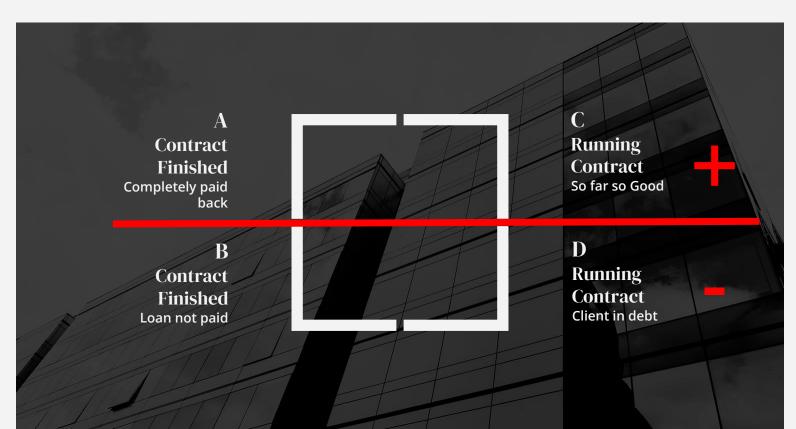❖ Missing Values

# Solutions

❖ Found some resources
❖ Aggregation

# Methodology

- ❖ **Synthetic Variables**
- ❖ **Existing Variables**
- ❖ **Categorical Variables**

Default Modelling

03

# 4 Loan Status

**A**
**Contract Finished**
Completely paid back

**B**
**Contract Finished**
Loan not paid

**C**
**Running Contract**
So far so Good

**+**

**D**
**Running Contract**
Client in debt

**−**

# Precision = 1 !

## Logistic Regression

- ❖ Precision : 0.9916
- ❖ Recall: 0.9672

## Support Vector Classifier
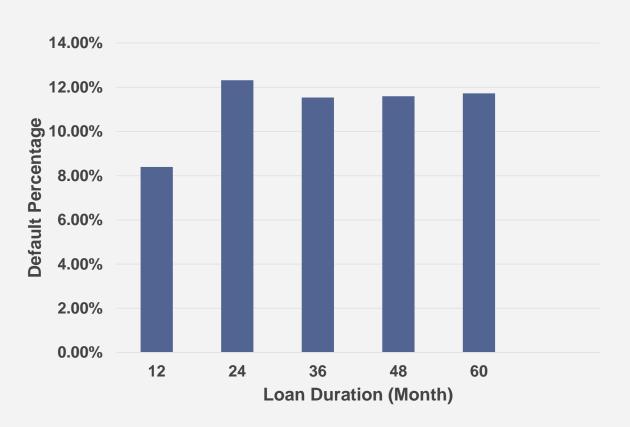
- ❖ Precision : 1
- ❖ Recall: 0.9672

# Top Positive Factors

- ❖ Average salary (A11)
- ❖ Credited interest
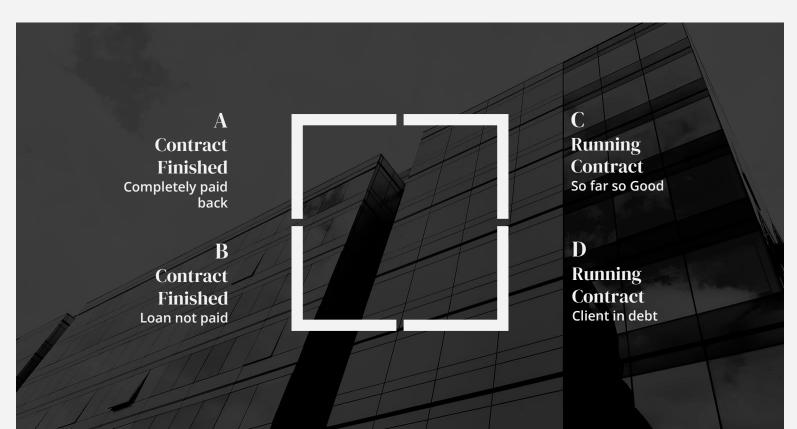- ❖ Received a money transfer

# Top Negative Factors

- ❖ Sanction interest
- ❖ Gender
- ❖ Loan duration

# 04

## Model Extension

# 4 Loan Status

**A**
**Contract Finished**
Completely paid back

**B**
**Contract Finished**
Loan not paid

**C**
**Running Contract**
So far so Good

**D**
**Running Contract**
Client in debt

# Benefits

Categorize customers

Various loan package

More potential revenue

# Pretty Good Results

❖ Precision and Recall =0.88

❖ Top influential positive feature for B: A12

❖ Top influential positive feature for D: Money transfers

# Summary & Recommendations

**05**

# Summary

- ❖ **2 Classification models**
- ❖ **Quality of the models**
- ❖ **Important variables**

❖ **Models usage**

❖ **Risk Policy**

❖ **Design loan packages**

# Future Works

**06**

# Future Works

❖ **More Data**

❖ **Clustering Analysis**

❖ **Credit Card**

# Thanks

Does anyone have any questions?

# Appendix

# Confusion Matrix

# Combinations of metrics

| | Micro | Macro |
|---|---|---|
| **Precision** | $\dfrac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i+fp_i)}$ | $\dfrac{\sum_{i=1}^{l}\frac{tp_i}{tp_i+fp_i}}{l}$ |
| **Recall** | $\dfrac{\sum_{i=1}^{l} tp_i}{\sum_{i=1}^{l}(tp_i+fn_i)}$ | $\dfrac{\sum_{i=1}^{l}\frac{tp_i}{tp_i+fn_i}}{l}$ |