UNIVERSIDADE DE PERNAMBUCO

KDMILE 2020

FACEPE
Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco

Impact of Unusual Features in Credit Scoring Problem

Luiz F. V. Vercosa

Rodrigo C. Lira

Rodrigo P. Monteiro

Kleber D. M. Silva

Jailson O. L. Magalhaes

Alexandre M. A. Maciel

Byron L. D. Bezerra

Carmelo J. A. Bastos-Filho

Before

ANALYST

DECIDES

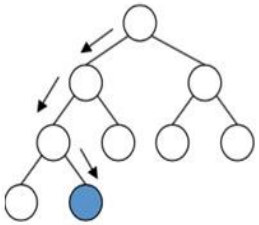
OR

GRANT

DENY

After

ANALYST



MODEL

DECIDES

OR

GRANT

DENY

Features

- Hard features

INCOME

SAVINGS

MARITAL

EMPLOYMENT

AGE

- Soft features

GEOLOCATION

WEB

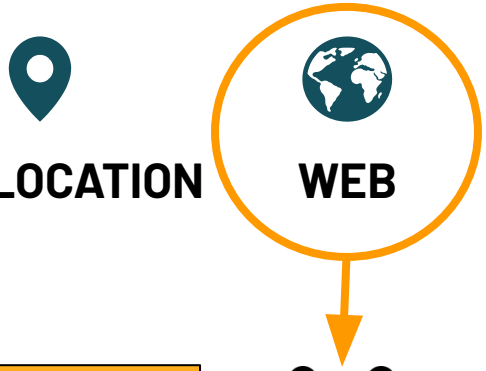
DEMOGRAPHIC

HISTORY

SOCIAL PROGRAM

Previous Works

- Soft features



Our Contribution

- Soft features

Vercosa *et al.* 2020


GEOLOCATION


WEB


DEMOGRAPHIC


HISTORY


SOCIAL PROGRAM

Niu *et al.* 2019

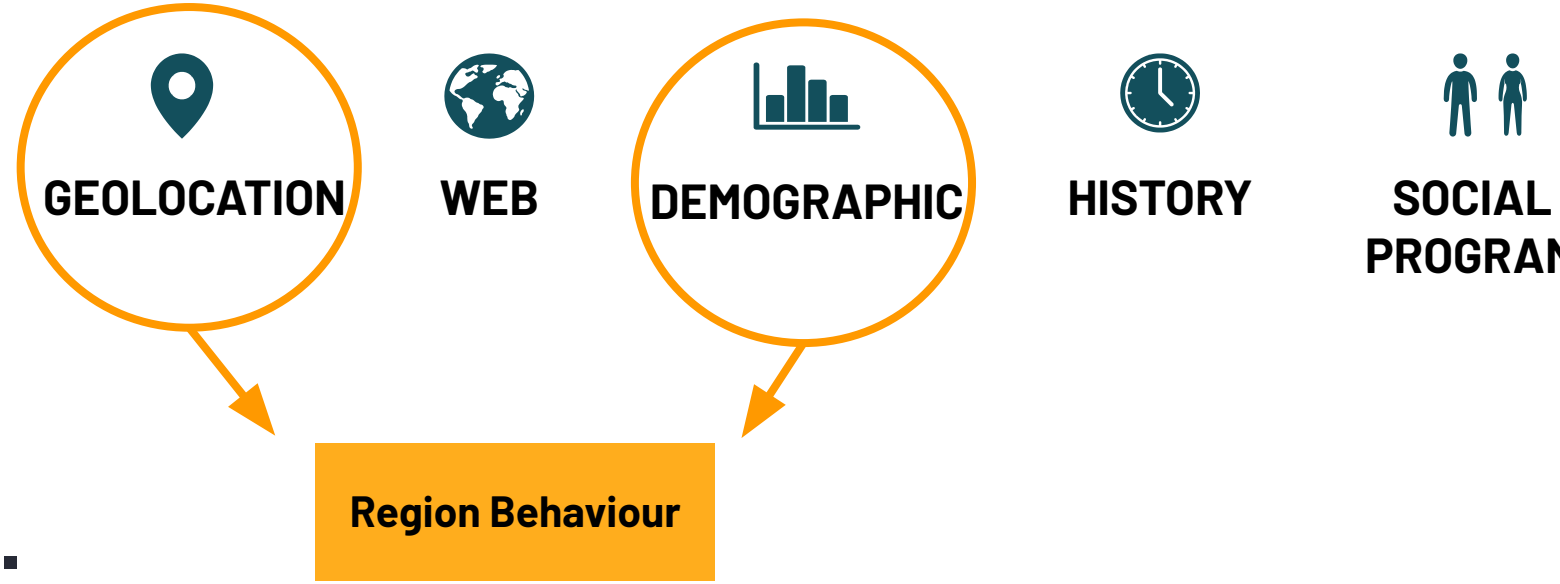

SOCIAL NETWORK


PSYCHOLOGY

Liberati *et al.* 2018

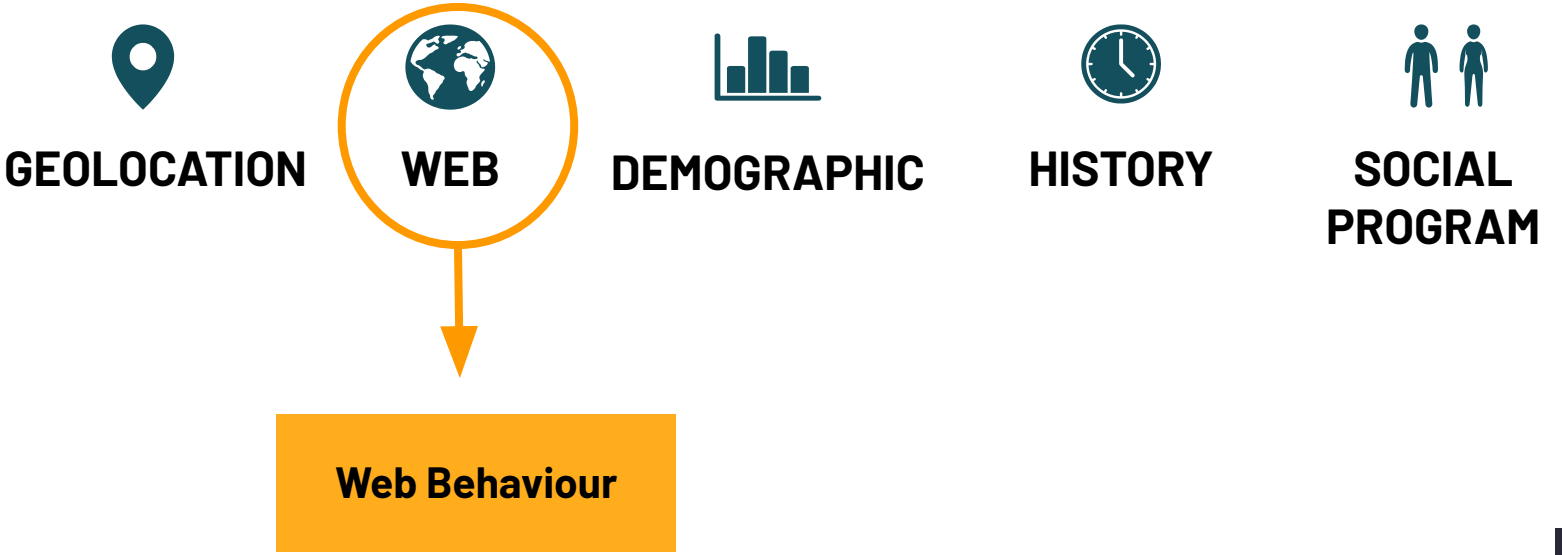
Hypothesis

- Soft features



Hypothesis

- Soft features



Hypothesis

- Soft features

GEOLOCATION

WEB

DEMOGRAPHIC

HISTORY

SOCIAL PROGRAM

Purchase Power

Hypothesis

- Soft features

GEOLOCATION

WEB

DEMOGRAPHIC

HISTORY

SOCIAL PROGRAM

CREDIT CARD GRANTING

CAR INSURANCE

Methodology

- Experiment 1

GOAL

Identify most promising features

METRIC

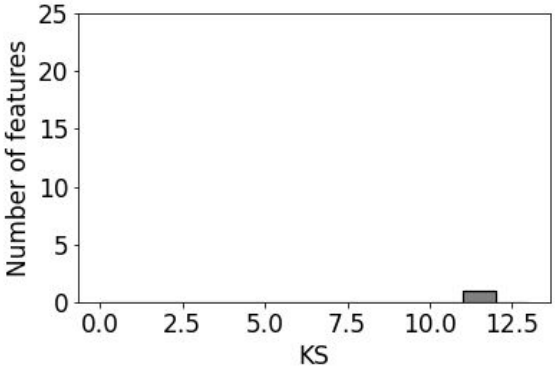
Kolmogorov-Smirnov (KS)

METHOD

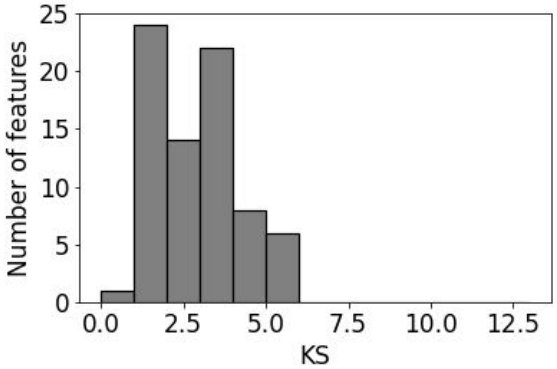
Bivariate Analysis

Results

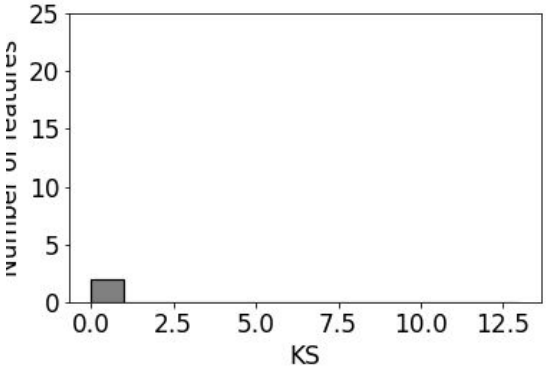
1 REGISTRATION



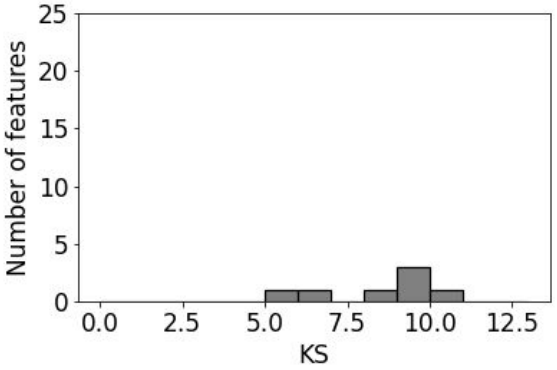
3 GEOLOCATION



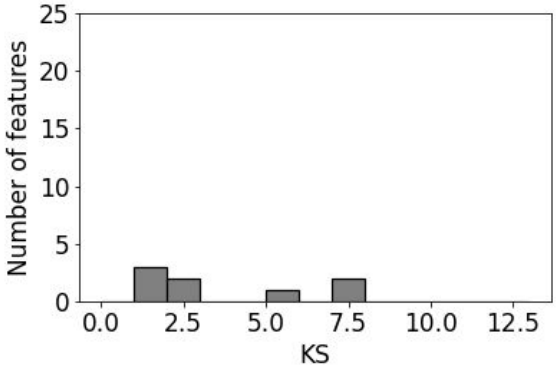
8 GOVERNMENT



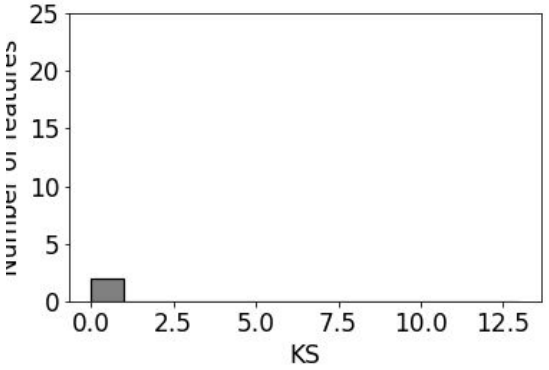
2 HISTORICAL



4 FINANCIAL



9 POLITICS



Methodology

- Experiment 1: part 2

GOAL

Identify most promising features groups

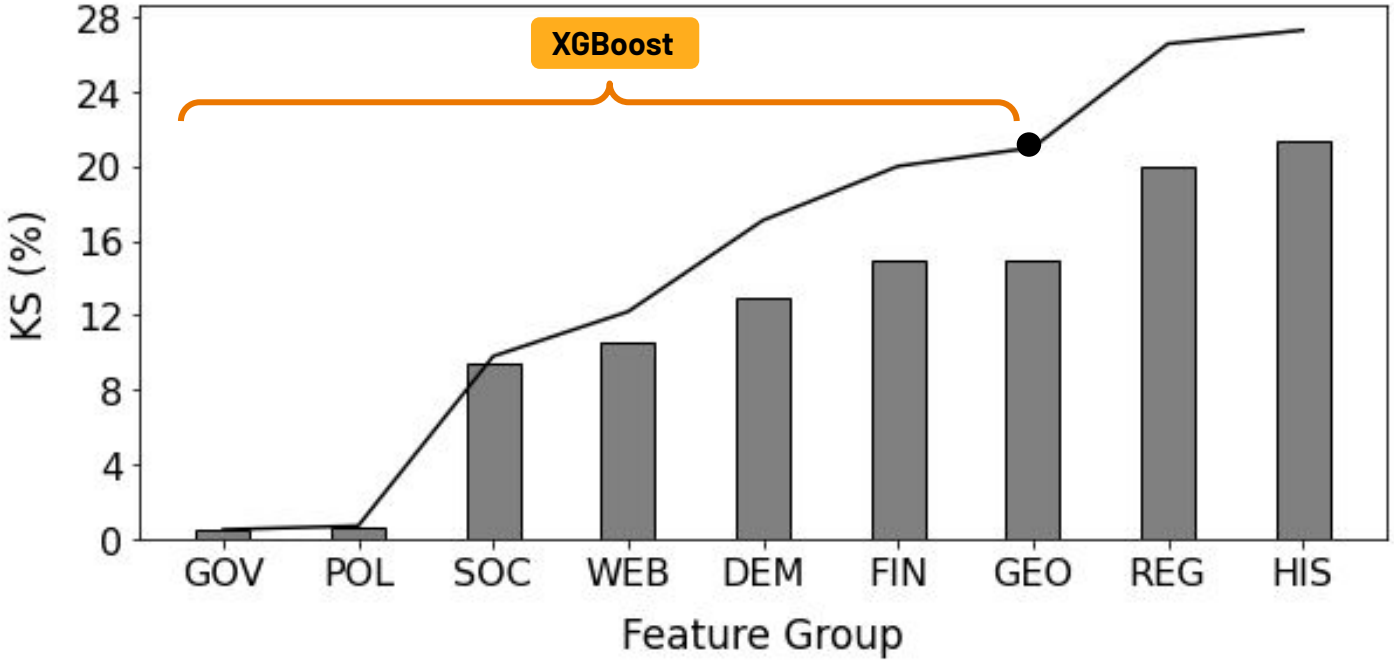
METRIC

Kolmogorov-Smirnov (KS)

METHOD

Extreme Gradient Boosting (XGBoost)

Results



GOV: Government

POL: Politics

SOC: Social Prog

WEB: Web

DEM: Demographic

FIN: Financial

GEO: Geolocation

REG: Registration

HIS: Historical

Methodology

- Experiment 2

GOAL

Identify best performance

METRIC

KS, Area Under Curve (AUC), MSLE, Lift

METHOD

XGBoost, AdaBoost, Multi-layer Perceptron

Methodology

- Experiment 2
  - Lift metric

$$\text{lift} = \frac{\text{GP in 90-100 percentile range}}{\text{GP in 0-10 percentile range} + 1}$$

  - Models fine-tuning
    - GridSearchCV

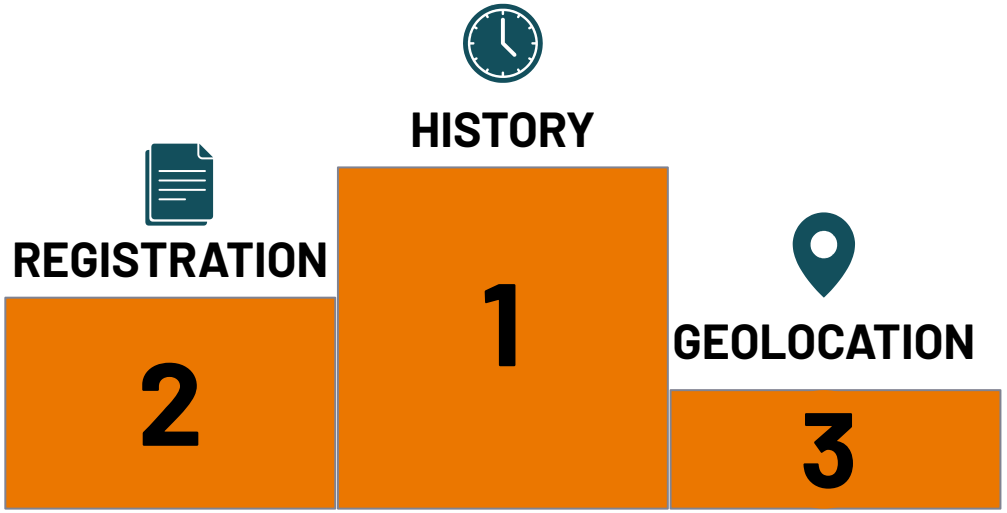
Results

- Experiment 2

Models	Lift	MSLE	AUC (%)	KS (%)
MLP	0.6044	0.12	66.93	24.02
AdaBoost	0.6308	0.1240	68.08	26.34
XGBoost	**0.1099**	0.1099	**68.81**	**27.24**
Company	0.6296	**0.1095**	68.30	26.61

Conclusions

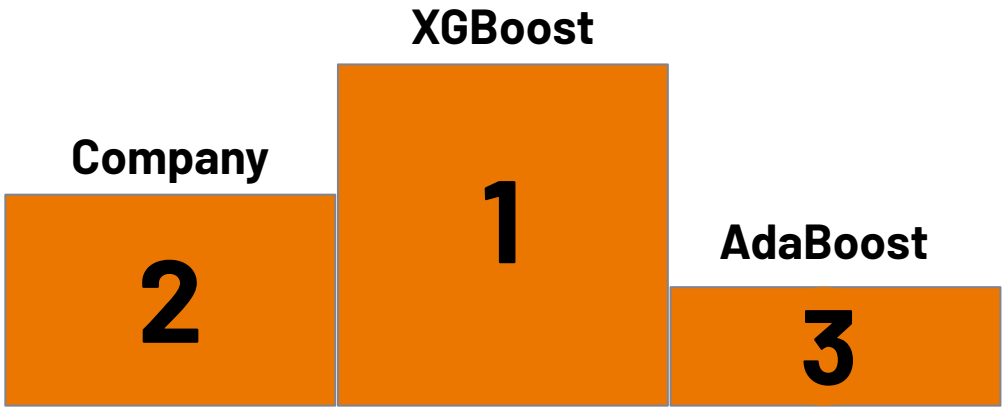
- Promising feature groups



Financial, Demographic, ..., Politics, Government

Conclusions

- Promising models



Limitation

- Available Registration features

Future Work

- Correlation of features
- Feature selection