FINAL PROJECT MODULE 3

Executive Summary

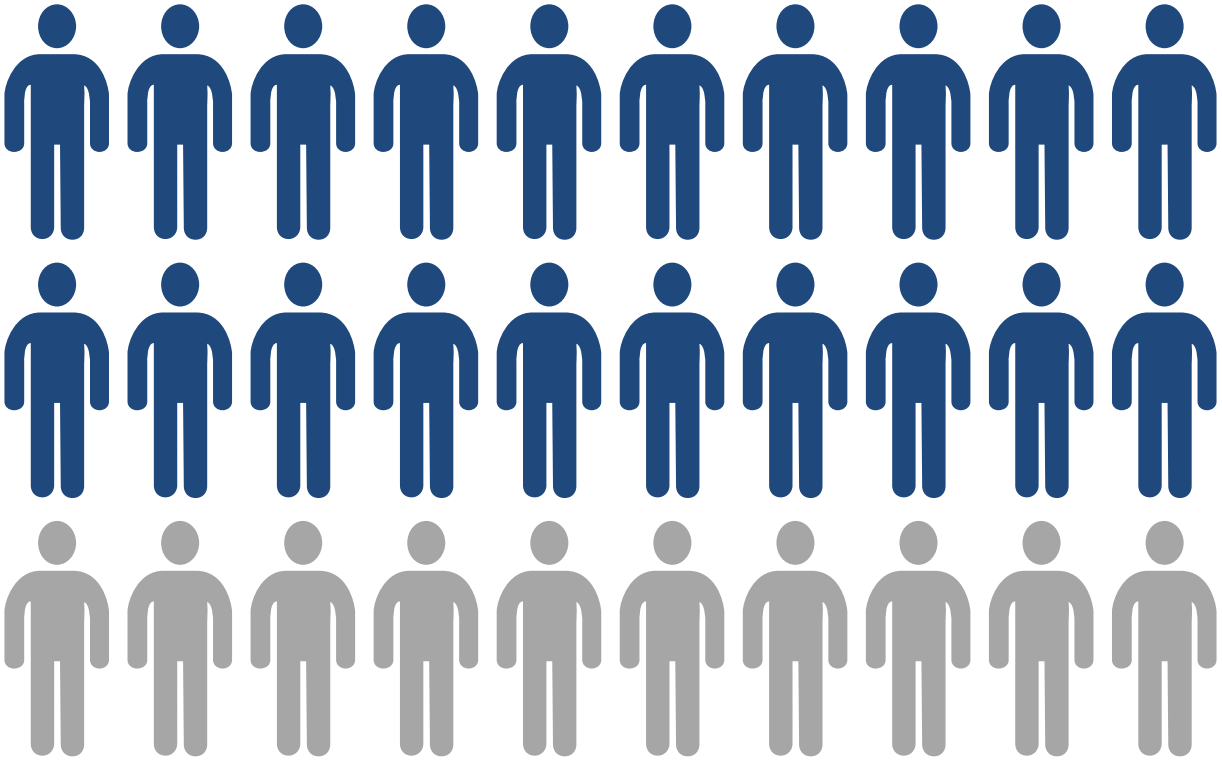
Made with PASSION by Britta Bettendorf
>>github.com/BrittaInData >>datenlage.blog

BASIC INCOME DATASET

- Representative survey by Dalia Research
- Conducted in April 2016
- Spans 28 EU member states
- Covers 9,649 records x 15 columns
- Includes demographics* and opinions**
- Available on kaggle.com

* age, gender, education, living area (city or rural), having children, having a full time job
** awareness of concept, effect on work choices, convincing arguments for and against

IF A REFERENDUM ON INTRODUCING BASIC INCOME TOOK PLACE TODAY, HOW WOULD YOU VOTE?



2/3

„I would vote for it“ +
„I would probably vote for it“

Made with PASSION by Britta Bettendorf
>>github.com/BrittaInData >>datenlage.blog

CHALLENGE

Could we possibly predict

whether someone is

for or against

the idea of a basic income?

DATA PREPARATION STEPS

9,649 records

Data Cleaning:

> Handled duplicates and missing values

Votes:

> Kept only „yes and „no“ votes

Age:

> Created new age groups

Arguments:

> Pulled out most common arguments

8471 records

Made with PASSION by Britta Bettendorf
>>github.com/BrittaInData >>datenlage.blog

DIFFERENT CLASSIFICATION MODELS

>> Similar Accuracy Scores

Logistic Regression

Accuracy: 78%

SVM

Accuracy: 78%

XGBoost

Accuracy: 78%

Random Forest

Accuracy: 77%

NEXT: BALANCING THE DATA

What was done?

The imbalance in the dataset - caused by far more „yes“ voters - was accounted for by a sophisticated algorithm called SMOTE.

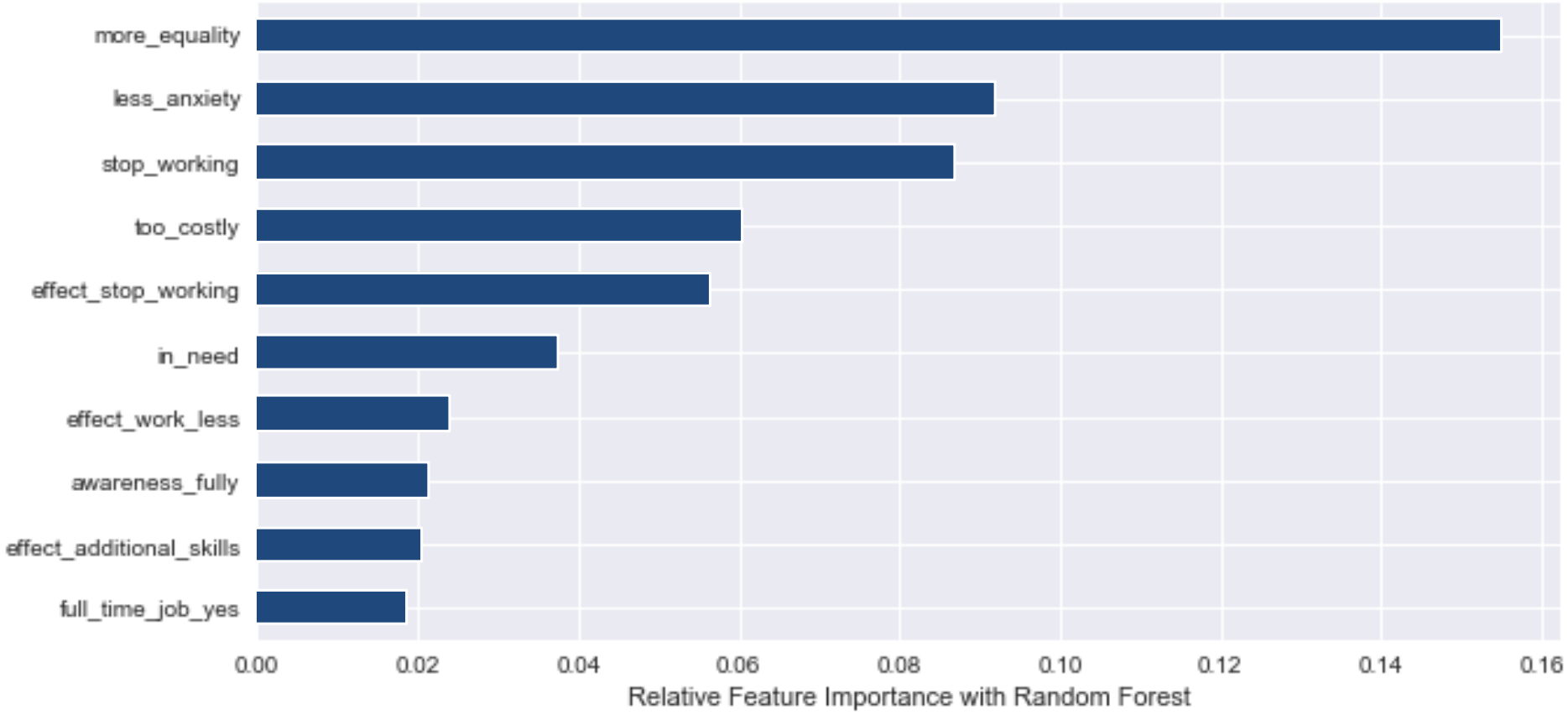
What was the outcome?

The accuracy improved by 4%.

The bias disappeared.

SUCCESS!

10 IMPORTANT FEATURES FOR VOTE PREDICTION



Made with PASSION by Britta Bettendorf
>>github.com/BrittaInData >>datenlage.blog

SUMMARY

Can we predict whether someone is

for or **against** the idea of a basic income?

Yes, we can.

With a certainty of almost 84%.


Thanks for having me!