
VDSG
VIENNA DATA SCIENCE GROUP
Data4Good
Hackathon
Vienna 2019
HILFSWERK
ÖSTERREICH
April 27-28, 2019
HILFSWERK

Problem

Within the framework of their 24-hour care services for people with physical disabilities, Hilfswerk Österreich provide a matching service between patients and care providers. Cancellations are undesirable.

There are two ways to reduce them: better matchmaking and better relationship maintenance. We were focusing on the latter.

VDSG

Goal

Our goal during the hackathon was to examine historical cases, in order to get insights on which data features are associated with cancellations.

Another goal was interpretability of the results: we love algorithms and models, but we also aimed at providing some valuable clues to Hilfswerk.

VDSG

Objectives

- Text mining and sentiment analysis of the text data (conversations with the clients and caregivers)
- Time-series analysis of individual cases (conflict trends)
- Creating a machine learning model to predict status (normal/conflict)

VDSG

Accomplishments

- Created multiple random forest classification models
- Performed text mining using different techniques (word frequency, tf-idf via Tidytext, Quanteda, tm)
- Carried out sentiment analysis of conversations with clients and caregivers (using tf-idf analysis and SentiWS – an external sentiment lexicon)
- Used text mining and sentiment analysis findings to engineer additional input variables which increased the performance of the models
- Assessed the feasibility of time-series analysis on the available data (not applicable)

VDSG

Challenges Encountered

- Labor-consuming data extraction, anonymization and cleaning
- Very few observations (742 observations of 90 cases in total)
- Dealing with data heterogeneity: time-ordered texts and categorical inputs for each case; two sources of data for each case
- Different results of tf-idf analysis via different text mining packages
- Hard to evaluate reliability of the results (e.g. possible overfitting of some models)

VDSG

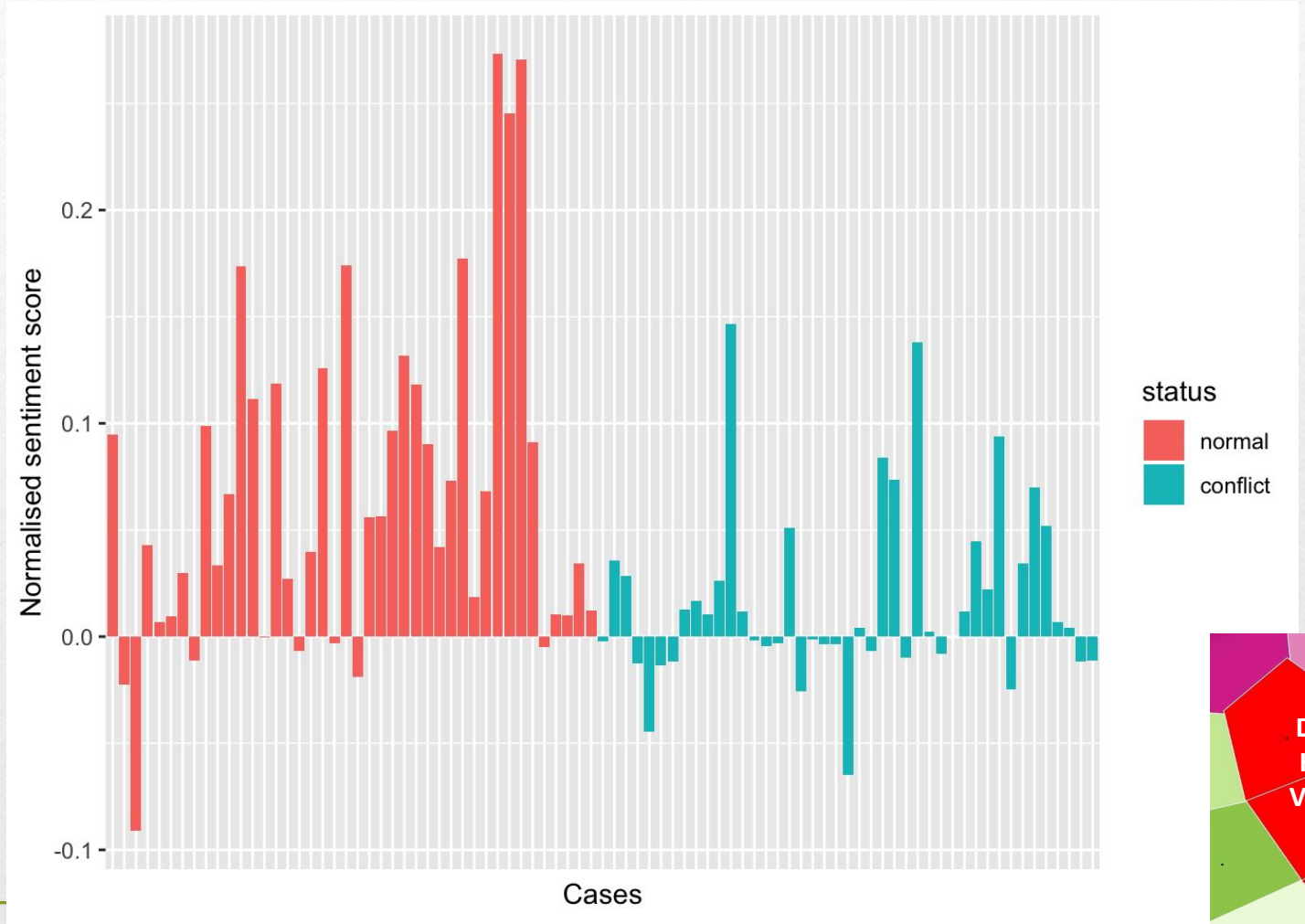
Insights

- Most normal and conflicted cases have positive SentiWS sentiment, but the average sentiment score is lower for conflicted cases
- Records of telephone calls have more negative sentiments than records of visits. Larger numbers of telephone calls are associated with conflict
- Sentiments calculated based on the tf-idf analysis are the most important predictors of whether the case is a conflict or not (overfitting possible!)
- Some possible signs of conflict:
  - frequent use of the word ‘nicht’ (especially by the caregiver)
  - usage of family words (‘Tochter’, ‘Mutter’, ‘Gattin’ etc)
- Many conflicts happen in the first four months of care

VDSG

Sentiment scores for cases

Calculated based on the SentiWS lexicon - a publicly available German-language resource for sentiment analysis

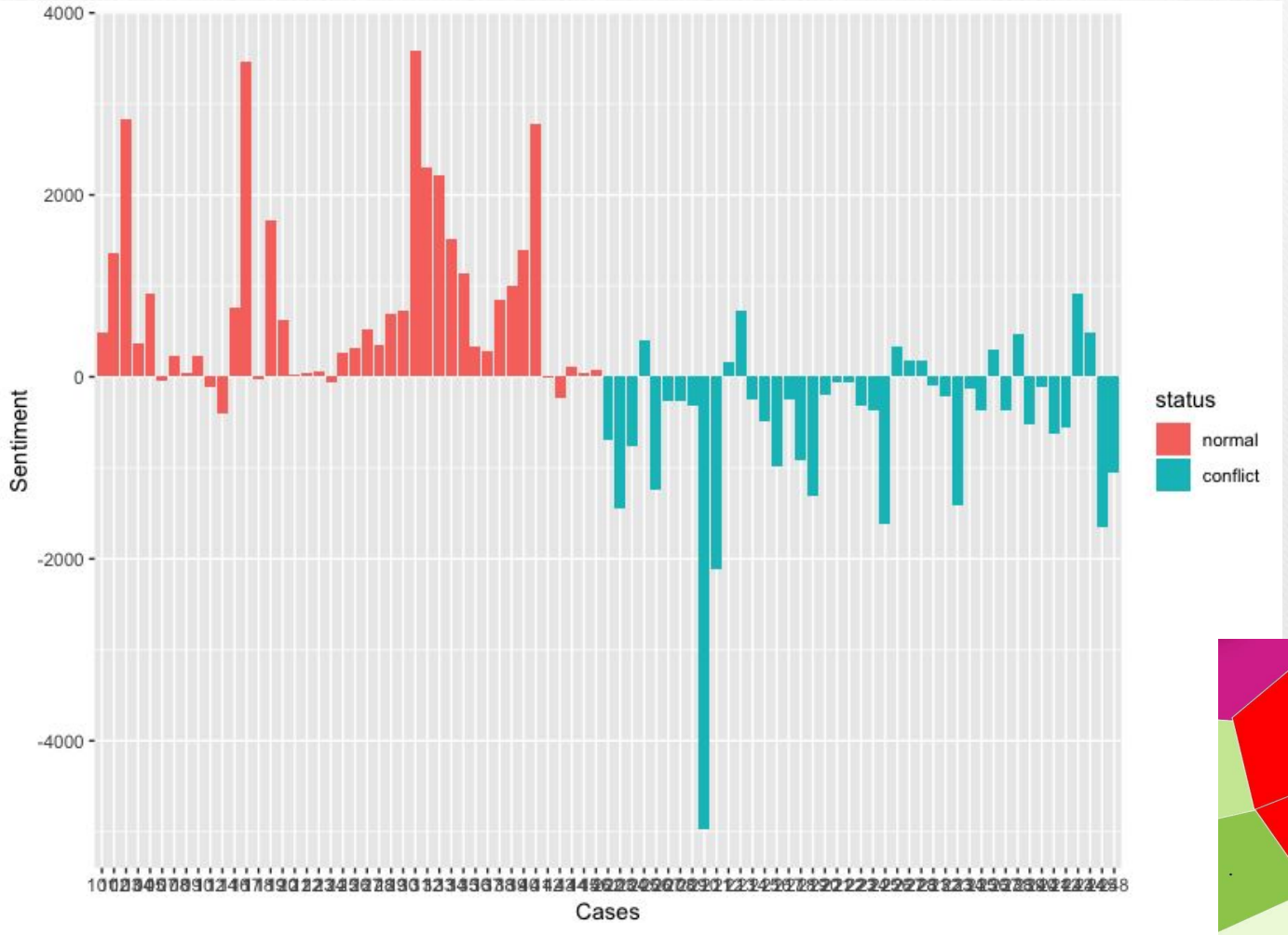


Data4Good Hackathon Vienna 2019

VDSG

Sentiment scores for cases

Calculated based on the tf-idf analysis of the normal and conflict training cases via Quanteda



Data4Good Hackathon Vienna 2019

VDSG

Most prominent words

Positive and negative words based on the word frequency analysis

POSITIVE

NEGATIVE

Data4Good Hackathon Vienna 2019

VDSG

Most prominent words

Positive and negative words based on the tf-idf analysis via Quanteda

POSITIVE

NEGATIVE

VDSG

The Road Ahead

- More thorough exploration of time trends in the data features, especially the texts
- Further text mining, e.g. types of important words for normal/conflict cases, caregivers/clients etc.
- There seems to be a potential for further improvement of the model, e.g. trying different feature sets
- Validate hypotheses on more data

VDSG

The Team

Project Lead:

- **Maria Ronacher**, https://www.linkedin.com/in/mariaronacher/

DS Team:

- **Bellinda Akwa-Asare**, https://www.linkedin.com/in/bellinda-akwa-asare-4538b52/
- **Jillian Augustine**, https://www.linkedin.com/in/jillianaugustine/
- **Matjaz Kovse**, https://www.linkedin.com/in/matjaz-kovse/
- **Roland Krasser**, https://www.linkedin.com/in/roland-krasser-5a2199139/

Data4Good Hackathon Vienna 2019

VDSG