Praktikum - Machine Learning for Information Systems Students

# ML Project – Final Presentation

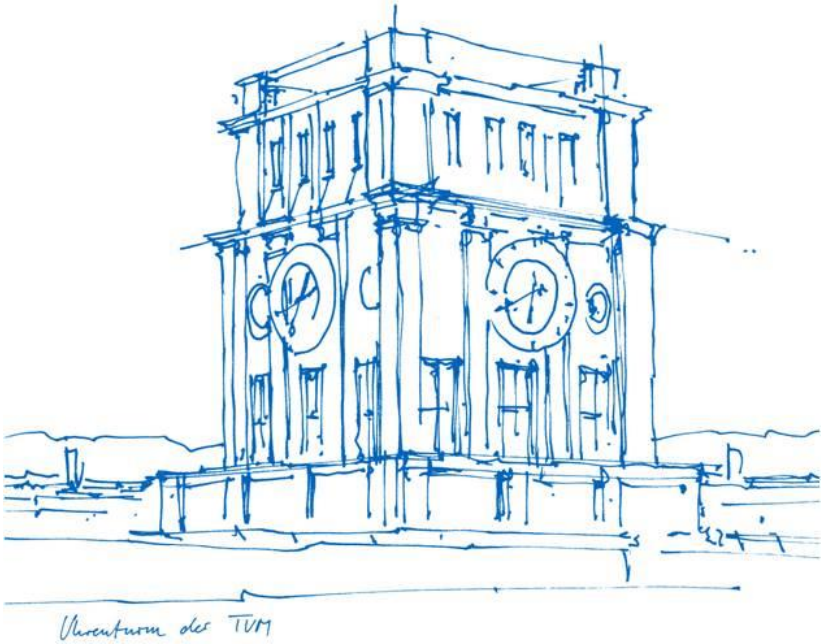
Group 3

Lucas Sanz Gorostiaga, Felix Böhm, Josep Quintana, Markus Linnenberg

Lehrstuhl I17 – Forschungsgruppe Wittges – SAP University Competence Center (UCC)

Technical University of Munich

Munich, 13 July 2021



# Table of Contents

# Project Background

**Status Quo and Goal**

- The UCC uses a SAP ticket system to handle incoming customer requests
- Requests are manually assigned to agents based on indicative information
- Idea: Make this support system more efficient with the help of machine learning

**Desired Outcome**

- Improve user experience: Reduce average ticket resolution time
- Improve efficiency: Reduce number of forwards per ticket (could indicate wrong classification/ticket assignment)
- These metrics should be tracked in a live deployment to evaluate the ultimate **business impact**

# Real-world challenges

- The UCC distinguishes between level 1 and level 2 requests, but this is not explicitly captured in the data (labelled)
- Individual operators are used for ticket assignment as opposed to Groups (support team) – operators can be part of many groups
- Distribution of tasks is not explicitly represented in the system – Ideally, a ticket is automatically assigned to a support team based on factors such as category, incident level, SLAs etc.
- Categories are assigned by the requestor and therefore not 100% reliable as they are not cleaned after ticket completion
- Assignment of operators is manual but ultimately reliable

# Approach - Classification

## Output

Predict the **operator** (Bearbeiter) of a ticket, as the data can be relied upon and the UCC assigns tickets ultimately to individual operators (as opposed to groups).

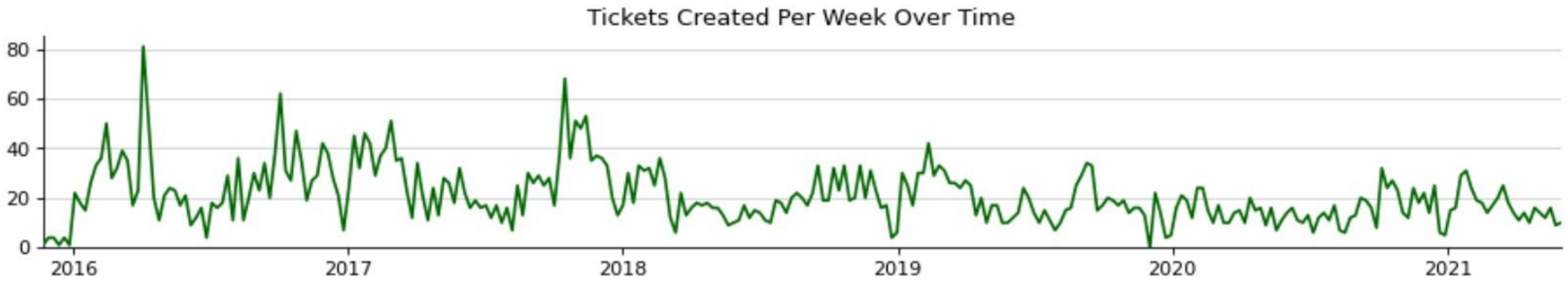
Predict the **category** of a ticket, as the category can help to infer the appropriate operator.

## Input

Leverage the **ticket text** of the **first message**, as this is in practice the primary information used by human agents to assign a support level and the operator.

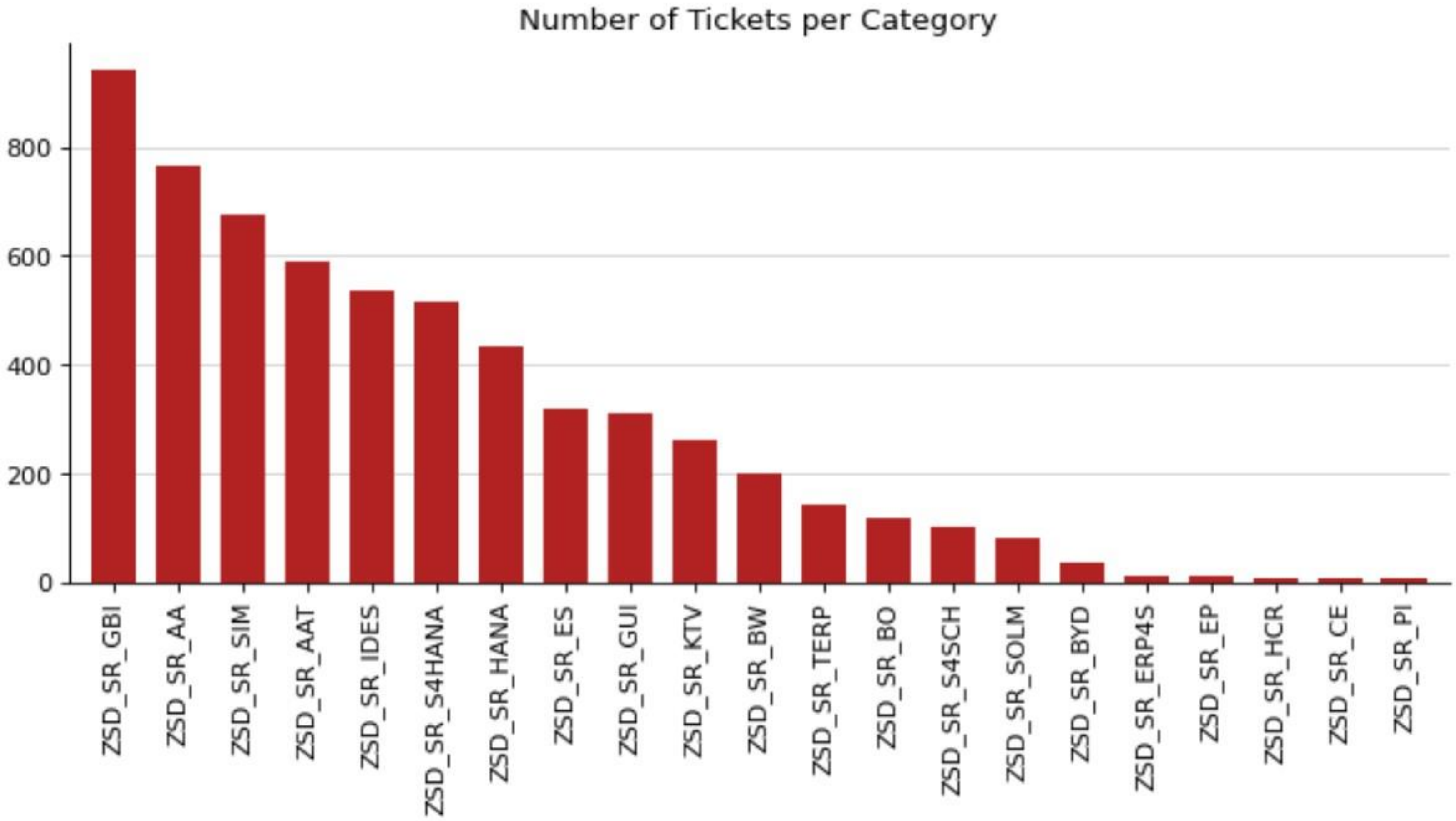
# Data Analysis - Tickets

- There are 12.184 tickets in total, of which we use **6.000** for prediction purposes
- 65% of tickets are in German and 35% in English



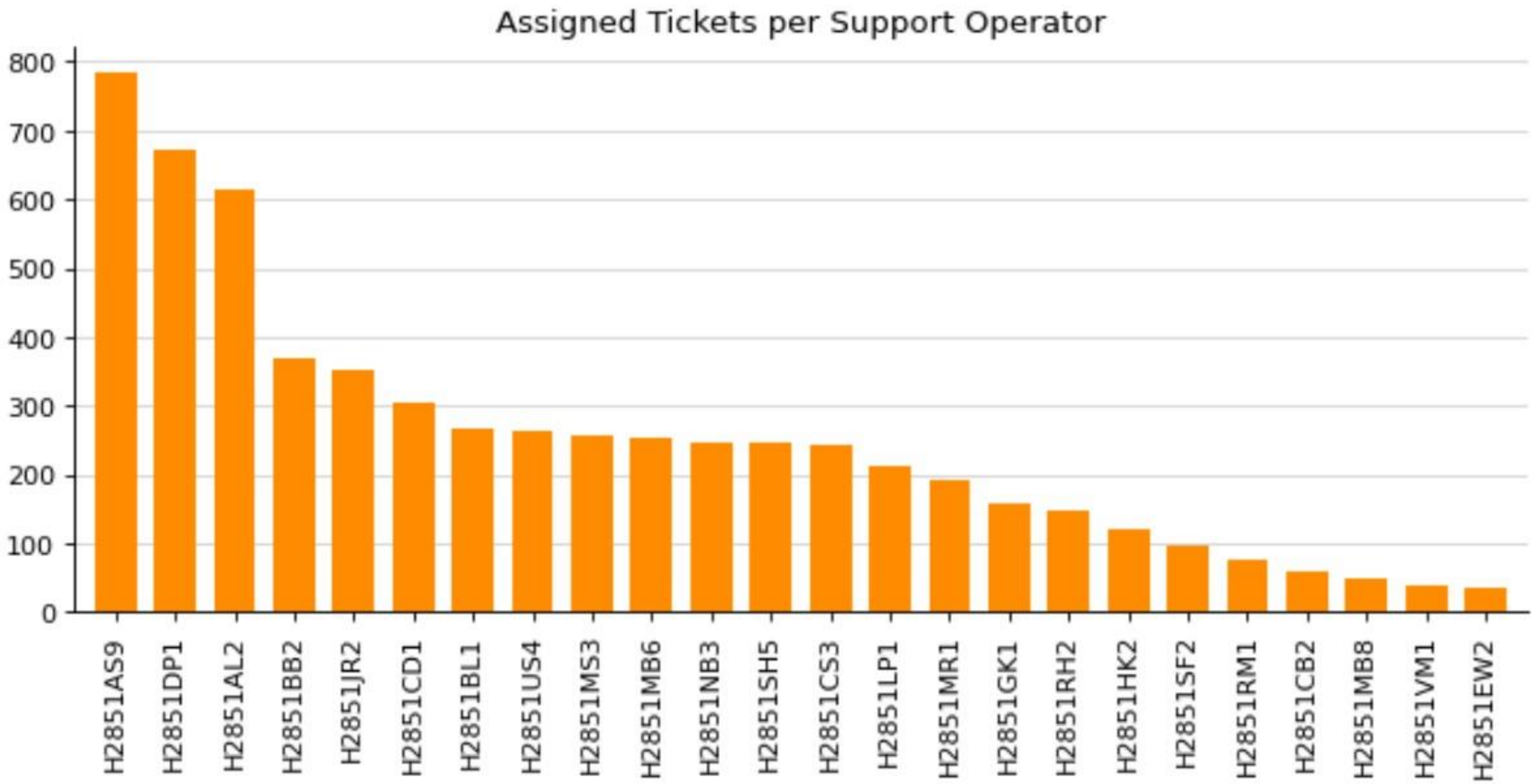
# Data Analysis - Categories

There are **21** different categories



# Data Analysis - Operators

In total 34 operators, of which **24** with meaningful ticket assignments



# Data Challenges

1. Languages are mixed (English and German)
2. Data includes autogenerated tickets without real customer interaction
3. Tickets without any data
4. Ticket data and meta data is stored separately

# Data Challenges

5. Operators / Bearbeiter change over time

# Data pre-processing

1. Select only the relevant columns from `tickets.csv`
2. Match the `ticket ID` from `tickets.csv` with the corresponding ticket file (CSV) and read in the data
3. Filter out auto-generated tickets and tickets without interaction
4. Extract the first message of the ticket file which contains the ticket description
5. Clean the message of specific characters ('/-_<>&‘)
6. Filter out `NaN` values and empty strings etc. in the dataframe
7. Detect the ticket language
8. Translate English messages to German via Google Translate API

# Model

1. Convert the message text into a token count vector with *CountVectorizer* from *sklearn*
   1. This removes stop words like ”und” or ”die” and makes everything lowercase
   2. We use stop words from the *spacy* package
2. Apply TF-IDF (term-frequency inverse document-frequency) to these counts via *TfidfTransformer* from *sklearn*
   => TF-IDF identifies important words in a specific ticket that occur less frequently in other tickets
3. Use Naive Bayes classifier *MultinomialNB* to classify the input data onto discrete classes (the ticket operator or the ticket category)

# Results

| Class | Metric | MultinomialNP | Baseline | Difference |
|---|---|---|---|---|
| Operator | Accuracy | 26.4% | 12.9% | + 13.5% |
| | Top 3 accuracy | 44.7% | 34.1% | + 10.6% |
| Category | Accuracy | 43.1% | 15.6% | + 27.5% |
| | Top 3 accuracy | 69.1% | 39.3% | + 29.8% |

1. We use accuracy and top 3 category as performance measures as these would be the most helpful in a practical setting (e.g. give suggestions to the agent)
2. Definition of ***Baseline***:
   Assignment of the operator/category based on the most frequent classes (e.g. share of top 3 categories account for 39.3% of the data) => “Does our model perform better than assigning tickets by random chance?”

# Discussion of results

Prediction by **operator**

- Relatively high number of classes (24)
- Practical issue: Operators change often (see analysis)
- We only use the ticket text for predictions, but ultimately that is unfair to the model, as e.g. one year a specific operator is assigned tickets of a specific type, and later that same ticket would be assigned to a replacement operator

Prediction by **category**

- Our model performs very well considering the number of classes (21)
- Model could be used to predict any predefined categories or support teams
- Ideally, the usage of the category leads to assignment to support teams (i.e. operators change but support team is the same) – This could help with training the model and assigning the correct support team or operator in the future

# Ideas for improvements

- Ideal predictor would be the **support team** rather than individual operators or categories. For this, the overall support process needs to be improved
- Due to the change in operators, including the **time component** into the input would probably increase the model performance significantly
- Fine-tune model and model selection – However: It makes sense to **address data quality/structural issues** first
- Our model could be deployed as a **cloud function** (Google cloud, AWS Lambda) which receives a request for every new ticket and that sends the top n predictions for the category / operator
- In production we would retrain the model on current data continuously (**online learning)**, which also could help to address the issue of changing operators
- Track **business impact** of this deployment by measuring average handling time, NPS or similar metric.

# Quick Demo

Thank you!