

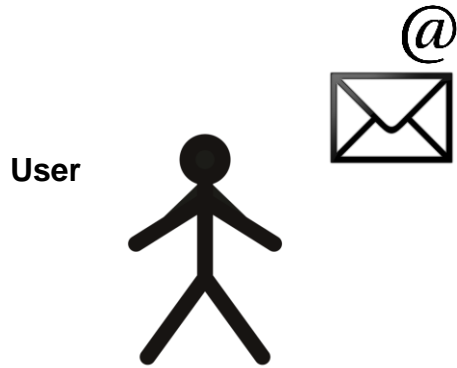


# Schnellerer Servicedesk-Support durch automatische Textklassifikation

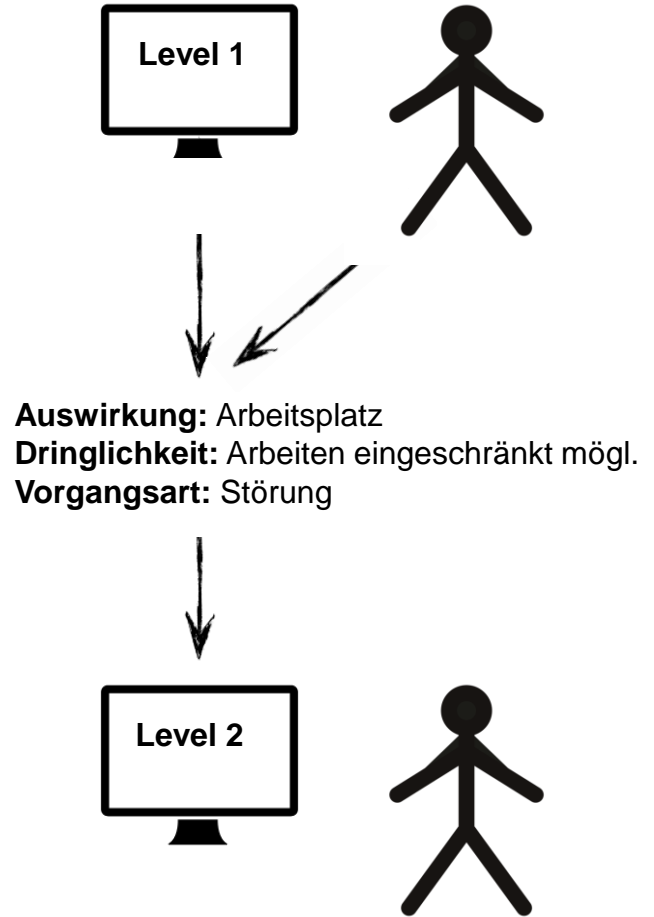
Dr. Kathrin Björkelund

Hochschule Karlsruhe, 3. März 2016

# Problemstellung



**Betreff:** Anmeldung am PC schlägt fehl  
**Beschreibung:** Max Mustermann kann sich nicht am PC anmelden. Fehlermeldung: Stammsatz fehlt.



# Modellierung: Dimensionen und Klassen

## Auswirkung

- Arbeitsplatz 98.08 %
- Abteilung 1.73 %
- Hauptabteilung 0.20 %

## Dringlichkeit

- Arbeiten möglich 90.39 %
- Arbeiten eingeschränkt möglich 6.16 %
- Arbeiten stark eingeschränkt 3.45 %

## Vorgangsart

- Service-Request 45.42 %
- Störung 35.47 %
- Administration 10.76 %
- Beschaffungsanfrage 3.58 %
- Monitoring 3.06 %
- Hardwarestörung 1.67 %
- Bedienungsfehler 0.02 %
- ...

# Exkurs: Baseline

- Wie gut ist ein extrem simples Modell?
- **Mehrheits-Baseline** sagt immer häufigste Klasse vorher
- Falls Trainingsdaten repräsentativ:
  - Auswirkung=Arbeitsplatz: 98 % Genauigkeit
  - Dringlichkeit=Arbeiten möglich: 90.39 %
  - Vorgangsart=Service-Request: 45.42 %

# Modellierung: Features

- Wörter („bag of words“) aus Betreff und Beschreibung
- Wortkombinationen (Bigramme)
- Filtern von Stoppwörtern, Daten, Emailadressen, ...
- Stemming

# Beispiel: Featurevektor

**Betreff:** Anmeldung am PC schlägt fehl

**Beschreibung:** Max Mustermann kann sich nicht am PC anmelden. Fehlermeldung: Stammsatz fehlt.

## Betrefffeatures

### Feature-Unigramme

**betreff:**anmeldung    **betreff:**pc    **betreff:**schlägt    **betreff:**fehl

**betreff:**anmeldung&**betreff:**pc    **betreff:**pc&**betreff:**schlag  
**betreff:**schlag&**betreff:**fehl

### Feature-Bigramme

**beschreibung:**max    **beschreibung:**mustermann    **beschreibung:**kann  
**beschreibung:**pc    **beschreibung:**anmelden    **beschreibung:**fehlermeldung  
**beschreibung:**stammsatz    **beschreibung:**fehlt

### Feature-Unigramme

**beschreibung:**max&**beschreibung:**mustermann  
**beschreibung:**mustermann&**beschreibung:**kann    ...

### Feature-Bigramme

## Beschreibungsfeatures

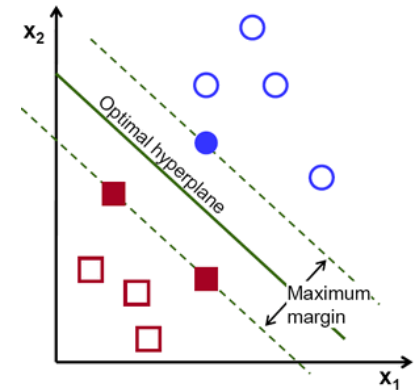
# Modellierung: Algorithmen

- **Naive Bayes:**

- $\hat{y} = \arg \max_{k \in \{1, \dots, K\}} p(C_k) \prod_i p(x_i | C_k)$

- MLE

- **Lineare SVM:** sep. Hyperplane



# Evaluation: Datensätze

- Aufteilung der Daten:
  - Trainingsdaten: Lerngrundlage (56k Tickets)
  - Entwicklungsdaten: Parametertuning ( 2k Tickets)
  - Testdaten: Finale Evaluation ( 2k Tickets)
- Zufällige vs. zeitliche Aufteilung

03.2014



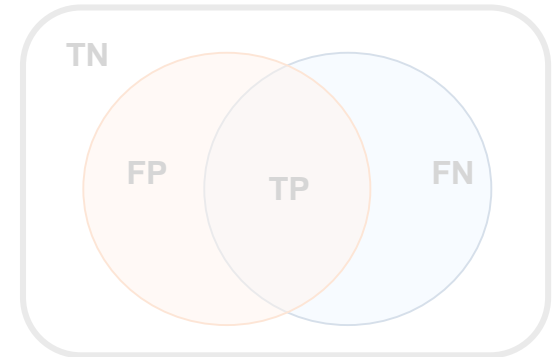
03.2015



# Evaluation: Metriken

- Accuracy:  $100 * \#TP / \#Testinstanzen$
- Precision:  $100 * \#TP / (\#TP + \#FP)$
- Recall:  $100 * \#TP / (\#TP + \#FN)$

Blau: tatsächlich Klasse c  
Orange: Klasse c vorhergesagt



# Evaluation: Ergebnisse

- DE besser als EN
  - Trainingsdatenmenge
- SVM besser als Naive Bayes
  - Abhängigkeit der Features
- Stemming/Normalisierung hilft für DE nicht
  - Annahme: kodiert relevante Information

# Evaluation: Ergebnisse

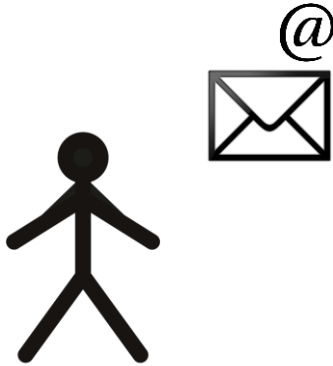
	Baseline	SVM
<b>Vorgangsart</b>		
	49.18	85.90
<b>Auswirkung</b>		
	98.36	98.36
<b>Dringlichkeit</b>		
	92.35	93.22

# Integration in Produktivsystem

- **Einbindung** in Ticketsystem
  - als Webservice
  - momentan Übernahme per Click
- **Caching** des Modells
  - Liblinear-Bibliothek dateibasiert
  - ursprünglich ohne Modell-Caching

# Problemstellung

User



**Auswirkung:** Arbeitsplatz  
**Dringlichkeit:** Arbeiten eingeschränkt mögl.  
**Vorgangsart:** Störung



**Betreff:** Anmeldung am PC schlägt fehl  
**Beschreibung:** Max Mustermann kann sich nicht am PC anmelden. Fehlermeldung: Stammsatz fehlt.

A photograph of a modern building facade with a grid of windows and white panels. A blue semi-transparent rectangle is overlaid on the left side, containing white text. A bright green horizontal bar is at the bottom of the blue rectangle.

# Vielen Dank!

**Dr. Kathrin Björkelund**  
**Data Management & Analytics**

inovex GmbH  
Karlsruher Straße 71  
75179 Pforzheim

+49 (0)173 3181-085  
[kathrin.bjoerkelund@inovex.de](mailto:kathrin.bjoerkelund@inovex.de)