Sentiment Analysis of Twitter Posts on Covid-19 Cultural Dimensions: Collectivist vs. Individualist

Authors: D. Dobler, L. Dorner, M. Köppel, P. Brockmann

Presenter: Patricia Brockmann
Email: Patricia.brockmann@th-nuernberg.de

Nuremberg Institute of Technology, Germany

Resumé: Patricia Brockmann

- Bachelor of Science Degree in Management Information Systems
  University of Colorado, Boulder
- German Diplom in Information Systems
  University of Regensburg, Germany
- Dr. rer. Nat. in Information Systems
  University of Regensburg, Germany
- Current: Professor for Information Systems
  Computer Science Department
  Nuremberg Institute of Technology in Germany

Research Interests

- Bionic Computation
  - Algorithms based on natural processes
  - Sentiment analysis, swarm intelligence, neural networks
- Global Software Engineering
  - Geographically distributed software development
  - Agile project management methods
  - Cultural dimensions in international software teams
- e-Learning
  - Gamification
  - Project-based learning

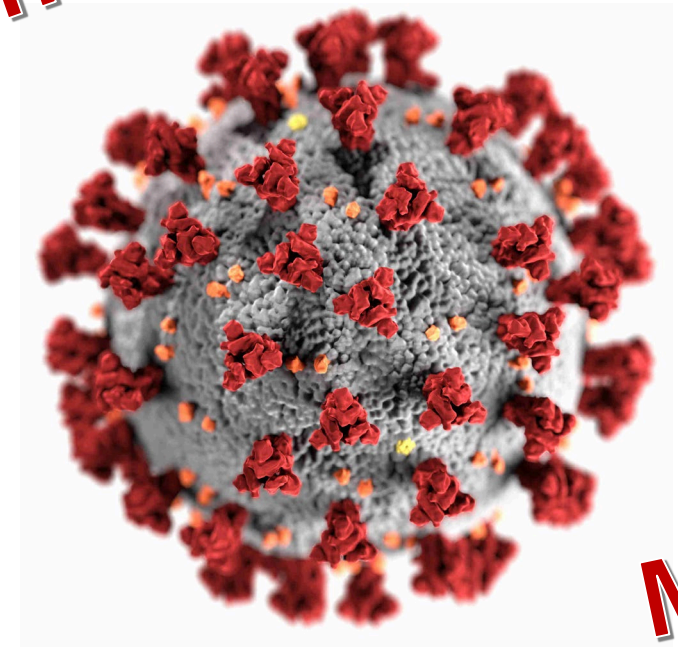
Contents

- Motivation
- Research Questions
- Methods
- Results
- Conclusions
- Future Work

Motivation: Covid-19 Pandemic

Lockdowns

Social Distancing



Travel Restrictions

Mental Health Issues

Research Questions

1. **Sentiment Analysis:** Can sentiment analysis deliver meaningful insights about opinions on the COVID-19 pandemic as expressed in Twitter posts?

2. **Cultural Dimensions:** Do cultural dimensions of people from individualist cultures vs. collectivist cultures affect their expressed opinions?

Sentiment Analysis

- Identify and classify the context of how terms are used
- Prediction of positive, neutral or negative sentiments

Positive	Neutral	Negative
• optimistic • lucky • calm	• no difference • doesn‘t matter • work from home	• difficulty • withdrawn • panic

Source: Pang, B. Lee, L., „Opinion Mining and Seniment Analysis“, Now Publishers, Boston, 2008.

Cultural Dimensions

Power Distance

Collectivist vs. Individualist

Assertiveness vs. Cooperation

Long-Term vs. Short-Term

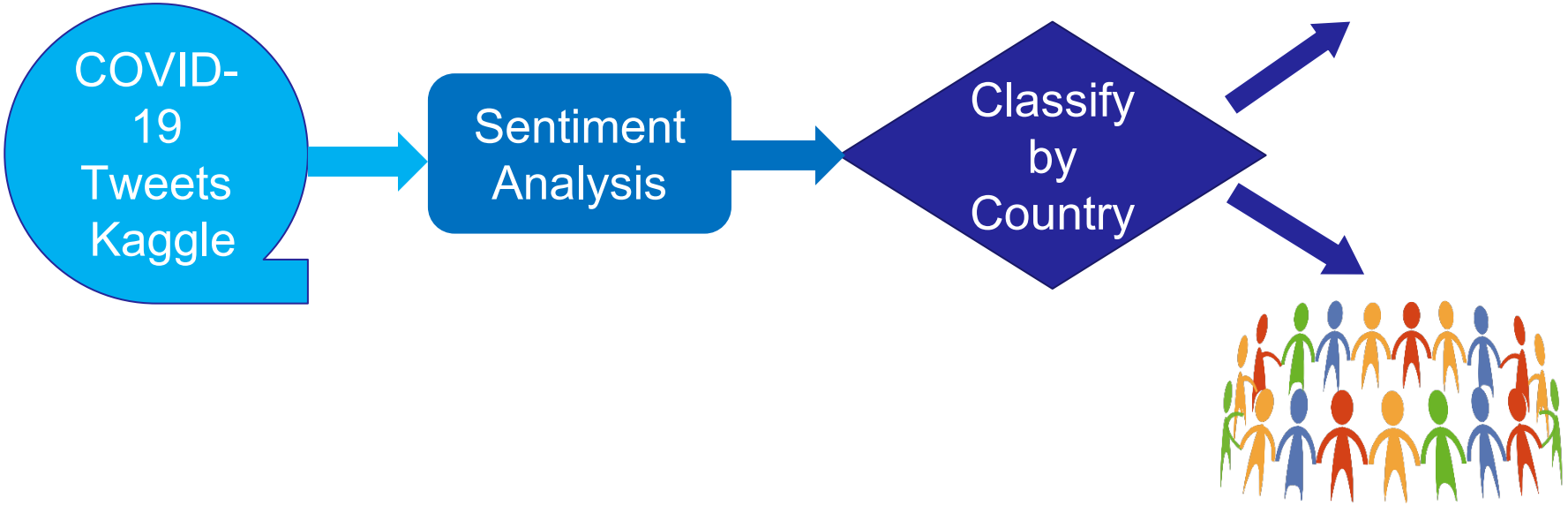
Uncertainty avoidance

Indulgence vs. Reststraint

Source: Hofstede, G. et.al, (2010) „Culture and Organizations – Software of the Mind“, McGraw-Hill,

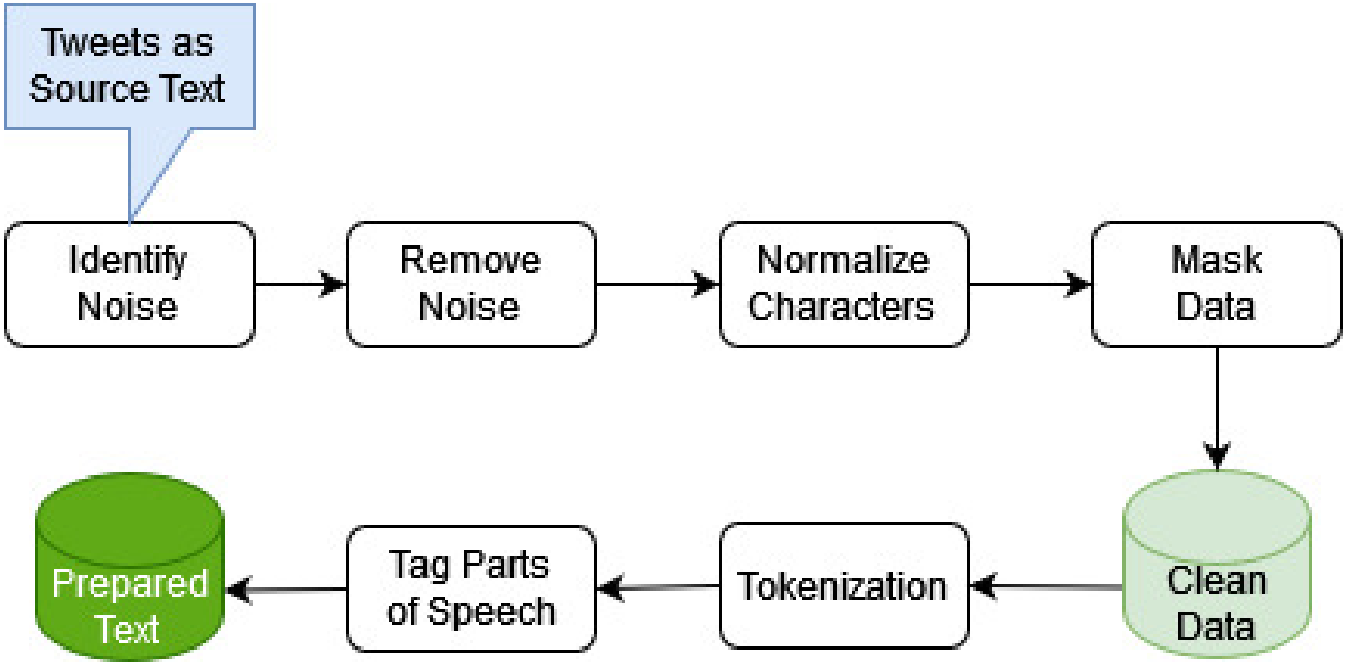
Source: www.wikimedia.org Creative Commons License 2.0

Methods: Experiment Design



Source: A. Miglani, "Coronavirus tweets nlp - text classification," https://www.kaggle.com/datatattle/covid-19-nlp-text-classification

Methods: Preprocessing Pipeline



Preprosessing: Data Cleansing

Original Data from Twitter API:

```
'The dominos fall:  #coronavirus &gt; world  #healthemergency &gt; #consumer and #supplychain
resources shut down &gt; #economy stalls &gt; #credit cracks ( cc )&gt; #liquidity shock ( ls ) &gt;
 #unemployment &gt; 2x ( cc &gt; ls ) &gt; $spx and #bonds crash ( sbc ) &gt; 3x ( -1 ( sbc *
(2x(cc&gt;ls)) = dominos )'
```

Data Cleansing:

- Delete special formatting characters and program code

Result:

- Small amount of stand alone characters
- Huge amount of white space

Methods: Preprocessing Text Data Remove Stop Words

Before Stopwords	After Stopwords
The product is really very good **(Positive)**	product really good**(Positive)**
The products seems to be good.**(Positive)**	products seems good **(Positive)**
Good product I really liked it**(Positive)**	Good product really liked **(Positive)**
I didn't like the product **(Negative)**	like product **(Positive)**
The product is not good **(Negative)**	product good **(Positive)**

Source: www.rapidminer.com

Methods: Generate n-Grams

- Groups of **n** tokens
  1. Unigram
  2. Bigram
  3. Trigram

- Adjective – Noun Phrases Adverb Phrases

	adj_noun_phrases	adv_phrases
116	[fortunate families]	[less fortunate]
117	[important message, safe #, safe # aady...	[very important]

Methods: Term Frequency * Inverse Document Frequency

- Term Frequency: $$\frac{\text{number of times a word appears in an interview}}{\text{total number of words in the interview}}$$
- Inverse Document Frequency: $$\log\left(\frac{\text{total number of interviews}}{\text{number of interviews a term appears in}}\right)$$
- TF * IDF: Term Frequency * Inverse Document Frequency

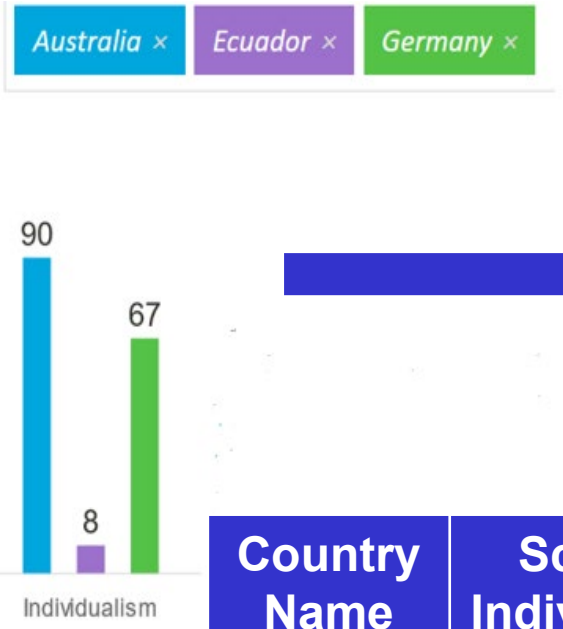
  → Shows how significant an n-gram is in a set of interviews

Source: Goodfellow, I., et al., „Deep Learning“, MIT Press, 2016.

Location Mapping of Countries to Cultural Dimensions

```
                        NYC
                Seattle, WA
                Chicagoland
        Melbourne, Victoria
                Los Angeles
                        ...
                Brooklyn, NY
            Toronto, Ontario
                        OHIO
Wellington City, New Zealand
i love you so much || he/him
```

Extract locations



Map countries to cultural dimensions

Country Name	Score for Individualism	Classification
Australia	90	individualist
Ecuador	8	collectivist
Germany	67	individualist

Source: Hofstede, G. et.al, (2010) „Culture and Organizations – Software of the Mind“, McGraw-Hill,

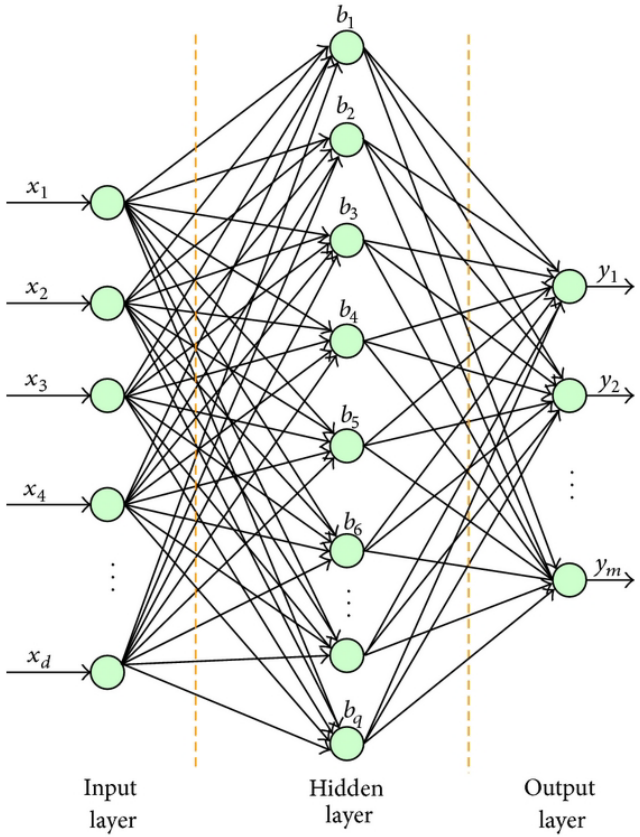
Methods: Neural Network

- Multi-Layer Perceptron Classifier (Sci-Kit Learn)

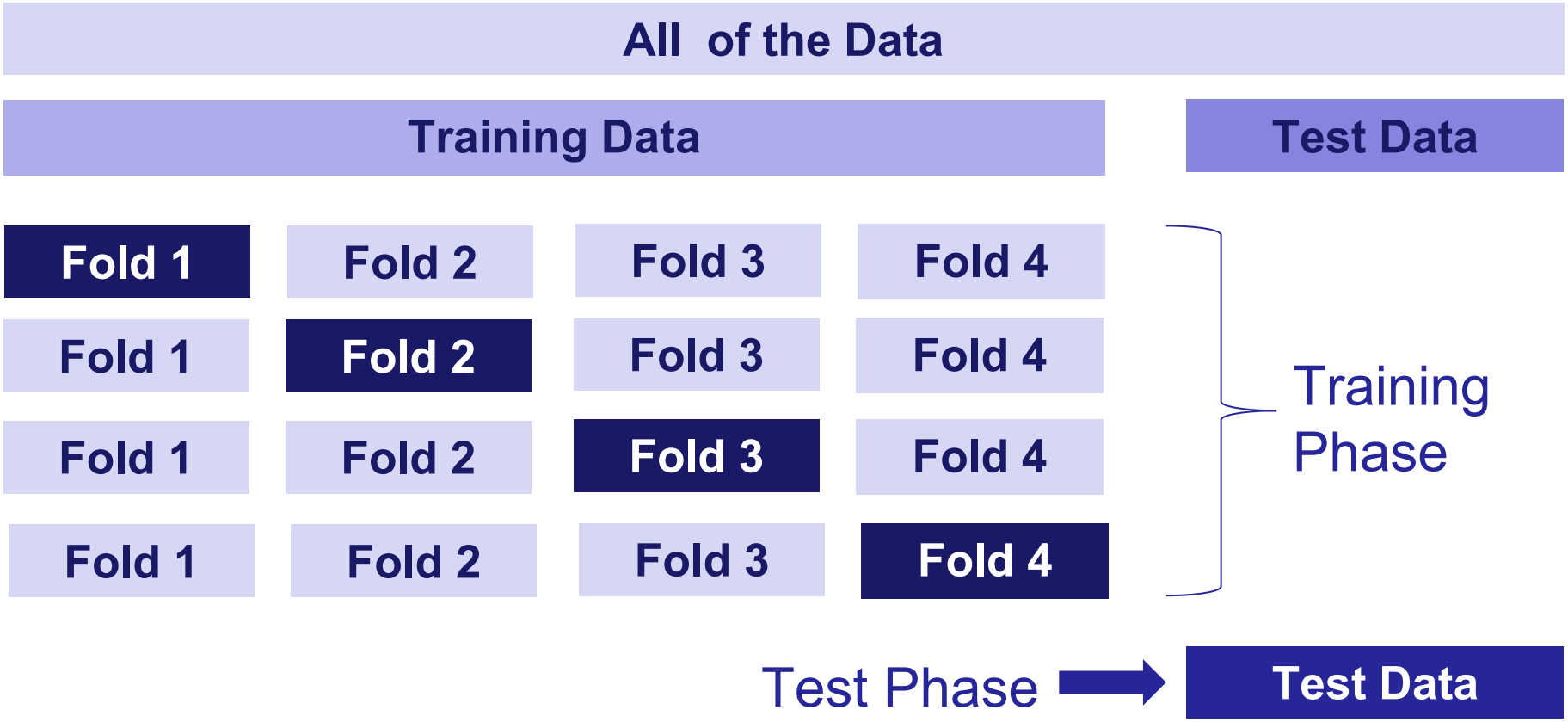
- Single hidden layer
- 100 neurons

- Output layer with 3 sentiments:

    1. Positive
    2. Neutral
    3. Negative

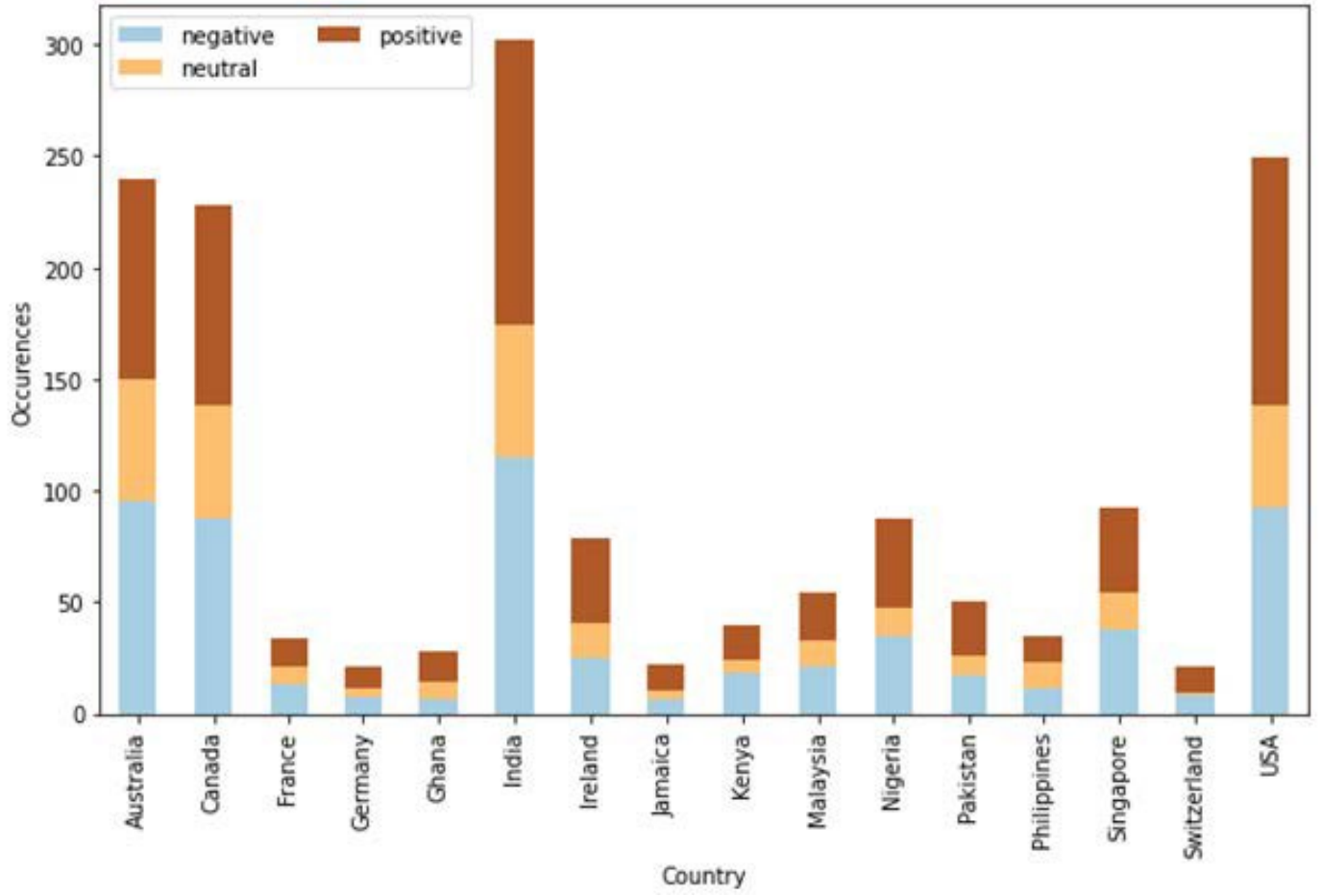


Methods: K-Fold Cross-Validation



Source: Goodfellow, I., et al., „Deep Learning“, MIT Press, 2016.

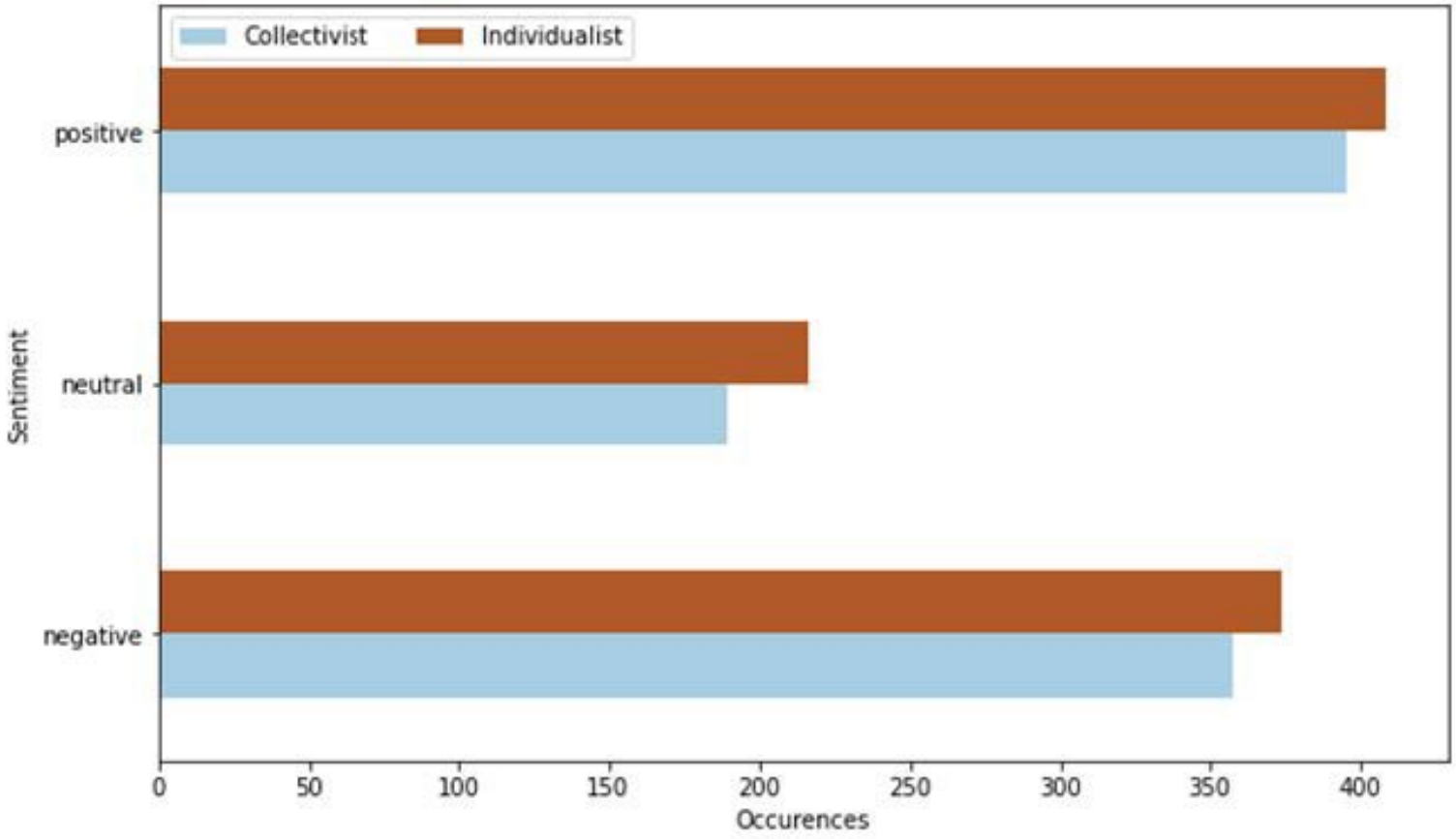
Results: Sentiment Analysis by Country



Performance Metrics

Metric	Score	Metric	Score
Accuracy	73.5%	Precision	71.0%
Recall	72.2%	F1	71.6%
Sensitivity	72.2%	Specificity	74.6%

Results: Sentiment Analysis Collectivist vs. Individualist



Limitations

- Data collected over a short period of time at the beginning of the pandemic
- Relatively small subset of the data for initial prototype
- Not representative of the entire population
- Not statistically significant
- Choice of preprocessing methods potentially skew results
- Manual labelling process can bias results

Conclusions

- **Research Question 1**:

  Sentiment analysis of Twitter posts can deliver meaningful insights into opinions about the COVID-19 pandemic.

- **Research Question 2:**

  No significant difference in Twitter posts of people from individualist vs. collectivist cultures could be discerned.

Future Work

- Conduct further analysis with larger data sets over longer time period to increase confidence levels.

- Investigate correlations between sentiments of further cultural dimensions:

  - Power distance
  - Masculine vs. feminine
  - Uncertainty avoidance
  - Long-term vs. short-term orientation
  - Inhibition vs. restraint

Acknowledgements & Contact Information

- We would like to express our thanks to the conference organizers

- Support from a research grant German Academic Exchange (DAAD) program for International Virtual Academic Collaboration (IVACS)

- Contact Author: Patricia.Brockmann@th-nuernberg.de