DATA SCIENCE CENTER

University of Bremen

Applications of Text Mining and NLP

CONTACT

Dr. Maryam Movahedifar

movahedm@uni-bremen.de

Website

Outline

Introduction to Text Mining

Text Mining Process

Applications and Ethics

Fake News Detection

Hate Speech Recognition

Media Content Analysis

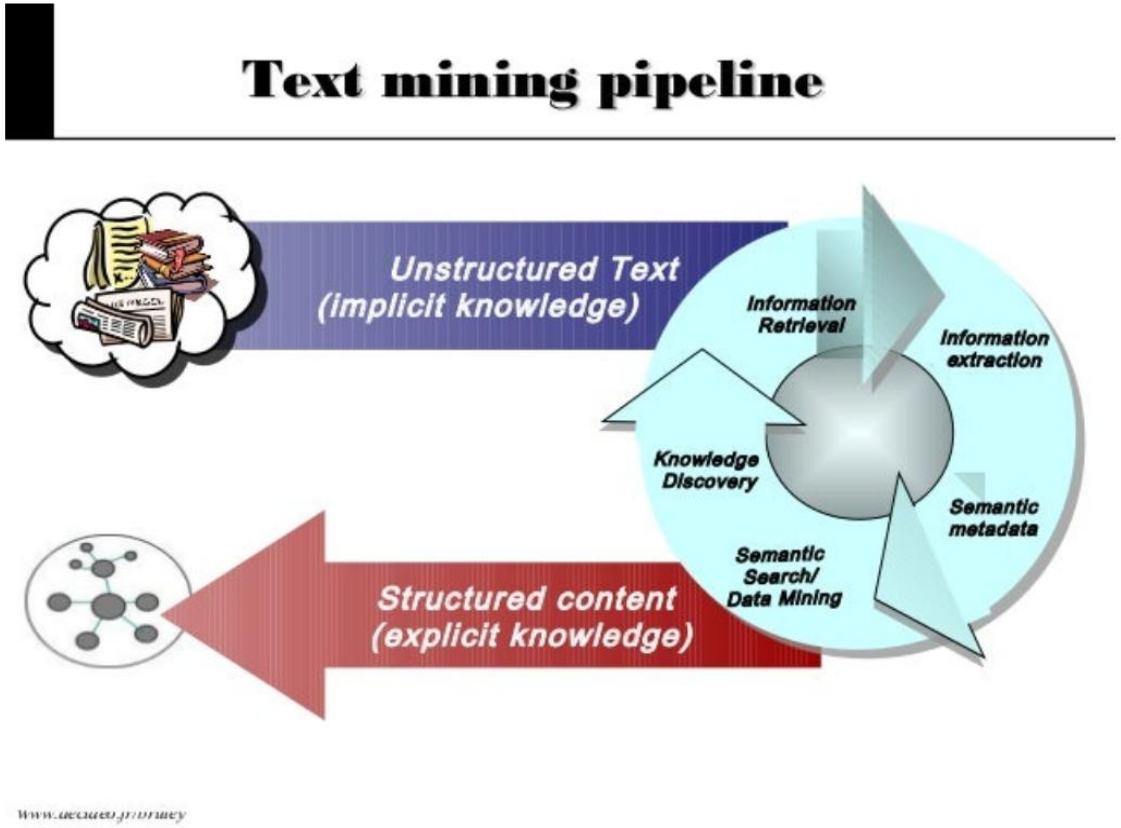
Healthcare Risk Prediction

Responsible AI

Summary Of Text Mining pipeline and its Challenges

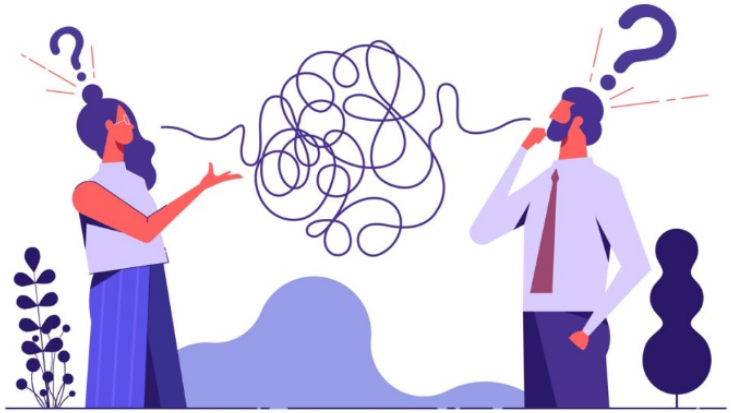
What is Text Mining?

- Extracting patterns from unstructured text data
- Used in **law, medicine, education, media**
- Key tasks: sentiment analysis, topic discovery, text classification
- Driven by the growth of online textual data



Challenges in Text Mining

One of the biggest challenges in text mining is that language is complex and hard.



- Different things can mean more or less the same (e.g., *"data science"* vs. *"statistics"*).
- Context dependency (e.g., *"You have very nice shoes"*).
- Same words with different meanings (e.g., *"to sanction"*, *"bank"*).
- Irony and sarcasm (e.g., *"That's just what I needed today!"*).
- Figurative language (e.g., *"He has a heart of stone"*).
- Negation and spelling variations (e.g., "not good" vs. "good" and "color" vs. "colour").

Applications and Ethics

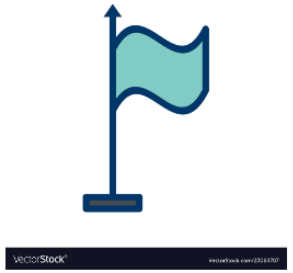
Applications of Text Mining

Similarity

- Find authors of an anonymous book
- Find duplicates and link records
- Find relevant documents given a user query

Clustering

- Targeted advertisement or learning
- Recommendation systems (e.g. similar books)
- Clustering stories (fiction works, diagnoses, misinformation)
- Track evolution of topics in discourse

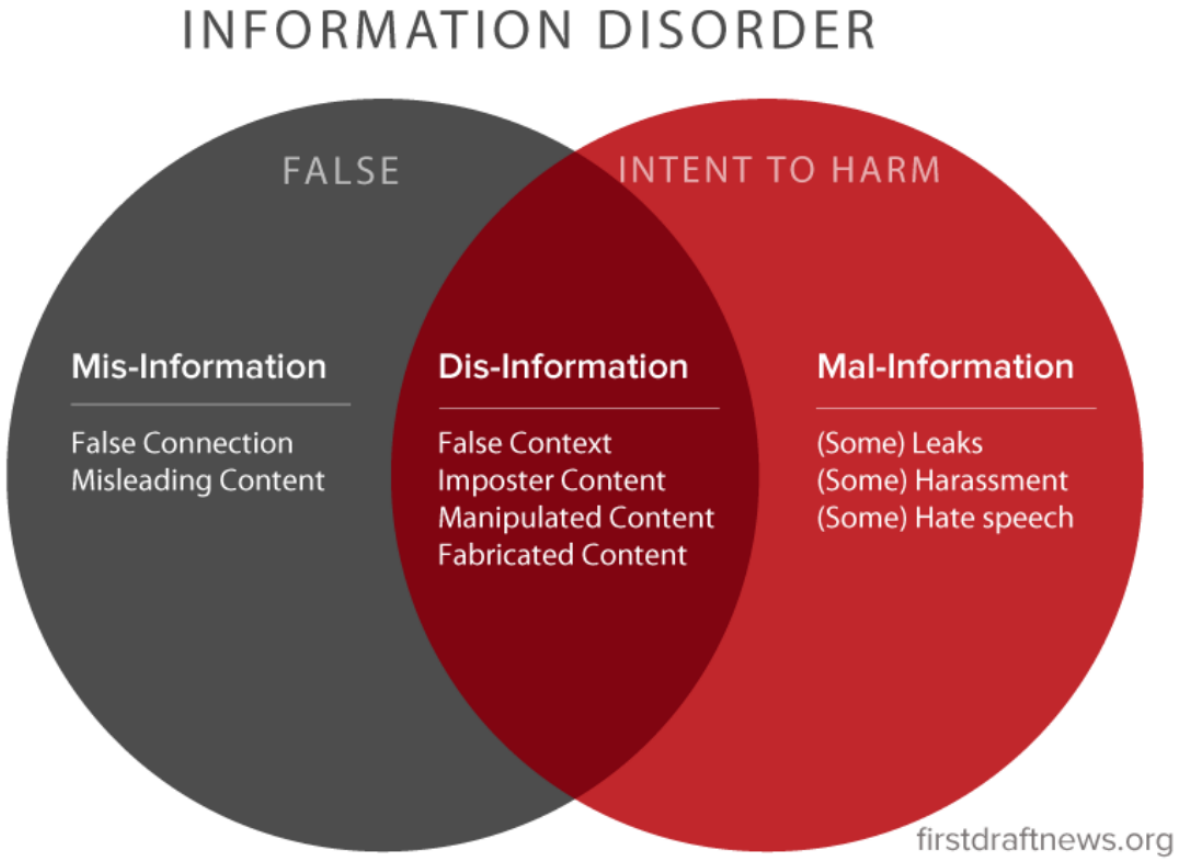


Classification/Regression

- Hate speech classification (spam, news)
- Sentiment and emotion analysis
- Predict student performance
- Probability of re-hospitalization
- Classifying reports

Information Disorder Online

- **Misinformation:** False information shared without intent to cause harm.
- **Disinformation:** Deliberately false information shared with the intent to harm.
- **Malinformation:** Genuine information shared to cause harm, often by breaching privacy.



These forms of information disorder threaten democracy, public health, and societal safety.

Real-World Application Areas

Healthcare Risk Prediction

Hate Speech Recognition

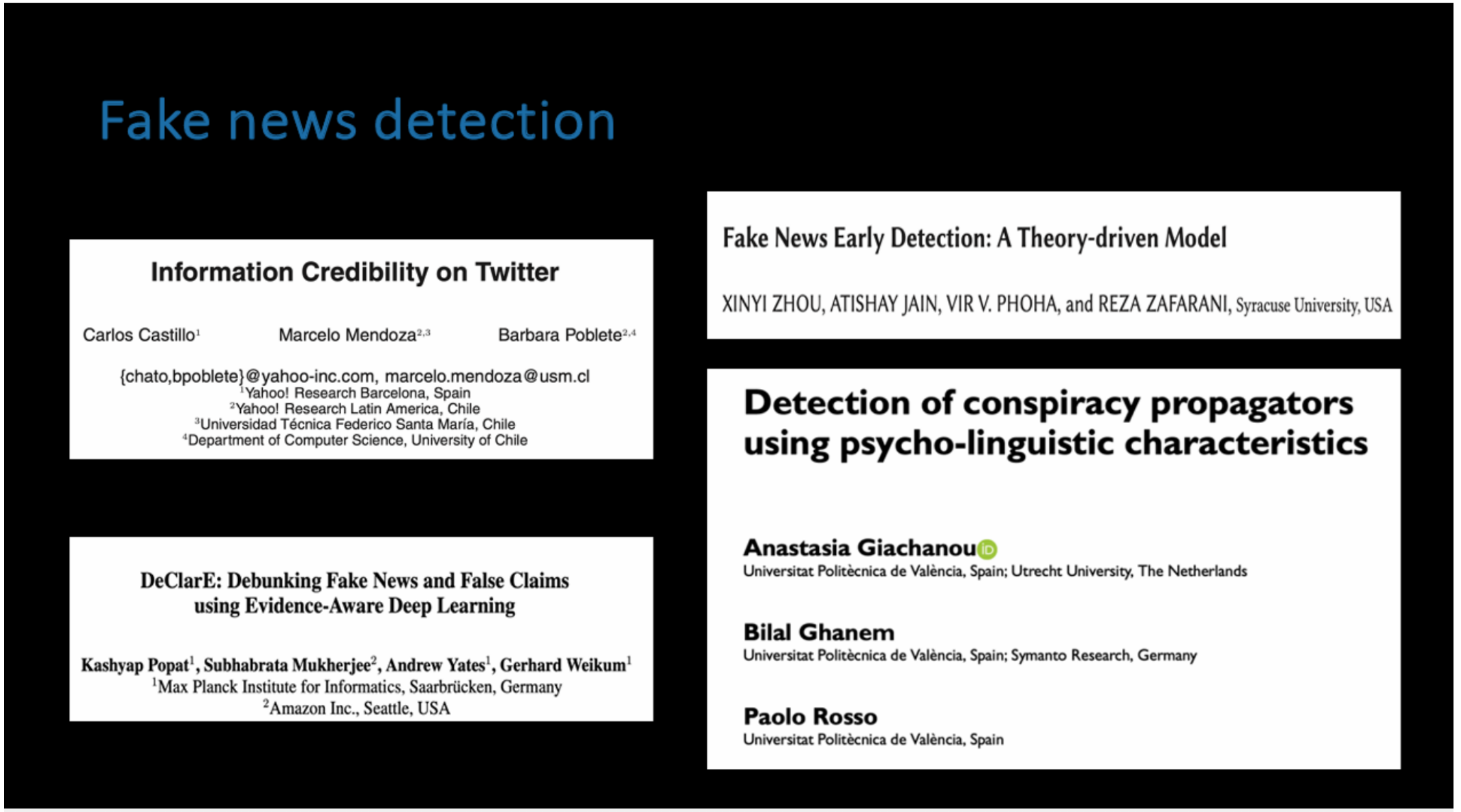
Media Content Analysis

Fake News Detection

Responsible AI

Fake news detection

Examples of Fake News Detection



Understanding Fake News

Definition:

- Intentionally and verifiably false content presented as news.
- Covers claims, speeches, posts, etc., involving public figures or institutions.
- Emphasizes both falsehood and intent, often tied to recognizable news outlets.

Difficult to Detect:

- Detection accuracy by humans: **55–58%**.
- Perceived truth increases with repetition (*validity effect*).
- Influenced by confirmation bias, desirability bias, and peer pressure.

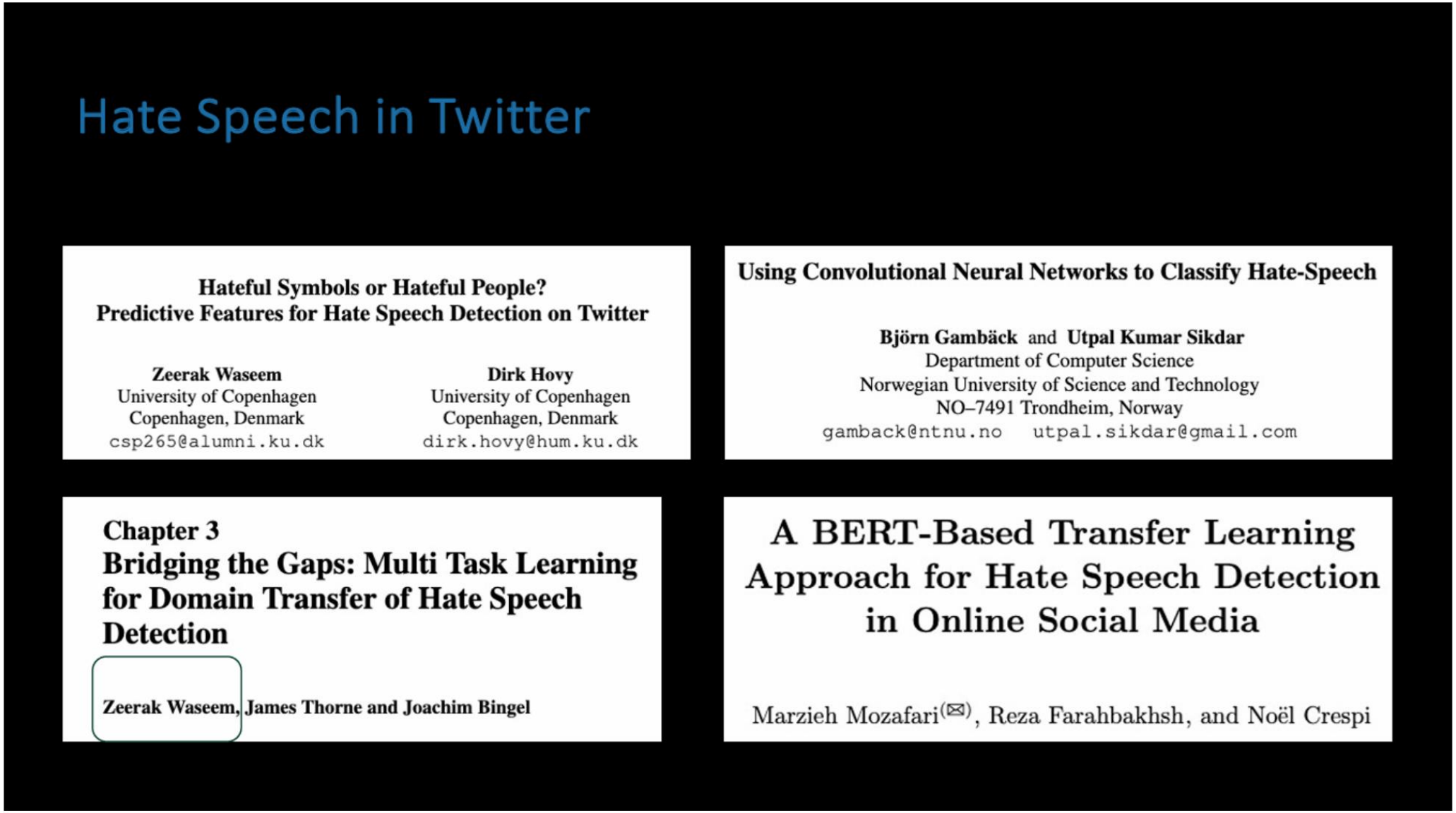
Feature Extraction and Classification

Features	Classifiers	Datasets
Sentiment	Decision Trees	Twitter
Punctuation	SVM	Facebook
Word usage	Transformers	BuzzFeed

Implemented using **Python** libraries like `scikit-learn`, `nltk.tokenize`, `sklearn.dummy`, `sklearn.decomposition`, `transformers`, `nltk`, etc.

Hate Speech Recognition

Examples of Hate Speech



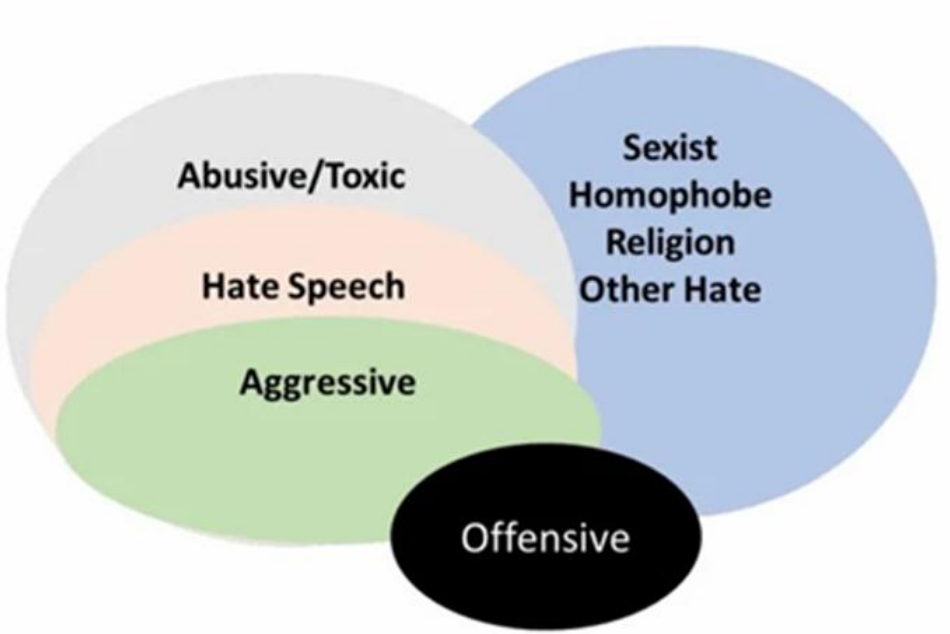
Hate Speech on Social Media

Why It Matters:

- Social media amplifies the spread of hate speech.
- Detecting hate speech is essential to reduce harm and protect individuals' beliefs.

Real-World Example:

On July 13 2023, D66 leader **Sigrid Kaag** announced her departure from politics via Twitter. She cited *"hate, intimidation, and threats"* as key reasons, especially due to its toll on her family.

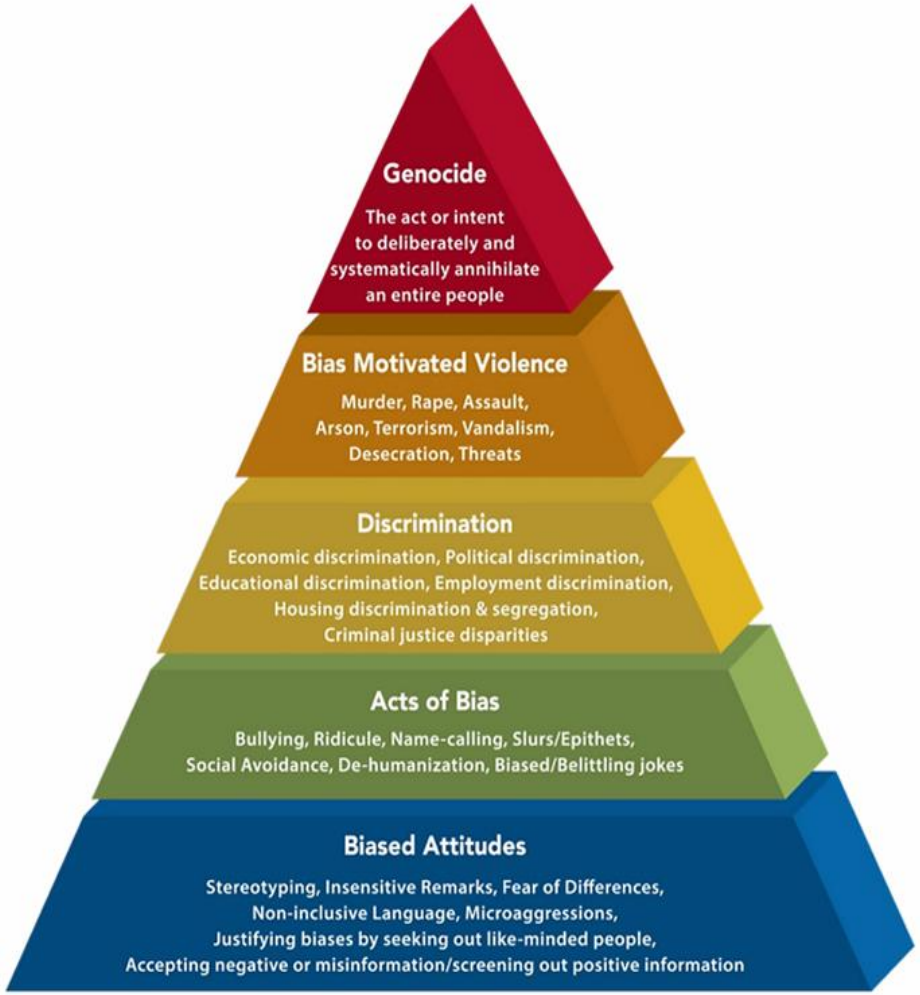


Hate Speech Pyramid (UN, 2019

United Nations Definition of Hate Speech

"Attacks or use of pejorative or discriminatory language with reference to a person or group based on identity factors like religion, ethnicity, race, gender, or nationality."

- Introduced in the **2019 UN Strategy and Plan of Action on Hate Speech**.
- Minor expressions (e.g., slurs, stereotypes) form the **foundation** of hate.
- Genocidal acts evolve from normalized, tolerated hate speech.
- The model emphasizes early intervention to prevent escalation.



Hateful Symbols Detection Study

Dataset: 16k tweets annotated for hate based on gender and race (from CRT and Gender Studies).

Method: TF-IDF with character **uni-, bi-, tri-grams** to capture spelling variation (e.g., "w0m3n").

Preprocessing: Removed stopwords **(except "not")**, usernames, and punctuation.

Classifier: Logistic Regression

System Setup	Precision	Recall	F1-score
Char n-gram + LR	0.83	0.77	0.74

Media Content Analysis

Examples of Media Content

Application of Text Clustering in Media

RESEARCH ARTICLE

Framing COVID-19: How we conceptualize and discuss the pandemic on Twitter

Philipp Wicke[1]*, **Marianna M. Bolognesi**[2]

Media Framing Dynamics of the 'European Refugee Crisis': A Comparative Topic Modelling Approach

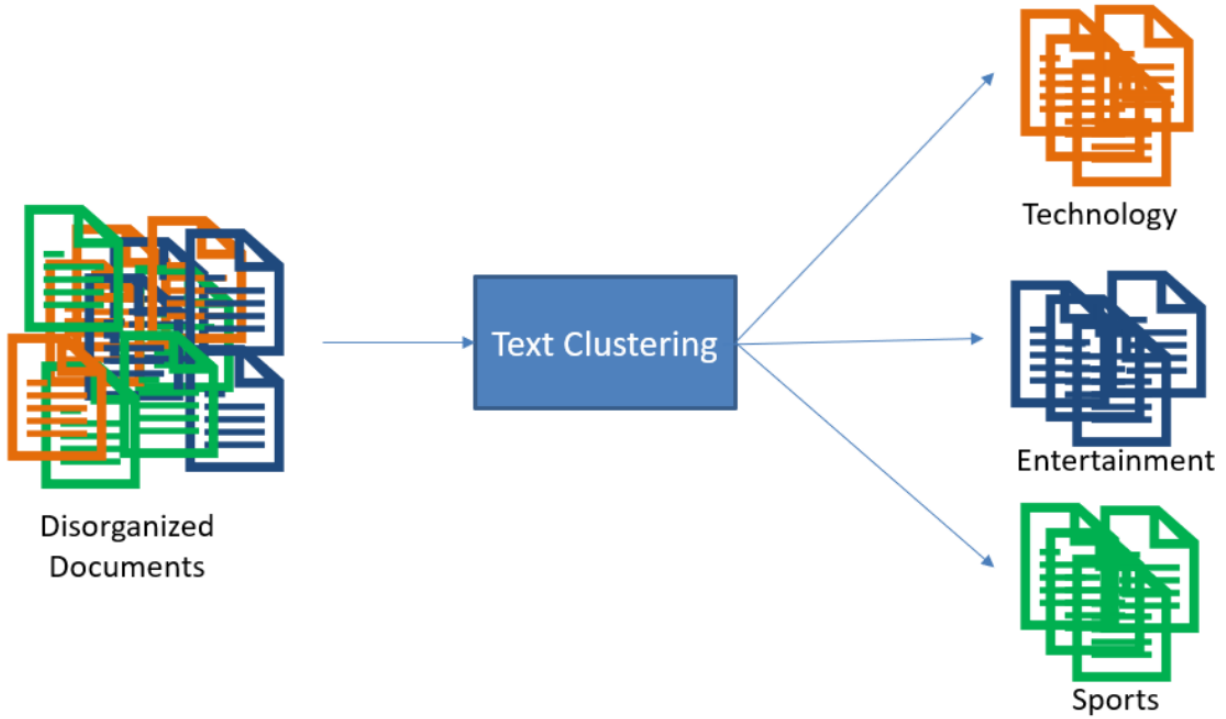
Tobias Heidenreich, Fabienne Lind, Jakob-Moritz Eberl, Hajo G Boomgaarden

Journal of Refugee Studies, Volume 32, Issue Special_Issue_1, December 2019, Pages i172–i182, https://doi.org/10.1093/jrs/fez025

Published: 27 December 2019 **Article history**

What is Media Content Analysis?

- Media content analysis uncovers how topics are **structured and framed** across different media.
- **Text clustering** helps group large volumes of unstructured content (e.g., articles, tweets) into meaningful themes.
- Used to identify trends, topics, or sentiment in domains like technology, entertainment, and sports.
- Enables journalists, researchers, and policymakers to track media focus and narrative shifts.



Example: CIVID-19 Framing on Twitter (Wicke & Bolognesi (2020))

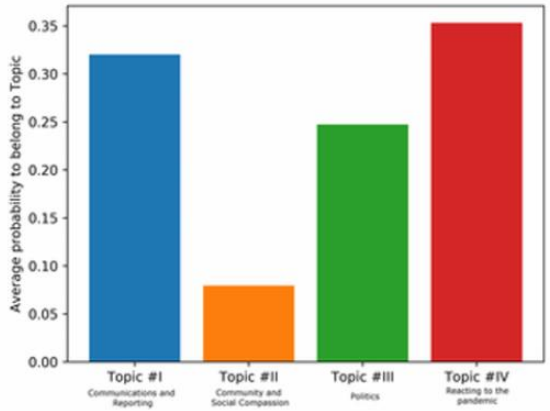
Aim: Examine how metaphorical frames like *WAR* are used in COVID-19 discourse on Twitter.

Data: 25,000 tweets/day, 80 hashtags.

Method: LDA Topic Modeling (4/16 topics), with frame correlation.

Main Frames: WAR, STORM, MONSTER, TSUNAMI.

Key Finding: 5.32% of tweets include *war-related terms*.



LDA-predicted average probability of WAR term contributing to one of 4 topics.

The results show that 5.32% of all tweets contain war-related terms

Healthcare Risk Prediction

Examples of Applications in Health

Applications in Health: Automating coding

ICD-10 Coding of Spanish Electronic Discharge Summaries: An Extreme Classification Problem

Publisher: IEEE | Cite This | PDF

Mario Almagro ; Raquel Martínez Unanue ; Víctor Fresno ; Soto Montalvo **All Authors**

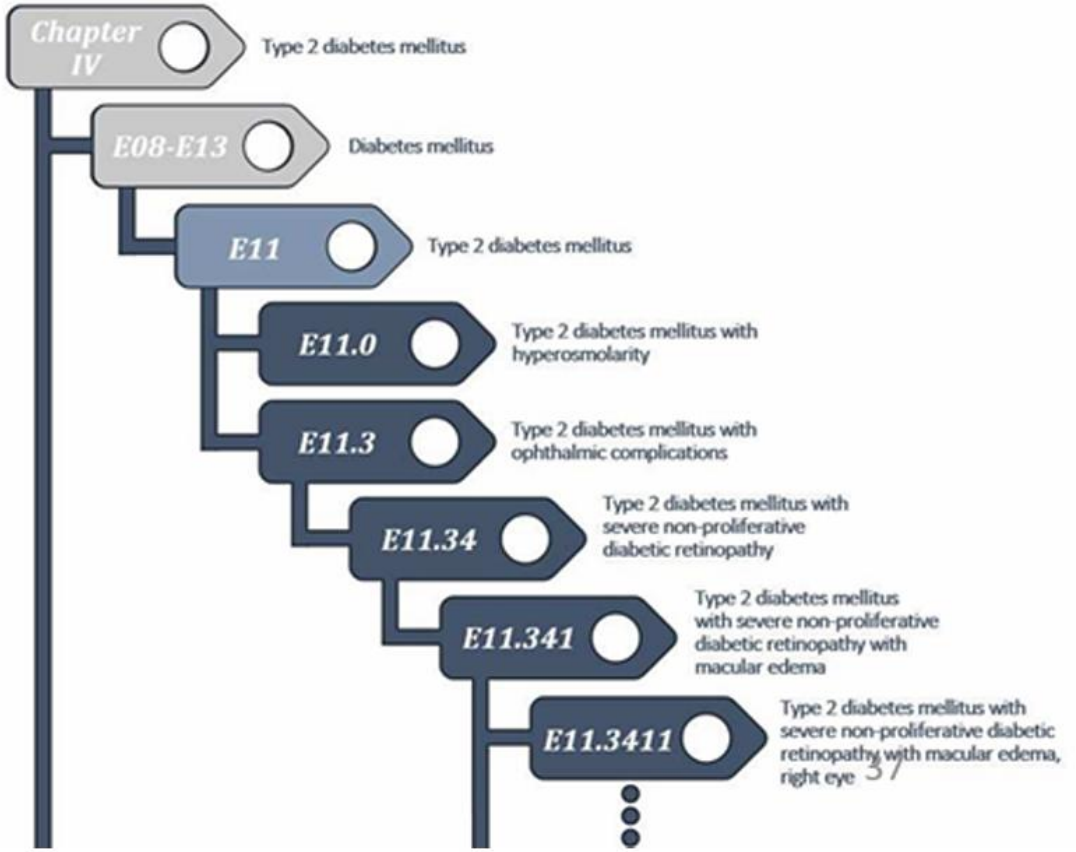
Automatic multilabel detection of ICD10 codes in Dutch cardiology discharge letters using neural networks

Arjan Sammani, Ayoub Bagheri, Peter G. M. van der Heijden, Anneline S. J. M. te Riele, Annette F. Baas, C. A. J. Oosters, Daniel Oberski & Folkert W. Asselbergs

npj Digital Medicine **4**, Article number: 37 (2021) | Cite this article

ICD-10 Coding

- Medical coding is used to identify and standardize clinical concepts in the records collected from healthcare services.
- The ICD-10 is the most widely-used coding system, with more than 11,000 different diagnoses.
- This coding system affects research, reporting, and funding.



ICD-10 Coding Example: Almagro et al., 2020

Objective:

- Suggest top 10 ICD-10 codes per case for expert review.

Dataset:

- 7,000 discharge reports with ICD-10 codes.
- Avg. 10 codes per report (cardinality = 10).

Preprocessing:

- Remove non-technical sentences (tagging-based)
- Strip accents, punctuation
- Apply stemming

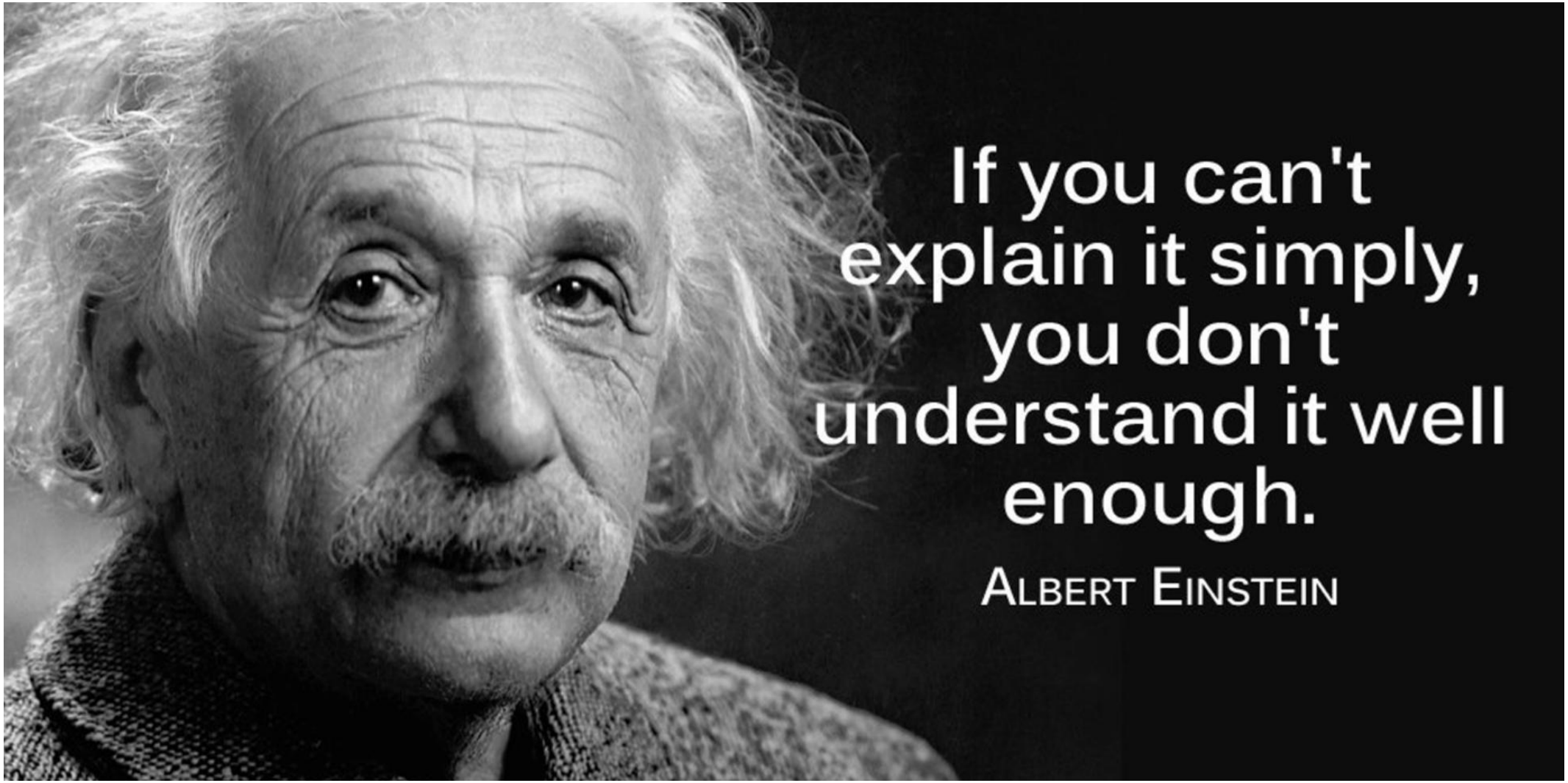
Evaluation:

- **P@10:** Precision in top 10 predicted codes.

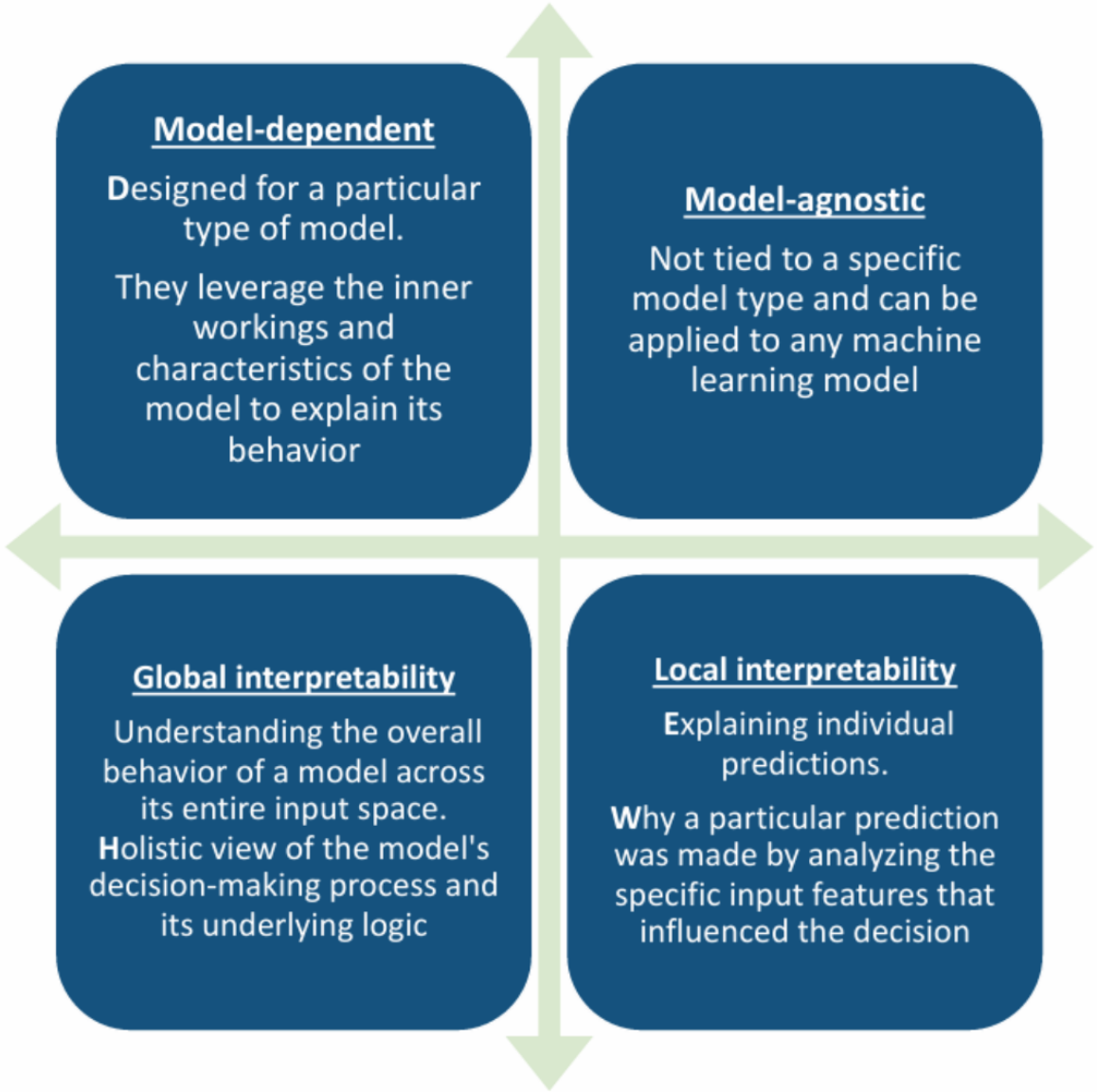
Method	**P@10**
Baseline	14.59
SVMs	37.06
MLPs	35.28
AdaBoost	36.36
Gboost	40.88
KLD	16.52
Document-Similarity	29.37
LSTM	15.08
XML-CNN	24.99
FastXML	29.87
SLEEC	27.00
Dependency-LDA	31.96
Voting (Final)	**0.xx**

Responsible AI

Interpretability in Supervised Learning



Interpretability: Being Right for the Right Reasons

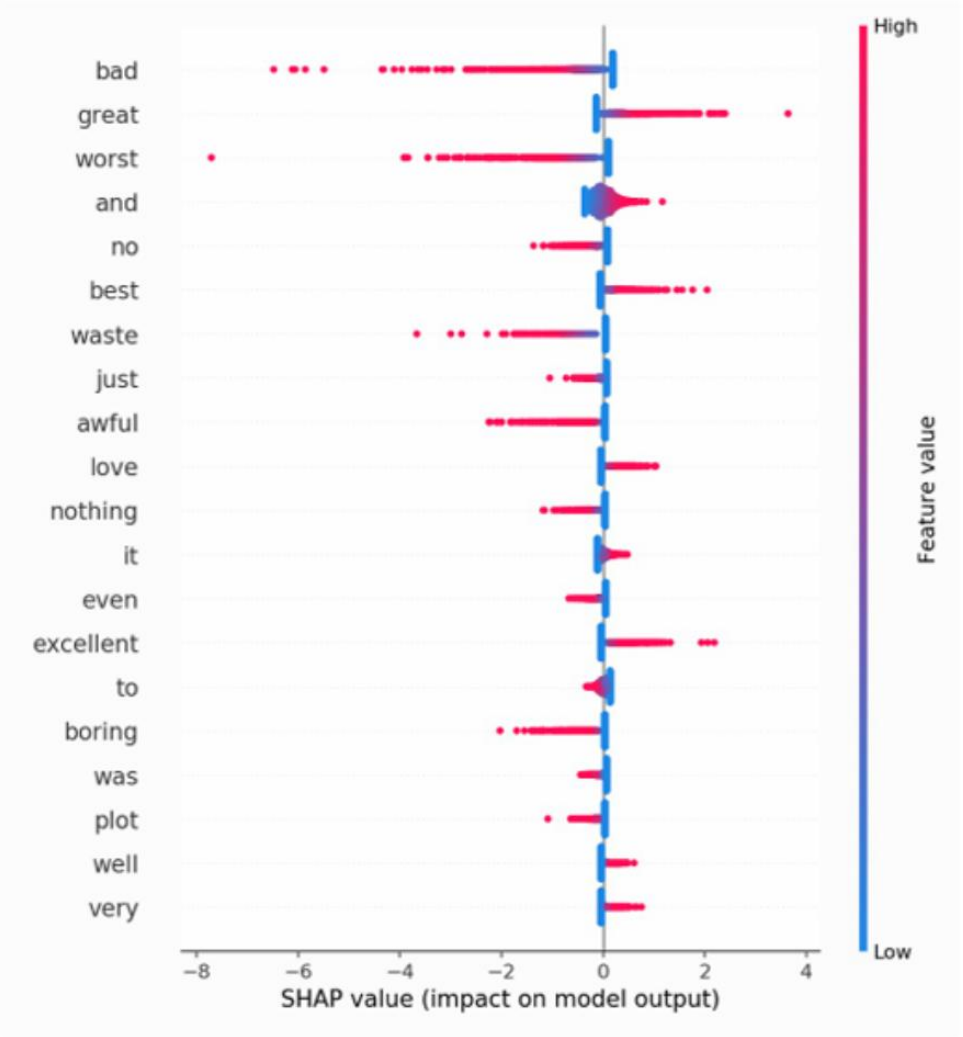


Global Interpretability

University of Bremen

Example: Sentiment Analysis Model

- Train a model to classify text sentiment (positive/negative).
- Analyze the model's feature importance or coefficients.
- Discover that emotionally charged words (e.g., "happy", "angry") have the highest importance scores.
- This reveals general patterns the model uses for classification — not just on one instance, but across the whole dataset.



References

- Jurafsky, D., Martin, J.H. (2024). *Speech and Language Processing*, third edition. Find online chapters here
- Eisenstein, J. (2018). *Natural Language Processing*. Find online chapters here
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake news detection on social media: A data mining perspective*. doi: 10.1145/3137597.3137600
- Waseem, Z., & Hovy, D. (2016). *Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter*. aclanthology.org/N16-2013

DATA SCIENCE CENTER

University of Bremen

Thank you for your attention!

MORE INFOS

www.dsc-ub.de

@DSC_unibremen