



Tackling Climate Change with Machine Learning

NeurIPS 2020 Virtual Workshop

CLIMATEX: A Dataset for Climate Change Topic Detection

Francesco Saverio Varini

Department of Computer Science
ETH Zurich
Zurich, Switzerland
fvarini@student.ethz.ch

Jordan Boyd-Graber

CS, iSchool, LSC, and UMIACS
University of Maryland
College Park, MD, USA
jbg@umiacs.umd.edu

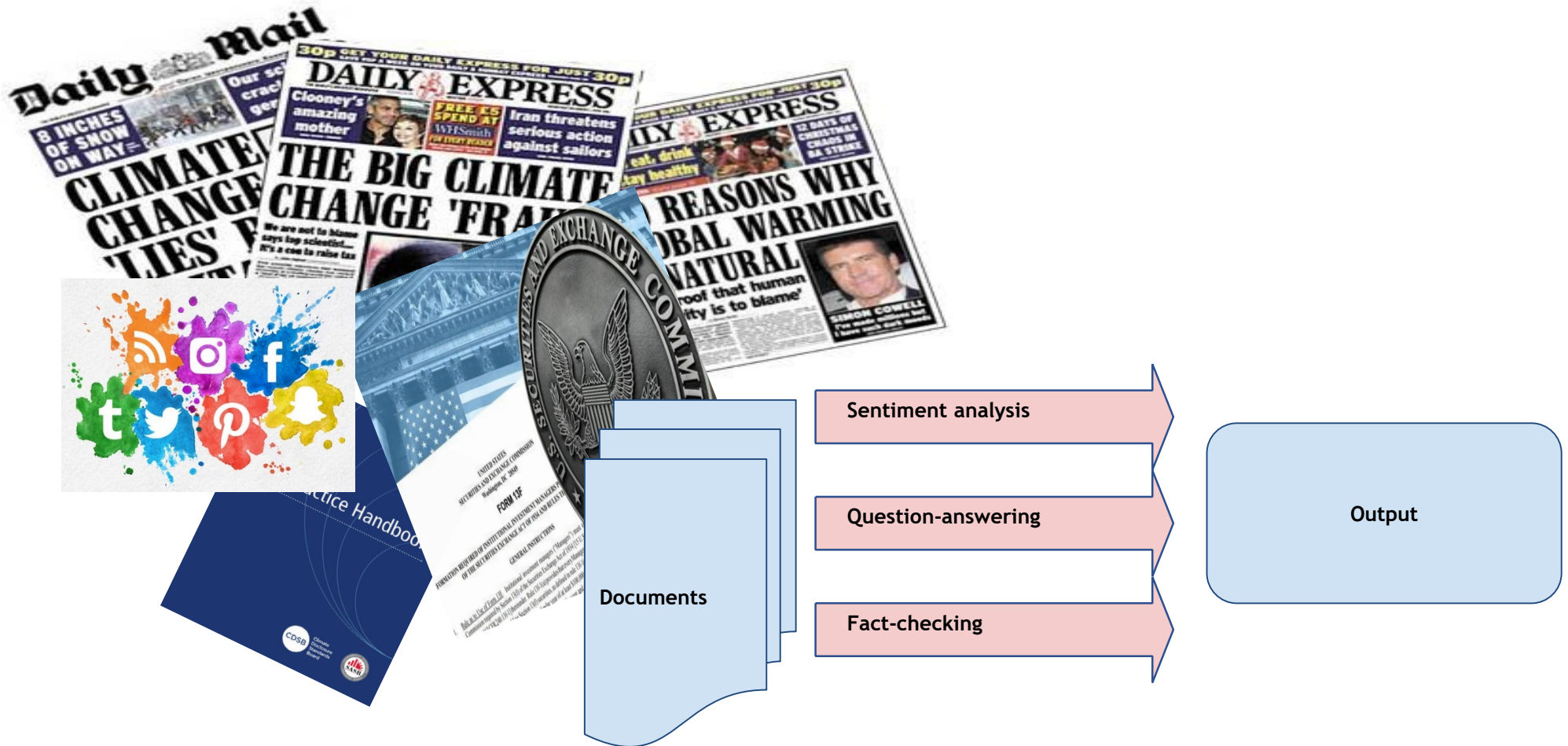
Massimiliano Ciaramita

Google Research
Zurich, Switzerland
massi@google.com

Markus Leippold

Department of Banking and Finance
University of Zurich
Zurich, Switzerland
markus.leippold@bf.uzh.ch

Why topic detection for climate change?



Not as easy as it seems...

- Some examples:

1. Compliance with these laws and regulations could require significant commitments of capital toward environmental monitoring, renovation of storage facilities or transport vessels, payment of emission fees and carbon or other taxes, and application for, and holding of, permits and licenses.
 2. Al Gore's book is quite accurate, and far more accurate than contrarian books.
 3. The temperature is not rising nearly as fast as the alarmist computer models predicted.
 4. The parties also began discussing the post-Kyoto mechanism, on how to allocate emission reduction obligation following 2012, when the first commitment period ends.
 5. The rate of Antarctica ice mass loss has tripled in the last decade.
 6. Globally about 1% of coral is dying out each year.
 7. Our landfill operations emit methane, which is identified as a GHG.
 8. Due to concerns about the de-forestation of tropical rain forests and climate change, many countries that have been the source of these hardwoods have implemented severe restrictions on the cutting and export of these woods..
 9. The 2015 conference was held at Le Bourget from 30 November to 12 December 2015.
 10. Polar bear numbers are increasing.
-

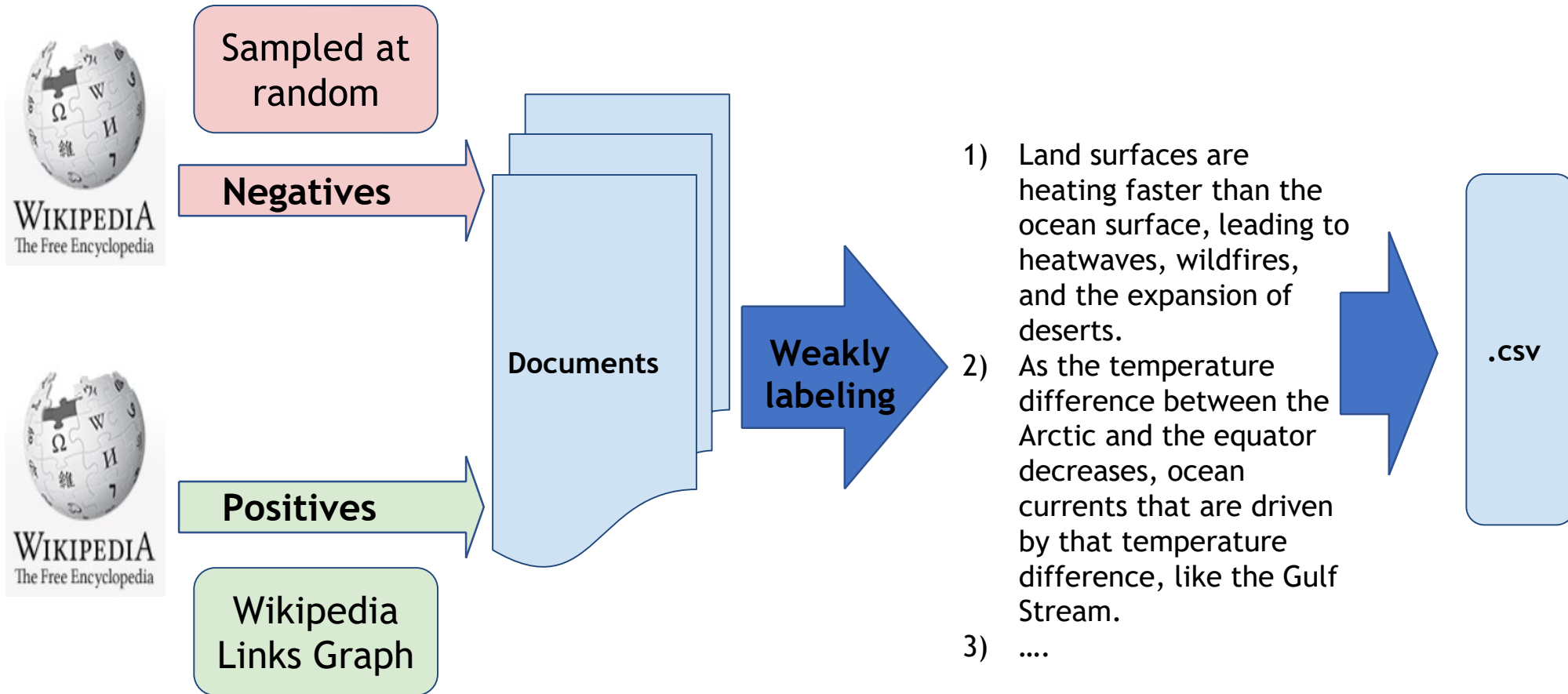
CLIMATEXT

Data	Tag	Sentences
Train split	Wiki-Doc-Train	115854 (57927 vs 57927)
Development split	Wiki-Doc-Dev	3826 (1913 vs 1913)
Test split	Wiki-Doc-Test	3826 (1913 vs 1913)

Data	Tag	Sentences
Wikipedia	AL-Wiki	3000 (261 vs 2739)
10-Ks	AL-10Ks	3000 (58 vs 2942)

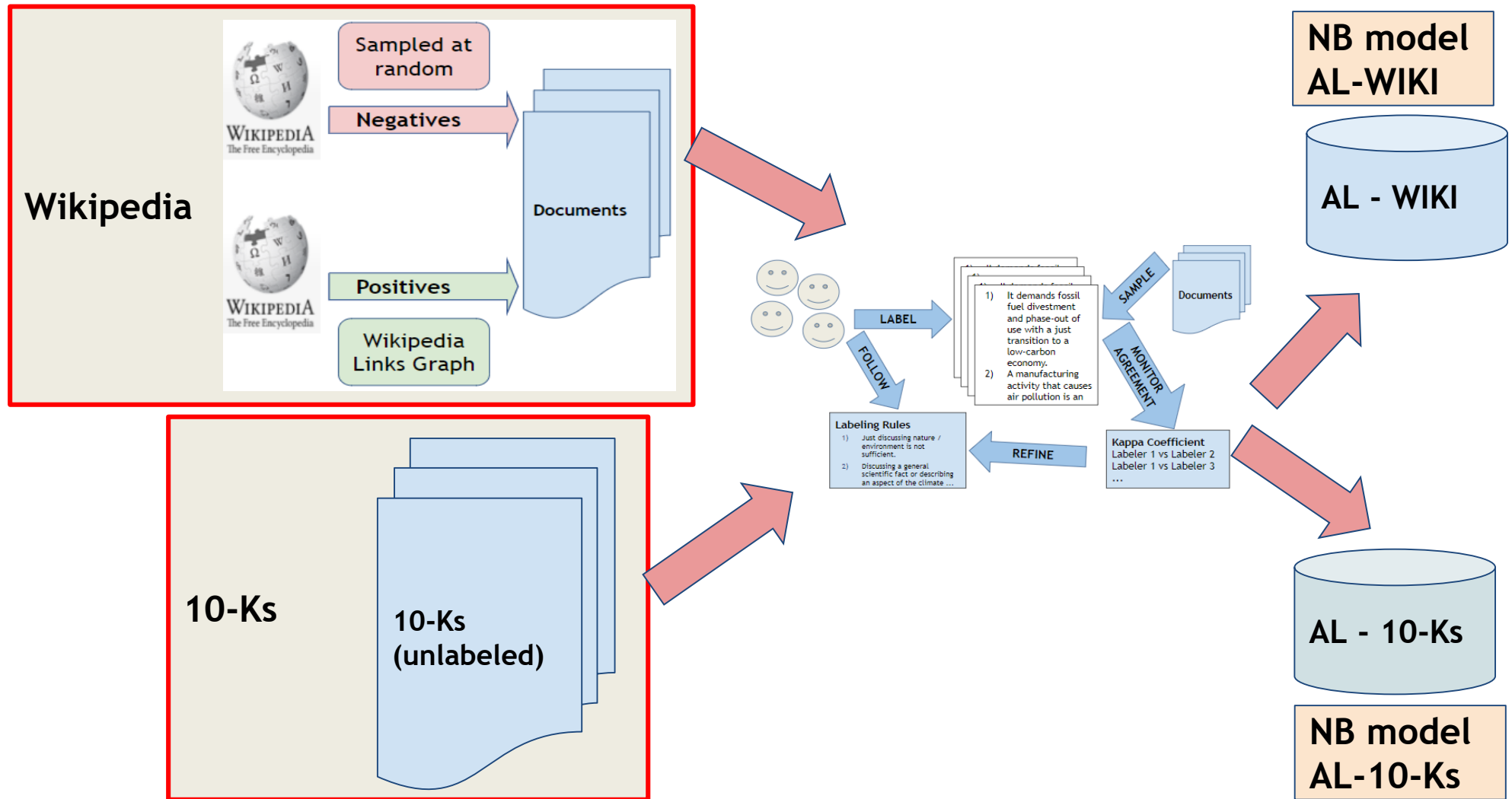
Data	Sentences
Wikipedia (dev)	300 (79 vs 221)
Wikipedia (test)	300 (33 vs 267)
10-Ks (2018, test)	300 (67 vs 233)
Claims (test)	1000 (500 vs 500)

Wikipedia data (Wiki-Doc-Train/Dev/Test)

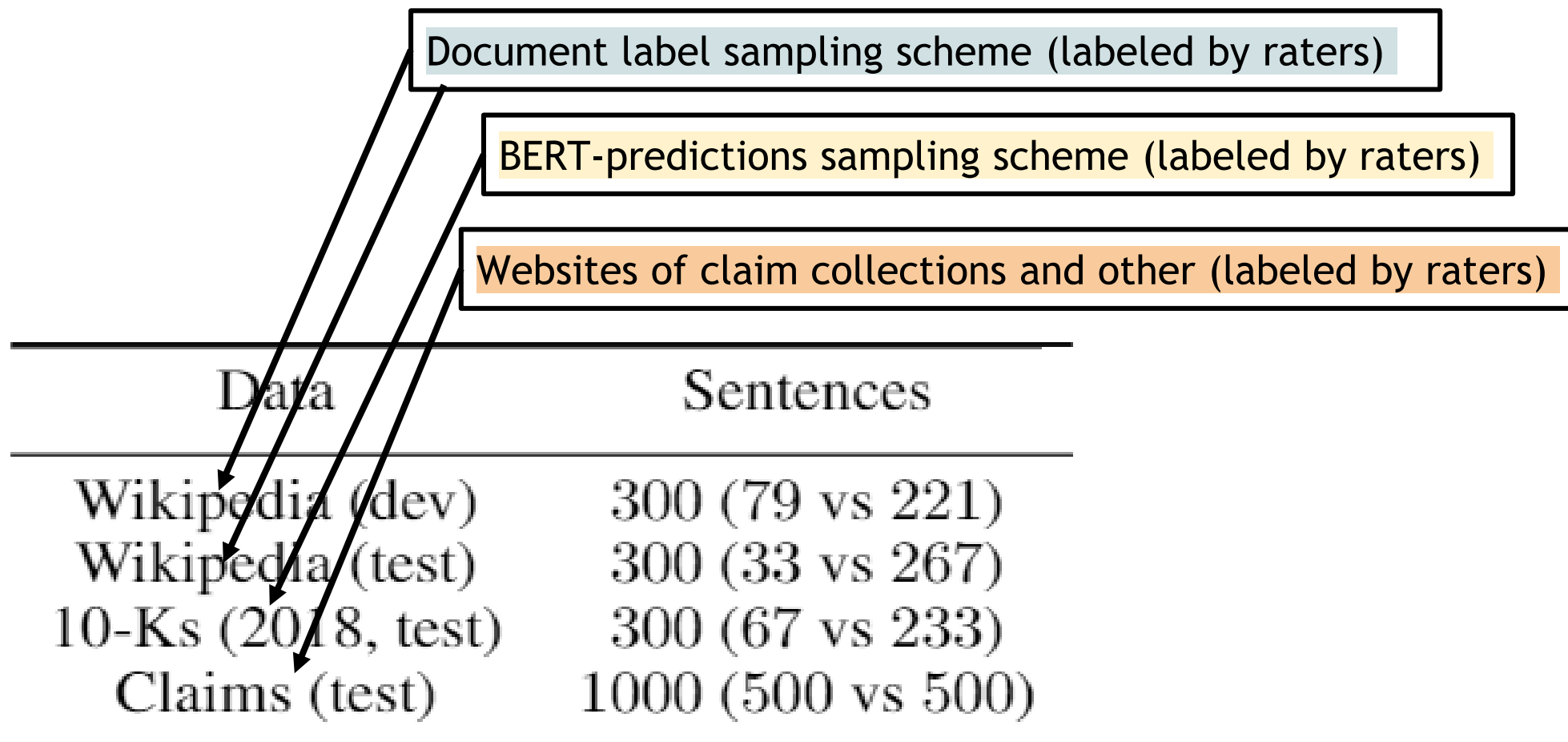


Tackling Climate Change with Machine Learning

AL-Wiki and AL-10Ks train data



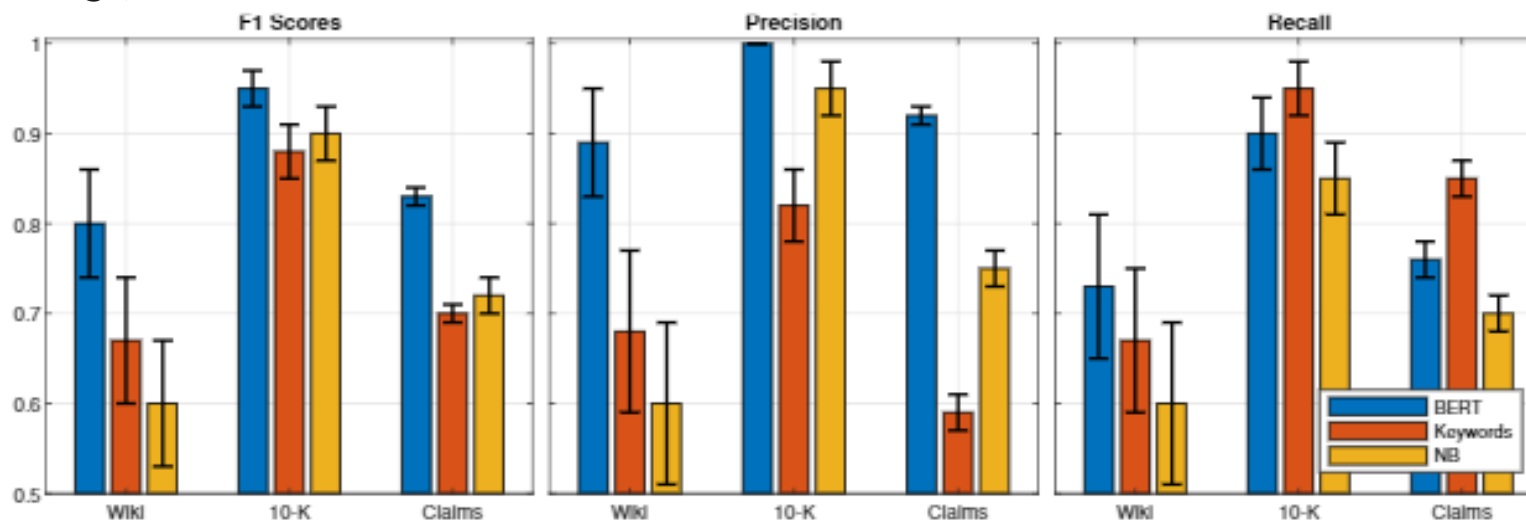
Wikipedia, 10K, and Claims evaluation data



Main Result

Our contributions:

1. We introduce CLIMATEX, a dataset for sentence-based climate change topic detection, which we make publicly available.
2. We analyze keyword-based, naïve-Bayesian, and a BERT-based approach to explore their performance in identifying climate-change relevant text in Wikipedia, 10-K filings, and in climate-related claims database.



Going forward:

- Make the annotated data public.
- Improve algorithms to detect climate-change topic in a wide range of text sources.