

ML4CC: Lecture 7

Sit with your discussion groups (same as last time)!

Assignments reminder

No PMIRO+Q this week

You have an exam on **March 14th (8am)**

Your third coding assignment will be due **Friday, March 15th** by 11:59pm.

Recap of previous paper

P: Want to be able to find rice paddies in countries without a lot of labeled data

M: Use transfer learning - pretrain a U-net for segmentation on Korean data, then fine tune it for France and Spain

I: Multiple forms of fine-tuning, including adding another input channel

R: There were some situations where the transfer learning helped but Random Forest models trained on small data did pretty well

O: How well will this approach generalize to other crops (that aren't flooded)

Climate Change in the News



Fury after Exxon chief says public to blame for climate failures

Darren Woods tells Fortune consumers not willing to pay for clean-energy transition, prompting backlash from climate experts



As the [world's largest investor-owned oil company](#), Exxon is among the [top contributors](#) to global planet-heating greenhouse gas emissions. But in an interview, published on Tuesday, Woods argued that big oil is not primarily responsible for the climate crisis.

The real issue, Woods said, is that the clean-energy transition may prove too expensive for consumers' liking.

"The dirty secret nobody talks about is how much all this is going to cost and who's willing to pay for it," [he told Fortune last week](#). "The people who are generating those emissions need to be aware of and pay the price for generating those emissions. That is ultimately how you solve the problem."

"We have opportunities to make fuels with lower carbon in it, but people aren't willing to spend the money to do that."

Experts say Woods's rhetoric is part of a larger attempt to skirt climate accountability. [No new major oil and gas infrastructure can be built](#) if the world is to avoid breaching agreed temperature limits but Exxon, along with other major oil companies currently [basking in record profits](#), is pushing ahead with aggressive [fossil-fuel expansion plans](#).

"It's like a drug lord blaming everyone but himself for drug problems," said Gernot Wagner, a climate economist at Columbia business school.

"I hate to tell you, but you're the chief executive of the largest publicly traded oil company, you have influence, you make decisions that matter. Exxon are at the mercy of markets but they are also shaping them, they are shaping policy. So no, you can't blame the public for the failure to fix climate change."

Paper 6 Discussion

MINING EFFECTIVE STRATEGIES FOR CLIMATE CHANGE COMMUNICATION

Aswin Suresh

EPFL

Lazar Milikic

EPFL

Francis Murray

EPFL

Yurui Zhu

EPFL

Matthias Grossglauser

EPFL

Tackling Climate Change with
Machine Learning: workshop at
ICLR 2023.

ABSTRACT

With the goal of understanding effective strategies to communicate about climate change, we build interpretable models to rank tweets related to climate change with respect to the engagement they generate. Our models are based on the Bradley-Terry model of pairwise comparison outcomes and use a combination of the tweets' topic and metadata features to do the ranking. To remove confounding factors related to author popularity and minimise noise, they are trained on pairs of tweets that are from the same author and around the same time period and have a sufficiently large difference in engagement. The models achieve good accuracy on a held-out set of pairs. We show that we can interpret the parameters of the trained model to identify the topic and metadata features that contribute to high engagement. Among other observations, we see that topics related to climate projections, human cost and deaths tend to have low engagement while those related to mitigation and adaptation strategies have high engagement. We hope the insights gained from this study will help craft effective climate communication to promote engagement, thereby lending strength to efforts to tackle climate change.

Attendance

Select one person from the group to go to this Google Doc and write down the names of all people present in the group (remember to mark who took attendance!). **If someone is virtual, mark it with a V.**

<https://docs.google.com/document/d/1PKhw9E2IJpAnFrFO88DOc2rscZFVlclv47QNa5h1sGs/edit?usp=sharing> (link is in Brightspace under Syllabus content)

Discussion Question 1

How precisely is tweet “engagement” defined? Do you think this is a good way to measure how effective communication is?

Sum of likes, retweets, and replies

and obtain pairs of tweets that were created within 7 days of each other and have a difference in engagement (sum of likes, retweets and replies) of 100 units or 10%, whichever is higher. This

People can engage with a tweet for many reasons, including to support it but also argue against it. This method collapses over different styles of engagement and has no means of measuring positive vs negative engagement.

Discussion Question 2

Why do the authors focus on the pairwise comparison of two tweets? What are the constraints they use to pick pairs? Do these constraints significantly impact the amount of data they can train on?

By comparing similar tweets with different content, they can isolate the impact of content

However, a significant challenge in building such models is the presence of confounding factors such as author popularity. A tweet might generate strong engagement because its author is popular rather than because of its engaging content. Another potential confounder is the change in public interest in climate change over time. For instance, tweets about climate change made around the time of extreme weather or a major climate change conference might receive greater attention than tweets at other times. We minimise the effect of such confounders by defining the task as comparing the engagement within a pair of tweets rather than predicting the engagement for a given tweet. The pair of tweets are chosen to be from the same author and from the same window in time.

A model trained to predict a single tweet's engagement would probably do well simply by guessing based on author popularity. By forcing the model to predict which of two tweets by the same author in the same time period gets more engagement, the model needs to learn the impact of tweet content

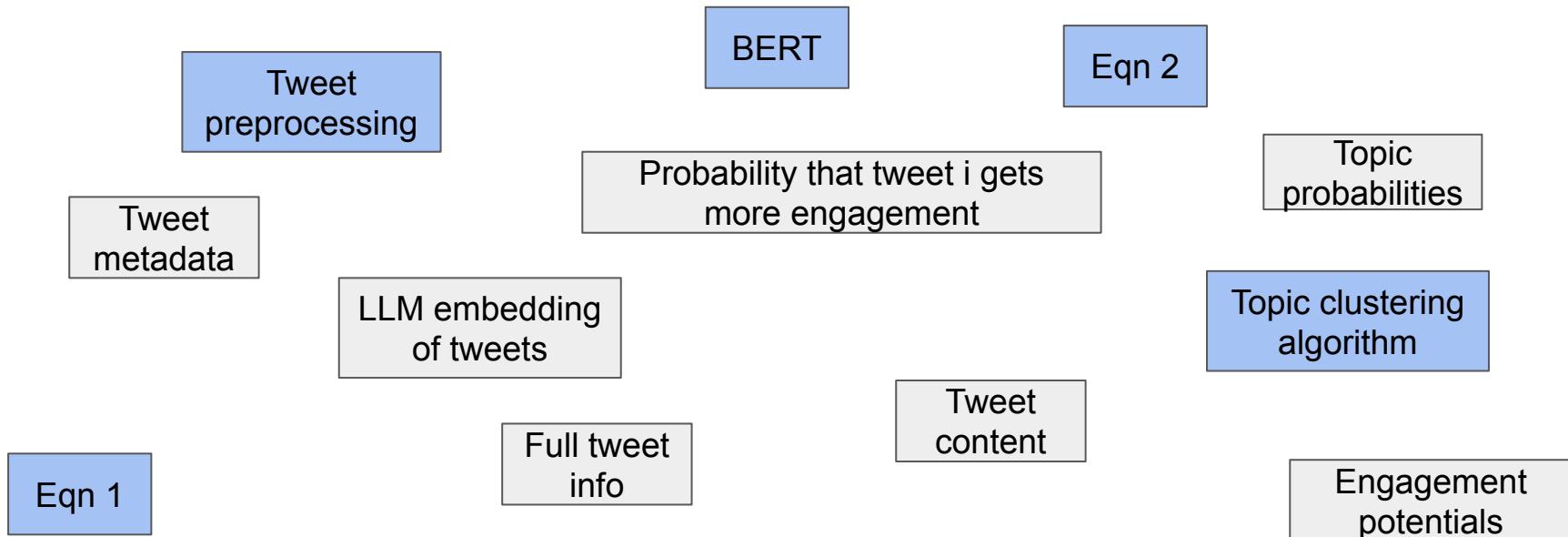
We use the Twitter API to obtain 8,041,921 tweets related to climate change created between January 1st 2021 and November 4th 2022. To decide whether a tweet is related to climate change, we check if it contains one of the keywords in the ‘General’ topic category of UN Global Pulse (2014), a taxonomy for studying climate change tweets. We keep only the tweets in English (94.28% of the dataset). For each tweet, we keep its full text, author and information about whether it contains URLs, hashtags, animated GIFS, images or videos. We also keep the public engagement metrics about the tweet, i.e. the number of likes, retweets and replies that it obtained.

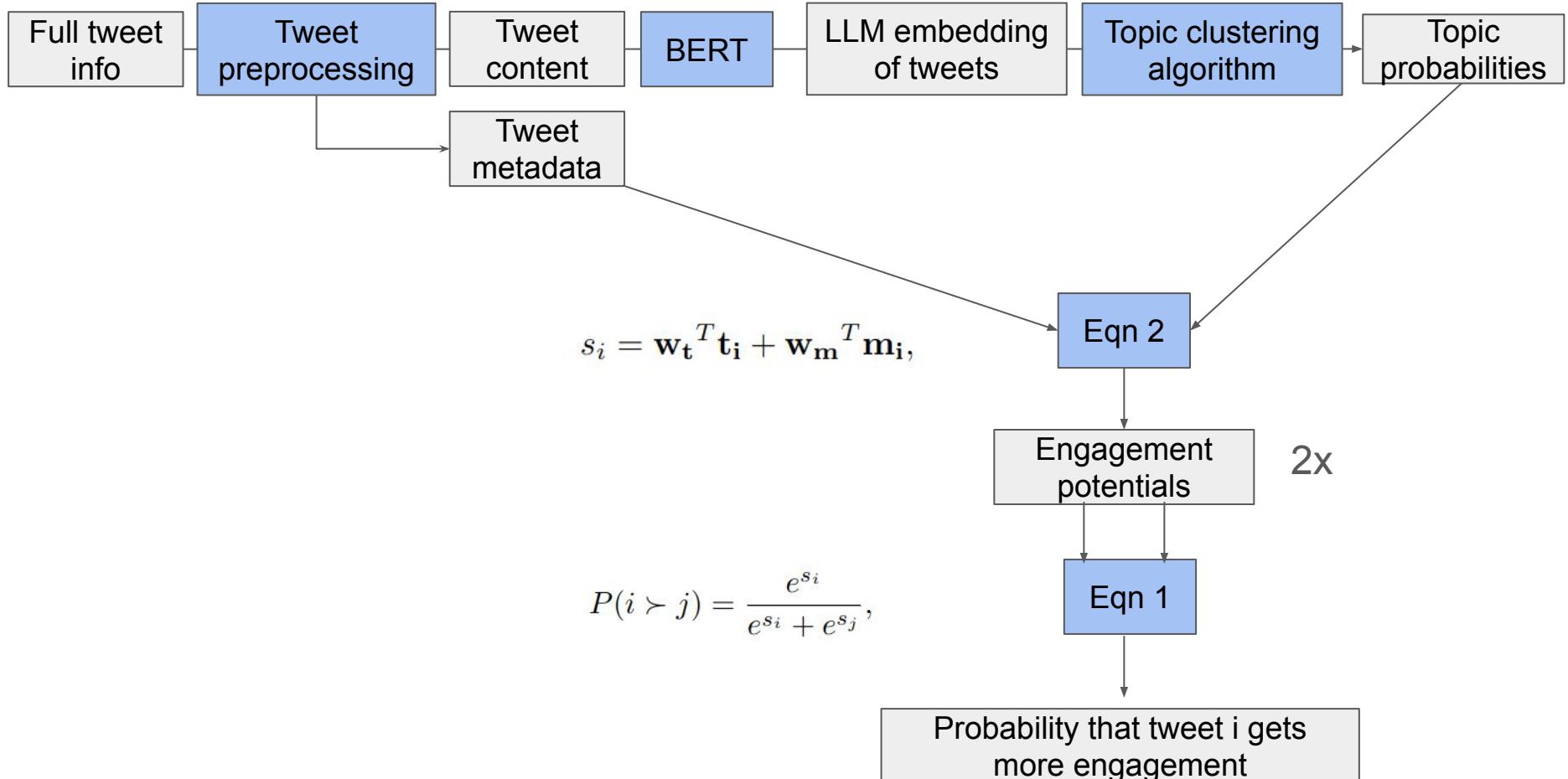
We then construct the pairs of tweets to compare. For each author, we go through their history and obtain pairs of tweets that were created within 7 days of each other and have a difference in engagement (sum of likes, retweets and replies) of 100 units or 10%, whichever is higher. This method of creating pairs ensures that we avoid confounding factors related to the author and time and minimise the noise in the comparison. We finally end up with 774,507 pairs of tweets which we use for training our engagement prediction model.

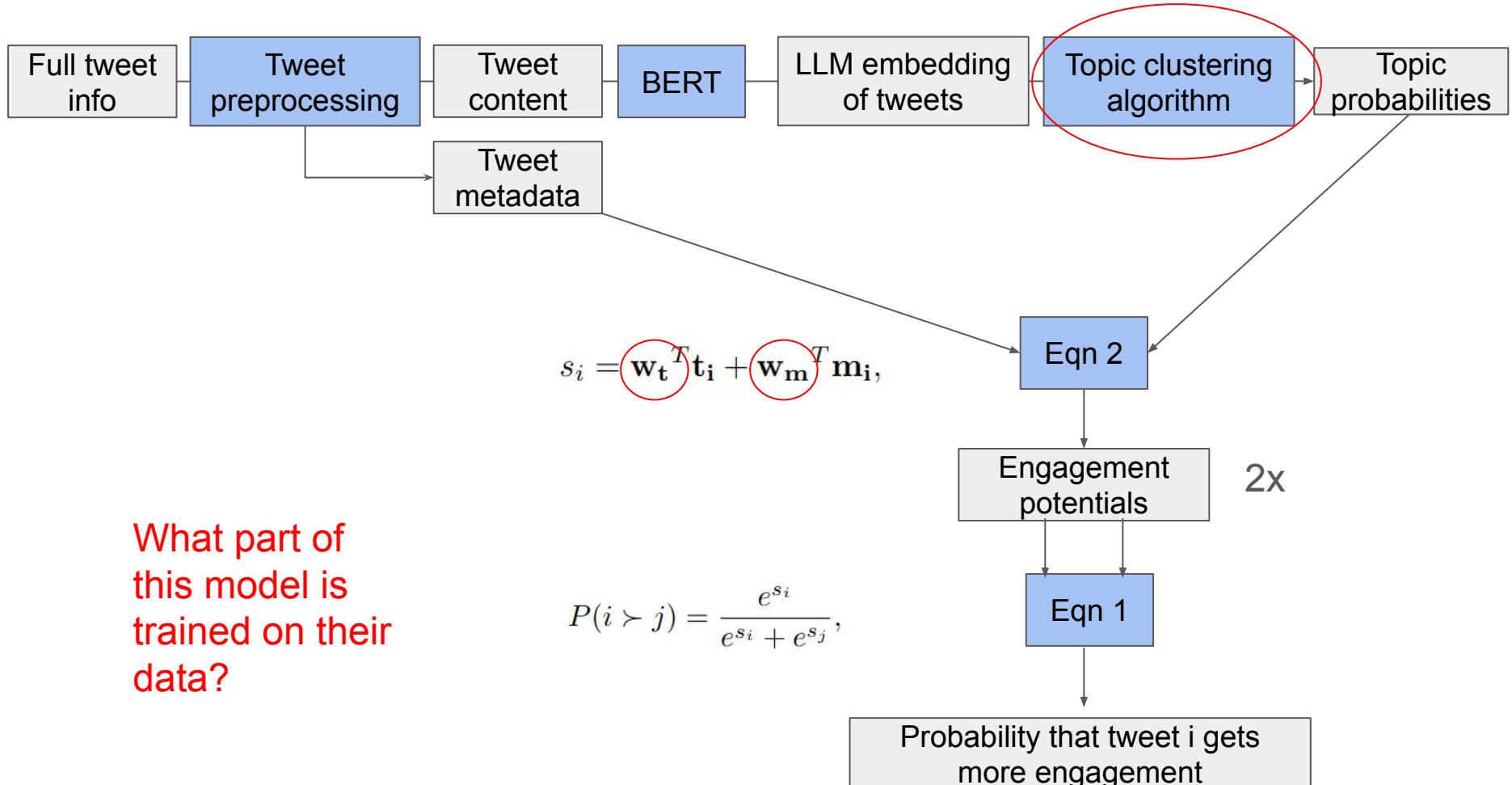
774,507 pairs is a lot less than the $\sim 8,041,921^2$ possible pairs.
(but still a good amount of pairs)

Discussion Question 3

Draw a box and arrow diagram that explains the conceptual architecture of the model built in the paper. Make sure it includes the following boxes:







Discussion Question 4

The authors describe their model as “interpretable”. What about the design makes the model easy to interpret/learn from?

Understandable features + linear model

The output of the topic clustering algorithm is a vector of probabilities associated with each topic. The metadata features are also easily interpretable. The weights assigned to these features by the linear model directly indicate how important each feature is for tweet engagement.

Table 2: Feature coefficients with 96% confidence intervals, computed from 300 bootstrap samples

$$s_i = \mathbf{w_t}^T \mathbf{t_i} + \mathbf{w_m}^T \mathbf{m_i},$$

Feature	Coeff.	Feature	Coeff.	Feature	Coeff.
President	7.71 ± 0.78	Geology	0.79 ± 0.77	Investment	-6.70 ± 0.81
Clean Energy	4.25 ± 0.82	Mixed	0.00 ± 0.10	Human cost	-8.41 ± 0.90
Drought-resistant	4.00 ± 1.12	Low Water	-0.38 ± 0.90	Projections	-10.68 ± 0.85
Africa	2.79 ± 0.90	Conference	-1.65 ± 0.99	Links/Promo	-13.71 ± 1.26
Planet	2.77 ± 0.84	Research	-2.12 ± 0.84	Meta:URL	-1.54 ± 0.13
Fossil fuels	1.56 ± 0.82	Youth	-2.70 ± 1.12	Meta:Hashtag	-0.12 ± 0.12
Politics	1.51 ± 0.79	Health	-3.58 ± 0.94	Meta:GIF	0.47 ± 0.28
Global warming	1.18 ± 0.84	Rain	-3.62 ± 1.03	Meta:Video	0.76 ± 0.13
Geopolitics	1.09 ± 0.77	News	-5.07 ± 0.93	Meta:Image	0.58 ± 0.10
Anger	0.79 ± 0.90	Deaths	-5.43 ± 0.89	Meta:WorkHr	-0.24 ± 0.10

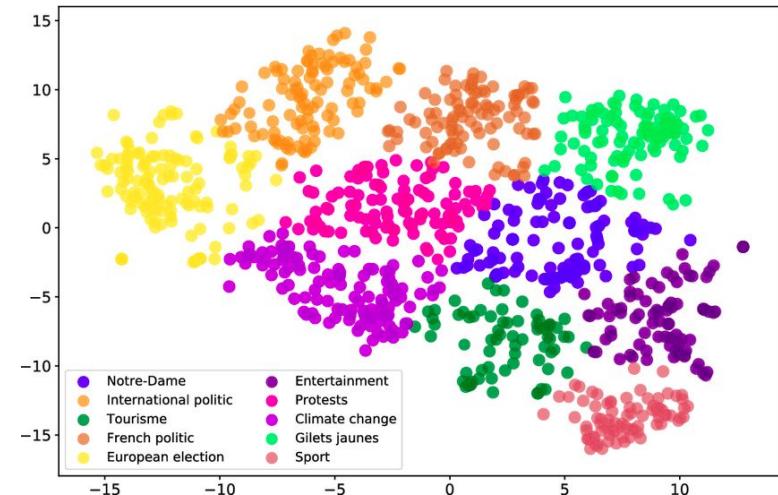
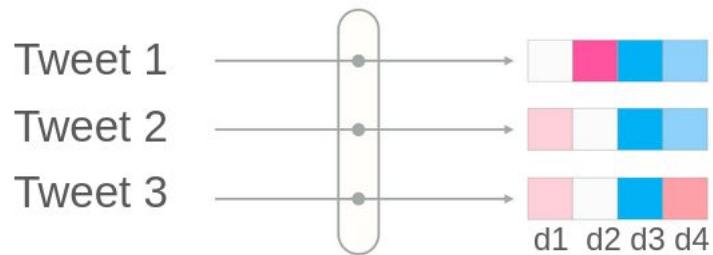
Discussion Question 5

How were the topic labels determined?

Topic labels determined by hand

variable model to cluster these representations into topics in an unsupervised fashion. The learned topics can be interpreted by looking at the tweets whose representations are closest to the centre of the clusters. Once the model is trained on a set of tweets, it can be used to infer the probability that

features. The names of the topics were manually assigned by examining the top 500 tweets with the highest probability for the topic. A random sample of 10 tweets from this set for each of the topics in the table is given in the Appendix for a more detailed interpretation.



Discussion Question 6

Which source of information is more responsible for good model performance:
tweet topic or metadata? How do you know?

Topic contributes more to performance

Table 1: Test accuracies of different models with 95% confidence intervals

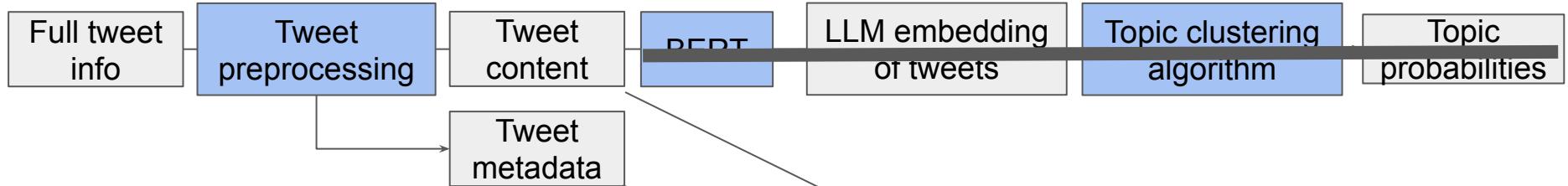
Random	Meta	Topic	Topic+Meta	Human
50.17 ± 0.50	58.90 ± 0.49	64.54 ± 0.48	66.53 ± 0.47	65.00 ± 6.61

But metadata contains extra information such that the union of both inputs performs better.

Discussion Question 7

What are the authors describing here?

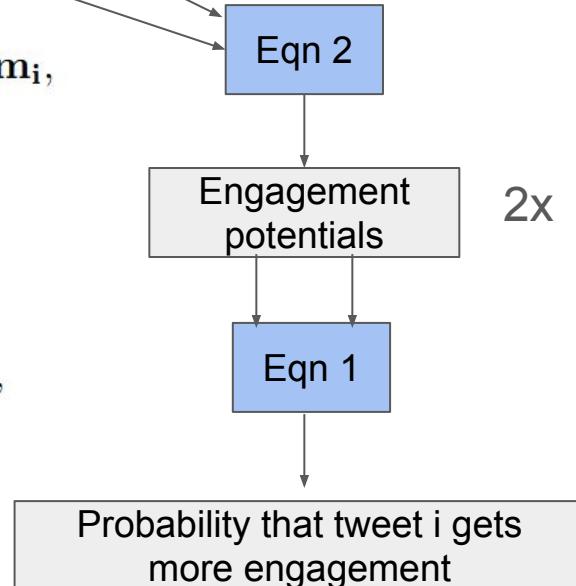
engagement of tweets containing the same URL. We also tried models using the words as features, instead of topics. However the accuracy was slightly lower, and the interpretation was more difficult for those models as a clear pattern couldn't be seen among the most predictive words. The accuracy for word-based models could possibly be increased by using contextual word embeddings and state-of-the-art sequence models like Transformers (Vaswani et al. (2017)), but their interpretation is likely to still be difficult.

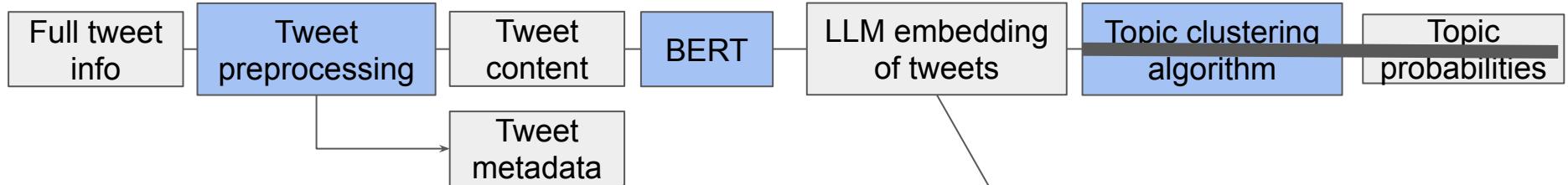


“Words as features” means the presence or absence of words (in a pre-chosen vocabulary) are used as features

$$s_i = \mathbf{w}_t^T \mathbf{t}_i + \mathbf{w}_m^T \mathbf{m}_i,$$

$$P(i \succ j) = \frac{e^{s_i}}{e^{s_i} + e^{s_j}},$$





“Words embeddings” refers to using the activity of an LLM layer as the features

$$s_i = \mathbf{w_t}^T \mathbf{t}_i + \mathbf{w_m}^T \mathbf{m}_i,$$

$$P(i \succ j) = \frac{e^{s_i}}{e^{s_i} + e^{s_j}},$$

Eqn 2

Engagement potentials

2x

Eqn 1

Probability that tweet i gets more engagement

Discussion Question 8

Share what questions you wrote in your PMIRO+Q and decide as a group what you'd like to ask.

Update your PMIRO+Q

Submit a second file to the Brightspace assignment (don't overwrite the original):

It should:

Update your PMIRO as needed

Answer your own Q

You can be talking with your group during this!

Summary

P - Want to know what climate change content in a tweet makes it engaging

M - Use an LLM plus topic clustering to get a tweet's topic and use that (plus metadata) to predict which of two tweets is more engaging

I - predicting engagement specifically for climate topics

R - the model performs above chance and on par with a human

O - Is this a good metric of engagement? Would writing tweets based on these topics actually help?

15 min break

Lecture

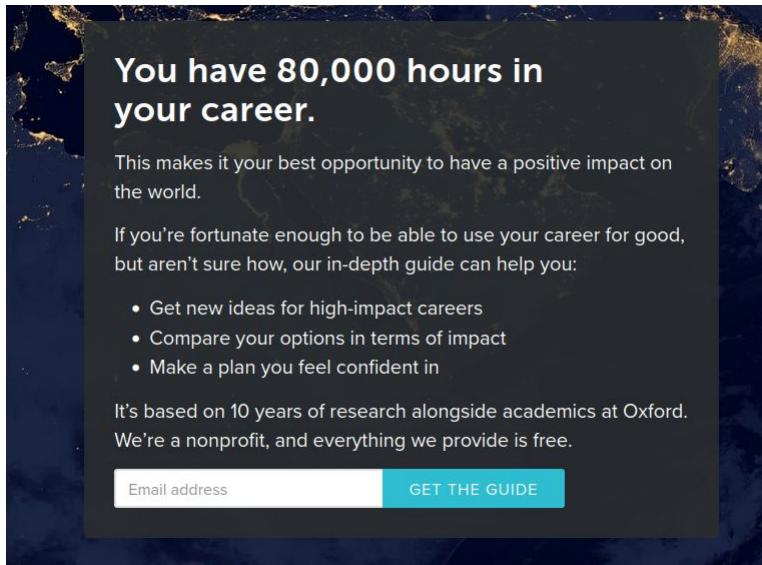
Career Day, Project Info, Exam Review

Alex Steffen, Climate Writer - *The Snap Forward*

“We don’t get to choose whether the context of all our work is an unprecedented, all-encompassing planetary crisis. It simply is. There is nowhere to stand outside of it. We can pretend that’s not true — insist that our lives or work or special concerns will remain untouched by ecological catastrophe and societal upheaval — but in the long run, we’re just fooling ourselves.”



Your career is your biggest chance to make an impact



<https://80000hours.org/>

Climate career pathways

- Work directly on climate change
- Work directly on something else but bring awareness and action on climate to your job
- Regardless of your job, engage in climate action and activism

Working on climate

WORK ON CLIMATE Programs ▾ Resources Blog About Us ▾ Donate

Join us

We help you find a way to work on climate

We are an action-oriented Slack community for people serious about climate work. Find climate jobs. Build climate companies. Find your people.

Join us on Slack



How to use the community



Join the community



Explore our programs



Meet climate experts



Find a job in climate

Share your journey, find collaborators, and celebrate each other's wins.

Find the right community offering for where you are on your path.

Drop in to meet our friendly climate experts during regular office hours.

Find a climate job via our #jobs, #jobs-alerts and #gigs channels.

Work On Climate

Threads
Direct messages
Mentions & reactions
Drafts & sent
Saved items
Slack Connect
More

Channels
announcements
community-feedback
community-organisers
events-external
events-official
general
gles
i-got-a-job
jobs
jobs-alerts
jobs-discussion
learn-psychology
learn-the-space
meet-nyc
need-help
project-wed-deck

jobs

Check guidelines before posting. Join jobs- subchannels like #jobs-software-engineering #jobs-remote #jobs-software-engineering #jobs-regen

Today

Vlad Mikhayukh 8:31 AM

Stem is hiring a Technical Product Manager for the AI Optimization team! The role is focused around working with cross functional stakeholders and SMEs to ensure Stem continues to deliver industry-leading optimization capabilities in behind-the-meter and front-of-the-meter energy storage operations and simulations. The right candidate will own and extend optimization driven services and applications at Stem.

- Location: Remote / San Francisco
- Time commitment: Full time
- Compensation: Not known
- Connection to company: Currently report to hiring manager

More details found here: <https://www.linkedin.com/jobs/view/technical-product-manager-ai-optimization-applications-at-stem-inc-3520450371/>
If you are interested, feel free to dm me directly with resume or reach out with any questions!

3

5 replies Last reply today at 12:08 PM

Laura Rosenshine 10:44 AM

NEW ROLE ALERT! TO HELP THE

WATS helps businesses streamline their waste operations, report on carbon impacts, and comply with shifting regulations. We are already making an impact at the intersection of sustainability and commercial waste management, and we are hiring our third engineer. This new addition to our team will help us ship features directly to users and join the foundational early days of our technology organization. About the role:

Builder role who will be able to solve a wide range of technical challenges directly tied to the user and join a team with low technical debt.

Top team including two co-founders, Meredith Danberg-Ficarelli & Laura Rosenshine, who have combined over two decades of experience in the commercial waste industry

Full-Stack opportunity with preferred experience for Javascript & Node.js (targeting at least four years of experience)

Open to remote work in USA or Canada, NYC preferred

Solving a critical mission for our environment and society

Please check out the JD for more information [here](#). Send your resume to hello@getwats.com (edited)

Database.org

Full Stack Developer | WATS

WATS is a climate tech company working to decarbonize commercial real estate with

Central Hub: <https://workonclimate.org/>

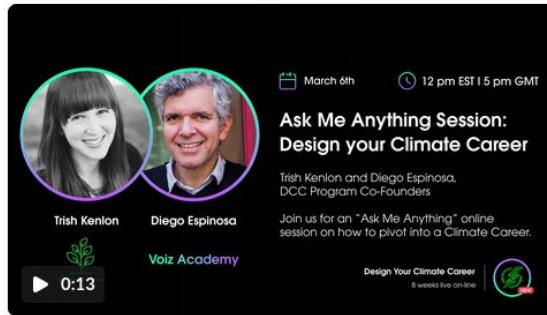
Working on climate

Communities help with resumes, interviews, skill learning etc.

- ⌚ community-organisers
- # events-external
- # events-official
- # general
- # gigs
- # i-got-a-job
- # jobs
- # jobs-alerts
- # jobs-discussion
- # learn-psychology
- # learn-the-space
- # meet-nyc
- # need-help
- # resume
- # role-academics
- # role-content-creators
- # share-your-path
- # topic-ai
- # volunteer-with-nonprofits

#ClimateCareers #Sustainability #CareerDevelopment

Climate Careers Events by Voiz Academy.mov ▾



Mike Heavers 7:09 PM

👉 - anyone have any recommendations for good training courses for sustainability education in the tech space? (e.g. green software principles)



3 replies Last reply 2 days ago

Yesterday ▾

Working on climate

You can also look for volunteer or internship opportunities to skill up

announcements

community-feedback

community-organisers

events-external

events-official

general

glgs

i-got-a-job

jobs

jobs-alerts

jobs-discussion

learn-psychology

learn-the-space

meet-nyc

need-help

project-wocl-deck

resume

role-academics

role-content-creators

share-your-path

topic-ai

volunteering

introductions

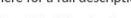
+ Add channels

1

 **Anna K** 10:49 AM
Has anyone here volunteered for Climate Neutral? I'd love to hear your thoughts on it!

 3 replies Last reply 1 month ago

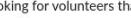
 **Geoff Skiggen** 1:58 PM
Calling all ocean activists. COARE (Center for Oceanic Awareness Research and Education), a global leader in conservation policy is looking for interns and volunteers.
I've been a volunteer for 1 year+ now and have loved every minute. I can say this org may be tiny, but it is mighty. They've achieved more than most marine conservation orgs triple their size.
COARE has written and passed substantial anti-shark finning policy at the United Nations and are also at the forefront of plastic pollution policy.
We're looking for Ocean Policy Interns, Ocean Program Interns, and Social Media Intern(s).
Click here for a full description: <https://www.coare.org/jobs/intern/>
Questions? Feel free to ping me on Slack or here: geoff@skigencreative.com (edited)



Wednesday, March 8th ▾

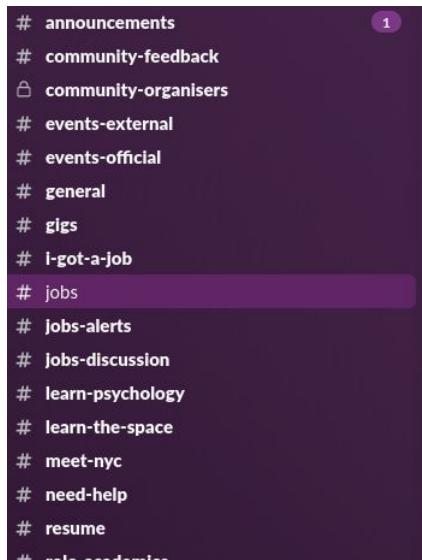
 **Michael Hawk** 4:54 PM
Hi Folks! I recently left Google to start a next-generation conservation organization (Jumpstart Nature) focused on biodiversity and climate. We're using technology and a multi-disciplinary approach to empower everyone to make a difference for the environment. In particular, we're looking at ways to leverage "new media" (apps, podcasts, games, etc) and behavioral science to get people off the sidelines. We're doing things that the large NGOs are ignoring or unable to do, and we need your help to keep pushing the boundaries of conservation.

I'm looking for volunteers that can help with podcast production (I can train you), campaign coordination across platforms, and marketing help in general.
We're in the early stage (filing for 501c3), so I'm also looking for a couple more board members - especially those with 501c3 experience and mobile app development experience.
We have a big vision and some great volunteers already. I'm happy to tell you more! Thanks!


 8 replies Last reply 13 days ago

Working on climate

Also job offers



A screenshot of a social media feed. At the top, it says "Powered by Teamleader". The feed shows a message from Sabyasachi Kar at 11:25 AM: "I have sent you a direct message. Please respond." Below this, a message from Andrea Dennis at 11:26 AM: "Hi Community! The science and technology team of Earthshot Labs is hiring a Senior Software Engineer. View the role description and apply here: <https://bit.ly/3SrjWu>". She continues: "We are looking for someone to take ownership of the front end of LandOS, our web mapping and carbon development project information system. Key skills to be successful in the role include React, Python, MapBox, and intuitions about geospatial data. Having solid back-end skills is a plus. Earthshot provides tech-enabled carbon development support and cutting-edge scientific predictive analytics to enhance every stage of a carbon project. We work with global clients by leveraging technology to bring science-backed rigor to carbon markets to accelerate the development and financing of nature-based solutions. Our vision at Earthshot is to scale reforestation and conservation globally to a level that meets the ecological and climate crises challenge. The role is 100% remote." She adds: "Thanks in advance for sharing this opportunity with your networks. If you have any questions, please connect with me directly at andrea@earthshot.eco. [Apply here](#)". Below the message, there are engagement icons: 2 hearts, 1 like, 2 video views, and 1 comment. It also shows "2 replies" and "Last reply 13 hours ago".

Working on climate

Job boards:

<https://climatebase.org/>

<https://terra.do/climate-jobs/job-board/>

<https://www.climatejobslist.com>

<https://www.climatepeople.com/>

Freelance: <https://www.leafr.work/>

and more!

<https://www.climatetechcareers.com/#jobboards>

<https://climateaction.tech/> slack

The screenshot shows a job search interface with the following details:

Find a job: Data Science

Location: New York, NY, USA

Refine Search: Remote

Set alerts

Job Listings:

- Senior Data Scientist** at Perennial (Remote - Boulder, CO, US) posted 2 months ago. \$130k-\$180k/Year. Tags: Food, Agriculture, & Land Use, Carbon Removal Tech.
- Data Science** at Thalo Labs (Brooklyn, NY, US) posted about 1 month ago. \$130k-\$180k/Year. Tags: Carbon Removal Tech.
- Machine Learning Engineer** at Pendulum Systems (Remote) posted about 2 months ago. Tags: Carbon Removal Tech.
- Senior Data Scientist** at KoBold Metals (Remote) posted 4 months ago. Tags: Energy, Climate Adaptation.
- Open Call (Remote)** at Yard Stick PBC (Remote - Oakland, CA, US) posted 15 days ago.
- Data Science Manager** at Rewiring America (Remote - United States) posted 2 months ago.

Learning more and networking

Climate Change AI summer school -

https://www.climatechange.ai/events/summer_school

Terra.do School - <https://terra.do/>

One Point Five Academy - <https://www.opf.degree/opf-academy>

Airminers Boot-up - <https://airminers.org/learn>

Climatematch Academy - <https://academy.climatematch.io/>

Starting a company

My Climate Journey community and podcast: <https://www.mcjcollective.com/>

Subak fellowship and accelerator: <https://subak.org/get-involved>

Resources for all stages: <https://climatefounder.org/>

“Every job is a climate job”



**CLIMATE SOLUTIONS
AT WORK**

Unleashing your employee power ►

CLIMATE
SOLUTIONS
AT WORK

INTRODUCTION

Before you begin

Employees hold tremendous power—you and your colleagues are instrumental to how your company functions, innovates, and survives through uncertainty.

For many employees, the importance of climate action has taken center stage, but it can be hard to know where (and at which levels) to accelerate climate action in the workplace, and how to bring your own skills to the table. If your company leadership is serious about its climate ambition, then they will welcome employees to the work of helping them get there and holding them accountable. We see the

● – Before you begin
– Getting to work
– Moving beyond net zero

Drawdown-aligned Framework

Emissions Reductions

Stakeholder Engagement and Collaboration

WHO THIS GUIDE IS FOR

Employees concerned about climate change, from those just starting their climate journeys to those ready to take concrete action in their workplace. This guide will help you understand whether your company is serious about addressing the climate crisis at scale—and how you can work with your colleagues to go further on climate.

WHAT THIS GUIDE IS

An illustrative suite of

<https://drawdown.org/publications/climate-solutions-at-work>

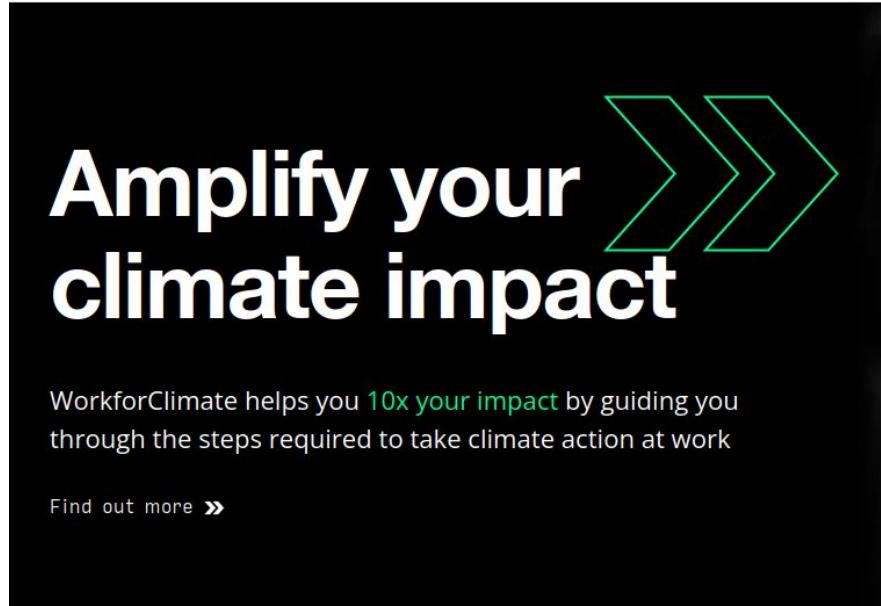
“Every job is a climate job”

	STATUS QUO	DRAWDOWN-ALIGNED
Emissions Reductions	Promises distant “net zero by 2050” commitment with reliance on offsets and no interim targets.	→ Moves on an accelerated timeline with interim targets and reduced reliance on offsets.
Stakeholder Engagement and Collaboration	Internal climate work is limited to sustainability teams and external engagement is primarily focused on investors.	→ Employees, communities, board members, and more are included.
Products, Partnerships, and Procurement	Products, partnerships, and procurement (the three Ps) are misaligned with climate commitments.	→ The three Ps are aligned with a company’s sustainability strategy.
Investments and Financing	Corporate investments and employee retirement plans are misaligned with climate goals.	→ Business is fully divested from fossil fuels and invested in climate solutions.
Climate Disclosures	A lack of transparency on emissions and climate-related risks is the norm.	→ Public emissions and climate risk disclosures occur on a regular basis.
Climate Policy Advocacy	Corporate and trade associations are known to lobby against climate policy.	→ Lobbying and contributions support bold climate policy, including alignment from trade associations.
Business Model Transformation	Sustainability is an under-resourced add-on to the core business of the company.	→ Business models always shift center on scaling climate solutions.
Long-term Thinking	Business is focused on growth and quarterly returns.	→ Uses business influence to promote justice and shift economic paradigms to fit within planetary boundaries.

Where we are and where we need to go

Private-sector climate targets need to meet the global climate crisis at scale. Companies can become “drawdown-aligned” by meeting all of the requirements outlined in this guide.

“Every job is a climate job”

The banner features the WorkforClimate logo at the top left. Below it, the main headline "Amplify your climate impact" is displayed in large white font, accompanied by three green chevron arrows pointing right. A subtext below the headline reads: "WorkforClimate helps you 10x your impact by guiding you through the steps required to take climate action at work". At the bottom left, there is a "Find out more" button.

Work
forClimate

News Resources Courses About

Amplify your climate impact

WorkforClimate helps you **10x your impact** by guiding you through the steps required to take climate action at work

Find out more ➞

• CLIMATE CHANGEMAKERS

Case study: How one employee got a climate-friendly option added to her company's 401(k)

When Camille Smith realised that her company's default 401(k) portfolio (essentially, the American version of a superannuation fund) was lacking in environmentally conscious fund options, she decided to do something about it. Here's how her persistence paid off in making climate-friendly investing available to everyone at her company.

• GETTING STARTED

4 pressing climate questions to ask your CEO right now

Ask questions. Get your boss to take climate action seriously. Simple.

“Every job is a climate job”

Neuron

NeuroView

Profession-specific resources are becoming more common

How Can Neuroscientists Respond to the Climate Emergency?

Adam R. Aron,^{1,*} Richard B. Ivry,² Kate J. Jeffery,³ Russell A. Poldrack,⁴ Robert Schmidt,⁵ Christopher Summerfield,^{6,*} and Anne E. Urai⁷

¹Department of Psychology and Neuroscience Graduate Program, University of California San Diego, San Diego, USA

²Department of Psychology, University of California, Berkeley, USA

³Division of Psychology and Language Sciences, University College London, UK

⁴Department of Psychology, Stanford University, Stanford, CA, USA

⁵Department of Psychology, University of Sheffield, Sheffield, UK

⁶Department of Experimental Psychology, University of Oxford, Oxford, UK

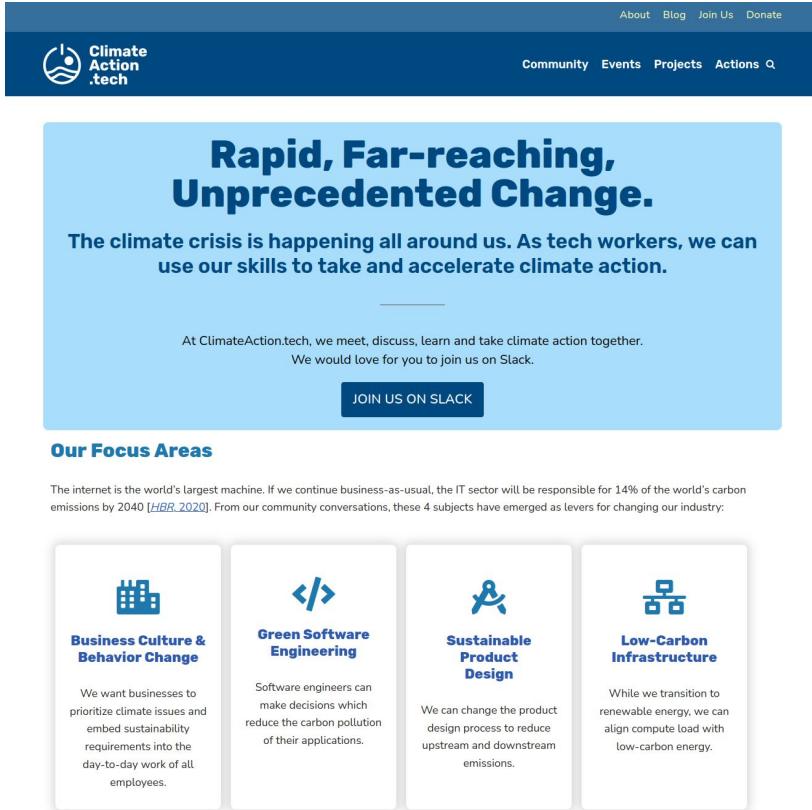
⁷Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

*Correspondence: adamaron@ucsd.edu (A.R.A.), christopher.summerfield@psy.ox.ac.uk (C.S.)

<https://doi.org/10.1016/j.neuron.2020.02.019>

The world faces a climate emergency. Here, we consider the actions that can be taken by neuroscientists to tackle climate change. We encourage neuroscientists to put emissions reductions at the center of their everyday professional activities.

“Every job is a climate job”



The screenshot shows the homepage of ClimateAction.tech. At the top, there's a dark blue header bar with the website's logo on the left and navigation links for About, Blog, Join Us, and Donate on the right. Below the header is a light blue main content area. In the center of this area, there's a large, bold, dark blue title: "Rapid, Far-reaching, Unprecedented Change." Below the title, a subtext reads: "The climate crisis is happening all around us. As tech workers, we can use our skills to take and accelerate climate action." Further down, there's a smaller text block: "At ClimateAction.tech, we meet, discuss, learn and take climate action together. We would love for you to join us on Slack." A dark blue button labeled "JOIN US ON SLACK" is positioned below this text. At the bottom of the main content area, there's a section titled "Our Focus Areas" with four white cards. Each card has a blue icon and a title: "Business Culture & Behavior Change" (building icon), "Green Software Engineering" (code icon), "Sustainable Product Design" (person icon), and "Low-Carbon Infrastructure" (server icon). Each card also contains a brief description of its focus area.

Rapid, Far-reaching, Unprecedented Change.

The climate crisis is happening all around us. As tech workers, we can use our skills to take and accelerate climate action.

At ClimateAction.tech, we meet, discuss, learn and take climate action together. We would love for you to join us on Slack.

JOIN US ON SLACK

Our Focus Areas

Business Culture & Behavior Change

We want businesses to prioritize climate issues and embed sustainability requirements into the day-to-day work of all employees.

Green Software Engineering

Software engineers can make decisions which reduce the carbon pollution of their applications.

Sustainable Product Design

We can change the product design process to reduce upstream and downstream emissions.

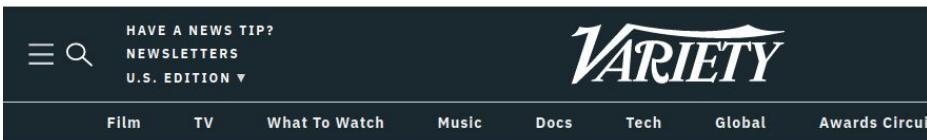
Low-Carbon Infrastructure

While we transition to renewable energy, we can align compute load with low-carbon energy.

Profession-specific resources are becoming more common

<https://climateaction.tech/>

“Every job is a climate job”



A screenshot of the Variety website. At the top left is a search bar with the placeholder "HAVE A NEWS TIP?". Below it are links for "NEWSLETTERS" and "U.S. EDITION ▾". The Variety logo is prominently displayed in the center. Below the logo is a navigation bar with categories: Film, TV, What To Watch, Music, Docs, Tech, Global, and Awards Circuit.

[HOME](#) > [MUSIC](#) > [NEWS](#)

Jul 12, 2022 7:55am PT

Here's How the Music Industry Can Help Fight Climate Change (Guest Column)

By Jem Aswad ▾



Tech Marketing Media Executive Transactions All Topics ▾

Daily Editions ▾ | Newsletters ▾ |

FASHION INDUSTRY CHARTER FOR CLIMATE ACTION

Climate Action Playbook



Sports industry needs to make climate change a priority

BY AYELET MAVOR

Monday, January 10, 2022

Profession-specific resources are becoming more common

Climate action outside of work

Non-profits working on climate issues

- Clean Air Task Force
- Earthjustice
- Conservation International
- CarbonFund
- Environmental Defense Fund
- many more!

Climate action outside of work

Political organizing

- Citizens' Climate Lobby
- League of Conservation Voters
- Extinction Rebellion
- People's Climate Movement

REPORT | MAY 26, 2021

Gen Z, Millennials Stand Out for Climate Change Activism, Social Media Engagement With Issue



ANNALS OF A WARMING PLANET

THE ANSWER TO CLIMATE CHANGE IS ORGANIZING

Dealing with global warming is always going to be about the balance of power.



By Bill McKibben
September 1, 2021

Gen Z, Millennials more active than older generations addressing climate change on- and offline

% of U.S. adults who say ...

Climate should be top priority to ensure sustainable planet for future generations

Gen

Z

67%

71

Gen

X

63

Boomer

& older

57

Addressing climate change is my top personal concern

37

33

27

29

Have personally taken action to help address climate change within the last year

32

28

23

21



Note: Respondents who gave other responses or did not give an answer are not shown. Seen content on social media and engaged with climate content based on social media users.
Source: Survey conducted April 20-29, 2021.

"Gen Z, Millennials Stand Out for Climate Change Activism, Social Media Engagement With Issue"

PEW RESEARCH CENTER

Climate action outside of work



FOSSIL FREE FESTIVAL DIVEST NYU ABOUT



TAKE ACTION

Sunrise NYU is a community of powerful young people fighting for a Green New Deal.

When people rise up they can be an unstoppable force for good. We are the official Sunrise Movement hub for New York University; a diverse group of dedicated students pushing for change.



Affordability Steering Committee

Academic Space Projects

2040 Now

2040 Now Events
Calendar

Apply Now

2040 Now Center

Resources & More

NYU's Commitment

NYU In Brooklyn

NYUWomxn100

Be a climate optimist!

NYU plans to be climate neutral by 2040. Join us April 17-21 for **2040 Now**—a University-wide initiative where we explore our community response to the challenge of climate change through community challenges, big events, exhibitions, competitions, and much more.

Through meaningful climate action, including both individual choices and societal involvement, our community can help create a climate-stable future, be healthier and happier, and have fun making change—together.

Sustainability is not a spectator sport. We hope to see you at one of the many events and programs below!



Projects Overview

The class project is your chance to go deeper into a question, dataset, and/or method you are interested in.

You will be evaluated on your ability to pose a good machine learning question in the domain of climate change, the selection and execution of relevant methods, your ability to work on a team, and your presentation skills both written and oral.

Project teams

Everyone must work in a team (no individual projects)

Teams should be 4 people (though there will need to be one team with 5, or a couple with 3)

A single grade will be given to the team.

Projects

Projects need to go beyond what you have done in homework assignments.

Possible approaches:

- replicate/extend an existing paper
- contribute to an established competition/benchmark
- Identify an underexplored problem where ML could help and try to solve it

Projects

Rough guidelines for creating a substantial project:

- Compare multiple methods and/or hyperparameters
- Use at least one method that isn't in scikit-learn (Kartik will be going through PyTorch code in labs)
- Explore feature importance methods
- Create your own dataset from multiple sources
- Use multiple evaluation methods
- Include advanced visualizations and/or a dashboard (but it better be really good)
- etc

Possible data sources

Here are locations to look for relevant data. Not all datasets from these sources will be climate change-related.

Note: these are real datasets. They may have missing data, poor documentation, or other challenges. You will probably have to teach yourselves some new skills to get what you want from them. These are all things data scientists have to face! Look into a dataset before committing to using it.

Starting Points

Climate Change AI

<https://www.climatechange.ai/papers>

Climate Change AI Workshop Papers

On this page, we show accepted works from all our [workshops](#) on "Tackling Climate Change with Machine Learning."

Venue	Text Search	Subject Areas	Award?
All Venues <input type="button" value="▼"/>	<input type="text" value="Enter search terms..."/>	<input type="text" value="Select some options..."/>	<input type="checkbox"/>

Showing 1 to 728 of 728 entries

Venue	Title	Subject Areas
NeurIPS 2023	Machine learning for gap-filling in greenhouse gas emissions databases (Papers Track) ►Abstract and authors: (click to expand)	Public Policy
NeurIPS 2023	EarthPT: a foundation model for Earth Observation (Papers Track) ►Abstract and authors: (click to expand)	Earth Observation & Monitoring
NeurIPS 2023	Towards Understanding Climate Change Perceptions: A Social Media Dataset (Papers Track) ►Abstract and authors: (click to expand)	Behavioral and Social Science

Starting Points

Climate Change AI

https://www.climatechange.ai/subject_areas

Subject Areas

Here, we provide a window into our content by subject area. Click on an area to see related papers, blog posts, webinars, and tutorials.

Applications

- Agriculture & Food
- Behavioral and Social Science
- Buildings
- Carbon Capture & Sequestration
- Chemistry & Materials
- Cities & Urban Planning
- Climate Finance & Economics
- Climate Justice
- Climate Policy
- Climate Science & Modeling
- Disaster Management and Relief
- Earth Observation & Monitoring
- Ecosystems & Biodiversity

AI Methods

- Active Learning
- Causal & Bayesian Methods
- Computer Vision & Remote Sensing
- Control Systems
- Data Mining
- Forecasting
- Generative Modeling
- Hybrid Physical Models
- Interpretable ML
- Meta- and Transfer Learning
- Natural Language Processing
- Optimization
- Recommender Systems

Starting Points

Climate Change AI

<https://www.climatechange.ai/tutorials>

Tutorials

Text Search	Topic	Difficulty	Subject Areas
<input type="text" value="Enter search terms..."/>	<input type="button" value="All Topics"/> 	<input type="button" value="All Levels"/> 	<input type="button" value="Select some options..."/>



Agile Modeling for Bioacoustic Monitoring



Jenny Hamer, Rob Laber, and Tom Denton, NeurIPS 2023

Bioacoustic monitoring promises to help unlock the ability to monitor biodiversity, ecosystem health, and endangered species cost effectively. This tutorial presents an "agile modeling" approach that enables users to build custom classifier systems efficiently for species of interest using transfer learning, audio search, and human-in-the-loop active learning.

Ecosystems and Biodiversity Intermediate Python Bioacoustics Audio/Acoustics

Search Human-in-the-Loop Ecosystems & Biodiversity Active Learning 



Aquaculture Mapping: Detecting and Classifying Aquaculture Ponds using Deep Learning



Joshua Cortez and John Christian Napil, NeurIPS 2023

Managing aquaculture ponds is vital for environmental monitoring and conservation. This tutorial presents how to leverage satellite imagery and semantic segmentation models to detect and map aquaculture ponds based

Places to find data

Kaggle - can look at competitions and data sets (but pay attention to quality!)

www.kaggle.com

Twitter Climate Change Sentiment Dataset

44k tweets pertaining to climate change



[Data Card](#) [Code \(8\)](#) [Discussion \(0\)](#)

About Dataset

The collection of this data was funded by a Canada Foundation for Innovation JELF Grant to Chris Bauch, University of Waterloo.

This dataset aggregates tweets pertaining to climate change collected between Apr 27, 2015 and Feb 21, 2018. In total, 43943 tweets were annotated. Each tweet is labelled independently by 3 reviewers. This dataset only contains tweets that all 3 reviewers agreed on (the rest were discarded).

Each tweet is labelled as one of the following classes:

- 2(News): the tweet links to factual news about climate change
- 1(Pro): the tweet supports the belief of man-made climate change
- 0(Neutral): the tweet neither supports nor refutes the belief of man-made climate change
- -1(Anti): the tweet does not believe in man-made climate change

The distribution of the data:

Waste Classification data

This dataset contains 22500 images of organic and recyclable objects

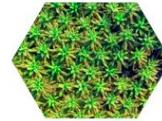
[Data Card](#) [Code \(71\)](#) [Discussion \(0\)](#)

<https://www.kaggle.com/datasets/techsash/waste-classification-data>

Places to find data

Zindi (can use closed or open competitions)

<https://zindi.africa/competitions>



Digital Africa Plantation Counting Challenge

Can you create a semi-supervised algorithm to count trees in plantations in Côte d'Ivoire?

Advanced | Prediction | Computer Vision | Agriculture

\$10 000 USD

~1 month to go
76 active participants
268 enrolled

Helping Côte d'Ivoire



Africa Biomass Challenge

Can you predict biomass in cocoa plantations in Côte d'Ivoire?

Advanced | Prediction | Agriculture

\$10 000 USD

3 months to go
106 active participants
666 enrolled

Helping Côte d'Ivoire



Intron AfriSpeech-200 Automatic Speech Recognition Challenge

Can you create an automatic speech recognition (ASR) model for African accents, for use by doctors?

Advanced | Automatic Speech Recognition | Media | Health

\$5 000 USD

3 months to go
5 active participants
148 enrolled

Places to find data

Driven Data competitions

<https://www.drivendata.org/competitions/>



Challenge Summary

Can you predict local epidemics of dengue fever?

Dengue fever is a mosquito-borne disease that occurs in tropical and sub-tropical parts of the world. In mild cases, symptoms are similar to the flu: fever, rash, and muscle and joint pain. In severe cases, dengue fever can cause severe bleeding, low blood pressure, and even death.



LEADERBOARD

DATA DOWNLOAD

SUBMISSIONS

Possible data sources

<https://opendata.cityofnewyork.us/>

Has an “environmental” section.
Contains data on trees, air quality,
energy use, sea level, etc.

The screenshot shows the NYC OpenData homepage. The top navigation bar includes links for Home, Data, About, Learn, Alerts, Contact Us, Blog, and Sign In. A search bar at the top right contains the word "climate". On the left, there are two dropdown menus: "Categories" (listing Business, City Government, Education, Environment, Health) and "View Types" (listing Data Lens pages, Datasets, External Datasets, and Files and Documents). The main content area displays a search result for "climate" with 60 results. The first result is "Sea Level Rise Maps (2050s 500-year Floodplain)" under the Environment category. It has a "Map" link, was updated on September 16, 2021, and has 33,946 views. The second result is "Sea Level Rise Maps (2020s 100-year Floodplain)" also under Environment, with a "Map" link, updated on September 16, 2021, and 21,768 views. The third result is "Natural Gas Consumption by ZIP Code - 2010" under Environment, with a "Dataset" link, updated on May 9, 2022, and 1 API Docs. Each dataset card includes a brief description, tags, and a "More" link.

Possible data sources

Energy information administration

<https://www.eia.gov/totalenergy/data/annual/index.php>

[SEE ALL REPORTS](#)

Annual Energy Review

Superseded -- see MER for key annual tables

Annual Ei
Year

EIA has expanded the *Monthly Energy Review* (MER) to include annual data as far back as 1949 for those data tables that are found in both the *Annual Energy Review* (AER) and the [MER](#). In the list of tables below, grayed-out table numbers now go to MER tables that contain data series for 1949 forward. New [INTERACTIVE tables and graphs](#) have also been added.

DATA CATEGORIES

[+ EXPAND ALL](#)

- + Energy overview
- + Energy consumption by sector
- + Financial indicators
- + Energy resources
- + Petroleum and other liquids
- + Natural gas
- + Coal
- + Electricity
- + Nuclear energy
- + Renewable energy
- + Environment
- + Appendices (heat content, conversion factors, and more)

Possible data sources

Emissions estimations from NEI

<https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei>

ClimateTRACE emissions data

<https://climatetrace.org/downloads>

Carbon dioxide data from remote sensing

https://disc.gsfc.nasa.gov/datasets/OCO2GriddedXC_O2_3/summary?keywords=oco-2

On road emissions data

https://daac.ornl.gov/CMS/guides/CMS_DARTE_V2.html

Air Emissions Inventories

CONTACT US

Air Emissions Inventory
Home

Data and Documentation

2020 NEI Data

2017 NEI Data

2014 NEI Data

2011 NEI Data

2008 NEI Data

National Emissions Inventory (NEI)

The National Emissions Inventory (NEI) is a comprehensive and detailed estimate of air emissions of criteria pollutants, criteria precursors, and hazardous air pollutants from air emissions sources. The NEI is released every three years based primarily upon data provided by State, Local, and Tribal air agencies for sources in their jurisdictions and supplemented by data developed by the US EPA. The NEI is built using the [Emissions Inventory System](#) (EIS) first to collect the data from State, Local, and Tribal air agencies and then to blend that data with other data sources.

DOWNLOADS

Climate TRACE emissions data is free and publicly available for download below. Each download package includes annual country-level emissions by sector and by greenhouse gas from 2015-2021, the applicable inventory of facility-level emissions, and facility-level ownership data where available.

VIEW DOWNLOADS BY SECTOR COUNTRY		
SECTOR	DOWNLOAD	METHODOLOGY
 Agriculture	CSV	Download (18 MB)
 Buildings	CSV	Download (0.8 MB)
 Fluorinated gases	CSV	Download (0.8 MB)
 Fossil fuel operations	CSV	Download (4 MB)
 Forestry and Land Use	CSV	Download (0.2 MB)
 Manufacturing	CSV	Download (15 MB)
 Mineral Extraction	CSV	Download (2 MB)
 Power	CSV	Download (0.7 MB)
 Transportation	CSV	Download (12 MB)
 Waste	CSV	Download (7 MB)

Possible data sources

Open Street Map for general land and road features

<https://www.openstreetmap.org>

https://wiki.openstreetmap.org/wiki/Environmental_OSM

Environmental OSM

[English](#) [español](#) [русский](#) [日本語](#)

Environmental OSM - Other languages

[Other languages](#) • [Translate](#)

Environmental OSM is a project to map out various systems that have an impact on the environment in hopes of helping to solve environmental problems.

Please use the [discussion page](#) to suggest and discuss other data we might survey and other uses for the data.

Contents [hide]

- 1 Objectives
- 2 Tags
 - 2.1 Energy
 - 2.2 Environmental monitoring
 - 2.3 Waste/recycling
 - 2.4 Food
 - 2.5 Transport
 - 2.6 Natural habitats / land use
 - 2.7 Green space
- 3 Environmental features in development
- 4 Related initiatives
 - 4.1 OpenStreetMap
 - 4.2 Other
- 5 People interested
- 6 Contact

Possible data sources

Bureau of Transportation
Statistics

https://www.bts.gov/product/national-transportation-statistics?keys=miles&field_topic_target_id>All



[Amtrak Fuel Consumption and Travel](#)



[Amtrak Fuel Consumption and Travel \(Metric\)](#)



[Amtrak On-Time Performance Trends and Hours of Delay by Cause](#)



[Annual Roadway Congestion Index](#)



[Automobile Profile](#)



[Average Cost of Owning and Operating an Automobile](#)



[Average Fuel Efficiency of U.S. Light Duty Vehicles](#)



[Average Fuel Efficiency of U.S. Light Duty Vehicles \(metric\)](#)



[Average Length of Haul, Domestic Freight and Passenger Modes](#)



[Average Length of Haul, Domestic Freight and Passenger Modes \(metric\)](#)



[Average Passenger Revenue per Passenger-Mile](#)



[Bus Fuel Consumption and Travel \(metric\)](#)

Possible data sources

UN agriculture data (crop and livestock, yields, land area, etc)

<https://www.fao.org/faostat/en/#data>

Data on what crops are where:

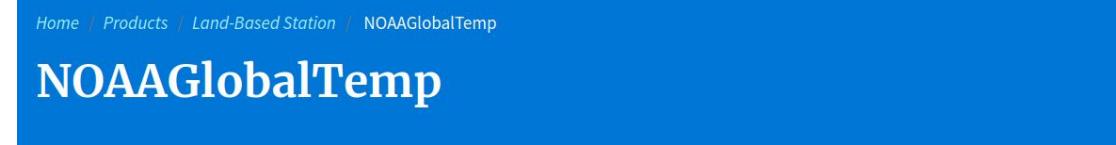
https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php#Section1_1.0

The screenshot shows the homepage of the FAOSTAT website. At the top, there is a navigation bar with links to About FAO, In Action, Countries, Themes, Media, Publications, Statistics, and Partnerships. Below this, there are language links for English, Français, Русский, and Español. The main content area is titled "FAOSTAT" and features a "Data" tab selected. Other tabs include Selected Indicators, Compare Data, Definitions and Standards, and FAQ. A search bar at the top right allows users to "Search an Indicator or Commodity". The main content area is divided into two columns under the heading "Data". The left column contains links to "Production", "Food Security and Nutrition", "Food Balances", "Trade", "Prices", "Land, Inputs and Sustainability", "Cost and Affordability of a Healthy Diet", and "Population and Employment". The right column contains links to "Investment", "Macro-Economic Indicators", "Food Value Chain", "Climate Change", "Forestry", "SDG Indicators", "World Census of Agriculture", and "Discontinued archives and data series". Each link is accompanied by a small icon.

Possible data sources

Global temperatures

<https://www.ncei.noaa.gov/products/land-based-station/noaa-global-temp>



The NOAA Merged Land Ocean Global Surface Temperature Analysis (NOAAGlobalTemp, formerly known as [MLOST](#)) combines long-term sea surface (water) temperature (SST) and land surface (air) temperature datasets to create a [complete, accurate depiction of global temperature trends](#). The dataset is used to support climate monitoring activities such as the Monthly Global Climate Assessment, and also provides input data for a number of climate models.

Possible data sources

Satellite datasets

<https://deepcube-h2020.eu/technology/earth-system-data-cube/>

<https://deepcube-h2020.eu/datasets/>

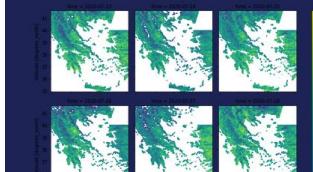
<https://github.com/satellite-image-deep-learning/datasets>

Data cube for drought forecasting in Africa

This is a dataset of minicubes suitable for Earth Surface Forecasting, created by the DeepCube research team of the Max Planck Institute for Biogeochemistry. The dataset is focused on Africa and the task is centered around drought impact forecasting.

To learn more about the Africa minicubes please visit <https://www.earthnet.tech/docs/ds-africa/>.

Data cube for the wildfire research community



This dataset has been created by the DeepCube research team of the National Observatory of Athens and is meant to be used to develop models for next day fire hazard forecasting in Greece.

The dataset includes dynamic variables, such as previous day Leaf Area Index, evapotranspiration, Land Surface Temperature, meteorological data, fire variables and Fire Weather Index, resampled at daily temporal resolution and 1km spatial resolution. It also includes static variables, such as roads density, population density and topography layers.

Possible data sources

Extreme weather events

<https://github.com/andregraubner/ClimateNet>

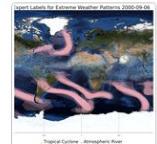
<https://gmd.copernicus.org/articles/14/107/2021/>

Model description paper |  

ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather

Prabhat ★, Karthik Kashinath ✉ ★, Mayur Mudigonda ★, Sol Kim, Lukas Kapp-Schwoerer, Andre Graubner, Ege Karaismailoglu, Leo von Kleist, Thorsten Kurth, Annette Greiner, Ankur Mahesh, Kevin Yang, Colby Lewis, Jiayi Chen, Andrew Lou, Sathyavat Chandran, Ben Toms, Will Chapman, Katherine Dagon, Christine A. Shields, Travis O'Brien, Michael Wehner, and William Collins

08 Jan 2021



Abstract

Identifying, detecting, and localizing extreme weather events is a crucial first step in understanding how they may vary under different climate change scenarios. Pattern recognition tasks such as classification, object detection, and segmentation (i.e., pixel-level classification) have remained challenging problems in the weather and climate sciences. While there exist many empirical heuristics for detecting extreme events, the disparities between the output of these different methods even for a single event are large and often difficult to reconcile. Given the success of deep learning (DL) in tackling similar problems in computer vision, we advocate a DL-based approach. DL, however, works best in the context of supervised learning – when labeled datasets are readily available. Reliable labeled training data for extreme weather and climate events is scarce.

We create “ClimateNet” – an open, community-sourced human-expert-labeled curated dataset that captures tropical cyclones (TCs) and atmospheric rivers (ARs) in high-resolution climate model output from a simulation of a recent historical period. We use the curated ClimateNet dataset to train a state-of-the-art DL model for pixel-level identification – i.e., segmentation – of TCs and ARs. We then apply the trained DL model to historical and climate change scenarios simulated by the Community Atmospheric Model (CAM5.1) and show that the DL model accurately segments the data into TCs, ARs, or “the background” at a pixel level. Further, we show how the segmentation results can be used to conduct spatially and temporally precise analytics by quantifying distributions of extreme precipitation conditioned on event types (TC or AR) at regional scales. The key contribution of this work is that it paves the way for DL-based automated, high-fidelity, and highly precise analytics of climate data using a curated expert-labeled dataset – ClimateNet.

More paper examples

<https://ai4good.org/fragile-earth-2022/>

<https://www.cambridge.org/core/journals/environmental-data-science>

<https://www.hadr.ai/accepted-papers>

https://openaccess.thecvf.com/CVPR2023_workshops/EarthVision

Possible data sources

Want a certain type of data? Check if there is a government agency about it

Or google (particularly google scholar)

Or ask me

Possible methods

Supervised classification or regression (classic methods, simple artificial neural networks, advanced artificial neural networks)

Computer vision (simple feature extraction, convolutional neural networks)

Self- or Unsupervised methods for representation learning (pca, clustering, artificial neural networks)

Feature importance

Reinforcement learning

Project report

Abstract - a one paragraph summary of the full report

Background - what problem are you addressing, how does it relate to climate change, how has it been approached in the past, how are you approaching it and why

Methods - what data are using, how did you process it, what methods are using, how did you decide hyperparameters, how are you evaluating your model

Results - what did your methods show, what worked and what didn't work, includes pointers to figures

Discussion - what do your results mean, what would you do in the future if you were to continue working on this project, how do your results relate to other studies

References - bibliography of sources you used and cited

~3 pages min for first five sections, plus extra pages for references and figures. Plus pre-run python notebook (should replicate your figures from data).

Project Presentations

~10 minutes (exact time will depend on the number of groups). Summary of your project report. Should include all the same sections.

All team members must speak in the project presentation.

There will be time for questions from other students (required) and me

Advice

Think big (could even result in a publication). This is a growth opportunity.

Choose something meaningful and interesting to you.

Expect things to take awhile and be frustrating, especially as you get started with new data or methods

Document what you are doing for yourself and for your final report

Manage your time! Work steadily, don't wait until the end

Research Quality

How to avoid machine learning pitfalls: a guide for academic researchers

Michael A. Lones*

Abstract

This document is a concise outline of some of the common mistakes that occur when using machine learning, and what can be done to avoid them. Whilst it should be accessible to anyone with a basic understanding of machine learning techniques, it was originally written for research students, and focuses on issues that are of particular concern within academic research, such as the need to do rigorous comparisons and reach valid conclusions. It covers five stages of the machine learning process: what to do before model building, how to reliably build models, how to robustly evaluate models, how to compare models fairly, and how to report results.

<https://arxiv.org/pdf/2108.02497.pdf>

Contents

1	Introduction	1
2	Before you start to build models	3
2.1	Do take the time to understand your data	3
2.2	Don't look at <i>all</i> your data	3
2.3	Do make sure you have enough data	3
2.4	Do talk to domain experts	4
2.5	Do survey the literature	4
2.6	Do think about how your model will be deployed	5
3	How to reliably build models	5
3.1	Don't allow test data to leak into the training process	5
3.2	Do try out a range of different models	6
3.3	Don't use inappropriate models	7
3.4	Do keep up with recent developments in deep learning	8
3.5	Don't assume deep learning will be the best approach	8
3.6	Do optimise your model's hyperparameters	9
3.7	Do be careful where you optimise hyperparameters and select features	9
3.8	Do avoid learning spurious correlations	11
4	How to robustly evaluate models	11
4.1	Do use an appropriate test set	11
4.2	Don't do data augmentation <i>before</i> splitting your data	12
4.3	Do use a validation set	12
4.4	Do evaluate a model multiple times	12
4.5	Do save some data to evaluate your final model instance	14
4.6	Don't use accuracy with imbalanced data sets	14
4.7	Don't ignore temporal dependencies in time series data	15
5	How to compare models fairly	16
5.1	Don't assume a bigger number means a better model	16
5.2	Do use statistical tests when comparing models	16
5.3	Do correct for multiple comparisons	17
5.4	Don't always believe results from community benchmarks	17
5.5	Do consider combinations of models	17
6	How to report your results	18
6.1	Do be transparent	18
6.2	Do report performance in multiple ways	19
6.3	Don't generalise beyond the data	19
6.4	Do be careful when reporting statistical significance	19
6.5	Do look at your models	20
7	Final thoughts	20

Writing evaluation

Your writing should be clear and concise, with minimal grammatical errors.

To write your background, find some example papers that may have tackled a similar problem so that you can situate your work relevant to past work (probably should do this before you start!).

For the methods, be clear and complete enough that someone could replicate what you did just from reading your methods section.

Figures should be easy to understand, with clearly labeled axes and not too many different colors or line styles.

Presentation Evaluation

Slides should have clear and informative titles

Figures should be fully visible with labeled axes

Slides should not have too much written text; rather, you should have one or two main points on the slide with one or two major visuals. Elaborate on what is on the slides as you speak.

Speak loud enough for everyone to hear

Go at the right pace throughout (time yourself beforehand!)

Advice: imagine yourself in the audience listening to a project you know nothing about. How would you want the speaker to communicate to you?

Chat GPT

Chat GPT



You

What are three papers that use machine learning to help with building energy efficiency?



ChatGPT

Certainly! Here are three papers that utilize machine learning techniques to improve building energy efficiency:

1. **"Deep Learning for Predictive Modeling of Energy Use of Appliances in Smart Buildings"**

- Authors: Hongyu Chen, Xudong Wang, Chuan Qin, Wei Xu
- Published in: IEEE Transactions on Industrial Informatics, 2017
- Summary: This paper presents a deep learning-based approach for predictive modeling of energy use of appliances in smart buildings. The authors utilize Long Short-Term Memory (LSTM) networks to capture the temporal dependencies in energy consumption data, enabling accurate prediction and optimization of energy use within buildings.

2. **"A Review on the Applications of Artificial Intelligence Techniques in Building Energy Systems"**

- Authors: Yaghoob Jafari, Somayeh Asadi, Shuli Liu, Xuhui Guo
- Published in: Applied Energy, 2019
- Summary: This review paper provides an overview of various artificial intelligence techniques, including machine learning, applied in building energy systems. It discusses the use of machine learning algorithms for energy demand prediction, optimization of HVAC systems, occupancy detection, and other aspects of building energy management.

3. **"Predicting Building Energy Consumption: A Review of Data-Driven Methods"**

- Authors: Baomin Wang, Qian Wang, Yu Liang, Zhiwei Xu, Jizhen Liu, and Yulin Jia
- Published in: Energy and Buildings, 2018
- Summary: This paper presents a comprehensive review of data-driven methods for predicting building energy consumption. It covers various machine learning techniques such as support vector machines, decision trees, neural networks, and ensemble methods, highlighting their strengths and limitations in modeling ↓ predicting energy usage patterns in buildings.

Nope

Nope

Nope



About 103,000 results (0.46 sec)

Chat GPT



You

What are three papers that use machine learning to help with building energy efficiency?



ChatGPT

Certainly! Here are three papers that utilize machine learning techniques to improve building energy efficiency:

1. **"Deep Learning for Predictive Modeling of Energy Use of Appliances in Smart Buildings"**

- Authors: Hongyu Chen, Xudong Wang, Chuan Qin, Wei Xu
- Published in: IEEE Transactions on Industrial Informatics, 2017
- Summary: This paper presents a deep learning-based approach for predictive modeling of energy use of appliances in smart buildings. The authors utilize Long Short-Term (LSTM) networks to capture the temporal dependencies in energy consumption, enabling accurate prediction and optimization of energy use within buildings.

2. **"A Review on the Applications of Artificial Intelligence Techniques in Building Energy Management"**

- Authors: Yaghoub Jafari, Somayeh Asadi, Shuli Liu, Xuhui Guo
- Published in: Applied Energy, 2019
- Summary: This review paper provides an overview of various artificial intelligence techniques, including machine learning, applied in building energy systems. It discusses the use of machine learning algorithms for energy demand prediction, optimization of HVAC systems, occupancy detection, and other aspects of building energy management.

3. **"Predicting Building Energy Consumption: A Review of Data-Driven Methods"**

- Authors: Baomin Wang, Qian Wang, Yu Liang, Zhiwei Xu, Jizhen Liu, and Yulin Jia
- Published in: Energy and Buildings, 2018
- Summary: This paper presents a comprehensive review of data-driven methods for predicting building energy consumption. It covers various machine learning techniques such as support vector machines, decision trees, neural networks, and ensemble methods, highlighting their strengths and limitations in modeling predicting energy usage patterns in buildings.

[HTML] **A review of data-driven building energy consumption prediction studies**

K Amasyali, NM El-Gohary - Renewable and Sustainable Energy Reviews, 2018 - Elsevier
... consumption prediction. This paper offers a review of the studies that developed data-driven building energy consumption prediction models, with a particular focus on reviewing the ...
☆ Save 99 Cite Cited by 1499 Related articles All 4 versions Web of Science: 885 ⓘ

[HTML] **A review of data-driven approaches for prediction and classification of building energy consumption**

Y Wei, X Zhang, Y Shi, L Xia, S Pan, J Wu... - ... Energy Reviews, 2018 - Elsevier
... algorithms for broad application throughout the building industry. This article reviews the prevailing data-driven approaches used in building energy analysis under different archetypes ...
☆ Save 99 Cite Cited by 654 Related articles All 2 versions Web of Science: 401 ⓘ

[HTML] **Modeling and forecasting building energy consumption: A review of data-driven techniques**

M Bourdeau, X qiang Zhai, E Nefzaoui, X Guo... - Sustainable Cities and ..., 2019 - Elsevier
... and forecasting tools. Consequently, data-driven building energy consumption modeling ... Thus, we report in the present paper a review on data-driven building energy modeling ...
☆ Save 99 Cite Cited by 596 Related articles All 5 versions Web of Science: 380 ⓘ

Nope

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumian
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
amin.mazloumian@zhaw.ch

Matthias Rosenthal
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
matthias.rosenthal@zhaw.ch

Hans Gelke
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
hans.gelke@zhaw.ch

Abstract—One third of food produced in the world for human consumption – approximately 1.3 billion tons – is lost or wasted every year. By classifying food waste of individual consumers and raising awareness of the measures, avoidable food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. We specifically designed a deep neural network that classifies food waste for every time food waste is thrown in the waste bins. Our method presents how deep learning networks can be tailored to best learn from available training data.

Index Terms—deep learning, neural networks, machine learning, food waste

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.

In this research, we adopt a deep convolutional neural network approach for classifying food waste in waste bins [9]. Deep convolutional neural networks are supervised machine

My Research Paper

Grace Lindsay

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes. Here, I will test three different neural networks trained to classify food waste

Writing integrity

Deep Learning for Classifying Food Waste

Academic Disciplinary offenses include but are not limited to:

- cheating, plagiarism, falsification of data or sources, forgery of academic documents in attempt to defraud;
- destruction, theft, or unauthorized use of laboratory data, or research materials.

The following penalties may be imposed by the faculty for disciplinary infractions:

- Warning
- Disciplinary Probation
- Suspension
- Dismissal

Abstract—One third of food waste in the United States is discarded every year. By classifying food waste, we can reduce the amount of food waste significantly. In this paper, we propose a deep learning model to classify food waste in half a mill second. The model is trained on a dataset installed on top of food waste. The model uses a deep neural network that classifies food waste into three categories: fruits, vegetables, and meat. The model is able to classify food waste in a wide variety of forms, such as soups and salads, which can largely hide the true nature of the food waste.

Index Terms—deep learning, food waste

technically a difficult task
/ hard to differentiate
consider distinguishing
time in a wide variety
waste, e.g., soups and
salads, can largely hide
the true nature of the food waste
in three different ways

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumian
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
amin.mazloumian@zhaw.ch

Matthias Rosenthal
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
matthias.rosenthal@zhaw.ch

Hans Gelke
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
hans.gelke@zhaw.ch

Abstract—One third of food produced in the world for human consumption – approximately 1.3 billion tons – is lost or wasted every year. By classifying food waste of individual consumers and raising awareness of the measures, avoidable food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. We specifically designed a deep neural network that classifies food waste for every time food waste is thrown in the waste bins. Our method presents how deep learning networks can be tailored to best learn from available training data.

Index Terms—deep learning, neural networks, machine learning, food waste

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.

In this research, we adopt a deep convolutional neural network approach for classifying food waste in waste bins [9]. Deep convolutional neural networks are supervised machine

My Research Paper

Grace Lindsay

According to [Mazloumian et al. \(2020\)](#): “The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.” Here, I will test three different neural networks trained to classify food waste

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumian
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
amin.mazloumian@zhaw.ch

Matthias Rosenthal
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
matthias.rosenthal@zhaw.ch

Hans Gelke
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
hans.gelke@zhaw.ch

Abstract—One third of food produced in the world for human consumption – approximately 1.3 billion tons – is lost or wasted every year. By classifying food waste of individual consumers and raising awareness of the measures, avoidable food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. We specifically designed a deep neural network that classifies food waste for every time food waste is thrown in the waste bins. Our method presents how deep learning networks can be tailored to best learn from available training data.

Index Terms—deep learning, neural networks, machine learning, food waste

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.

In this research, we adopt a deep convolutional neural network approach for classifying food waste in waste bins [9]. Deep convolutional neural networks are supervised machine

My Research Paper

Grace Lindsay

According to [Mazloumian et al. \(2020\)](#): “The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and not-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.” Here, I will test three different neural networks trained to classify food waste

Not plagiarism, but bad writing!

Writing integrity

Deep Learning for Classifying Food Waste

Amin Mazloumian
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
amin.mazloumian@zhaw.ch

Matthias Rosenthal
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
matthias.rosenthal@zhaw.ch

Hans Gelke
Institute of Embedded Systems
Zurich University of Applied Sciences
Zurich, Switzerland
hans.gelke@zhaw.ch

Abstract—One third of food produced in the world for human consumption – approximately 1.3 billion tons – is lost or wasted every year. By classifying food waste of individual consumers and raising awareness of the measures, avoidable food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. We specifically designed a deep neural network that classifies food waste for every time food waste is thrown in the waste bins. Our method presents how deep learning networks can be tailored to best learn from available training data.

Index Terms—deep learning, neural networks, machine learning, food waste

The automatic classification of food waste in waste bins is technically a difficult computer vision task for the following reasons. a) It is visually hard to differentiate between edible and non-edible food waste. As an example consider distinguishing between eggs and empty egg shells. b) Same food classes come in a wide variety of textures and colors if cooked or processed. c) Liquid food waste, e.g. soups and stews, and soft food waste, e.g. chopped vegetables and salads, can largely hide and cover visual features of other food classes.

In this research, we adopt a deep convolutional neural network approach for classifying food waste in waste bins [9]. Deep convolutional neural networks are supervised machine

My Research Paper

Grace Lindsay

According to [Mazloumian et al. \(2020\)](#), visually classifying food waste in bins faces three main challenges: distinguishing edible from non-edible waste, overcoming how diverse single food classes can be, and classifying images with liquid and soft waste that hide other waste classes. Here, I will test three different neural networks trained to classify food waste

Lovely!

Citations and references

≡ Google Scholar Deep Learning for Classifying Food Waste 

Articles About 76,900 results (0.18 sec)

Any time Since 2023 Since 2022 Since 2019 Custom range... Sort by relevance Sort by date Any type Review articles include patents include citations Create alert

Deep learning for classifying food waste
A Mazloumian, M Rosenthal, H Gelke - arXiv preprint arXiv:2002.03786, 2020 - arxiv.org
... food waste can be significantly reduced. In this research, we use deep learning to classify food waste in half a million images captured by cameras installed on top of food waste bins. ...
☆ Save  Cite  Cited by 4 Related articles All 3 versions 

[HTML] A Novel Model to Detect and Classify Fresh and Reduce Food Waste Using a Deep Learning Technique
TB Kumar, D Prashar, G Vaidya, V Kumar... - Journal of Food ..., 2022 - ... to be blamed for the majority of food that is wasted. Families are the waste, and we sought to reduce this by identifying fresh and damaged food. ...
☆ Save  Cite  Cited by 2 Related articles All 5 versions 

[HTML] An Application of Deep Learning Models to Auton Classification
AZ Espinoza - 2019 - search.proquest.com
... are trained and evaluated on a novel food waste dataset to assist in a ... of food image classification, rather than the task of food waste clas...
☆ Save  Cite  Related articles All 4 versions

[PDF] An efficient classification of kitchen waste using techniques
APS Kumar, K Buelaevanzalina - Turkish Journal of Computer and ..., 2020 - ... removal of knives and chemical waste. • In this project we have to build different types of waste. We used transfer learning models to classify th...
☆ Save  Cite  Cited by 2 Related articles All 2 versions 

Identification of Food Waste through Object Recognition
L Farinella, E Fernandes, N Michener... - 2020 11th IEEE ..., 2020 - ieexplore.ieee.org 

X Cite

MLA Mazloumian, Amin, Matthias Rosenthal, and Hans Gelke. "Deep learning for classifying food waste." *arXiv preprint arXiv:2002.03786* (2020).

APA Mazloumian, A., Rosenthal, M., & Gelke, H. (2020). Deep learning for classifying food waste. *arXiv preprint arXiv:2002.03786*.

Chicago Mazloumian, Amin, Matthias Rosenthal, and Hans Gelke. "Deep learning for classifying food waste." *arXiv preprint arXiv:2002.03786* (2020).

Harvard Mazloumian, A., Rosenthal, M. and Gelke, H., 2020. Deep learning for classifying food waste. *arXiv preprint arXiv:2002.03786*.

Vancouver Mazloumian A, Rosenthal M, Gelke H. Deep learning for classifying food waste. arXiv preprint arXiv:2002.03786. 2020 Feb 6.

BibTeX EndNote RefMan RefWorks

Put all full citations in the reference section.

Project Timeline

Now - Start thinking about your teams and project ideas

March 14 - Project Plan Homework assigned

April 4 - Project Plan Homework due, feedback the following day (teams cannot be changed after this point)

April 18 - In class project check-in

April 25 - In class project work time

May 2 - In class project presentations

May 9 - Project reports due

Exam Review

Exam will include multiple choice, T/F, and short answer questions

You will be tested on knowledge of climate change topics as well as machine learning topics

There will be questions about the papers we've read

You will not need to write code

To study for the exam:

For climate content, know:

- the basics of what is causing climate change
- where emissions are coming from
- changes we expect to happen as a result of warming
- basics of remote sensing
- SSPs
- human attitudes

To study for the exam:

For ML content, know:

- basics of the simple regression models
- the basics of artificial neural networks
- what architectures are used for what tasks and why
- what loss functions are used for different problems
- what balanced data means
- pros and cons of different evaluation metrics
- validation/generalization
- types of data (image, time series, etc)
- Transfer learning, representation learning

To study for the exam:

Make sure for each of the papers we've read, you know:

- PMIRO answers
- The general architecture of the model used, what the inputs to the model were, any preprocessing discussed, what the output was, what the objective function was, how the performance of model was evaluated, what kind generalization did they test for

In all cases, put emphasis on content in the slides