Welcome to the DataFirst kick off event!

Fall 2023

DataFirst (formally DataFest) is a semester-long event at USC where students from different backgrounds and programs get hands-on experience in real projects involving data science.

The projects are proposed by USC faculty and researchers. Students are assigned to a project of their choice. Students work in a small team supervised by a faculty mentor.

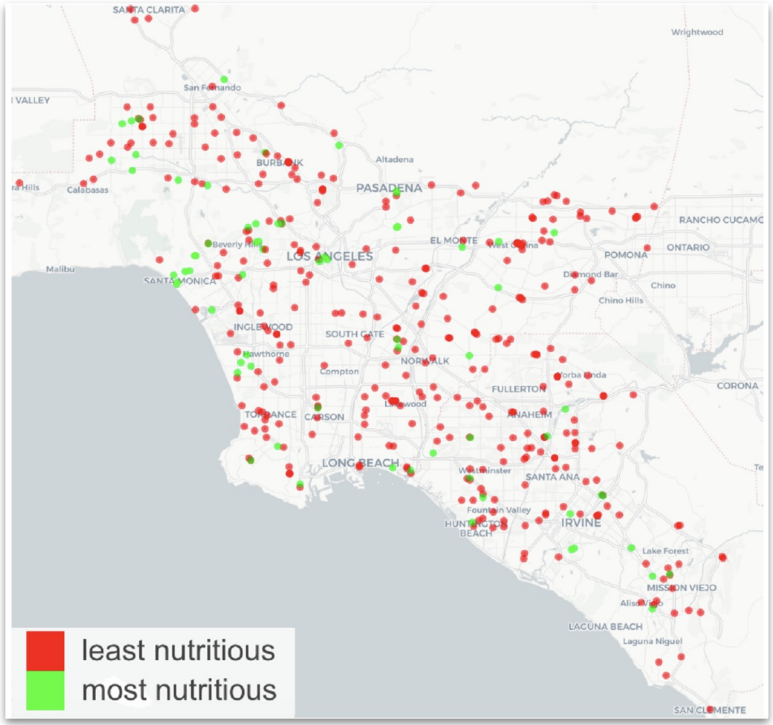
Started in 2019
120+ projects successfully completed

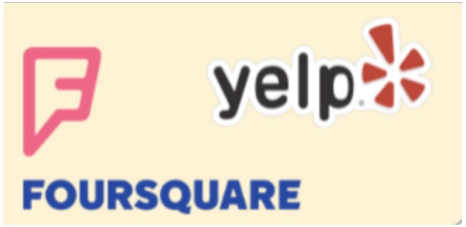
USC CKIDS Center for Knowledge-Powered Data Science

GRIDS @ USC

Example past projects

Ex. Scraping and analyzing menu data to score healthiness of restaurants in Los Angeles





- Use menu data + NLP to estimate nutrients and develop nutritional score metric
- Found that “most nutritious”restaurants are highly clustered in wealthy neighborhoods
- Whereas the “least nutritious” restaurants are everywhere

https://www.sites.google.com/view/continuous-food-indicator

Ex. Characterizing Online Attitudes, Expectations, and Concerns about Novel Medical Treatments

*What are topics & sentiment expressed about **male birth control** on **Reddit**, and how do they vary across the US, and over time?*

Findings

- Sentiment analysis demonstrates a slight positive to neutral sentiment towards male contraceptives.
- Sentiment score remained stable throughout the 12 years time frame, which is likely due to the absence of a male contraceptive product.
- Topics were focused on possible side effects, alternative treatments, and family planning

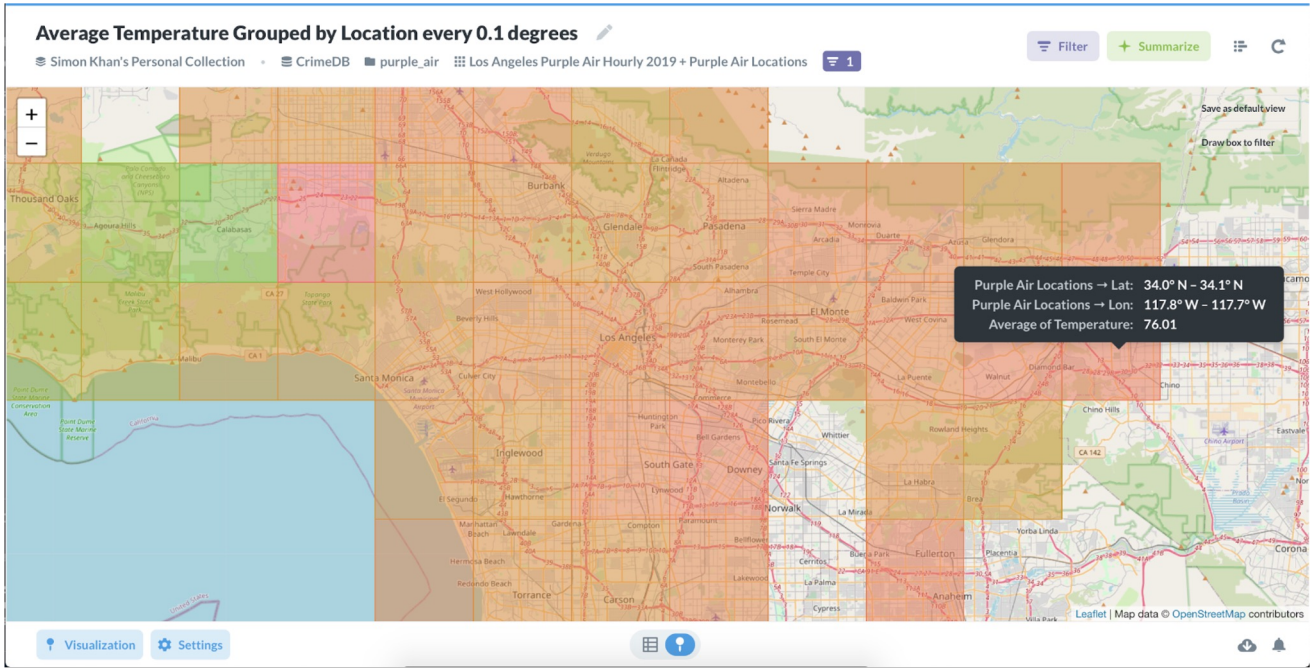
https://cllei12.github.io/

Ex. Machine learning to automatically detect and classify homeless encampments in Los Angeles

- Used the YOLOv5 Small object detection model, an end-to-end neutral network that makes predictions of bounding boxes and class probabilities all at once.
- Can successfully detect tents classify types of encampment

Ex. Heat Map - Los Angeles

- Show where heat islands (and cooler islands) exist in Los Angeles
- Build an interactive heat map of LA county that has an intuitive UI
- Experiment with different correlations of data such as temperature/crime activity, temperature/average income



Participating in DataFest

Next Steps
Why should I participate?

Why should you participate?

- Improve your data science skills by working with real data

- Get closer experience working with a faculty mentor than you do in classes

- Hands-on learning of the research and/or production process

- Contribute towards exciting research that you can add on your resume!

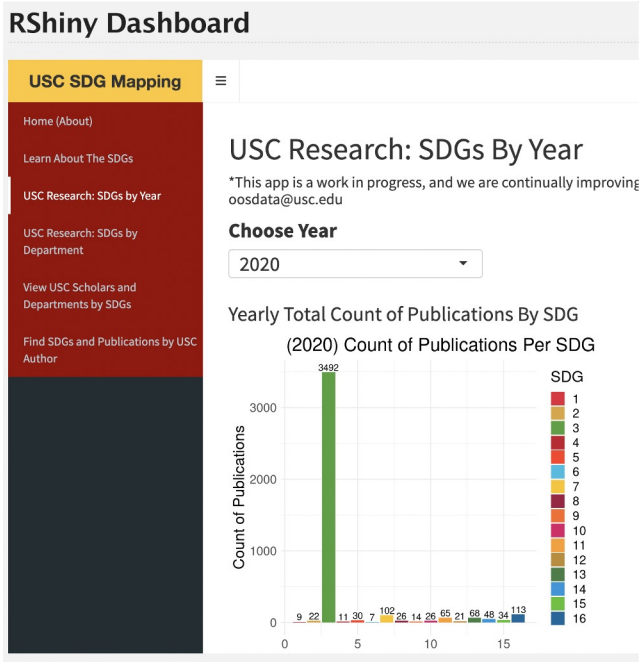
- Work on real world, interdisciplinary problems that combine Data Science with fields like Geology, Neuroscience, Sports, and Health!

Why should you participate?

Get a paid research position in the future!

Documented by a quote from a previous mentor:

> *I wanted to just shout out a huge thank you for including me in this CKIDS DATAFEST as a mentor this semester.*
>
> *The students were absolutely fabulous and helped us make huge headway on our Asgmt: Earth Research and Engagement Goals.*
>
> *We will also be hiring one of the students - as a PSIP intern - specializing in data analytics to finish this research dashboard and to help us with all of our Asgmt: Earth Power BI dashboards as well!*



Incentives for Participation: Awards!

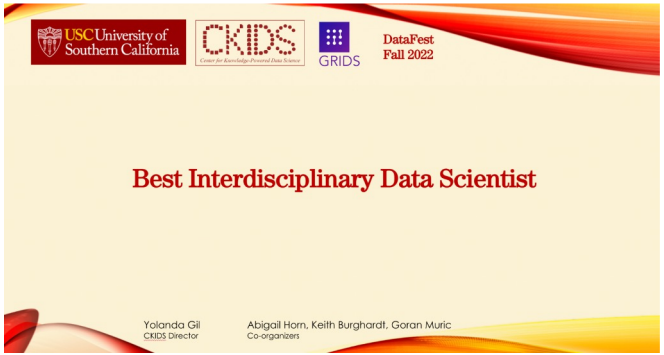
- At the end of the semester, faculty mentors nominate their students for awards

- There are 9 types of awards, for individuals and for groups

- Awards are provided when there are ties as well

- Each award receives an **$100 Amazon gift card**

- Something else to add to your CV!



USC University of Southern California

CKIDS Center for Knowledge-Powered Data Science

GRIDS

DataFest
Fall 2022

Best Data Science Teamwork

Awarded for attacking a large and diverse problem area by dividing and conquering, effectively creating "pods" of two students working on different topics, and on two different data sets to ensure generality of their approaches. This distributed yet redundant approach allowed the group to explore many different approaches and help each other out.

Yolanda Gil
CKIDS Director

Abigail Horn, Keith Burghardt, Goran Muric
Co-organizers

Create a project website!

Create a digital and visual archive for your project by creating a project website

Award for Best Project Website Spring 2022:
https://sites.google.com/usc.edu/social-media-habits

GRIDS

Graduates Rising in Informatics and Data Science (GRIDS) is the first graduate student organization within the Department of Computer Science at the USC Viterbi School of Engineering.

Our mission is to advance the academic and professional interests of Informatics and Data Science students at USC through experiential learning, corporate partnerships, and engaging networking opportunities that build community.

GRIDS

Industry and Professor Talks

Anything that you want to learn about data science, there is someone here at USC who knows it.

Projects

Working on real-world projects that address problems you care about is the best way to learn data science.

Conversations

Interacting with people from diverse backgrounds who share the same passion for data science as you is the best part about it.

Fall 2023 plans

GRIDS

- Work closely with CKIDS over the semester
- GRIDS will helps pair students and professors with similar interests
- We also provide day-to-day mentorship to students as they work on projects
- Providing office hours every 2 weeks for DataFirst Projects

Connect with us!!

GRIDS

GRIDS @ USC

LinkedIn

@GRIDSUSC

Instagram

GRIDS @ USC

Twitter

GRIDS @ USC

GRIDS Membership Form

So I’m interested, what now?

Presentations from faculty:

- **3 min** pitch, followed by **1 min Q&A**
- Follow up with faculty mentors after the event to answer further questions
  - **Mentors**: let students know how they can best reach you (email, set up a Zoom office hour, Slack, etc.)
- How do I sign up?
  - CS Student Affairs and GRIDS will circulate form → Respond by ***Sunday September 3rd***
    - **Form: >>> https://forms.gle/VREHxGuJUqkbbTWS8 <<<**
  - In this form, you’ll be asked to rank the projects by level in interest
  - Based on interest, we will assign you to projects
  - Assignments will be made by **Saturday September 9th**
- FAQ:
  - What if I’m only interested in one project? Only rank that project but it means that you may not be chosen for a project at all
  - Will everyone be assigned to a project? Unfortunately, not everyone will be assigned to a project.

Summary of deadlines

- Student application deadline Sunday, September 3, 2023
- Faculty feedback Thursday, September 7, 2023
- Student assignments Saturday, September 9, 2023
- Midterm
  Wednesday October 18, 2023

Project presentations

1. **Pyleoclim: A Python Package for the Analysis of Paleoclimate Data** (Prof. Deborah Khider and Prof. Julien Emile-Geay)
2. **Learning and forgetting in neural networks** (Prof. Marcin Abram)
3. **Utilizing AI Generated Images for Object Detection and Classification** (Prof. Seon Ho Kim)
4. **A Knowledge Graph of a Crowdsourcing Event** (Prof. Daniel O'Leary)
5. **Urban Futures Data Core** (Prof. Alice Chen)
6. **AI/ML assisted fault detection in foundry processed devices** (Prof. Andrew Rittenbach and Prof. JP Walters)
7. **California Public Sector (CAPS) Job Market: Online Dashboard** (Prof. William Resh)
8. **Does Municipal Broadband Deliver as Promised?** (Prof. Hernan Galperin)

Project presentations (cont.)

9. **Automated question type coding of forensic interviews** (Prof. Thomas D. Lyon)
10. **Building a Platform for NFL Data Insights** (Prof. Jeremy Abramson)
11. **Understanding the Relation Between Noise and Bias in Annotated Datasets** (Negar Mokhberian)
12. **Federated Learning for Neuroscience** (Prof. Jose-Luis Ambite)
13. **Bad Writing is "Fine": Tuning an LLM to Suggest Improvements** (Prof. Benjamin Nye)
14. **Analyzing Open Source Software Ecosystems** (Prof. Jeremy Abramson)
15. **Build a multilingual decipherment system** (Prof. Jonathan May)
16. **Natural language processing of safety reports in nuclear power plants** (Prof. Najmedin Meshkati)
17. **Application of AI, ML and NLP in Runway Safety** (Prof. Najmedin Meshkati)
18. **AI Ethics for Smart Health through Smart Watches** (Prof. Yolanda Gil)

Contact mentors for Q&A on Piazza

- DataFirst student info (includes all other links below): https://ckids-datafirst.github.io/website/info-students/

- Kickoff resources: https://www.isi.edu/~ulf/DataFirstFall2023/index.html
- DataFirst Piazza home: https://piazza.com/usc/fall2023/dsci591/home
  - Sign-up: https://piazza.com/usc/fall2023/dsci591
- Pointer to all projects: https://www.isi.edu/~ulf/teaching/DataFirstProjectsFall2023.html
- Project application form: https://forms.gle/ABM5caP55kCe1eqz8