2021 Data Science Competition Information Kick-Off Session

Nick Duffield

Director, Texas A&M Institute of Data Science

https://tamids.tamu.edu   **https://u.tamu.edu/tamids2021**

Welcome To The New Competition!

- New Organizing Team

- New Sponsor

- New Problem Setting

- New Categories and Prizes

Some Things That Haven't Changed

- Still open-end problem giving flexibility to define your approach
- Still operating the competition online
- Still using e-campus for submission

Tonight's Schedule

- 6:00-6:30: **Nick Duffield** (TAMIDS Director, Organizing Team)
  - Competition Organization and Setting

- 6:30-6:50: **Allyson King** (2020 Winner)
  - Experiences and best practices from the 2020 Data Science Competition

- 6:50-7:00: **Darren Homrighausen** (Statistics, Organizing Team)
  - Competition Categories, Events and Prizes

- 7:00-7:10: Break

- 7:10-7:30: **Josiah Coad** (2019 Winner)
  - Experiences and best practices from the 2019 Data Science Competition

- 7:30-7:50: **Guy Whitten** (Political Sciences, Organizing Team)
  - Perspectives on Analysis of Election Data

Organizing Team
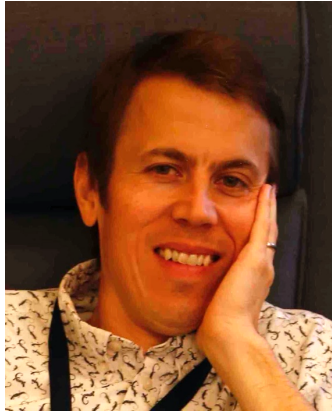
James Caverlee
Computer Science & Eng.

Darren Homrighausen
Statistics

Guy Whitten
Political Science

Nick Duffield
Electrical & Computer Eng.
TAMIDS Director

Jennifer South
TAMIDS

https://u.tamu.edu/tamids2021

Experiences From Previous Years

Allyson King

- 2020 Winner (Undergrad Team Inflightful)
- Set to graduate May 2021 in Stat and Comp Sci
- Started TAMU Datathon with Josiah Coad
- Will work as software engineer / data analyst

Josiah Coad

- 2019 Winner (Grad Team Superficial Intelligence)
- Set to graduate May 2021 on Stat, Math, Comp Sci
- Started TAMU Datathon with Allyson King
- Planning to take Master's degree in Machine Learning

Sponsors

TEXAS A&M
Institute of
Data Science

Chevron

TEXAS A&M UNIVERSITY
Department of Computer
Science & Engineering

TEXAS A&M UNIVERSITY
Statistics

TEXAS A&M UNIVERSITY
Department of Electrical
& Computer Engineering

TEXAS A&M UNIVERSITY
Wm Michael Barnes '64 Department of
Industrial & Systems Engineering

TAMIDS will arrange a Data Science student information session with Chevron after the competition completes

Competition Logistics

Getting Started

- **Register individually as a competitor by 11:59pm Monday March 1**
  - Google form: https://forms.gle/JCFdiv4KZAYK69ZB6
  - Include *team name* and *known team member names*
    - Team of 1-5 members, graduate or undergrad students enrolled in TAMU during Spring 2021
    - Team may specify a mentor (TAMU faculty member or postdoc)
  - Undergraduate and graduate competition divisions
  - All team members must register individually by March 1 deadline
- **Need teammates?**
  - Introduce yourself on the slack channel
  - https://tamids2021.slack.com/channels/find-team-members
    - Also available by visiting the registration page
- **Teams & members will be setup in e-campus course for the competition**
- **Organizational questions or problems?**
  - Contact Jennifer South, TAMIDS, jsouth@tamu.edu

Competition Period

- **Start: Release of Data: Wednesday March 3:**
  - Release of data to competitors through e-campus

- **Midpoint Event: Monday March 15**
  - Technical help session
  - *Midpoint competition prizes (more about this later)*

- **End: Submission of Entries: Wednesday March 31 11:59pm**
  - Deadline for submission of entries through e-campus

- **Technical questions?**
  - Submit to discussion board in e-campus

https://u.tamu.edu/tamids2021

Submission format

- **Report**
  - Up to 10 pages, 10 pt arial, 1-inch margins all round, pdf format
    - Should be self-contained, but supplementary materials allowed (see below)
  - We will supply a rubric to specify required reports elements

- **Executable code**
  - e.g., a Jupyter notebook, github repository
  - Don't just submit a printed code listing
  - Ask in e-campus if in any doubt about your proposed executable code format
  - Reuse OK, but acknowledge any external sources used

- **Optional supplementary materials**
  - Such as supplementary figures, a dashboard, a website, or an app
  - Ask in e-campus if in doubt about proposed supplementary materials format
  - *Special Team Prize for Best Supplementary Materials (more about this later)*

Judging and Final Events

- **Wednesday March 31: Submission Deadline**

- **Wednesday April 7: Finalists Announced**
  - Finalist teams will prepare a 10-minute video presentation for the final

- **Wednesday April 14: Online Finalist Event**
  - Finalist teams video presentation + 5 minutes live Q&A with judges
  - Announcement of winners
  - Awarding of Prizes (more details later)

The Competition Setting

Competition Theme

- **2018 Competition**
  - Chicago Taxi Rides
- **2019 Competition**
  - LA Bike Share
- **2020 Competition**
  - Airline Performance Statistics
- **2021 Competition**
  - US Presidential Election Spending

Competition Approach (2020)

From the 2020 Competition Announcement

In this year's Data Science Competition, students will use airline data from the US Bureau of Transportation Statistics to develop performance measures for airlines and models to forecast expected performance. Students are encouraged to identify one or more performance measures that they would like to model. Examples include:

- Customer satisfaction implied from travel delays, lost baggage, or other service metrics.

- Business performance measured by net income

Students should not feel limited to these measures, and the competition provides them with the opportunity to showcase novel insights and analyses of the data.

New theme, same approach

- **Open ended**
  - Formulate your specific approach with a broad problem context
  - Find and use additional data if helpful
- **Define your metrics**
  - Identify relevant performance metrics for the area
- **Develop your model**
  - Identify determining features and develop predictive model for performance
- **Show insight and make recommendations**
  - What operational step can be taken to improve performance in future?

US Presidential Election Spending

- **Enormous amounts of money are spent in US elections**
  - Donations by individuals to candidate campaigns
  - Union and corporation spending through Political Action Committees (PACs)
- **Supporting diverse campaign activities**
  - From advertisement and broadcasts to door-to-door meeting with voters
- **Data**
  - US Federal Election Commission maintains comprehensive and detailed data on election spending and outcomes

Data Challenge Questions

- **What is the effect of different type of donation on election outcomes?**

- **How should money be spent efficiently have the greatest effect?**
  - Where and through which channels should it be directed?
  - Is this effectiveness getting more or less pronounced over time?
  - What other factors and data may be important in understanding ?

- **What are the best metrics for electoral impact?**

The Operational Recommendation

- **Imagine yourself as a political consultant**

- **Develop models**
  - Predict the impact on electoral outcomes of contributions and spending by campaigns and donors.

- **Develop recommendations**
  - What level of funds are needed to have impact?
  - Where should the funds be directed?

Competition Data Release: March 3rd, 2021

Tonight's Schedule

- 6:00-6:30: **Nick Duffield** (TAMIDS Director, Organizing Team)
  - Competition Organization and Setting
- 6:30-6:50: **Allyson King** (2020 Winner)
  - Experiences and best practices from the 2020 Data Science Competition
- 6:50-7:00: **Darren Homrighausen** (Statistics, Organizing Team)
  - Competition Categories, Events and Prizes
- 7:00-7:10: Break
- 7:10-7:30: **Josiah Coad** (2019 Winner)
  - Experiences and best practices from the 2019 Data Science Competition
- 7:30-7:50: **Guy Whitten** (Political Sciences, Organizing Team)
  - Perspectives on Analysis of Election Data