

COGS 9 - Introduction to Data Science

Winter 24

Week 3 Discussion

About me..



- 2nd Year CSE Grad Student
- B.E in Computer Science and Engineering
- Current research focuses on Computational Neuroscience, AI/ML

Agenda for Today

- Project Group Formation
- Quick overview of the Reading for this week
- Questions

Project: Groups

- Form groups of 4-5 and come up with a group name
- Discuss about topics that interests you within the group
 - You will be focusing on how to integrate a variety of Data Scientific methods and theories to address a problem
 - **No** coding
 - Topics can include churn prediction, how does Black Friday sales affect the economy, is social media spreading more positive or negative information and so on.
 - The project is open ended

Let's take 5 - 10 minutes and get back

[Link](#) to form

Fill this form once you have finalized your team!!!



Project: **Directions**

Once you have decided on your topic -

- Look for appropriate datasets. Kaggle, Google Dataset Search, Data.gov, UCI ML repository are few sites you can check out
- Analyse what's given in the data and what inferences can be made
 - What kind of problem are you trying to solve?
 - Can you derive useful information from raw data provided?
 - What impacts does solving the problem have?

You can change the question or dataset later!

Attendance for the session

Fill out the form below to mark your attendance

[Link:](#)



Project: Quick tips

- Share your email IDs with your team members
- Establish communication channels with your team members. Example: Slack, Discord, WhatsApp, Messenger etc.
- Maintain a Drive folder to store your project documents and files
- Go over your schedule and set up weekly meetings
- Coordinate with your team regularly
- Divide your work equally

Reading: 50 Years of Data Science

- Talks about the advancements in the field over the last 50 years
- How is data science different from statistics
- Difference between generative and predictive modeling
- Common Task Framework
 - Consider a publicly available training dataset
 - Competitors competing to infer a class prediction rule from the training data
 - Scoring referee reports the score achieved by the submitted rule of unseen test data

Reading: 50 Years of Data Science (contd.)

6 major divisions of Greater Data Science (GDS)

1) Data Gathering, Preparation and Exploration

- Gathering relevant information
- Preparing data for analyses
- Exploratory Data Analysis

2) Data Representation and Transformation

- Modern Databases: SQL, noSQL, Live data streams
- Mathematical representation: Fourier Transforms, network data

Reading: 50 Years of Data Science (contd.)

6 major divisions of Greater Data Science (GDS)

3) **Computing with Data**

- Cluster and Cloud Computing
- Python, R

4) **Data Visualization and Presentation**

5) **Data Modeling**

- Generative and Predictive Modeling

6) **Science about Data Science**

- Analyse effectiveness of workflow using various metrics

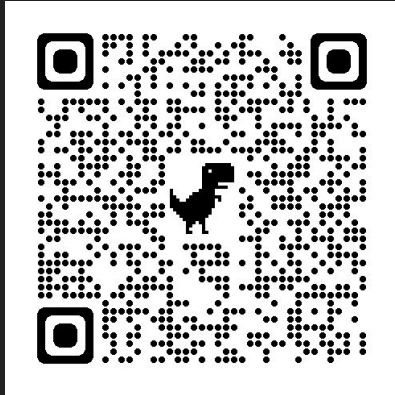
Reading: 50 Years of Data Science (contd.)

- Cross-Study Analysis
 - Different teams produce different predictions for the same problem. Example: medical domain
- Cross-workflow analysis
 - Different workflows can lead to different conclusions
- Prediction
 - Concerned about performance (metrics)
 - More CS focused
- Inference
 - Understanding input-output relationships
 - Central to traditional statistics

Questions?

Make sure you filled out the group information and attendance forms!

Groups



Attendance

