

Course grading

- Assignments 30/100
 - Mid semester & Final examination 50/100
 - Course project and presentation 20/100
-
- Assignments 20/100
 - Mid semester & Final examination 50/100
 - Course project and presentation 30/100

Class Projects

- Data Quality
- Data Privacy
- Mlops

- Due date – November 17, 2023
- Midpoint check – November 3, 2023

- Expectation – A paper (single spaced, 6 pages)
- Example implementation

Class Project – Data Quality

- Definition of Data Quality
- How to measure Data Quality ?
- What are the challenges ?
- Review of some of the techniques
 - Data Deduplication
 - Data Consistency
 - Data Accuracy
 - Information completeness
- How can ML algorithms ML help for Data quality ? n relation to traditional data quality frameworks ?
-

Class Project – Data Quality

○

- An overview of data quality frameworks
- Corinna Cichy and Stefan Rass.
- Lisa Ehrlinger and Wolfram Wöß. A survey of data quality measurement and monitoring tools. Frontiers in big data, page 28, 2022.
- A Survey on Data Quality: Classifying Poor Data
- Nuno Laranjeiro; Seyma Nur Soydemir; Jorge Bernardino

Class Project – Data Privacy

- Definition of Data Privacy
- Process need to preserve Data Privacy
- What is differential data privacy ?
- Review of some of the techniques

Class Project – Data Privacy

- <https://www.scientificamerican.com/article/privacy-by-the-numbers-a-new-approach-to-safeguarding-data/>
- Cynthia Dwork. A firm foundation for private data analysis. Communications of the ACM, 54(1):86–95, 2011.

An example implementation

- <https://github.com/google/differential-privacy>
- <https://github.com/tensorflow/privacy>
-

Class Project – MLops

- Definition of Mlops

MLOps stands for Machine Learning Operations. MLOps is a core function of Machine Learning engineering, focused on streamlining the process of taking machine learning models to production, and then maintaining and monitoring them.

- DevOps - DevOps is a combination of software development (dev) and operations (ops).
- practices and tools designed to increase an organization's ability to deliver applications and services faster than traditional software development processes
- Agile, CI/CD

Class Project – MLops

- Why do we need MLops ?
- What are the benefits ?
- MLops – Pipeline
- What are the challenges ?

Class Project – MLops

- =
- Why Machine Learning Models Fail In Production ?
- Hidden Technical Debt in Machine Learning Systems - D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
- Scaling Machine Learning as a Service - Li Erran Li