

Leo Ding

NEW YORK CITY, NY

ld2425@nyu.edu

[Personal Website](#)

[LinkedIn](#)

[GitHub](#)

Experience

Data Science Lead – NASA SPOCS

- automated experiment regarding virulence and antibiotic resistance in E. coli in microgravity environments
- decided data collection and analysis needs for proposed project
- collaborated with university students from USC, UCLA, and UC Berkeley

Projects

- [Hidden Markov Model Part Of Speech Tagger](#)
 - trained algorithm to tag word parts of speech
 - made decisions based on prior and word transition probabilities
- [Spam Classifier](#)
 - used TfidfVectorizer to determine word importance in SMS messages
 - created a Logistic Regression model to classify spam messages
- [Movie Ratings Analysis](#) (Introduction to Data Science Capstone)
 - applied Principal Component Analysis to effectively reduce features for analysis
 - used Multiple Linear Regressions to evaluate relationships between movies

Technical Skills

- Python: pandas, NumPy, Matplotlib, Scikit-learn
- Java
- PostgreSQL
- MongoDB
- Apache Ecosystem: Hadoop Distributed File System (HDFS), Hadoop MapReduce, HBase, Hive, Impala, Pig, Spark SQL
- GitHub
- HTML, CSS, JavaScript

Education

September 2020 - Present (Expected May 2023)

New York University – Joint Computer Science and Data Science Degree

GPA: 3.89