Martin Thomas Durkin

412-889-8244 | mthomasdurkin@gmail.com | https://thomasdurkin.github.io/

EDUCATION

University of Rochester — **Rochester, NY**

M.S. in Data Science — *Class of Fall 2022*

GPA: 3.79/4.0

Coursework: Computational Statistics, Data Mining, Database Systems, Tools for Data Science, Deep Learning, Introduction to Statistical Machine Learning, Predictive Analytics using Python,

Rensselaer Polytechnic Institute — **Troy, NY**

B.S. in Computer Science — *Class of 2021*

GPA: 3.54/4.0

Coursework: Application Programming using Java, Computer Organization, Data Structures, Introduction to Algorithms, Large Scale Programming, Network Programming, Open Source Software, Operating Systems

SKILLS

Programming Languages: Python (PyTorch, PySpark, Pandas, NumPy, Matplotlib, Seaborn, SciKit-Learn), Bash, C, C++, C#, Java, JavaScript, R, SAS, SQL, Visual Basic

Web Development: CSS, HTML, MongoDB, .NET Framework, Node.js, PostgreSQL, Vue.js

Other: Linux Ubuntu, Git, Confluence, Jira, Jupyter Notebook, Databricks

Certifications: Databricks Data Engineer Associate

WORK EXPERIENCE

Harris School Solutions

Research and Development Intern — **(January 2020 – October 2021)**

- Engaged in an Agile workflow to modernize the BOCES module for the purpose of providing a better user experience over the legacy WinCap application, which manages contracts, shared programs, and services used by school districts
- Implemented the entire Actual Use Bill Schedule financial management module, using Blazor, C#, and Harris' Cheyenne Framework, allowing clients to issue, process, and post bills
- Actively participated in code reviews and communicated with QA to accelerate the release of new features to beta users
- Utilized Jira to manage tasks throughout two week sprints and prioritize items in the backlog with the Product Owner

RELEVANT PROJECTS

NASA Capstone Researcher — **Fall 2022**

- Worked in a team of students alongside experts from NASA and Coral Vita to build a machine learning pipeline using Python that is able to determine coral presence and bleaching levels in order to assist large-scale coral farming
- Employed data from two NASA satellites that was spatially and temporally aligned with coral databases resulting in the ability detect coral at a specific location and date with an overall accuracy of 91 percent
- Dataset consisted of 31 LiDAR features plus an additional 21 features were engineered to further enhance the model's capability of correctly identifying moderate/severely bleached coral

Classification of Cancer Discussion Posts — **Fall 2022**

- A comparative study of deep learning models, conducted with a partner, to correctly identify the cancer a patient has, creating a more streamlined process when making a post on the Cancer Survivors Network website
- Scraped data using Beautiful Soup resulting in over 100,000 total posts and 13 unique classes
- Used PyTorch to create and train multiple deep learning models and determined that a stacked model consisting of a Bi-LSTM and transformer encoder provided the best results at nearly 71 percent

Trending Research Topics — **Fall 2021**

- Utilized Dimensions.ai to construct a dataset consisting of over 51,000 grants from 131 R1 universities in order to compare trends in Computer Science research to those on CSRankings.org by applying LDA and BERT topic modeling
- Data was cleaned using numerous NLP techniques in order for the models create distinctive and identifiable topics
- Additional models were created for only the University of Rochester, seeing a shift in research towards lasers, cancer treatment, and psychology