Matthew Ghannoum

LinkedIn GitHub Portfolio Website Medium Blog

Data Science Experience

Centre for Skills Scholarship - Research in collaboration with the Akkodis Group

05/2024 - 10/2024

Achievements

- Implemented a model for identifying and quantifying skill mismatches and alignments between job postings, courses and students. To improve product offerings and make better recommendations.
- Over 15000 times faster execution of an econometrics similarity score (~50 minutes to ~250 milliseconds) by transforming the original algorithm into a series of matrix operations and utilizing CuPy (NumPy-like API built on Cuda). This enabled much quicker experimentation on our models.
- Processed over 2.5 million text documents to extract and map skills, using Named Entity Recognition and Semantic Textual Similarity models. To utilize the High Performance Computing hardware available (multiple CPU and GPU), both SpaCy pipelines and Python Multiprocessing were implemented.

Responsibilities

- Reading through relevant scientific literature to learn how to implement specific models and to improve the understanding of the problem space.
- Unit testing an internal Python package to ensure updates didn't cause bugs or regressions.
- Carefully defined requirements with non-technical stakeholders to ensure we met their intended goals.
- Organizing and prioritizing tasks to keep track of progress and manage blockers.
- Communicating and presenting research findings to technical and non-technical stakeholders.

Research Assistant at the UTS Behavioral Data Science Lab with the Burning Glass Institute

02/2023 - 03/2024

Achievements

- Developed and scaled a Labour Market Graph in Neo4j to over 3 million nodes and over 70 million relations by aggregating and processing data from multiple sources.
- Developed novel and efficient skill inference models in Cypher (graph query language) and Python.

Responsibilities

- Validating data and enforcing constraints using built in Neo4j profiling tools such as Meta-Graphs.
- Containerized all services (Docker) and provided extensive documentation to streamline deployment.

Education

Bachelor of Computer Science (Honours), University of Technology Sydney

Centre for Skills Scholarship | Dean's List 2024 | Weighted Average Mark (WAM): 85.73
Grade Point Average (GPA): 6.47 | Major: Artificial Intelligence and Data Analytics

Noteworthy Subject Marks

High Distinctions: Natural Language Processing, Machine Learning, Software Engineering Studio 1A & 2A, Honours Project, Application Development in the iOS Environment, Project Management and the Professional.

Distinctions: Deep Learning and Convolutional Neural Networks, Mathematical Modelling 2, Social and Information Network Analysis, Data Structures and Algorithms, Web Systems and Network Fundamentals.

Software Development Experience

Web Developer at Infrastructure Partnerships Australia

03/2021 - 10/2023

- Updated and maintained the Single Page Application built in React.js for the member's section.
- Developed content pages based on Figma designs using WordPress, PHP, JS, HTML and CSS.
- Coordinated projects with technical and non-technical stakeholders to ensure deliverables met expectations before set deadlines.

Web Development Intern at Stomble

07/2020 - 01/2021

- Built components for authentication interfaces and parts of the video player using React and Redux.
- Learned several AWS technologies, mainly API Gateway and Lambda with the Serverless Framework.

Projects

Movie Recommendation Engine: Built a Latent Factor Model that recommends movies by first creating a matrix of users and movies, which is then factored into two separate matrices representing latent features of users and movies. These matrices are initially filled with random values, which are then optimized using Gradient Descent to closely match the original ratings matrix.

Melanoma Detection App: Using the NasNet CNN model from Keras/Tensorflow, I trained and deployed a melanoma classification model into a desktop application (React.js and Electron) intended for CAD.

Course Recommendation App: Built a full stack web app to recommend university subjects (data collected through web scraping) using the Serverless Framework with AWS API Gateway, Lambda, Amplify and RDS.

Regression ML Web App: Implemented Linear, Polynomial and Decision Tree regression from scratch in Typescript and visualized them in a React.js application deployed here: Link to this web application.

Disease Classification Model: Trained, validated and tested a Decision Tree classifier (Scikit-Learn) on a dataset from Kaggle containing boolean values for a set of symptoms/diagnostics and a prognosis given by a doctor (the target class). After analyzing the results of the model, I explained several factors (both positive and negative) that contributed to the model's high performance. Link to my blog article on this project.

Skills

Data Science Skills

- Concepts and Theory: Descriptive Statistics, Inferential Statistics, Hypothesis Testing, Cross Validation, Machine Learning Evaluation Metrics, Optimisation, Regression, Classification and NLP.
- Data Processing: Pandas, Numpy, Python Standard Library (mainly os, json, csv) and Multiprocessing.
- Data Visualisation and Reporting: Matplotlib, Seaborn, Plotly and Y-Data Profiling (Pandas Profiling).
- Statistics and ML Modelling: Scipy, Scikit-Learn, Tensorflow, CuPy and MLX.

Software Development Skills

- Deployment and CI/CD: Docker, Docker Compose, Setting Up Environments and Git.
- Databases: MongoDB, MySQL, PostgreSQL, SQLite and Neo4j.
- Backend Frameworks: Python FastAPI, Python Flask, Node.js Express and Go Fiber.
- Frontend Technologies: React.js, TailwindCSS, Next.js and Astro.