Max Matkovski

✉ maxmatkovski@gmail.com ⬡ github.com/maxmatkovski

Education

Georiga Institute of Technology   **2024**
Master of Science (M.S.), Computer Science

University of California, Los Angeles   **2023**
Bachelor of Science (B.S.), Cognitive Science, Computer Science

Skills

General: Data Engineering, Machine Learning, Data Cleaning, ETL, PySpark, Redshift, Git, Lambda, Node.js, Snowflake, Airflow, CloudWatch, TerraForm, AWS, GCP, DataBricks, Tableau, PowerBI, dbt
Technologies: Python 3, R, SQL, C++, C, HTML/CSS, React Native, JavaScript, Pandas, NumPy, Matplotlib, SkLearn, TensorFlow, MySQL, PostgreSQL
Languages: **Native:** English – **Proficient:** Russian, Spanish, Portuguese – **Elementary:** Italian, Hebrew, Farsi

Experience

Data Science Intern   **2021 – 2022**
Finphil
- Helped create customized recommendation systems using Content Filtering and Collaborative Filtering to recommend retail investors articles across 90 different investment sectors using Python 3 and TensorFlow.
- Compared Collaborative Filtering models by utilizing different recommendation techniques (user-user vs. item-item), correlation metrics (cosine similarity vs. Pearson's), and neighborhood sizes.
- Optimized engagement with platform and improved correct content prediction (explicit rating) percentage from 40 percent to 68 percent via transition to item-item filtering, Pearson Correlation Function, and neighborhood size of five.
- Managed SQLite data set of over 100,000 data points collected from 2,000 users during prototype launch of Finphil customized wealth management platform to understand user behavior.
- Performed data ingestion using PySpark and contributed to ETL process using AWS S3 and AWS Kinesis.
- Generated over 50 concrete data visualizations and tracked a variety of KPI's using PowerBI.

Data Engineer Intern   **2020 – 2021**
LeNgineer
- Scraped and imported data into an unstructured AWS S3 data lake and later extracted unstructured data into a structured data warehouse using lambda function.
- Structured data was sent with api to Python for analysis and machine learning, using PySpark and Spark.ml.
- New data was stored in previous AWS data warehouse as new table and used for data visualization on web platform.
- Built data pipeline and ETL process using Python, Selenium, AWS S3, Lambda, Cloudwatch, TerraForm, etc...

Software Engineer Intern   **2019 – 2020**
LeNgineer
- Developed a full-stack web based search engine for used cars using HTML, CSS, JavaScript, SQL, and React.js
- Created web scraper which used Python (Selenium) to interact with web page elements and gather used car data.
- Designed several product features eg. customized used car recommendations and location based query results.
- Conducted market research via surveying of potential consumers, while collaborating with fully remote team three hours ahead.

Research Data Analyst Intern   **2018 – 2019**
Sanford Burnham Prebys Medical Discovery Institute
- Utilized machine learning and data clustering techniques (Sklearn) to detect Alzheimer's disease in lab data.
- Hypothesis tested data to investigate possible correlation of Protein ATF6a with Alzheimer's Disease.
- Cleaned data to ensure uniformity and consistency across various samples, maintained data quality and currency.
- Helped write grant approvals to the National Institute of Health to support funding for our research project.
- Assisted with essential lab duties, including genoytyping, gel electrophoresis, autoclaving, etc...

Open Source

Natural Language Processing of 100,000 Climate Change Tweets | *BERT Model*
- Scraped 100,000 live Climate Change related tweets using snscrape module, performed text data ingestion and cleaning using pandas and regular expressions.
- Processed scraped text data with NLTK and WordPunctTokenizer, using techniques such as tokenization, lemmetization, and stopwords, generated visual Word Cloud with most commonly occurring words in twitter data set.
- Performed sentiment analysis using Bidirectional Encoder Representation from Transformers (BERT) Classifier, analyzed findings to discover slightly negative (2.78/5) average sentiment score for climate change tweets.