Aman Chopra

Seeking: Summer 2022 internship in Data Science/Machine Learning/Data Engineering

Interests: Nutrition, productivity tools, guitar and singing , running, basketball, listening to entrepreneurship podcasts, cross-country road-tripping.

Skills: Python, Java, R, SQL, HTML, CSS, Flask, Hadoop, Spark, HIVE, Microsoft Azure, Airflow, Docker, Jupyter, ML/Deep Learning (scikit-learn, tensorflow, pytorch), data visualization (plotly, seaborn, matplotlib, ggplot), webscraping/automation (bs4, selenium).

Honors: Eagle Scout, IBM Data Science Professional Certificate Recipient, Dean's List

Email: chopra21@purdue.edu

EDUCATION

Purdue University - B.S. Data Science (GPA: 3.89) **Expected May 2023**

EXPERIENCES

Tamr, Inc. – DataOps Engineer **January 2021 – Present**

Aarete Consulting – Data Scientist **August 2021 – November 2021**

- Supported clients like Blue Shield and Molina Healthcare in various ad-hoc requests including data cleaning/integration, identifying savings opportunities in claims by validating medical service pricing with CMS guidelines, and building RPA solutions for day-to-day tasks.
- Constructed a web-scraping bot to collect prices for medical services at over 50,000 providers, resulting in third-party enrichment to client datasets.
- Designed a greedy algorithm to assign MDs, Consultants, and Analysts to client teams while maintaining skillset diversity and various other constraints.

Munich Reinsurance – Data Science Intern **June 2021 – August 2021**

- Implemented ML algorithms (GMM, isolation forest outlier detection) to quantify risk during underwriting and benchmark employees selling to risky customers based on demographics, mortality predictions, and likelihood of lapse.
- Designed models (GLM, HDBSCAN, XGBoost) to run quarterly audits for clients (AAA, Prudential Financial, Principal) by clustering and detecting abnormal policies issued to individuals with elevated BMI, abnormal lab tests, and family history.
- Constructed an ETL pipeline using Airflow to process recurring data from clients and load it into production databases.

Sagepointe Advisors - Quantitative Analyst Intern **November 2020 – January 2021**

- Analyzed financial market data from various sources to construct diversified investment portfolios with stocks, bonds, treasuries, REITs, and precious metals. Implemented strategies including inverse volatility and risk parity to minimize risk. Measured performance using metrics such as CAGR, maximum drawdown, and Sharpe Ratio.
- Utilized sqlite3 for data storage and retrieval, pandas for processing and analytics, and plotly for data visualization.

Ford Motor Company (Global Data Insights & Analytics Team) - Data Science Intern **May 2020 – July 2020**

- Analyzed connected vehicle data to reduce warranty spending and gain insights about customer complaints regarding the AC system. Adopted an agile workflow and delivered ad-hoc reports to business partners.
- Produced a scalable ETL pipeline and dashboard for Climate Control to visualize AC system defects. Utilized techniques including KNN, anomaly detection, clustering, and regression.
- Leveraged HPC tools (Spark, Hadoop, HIVE, SQL) to query, merge, aggregate, and load high frequency data from HDFS. Compared Spark job execution times on PySpark, spark-submit, HIVE, and Apache Ambari.

PROJECTS

Health and Fitness Data Analysis **February 2021 – Present**

- Extracted and cleaned personal data from various fitness apps/wearables. Trained regression models predicting body fat percentage and total daily energy expenditure to help reach my fitness goals.
- Performed diagnostic testing to assess variance, residuals, outliers, and multicollinearity. Utilized cross-validated prediction and confidence intervals, R^2, F-tests, PRESS values for model selection.
- Designed and deployed a web application on Heroku using Python (Flask), HTML, CSS, SCSS, and JS to summarize my fitness achievements in 2021 (inspired by "Spotify Wrapped").

Brain MRI Image Reconstruction **September 2021 – December 2021**

- Utilized deep learning to reconstruct brain images and accelerate MRI scan times. Designed a U-Net-based generative adversarial network with dense and residual connections to optimize performance with regard to perturbations, distribution shift, and the ability to recover small details in the image.

Type 1 Diabetes (Newly Diagnosed) Management Dashboard **April 2021 – July 2021**

- Designed a dashboard streaming real-time data from in-body blood glucose sensors, the Apple Watch, and Fitbit to help manage diabetes and identify patterns in my blood glucose readings. Developed a pipeline in Python to automate the collection, preprocessing, and storage of data related to blood glucose, insulin intake, nutrition, and fitness.
- Implemented an LSTM recurrent neural network to predict future blood glucose readings and prevent hypoglycemia and hyperglycemia.

Sign Language Hand Gesture Recognition **February 2021 – May 2021**

- Trained and implemented a CNN (image processing) to translate hand gestures to audio and enable communication with students at the California School for the Deaf. Constructed a dataset with 100,000 labeled sign language images in different angles, lightings, and settings.

Gym Appointment Scheduling Automation **February 2021 – March 2021**

- Designed a bot to automate scheduling appointments at Purdue University's Recreational Center using Selenium and Cron jobs.