NEIL CHITRE

San Francisco, CA | chitreneil5@gmail.com | +1 (341) 699 6496 | Github | LinkedIn |Portfolio

PROFILE

- Recent Master's in Data Science graduate with over 5 years of expertise in Full stack software engineering, actively seeking challenging opportunities to leverage my skills. Highly proficient in various technologies such as Python and Java, with a strong skill set in creating ETL data pipelines.

PROFESSIONAL EXPERIENCE

Boost Sport AI, San Francisco | ***Data Engineer - Internship*** **Nov 2022 - Jun 2023**

- Implemented a scalable ETL pipeline using Python, Spark, and Apache Airflow to fetch sports data from APIs and store it in PostgreSQL database, resulting in a 30% reduction in data errors.
- Built a Natural Language Generation (NLG) system using directed acyclic graphs (DAGs) to transform various API data feeds into auto generated, real-time game recaps for the English Premier League.
- Developed an accurate sports prediction model using the ELO ratings algorithm, achieving an 83% accuracy in forecasting win probabilities for teams across multiple leagues, including EPL, NBA, and NFL.
- Created comprehensive test cases using Python's unittest framework to achieve 100% code coverage ensuring reliability of critical application components.

Philips Healthcare, Baltimore | ***Senior Software Engineer - Full Stack*** **Jul 2018 - May 2022**

- Led a team of 3 developers in engineering a Python application using Flask and RabbitMQ, that streamlined the onboarding process of new customers, reducing onboarding errors by 50%.
- Designed a highly scalable database schema on PostgreSQL for a medical platform called ROCC, reducing query times by 40%.
- Designed and developed secure REST APIs using Java SpringBoot and Python, connecting the ROCC user interface with backend infrastructure; reducing API response time by 25%.
- Implemented the entire frontend application for ROCC using React.JS, Typescript, Redux and Node.JS, enabling real-time communication between radiologists and technologists.
- Deployed over 5 microservices on Kubernetes using Docker containers, ensuring high availability.
- Wrote fast and efficient queries using GraphQL API, reducing the application load time by 24% and patient connect time from 8 seconds to 3 seconds.
- Deployed a Flask application on an AWS EC2 instance, ensuring seamless access to the application. Configured Nginx as a reverse proxy to enhance performance, scalability, and security.

EDUCATION

MS in Data Science, University of San Francisco **Jul 2022 - Jun 2023**

Relevant Coursework: Advanced Machine Learning, MLOps, Distributed Computing (Apache Spark), NoSQL, Data Acquisition, Linear Regression, A/B Testing, Data Engineering, Data Modeling, NLP, Probability and Statistics

BS in Computer Science, Manipal Institute of Technology **Jul 2014 - May 2018**

Relevant Coursework: Data Structures & Algorithms, Database Management Systems (SQL), Object Oriented Programming, Operating Systems, Artificial Intelligence, Software Designing

ACADEMIC PROJECTS

Only-Stats (Sports Analytics startup) Website | Github Repo

- Led a team of 5 engineers in successfully developing an MVP for a sports analytics startup using React, Flask, MongoDB and machine learning, providing data-driven insights to sports teams and fantasy league players.
- Selected among the top 3 performing startups in the cohort and presented the product to venture capitalists

E-Commerce Product Recommender for ASOS (Github Repo)

- Built a scalable ETL pipeline using Airflow, Spark, SparkML and MongoDB to handle large-scale data and process it into rich word embeddings using NLP (BERT) and recommend top 10 similar products.

SKILLS

Programming Languages:	Python, Scala, Java, PostgreSQL, REST API, Javascript, React, Bash, Shell Script
Technologies:	Flask, MongoDB, Redis, Docker, Kubernetes, AWS S3, EC2, Git, Google Cloud Platform (GCP), CI/CD, Node.JS, Jenkins, SpringBoot, RabbitMQ, GraphQL
Libraries:	Pandas, Numpy, Matplotlib, Scikit-learn, PyTorch, TensorFlow, NLTK
Big Data Tools:	Hadoop, Kafka, Airflow, Elasticsearch, Spark, Flink, Hive, Snowflake
Machine Learning:	Regression, Decision Trees, Random Forest, Boosting, MLFlow, DVC, Metaflow