**ADIT SHRIMAL**

[adit.p.shrimal@gmail.com](mailto:adit.p.shrimal@gmail.com) | +1 (408) 203 2174 | San Francisco, CA, USA | [www.aditshrimal.com](http://www.aditshrimal.com) | [LinkedIn](http://www.linkedin.com/in/aditshrimal) | [Github](https://github.com/aditpshrimal)

**EDUCATION**

**M.S. in Data Science, University of San Francisco Jul 2022 - Present**

Relevant Coursework: Machine Learning, Linear Regression, A/B Testing, Statistical Modeling, Distributed Computing (Spark), Relational Databases (SQL), NoSQL, MLOps, Data Engineering (Kafka, Hadoop, HBase, Hive, Spark, ETL, Schema design and Dimensional Data Modeling), Deep Learning

**B.E. in Computer Engineering, University of Mumbai Jul 2015 - May 2018**

Relevant Coursework: Data Structures & Algorithms, Computer Networks, Machine Learning

**PROFESSIONAL EXPERIENCE**

**Amazon Web Services Nov 2022 - Present**

*Machine Learning Intern*

* Designed and implemented Recurrent Neural Networks (**RNNs**), specifically the **GRU4Rec** model using **PyTorch** and the **ReChorus** library, achieving a notable **NDCG@10** of **0.42**.
* Utilized **SHAP** and **KernelSHAP** to explain model predictions and identify features with the highest impact on recommendations, improving **model interpretability**.
* Integrated **TimeSHAP** to provide time-based explanations of recommendations, enabling users to understand better why specific items are being recommended at a given time.

**Fractal Analytics Nov 2021 - Jun 2022**

*Data Engineer*

* Designed **data-driven solutions** for client data warehouses using **GCP** (Google Cloud Platform) products (Airflow, Dataflow, BigQuery, Cloud Storage), reducing data processing timeby **30%**.
* Implemented an end-to-end **data encryption strategy** with Google Tink and KMS, reducing security incidents by **80%** and ensuring compliance with data protection regulations.
* Won **Star Award 2022** for delivering outstanding solutions and solving complex technical challenges.

**Bewakoof Brands Pvt. Ltd. Feb 2020 - Nov 2021**

*Data Scientist*

* Worked within an **Agile** framework to construct a cost-efficient **GCP** data pipeline (**40%** cost reduction) with improved scalability and fault tolerance.
* Devised a **product ranking** auto-sort algorithm, boosting CTR (click-through rate) by **26%** and A2C (add-to-cart rate) by **12%**.
* Optimized personalized recommendations employing **Alternating Least Squares** (used implicit ratings), resulting in a significant **14%** increase in RPTI (revenue-per-thousand impressions).
* Recognized as **Employee of the Month** for October ‘20 and **Employee of the Quarter** for Q4 FY 20-21.

**Reliance Jio Jul 2018 - Feb 2020**

*Data Engineer*

* Built a **streaming data pipeline** using Kafka, Flink & Elasticsearch, handling **~150k/sec** events in real-time with zero data lag and outperforming existing pipelines with a data lag of two hours.
* Utilized **Agile** methodologies to construct a **data warehouse** with HDFS, Spark, and Cassandra, enabling systematic computational analysis for Product Managers and Leaders.
* Conducted a successful Proof of Concept (PoC) for deploying analytics pipelines on **Kubernetes** for two Jio media applications, demonstrating the feasibility and efficiency of containerized data processing.

**TECHNICAL SKILLS**

* **Programming Languages**: Python, Java, Bash
* **Databases**: MySQL, MongoDB, Cassandra, Elasticsearch
* **Big Data & Cloud**: Kafka, Spark, Flink, HDFS, Airflow, GCP (Pub/Sub, BigQuery, Dataflow, Key Management Service, Cloud Functions), AWS (Kinesis, S3, Lambda, Redshift, Glue, Athena)
* **Machine Learning**: Linear Regression, Logistic Regression, Decision Trees, Random Forest, Kmeans, Gradient Boosting (XGBoost), SVD, PCA, Collaborative Filtering (ALS), Feature Engineering, and Model Evaluation, PyTorch
* **MLOps**: MLFlow, Data Version Control (DVC), Git, CI/CD, Docker, Kubernetes, API design (Flask)
* **Data Visualization**: Tools - Tableau, Metabase, Kibana, Libraries - Matplotlib, Seaborn