Sarthak Kothari

sarthakkothari.me | Austin, TX | sarthakkothari@hotmail.com | +1 (617) 840-3450

EDUCATION

Northeastern University, Boston, MA Sept 2017 – Dec 2019
College of Computer and Information Science
Candidate of Master of Science in Data Science
Relevant Courses: Parallel Data Processing, Algorithms, Data Mining.

K. J. Somaiya College of Engineering, Mumbai, India Aug 2013 – Jun 2016
Secured a Bachelor of Engineering in Computer Science.

WORK EXPERIENCE

ClosedLoop.ai, Austin, TX Feb 2020 – Present
Software Engineer.

- Designed, developed, and deployed a scalable microservice using docker on AWS EC2 instances. This service is responsible for managing models, triggering predictions, and generating reports for new customer data using Scala.
- Redesigned automatic report generator to include predictions from multiple models by tracking model completion via MongoDB & AWS SQS, significantly increasing platform's reporting efficiency (≥ 50%) & reducing customer overheads.
- Established a core library for streamed API calls that standardized authentication, logged API usage, and internally notified developers of errors thus improving efficiency, bug detection and system logging.
- Implemented feature drift tracking by calculating population stability index between model populations in Python & Scala helping customer detect stale models and re-train them for better accuracy.
- Boosting platform's model training & testing time by 25% by parallelizing data preparation leveraging AWS SQS.
- Designed to platform's UI pages that enabled customers to monitor as well as trigger models & reports using React.

Center for Complex Network Research - Northeastern University, Boston, MA Jun 2019 – Dec 2019
Data Science Research Assistant.

- Developed high dimensional embeddings for more than 3 million authors and 1 million publications to quantify the impact of an author's presence in a paper from a network graph of scientific publications.
- Engineered metapath2vec approach in TensorFlow for generating the high dimensional embeddings.

Staples Inc., Framingham, MA Jan 2019 – Aug 2019
Data Science Intern – Operations.

- Improved weekly demand forecasting by 23% for 550,000+ SKU's, by developing an ensemble of weak models.
- Visualized a webpage showing Staple's delivery footprint to recognize key areas of interest with Python and D3.js.

Fidelis Cybersecurity, Bethesda, MD Jul 2018 – Aug 2018
Software Engineer – Data Intern.

- Increased data retrieval efficiency by 40% via migrating to Neo4j, a graph database, from MongoDB and Spark with Python.
- Executed clustering analysis on malware alert data by harnessing the intelligence obtained through analyzing relationships between nodes in Neo4j.

Hansa Cequity, Mumbai, India Aug 2016 – Aug 2017
Associate Analyst

- Modeled a logistic regression for lead scoring to rate a lead's likelihood of becoming a customer with an accuracy of 90%.
- Developed efficient data ingestion pipelines using Python and SQL for 100K+ records and build descriptive dashboards in SSRS enabling clients to evaluate product and campaign performance, boosting sales by 15%.

ACADEMIC PROJECTS

DeepDrug Repurposing, Northeastern University, Boston, MA Sept 2019 – Dec 2019

- Predicted the probability of interaction between drug molecule and target protein with WideDTA based neural network.
- Deployed the model using a web-based application with Flask, HTML and Python.

Matrix Multiplication in Distributed Environment, Northeastern University, Boston, MA Nov 2018 – Dec 2018

- Executed a comparative study between Cannon's vs Simple Block Partitioning for Distributed Matrix Multiplication algorithm.
- Implemented the algorithms in MapReduce & benchmarked them on AWS EMR cluster on matrices of different sizes, with Canon's algorithm outperforming Simple Block Partitioning.

TECHNICAL KNOWLEDGE

Statistical Programming Languages : Scala, Python, R, Java, JavaScript, React, etc.
Database tools & technologies : Spark, MongoDB, Hadoop, Neo4j, SQL Server, etc.
Tools, frameworks & Cloud Technologies : Docker, Jenkins, Lagom, Play, AWS, GCP, Eclipse, Git, Jira.