Tarun Thomas Eapen

414-397-2599 | eapentarunthomas@gmail.com | LinkedIn | Github | Persoanl Website

Education

University of Wisconsin-Milwaukee — Milwaukee, WI
Master of Science in Computer Science — *Sept. 2022 – May 2024*

College of Engineering Trivandrum (CET) — Trivandrum, India
Bachelor of Technology in Computer Science — *July 2012 – June 2016*

Experience

Software Engineer — July 2016 – May 2022
SAP Labs India (Analytics Cloud) — *Bangalore, India*

- Implemented FastText model to improve retrieval of articles in component-issue search used by SAP consultants.
- Worked alongside research scientists in redesigned the big data analytics pipelines to make them schema-less, migrated from Teradata to separate the compute(**Spark**) and storage(**AWS S3**) to scale them independently which led to a **90% reduction** in operational costs.
- Created dashboards to serve the reporting needs for three business areas: Invoicing, Inventory and Accounts.

SAP Labs India (Business ByDesign) — *Bangalore, India*

- Designed, and developed reusable and robust **REST APIs** using **Java** and **Spring Boot** for Indian and South Korean markets to support tax and invoicing requirements, **increasing adoption in these markets by 25%**.
- Designed and developed **OData/REST microservices** for updating transactional data in real-time which **increased invoicing throughput for customers by 45**%.
- Collaborated with cross-functional teams to identify and implement modifications to the **database design** and optimized performance of **database queries** by leveraging capabilities of SAP HANA query execution engine resulting in an average **45% reduction in response time** of web-based business applications.
- Designed, developed, and maintained enterprise cloud applications and API gateways for the engineering change management module.

Projects

New York Taxi Fare Prediction | *Python, TensorFlow, pandas, scikit-learn, NumPy, XGBoost* — Dec. 2022

- **Performed exploratory data analysis** and modelling to predict New York city taxi fares using NYC Taxi fare and weather datasets | *seaborn, mathplotlib, statsmodels*
- Performed a grid-search across numerous Machine Learning models (LR, DT, RF, XGBoost) and data features to identify XGBoost with hyperparameter tuning as the winning model with an accuracy of 98.6%

Extracting context from Twitter hashtags | *Python, TensorFlow, Keras, Twitter API* — Oct. 2022

- Developed a ML pipeline to extract conversational context from Twitter hashtags using pretrained BERT models.
- Pretrained BERT models for Named Entity Recognition and Sentiment Analysis were trained using tweets collected by developing TweePy **Python** script to access Twitter API.
- Created an evaluation against a similar pipeline created using Stanford Stanza's NER and Sentiment Analysis models, to provide a comprehensive evaluation and comparison of the performance of the two models.

Cognitive Popularity Tracking using Sentiment Analysis | *Java, NodeJS, TensorFlow* — April 2016

- Mined opinions of celebrities from social media and news outlets, calculating sentiment and attributing it to corresponding entities. Used number of mentions as an index of exposure and calculated sentiment as an index of popularity.
- Sentiment analyzer and named entity recognition models from Stanford CoreNLP package were used and coded using **Java**.
- Developed a frontend using a **NodeJS** application, which uses Koa as the server framework and provides a **REST API** service to query the Elasticsearch database. This allowed for easy access and visualization of the mined data for further analysis.

Technical Skills

Programming Languages: Java, Python, C/C++, Scala, SQL(Postgres), JavaScript
Frameworks & Tools: Spring Boot, Django, AWS(S3, EC2, EMR, DynamoDB, SWF), Microsoft Azure, Maven, NodeJS, React, Angular, Flask, Spark, MongoDB, REST, OData, Docker
Libraries: PyTorch, pandas, NumPy, Matplotlib, TensorFlow, Keras, scikit-learn
Courses Taken: Machine Learning, Natural Language Processing, Deep Learning, Data Science, Data Structures and Algorithm Design, Data Mining, Cloud Computing