

Aritra Kumar Lahiri

Toronto, ON M6M 0B9, CA

Ph: +1 5142102763; E-mail: aritra.lahiri@torontomu.ca

LinkedIn: <https://www.linkedin.com/in/aritrakumarlalhari/> GitHub: <https://github.com/Aritra23>

Portfolio: <https://aritra23.github.io/>

PROFESSIONAL EXPERIENCE

Software Engineer II, TD Bank, Toronto, ON (June 2022 – May 2024)

- Implemented payment and digital credit offer API using Springboot, Java and IBM DB2 database server for enhancing TD Easy Web offer acceptance workflow, one of its primary user enablement products.
- Spearheaded a group of developers to establish the API framework standards across the bank in Java and NodeJS technologies.
- Achieved success in developing the end-to-end API developer portal solution using GCP Apigee policies. Utilized Python scripts for designing the CI/CD pipelines.

Software Engineer, Ford Motors, Detroit, MI (May 2019 – September 2020)

- Implemented Rest Api using Spring Data JPA in Java to query Hive tables.
- Productionized application with IIOT platform saving huge business cost, won Tech award for innovation.
- Optimized batch jobs with Spring scheduler, integrated Python ML model with Alteryx workflow utilizing Rest Api and designed micro-service notifying callback response with front end. Deployed containerized application using Docker.
- Tools - Angular 6, Springboot in Java at backend, MS SQL server, PCF, AWS S3, LDAP and Email as micro-services.

Software Development Engineer, Pearson, Chandler, AZ June 2016 – May 2019

- Implemented Service Oriented Applications using web services with XML, SOAP, WSDL, JAX-WS. Developed REST APIs utilizing Spring and Hibernate for persistence. Fixed integration endpoint mapping following open API spec in Swagger, consumed REST APIs using Swagger UI. Verified the Web Service response using Apache Axis, SOAP UI.
- Designed and implemented Spring Boot Microservices for search and indexing data into backend MongoDB. Extensively used Java Collections API like Map, Set, Lists, Java concurrency package for multithreaded application calls.
- Used JUnit for TDD integrated with Ant, Jenkins, improved unit test coverage by 80 %, collaborated through Bitbucket.

INTERNSHIP

Web Developer Intern, Arizona State University. Scottsdale, AZ May 2015 – August 2015

- Prototyped a health analytics recommender system, that maps disease risk factors with biosensor device features and generate a mapping algorithm to prescribe devices based on patient profile.
- Designed MVC architecture using Angular JS at front end, Node JS Server, Express JS for connecting with the DB at back end and MongoDB (NoSQL) database. Installed Ruby for running Sass CSS, Compass, Git for version control.

ACADEMIC & RESEARCH PROJECTS

Title: GameOfThronesQA: Answer-Aware Question-Answer Pairs for TV Series

Publication: April 2022

Project Summary: Presented a corpus of QA dataset GameofThronesQA, using Answer Aware Question Answering techniques and provided a novel pipeline framework for Answer Aware question Generation.

Tools & Technology: LLMs like fine-tuned variant of BERT, Python.

Title: Auto-Answer Aware QA generator

Publication: October 2022

Project Summary: Proposed an automated Question-Answer (QA) pair generation tool Auto-AAQA by leveraging transfer learning techniques to simplify the question-answer generation. The tool utilizes the Answer Aware Question Generation (AAQG) method using the fine-tuned transformer-based model and task-based prefixes to generate high-quality QA pairs. The tool is developed as a Python and Flask-based web application that provides a user interface to visualize the QA pair generation task and serves the client request by rendering the final output via API calls. The application is hosted on DigitalOcean with Platform-as-a-Service (PaaS) support.

Tools & Technology: AAQG model (fine-tuned BERT-HLSQG), Python, Flask.

Title: OTTQA V1.0 – Entity level QA dataset for sentiment analysis

Publication: November 2022

Project Summary: Implemented an entity level QA dataset OTTQA V1.0 containing primary TV series character opinions along with their sentiment polarities from the tweets generated from answer span extraction. Utilized supportive tweets by extracting them according to their relevancy with answer span keyword which is used to gauge opinion changes of OTT series characters over a given period.

Tools & Technology: Python, Anaconda, Pandas, Tweepy API, Textblob, Scikit learn ML library, NLTK, TF-IDF vectorizer.

Title: TREC Clinical Trials Track

Publication: November 2023

Project Summary: Retrieval of most relevant clinical trials from PubMed Data based on provided topic tracks utilizing natural language processing techniques and neural language models. Performed data preparation tasks using PubMed Parser for article extraction and combine cleaning and data preprocessing for model input for information retrieval. Employed Sentence Transformer and Doc2Vec models for feature extraction and then reranked the articles based on Cosine similarity scores.

Tools & Technology: Python, SentenceTransformer, Doc2Vec, NLTK, TF-IDF vectorizer, NDCG

Title: Implementation of Data Partitioning techniques, Query processing operations

Duration: August 2015 – December 2015

Location: Arizona State University

Project Summary: Analysed Movie-Lens Dataset with 10M records to implement range partitioning and round – robin partitioning. Implemented parallel sort and parallel join operations on partitioned tables.

Tools & Technology: Python and PostgreSQL

INDEPENDENT PROJECTS

Title: Sentiment Analysis on Game of Thrones Season 8 Tweets

Duration: May 2019

Team Size: 3

Project Summary: Analysed sentiments of different characters by calculating polarity scores and predicted season outcome by applying logistic regression over the model dataset sample using train test split.

Tools & Technology: Python, Anaconda, Pandas, Tweepy API, Textblob, Scikit learn ML library, NLTK, TF-IDF vectorizer.

Title: Spotify Data Visualization

Duration: March 2019

Team Size: 3

Project Summary: Analyzed Spotify Million Playlist Dataset to extract and visualize top artists and tracks by each region.

Tools & Technology: JVectorMaps(native JS library), Python, Pandas, Jupyter Notebook, Plotly.JS.

Title: Mission to Mars

Duration: January 2019

Team Size: 1

Project Summary: Implemented Flask App to render scraped Mars related data from NASA website and display the information in a single page application.

Tools & Technology: Jupyter Notebook, BeautifulSoup, Pandas, Splinter, PyMongo and Flask.

Title: DOM Manipulation using UFO Sighting Dataset

Duration: January 2019.

Team Size: 1

Project Summary: Manipulate Document Object Model to filter dataset based single or multiple search categories.

Tools & Technology: Python, JS.

Title: Data Journalism using D3

Duration: January 2019

Team Size: 1

Project Summary: Interactive Data Visualization using D3.JS on U.S. Census Bureau and the Behavioral Risk Factor Surveillance System dataset.

Tools & Technology: D3.JS

Title: Inspection Ratings of Chicago Restaurants

Duration: January 2019

Team Size: 3

Project Summary: ETL workflow for calculating food inspection ratings on Chicago Restaurants.

Tools & Technology: Pandas for data munging, SQL for storing transformed data and Python, Mongo and Flask App for the load, Kaggle Dataset source.

Title: Climate Analysis and Exploration

Duration: December 2018

Team Size: 1

Project Summary: Analysis of Weather across 500+ cities comprising of different parameters like cloudiness, humidity, max temperature, and wind speed.

Tools & Technology: Python, SQL Alchemy ORM queries, Pandas, Matplotlib and Open Weather Map API.

Title: Analyse video game data and provide prediction for purchasing trends

Duration: December 2018

Team Size: 1

Project Summary: Analyze the data for a recent fantasy game Heroes of Pymoli and generated a report that breaks down the game's purchasing data into meaningful insights which includes Player Count, Purchasing Analysis, Gender Demographics, Purchasing Analysis based on Gender, Age Demographics, Top Spenders, Most Popular Items, Most Profitable Items

Tools & Technology: Python, Anaconda, NumPy, Pandas, Matplotlib, CSV

Title: Mini Messaging Service (Twitter)

Duration: December 2017

Team Size: 1

Project Summary: Implemented a backend RESTful API of a mini message service like Twitter supporting features like view, write, search tweets, follow and unfollow users as well as list of all users following the current user and their most popular follower.

Tools & Technology: Spring Boot, IntelliJ IDE, Java, H2 database

ACADEMIC QUALIFICATION

PhD in Computer Science Toronto Metropolitan University, Toronto, ON, CA, Expected Completion – May 2025

MS in Computer Science, Arizona State University, Tempe, AZ, USA, Year of Completion - 2016

BTech in Computer Science & Engineering, Meghnad Saha Institute of Technology, WBUT, Kolkata, India, Year of Completion - 2013

PUBLICATIONS

- Lahiri, Aritra Kumar, and Qinmin Vivian Hu. "GameOfThronesQA: Answer-aware question-answer pairs for tv series." European Conference on Information Retrieval. Cham: Springer International Publishing, 2022.
- Lahiri, Aritra Kumar, Emrul Hasan, Qinmin Vivian Hu, and Cherie Ding. "TMU at TREC Clinical Trials Track 2023."
- Lahiri, A. K., & Hu, Q. V. (2022, November). Entity Level QA Pairs Dataset for Sentiment Analysis. In *2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (pp. 270-276). IEEE.
- Lahiri, Aritra Kumar, and Qinmin Vivian Hu. "Named Entity-based Question-Answering Pair Generator." In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 4902-4906. 2022.
- A Game Theoretic Approach to Demand Side Management in Smart Grid with Multiple Energy Sources and Storage. *International Journal of Advanced Computer Science and Applications (IJACSA)* , *The Science and Information (SAI) Organization* [Volume 9, Issue 2, February 2018](#)

TRAINING & CERTIFICATIONS

- Data Analytics from University of Arizona, Tucson, Phoenix, 2018.
- Machine Learning from Coursera 2018