Bidisha Das Baksi

424-312-2673 | Los Angeles, CA 90007 | baksi@usc.edu | https://www.linkedin.com/in/bidishadasbaksi/ | github.com/Bidisha010496

EDUCATION

Master of Science, Computer Science — December 2022
University of Southern California , Los Angeles , CA — 3.8 GPA

Bachelor of Engineering in Computer Science and Engineering — May 2018
The National Institute of Engineering, Mysore — 9.56/10.00 CGPA

TECHNICAL SKILLS

Programming Languages : Java, Python, C++, SQL.
Databases : MySql , Vertica Columnar Database (vSql), Distributed Database Systems.
Framework, Libraries and Tools : Machine Learning, Natural Language Processing, Hadoop MapReduce, AWS, Big Data, QlikSense, Numpy ,PyTorch,Pandas, Scikit-learn, Jenkins, Git, REST API, Spring MVC, React JS, Flume, Guice.

WORK EXPERIENCE

SWE Intern , Google — May 2022 - Present

- Enhanced update icon API to implement icon removal feature on the OAuth consent page in Pantheon.
- Developed Flume Data Processing Pipeline by leveraging data from token dump and OAuth service logs to build statistics reporting tool on usage of OAuth application by client apps. (Technology : Flume , Java)

Information Science Institute, USC, USA - Graduate Research Assistant — August 2021-Present

- Managing and contributing to open source software Web Karma, a GUI based data integration tool to combine heterogeneous datasets using ontologies. (Technology : Java , JSP)
- Implemented and integrated domain independent automatic semantic labelling feature. (Technology : NLP, Java)

Intuit, Bangalore - Software Engineer 2, Full Stack Engineer — January 2020-April 2021

- Designed and Engineered system workflow for CRM tool to generate critical real-time feedbacks, improved team management efficiency by 80% . (Technology : Java , Apache Beam, Drools API , Qliksense Analytics)
- Developed parallel processing API to process up to 1000 transactions in a single request, slashed latency to less than 1 minute and accomplished an overall performance gain of 30 minutes. (Technology : Java , SpringBoot)

Intuit, Bangalore - Software Engineer 1, Big Data — July 2018-December 2019

- Collaborated with team and migrated 70+ data pipelines from on-premise infrastructure to AWS cloud.
- Redesigned business-critical data pipeline to process over 50 GB of daily transaction data in batches with a performance gain of ~ 40 minutes. (Technology: Apache Spark, Apache Kafka, Hive, AWS S3, Vertica)
- Designed, and headed a team of 4 to build an optimised data parity tool, achieved a scanning speed of about 90 minutes for over 400 tables and several petabytes of data. (Technology : Python, S3, Tidal Job Scheduler)
- Restructured and tuned ETL queries to save ~100 minutes of query runtime.

Morgan Stanley, Bangalore - Spring Intern, Technology Division — February 2018-June 2018

- Developed a Typeahead component for a query-based search engine. (Technology : AngularJs, Java,HTML,CSS)
- Devised and implemented syntax parser to detect errors in syntax-semantics and suggest corrections.
- Achieved reduction in average typing duration of a query from 30 seconds to 5 seconds and improved accuracy.

ACADEMIC PROJECTS

Evaluating the sensibility of Chatbots by using existing common sense knowledge graphs — Fall 2021

- Fine Tuned Language Model (BERT) to transfer common sense knowledge understanding from knowledge bases (COMET, ATOMIC) and evaluated sensibility of a turn-level chatbot conversation.

Local Market Analysis for a Successful Restaurant Yield — Final Semester, 2018

- Market Analysis of a defined area using Machine Learning to predict the growth of a newly established restaurant business in the area. Published Survey Paper : https://link.springer.com/chapter/10.1007/978-981-13-1498-8_22.
- Implemented Random Forest Regression and Multi-Output Regression models to predict success probabilities.

LEADERSHIP & INVOLVEMENT

- Member of Technical Development Leadership Team of GRIDS USC.
- Best Project Award in GED Hackathon Intuit (2019) , presented by Marianna Tessel (Intuit CTO).