ANKAN MAZUMDAR

San Jose, CA|+1-8726642297 | [ankanmazumdarofficial@gmail.com](mailto:ankanmazumdarofficial@gmail.com) | <https://www.linkedin.com/in/ankanmazumdar/>

# SUMMARY

Data Engineer with over six years of experience with proficiency in Data analysis, Data Engineering, ETL, cloud technologies, and IT DevOps. Expertise in collecting, transforming, and organizing data to enable informed decision-making through data analysis processes and researching projects related to AI, ML, NLP, LLMs, Computer vision, Recommendation systems.

# SKILLS

**ML frameworks and libraries**: TensorFlow, Keras. Pandas and NumPy, Sklearn.

**Programming**: C/C++, Python, Java, Go Lang

**ETL**: Ab Initio, SSIS, Kafka, Informatica

**DBs**: SQL, and NoSQL DBs (Mongo), Linux

**Cloud:** AWS, GCP, Azure

**Apache Hadoop Framework and Technologies:** Scala, Spark, Hive.

**Reporting**: Tableau, MS Excel, Murex Trading Platform, DevOps: Jenkins, Docker. Monitoring: Control-M, Airflow, Grafana

# WORK EXPERIENCE

## System Engineer

Tata Consultancy Services October 2020 - August 2023

* Collaborated in building workflow pipelines for real-time transactions for trading application. (Murex )
* Led process enhancements to resolve data mismatches, system defects, and functional flaws, and trained team members to enhance overall productivity. Conducted RCA to remediate system vulnerabilities, resulting in a 15% reduction in system issues.
* Delegated and led team in client reporting, gathering business requirements and ensuring timely project delivery, highlighting effective communication and leadership abilities.
* Designed dashboards in Tableau reducing data reporting anomalies by 20%.

## ETL Developer

Bitwise Solutions June 2017 - July 2020

* Accomplished data migration project from On-premise to Google Cloud using Ab Initio, Hadoop, Java, Unix, SQL, and GCP services, showcasing technical expertise. Completed the project 1 month ahead of schedule.
* Implemented data security layers using Encryption, Redaction, and Tokenization concepts, emphasizing attention to data security and risk assessment. No data security incidents occurred related to it after that.
* Developed a housekeeping script to optimize server space and memory usage, showcasing problem-solving skills, led to 20% reduction in maintenance costs.
* Optimized SQL queries for quicker real-time report generation, resulted in 40% reduction in report generation time.

Research Journals:

* IJARESM paper on Automated Image Labeling through active and transfer learning techniques: developed an image predicting UI,improved accuracy and confidence score by 12% after manual labeling and retrained model. <http://www.ijaresm.com/uploaded_files/document_file/Ankan_MazumdarubDT.pdf>

<https://github.com/ankan-mazumdar/Active-Learning>

Academic Projects:

* Stock Market ETL, Analytics and Predictions and LLM:

<https://github.com/ankan-mazumdar/StockMarketAnalytics-Prediction/blob/main/README.md> For ETL, used Python, Amazon Web Services (AWS), Apache Kafka, Glue, Athena, and SQL.

For Prediction, Compared performance of statistical and deep learning models for stock price prediction, found LSTM model which outperforms ARIMA, moving average with MAPE of 1.1%.

For Analytics and dashboard, used Pandas, Finance API, and Streanmlit For LLM chatbot, used RAG based openAI API .

* Harmful-website-Detector using active learning: [https://github.com/ankan-mazumdar/Harmful-website-Detector](https://github.com/ankan-mazumdar/Harmful-website-Detector%20Crafted%20a%20harmful-website-detector%20) [Crafted a harmful-website-detector](https://github.com/ankan-mazumdar/Harmful-website-Detector%20Crafted%20a%20harmful-website-detector%20) using active learning, achieved 95% accuracy with Log. Regression and SGD model, to improve harmful website detection by reducing error rate.
* Amazon-Food-reviews-sentiment-Analysis:

<https://github.com/ankan-mazumdar/Amazon-Food-reviews-sentiment-Analysis> Compared performance of Logistic Regression models with Bag-of-Words, TF-IDF, and Word2Vec features on Amazon food reviews dataset, found LR(l2|GridSearchCV) for TF-IDF achieves the best test accuracy of 92.29%. Deployed it on AWS.

# EDUCATION

Illinois Institute of Technology August 2023-Present Masters in Artificial Intelligence

University of Hyderabad, India November 2022

Post Graduation Diploma in Artificial Intelligence-Machine Learning, India (completed degree with distinction grade).

Netaji Subhash Engineering College, Kolkata, India June 2017

Bachelors in Technology in Computer Science and Engineering

# Recognitions

* TCS Gems for Collaborations & Contributing Outstanding Team Performance.