Tamajyoti Dhar

Results-oriented Data Engineer with **10+ years** of experience and a proven track record of designing and implementing scalable **data and machine learning solutions**.Experienced in **leading cross-functional teams** and **setting up analytics functions** from scratch to drive data-driven decision-making. Skilled in **data architecture, ETL development, and developing analytics frameworks,** with a strong focus on delivering high-quality actionable insights to drive business decisions. Seeking to leverage the technical and leadership experience to drive innovation and business growth through data driven strategies

+91-9008054298 (**Whatsapp**)
tamajyoti.dhar@gmail.com
https://www.linkedin.com/in/tamajyoti-dhar-6494022b/

EXPERIENCE

Freelance — *Senior Data Engineer*:OCT 2021 – PRESENT

- **Lead a team** of 4 data analysts and scientists for a gametech startup building **analytics and experimentation framework** along with the end to end **data flow architecture**
- Developed the **data collection, ingestion and delivery system** to deliver artifacts for cross functional teams to aid them with insights for decision making and **KPI monitoring**.The project involved ingestion and warehousing of data along with subsequent data transformation for serving downstream users. The techstack for the project involved *Kinesis Firehose, S3, Airflow, Lambda, Glue, EMR, Python, DBT Core, Redshift and Tableau*
- Developed end to end **near real time alerting mechanism** and player behavior metrics to identify game play. The techstack for the project involved *Kinesis Firehose, S3, Lambda, Kinesis Data Stream, Apache Flink (KDA)* improving event handling efficiency by **30%**.
- Designed and developed the entire **single source of truth data architecture** for a Web3 startup involving end to end data ingestion from multiple external and internal sources and large scale data transformation both at ingestion side and data warehouse side. The tech stack used involved *Kafka, Spark Streaming EMR Jobs, S3, Airflow, Python, Fivetran, Snowpipe, Snowflake, DBT Core, Hex analytics*
- Designed, trained, and deployed an LLM for cybersecurity threat detection and classification, **fine-tuning models like GPT-3 and BERT** on domain-specific data. Applied NLP techniques, including **Named Entity Recognition and token classification**, to identify malicious entities and analyze threat intelligence effectively
- Developed data pipelines capable of processing over **TBs of threat data per day**, scouting data from sources like security logs, cloud storages etc..
- Deployed the fine-tuned LLM model on **Kubernetes using PyTorch**, optimizing inference times by 25%**.** Integrated the model within existing cybersecurity monitoring tools, achieving scalable, real-time threat intelligence access through containerized microservices.

Tech Stack: Python, Redshift, MongoDB, Snowflake, DBT Core, Kinesis, Kafka, Lambda, Glue, Pyspark, Airflow, AWS Services, Pytorch, Terraform, BERT

Complyadvantage, London — *Senior Data Engineer*:JULY 2019 – SEP 2021

- Built an end to end pipeline to consume ~**100K records** hourly from Kafka stream and transform them using **Spark** and ingesting them in **Elasticsearch** for downstream recommendation
- Created the entire **reporting layer** leading a team of 3 data scientists to reduce analytics effort on ad-hoc requests and also built models for **lead scoring, user usage pattern, recommendation tracking**
- Developed **Rest APIs** for information extraction from raw texts using **multilingual and multiclass classification models,** built using language models like **BERT** alongwith model integration in the main data pipeline
- Driving the ideation and solutioning of **entity resolution** as part org wide risk information graph for resolving context of entities using Graph Frameworks
- Developed and deployed **ML models** to identify information of interest using TFIDF, Vector Embeddings and supervised classification techniques in **PySpark**

Tech Stack: Python,Flask, Swagger, PySpark, Sparql, Elastic Search, AWS, Pytorch, OWL, Docker, Kubernetes

AREAS OF EXPERTISE

ETL/ELT, Data Lakes, Data Warehousing, Schema Design, Data Modelling, API Development

Natural Language Processing, Graph Learning & Analytics. Knowledge Graphs, Ontology, Semantic Web Stack

A/B Testing Framework, Churn Modelling, Pattern Mining, Product Analytics, KPI Design, Data Compliance & Governance

SKILLS

Programming: Python, Scala, R, Spark, VBA.

ETL Tools: Airflow, Fivetran, Helpshift

Query Language: SQL, Presto, Hive, SPARQL

Databases: Elastic Search, MongoDB, Snowflake, Redshift, PostGres, MySQL, Teradata, Druid

Frameworks:DBT Core, Pytorch, Flask, Swagger, Spark, Kafka

Visualization: Tableau, Looker, Periscope, Metabase, D3.js

DevOps : Git, Docker, Terraform, Helm, Kubernetes, CI/CD, AWS,GCP

Bluejeans, Bangalore — *Data Scientist*: MAY 2018 – JUNE 2019
- Set the **strategy for data science** function with the CTO's office and **lead a team** of 3 to develop and deliver data projects across organization along with mentoring and guiding them
- Built a **fraud detection** algorithm to detect scheduling fraud using text similarity and user pattern mining
- Built a **load forecasting** algorithm to identify anomalous load across the various POPs (point of presence) on a 30 min sliding window with an accuracy of ~82%

Tech Stack: Python, Redshift, MongoDB, Kafka, MemSQL,Periscope

Belong.co, Bangalore — *Lead Analytics Engineer*:DECEMBER 2015 – APRIL 2018
- **Built the analytics function** from scratch designing and developing ETL, KPI's , A/B test frameworks for internal teams along with **leading a team** of 4.
- Built a **recommendation system** to match profiles of people to the jobs with the best match after learning hiring trends using **collaborative filtering**
- Developed **customer churn modeling** framework to predict churn with almost ~85% accuracy along with monitor key **company metrics** to identify areas of concern. Created an automated alerting mechanism framework based on user usage to make CSM aware of customer satisfaction.

Tech Stack: Python, Postgres, Elasticsearch, Flask, LDA, Presto, Hive, Metadata, Looker

EY, Bangalore — *Consultant*:APRIL 2015 – NOVEMBER 2015
- **Lead a team** of 4 consultants to design & deliver AML models for a US brokerage client
- **Developed user classification algorithms** to detect money laundering using clustering algorithms, neural networks, outlier and anomaly detection techniques and ensemble learning
- Improved detection rates by 71 basis points Year-on-year after implementing the new scoring layer incorporating localized scoring with 28.5% decrease in false positives

Tech Stack: Python, MySQL, R

HSBC, Bangalore — *Senior Business Analyst*:JANUARY 2014 – APRIL 2015
- Developed a framework based on user clustering and risk scoring for identifying suspicious transactions for HSBC affiliates along with Scenario Development, Threshold Tuning, Segmentation as well as conducting independent validation of AML models for Retail Banking.
- SAR Detection rates improves by 63 basis points Year-on-year after implementing the new scoring layer incorporating localized scoring

Tech Stack: Python, Teradata, SAS

Infosys, Bangalore — *Senior Associate Consultant*:JULY 2012 – JANUARY 2014
- Built a mortgage delinquency predictor for a mortgage client.
- Identified and developed algorithms for user clustering and risk of default probability

Tech Stack: R, Oracle SQL, VBA

EDUCATION

Welingkar Institute of Management, Bangalore — *PGDM*

JULY 2010 - MAY 2012

Ranked 1 in PGDM (Post Graduate Diploma in Management) program

NERIST, — *B.Tech (Electronics and Communications)*

JULY 2006 - MAY 2010

Ranked among top 5 in Bachelors engineering program