Ruben Mukherjee

Product Creator with 6 years of broad-based experience in building data-intensive applications, overcoming complex architectural and scalability issues. Proficient in predictive modeling, data processing and data mining algorithms as well as in web and scripting languages such as Python. Capable of creating, developing, testing and deploying highly adaptive diverse services to translate business and functional qualifications into substantial deliverables.

rubenmukherjee@outlook.com | +353-89-491-1524 |www.linkedin.com/in/rubenmukherjee|github.com/mukherjeeruben| Dublin, Ireland

Education

Dublin City University, Dublin, Ireland *Sept 2021- Aug 2022*
Master of Science in Computing (Artificial Intelligence)
Relevant Coursework: Data Analytics and Mining, Machine Learning, Neural Networks, Statistics, Data Visualization, Data Structure and Algorithms
Cochin University of Science and Technology, **Kochi, Kerala, India** *Aug 2013-May 2017*
Bachelor of Technology in Information Technology
Relevant Coursework: Data Structures & Algorithms, Cryptography, Computer Graphics, Distributed Computing, Computer Networking.

Technical skills

Languages	**:** Python, Typescript, C#
Frameworks	**:** Flask, Django, Dash, Angular, React , .NET
Library	**:** scikit-learn, pandas
Cloud	**:** AWS, Pivotal Cloud Foundry, Azure
Database	**:** MySQL, Postgres, Influx, Amazon-DynamoDB, Mongo DB, Amazon-Redshift
Data	**:** Airflow ,Kafka, RabbitMQ, Celery, AWS DMS
Version Control	**:** Jira(Big-Bucket), GitHub, GitLab
DevOps	**:** Octopus, Azure DevOps
Exposure to	**:** Amazon S3, AWS Lambda, AWS Glue, Docker
Methodologies	**:** Agile (Scrum Model - Antipatterns)

Experiences

Retail in Motion, Dublin, Ireland *Jan 2023 – Present*
Data Engineer (Analytics)

- Building data pipeline for transforming data from legacy systems to data warehouse via modern data orchestration.
- Create and maintain data pipelines with **Airflow (MWAA).**
- Create table and task mappings with **AWS DMS (Custom Templating)**
- Consume real time data from external data source and process in **Airflow**.
- Design and create reports on **AWS Redshift** (Data Warehouse) via **S3 bucket** (Data Lake)
- Work cross-functional stakeholder for generating ad-hoc reports on **Power-Bi** and **Dash** App (**Python**)
- Create and maintain ETL data pipeline in **.NET** (**C#**). With **MsSQL, Redshift** (Data Warehouse)
- Create **DevOps** Pipeline on **Octopus** (**CICD)**

Danalto, Dublin, Ireland *Jul 2022 – Dec 2022*
Senior Software Development Engineer (Data Engineering and Analytics)

- Building interfacing applications for concept patent technologies – Internet of Things – Location based Tech with **ETL** (Data Processing)
- Maintain Data pipelines in **Airflow** (Team - Member)
- Developing and enhancing applications in **Django, Flask** and **React** Framework
- Creating custom visualization for devices Error and pression on **Dash** (Framework) - Python
- Data Pipelining/ ETL: Developing real time data ingest pipelines using **Kafka** – **RabbitMQ** and **Celery**
- Creating microservices applications in Flask (API)
- Deployment to **EC2** (AWS) instances on **Docker** via **DevOps** pipelines via **GitLab**. **CICD**
- Creating Unit Test Cases in pytest for modules

Tata Consultancy Services Limited (Digital), Mumbai, India *Jul 2021 – Nov 2021*
Data Analyst (Data Analyst for Aerospace Manufacturing Domain)

- Led the module technical team for multiple releases for enhancements resulting into value additions to the Project Management Application
- Develop algorithms using Natural Language Processing and Deep Learning models for predictive maintenance.
- Design algorithms to track and detect anomalies in multiple sensors data in an airplane.
- Develop chatbot to generate ad-hoc queries for dynamic report generation.

System Engineer (Software Developer for Aerospace Manufacturing Domain) *Nov 2017 – Jun 2021*

- Research and development team member for creating complex full-stack web and desktop business applications for a major American airplane manufacturing company
- Implemented business process architecture and design in IT for managing the manufacturing fleet services of an aircraft using latest technology stack (**Flask – Angular – React**) which improved the performance of the existing application by 70-80%
- Infrastructure lead for security and virtual workspace.

Projects

Location Based IOT Product – Danalto *Jul 2022 – Present*

Developed and Enhanced location based IOT application with Lora wan and Ultra-Wide Band range devices with product stack on Kafka, Rabbit MQ, Django and React frameworks. Creating endpoints in **Django** (**Python**) frameworks, **Python** data migration scripts from **DynamoDB** to **Influx dB**, Created Authentication with JWT in Flask (Python) API endpoints. Deployment on Amazon EC2 Instances via Gitlab Dev Ops pipeline.

- Stack: **PostgreSQL | DynamoDB | Influx DB | Django – Flask (Python) - React**
- Additional Services: **Redis Cache, Amazon S3 bucket, Kafka, RabbitMQ, Celery**
- **Cloud: CICD – Dev ops (GitLab, AWS)**
- Backend: **DynamoDB** and **Influx DB** as primary database for data fetch operations regular transactions, **PostgreSQL** for regular operations
- Server: Created exposed endpoints with authentication on **Django** web-framework with storage service as **Amazon S3 bucket**, **Redis cache** for recurrent static service calls, real-time data extraction and transformation and microservice apps for better management and real time database synchronization with auto recovery from server failovers
- Worked in Agile Methodology

Academic: Similarity of e-commerce products based on

Developed model to find product matches for a competing ecommerce platform *Jan 2022 – Mar 2022*

- Stack: **Script (Python)**
- Implementation: BERT Model with bagging and boosting on products with sentence embeddings and pre-trained CNN model based on 'clip' dataset. Similarity matrix based on Vector Space model and pre-trained BERT Sentence transformation with cosine similarity.

Academic: Analytics Application for analysis of public transport.

Designed and developed Visualization graphs for New-York taxi data (Open Source)- preprocessing with pandas *Nov 2021 – Jan 2022*

- Stack: **Dash (Python)**
- Backend: API Data Sources from NYC open data source. Created URL based queries for data retrieval.
- Application: Used pandas for data preprocessing. Plotly for data visualization on Dash and deployed on **Heroku**.

Analytics Application for parts similarity based on Image and Description

Designed and developed analytical and predictive modules to detect the age and condition of a part based on CNN *Mar 2021 – Nov 2021*

- Stack: **Flask | Dash – Script (Python)**
- Backend: SQL and API Data Sources from in-house applications.
- Application: Custom Model built on pretrained CNN with OpenCV with images of damaged and new airplane parts. Image annotation comparison based on vector space model, ranking to find the similarity match based on cosine similarity.

Dashboard Application for Ad-hoc Query Creation and Data Visualization - TCS Digital *Dec 2019 – Mar 2021*

*Developed common application platform for visualizing data from Tableau and creating ad-hoc queries based on real time selection criteria. Developed API end points in **Flask** and **Django** (**Python**), created Ad-hoc queries for report generation via background schedulers in **Flask** (**Python**). Azure Dev Ops pipeline used to deploy to Azure VM instances. Developed SyncUP code in **Python** for Live database sync.*

- Stack: **MsSQL | SAP – Hana | Teradata – Flask - Angular**
- Additional Services: **Redis Cache, Amazon S3 bucket, Pivotal Cloud Foundry**
- Backend: **MsSQL** as primary database for regular transactions, SAP-Hana and Teradata for big data fetch operations
- Server: Created exposed endpoints with authentication on flask web-framework with storage service as **Amazon S3 bucket**, **Redis cache** for recurrent static service calls, real-time file to data extraction and microservice apps for better management and real time database synchronization with auto recovery from database service failovers
- Client: Created Angular app for rendering different tableau dashboard for users on a single page with customization, Ad-Hoc real time query creation page for dynamic data, Batch scheduler for recurrent data customized on a time basis and store in personal storage bucket and real time file data extraction for old-reports
- Worked in **Agile** Methodology

Airplane Parts, Cost Estimation and Workflow Matrix Application - TCS Digital *Dec 2018 – Nov 2019*

Developed Hybrid application to trace parts and subparts, cost and time estimation with workflow allocation matrix.

- Stack: **Oracle – MVC (.NET Framework C#) - Angular**
- Architecture: MVC model (Onion Architecture)
- Backend: Created Views and Stored procedures and triggers in Oracle database to execute in a single transaction. Created batch for auto transaction based on configured time schedule
- Server: Created individual exposed API endpoints for asynchronous service calls on **MVC** with .**NET** framework
- Client: Hybrid – Server View on html and Angular client app for workflow management for better performance
- Worked in Agile Methodology
-

Airplane Aftersales Cross Team Project Management Platform - TCS Digital *Feb 2018 – Nov 2018*

*Developed Architecture and features for managing independent applications from a standalone platform. Developed API endpoints for features in frontend application in **Flask (Python).***

- Stack: **Oracle – WCF – WPF (.NET Framework C#), Microservices Application - Flask (Python)**
- Architecture: MVVM model (Onion Architecture)

- Backend: Created Views and Stored procedures and triggers in **Oracle** database to execute in a single transaction
- Server: Created individual exposed endpoints for asynchronous service calls on **WCF (.NET C#)**
- Client: Created user interfaces of forms and presentation on grid, charts and document layout
- Worked in Agile Methodology

Concept Applications

Standalone research applications for reusability of code and proof of concepts

- Developed package for **Microsoft Azure Dev-Ops Team Foundation Server** on **python-flask**
- Developed package for real-time multiple service base database synchronization on **python-flask**
- Developed sensor-based concept application to detect anomaly detection in airplane parts with python script files hosted on **Raspberry-Pi**. Wireless RF transmission implemented via HT-12E/D IC