

Sofia Dutta

Software Engineer, Tech Lead, Big Data Engineer

(443) 554-4170 | sofia.dutta17@gmail.com | <https://sofiadutta.github.io>

WORK EXPERIENCE

Tech Lead @ NewWave Telecom & Technologies, Inc.

Jan 2021 – Present

Windsor Mill, MD, USA

- Worked on the Imersis project, creating the end-to-end system architecture design, database schema design, cloud infrastructure setup and Looker data visualizations dashboards for the Imersis data quality analytics platform.
- Handled hundreds of millions of healthcare records from Centers for Medicare & Medicaid.
- Worked with unstructured customer data, and reduced data pre-processing time by a factor of ten.
- Leading data-processing efforts, guiding new employees to quickly ramp up on and start delivering on project goals.
- Built Apache Airflow data orchestrators for automated data analysis workflow for Imersis project.
- Built data analysis pipelines using Databricks data lake to further speed up analysis work.
- Built proof-of-concept solutions for customer demonstration to win project contracts on behalf of NewWave.

Data Scientist Intern @ NewWave Telecom & Technologies, Inc.

May 2020 – Dec 2020

Windsor Mill, MD, USA

- Carried out data visualization tasks using LookML, Matplotlib, and Seaborn to present quality measures based on chosen computation metrics.
- Successfully improved computation speed by 10-fold by deploying data analysis workflow in Google Cloud Platform (GCP) clusters and using Apache Spark for quality metrics computations.
- Collaborated with the Product Owner and other engineers in creating mechanisms for generating fake training data using Python programming to test out the efficacy of machine learning algorithms used in the project.
- Carried out necessary DevOps tasks for setting up Big Data Analytics environment by configuring GCP environment to execute Python programs and connected the cloud infrastructure with Looker's dashboards for delivering computed results to be presented to customers.

Graduate Student Researcher @ Ebiquity Research Lab, UMBC

Sep 2019 – May 2020

Baltimore, MD, USA

- Built [Ontology](#) for Smart Home Access Control and developed an [Android app](#) for handling context-sensitive access control in a Smart Home Environment.
- Created [YouTube videos](#) for presentation to the National Institute of Standards and Technology and published paper: "[Context Sensitive Access Control in Smart Home Environments](#)" at IEEE Big Data Security 2020 conference.

Software Engineer, Tech Lead @ Tata Consultancy Services

Nov 2010 – Feb 2018

Kolkata, India

- Led a team of developers working on design, development, and delivery management of seven different projects for various clients of TCS.
 - Created API interfaces using PL/SQL stored procedures for daily usage for clients of TCS.
-

- Carried out change based regression testing and documented software functional specifications.
- Prepared test plans and executed system integration testing and user-acceptance testing.
- Ensured client systems were up in four hours after migration activities saving millions of dollars in potential revenue lost.
- Wrote scripts for managing customer data migration task of over a billion records.
- Managed continuous integration and continuous deployment to production environments.

SKILLS

Programming	Python, Java, SQL, PL/SQL, Go
Data Science tools	PyTorch, Sci-kit Learn, Apache Spark, Keras, Tensorflow, Hive, Hadoop, Looker, LookML, OpenCV
Enterprise tools	Databricks, Google Cloud Platform, Apache Airflow, Google Dataproc, Google Compute Engine, Google Cloud Storage, Google Cloud SQL, Google Big Query, AWS S3
Back-end tools	Google BigQuery Table, Oracle Databases, PostgreSQL, Microsoft SQL Server, MongoDB, JSON
IDEs/Dev tools	Git, Jupyter Notebook, Google Colab
Domain	Big Data Analytics, Deep Learning, Machine Learning

EDUCATION

Udacity Nanodegree: Deep Learning		2021
University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA		
Master's in Data Science	GPA: 4.0	2019 – 2020
West Bengal University of Technology, Kolkata, India		
Bachelor's in Computer Science	GPA: 3.5	2006 – 2010

RELEVANT COURSEWORK

Udemy course – Mastering Go fundamentals	2021
------------------------------------------	------

FAVORITE MACHINE LEARNING PROJECTS

QABot: A Chatbot for Open Question Answering Using Neural Networks - Built “QABot”, a Chatbot using the sequence-to-sequence Deep Learning model that utilizes the Encoder Decoder Neural Network architecture combined with Attention Mechanism to respond to user search queries.

Best Deep Learning project exploring Image Processing: Image-to-Image Translation Using CycleGAN - Implemented CycleGAN for an image-to-image translation. Trained an unsupervised image translation model via the Generative Adversarial Network (GAN) architecture using unpaired collections of images from two different domains. Ref: https://sofiadutta.github.io/datascience-ipynbs/pytorch/CycleGAN_Img_Translation_PyTorch_Horse2Zebra.html