

Sofia Dutta

Data Scientist

sofia.dutta17@gmail.com | (443) 554-4170

[GitHub](#) | [LinkedIn](#) | [Personal homepage](#) | [YouTube](#) | [Scholar](#) | [Twitter](#)

WORK EXPERIENCE

Senior Data Scientist @ NewWave Telecom & Technologies, Inc. Jan 2021 – Present
Windsor Mill, MD, USA

- Performing data quality analysis on millions of healthcare records from Centers for Medicare & Medicaid.
- Built a system for computation of data quality analysis and exploration in a record time of a month.
- Trained a machine learning model using 3000 rules for quality computation metrics.
- Improved computation speed 10-fold by deploying analysis workflow in Google Cloud Platform (GCP) clusters and using Apache Spark for quality metrics computations.
- Exported results to Looker for creating intuitive dashboards visualizations.

Data Scientist Intern @ NewWave Telecom & Technologies, Inc. May 2020 – Dec 2020
Windsor Mill, MD, USA

Software Engineer @ Tata Consultancy Services Nov 2010 – Feb 2018
Kolkata, India

- Led the design, development, and delivery management of seven projects for clients of TCS.
- Created API interfaces using PL/SQL stored procedures for daily usage for clients of TCS.
- Led meetings to capture requirements from DHL UK, Staples USA, Hyatt USA, Kaiser-Permanente USA.
- Carried out change based regression analysis 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.
- Implemented scripts for data migration of a billion records while adhering to strict time SLA bounds.
- Completed client data migration from legacy Oracle Apps (11i) to Oracle ERP Suite (R12).
- From 2013 - 2018, managed continuous integration and continuous deployment in production environments.
- Got certified in seven Oracle Apps competencies while working for client projects.

SKILLS

Programming	Python, Java, SQL, PL/SQL, T-SQL
Data Science tools	PyTorch, Sci-kit Learn, PySpark, Keras, Tensorflow, Hive, Hadoop, MLlib, LookML, Azure AI
Enterprise tools	Google Cloud Platform, Amazon Web Services S3, Oracle Applications
Back-end tools	Oracle Databases, PostgreSQL, Microsoft SQL Server, MongoDB, JSON
IDEs/Dev tools	Jupyter Notebook, Google Colab, PL/SQL Developer, Git

EDUCATION

University of Maryland, Baltimore County (UMBC), Baltimore, MD, USA
Master's in Data Science GPA: 4.0 **2019 – 2020**

Coursework: Practical Deep Learning, Machine Learning & Data Analysis, Big Data Processing, Databases: SQL, NoSQL

West Bengal University of Technology, Kolkata, India
Bachelor's in Computer Science GPA: 3.5 **2006 – 2010**

Coursework: Data Structures & Algorithms + Labs, Operating Systems + Labs, Artificial Intelligence + Labs

RESEARCH

Ebiquity Research Lab, UMBC | Graduate Student Researcher Sep 2019 – May 2020
Baltimore, MD, USA

- Authored an [Ontology](#) for Smart Home Access Control by extending earlier research in Semantic Web.
- 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.
- Published a [paper](#) at the IEEE Big Data Security 2020 conference.

PROJECTS

Master's degree capstone project: [QABot: A Chatbot for Open Question Answering Using Neural Networks](#)

Best deep learning project: [Image-to-Image Translation Using CycleGAN](#)

PUBLICATION

Sofia Dutta et. al., "Context Sensitive Access Control in Smart Home Environments", InProceedings, 6th IEEE International Conference on Big Data Security on Cloud (Big Data Security 2020), May 26, 2020, Baltimore, MD, USA.