

Sofia Dutta

Software Engineer

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

🌐 <https://linkedin.com/in/sofiadutta> 🌐 <https://sofiadutta.github.io>

Software Engineer with over a decade of experience working with large quantities of data. Worked on every aspect of the software development life cycle, from design & development of core functionality to production deployments in cloud environments. Seeking software engineering opportunities in the US.

SKILLS

Programming – Python, Java, SQL, PL/SQL

Data Science – Apache PySpark, Looker, LookML, PyTorch, Sci-kit Learn, Tensorflow, Hadoop

Data/Software Engineering – Apache Airflow, Databricks, Jupyter Notebook, Google Colab, Git

Cloud – Google Cloud Platform, Google Compute Engine, Google Cloud Storage, Google Cloud SQL, AWS S3

Back-end – Google BigQuery Table, Oracle Databases, PostgreSQL, MySQL, MongoDB, JSON

Domain knowledge – Big Data Analytics, Deep Learning, Machine Learning

WORK EXPERIENCE

Technical Leader @ NewWave Telecom & Technologies, Inc., Windsor Mill, MD, USA Jan 2021 – Present

- Designed the system architecture, database schema, and data workflow for the [Imersis](#) data quality analytics platform.
 - Built the project's cloud infrastructure setup and processed hundreds of millions of healthcare records using Apache PySpark to compute data quality metrics.
 - Created Looker dashboard visualizations with drill-down options that “explains” why the data quality came out to a certain value.
 - Due to the lack of large quantities of data, re-designed the machine learning goal of the project into a data quality “Explainable AI” system.
 - Built a data quality system that allows clients to understand how to improve their data quality over time.
 - Using a four-fold workflow automation strategy saved tens of thousands of dollars in cloud costs for the project.
-

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

- Helped speed up the project's data processing pipeline 10-fold by moving to Google Cloud Platform (GCP).
 - Managed DevOps on GCP and connected the cloud infrastructure to back-end Big Data tables for delivering results to dashboards.
-

Software Engineer @ Tata Consultancy Services, Kolkata, India Nov 2006 – Feb 2018

- Led the design, development, and delivery of API interfaces using PL/SQL stored procedures for several projects of TCS.
 - Carried out change-based regression impact analysis, created software functional specifications, and prepared test plans for several projects of TCS.
 - Performed system integration, user acceptance, and performance testing and ensured client systems had very high uptime even when carrying out data migration activities.
 - Saved millions of dollars in potential revenue lost to the client and was awarded for the said effort by clients.
-

EDUCATION

Udacity Nanodegree: Deep Learning

Oct 2021

Masters in Data Science from the University of Maryland, Baltimore County, Baltimore, MD

Jan 2019 – Dec 2020

Bachelor's in Computer Science from West Bengal University of Technology, Kolkata, India

Jun 2006 – Nov 2010