

# Sofia Dutta

SOFTWARE DEVELOPER · DATA ENGINEER · DATA SCIENTIST

Baltimore, MD, USA

☎ 443-554-4170 | ✉ sofia.dutta17@gmail.com | 🏠 sofiaadutta.github.io | 📷 sofiaadutta | 🌐 sofiaadutta | 🎓 Publication

“Our intelligence is what makes us human, and AI is an extension of that quality.” - Yann LeCun

## Summary

Working as a Data Scientist at NewWave Telecom & Technologies, Inc. Building Machine Learning solutions to improve healthcare data quality of millions of records. Since 2019, building skills in data science domain through projects done on Deep Learning and Machine Learning domains at graduate school. Before that, worked as a Software Developer from 2010 to 2018, leading teams of up to five engineers between 2013 and 2018. Performed activities for end to end software development process, including software requirement gathering from clients, design, development and testing of software etc. Additionally, was a key resource during large scale data migration activities for several projects.

## Work Experience

### NewWave Telecom & Technologies, Inc.

Baltimore, MD

DATA SCIENTIST INTERN

May 2020 - Present

- Built a system for computation of quality analysis and data exploration in record time of a month.
- Trained a machine learning model using 3000 rules for quality computation metrics defined by domain experts.
- Improved computation speed by 10 fold by deploying analysis workflow in Google Cloud Platform (GCP) clusters.
- Used Apache PySpark to perform computation in GCP and exported results to Looker for creating intuitive dashboards visualizations.
- Performing data quality analysis tasks on about half a million healthcare records from Centers for Medicare & Medicaid Services.

### Ebiquity Research Group, UMBC

Baltimore, MD

STUDENT RESEARCHER IN SEMANTIC WEB

Sep 2019 - May 2020

- Authored an *Ontology* for Smart Home Access Control by extending earlier research in Semantic Web.
- Developed Android *app* for handling context-sensitive access control in a Smart Home Environment.
- Uploaded YouTube demo *videos* for presentation to the National Institute of Standards and Technology by principal investigator of project.
- Published a *paper* at the IEEE Big Data Security 2020 conference.

### Tata Consultancy Services (TCS)

Kolkata, India

SOFTWARE DEVELOPER AND DATA ANALYST

Nov 2010 - Feb 2018

- Led the design, development and delivery management of several projects for clients of TCS.
- Created API interfaces using PL/SQL stored procedures for daily usage by clients.
- Held several meetings over the years 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.
- Completed client data migration from legacy Oracle Apps (11i) to Oracle ERP Suite (R12).
- Implemented scripts for managing customer data migration task of over a billion records while maintaining strict time frames.
- Ensured client systems were up within four hours after migration activities saving millions of dollars in potential revenue lost.
- From 2013 - 2018, managed continuous integration and continuous deployment in production environments.
- Got certified in seven Oracle Apps competencies while working for client projects.

## Education

**Masters in Data Science**, University of Maryland, Baltimore County, GPA: 4.0

Baltimore, MD

**Bachelors in Computer Science and Engineering**, West Bengal University of Technology, GPA: 3.5

Kolkata, India

## Skills

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

References available upon request