Sofia Dutta
443-554-4170 • sofiad1@umbc.edu • linkedin.com/in/sofiadutta • github.com/sofiadutta

Career objective: *A software developer with a knack for data science, machine learning, and deep learning, I want to work for a software company that understands the value of data and uses it to achieve interesting outcomes.*

Education

University of Maryland, Baltimore County (UMBC), Baltimore, MD *Spring 2019 – Fall 2020*
Master's Professional Studies, Data Science, *GPA: 4.0*

West Bengal University of Technology, Kolkata, India *Fall 2006 – Spring 2010*
Bachelor of Technology, Computer Science, *GPA: 3.5*

Technical Skills

Programming Languages: Python, SQL, PL/SQL, Java, Visual Basic, C#, JavaScript, HTML, C++, C
Development Tools: PL/SQL Developer, Toad for Oracle, JDeveloper, Docker, Jupyter Notebook, Visual Studio, Google Colab, Oracle Report Builder, Oracle Form Builder, Oracle Workflow Builder, Git, CVS
Enterprise Tools: Google Cloud Platform, Amazon Web Services S3, Oracle Applications
Backend Tools: Oracle (9i, 10g, 11g), Microsoft SQL Server, MongoDB, JSON
Data Tools: Sci-kit Learn, Apache Spark 2.4, MLlib, Keras 2.2, Tensorflow 1.15, PyTorch 1.4
Operating Systems: Mac OSX, Windows, Linux (Ubuntu, RHEL)

Work Experience

Ebiquity Research Group, UMBC, Student Researcher, Baltimore, MD *Sep 2019 – Present*

- Working on research in the area of Semantic Web, Context-based Access Control and Smart Home Automation
- Submitted a paper to IEEE Big Data Security 2020 conference.

Tata Consultancy Services (TCS), Kolkata, India *Nov 2010 – Feb 2018*

- Led a team of developers in preparing PL/SQL stored procedures.
- Designed, developed and tested API interfaces for PL/SQL stored procedures.
- Prepared functional specification documents.
- Performed requirement and change based regression analysis.
- Prepared test plans and performed system integration testing and user-acceptance testing.
- Worked on Oracle Fusion HCM (Core HR) functionalities for clients
- Wrote migration scripts for customer data migration projects.
- Completed client data migration from legacy Oracle Apps (11i) to Oracle ERP Suite (R12).
- Managed production environment deployments.

Projects

Big Data Twitter Sentiment Analysis @ UMBC *Fall 2019*

- Compared performance of traditional machine learning algorithms like support vector machines, logistic regression, and neural networks created using Keras CNN, Keras Bidirectional LSTM to empirically prove neural networks are better at sentiment classification

Sentiment Analysis on user review datasets from Amazon and IMDb @ UMBC *Spring 2019*

- Compared performance of traditional machine learning algorithms like support vector machines, logistic regression, and neural networks created using Keras CNN, Keras Bidirectional LSTM to empirically prove neural networks are better at sentiment classification

Data characterization projects using Python Sci-Kit Learn @ UMBC *Spring 2019*

- Analyzed Baltimore City Employee Salary data to prove there is no income inequality in Baltimore City Government
- Studied New York City Film Permits data to figure out top filming locations for popular movies
- Combined two different datasets from the New York City Fire Department and showed that it is possible to use data analysis techniques to find high impact incidents

Commercial Bank of Dubai, UAE @ TCS Feb 18 – Jan 17

Relevant Coursework

Data 601: Introduction to Data Science *Spring 2019*

- Performed data analysis projects using supervised and unsupervised machine learning packages.
- Worked on data collection, storage, transformation, cleaning, analysis, and visualization.

Data 602: Introduction to Data Analysis and Machine Learning *Spring 2019*

- Worked on practical machine learning and data analysis problems.
- Worked on end-to-end processing pipeline for extracting and identifying useful features that best represent data, applying machine algorithms, and evaluating their performance for modeling data.
- Learned machine learning APIs like Sci-kit Learn, Keras, Tensorflow.
- Learned machine learning algorithms like decision trees, logistic regression, support vector machines, convolutional neural networks, recurrent neural networks, bidirectional LSTM.

Data 603: Platforms for Big Data Processing *Fall 2019*

- Using Apache Spark performed Map-Reduce operations on streaming data
- Learned Big Data technologies like PySpark, Spark SQL, MLlib, Spark Streaming, Hive, Hadoop
- Worked on practical projects with large datasets
- Used NoSQL storage system (MongoDB) to manage large datasets collected from Twitter Data APIs

Awards and recognitions

Awarded certificate: “Recognition of exceptional leadership and teamwork skills” at school *2019*
Awarded TCS Gems “Champions of ILP” for contributions in training new employees *2017*
Awarded certificate: “Most likely to Slay a Dragon” for efforts in Finance Data Migration *2015*

Certifications completed

- Oracle 9i SQL (1Z0-007)
- Oracle 9i PL/SQL (1Z0-047)
- Oracle Database 11g: Advanced PL/SQL (1Z0-146)
- Oracle E-Business Suite 12 Financial Management Implementation Specialist: Receivable (1Z0-518)
- Oracle Fusion HCM Base Cloud Service 2016 Implementation Essentials (1Z0-329)
- Oracle Global Human Resources Cloud 2017 Implementation Essentials (1Z0-965)
- Oracle Talent Management Cloud 2017 Implementation Essentials (1Z0-966)