|  |
| --- |
| Sofia Dutta |
| Aspiring Data Scientist and Machine Learning Engineer | Email: [sofiad1@umbc.edu](mailto:sofiad1@umbc.edu) |
| Summary |
| * Aspiring Data Scientist and Machine Learning Engineer * 7 years of experience in Software Consultancy and Data Analysis |
|  |

|  |  |  |
| --- | --- | --- |
| Education |  |  |
| Current – 2019 | Master’s in Data Science, UMBC, Baltimore, USA | 4.0/4.0 |
| 2010 – 2006 | Bachelor’s in Computer Science and Engineering, WBUT, India | 3.5/4.0 |

|  |
| --- |
| DATA SCIENCE PROJECTS |
| * Comparison of Word2vec and Doc2Vec model driven Sentiment Analysis using SVM, LR, Keras CNN, Bidirectional LSTM with and without pre-trained Word and Document Embeddings. * Sentiment Analysis on Amazon laptop reviews: <https://sofiadutta.github.io/datascience-ipynbs/sentiment-analysis/Sentiment_Analysis_Amazon_Laptop_Review.html> * Sentiment Analysis on IMDb movie reviews: <https://sofiadutta.github.io/datascience-ipynbs/sentiment-analysis/Sentiment_Analysis_IMDB_Movie_Review.html> |
| * Data characterization of Baltimore City Employee Salaries to study income inequality in Government Link: <https://sofiadutta.github.io/datascience-ipynbs/EDA/Data_Analysis_Baltimore_City_Salaries.html> |
| * Data characterization of NYC Film Permits to figure out where my favorite films being shot at   Link: <https://sofiadutta.github.io/datascience-ipynbs/EDA/Data_Analysis_NYC_Film_Permits.html> |
| * Data characterization of NYC Fire Department data to find impactful events via a data analytics path   Link: <https://sofiadutta.github.io/datascience-ipynbs/EDA/Data_Analysis_NYC_Fire_Department.html> |
| * Big-Data Analytics and Sentiment Classification with Twitter Stream Data. See project links below: * [Big data analytics: Tweet Sentiment Classification and Big Data Analytics using Spark Dataframe](https://sofiadutta.github.io/datascience-ipynbs/big-data-analytics/Using_MyClassifier_Twitter_Data_Sentiment_Classification_and_Big_Data_Analytics_on_Spark_Dataframe.html) * [Big data analytics: Tweet Stream Analysis using PySpark](https://sofiadutta.github.io/datascience-ipynbs/big-data-analytics/Twitter_Stream_Analysis_Using_Spark_Streaming.html) * [Big data analytics: Tweet classification via Word2Vec models, Spark MLlib and my own classifier](https://sofiadutta.github.io/datascience-ipynbs/big-data-analytics/Using_SparkMLLib_Word2Vec_Model_Driven_LogReg_Sentiment_MyClassifier_Creation.html) |

|  |  |
| --- | --- |
| Technical Skills | |
| Platforms | Mac, Unix, Windows | |
| Languages | Python, Java, SQL, PL/SQL, TSQL | |
| Tools & stacks | Scikit-learn, Keras, PyTorch, Tensorflow, MLlib, PySpark, PL/SQL Developer, Oracle Report/Forms/Workflow Builder | |
| AI Technologies | Knowledge Graph, Semantic Web, Protégé, SPARQL, OWL | |
| Back-end | MongoDB, Oracle (9i, 10g, 11g) | |
| Enterprise tools | Google Cloud, AWS S3, Oracle Fusion 11, Oracle Applications (R12 & 11i) | |

|  |  |  |
| --- | --- | --- |
| Work Experience | | |
| Current – 2019 | Student researcher in Semantic Web and Smart Home Access Control, Ebiquity Group, UMBC, USA |
| 2018 – 2010 | Software Consultant and Data Analyst at Tata Consultancy Services (TCS) Ltd., India |

|  |  |
| --- | --- |
| TCS Projects | |
| Feb’18–Jan’17 | Commercial Bank of Dubai, UAE |
|  | * Was a key project resource trained to work with Oracle Fast Formulas * Led a team of developers working on Oracle Fusion Human Capital Management (HCM) functionalities like resignation, transfer, promotion, performance rating etc. * Prepared software design and test plans documents * Performed requirement and change analysis. |
| Nov’16–Sep’16 | Kaiser Permanente, USA |
|  | * Led a team of developers working on Oracle Fusion (HCM) functionalities * Prepared software design and test plans documents * Performed requirement and change analysis. |
| Jun’16–Feb’16 | Hyatt-Oracle R12 Rollout, USA |
|  | * Ensured efficient execution of data queries using PL/SQL stored procedures. * Worked on ACH payment, invoice management and HRMS configuration. * Performed requirement and change analysis. |
| Feb’16–Jun’15 | Staples, USA |
|  | * Ensured efficient execution of data queries using PL/SQL stored procedures. * Worked on ACH payment, invoice management and HRMS configuration. * Performed requirement and change analysis. |
| Jun’15–Jan’14 | DHL, UK |
|  | * Ensured efficient execution of data queries using PL/SQL stored procedures. * Handled HRMS and Payroll configuration and administration. * Seamlessly managed data migration activities from Oracle Apps to ERP Suite. * Performed requirement and change analysis. * Tested system functionality and bug fixes. |
| Jan’14–Sep’13 | Tata Steel, India |
|  | * Ensured efficient execution of data queries using PL/SQL stored procedures. * Worked on software design and test plans documents |
| Sep’13–May’13 | IHCL - R12 Reimplementation, Taj Hotel, Mumbai |
|  | * Ensured efficient execution of data queries using PL/SQL stored procedures. * Handled change management activities for Oracle reports, forms and databases. * Seamlessly managed data migration activities. * Worked on software design and test plans documents |
| May’13–Apr’13 | CMCL - Oracle ERP R12 Implementation, Star Cement, Kolkata |
|  | * Ensured efficient execution of data queries using PL/SQL stored procedures. |
| Apr’13–Jun’11 | Bangladesh Railway Oracle Apps, Bangladesh Railway, Bangladesh |
|  | * Ensured efficient execution of data queries using PL/SQL stored procedures. * Developed report generation packages for payroll management * Worked on software design and test plans documents |

|  |
| --- |
| 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) |