Sarang Patil *(he/him)*

4760 Chapel Square, Halethorpe, MD 21227 | 443-527-7295 | sarangp2402@gmail.com

EDUCATION:

University of Maryland Baltimore County**,** Baltimore, MD **Jan 2021 - Dec 2022**
Mater's Degree in Data Science, *GPA – 4.0/4.0*

Savitribai Phule Pune University**,** Pune, India **Jun 2016 - Aug 2020**
Bachelor of Engineering in Computer Engineering, *GPA – 3.7/4.0*

EXPERIENCE:

Data Science Research Assistant, *University of Maryland Baltimore County*, Baltimore, MD **Jan 2022 – Present**

- Implemented **Keras** and **TensorFlow** predictive models using **Python 3.10** to semantically locate the user at home solely with the help of non-intrusive sensor data.
- Analyzed the historical data of energy/water/gas consumption to draw inferences and predict the user's location at home and reduced hardware costs of sensors for data collection by 50-60% approximately by using non-intrusive sensors.
- Utilized **SAS** and **R** for statistical analysis and **Tableau** for visualization dashboards.
- Contrasted the machine learning approach with the static rule approach to compare which approach performs better, and found out that the ML approach gave 5 times more accurate results than the static rules approach.
- Prototyped data visualizations using Charts, drill-down, and parameterized controls using **Tableau** to highlight the value of analytics in executive decision support control.
- Executed data manipulation and then carried out exploratory data analysis and statistical analysis of the manipulated data.
- Troubleshoot and tune machine learning algorithms in **Spark**.
- Actively involved in code review and bug fixing for improving performance.
- Testing and validating machine learning models.
- Developed **Hive** queries to process the data and generate the data for visualizing.

Environment: *Python, SQL, Keras, R, SAS, Spark ML, Apache Spark, Databricks, Hive, Athena, AWS Glue*

Data Science Intern, *CoReCo Technologies*, Pune, India **Aug 2019 - Jun 2020**

- Initiated a project in an Agile team environment (**SCRUM**) implementing the Naïve Bayes model for detecting inappropriate tweets using Twitter API which had a prediction accuracy of 76% and precision of 89%.
- Administrated database management and data querying of approximately 2000 records using **SQL Server**.
- Demonstrated the impact of different algorithms on data in more business terms and less technical jargon by mediating between the project team and management team.
- Involved in communicating with the investors to understand the business logic implementation.
- Trained analytical models with **Spark ML** estimators including linear regression, decision trees, logistic regression, and k-means.
- Used collections in **Python** for manipulating and looping through different user-defined objects.
- Designed and created backend data access modules using **PL/SQL** stored procedures.
- Ingested data from disparate sources using **SQL** and **Google Analytics API** to construct data views for BI tools like **Tableau**.
- Performing feature engineering to add new features/products to train existing models.
- Performed data cleansing and applied transformations using **Databricks** and **Spark** data analysis.
- Developed and executed the User Acceptance Testing portion of the test plan.

Environment: *Python, SQL, PowerBI, Tableau, RESTFul API, SFTP, PySpark, Spark ML, Spark MLLib*

Project Intern, *Aalborg University*, Copenhagen, Denmark **Jan 2018 - Feb 2018**

- Managed an SDLC project team as a team lead at Aalborg University in Denmark regarding the Steam Game Distribution Platform.
- Translated business questions into technical requirements while working with diverse audiences overseas.
- Spearheaded Technical Analysis and Business Analysis of the project and built innovative system architecture to satisfy the requirements of 47% of steam video game customers.
- Experience in dimensional modeling (Star schema, Snowflake schema), transactional modeling, and SCD (Slowly changing dimension).
- Leveraged cloud and GPU computing technologies for automated machine learning and analytics pipelines, such as **AWS**, and **GCP**.
- Implemented a Continuous Delivery (CI/CD) pipeline with **Docker**, **GitHub**, and **AWS**.
- Involved in preparing associated documentation for specifications, requirements, and testing.

Environment: *AWS, Python, Docker, GitHub, Python, Java, JavaScript, SAS, Cloud SQL*

SKILLS:

Languages: Python, R, Java; **OS**: Windows and Linux; **Machine Learning and Deep Learning**: TensorFlow, Keras, Torch, Spark MLlib; **Databases**: SQL, PL/SQL, NoSQL, MySQL, MongoDB, Hive, Cassandra; **Agile Methodologies**: Scrum, Kanban; **Packages**: Pandas, NumPy, Seaborn, Matplotlib, NLTK, Spark; **Container**: Kubernetes, Docker; **Cloud Computing**: AWS Cloud and Google Cloud; **Big Data**: Databricks, Apache Hadoop, Apache Spark; **Microsoft**: Word, Excel, Teams, PowerPoint; **Data Visualization**: Tableau, SAS, Athena; **Web Scraping**: Selenium; **ETL**: AWS Glue, SAS, Apache Airflow

CERTIFICATIONS:

- **Certified AWS Cloud Practitioner**, *Amazon Web Services*
- **Machine Learning**, *Stanford Online*
- **Big Data Modelling and Management Systems**, *UC San Diego*

PROJECTS:

- **Disease Diagnosing Chatbot and Hospital Recommender System**
  Technologies: Python, Keras, Pandas, BERT, Recommendation System, NLTK, Natural Language Processing, Deep Learning
- **Steam Video Game Recommendation Engine**
  Technologies: Python, TensorFlow, Cosine Similarity, Tableau, Selenium, AWS Kinesis, AWS Lambda, Topic Modelling
- **Stock Market Prediction using Twitter Sentiments**
  Technologies: Python, NumPy, PySpark, Time-Series, Sentiment Analysis, TensorFlow, Statistical Modeling
- **Market Basket Analysis and Customer Segmentation**
  Technologies: Python, Apache Spark, Databricks, SQL, Spark MlLib, Text Mining, EDA, Clustering, Association Rules
- **Weather Forecasting and Weather Data Analysis using ML Algorithms**
  Technologies: Python, JupyterLab, Text Mining, ETL, Tableau, Time-series model, Neural Network

PUBLICATIONS:

- "**An Effective Analysis of Anti Troll System using Artificial Intelligence**", International Research Journal of Engineering and Technology (IRJET), Volume: 09, Issue: 12, Dec 2019