SUSMITHA ARIKATLA

+1(281)-750-6048|susmitha.usk7@gmail.com |[GitHub](https://github.com/Susmithareddy1994?tab=repositories) | [LinkedIn](https://www.linkedin.com/in/smithareddya/) | Houston, Texas

**EDUCATION**

**MSc in Data Science** | University of Houston | Houston, Texas | GPA: 3.6

* Awards & Scholarships: Dean’s Honors List, Engineering Dean’s Master Scholarship, Masters Competitive Scholarship
* **Relevant Coursework:** Machine Learning, Statistics, Data Analytics, Database Management, Big Data, Time Series Forecasting, Artificial Intelligence, Data Visualization, Time Series Analysis, Data Analysis

**Bachelors in Electronics and Communications Engineering**| India| C.G.P.A: 8.6

**WORK EXPERIENCE**

**Information Visualization – Teaching Assistant |** University of Houston | Houston, Texas

* Highlighted the utilization of various software tools including Excel, Power BI, R, Tableau to demonstrate the implementation of data science techniques and machine learning models.

**Project Planning & Implementation Engineer |** Tata Communication **|** Chennai, India

* Designed project plans that include project scope, timelines, budget, and resource requirements.
* Coordinated with vendors, contractors, and other stakeholders to ensure that projects are executed according to plan.
* Monitored project progress and adjustments as necessary to keep projects on track.
* Implemented data visualization techniques for creating Dashboards and reports using Python and Tableau to present complex data insights to stakeholders, resulting in improved decision-making and a 20% reduction in time spent on data analysis.

**SKILLS**

* Tableau, Power BI, Qlik Sense, MS Excel, Python, R
* Word, PowerPoint, DAX, Hypothesis Testing, SQL, ETL
* Pandas, Numpy, TensorFlow, Keras, Seaborn, Matplotlib, Scikit-learn, PyTorch.
* AWS, GCP, MYSQL, Snowflake, Microsoft Azure
* Logistic Regressions, Decision Trees, Clustering, Neural Networks, Random Forests, SVM

**PROJECT EXPERIENCE**

**Visualization of Data Scientist Job Salaries**

**Tech Stack:** Snowflake, S3Bucket, Power BI, SQL

* Performed extensive data analysis on job salary data for data scientist positions, identifying key industry trends and patterns, resulting in actionable insights for optimizing compensation strategies.
* Utilized advanced cloud-based technologies such as Snowflake and S3Bucket to efficiently store and retrieve large volumes of job salary data
* Designed and Crafted interactive dashboards using Power BI to visually depict the distribution of Data Science job salaries based on experience levels and job titles for 2023.

**Construction Safety Analysis using OSHA Dataset**

**Tech Stack:** Numpy, Google Collab, Power BI, Excel

* Extracted and analyzed 100k records from OSHA Website using advanced web-scraping techniques, resulting in enhanced data quality and improved understanding of safety trends within the construction industry.
* Applied Principal Component Analysis (PCA) to effectively reduce the dimensionality of the dataset, resulting in more efficient data representation and analysis.
* Employed K-means Clustering technique to determine the optimal number of clusters (K) and identify distinct safety profiles among construction companies, allowing for targeted safety interventions and improvements.

**Seoul Bike Sharing Demand**

**Tech Stack:** Regression, Google Collab, MySQL, Python, Tableau

* Collected and aggregated data from various sources to create a comprehensive dataset for analysis.
* Cleaned, normalized, and engineered features in the dataset to ensure data integrity and suitability for modeling.
* Applied predictive models, including regression and machine learning algorithms, to accurately forecast the demand for bikes at different stations. Achieved 80% accuracy rate, significantly improving bike sharing demand forecasting in Seoul.

**Comparative study of supervised learning algorithms for Intrusion Detection**

**Tech Stack:** Numpy, Algorithms, Python, Data Analysis

* Gathered Data from external sources and processed it using techniques PCA, Label encoding, and Normalization.
* Trained and evaluated Decision Tree, Random Forest, Support Vector Machines, Boosting, Bootstrap Aggregation, Naive Bayes, and Advanced Neural Networks algorithms.
* Assessed the best algorithm based on performance metrics like Accuracy Score, Execution Time, F1 score.

**CERTIFICATIONS**

* Snowflake Certificate
* Google Data Analytics Professional Certificate
* Tableau Desktop Specialist (Tableau Software, LLC)
* Machine Learning-HarvardX