**Susmitha Arikatla**

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

**EDUCATION**

**University of Houston** **Houston, Texas**

Masters in Data Science (CGPA: 3.6/4.0) *May 2023*

* Awards & Scholarships: Dean’s Honors List, Engineering Dean’s Master Scholarship, Master’s Competitive Scholarship
* **Relevant Coursework**: Machine Learning, Statistics, Data Analysis, Database Management, Big Data, Time series Forecasting, Artificial Intelligence, Data Science for Security, Data Visualization
* **Certifications**: Google Data Analytics Professional Certificate, Tableau Desktop Specialist (Tableau Software, LLC)

Machine Learning-HarvardX,Python & Machine Learning for Financial Analysis

**SKILLS**

* **Programming Languages:** Python, R, SQL
* **Databases:** Oracle, MongoDB, AWS S3, Microsoft Office (Word, Excel, PowerPoint), MYSQL, Snowflake
* **DW Skills:** Apache Spark, Apache Kafka, Data Pipelines, Hadoop
* **Python libraries: -** Pandas, Numpy, TensorFlow, Keras, seaborn, matplotlib, Scikit-learn, PyTorch, NLTK
* **Data Modeling -** Logistic Regressions, Decision Trees, Linear Regression Neural Networks, Random Forests
* **Additional Skills**: -Storytelling, Statistical Skills, Strategic Thinking, Decision-making, ad-hoc report
* **Tools**: Data Mining, Statistics, ETL, Hypothesis testing, Qlik Sense, Alteryx, DAX, Informatica, PowerBI, Tableau, SSRS, Agile, SAP, financial analyses

**PROFESSIONAL EXPERIENCE**

**Cullen College of Engineering, University of Houston Houston, Texas**

Data Visualization Teaching Assistant *Aug 2022 – May 2023*

* Showcased the utilization of various software tools including Excel, SAS, R, Tableau to demonstrate the implementation of data science techniques and machine learning models.
* Generated visualizations for professors to use in class, increasing student comprehension by 17% in 4 months.

**Freelance Remote**

Data Analyst *June 2020 – Apr 2022*

* Designed custom dashboards daily using Power BI to illustrate Customer insights, resulting in a 15% increase in conversion rates.
* Partnered with Cross-functional teams and identified opportunities for improvement through analysis of key metrics and KPIs related to user acquisition.
* Created training materials and processed documents using Microsoft Office for implementing predictive analytics.

**ACADEMIC PROJECTS**

**Construction Safety Analysis using OSHA Dataset** [**Kaggle**](https://www.kaggle.com/code/susmithareddya/construction-safety-analysis-using-osha-data/edit) **Link**

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

* Applied web-scrapping to collect the Construction Accidents records from OSHA Website.
* Preprocessed the data using advanced NLPtechniques, including stemming and lemmatization, to enhance data quality and trends.
* Applied Principal Component Analysis **(**PCA**)** to reduce dataset dimensionality and implemented TF-IDFVectorization for feature representation.
* Employed the K-means Clustering technique to identify the optimal number of clusters (K) and centroids.

**Seoul Bike Sharing Demand**  [**Kaggle**](https://www.kaggle.com/code/susmithareddya/seoul-bike-sharing/edit) **Link.**

**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.
* Developed 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**  [**Kaggle**](https://www.kaggle.com/code/susmithareddya/comparative-study-of-supervised-learning-algorithm/edit) **Link**

**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, SVM, XGBoost, Naive Bayes, and Advanced Neural Networks algorithms.
* Implemented the best algorithm based on performance metrics like Accuracy Score, Execution Time, and F1 score.

**VOLUNTEERING & INTERESTS**

* UH-ISSO organization | Student Mentor