SUSMITHA ARIKATLA

+1(281) -750-6048 | susmitha.usk7@gmail.com | [LinkedIn](#) | [Git Hub](#) | Houston, Texas

EDUCATION

MSc in Data Science | University of Houston | Houston, Texas | GPA: 3.6 *May 2023*

- 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

SKILLS

Visualization Tools	: Tableau, Looker, Power BI, Qlik Sense, Microsoft Excel, Python (Seaborn, Matplotlib), R
Tools	: Microsoft Office, DAX, PowerPoint, SharePoint, Hypothesis Testing, SQL, Spark
Database	: MYSQL, SQL Server, PostgreSQL, Azure ML Studio Python
Libraries	: Pandas, Numpy, TensorFlow, Keras, Scikit-learn, PyTorch.
Cloud Technologies	: AWS, GCP, Snowflake, Microsoft Azure
Environments	: GitHub, Google colab, PyCharm, VSCode, Jupyter, Rstudio
Data Modeling	: Logistic Regressions, Decision Trees, Multivariate Regression, Neural Networks, Random Forests

WORK EXPERIENCE

Data Science – Teaching Assistant | University of Houston | Houston, Texas *Aug 2022 – May 2023*

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

Data Analyst | Freelance **|** Remote *Jun 2020 – Apr 20222*

- Collaborated with clients to understand their data analysis needs and deliver tailored solutions.
- Conducted data cleaning, preprocessing, and exploratory data analysis to uncover insights.
- Performed statistical analysis and created data visualizations to communicate findings effectively.
- Developed predictive models using machine learning techniques to improve decision-making.

PROJECT EXPERIENCE

Construction Safety Analysis using OSHA Dataset *Apr 2023*

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

- Extracted and analyzed 100k records from OSHA Website using advanced web-scraping techniques, resulting in enhanced dataquality 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 amongconstruction companies, allowing for targeted safety interventions and improvements.

Visualization of Data Scientist Job Salaries *Dec 2022*

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, resultingin actionable insights for optimizing compensation strategies.
- Utilized advanced cloud-based technologies such as Snowflake and S3Bucket to efficiently store and retrieve large volumes of jobsalary data, streamlining the analytical process by 30%.
- Designed and Crafted interactive dashboards using Power BI to visually depict the distribution of Data Science job salaries based onexperience levels and job titles for 2023.

Seoul Bike Sharing Demand *Aug 2022*

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

- Collected and aggregated data from various sources to create a comprehensive dataset for analysis (EDA)
- 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 atdifferent stations. Achieved 80% accuracy rate, significantly improving bike sharing demand forecasting in Seoul.

Optimization of SVM Classifier using Kernel and Ensemble Techniques *Apr 2022*

Tech Stack: Machine Learning Algorithms, Python, Classification

- Designed custom kernel functions for SVM using Sine, Cosine, RBF, Gaussian, and Polynomial functions using kernel tricks to train unbalanced data. Used Boosting and Bagging methods on custom kernel SVM models to improve accuracy.
- Increased accuracy of the models by more than 15% when compared with inbuilt SVM functions.

CERTIFICATIONS

- Machine Learning (HarvardX)
- Google Data Analytics Professional Certificate
- Tableau Desktop Specialist (TableauSoftware, LLC)
- Python and Machine Learning for Financial Analysis