SUSMITHA ARIKATLA

+1(281)-750-6048|susmitha.usk7@gmail.com |GitHub | LinkedIn | 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

WORK EXPERIENCE

Information Visualization – 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 data science techniques and machine learning models.

Data Analyst | Freelance **|** Remote *Apr 2019 – Mar 2022*

- Implemented data visualization techniques for creating Dashboards and reports using Python and Tableauto present complex data insights to stakeholders, resulting in improved decision-making and a 20% reduction in time spent on data analysis.
- Collaborated with cross-functional teams to develop and implement machine learning algorithms for maintenance and collect and report KPIs.
- Combined advanced statistical techniques and predictive modeling to analyze a large dataset of customer behavior, resulting in a 15% increase in customer retention.

SKILLS

- Tableau, Looker, Power BI, Qlik Sense, Excel, Python, R, Data Analytics, Matillion, SAS, SPSS
- Microsoft Office, DAX, PowerPoint, SharePoint, Hypothesis Testing, SQL, SAS, ETL, Data Modeling
- Pandas, Numpy, TensorFlow, Keras, Seaborn, Matplotlib, Scikit-learn, PyTorch.
- AWS, GCP, MYSQL, Snowflake, Microsoft Azure
- Business Performance Analysis, Financial Modelling, Statistical Analysis, Reporting Tools, Price Testing and Analysis, Forecasting Trends, Analyzing Performance, Expense and Revenue Analysis

PROJECT EXPERIENCE

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.

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, 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 on experience levels and job titles for 2023.

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.

Optimization of SVM Classifier using Kernel and Ensemble Techniques

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 accuracyof the models by more than 15% when compared with inbuilt SVM functions.

CERTIFICATIONS

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