ANITA DURG

DATA SCIENTIST

Phone: (512) 954-7266 | Austin, Tx | Email | LinkedIn | GitHub | Tableau

SUMMARY

Data Analytics professional with background utilizing Python, SQL, Tableau for the storytelling, data visualization, exploratory data analysis, and applying machine learning models to data. Currently seeking Data Analysts opportunities in the healthcare, tech start-up, e-commerce.

SKILLS

LANGUAGES: Python, Pandas, Plotly, NumPy, SQL, SAS, HTML
DATA COLLECTION: JSON, CSV, API, Web-Scraping, Excel
DATA VISUALIZATION: Matplotlib, Seaborn, Tableau, Excel
MACHINE LEARNING: Linear Regression, Logistic Regression, Classification, KNN, Random Forest, Naive Bayes, Natural Language Processing (NLP), Decision Tree Classification, Time Series Analysis

RELEVANT EXPERIENCE

Cognidoc Healthtech Private, *Data Analyst*, Bangalore-India Jan 2023 - current

- Collaborated with a team of doctors, developers, and clinical research associates to analyze the effect of the Itoluzima drug on COVID positive patients.
- My role is to clean the data and provide meaningful visualizations using plotly, seaborn, and matplotlib which display the improved results of a new drug on few pro-inflammatory biochemical parameters to view a significant decrease.
- Vonisha Healthcamp Non-profit Organization Visualizations

Springboard Data Science Career Track, Certification - April 18th, 2023

Multiple Time Series on Crime Reports in Austin

- Conducted 20-time series analyses with 10 APD(Austin Police Department) sectors and about 2 major crimes.
- Implemented ARIMA modeling and a few simple forecast methods like the simple average method, and Holt-Winter's exponential smoothing method as a baseline for a single time series. Attempted XGBoost for single time series.
- Applied Prophet and AutoARIMA on all 20-time series. MAPE scores with a mean of 50% were obtained.

Yelp restaurant reviews - Sentiment Analysis

- Applied Logistic Regression, Random Forest Classifier, AdaBoost Classifier, and XGBoost to analyze Philadelphia restaurants using the Yelp reviews dataset.

- AdaBoost Classifier recall score improved from 50% to 83%. Feature Interpretability was shown using SHAP on the XGBoost classifier.

Heart Disease Prediction

- The logistic Regression algorithm is used to predict heart disease by using two features: Blood Pressure and cholesterol. By selecting the best hyperparameters accuracy score got improved from 0.59 to 0.81

PAST WORK EXPERIENCE

Digital Cheetah Solutions - *Web Developer,* Austin, Tx Apr 2007–May 2017

Developed many web applications with HTML, CSS, JavaScript, proprietary code (PERL-based), jQuery, Ajax, and MySQL to perform CRUD (Create, Read, Update, Delete) operations. Developed REST APIs, and many more applications. Extensively wrote complex SQL queries and advanced SQL aggregate functions.

EDUCATION

San Jose State University - MS Molecular Biology- '97
Osmania University - MS Genetics - '94
Osmania University – BS in Biological Sciences – '92