

Sai Varun Teja Mudumba

Email: saivaruntejamudumba@gmail.com

Ph: 9403058666

Professional Summary:

- Experience in analyzing data using **Python, SQL, Hive, PySpark, R** for **Data Ingestion, Data Analytics, Data Cleansing, Machine Learning** and some concepts of **Deep Learning**.
- Proficient in **MySQL** and **NoSQL (MongoDB)** databases with **Python**.
- Experienced in working with various Python IDE's using **Jupyter Notebook, Spyder, Visual Studio code**.
- Experience with **SQL Alchemy, NumPy, SciPy, Pandas, Matplotlib, Beautiful Soup, Urllib, Keras** python libraries during development lifecycle.
- Strong ability to conduct qualitative and quantitative analysis for effective data-driven decision making in analytics.
- Experience in data preprocessing, data analysis, machine learning to get insights into structured and unstructured data.
- Working knowledge of **Big data** technologies like **Kafka, Spark** and **YARN**.
- Experience in analyzing data using **Hive QL, Pig Latin** and custom **MapReduce** programs in Python.
- Excellent Interpersonal and communication skills, efficient time management and organization skills, ability to handle multiple tasks and work well in a team environment.
- Supporting for Git version control system merging, hooks, branching the files and creation of repositories.

TECHNICAL SKILLS:

Programming Languages	Python, Java, R, JavaScript, PL/SQL, Unix, C
Python Libraries/Packages	NumPy, Pandas, SQL Alchemy, SciPy, StatsModels, Matplotlib, Seaborn, Plotly, Bokeh, ScikitLearn, XGBoost, NLTK, TextBlob, Urllib, Beautiful Soup, TensorFlow, Keras.
Data Bases & Hadoop Ecosystem	MySQL, NoSQL (MongoDB), HDFS, Map Reduce, YARN, Hive, Spark-PySpark
Machine Learning	Supervised Learning (Linear Regression, Logistic Regression, Decision Tree, Random Forest, XGBoost, SVM, Classification), Unsupervised Learning (Clustering, KNN, Factor Analysis, PCA), Natural Language Processing
Analytical Tools	Tableau, Rapid Miner, SAS Enterprise Miner
Quant & Statistical Analysis	Hypothesis Testing, Anova, Ancova

CI-CD Tools	GitHub
-------------	--------

Professional Experience:

Capital One, Dallas, TX
Role: Software Engineer

March 2020-Present

- Worked on the Capital One's version of Jupyterlab Hub called Notebook One and
- implemented a code to detect if there are any hardcoded user credentials and blacklisted domains in it.
- Implemented a AWS Lambda function to perform an Offline scan of all the notebooks and
- sending an email with all the findings detected using the above code.
- Worked with the Notebook Transfer API and implemented an Online scan for the hardcoded user credentials and blacklisted domains when the user tries to share his notebook to another user.
- Created a python package which allows users to send an Email to the Capital One users just by a function call, scanning it for passwords and sensitive information, relieving users from generating Bearer Tokens and making a connection to the API.
- Created a python package which allows users to schedule their notebooks and run them at the specified time.

Environment: Python, Tableau, NotebookOne, AWS

Amigos Library Services, Dallas, TX
Role: Data Science Intern

September 2019-December 2019

- Implemented a Python script that automates the process of providing all the members that fall under a specified radius along with its characteristics when a library details are given.
- Developed a Logistic Regression classifier model to predict the score range for library based on the operating budget, type of library, location etc., and achieved an Accuracy score of 88%.
- Developed a Gradient Boosting classifier model to predict whether the given library is eligible to be a member or non-member or other and achieved an Accuracy score of 89%
- Performed preliminary data analysis and data preparation using descriptive statistics and handled anomalies such as removing duplicates and imputing missing values.
- Created an Interactive Visualization System using Google Visual Studio, Python libraries to filter data and create dynamic visualizations to determine the relations between the features.
- Designed a SQL database integrated from the different tables related to the company with normalizations.
- Extracted data from the MySQL server and analyzed data using Python in order to do real time stream analytics.

Environment: Python, Tableau, Machine Learning (ScikitLearn), Excel, MySQL workbench

Data Marshall, Franklin, TN
Role: Data Engineer Intern

May 2019-August 2019

- Automated the process of Classification of tickets and the Priority of the tickets that come up in the company.
- Used eXtreme Gradient Boosting Classifier for building the model to classify the tickets that come up in the company into High, Medium and Low and achieved an accuracy of 71%.
- Used a Random Forest Classifier model to classify the priority of the tickets achieving an Accuracy of 75%.
- Executed Data Analysis and Data Visualization on survey data using Python libraries as well as Compared respondent's demographics data with Univariate Analysis using Python (Pandas, NumPy, Seaborn, ScikitLearn, and Matplotlib).
- Applied techniques like the Dummy Coding and SMOTE sampling techniques to treat categorical values and dominance of the majority classes and treated the data to avoid the Multicollinearity Trap since they are all highly correlated.
- Extracted the company related tickets data from the MySQL server using SQL Alchemy a Python SQL toolkit to extract the database using pandas into a data frame and processed data according to treat duplicates, handling missing values.

Environment: Python, Tableau, Machine Learning (ScikitLearn), Excel, MySQL workbench

University of North Texas, Denton, TX

January 2019-December 2019

Role: Data Science Graduate/Graduate Assistant

- **WAZE TRAFFIC ANALYSIS FOR ACCIDENT PREDICTION :**

Achieved **70%** Accuracy score while predicting the occurrence of Accidents based on a sequence of multiple events that occurred on the road by developing a **Sequential Model** (with Embedding and LSTM layers) using **Keras** library. Developed a **Sequence Generator** script that gives us the sequences that happened on the road before an Accident occurred using some **Radial Filters** and Time criteria.



<https://github.com/saivaruntejamudumba/Waze-Traffic-Analysis-for-Accident-Prediction>



<https://medium.com/@saivaruntejamudumba/waze-traffic-data-analysis-part-1-1a9b31c721f6>

- **RECOMMENDER SYSTEM USING USER QUERY:**

Developed a Recommendation System on a restaurants data set that provides us the to provide the best results provided a User Search query based on the Reviews and the Stars. Applied the **NLP** (TF-IDF, Word2vec) and the concept of **Cosine Similarity** between the User Query and the word vectors to produce the best recommendations and used the Stars criteria to sort them before displaying them to the Users.



https://github.com/saivaruntejamudumba/Recommender_System_using_User_Query

- **AUSTIN AIRBNB PRICE PREDICTION:**

Achieved 76% R2 score while predicting the Austin homestay prices for renting and offering lodging based on the locality, number of nights to stay, availability and the reviews, developed by building **Multivariate Regression** and **Gradient Boosting Regression** models after treating the Missing values with **Imputation** and treating Categorical values with **One Hot Encoding**.



<https://github.com/saivaruntejamudumba/Austin-AirBnB-Price-Prediction>

- **CRICKETERS SCORE PREDICTION and ANALYSIS using CRICPY:**

Performed basic Data Analysis and Model building on the data acquired from cricpy package which extracts the cricket players statistics from the ESPN Cricinfo Statsguru. Achieved 96% R2 score while predicting the scores for a player based on the various attributes like the Position in which he bats, Minutes spent playing, Number of 4's and 6's and the number of Innings, after treating the Missing values with **Imputation** and treating Categorical values with **Dummy Encoding**.



<https://github.com/saivaruntejamudumba/Cricketers-Score-Prediction-and-Analysis-using-Cricpy>

Education Details:

Master of Science

University of North Texas

Bachelor of Technology

Jawaharlal Nehru Technological University