Sai santosh Eedupuganti

Irvine, California

☎ +1 518-512-9486 | ✉ saieedupuganti@gmail.com | ⌂ saisantosh97.github.io | saisantosh97

Summary

Aspiring Data Scientist with 3+ years of experience, enthusiastic in solving business problems. Skilled in Data Analysis, Data Visualization, Statistics and machine Learning models with keen eye to detail and estimation. Highly organised individual with proficiency in team collaboration and sound communication skills to engage with business and team members.

Skills

Programming languages :	Python, R programming
Database systems :	MySql,postgresql, MongoDB
Machine Learning :	Classification, Regression, Clustering, Dimensional reduction, Ensemble methods, Recommendation systems, Feature Engineering, Natural language Processing ,Time series analysis, Neural networks.
Libraries :	Numpy, Pandas, Matplotlib, Scipy, Scikit-Learn, NLTK, Seaborn, Keras, Tensorflow, Pyspark, Dask, Networkx, Beautifulsoup, Selenium.
BI and data visualization :	Tableau, Excel
Front-end Languages :	HTML, CSS
Others :	Problem solving, Statistical methods, ETL(Extract, Transform, Load), Machine learning pipelines, Topological data analysis, Web scraping, Github, Dockers, Heroku.

Work Experience

Research Assistant

Albany, New York

State University of New York — *Jan. 2019 - Jan. 2020*

- Worked on NYSDEC projects on water quality analytics and forecasting.
- Built data pipelines from extracting data through API's using Python , R and data wrangling to create datasets and visualized over 20 years of data using Tableau and RShiny.
- Analyzed 2 million data and built regression models to predict chemical value in water for safer consumption using Linear regression and Tree based models.
- Supported multiple data requests like managing , organizing and creating easy access dataframes for other team members.
- **Tech stack:** Python , Tableau , R , RShiny , Sklearn , Jupyter Notebook.

Data Analyst

Hyderabad, India

Navtech — *Jun. 2016 - Jul. 2018*

- Worked with Cloud product team to support daily data requests using Python , SQL and Tableau.
- Analyzed millions of rows to come up with meaningful business driven insights which helped the team to understand about the product users better.
- Built dataframes using PySpark with AWS EC2 cluster for big data requests within the org.Scraped data from 3rd party API's using Python BeautifulSoup and Urllib and aggregated multiple data sources for analysis and product insights.
- Developed clustering algorithm for customer segmentation by demographics using Python and Sklearn.
- Came up with classification algorithms to know customer retention and analysis using Python and JupyterNotebooks.
- Owned dashboards in tableau for weekly and monthly client business metrics reviews and updates.
- **Tech stack:**Python , SQL , Tableau , PySpark , Jupyter Notebooks, AWS EC2

Projects

Facebook friend recommendation using graph mining

- Developed a model to predict link whether two users are going to be friend in future or not.
- Created graph based features like shortest path, Jaccard, Cosine distance, Pagerank, Adar index, Katz Centrality and built random forest model, achieved a f-1 score of 93 and performed hyper-parameter tuning for best scores(Precision and recall)
- **Tools:** Python, Numpy, pandas, Scikit-learn, NetworkX, Random forest,Xgboost

Quora question pair similarity

- Developed a model to predict whether a pair of questions are duplicate or not, which is useful to instantly provide answer to questions that have already been answered
- Performed text feature engineering and implemented random forest, logistic regression, linear SVM model, XGBoost with Hyperparameter tuning and achieved a log loss of 0.313.
- **Tools:** Python,Numpy, Pandas, Scikit-learn, Plotly, NLTK, BeautifulSoup, XGBoost

Taxi demand prediction in NYC

- Cleaned and performed Exploratory data analysis and feature engineering on New York city taxi data using Python and Dask
- Segmented the data by using K-means clustering and applied baseline models, Linear Regression, Random Forest Regressor, XGBoost Regressor to predict the number of taxi pickups in a given location at a particular time interval and computed a Mean Absolute per error of 11.57
- **Tools:** Python, Pandas, Numpy, Matplotlib, Sqlite3, Dask, Folium, Gpxpy, Scikit-Learn, XGBoost

Stack overflow Tag prediction

- Collaborated with a team of 3 to design a model to suggest tags for question posted on Stack overflow
- Interpreted data, created statistical analysis of content of questions, framed a multi class classification problem and build logistic regressor and support vector regressor to predict tags for content of question.
- Achieved a average F-1 score of 0.51 after necessary hyper-parameter tuning.
- **Tools:** Python, pandas, Numpy, sqlite3, matplotlib, seaborn, scipy, NLTK, Scikit-learn.

Education

State University of New York *Albany, New York*

MASTER'S IN DATA SCIENCE *Aug. 2018 - May. 2020*

Jawaharlal Nehru Technological University Hyderabad *Hyderabad, India*

BACHELOR OF TECHNOLOGY IN MECHANICAL ENGINEERING *Jun. 2014 - Apr. 2018*