Los Angeles, CA, USA
5623915275
Saumya071294@gmail.com
www.linkedin.com/in/saumya-s
www.github.com/saumya-saumya

SAUMYA

Technical Skills

Languages
Python, R, C++, JavaScript, Java

Machine Learning Tools and Frameworks
Preprocessing, Information Extraction, Spacy, Standford CoreNLP, Keras/Tensorflow, Scikit learn, numpy, scipy, Matplotlib, pandas

Database Management and Visualization Tools
MySQL, Oracle SQL, Tableau, looker, plotly, Business intelligence

Research and Planning
Problem-solving ability, gathering information, Developing Evaluations, calculating Results.

Experience

05/2019 – PRESENT

Data Scientist / Institutional Research and Analytics, CSULB, California

- Ongoing research project on higher education market analysis, perform Data mining , data engineering and feature engineering tasks to extract information to create dataset for the model, visualize and compare university dataset at a location while also evaluating the community around the area like population, income and others to help the college understand the location trend and support business decisions.
- Develop Python Script to automate data download from website according to the user preference for web scrapping and storing in dataset.
- Worked on automation Script to get classifiers from the python module and tune hyper-parameters for each algorithm in Scikit-learn.
- Build infrastructure to automate the reports generation on the graduation rate of universities for each year according to Race, Ethnicity gender and demographics using plotly and seaborn.
- Another research project on Identifying top K universities based on the user criterion and visualize using clustering algorithms in Machine Learning, operationalized models to automate ETL, model training, Performance monitor and retraining models under continues monitored System.

12/2017 – 06/2018

Software Developer / Sprinklr, India

- Developed and implemented a module for managing messages on social media for clients like McDonald's and Microsoft, to interact with the customers and analyze business accordingly with 86% accuracy on streaming data.
- Built information extraction and text understanding pipelines, using statistical and machine learning techniques, to identify customer insights from multilingual and unstructured data using SQL. Coordinated effectively a team of 12 members with skills in engineering and design.

02/2017 – 08/2017

Software Developer Intern / Nvidia, India

- **Sentiment Analysis Project:** Worked to Analyze and predict the user utterances emotions with various algorithms and compare results.
- **Copilot Project:** Researched on project named “Contextual Domain classification in spoken language systems” for their autonomous driving project with Python, TensorFlow using different network architecture like CNN, RNN, LSTM and monitor performance and results for each and evaluate results.
- **Copilot Project**: Worked to build chatbot for the system using user utterances and also understand each for the existing systems.

Education

MAY 2020

Master's in Computer Science / California State University, Long Beach, California, USA

JUNE 2017

Bachelor's in Computer Science and Engineering / R. V. College of Engineering, Bangalore, India

Projects

- **Freshman Survey CIRP Project** [LINK]
  The CIRP Freshmen Survey data on incoming students, provide insights and institutional improvements through depth data analysis and data visualization on tableau, using python and MySQL.
- **Lipreading using Deep Learning**
  End to end Sentence level lip-reading using lipnet (TensorFlow network) which consists of convolution network and bi-gated Recurrent Unit network. Researched the network used and the architecture with accuracy.
- **Stream Processing Tweets Using Apache Flink** [LINK]
  Using the streaming twitter data online implemented stream processing using Apache Flink which was deployed on AWS (EMR, EC2, S2). The Project’s main aim was to use the tweets and analyze sentiment with natural language processing while showing real time insights of the emotions and storing the data. In addition, evaluated performance with the Apache Spark.
- **Market Analysis and Geolocation based Advertisement** [LINK]
  Based on the data available online about the area’s population, income, preferences and trends evaluating the demand for the product and Data Analytics for advertising the bidder’s product to target user’s only. Visualized targeted user’s location and evaluated the location demand for the product by intensity.