Sanjana Moodbagil Mallikarjuna

moodbagi@usc.edu | Los Angeles, CA | (213) 573-9842 | linkedin.com/in/sanjana-m-moodbagil | github.com/sanjanamm98 | sanjanamm98.github.io | GHC 22 Attendee

EDUCATION

University of Southern California, Los Angeles, CA **Aug 2021-Present**
Master of Science in Computer Science | *Analysis of Algorithms, Database Systems, Web Technologies, Natural Language Processing, Machine Learning for Data Science, Security Systems*| **GPA 3.6/4**

BMS College of Engineering, Bangalore, India **Aug 2020**
Bachelor of Engineering in Information Science and Engineering | **GPA 9.32/10**

SKILLS

- **Languages**: Python, JavaScript, Java, SQL, HTML/CSS
- **Frameworks & DBMS**: MongoDB, NoSQL, Tensorflow, Keras, Sklearn, NLTK, NodeJS, Angular, Flask, Ajax, Android
- **Utilities / Tools**: Anaconda, Git, Google Cloud Platform, AWS S3, EC2, AWS DynamoDB, Lambda, Google AppScripts
- **Data viz.**: Tableau, Excel, Matplotlib, Seaborn, Google Data Studio

WORK EXPERIENCE

Software Engineer Intern | Karsun Solutions LLC | Herndon, VA **June 2022-Aug 2022**

- Devised an application to generate PII anonymized 'fake' data from original data in a team of 5 for the National Science Foundation (NSF). Developed a python code to automate the generation of synthetic data for time series and relational databases using Synthetic Data Vault (SDV) in domain as per project goals.
- Extracted data pertinent to project goals. Conducted Exploratory Data Analysis (EDA) to interpret structure of data and compare format and statistical properties with the generated synthetic data.
- Automated the generation of metadata for relational tables using Python and MySQL and conducted PII check to preserve data privacy. Worked with AWS S3 Bucket, Lambda and Dynamo DB to integrate backend with the rest of the application.

Risk Data Scientist | **Information Technology Services (ITS) | USC**, Los Angeles, CA **Jan 2022-May 2022**

- Formulated IBM QRadar SIEM risk metrics to reduce threats severity, events and flows through logs. Using these key metrics to pull essential data into Panaseer (Data Lake).
- Designed dashboards for visualization on Panaseer. ML Pipeline and analytics to build statistical model to forecast future risks.

Data Scientist / Data Analyst Intern | **Caliper Business Solutions,** Bangalore, India **Jan 2021-Jun 2021**

- Created Data Transformation models using SQL on Holistics to utilize data for analysis and developing ML models.
- Constructed Analytics dashboards using SQL and Python libraries on Time Budgeting and Transporter Performance.
- Implemented Route Analysis on unsupervised spatial data using DBSCAN model to understand Driver behavior and Patterns in Route Taken to help companies have prior knowledge about expected stoppages, potential stoppage hubs, expected destination time etc

Research Intern | **Dalhousie University,** Halifax, Canada **Jun 2019-Aug 2019**

- Devised a python script on Jupyter Notebook to analyze the time series of groundwater levels and reconstruct the history of groundwater recharges to aquifers based on the nature of hydrographs.
- Executed data analysis and signal processing of time series data using python. Presented dashboards on Tableau to show data trend, understand how various parameters affect the nature of hydrographs, and forecasts of the recharge rates.

PROJECTS

Stock Search Application **Spring 2022**

- Built an Angular Application hosted on GCP connected to NodeJS proxy server that uses Finnhub API for stock details.
- The application monitors stock trends, price changes, top news articles and Highcharts to show stock variations plots. One can buy/sell stocks, wishlist favorite stocks, keep track of stocks bought or sold, post on Twitter or Facebook. HTML5 Local Storage is used to wishlists and purchased stocks. Implemented the same with Android Studio using Java.

Analysis of Speech Patterns in Children to Detect Depression **Spring 2020**

- Performed a comparative study between Logistic Regression, Random Forest, SVM-Gaussian and SVM-Linear to identify a supervised model that best analysis the features extracted from audio data to identify children with depression. Random Forest performed the best with an accuracy of **94%.**
- Applied transfer learning approach where a pre-trained WaveNet feature extracting model was used with CNN classifier that gave an accuracy of **78%.** This gives more and better input features but performs poorly due to small dataset size

Histopathological Cancer Detection **Fall 2019**

- Investigated the potential of ML algorithms to detect metastases in lymph nodes shown on body CT scan image data
- Trained supervised models such as MobileNet, Xception, and CNN in characterizing cancerous from normal cells using Jupyter notebook with TensorFlow and got an accuracy of **90%, 75% and 90%** respectively.

ACHIEVEMENTS & LEADERSHIP

- Selected for **Mitacs Scholarship** to participate in a research internship at Dalhousie University, Canada - One among four selected from Fall 2020 batch.
- Organized coding and debugging competition for college technical fest at BMS College of Engineering