THARUN KUMAR REDDY KARASANI

+1 (206) 532-9924 | karasani.tarunreddy@gmail.com| tharun-kumar-reddy-karasani |tharunkumarreddy5.github.io | Seattle, WA

- A Data Science enthusiast with over **2+ years** of **academic** and **2+ years** of **industrial** experience in **Data Analytics**, **Machine Learning, Prescriptive Analytics, Data Engineering** & **Visualization, Big Data**
- Problem solver and Data Savvy with strong **Analytical & Statistical skills** and knowledge of **Data's Impact on Business**
- Extensive hands-on experience with various Machine, Deep Learning Algorithms, ETL, Visualization tools & large data systems
- Proponent of **Diversity, Inclusion,** and ability to interact with peers and stakeholders with actionable insights that drive impact

My Portfolio

EDUCATION

Master of Science in Data Science | University of Washington, WA **(GPA: 4.0/4.0)** **Mar 2023 (Expected)**

Relevant UW Coursework - Applied Statistics, Statistical Machine Learning, Scalable Data Systems & Algorithms

Bachelor of Technology in Computer Engineering | VIT University, India **(GPA: 4.0/4.0)** **Apr 2019**

University Rank 3 - Program representative and **Merit Scholarship** winner for three consecutive years

TECHNICAL SKILLS

- **Programming Languages:** Python, R, SQL, Java, PySpark, Scala
- **Software & Tools**: Alteryx, Knime, Tableau, GitLab, GitHub, R-Studio, Anaconda, MySQL, Postman, Matlab, Microsoft Excel
- **Frameworks**: ETL, Scikit-Learn, Tensorflow, Keras, Pytorch, Flask, NLTK, Spacy, KerasRL, AutoML
- **Data Science & Machine Learning:** Regression, Classification, Pattern Mining, Ensemble Models, Bagging, Boosting, Cluster Analysis, Prescriptive Analytics, Time series Forecast, RNN, Computer Vision, Natural Language Processing, Reinforcement Learning

WORK EXPERIENCE

Data Science Engineer – Data Intelligence | Goldman Sachs – Bengaluru, India **Jan 2019 – Jul 2021**

- Built and deployed machine learning models and performed prescriptive analytics to generate business insights for operational efficiency
  - Achieved **70% reduction** in Trade Fail rate by identifying fail root causes using a Stacked generalization of **K-medoids** clustering, **Ensemble** models, and **FP-Growth** pattern mining algorithms
  - **Automated 30%** of the manual workflow by resolving trade flow bottlenecks using **unsupervised, distance models** in settlements
  - Accelerated model process time by **1.5 times** through dynamic multiprocessing and threading in the ML Pipeline
  - Enabled data-driven decision making by publicizing powerful **Tableau** dashboards among business teams with relevant **KPIs**
  - Integrated model predictions with GitLab Issue board to track the status of issue resolution for functional resource management

Business Intelligence Analyst - Advanced Analytics | Goldman Sachs – Bengaluru, India **Jan 2019 – Jun 2019**

- Performed data modeling, pipelining, processing using Python, Alteryx and visualized it through live, interactive Tableau dashboards
  - **Front to Back –** Developed huge data artifacts that stitch data from front-end trading and sales to back-end settlements systems
  - Improved ETL processing speed by **50%** after pipelining data through **PySpark refiners** and leveraging OLAP models
  - **Zero Tolerance Controls -** Enhanced the code fail tracking and control framework for threshold breaches by leveraging JIRA API
  - **Quick SDLC -** Automated and fastened the ETL jobs SDLC process by integrating with GitLab version control
- Partnered and cross-trained with four global teams to understand the product functionality and formulate business **OKRs**
- Goldman Sachs COVID-19 Data-thon winner for proposing an effective vaccine distribution algorithm

Machine Learning and Big Data Researcher | VIT University – Vellore, India **Jul 2018 – Oct 2018**

- Developed a **time series forecast** model for analyzing groundwater availability and usage in the city of Vellore, India
- Monitored the usage with seasonality and forecasted the availability in real-time using the **LSTM** model
  - Observed an average excess water usage reduction of **200 gallons** per family post model deployment
  - Automated the extraction, processing pipeline, and storage using **PySpark** and HDFS
- Productionized the Tableau dashboard to monitor the availability and launched a personalized website for user details and bill payments

RESEARCH AND PROJECT WORK

COVID-19 Infection Dynamics and Vaccine Distribution | Regression Analysis, ANN, Regularization, FIR **Mar 2021**

- Devised **epidemiology ML model** to forecast the Coronavirus spread in INDIA and USA with and without vaccination
- Proposed a vaccine distribution algorithm using **ANN** to understand the vaccine urgency in the states of India & prioritize vaccination
- Presented the research paper and published findings at ITNG 2021 conference

Smart Face Recognition and Activity Monitoring |CNN, Attention Model, Hadoop, SQL **Sep 2018**

- Developed a face recognition and activity description system using **CNN** and captioning model with **Attention**
- Leveraged **SparkML, HDFS,** and a few other big data components for data extraction, processing, and data model creation

PUBLICATIONS

- **Springer -** A Comprehensive Analysis of SARS-CoV-2 in India **Feb 2021**
- **IEEE -** Monitoring of Groundwater level & Development of Control Mechanism based on Machine Learning Algorithm