Ronak Sankaranarayanan

Houston, TX - 77084 | ronaksankar.it@gmail.com | 774-418-9871 | linkedin.com/in/ronaksitds/ | Personal Website

EDUCATION:

Master of Science, Data Science, 4.0/4.0 **May `22**
Worcester Polytechnic Institute, Worcester, MA, USA

Bachelor of Technology, Information Technology, 3.3 / 4.0 **May `19**
Anna University, Chennai, TN, India

WORK EXPERIENCE:

Data Engineer, Eficens Systems LLC, Houston, Texas **June `22 - Present**

- Coordinated with BI developers on automating workflows to transform and load data into the Snowflake data warehouse for live Tableau dashboard integration.
- Designed Transformation DAGs workflow in Airflow using SQL stored procedures for extracting, merging, transforming, and loading data into the warehouse.
- Created, maintained, and optimized multiple STAR schema databases in the Snowflake data warehouse and loaded the data from different sources.

Research Assistant, Worcester Polytechnic Institute, Worcester, Massachusetts **May `21 - May `22**

- Engineered *DOSGAN, GMM-UNIT, and STARGAN-V2* generative models for Synthetic Time Series data generation for Human Context Recognition (HCR) using Pytorch
- Evaluated the quality and diversity of the synthetic data generated by the models with Kullback–Leibler (KL) Divergence and Frechet Inception Distance (FID) score. The best model attained a low KL divergence and FID Score of **0.41 and 2.11**.
- Improved the performance of Human Context Recognition Classifier model by introducing Synthetic data in the training phase by **24%.**

Data Analyst, Market Simplified India Limited, Chennai, India **Jun `19 - Jul `20**

- Tracked application performance issues using *Firebase Crashlytics* and developed Tableau dashboards to visualize the reports.
- Developed targets and designed promotional campaigns for Banking applications as a part of the in-house campaign management team.
- Performed *A/B testing* for banking campaign on student loans towards the target audience and measure the success with a number of visits to the websites and other parameters. Observed a **15% increase** in website visits and other parameters regarding the specific student loan campaign.

ACADEMIC PROJECTS:

Emotion Recognition in Human Voice, Machine Learning **Aug `20 - Dec `20**

- Extracted and preprocessed audio handcraft features such as Periodogram, Intensity, Fundamental Frequency, and HNR from .mfcc files and formed them as Time series data.
- Divised Neural Network with Gated Recurrent Unit (GRU) to predict the human emotions from the audio features and achieved an Accuracy of 78% on the multi-class classification of 6 human emotions
- Analyzed the performance of the same model for a Binary classification and achieved an Accuracy of 81%.

Scalable COVID-19 analysis using Big Data, *Big Data Management* ***Jan '21 - May '21***

- Built multiple distributed and scalable map-reduce jobs in Hadoop for a dataset with 10 million data points containing personID and coordinates of seats.
- Calculated and displayed personIDs close contact with infected people using distance-based computation using SparkSQL.
- Applied SparkContext to retrieve data from the HDFS and implemented a clustering algorithm to find the number of clusters in the concert using PySpark MLlib.

Movie Prediction Rating System, Statistical Learning **Sep `20 - Dec `20**

- Preprocessed 450,000 records from multiple sources in JSON format and merged them into a single holistic dataset.
- Implemented, Trained, and Tested *Support Vector Machine (SVM), Random Forest, and XGBoost* using 3-fold Cross Validation in R.
- Accomplished an **F1 score of 0.75** on the XGBoost as the best performing and **0.41 F1 score** on SVM as least performing.

SKILL:

Languages:	Python, R, SQL, SAS, Java, C++, React-Native
Visualization:	Tableau, Power BI, Matplotlib, Seaborn
Database:	MySQL, SparkSQL, PostgreSQL, Oracle SQL, MongoDB, SQLAlchemy, Snowflake
Technologies:	Firebase Crashlytics, Hadoop, Apache Spark, Anaconda, Tensorflow, Pytorch, PySpark, Pandas, Numpy, AirFlow, Docker, Azure, AWS, Kubernetes, MLFlow.
Analytics & ML:	Classification, Regression, Classification, Boosting algorithm, Time Series Forecasting, CNN, RNN, NLP, Predictive modeling, Web Analytics, Hypothesis Testing, Data Augmentation, Big Data Management, Big Data Analytics