**Sai Praharsha Devalla**

315-832-0961 | [sdevalla@syr.edu](mailto:sdevalla@syr.edu) | [www.linkedin.com/in/sai-praharsha-devalla](http://www.linkedin.com/in/sai-praharsha-devalla) | <https://github.com/saipraharsha>

**EDUCATION**

**Master of Science in Data Science,** Syracuse University (4.0/4.0) **May 2021**

**Bachelor of Technology in Electronics & Communications**, KL University (3.9/4.0) **May 2017**

**TECHNICAL SKILLS & PUBLICATIONS**

**Publications:** “Ionospheric scintillation forecasting model based on NN-PSO technique” paper published in journal “Astrophysics and Space Science”.

**Languages:** Python (Sklearn, PyTorch, keras, matplotlib), RShiny, Spark (MLlib), SQL, R (caret, ggplot2, tidyverse), C/C++

**Concepts**: EDA, Statistics, Machine Learning & Deep Learning (Regression, Naïve Bayes, Decision Trees, Random Forests, GBTs, SVM, K-Means, Neural Networks), ETL, Data Modeling, Data Warehousing, Dashboarding, Data Visualization

**Tools**: MS Excel, Tableau, Visual Studio, Alteryx, Oracle, Power BI, MATLAB, MS SQL Server, AWS Services, Snowflake, Hive

**EXPERIENCE**

***Data Scientist, iConsult Collaborative, Syracuse University*** **Dec 2020 – May 2021**

* Partner with a Crouse Hospitals to build **Machine Learning models** for **binary classification** of patients.
* Created a serialized **Alteryx** **Data Pipeline** to execute **ETL** & **Data Mining** operations to predict heart stroke for patients which reduced manual workload by 40% monthly and provided insights based on which treatment plans can be adapted.
* Designed multiple **Tableau** Dashboards to discover and identify trends of patients who are diagnosed with the disease.

***Graduate Teaching Assistant,* *Syracuse University* May 2020 – May 2021**

* Collaborated in developing a management system for university recreation center that enables easy tracking of available services in recreational center and allows students to reserve the time slots for available services.
* Formulated and deployed a serverless, scalable architecture using **AWS services** like Lambda, S3, DynamoDB to support real-time updates of facilities management system.

***Software Engineer at Infosys Ltd, Bangalore***  **June 2017 - July 2019**

* Built data integration/engineering workflows on **Big Data technologies** and platforms (Hadoop, Spark, Hive, S3).
* Analyzed structured, semi-structured, and unstructured data by executing complex **PySpark** scripts.
* Developed data pipelines using **Spark framework** to store information into **AWS S3** layers, **Hive** database and **Snowflake**.
* Implemented statistical techniques for hypothesis testing to validate data attributes for building **tableau** dashboards.
* Re-architect the current data foundation to simplify data landscape. Automated the workflow using job scheduler **Airflow** and boosted the productivity by 30%.
* Engineered a digital wallet using Object oriented programming in **Python** and integrated with **SQL** Server which handles money transfers, movie bookings. Developed a functional design of movie booking module.

**ACADEMIC PROJECTS**

***Online News Popularity* | Python: pandas, matplotlib | Apache Spark: PySpark, MLlib | Machine Learning | Statistics**

* Performed classification modeling whether an article is going to be a hit before publishing based on feature selection.
* Engineered Random Forest, Gradient Boosting and Logistic Regression models to classify whether the article published by Mashable will be hit. Random forest was the best performing model with an AUC of 76%.

***Will this Car Be A Good Buy?* | Python: sklearn, pandas, matplotlib | R: caret, ggplot2, tidyverse | R Shiny | ML**

* Designed a web application (**RShiny** Dashboard) containing insightful visualizations and a ML prediction module.
* Programmed CRISP-DM methodology to construct ensemble Machine Learning models (**GBM & Random Forest**), to help auto dealers accurately predict if a car is going to be resold after buying from auction. [[Web Application](https://bhavishkumar.shinyapps.io/project_dashboard/)]
* Assessed model performance and fine-tuned Machine Learning models to maximize Accuracy & Recall.

***Retail Sales Forecasting* | Python: PyTorch, pandas, matplotlib | Deep Learning | ML**

* Aided a retail organization accurately forecast weekly sales at a Store-Department level using Deep Learning models.
* Devised Data Preprocessing & developed multivariate time series forecast models by training Long Short-Term Memory (LSTM) & Gated Recurrent Unit (GRU) Neural Networks on processed data, by utilizing PyTorch library.
* Trained Linear & Random Forest Regression models to compare performance with LSTM & GRU models.

***Data Warehouse for Business Intelligence* | SSIS: Fact and Dimensional tables | OLAP Data Sources | Power BI**

* Merged to two fictitious companies by designing an Integrated Star Schema Data Warehouse.
* Performed **ETL** operations using SSIS packages to populate the Fact & Dimension tables of warehouse with data from the 2 OLAP data sources. Also, a user interactive **Power BI** dashboard for Business Intelligence is devised.

**LEADERSHIP**

* Mentored as graduate teaching assistant to 40+ students by conducting one-one sessions and office hours for “Intro to Data Science” course, deals with Machine Learning concepts.