Prudhvi Vajja

v.prudhvi0001@gmail.com
LinkedIn | GitHub | Website
+1(812) 650 8597

Education

Master of Science in Data Science **May 2021**
Indiana University, Bloomington GPA: 3.78/4.0
Coursework: Machine Learning, Applied Algorithms, Statistics, Advance Database Concepts, Exploratory Data Analysis, Artificial Intelligence, Computer Vision, Cloud Computing.

Bachelors in Electronics and Communications **Aug 2014 – May 2018**
Jawaharlal Nehru Technological University, Kakinada, India GPA: 8/10

Technical Skills

Languages: Python, R, Scala. **Frameworks:** AWS-Sage Maker, GCP, Git, Apache Spark, Docker, Hadoop, TensorFlow, PyTorch, scikit-learn, etc. **Skills:** Machine Learning, Algorithm Design, NLP, CV, Visualizations, Data Processing, Statistical Analysis. **Web:** Django, Streamlit, Flask, HTML, REST architecture, Socket. **Databases:** SQL, NoSQL, Postgres, MongoDB, Redis. **Hobbies:** Blogging, Sketching, Reading Books, Sports.

Experience

Mesh Labs *(Indiana University, Bloomington)* **Nov 2019 – Present**
Research Assistant – Python, Web Applications, Open Source.
- Coordinated with Professor(RandyHeiland) and his research associates in developing open-sourced Jupyter notebook based web applications for NanoHub.org and did unit testing/validation on them.
- Increased website traffic by 10% by generating interactive visualization plots and improving UI/UX interface.

Research Assistant **(SPEA)** – R, HPC, Shell Scripting, Excel. **Aug 2019 – Nov 2019**
- Performed exploratory data analysis on a dataset with 1 million rows and identified the factors affecting opioid and narcotic overdoses across the US.
- Implemented a pipeline with 5 stages (preprocessing the data to model analysis) using shell scripting and R.

Tata Consultancy Services *(New Delhi, India)* **Nov 2018 – Jun 2019**
Data Analyst – Python, SQL, ETL, Tableau.
- Developed several ETL's to seamlessly load data from multiple sources to DataMart's using Informatica designer 8.6.
- Slashed the batch runtimes by 40% by optimizing complex SQL queries using relational algebra methods.
- Created interactive dashboards with quick filters and workflows for report scheduling in Tableau.

Extracurricular and Projects

Quantium Data Analytics Virtual Experience Program *(Virtual) Certification* **Sep 2020 – Oct 2020**
Data Analyst – Python.
- Conducted analysis on the client's transaction dataset and identified customer purchasing behaviors to generate insights and provide commercial recommendations.
- Extended analysis to help identify benchmark stores that allow you to test the impact of the trial store layouts on customer sales and prepare a report for the client.

Distributed Map-Reduce & Memcached-lite *[Code]* [GCP, gRPC, Python, Socket, Multi-Threading]
- Designed and deployed a fault-tolerant highly scalable cloud-based MapReduce library on GCP for processing and generating big data sets in parallel by dynamically spawning mappers and reducers.
- This Architecture can be used by higher-level applications for big data processing using a well-defined REST API.
- Implemented a Memcached based Key-Value storage to save intermediate files and final data from mappers and reducers (microservices) running separately on multiple VMs.

Explorer *[Code], [App]* [Python, Streamlit, Heroku, Ensemble]
- Streamlined an end to end web application to preprocess, visualize, and perform predictive analysis on user data.
- Integrated tools such as grid search, confusion matrix, and ensemble methods to increase the performance of the models by >20%.

Audio Denoising & Speaker Identification *[Code]* [Deep Learning, CNN, RNN, LSTM,]
- Implemented CNN, RNN, LSTM architectures for speech denoising and compared their SNR ratio *(LSTM = 19dB).*
- Trained a convolutional Siamese network with a contrastive loss on the STFT representations of audio from a subset of the Vox-Celeb dataset on AWS.

Kaggle and Academic projects *[*Python, TensorFlow, Kaggle, BERT]
- Twitter-Disaster tweet analysis [**code**], Parts of speech-Tagging [**code**], Disease classification in plants [**code**]
- Implemented a heuristic page rank algorithm using MapReduce on the google web graph dataset.