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

Skills

Languages & DB : Python, R, Scala, PostgreSQL, Redis, Hadoop.
Web Frameworks : Django, Streamlit, Flask, HTML, REST architecture, Apache Spark.
ML Frameworks : AWS-Sage Maker, TensorFlow, PYTorch, Scikit-Learn, Jupyter Notebooks, Git, Spark, Jira.
Statistics : A/B Testing, ANOVA, Hypothesis testing, Cross-Validation, Chi-Squared, Etc.
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(RH) and 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.

School of Public and Environmental Affairs *(Indiana University, Bloomington)* **Aug 2019 – Nov 2019**

Research Assistant – R, HPC, Shell Scripting, Excel

- Performed exploratory data analysis on dataset with 1 million rows and identified the factors effecting opioid and narcotic overdoses across 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 form 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.

Personal Projects

Explorer *[Code], [App]* [Python, Streamlit, Heroku, Ensemble]

- Streamlined an end to end web application to preprocess, visualize and perform predictive analysis of user data.
- Integrated tools such as grid search, confusion matrix and ensemble methods to increase the performance of the models by >5%.

Statistical Analysis of Heart Disease *[Code]* [R, ggplot, Tidy verse]

- Discovered that subjective features are the main cause of heart disease in US by binning features into objective, examination, subjective and applying multiple regression and hypothesis testing.
- Performed predictive analysis with 95% accuracy by modelling a linear regression (LOESS) to capture the non-symmetrical trends in the dataset.

Twitter Disaster Analysis *[Code]* [Python, TensorFlow, Kaggle, BERT]

- Extracted meta-features from tweets using Lemmatization, TF-IDF, and N-gram techniques to differentiate the meanings of similar words for disaster and non-disaster tweets.
- Finetuned BERT & Glove models to improve the accuracy from 92% to 97% and attained a rank - 651/2500 in Kaggle.

Disease Classification in Plants *[Code]* [Python, TensorFlow, Kaggle, Res-Net]

- Tripled 70% imbalanced dataset to generate balanced data using SMOTE and augmentation and increased the model speed by 40% using batches.
- Ensembled Resnet, Efficient net, EN-Noisy student and improved the model AUC score to 0.948.

Academic Projects *[Code]* [PostgreSQL, Spark, Hadoop]

- Implemented a heuristic page rank algorithm using MapReduce on Google web graph dataset.
- Performed key-value stores (MapReduce & Spark) on nested and graph database using PostgreSQL.