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

Bachelors in Electronics and Communications **Aug 2014 – May 2018**

Jawaharlal Nehru Technological University, Kakinada, India GPA: 8/10

Coursework: Machine Learning, Applied Algorithms, Statistics, Advance Database Concepts, Exploratory Data Analysis, Artificial Intelligence, Computer Vision, Cloud Computing.

Skills

Languages & DB : Python, R, Scala, PostgreSQL, Redis.
Web Frameworks : Django, Streamlit, Flask, HTML, REST architecture.
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 **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 **Aug 2019 – Nov 2019**

Research Assistant – R, HPC, Shell Scripting, Excel

- Worked on large datasets to clean and analyze the factors impacting opioid and other narcotic overdoses across the country through Exploratory data analysis and hypothesis testing.
- Implemented a data pipeline to automate the process of data cleaning and model analysis using shell scripting and R.

Tata Consultancy Services **Nov 2018 – Jun 2019**

Data Analyst – Python, SQL, ETL, Tableau.

- Developed several ETL's to seamlessly load data and programmatically analyze to discover business insights.
- 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.

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]

- By binning features into objective, examination, subjective and performing multiple regression and hypothesis testing found that subjective features are main cause of heart disease in US.
- A linear regression (LOESS) model was developed to capture the non-symmetrical trends and perform predictive analysis of data.

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

- Extracted meta features from tweets using Lemmatization, TF-IDF and N-gram techniques to differentiate 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]

- Implemented a data pipeline for data preprocessing, augmentation and applied SMOTE to generate balanced data from a 70% imbalanced data.
- 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.