Jayanth Dasamantharao

NJ - 08901 +1 (848)-313-5617 dasamantharao.jayanth@rutgers.edu
linkedin.com/in/JayanthDasamantharao github.com/JayanthDasamantharao

EDUCATION

Rutgers University, *Masters in Data Science* *Sept 2022 - May 2024 | New Brunswick, NJ*

Relevant Coursework: Statistics: Statistical Modeling & Computing, Stat Learning; Neural Networks, Machine Learning, Data Structures & Algorithms(Python), Data Wrangling (R), NLP, AI, Data Mining. Projects Overview: Portfolio.

Andhra University *2017 - 2021 | Visakhapatnam, India*

Bachelor of Technology in Electrical and Electronics Engineering

Research Paper: Self Driving Cars Using Image Processing and Deep Learning.

TECHNICAL SKILLS

Data Engineering: SQL, MySQL, SQLite, SQL Workbench, Data Warehousing, Postgres, MongoDB, Oracle DB, PySpark, Hadoop, EDA. **Data Visualization:** PowerBI, Tableau, Plotly, Shiny, Matplotlib, Seaborn, Data Wrangling.

Deep Learning & Machine Learning (ML): Clustering, NLP - NLTK, Keras, Hugging Face Transformers, spaCy, LLM's, NER, Computer Vision, Recommender Systems, Language Models, CNN, RNN, Transformers, LSTM, BERT, GPT.

Programming: Python, R, Scala **Tools & Frameworks:** Scikit-learn, Bokeh, TensorFlow, PyTorch, Numpy, Pandas, Dash, Altair, SAS - Proc ANOVA, Proc SQL, Proc GLM, Pearson correlation, SAS macros.

Cloud Services: AWS Athena, Amazon Lambda, Redshift, Amazon S3, RDS, RedCap, Dynamo DB, GCP: BigQuery, Google Cloud Storage, Google Cloud Functions, Google Cloud Dataproc.

PROFESSIONAL EXPERIENCE

Harvest Software Solutions, LLC, *Data Science Intern* *June 2023 - Sept 2023 | Jacksonville, Florida*

Real-time Ad Campaign Optimization:

- **Real-time Data Extraction using API's:** Extracted data from Aniview and CMCBeacons via APIs. Managed 5M+ rows, performed EDA, and incorporated CPM for critical performance evaluation.
- **Data Analysis:** Utilized visualization, summary stats, and hypothesis testing for insights into ad campaign performance.
- **Predictive Modeling:** Implemented **XGBoost for predictive modeling** with low MSE. Applied **5-fold cross-validation to assess and prevent overfitting**. Leveraged H2O's AutoML for automated model selection, optimizing CPM predictions.
- **Automation and Techniques:** Applied **AutoML for automated tasks**, such as feature engineering and hyperparameter tuning, to select the optimal model for CPM prediction. **Techniques employed:** XGBoost, K-means, BERT, plotly, matplotlib, NLTK, spacy, pandas, numpy, ARIMA.

Accenture Solutions Pvt. Ltd., *Application Development Associate* *June 2021 - Aug 2022 | Hyderabad, India*

Database Analyst - Regeneron Inc:

- **Data Analysis:** Utilized **Athena and Redshift for client data analysis**, identifying two-thirds of integrity exceptions and improving process efficiencies. Implemented **ETL for efficient data handling**, reducing support calls by 50%
- Designed Python **pandas data pipelines to automate scripts**, resulting in a 50% reduction in support calls and ensuring efficient storage and maintenance of customer campaigns.
- Led a cross-functional **Agile development team** of 3, driving improvements in data visualization.
- **Customer Data Warehousing:** Conducted **business analysis for customer data warehousing**, resulting in improved data warehousing, enhanced data accuracy, and a significant reduction in data processing duration.
- Formulated **SQL queries using SQL Workbench** to meet business needs, thereby enhancing data quality and precision.
- Implemented technological solutions, including DDL updates and the **deployment of data into production servers**.

RESEARCH PROJECTS

[Probabilistic Graphical Models, Bayesian Networks, PyMC3, Informative Prior Distributions]
Developed Bayesian logistic regression models using PyMC3 to predict diabetes status, incorporating informative priors. Evaluated performance with recall and f1-score metrics, comparing outcomes between uniform and normal distribution priors. *Bayesian Classification on Diabetes dataset.*

[Hugging Face transformers, Naive Bayes, Logistic Regression, BERT, N-Grams, GloVe embeddings]
Led Language Identification employing Naive Bayes, Logistic Regression, and DistilBERT, achieving an 89% baseline accuracy globally. Implemented DistilBERT, enhancing precision to 94%, and incorporated n-grams analysis for a more nuanced understanding of linguistic patterns. *Language Identification.*

[ML models - Random Forests, KNN, Logistic Regression, SVM, Decision Tree, and XGBoost]
Executed diverse ML strategies and time series analysis for auto insurance fraud, prioritizing reduced false negatives. Chose K-Nearest Neighbors (KNN) with an impressive 97% recall. *Auto Insurance Fraud Detection.*