Bharath Velamala

Tucson, Arizona, 85719 | (520) 543-9800 | bharathvelamala@arizona.edu | LinkedIn | GitHub

EDUCATION

University of Arizona, Tucson   Aug 2023 - May 2025
Master of Science, Data Science, **GPA 4.0/4.0**
Coursework: Deep Learning, Natural Language Processing, Data Mining, Machine Learning, Data Visualization, Big Data
Indian Institute of Technology, Patna   Aug 2017 - May 2021
Bachelor of Technology, Mechanical Engineering, **GPA 8.0/10.0**
Coursework: Data Science, Python Programming, Time Series Analysis, Statistical Analysis

TECHNICAL SKILLS

Programming : Python, R Programming, Java, SQL, C++, MATLAB
Packages   : Pandas, NumPy, Matplotlib, Tensorflow, Scikit-Learn, PyTorch, transformers, ggplot2, R Shiny, PySpark
Tools   : Git, Google Cloud, AWS, Apache Beam, Apache Spark, Apache Airflow, BigQuery, Tableau, Terraform,
                          Hadoop, Kafka, Pub Sub, Dataproc, Data Fusion, Cloud Storage, Cloud Functions, Docker, SnowFlake, Jira
Domain Skills : Machine Learning, Exploratory Data Analysis, Scalable ETL pipelines, Big Data, Statistical Modeling
Certifications : **Google Cloud Certified** Professional Data Engineer, **Google Cloud Certified** Associate Cloud Engineer

PROFESSIONAL EXPERIENCE

Biostatistician Intern (Data Science) - Mayo Clinic, Rochester, MN   May 2024 – Present
- Created medspacyV, a user-friendly graphical tool for non-programmers, reducing setup time for clinical NLP pipelines by 60% and improving workflow efficiency, leading to faster deployment and analysis of patient records.
- Engineered custom NLP components, including a tokenizer and sectionizer, boosting performance by 30% and enabling precise keyword identification in medical records.
- Led the integration of Excel-based rule editing, improving rule configuration efficiency by 40%, and increasing accuracy in detecting clinical concepts by 25%.

Data Engineer - HSBC, Hyderabad, India   Jan 2023 - Aug 2023
- Spearheaded full-scale data migration project data migration using Python and ETL pipelines to transition financial decision-making applications from on-premise to Google Cloud, achieving 100% migration success.
- Boosted downstream analytics by providing real-time access to transactional and account-level data via BigQuery views, orchestrated with Google Cloud Composer and Apache Airflow, improving decision-making efficiency by 30%.
- Collaborated with upstream and downstream teams to resolve over 50 data migration bugs using Control-M for scheduling and SQL queries for debugging, improving data consistency and reducing bug resolution time by 25%.

Associate Data Engineer - SpringML Inc., Hyderabad, India   Jan 2021 - Dec 2022
- Processed large volumes of diverse data (2M+ records) by developing scalable ETL pipelines using Google Cloud Dataflow and Apache Beam, reducing execution time to under 8 minutes, improving processing efficiency by 50%.
- Employed the Dataproc service to execute large-scale ETL pipelines using PySpark, optimizing data processing workflows and achieving a 60% reduction in processing time while increasing data throughput by 40%.
- Orchestrated an automated data warehousing workflow in BigQuery by creating materialized views with SQL queries, and developed interactive dashboards in Looker and Data Studio, enhancing operational visibility by 40%.
- Implemented Cloud Functions for event-based triggers, cutting application waiting time by 50% and enhancing performance, while ensuring 100% compliance with PII data privacy regulations through data masking protocols.
- Contributed to the development of an automated supply chain analytics platform integrating SAP IBP API with Google Cloud Data Fusion, resulting in 35% faster data retrieval and enhanced reporting capabilities.

PROJECTS

Multilabel Image Classification | *Python, PyTorch, Sklearn, Pandas, NumPy, CV2, PIL, Matplotlib*   GitHub
- Engineered and optimized multi-label image classification models, achieving over 90% average validation accuracy by addressing class imbalance and boosting the model's F1-score by 20% through hyperparameter tuning.
- Generated and analyzed comprehensive evaluation metrics, including precision, recall, and F1-score, offering insights into predictive capabilities across 40 attribute classes, and achieving an average F1-score of 0.85.

Sentiment Analysis from Text | *Python, NLP, Sklearn, NLTK*   GitHub
- Executed sentiment analysis using Logistic Regression, Linear SVM, and Naive Bayes, securing a 0.91 leaderboard score in a Kaggle competition while conducting thorough exploratory data analysis and text preprocessing.

From Takeoff to Touchdown Visualizing Data on Air Disasters | *R, Python, TensorFlow, ggplot2, Quarto, R Shiny*   GitHub
- Developed an R Shiny application to visualize patterns in aviation incidents in the US, leveraging TensorFlow to summarize 90% of the text data in reports, thereby identifying probable causes of aviation incidents.