Harish Nandhan Shanmugam

hash1366@colorado.edu | (720)261-7297 | Boulder, CO, United States
GitHub: HarishNandhan | LinkedIn: Harish Nandhan Shanmugam | Portfolio Website

Technical Skills

Languages: Python, R, C, SQL, NOSQL-MongoDB
Libraries: Numpy, Pandas, Matplotlib, Seaborn, Summarytools, SweetViz,Scikit-learn, OpenCV, Keras, Tensorflow, Pytorch, Pyspark
Technologies/Frameworks: Cloud - (AWS & Azure), ETL Operations, CRM (Salesforce), Agile, Microsoft Office Suite(Excel & Powerpoint), Dashboarding(PowerBI & Tableau), Flask, Docker, LLMs, RAG, Generative AI, Version Control(Git), Data Analytics, Business Understanding, Market Research Analysis, Risk Analysis, Data Modeling, Data Mining, Predictive Modeling, Statistical Analysis
Certifications: Git,Flask- LinkedinLearning, Statistical Learning - NPTEL, HTML - Infosys Springboard

Education

University of Colorado Boulder — Boulder, Colorado
Masters in Data Science — *August 2024 - May 2026*
GPA: 4/4

Anna University — Chennai, Tamil Nadu
Bachelor of Technology in Artificial Intelligence and Machine Learning — *June 2020 - May 2024*
GPA: 3.94/4

Experience

LetsGrowMore - An MSME Startup — Jasra, India(Remote)
Data Science Intern — *June 2023 - July 2023*

- Constructed an Admission Predictor Module by analyzing scores in Physics, Chemistry, and Mathematics, which effectively streamlined the admission prediction process for Tamil Nadu Engineering Admissions.
- Orchestrated the extraction of datasets from **AWS S3** and conducted a comprehensive automated EDA using **Summarytools** and **SweetViz**, resulting in actionable insights which enhanced the model performance by **25%**.
- Built and Deployed a prediction model utilizing KNN and Random Forest Classifier, which achieved a **94.3%** accuracy, providing reliable predictions for college admissions and addressing critical decision-making needs for a student.

Plumb5 Analytics (A Unit of Decisive Analytical Systems) — Bengaluru, India
Market Data Analyst Intern — *January 2023 - April 2023*

- Designed a custom email template builder inspired by Beefree.io, which **doubled** the company's email campaigning efficiency and reduced dependency on external tools, leading to significant cost savings of **133$** per month.
- Performed a comprehensive Competitor analysis using **Microsoft Excel** to identify gaps compared to Amberity and Treasure Data, and recommended actionable strategies to strengthen the company's competitive advantage.
- Developed an interactive **PowerBI** dashboard to visualize campaign data, enabling the marketing team to identify errors and optimize campaign strategies, resulting in a **40%** improvement in decision-making efficiency.

Projects

Task Insights and Workload Overview Analysis for DEN — Python, SQL, PowerBI, Pandas, Excel

- Conceptualized a data-driven approach to monitor task insights and workload distribution, catering to **Denver International Airport** employees' operational needs.
- Formulated interactive dashboards in **PowerBI**, enabling users to interact with real-time performance metrics and workload trends, saving **15 minutes** per analysis session.
- Extracted and organized data using **Python** and **SQL** to generate precise, actionable insights for better decision-making.

LLM Models Performance Comparison Analysis — Python, Excel, EDA, PowerBI, Huggingface

- Streamlined dataset size from **700k to 91k** records by applying advanced data cleaning techniques and removing redundancies, ensuring a high-quality dataset for analysis and enhancing processing efficiency.
- Resolved class imbalance in the 'gpt 4o' model using the **SMOTE** approach, improving the representation of minority classes and boosting model performance in underrepresented scenarios.
- Analyzed hypothesis testing results and created **PowerBI** visualization dashboard to uncover the most impactful parameters, providing meaningful insights into model performance for key attributes such as 'Code Generation' and 'Problem Solving.'
- Secured a multiclass classification **AUC score of 95%** by crafting a RandomForestClassifier to predict the best-performing LLM based on task-specific attributes like language preference and detail orientation.

BMI Prediction using Facial Images — OpenCV, VGG16, ResNet, Tensorflow

- Generated a predictive model by integrating advanced neural networks with **ResNet** and **VGG16** to improve sex prediction accuracy to 99% and BMI prediction accuracy to 86%.
- Examined with classmates' photos to increase accuracy using the Polk County prison dataset for model training; with Asian samples, accuracy was 94%.

Publications

- "A Deep Learning Approach for Non - Invasive Body Mass Index Calculation" in Communication in Computer and Information Science Springer Book Series (International Conference on Advances in Artificial Intelligence and Machine Learning in Big Data Processing)
- "Unmasking Pneumothorax: Deep Learning Strategies for Precise Classification and Segmentation" in IEEE Xplore (2nd International Conference on Artificial Intelligence and Machine Learning Applications, AIMLA – 2024