HITESH NARAYANA

Los Angeles, CA 90007 • +1 (213)68-12369 • hiteshna@usc.edu • linkedin.com/in/hiteshnarayan/ • github.com/Hit07

EDUCATION

UNIVERSITY OF SOUTHERN CALIFORNIA — **JAN 2024 – Expected DEC 2025**
Master of Science in Computer Science — ***GPA: 3.5/4.0***

Coursework: Analysis of **Algorithms**, **Database Systems**, **Machine Learning,** Information Retrieval

BANGALORE INSTITUTE OF TECHNOLOGY – India — **AUG 2019 – JUL 2023**
Bachelor of Technology in Information Science and Engineering — ***GPA: 9.0/10.0***

Coursework: Data Structures and Algorithms**,** Machine Learning, Artificial Intelligence, Big Data

SKILLS

Languages: Python (Proficient), **SQL**, C++, R **Machine Learning**: **PyTorch**, **TensorFlow**, **Keras, Scikit-learn**, NLTK **Libraries: Pandas**, **NumPy, Matplotlib,** Seaborn, Statsmodels **Big Data: Apache Spark**, Hadoop, Dash **Cloud:** AWS (S3, Glue), Google Cloud, Microsoft Azure **Technologies:** Git, RESTful APIs, **CUDA**, Flask, **Power BI,** Tableau, DAX

PROFESSIONAL EXPERIENCE

Research Assistant – HUMANS Lab (HUmans MAchines Networks Society) | ***Web Scraping, ETL, Apache Spark, Scikit-learn.*** — **JUL 2024 – Present**

- Collaborated with a team of researchers to explore the effects of social media discourse on public opinion and hate in the upcoming 2024 election led under **Prof. Emilio Ferrara.**
- Developed a ETL pipeline that streamlined the analysis of over **5M+ tweets**, enabling targeted insights on 'X' discourse related to voter sentiment during the election cycle.
- Implemented a distributed computing framework using **Apache Spark**; increased data processing **efficiency by 30%,** resulting in a significant reduction in project turnaround time, with tasks completed 20% faster than previously recorded
- Executed predictive analytics frameworks using time series methodologies to analyze sentiment trends and forecast user behaviour and sentiment shifts by extracting features from X data.

Business Analyst – Genpact, India | ***SQL, Power BI, DAX, ETL*** — **JUL 2023 – DEC 2023**

- Optimized SQL queries by rewriting complex joins and indexing critical tables; achieved a **30% reduction** in query runtime which enhanced data retrieval speed for **5+ reports** generated weekly.
- Created a fully automated reporting solution utilizing Power BI and DAX, which enabled real-time monitoring of payroll trends, increased decision-making capabilities and reducing turnaround time for **report delivery by 40%.**
- Facilitated to AWS Glue ETL pipelines by writing scripts for serverless data transformation, improving data availability and reducing operational overhead.

PROJECTS

FoodVision: Multiclass Image Classification using TinyVGG | ***Python, PyTorch, CUDA, torchvision, TensorFlow*** — **JUN 2024 – AUG 2024**

- Built an **deep learning** image recognition system for food classification, handling **500K+ data points**; integrated feedback loops that improves learning efficiency, resulting in a rapid adaptation to new food items within the existing database.
- Engineered a custom CNN model (TinyVGG) that surpassed the baseline accuracy by **20%,** enabling precise categorization of diverse food items and improving classification speed by **15% during testing phases**.
- Improved model performance through hyperparameter tuning techniques with **TensorFlow and Keras**, achieving a **15%** increase in accuracy while cutting training time by **40%** using automated scheduling tools for enhanced experimentation efficiency.

Predictive Modelling Using Multivariate Regression | ***Pandas, Scikit-Learn, Matplotlib, Seaborn, EDA*** — **APR 2024 - MAY 2024**

- Developed a predictive algorithm for residential property prices in 1970s Boston; model facilitated accurate pricing strategies for real estate evaluations in historical contexts.
- Constructed a multivariate regression model utilizing scikit-learn, validating predictive capability with an **R-squared score of 79.3%** on training sets and achieving a commendable **74.47% on unseen test sets**.
- Implemented **logit transformation** techniques on datasets, resulting in a **20% increase in model accuracy**, which minimized prediction errors.
- Engineered data manipulation processes using Pandas to clean and prepare datasets exceeding 5,000 rows, enhancing analysis efficiency and reducing preparation time by 40%.

Childbed Fever Analysis: Statistical Analysis | ***Python, Pandas, , SciPy, StatsModel*** — **APR 2024-MAY 2024**

- Conducted comprehensive statistical analysis on **20+** historical medical records to evaluate the impact of handwashing on childbed fever mortality rates.
- Designed data visualization techniques to create time series plots, kernel density estimates, and box plots, communicating complex mortality trends based on statistical outcomes from analysis involving over 5,000 historical medical cases.
- Performed comprehensive analysis of historical datasets using statistical methods; identified key trends leading to a notable reduction of 60% ($p < 0.0000003$) in deaths attributed to childbed fever after intervention implementation.

Data [Exploration, Cleaning, Manipulation, Visualisation] Analysis | ***NumPy, Scikit-Learn, SciPy, Statsmodels*** — **MAR 2024–JUL 2024**

- Spearheaded a comprehensive data science project that involved detailed analysis, exploration, visualization, and **time-series forecasting** across 15 diverse datasets; improved model accuracy by identifying key patterns in the data.
- Conducted time-series forecasting and anomaly detection, utilizing statsmodels and SciPy, to uncover inefficiencies and create data-driven stories based on historical and current information.