

GIRISH WANGIKAR

(+1) 984-944-3688 | Raleigh, NC / gwangik@ncsu.edu / [linkedin.com/in/girish-wangikar](https://www.linkedin.com/in/girish-wangikar/) / github.com/GirishWangikar | [Portfolio](#)

EDUCATION

North Carolina State University, Raleigh

Aug 2022 - May 2024

Master of Science, Electrical and Computer Engineering

GPA - 3.92

Coursework: Topics in Data Science, Advance Machine Learning, Automated Learning and Data Analysis, Pattern Recognition, Neural Networks and Deep Learning, Object Oriented Design and Development

Rajarambapu Institute of Technology, Maharashtra, India

Jul 2018 - Jul 2022

Bachelor of Technology, Electronics and Telecommunication Engineering

GPA - 3.83

TECHNICAL SKILLS

Programming Languages : Python, SQL, R, MATLAB, C++, DAX

Frameworks/Libraries : Scikit-learn, Pandas, PyTorch, TensorFlow, Keras, OpenCV, Seaborn, NumPy, Matplotlib

Tools : PowerBI, Apache Spark, Jupyter Notebook, Git, Docker, Kubernetes, AWS, Tableau, BigQuery, EDA, Jira

Additional : Effective Data Visualization, Data Wrangling With Attention to Detail, Proficiency in Handling ETL Processes, Predictive Modeling and Forecasting, Knowledge of Machine Learning Algorithms (XgBoost, RandomForest, Logistic Regression, SVM)

PROFESSIONAL AND RESEARCH EXPERIENCE

Data Science Co-Op

Chaska, Minnesota

Entegris

Jan 2024 - Present

- Implemented an end-to-end **Generative AI** Q&A system on **VertexAI (GCP)** using the **PaLM TextBison LLM** model and **LangChain**. Leveraged vector embeddings and vector search techniques to retrieve answers from BigQuery data, eliminating manual data sifting.
- Created and deployed a text summarization tool using the **Google Gemini 1.0 Pro LLM**, incorporating web scraping for text extraction, **prompt engineering**, and natural language processing (NLP) techniques to generate summary PDFs.
- Developed predictive model using the **SARIMA** technique and **time-series** historical data for time series analysis and forecasting, enabling managers to set quarterly targets and facilitating faster decision-making processes and increased work efficiency.

Research Intern

Raleigh, North Carolina

Sozzani Lab

Jun 2023 - Dec 2023

- Developed and implemented **LSTM RNN** models for multi-variate genomics **time-series** data, enhancing predictive accuracy over previous methods by over 10%; conducted data analysis of **30,000+** genomics samples using **SQL** to be harnessed in further research.
- Involved in the addition of an **AutoEncoder** framework tailored for **predictive modeling** of Differentially Expressed Genes (DEG's).
- Documented algorithms and methodologies in a **research paper** submitted for publication to serve as a reference for other researchers.

Data Science Intern

Maharashtra, India

Exposys Data Labs

Jan 2022 - Mar 2022

- Performed **ETL** on **500GB** of **unstructured data** to develop **ARIMA** model for forecasting car sales with an accuracy of 85%.
- Collaborated with **cross-functional** teams to translate data insights into actionable recommendations and created **PowerBI** dashboards, presenting findings in a simplified way to non-technical stakeholders for data-driven decision making.

ACADEMIC PROJECTS

- Deep Learning - Brain MRI Segmentation | Python (Jupyter Notebook, PyTorch)** Aug - Nov 2023
 - Implemented U-Net deep learning for precise glioma segmentation in MRI scans, achieving a **0.9466 Dice similarity** coefficient.
 - Conducted statistical analysis on 110 patients, revealing associations between shape features, genomic subtypes, and emphasizing the role of AI-driven radiomics in advancing personalized cancer treatments.
- Generative AI - Deepfake Images and Video Detection Algorithm | Python (TensorFlow, PyTorch)** Feb - May 2023
 - Generated **120,000** fake images from **CelebA** real image dataset for different Generative Adversarial Networks (GANs).
 - Leveraged **Amazon S3** for centralized storage of 120,000 fake images, ensuring data integrity and **version control**.
 - Utilized **Amazon EC2** GPU instances to expedite model training, with **Accuracy of 99.37%** and **F1 score of 98.97%**.
- Sensor Data Analysis - Terrain Identification using LSTM | Python, SQL (TensorFlow)** Jan - Feb 2023
 - Applied **Bidirectional LSTM** to identify terrain by making use of accelerometer and gyroscope measurements.
 - Achieved a notable **89.3% accuracy** on the test set and collaborated with a team member from the Department of Statistics at NCSU.
- Computer Vision - 2-D Object Detection for Autonomous Vehicle | Python (Keras, TensorFlow)** Nov - Dec 2022
 - Utilized YOLOv3 to train a Machine Learning model on **10,000+ images** for Object Detection in Autonomous Vehicles.
 - Compared the performance with other algorithms using model Frames Per Second (FPS) and mean Average Precision (mAP).

CERTIFICATIONS & EXTRACURRICULARS

- Winner of the Machine Learning track at the 3rd annual N.C. **PSI Hackathon** as a member of team of four students
- Google Cloud Training** – Data Engineering, Big Data and Machine Learning on GCP Specialization Certificate
- NVIDIA** - “Deep Learning Fundamentals”, “Image Segmentation Techniques”, “Time-Series Data Modeling with RNN”
- Co-authored the research paper “**Diabetes Detection - An Application of Machine Learning in Healthcare Industry**”
- Generative AI with Large Language Models**