

Ranjan Ravi

ranjan7796@gmail.com | [LinkedIn](#) | [Portfolio](#) | [GitHub](#) | (813) 729-1906 | San Francisco, CA

PROFESSIONAL SUMMARY

Experienced Machine Learning Engineer with **4 years** of expertise in **Python, SQL, Natural Language Processing, Data Science, Machine Learning, and Data Visualization**. Experience working with **Large Language Models** such as GPT, BERT, MPNet, and T5 in building an **information retrieval** application with Custom Ranking Algorithms. Hands-on experience in **building ML pipelines** from Data Collection to Model Deployment.

WORK EXPERIENCE

Data Science Intern, Bayer, New Jersey

June 2022 – March 2023

- Developed a pipeline for quick information retrieval for medical professionals, utilizing a **strong encoder model** to retrieve the most relevant PubMed articles for a given question. Integrated with **ChatGPT**, the pipeline provides effective answers by passing contexts along with the question.
- Built a **semantic search tool** allowing users to upload and access information from their corpus.
- Designed a **custom ranking algorithm** for search results, incorporating cosine similarity scores and fuzzy-matching distances. Wrote unit and integration tests for the API using Pytest.
- Assisted in **deploying ML models (T5 and BERT)** on Amazon **Sagemaker**.

Research Assistant and Tableau Mentor, University of South Florida, Tampa, Florida

November 2021 – June 2022

- Involved in ongoing research of parsing student resumes using **Named Entity Recognition**. Developed a spacy annotator, an interface to quickly label entities for NER using ipywidgets which is used to generate training data.
- By working as a Teaching Assistant, I supported Undergraduate students in understanding the significance of data-driven decision-making and problem-solving using **Tableau**.

Data Scientist, TRIMBLE Inc, Chennai, India

November 2020 – August 2021

- Built an efficient **time series forecasting** model for the Civil Construction division of Trimble that reduced **MAPE and RMSE by 24%** with the Facebook Prophet algorithm. Managed end-to-end data flow by designing and **implementing an ETL process** by loading the revenue data from various sources into the Snowflake data warehouse **using Python**, ensuring high data quality and accuracy for time series analysis and reporting purposes.
- Developed dashboards with monthly and quarterly revenue predictions using **Tableau**. Constructed an ML pipeline by deploying the model in AWS Sagemaker and implementing its inference function via AWS Lambda.

Associate Software Engineer, BNP PARIBAS, Chennai, India

June 2018 – November 2020

- Involved in an Agile software design and development of Reporting platform for generating tax reports for clients.
- Written **PL/SQL** procedures and SQL scripts for implementing the tax calculation based on the Double Tax Treaty.

EDUCATION

Master's in Business Analytics and Information Systems, University of South Florida, Tampa, Florida

May 2023

Bachelor's in Computer Science and Engineering, Madras Institute of Technology, Anna University, India

April 2018

TECHNICAL SKILLS

- Proficient in **Python and R** programming and its frameworks such as **FastAPI, Flask, Streamlit**, and libraries such as **PyTorch, TensorFlow, Keras, NLTK, SpaCy, Gensim, Scikit-Learn, Pandas, Numpy, Scipy**, and OpenCV. Experienced in analyzing text data using Natural Language Processing and building NLP models.
 - NLP Albuementation:** Explored various data augmentation techniques for text data and evaluated the impact of NLP albuementation on the performance of a classification model. [GitHub](#)
 - Text Analytics:** Conducted text analysis on the inaugural speeches of all US presidents and performed Topic Modelling to uncover common themes in each speech. [GitHub](#)
 - GitHub Analytics:** Built a **Classifier** to identify potential candidates based on their GitHub profiles in order to minimize the time and effort the recruiters put in during the recruitment process. | [GitHub](#)
- Good knowledge of Machine Learning Models such as SVM, Naive Bayes, Decision Forests, KNN, Ensembles, Neural Networks (CNN, RNN), NLP, Clustering, word2vec, doc2vec, and Web scraping. Built,
 - Sentiment Analyzer:** A big data project to capture the sentiment of E-Commerce user reviews stored in a **Hadoop** File System using Apache **PySpark**. [GitHub](#)
 - Linear Regression Model and Web Scrapping using Real Estate Data:** Designed and built a web scraper program to scrape house data from a real estate website and built a Linear Regression model | [GitHub](#)
- Experienced in developing time series forecasting models utilizing algorithms such as Facebook Prophet, Neural Prophet, ARIMA, Exponential Smoothing, and LSTM.
- Comfortable in analyzing live data such as Covid-19 and having good knowledge of **AWS, GCP**, and its services such as Sagemaker, EC2, S3, and Lambda. Good exposure to **Docker, Kubernetes, Git**, and **Snowflake**. Built,
 - Covid-19 Tracker:** Developed a web-based application using the **Python Dash** framework, featuring an interactive chart displaying daily trends of confirmed, recovered, and deceased coronavirus cases. [GitHub](#)
- Experienced in working with SQL and NoSQL databases such as **Oracle, Postgres, MS SQL, MongoDB, SQLite**.
- Experienced in Analytical & Reporting Tools like **Tableau, Power BI, Looker**, Jasper Reports, Microsoft Azure Machine Learning Studio, MS Excel, and Python visualization libraries like **Plotly, Matplotlib**, and Seaborn.