PURUSHOTHAM PARTHY

7828824002 | purushothamparthy@gmail.com | 5222 Green St, Halifax, NS B3H 1N7 | LinkedIn | GitHub

EDUCATION

Master of Applied Computer Science, Dalhousie University, Halifax, NS   *January 2024 - May 2025*

Bachelor of Technology (Computer Science), VIT University, Vellore, India   *July 2016 - May 2020*

TECHNICAL SKILLS

Languages   : Python, Java, JavaScript, SQL, C++, HTML, CSS

Frameworks & Libraries : Scikit-Learn, Keras, Pandas, NumPy, SciPy, TensorFlow, Matplotlib, OpenCV, Spark, Hadoop

Databases   : MySQL, PostgreSQL, Neo4j, MongoDB

Cloud Technologies   : Sagemaker, Comprehend, BedRock, Quicksight, Serverless

Concepts   : Machine Learning Algorithms, Deep Learning, Data Wrangling, Data Visualization, Statistical Analysis

Certifications   : AWS Certified Solutions Architect (AWS SAA)

PROFESSIONAL EXPERIENCE

Data Scientist | *Virtusa Consulting Services Pvt Ltd*   *January 2021 – January 2024*

Context Based Search| *An enterprise solution for efficient information retrieval.*   *July 2023 – January 2024*
- Developed an advanced **Question Answering** tool using open-source **LLM Llama2**, resulting in a 30% reduction in information retrieval time for employees.
- Used techniques like **in-context learning** to fit the organization's specific context, enhancing data precision by 25% in official document searches.
- Collaborated with cross-functional teams to identify key data extraction requirements, improving user satisfaction by 40%.

ColBERT| *A customized BERT algorithm enabling faster and better information retrieval.*   *April 2023 – June 2023*
- Collaborated with **Stanford AI Labs (SAIL)** enhancing information retrieval accuracy by 40% for medical datasets for clients.
- Used **text classification models** to clean and extract medical texts from over 1 TB of passage data from the **MS MARCO Dataset**.
- Implemented the ColBERT algorithm using python into a **Question Answering Prototype**, improving reducing response times by 30%.

SpyGlass – Johnson and Johnson| *Analytics solution for reducing unauthorised information release.*   *June 2022 – March 2023*
- Developed a Python-based **NLP solution**, reducing unauthorized information release incidents by 40%.
- Engineered a robust data pipeline with **AWS Sagemaker and Comprehend Medical**, enhancing data processing speed by 30%.
- Implemented **medical entity recognition** techniques that improved data extraction accuracy resulting in more reliable clinical trial data management.
- Designed a real-time analytics dashboard using **AWS Quicksight**, enabling stakeholders to access critical insights instantly.

Insurance Claims adjudication | *Solution to automate insurance claim-cost matching.*   *January 2021 – May 2022*
- Implemented **Entity recognition** to extract CPT codes, increasing data extraction accuracy by 35% in insurance claim forms.
- Automated the code and cost matching, reducing manual verification time by 50%.
- Utilized **PySpark** to process and index large datasets efficiently, reducing the adjudication processing time.

Project Vnet | *A no Code platform for developing an end-to-end ML pipeline.*   *February 2021- January 2022*
- Setup pipelines for a dynamic **ML development platform** allowing users to deploy models 20% faster than traditional methods.
- Developed modules for **model training & testing** using frameworks like **Scikit-Learn**, **Keras**, and **PyTorch,** reducing setup times by 50%.
- Implemented **PCA** and components of **data transformation modules**, enabling users to preprocess the data.
- Integrated **data visualizations** module using **Plotly** enhancing user experience and enabling data-driven decisions.

ACADEMIC PROJECTS

Information Retrieval Solution | *An enterprise solution for efficient information retrieval*
- Designed a lighter and more secure alternative to cloud based LLM services for information retrieval using python and RAG architecture.
- Integrating **Cohere-LLM model** which is open source with **Chroma vector DB** optimized for 100% on Prem deployment.
- Implemented **Document chunking** and **Embedding** using LangChain Framework.

RideSync | *A Ride Sharing app for trips and daily commuting*
- Developed a **full-stack application** using **Java Spring Boot, React** and **MySQL** as Database. Implemented **RESTful APIs** in backend following the **MVC** architecture.

Nutrifit | *An innovative All in One Fitness application*
- Built **full-stack** web application using **React.js, Node.js, Express** and **MongoDB**. Created **RESTful APIs** with Node.js and Express for efficient data retrieval and server-side validations.