

# Rajit Subin Puzhakkarezhath

✉ [rajitsp@gmail.com](mailto:rajitsp@gmail.com) ☎ +13122167563 🔗 [linkedin.com/in/rajit-subin-puzhakkarezhath/](https://www.linkedin.com/in/rajit-subin-puzhakkarezhath/) 🌐 [rajitsp.github.io](https://rajitsp.github.io)

## EXPERIENCE

### Machine Learning Engineer

*Squark, MA*

Oct 2023 - Current

- Implemented Retrieval Augmented Generation(RAG) with an **LLM**, translating realtime logs into comprehensible English. Utilized LangChain for integration, executing **ETL** processes to split documents, transform embeddings, and store in FAISS vector database.
- Spearheaded library updates, resolving dependencies conflicts for stability in a Python virtual environment. Enhanced code integrity by seamlessly integrating logging, model version control, debugging, error tracing, and exception handling for optimized service operation.
- Collaborating to implement **multivariate time series analysis**, employing algorithms such as Varmax and LSTM, covering preprocessing, denoising, outlier handling, automodeling, and prediction in an AWS environment with EC2 instances and S3 buckets.

### Software Engineer Intern (Data Science R&D)

*CCC Intelligent Solutions, Chicago*

June 2022 – Apr 2023

- Engineered a **multi-label image classification pipeline** feeding multiple images as a single tensor to a Vision Transformer (ViT) model attaining **97.8% accuracy** on 100k instances.
- Developed a **multi-modal transformer recommendation system** for predicting damaged auto parts, including **hypothesis testing** for result validation. Utilized GPU cluster model training with **Docker** and **Kubernetes**, monitoring via ClearML and Airflow.
- Created a **3D visualization** by **reducing the dimensionality** of 700+ dimensional embedding from a custom **transformer architecture** on **Tensorboard** projector to demonstrate correlation between damaged parts of a vehicle.
- Collaborated in a hackathon, leveraging a **GPT-based open source LLM** model to retrieve vehicle repair method documents, while also working cross-functionally with various back-end and front-end engineering teams.

### Associate Data Analyst

*BLR Logistiks*

Aug 2019 - May 2021

- Analyzed 10,000+ shipment and inventory records using Spark and Python, improving efficiency by 30%.
- Developed interactive **PowerBI dashboards** incorporating **SQL** queries for data visualization to gather trends from data enabling stakeholders to make business decisions.

### Data Science Intern

*Aditya Birla Group, Mumbai*

June 2019 - July 2019

- **Led** the process of **extracting, compiling, formatting, restructuring, and validating** data to ensure quality for designing a sales forecast model for a specific plant's upcoming month's sales in a region, utilizing the past 4 years' data.
- Built a foundational understanding of business priorities through cross-project collaborations while implementing novel machine learning methods. Incorporated non-linear regression, Bayesian methods, and LSTM Neural Network using TensorFlow, alongside Pandas and NumPy for data pre-processing, achieving a 71% accuracy.

## PROJECTS

### Liver Cirrhosis Prediction Analysis

- Analysed **healthcare data** from 400+ Mayo Clinic trial patients to build a liver cirrhosis prediction model. Employed stratified k-fold cross-validation, comparing Logistic Regression and **XGBoost** models and conducting feature analysis.
- Demonstrated model effectiveness with 74% accuracy and 0.74 ROC AUC using an optimized XGBoost model. Additionally, illustrated **feature importance** through insightful plots, enhancing predictive understanding in healthcare data analysis.

### MLOPs end-to-end pipeline using AWS

- Implemented end-to-end MLOps pipeline using **AWS SageMaker** to build and deploy an **CatBoost** model for detecting faulty states in wind turbines. Conducted **exploratory data analysis, model training**, performed model **evaluation**, and **deployed** it to an endpoint.
- Achieved **95% precision** in training and 93% in test datasets, leveraging AWS services such as **SageMaker Studio, S3 buckets, EC2 instances, CloudWatch** and prebuilt containers.

### Amigo –Smart Voice Controlled Bot

- Engineered a bot for multiple functions: image captioning, home automation, speech recognition, **sentiment analysis** via textual, facial, and audio information using concepts of **computer vision, NLP, and deep learning**.
- Utilized the **Random Forest Classifier** for speech emotion recognition, **Xception net** architecture to recognize facial emotions, the **resnet-152** model for the encoder and **LSTM** network for the decoder network and developed a GUI using Tkinter.

## EDUCATION

**Master of Science in Computer Science** | *University of Illinois at Chicago (UIC), Illinois*

2023

Coursework: Artificial Intelligence (AI), Cloud Computing, Advanced ML, Deep Learning for NLP, Data Mining

**Bachelor of Engineering** in Electronics and Telecommunication Engineering | *University of Mumbai (DJSCE), India*

2021

Relevant Coursework: Data base management system, Image Processing and Machine Vision, Big Data Analytics

## SKILLS

- **Languages, Databases, Software, OS:** Python, Scala, Java, HTML | **SQL** | VSCode, Git, Jupyter, MATLAB | Linux, Windows
- **Data Science:** NumPy, Pandas | Data Visualization (Tensorboard, Matplotlib, Tableau) | Regression, Classification, Clustering
- **Machine Learning:** Scikit-learn, SciPy | Deep Learning (PyTorch, TensorFlow, Keras) | NLP (NLTK, Transformers, Huggingface)
- **Cloud and Big Data:** AWS (EC2, EMR, S3, Lambda, SageMaker), Docker, Kubernetes, Apache Hadoop and Spark, Map-Reduce