Rajit Subin Puzhakkarezhath

rajitsp@gmail.com +13122167563 linkedin.com/in/rajit-subin-puzhakkarezhath/ rajitsp.github.io

EDUCATION

Master of Science in **Computer Science** | *University of Illinois at Chicago (UIC), Illinois* 2023
Coursework: Artificial Intelligence, 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

EXPERIENCE

Software Engineer Intern (Data Science R&D)

CCC Intelligent Solutions, Chicago **June 2022 - Present**

• Engineered a pipeline to feed multiple images as a single instance of tensors for a multi-label multi-class image classification problem. Modelled a Vision Transformer (ViT) using the above pipeline on 100k instances to attain a 97.8% accuracy.
• Developed a multi-modal custom transformer-based recommendation model to predict the number and type of damaged parts. Evaluated the model using a separate pipeline to ensure its effectiveness.
• Created a 3D visualization by reducing the dimensionality of 700+ dimensional embedding using t-SNE and PCA from a custom transformer architecture on Tensorboard projector to demonstrate correlation between damaged parts of a vehicle.
• Collaborated in a 3-day hackathon to leverage a GPT based LLM model to retrieve vehicle repair method documents.

Coding Department Co-Head

DJS Antariksh **Jan 2020 - Dec 2020**

• Built object recognition models using EfficientDet and YOLOv4 architectures and achieved an accuracy up to 82%.
• Participated in the European Rover Challenge (ERC) and won the 'Best Science Task' award. Secured 'Third' place in the world for all the tasks and design report in September 2020.

Data Science Intern

Aditya Birla Group, Mumbai **June 2019 - July 2019**

• Designed a sales forecasting model to predict the sales of a particular plant in given region on a specific date in the upcoming month based on previous 4 years of data.
• Researched and implemented novel machine learning techniques including LSTM Neural Network using TensorFlow as well as Pandas and NumPy in python for data pre-processing and obtained an accuracy of 71%.

PROJECTS

MLOPs end-to-end pipeline using AWS

• Implemented end-to-end MLOps pipeline using AWS SageMaker to build and deploy an XGBoost model for detecting faulty states in wind turbines. Conducted exploratory data analysis, trained the model, evaluated its performance, 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 to accomplish multiple functions such as object detection, image captioning, home automation, speech recognition, sentiment analysis through textual, facial, and audio information using concepts of 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.

eBay Delivery Date Prediction

• Constructed a model for the machine learning competition of eBay to estimate the delivery date of shipments of online purchases on real world dataset of 15 million records of C2C and B2C orders.
• Conducted extensive data pre-processing and transformation using pandas and swifter and modelled quantile regression and catboost to achieve an r2-score of 79%.

Log File Monitoring and Alert System

• An Apache Spark based log file processor that sends automated alerts to stakeholders based on log severity and frequency for cloud monitoring and troubleshooting.
• Consists of an Akka actor system deployed on an Elastic Kubernetes Service cluster with Redis database for persistent storage. Uses Amazon's Managed Streaming for Apache Kafka to perform inter-process communication and AWS SNS to send email notifications.

Sentiment Analysis using DistilBERT

• Applied a state-of-the-art BERT (Bidirectional Encoder Representations from Transformers) based model to analyze and classify sentiments in the electronics reviews from the Amazon review dataset.
• Sampled and cleaned data from the imbalanced raw dataset, tokenized the text and finetuned the distilled version of BERT on 60K reviews to achieve an accuracy of 93.13%.

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