Aravind S Data Scientist

aravind9722@gmail.com 9710082049 Bangalore, India https://github.com/aravind-selvam
https://www.linkedin.com/in/aravind-selvam/

PROFILE

Data Scientist with one year of experience in delivering successful full-scale projects. Having hands-on expertise in machine learning, deep learning, and transfer learning techniques. With a track record of delivering results, Highly motivated and has excellent problem-solving skills with deep understanding of technologies such as **Machine Learning** and **NLP**, and capable of utilizing them to drive business value and achieve desired outcomes.

SKILLS

Python | **Deep Learning** (PyTorch | TensorFlow | PyTorch Lightning) | **Database** (MongoDB | SQL) | **Linux** (Bash)

Natural Language Processing (Text Classification | NER | Summarization | Transformer | BERT | GPT | RNN | LSTM)

Machine Learning (Scikit-learn | XGBoost | Dask) | **MLOps** (Docker | MLFlow | Github Actions | CircleCI | Prefect)

AWS (EC2 | ECR | S3) | **GCP** (GCE | GCR | BigQuery) | **REST API** (Flask | FastAPI) | **Front End** (HTML | CSS)

PROFESSIONAL EXPERIENCE

Junior Data Scientist, *iNeuron Intelligence Pvt. Ltd.* 07-2022 – present | Bangalore

- Implemented an in-house chatbot for a financial firm using Rasa and Dual Intent Entity Transformer (**DIET**), resulting in a 30% reduction in onboarding time and 20% increase in employee productivity. Deployed on **GCP**.
- Developed a video transcript summarization tool using Whisper and **BART**, improving video content consumption and user engagement on the company's website.
- Utilized NLP **keyword extraction** to process customer feedback, resulting in a significant improvement in customer satisfaction.

Data Science Intern, *iNeuron Intelligence Pvt. Ltd.* 04-2022 – 07-2022 | Bangalore

- Implemented a live data visualization dashboard to monitor key performance indicators (**KPI**) for a restaurant franchise, resulting in improved decision-making through data-driven insights.
- Performed **data cleaning** and preprocessing for a machine learning project focused on **text classification**.

PROJECTS

Video Summarization, *STT and text summarization (POC)*

- Developed a web application for video summarization using **Flask** framework and Transcript, summarize, and video downloader components.
- Used Open AI's **Whisper** for transcription and **Fine-tuned BART** model for summarization.
- Achieved the goal of making it easy for users to understand the content of videos with minimal effort, by implementing it on the company's website.
- Implemented **CI/CD** using GitHub actions and deployed the web app on **Amazon** Elastic Compute Cloud (**EC2**).

Language Identification using CNN, *Audio Classification*

- Applied audio preprocessing techniques and utilized **CNN** architecture for image classification using **mel-spectrogram** images.
- Trained the model with a large Indian language dataset, resulting in a 90% accuracy in language classification.
- Used **PyTorch**, **Torch Audio**, and ffmpeg to develop the model.
- Implemented CI/CD with **Circle-CI** and deployed the web application on **GCP Compute Engine**.

ML Pipeline Standardization, *Pipeline (POC)*

- Implemented a standardized machine learning pipeline with 5 components to streamline future projects at iNeuron.
- The pipeline includes data ingestion, data transformation, data validation, data evaluation, and model training and deployment.
- Adopted **MLFlow** Experiment Tracker for improved experiment comparison and efficient parameter tracking.
- Stored processed data in **MongoDB** for quick and effective access, increasing data retrieval efficiency.
- Securely stored pipeline artifacts in an **S3 bucket** for reliable retrieval and protection against data loss.

EDUCATION

B-Tech : Biotechnology, *Sathyabama University* 2014 – 2018 | Chennai, India

12th Std, *Velammal Matriculation Higher Secondary School* 06-2013 – 05-2014 | Chennai, India