

Arthi Murali

9940627763 · meetarhi@gmail.com · meetarhi.com

linkedin.com/in/meetarhi · github.com/meetarhi · medium.com/@meetarhi · kaggle.com/arthimurali

As a skilled Python Programmer with a strong foundation in bioinformatics, I specialize in data extraction, transformation, and automation. I build efficient data pipelines using SQL, MongoDB, and tools like Apache Airflow and AWS. My experience extends to machine learning, enabling me to apply analytical techniques and predictive models to derive actionable insights from complex datasets. I am also proficient in developing user-friendly interfaces with Streamlit and Chainlit, and excel in data manipulation and analysis using Pandas. I'm eager to leverage my technical expertise to drive innovation and deliver impactful results in data-driven environments.

WORK

Heyram infrastructure

ML engineer Intern

During my internship, I developed an interactive learning application using Streamlit, hosted on Hugging-Face Spaces, for middle school children. It featured entity recognition, parts of speech tagging, image-to-text, and audio-to-text conversion powered by Spacy, Tesseract, and OpenAI APIs

iMorph

Product Development Intern

During my product internship tenure at iMorph, I was responsible for prototyping and building micro products. I led the initiative for python assessment bot which is based on openai assistant used to screen candidates and tech news aggregator tool that is used for building datasets.

TALKS AND PRESENTATIONS

Talk on Disease Prediction using CNNs

Presented at ChennaiPy, covering neural network fundamentals and CNN structure for medical data predictions. Demonstrated a code example predicting lung cancer using a Kaggle CT scan dataset. Engaged in discussions about the potential of CNNs for early disease detection and the challenges in sourcing real medical data.

EXPERIENCE / PROJECTS

1. YouTube Data Harvesting and Warehousing

The YouTube Data Harvesting and Warehousing app enables users to access and analyze data from multiple YouTube channels. Built with Streamlit and hosted on Hugging Face Spaces, the app uses the YouTube Data API to extract channel, playlist, video, and comment data. After users input a YouTube Channel ID, data is retrieved and stored in MongoDB. Users can then select a channel to migrate data to SQL for further analysis. SQL queries enable viewing results in tables or Dataframes, helping users understand channel performance, engagement, and trends over time for strategic content planning.

Technology - Python, python-youtube, pandas, MySQL, MongoDB, Streamlit, SQLAlchemy, Visual Studio, JupyterLab

2. Phonepe pulse visualization App

Developed a visualisation app for analyzing PhonePe's digital payment data. Extracted and transformed JSON data from GitHub into CSV format, organising it into Dataframes by year, quarter, and category. Merged dataframes for simplified retrieval and built dynamic visualizations, including graphs, charts, and maps, to display transaction trends, payment categories, and user activity across India. The app provides valuable insights for market analysis and geolocation-based performance tracking. Deployed on Streamlit Community Cloud for broad accessibility.

Technology - Python, Streamlit, Pandas, Plotly Express, Matplotlib, Geopandas, and Plotly Graph Objects, Visual Studio, Jupyterlab

3. Tech news aggregator tool

Tech News Aggregator is a tool that gathers news articles from various technology news sources onto a single platform. It scrapes articles from a list of RSS feeds in a text file, parses key information, and stores it in a database for users to easily search and consume tech news content. A command-line interface enables user to add and delete website urls from the text files, the url will fetch rss feed using requests.

Technology - click, feedparser, sqlite3, Requests, Visual Studio, Command-Line Interface (CLI)

4. AWS Lambda leveraged Server Monitoring and Slack Notification System

Designed an Automatic Data Collection and Storage System with AWS Lambda and Slack Integration for Server Availability Monitoring and Slack Notification using AWS Lambda, CloudWatch, Slack API.

Technology - Python, AWS Lambda, CloudWatch, Slack API, psycopg2, AWS SNS, Slack workspace

MINOR PROJECTS

1. Forecasting whether the patient has Diabetes based on symptoms from Pima Indian Diabetes dataset, using comparative analysis of **machine learning** algorithms
 2. Amazon privileges access prediction based on the employer role using **H2O AutoML**.
 3. Tweet's scrapping app to extract tweet data from User-Specified Details.
 4. Backend API for E-commerce Website with User Registration, Product Rating, Sorting using **Flask**.
 5. Predicting the need of specific target drug for a patient from Longitudinal Record using **Xgboost**.
 6. Innovative AI Playground for Interactive Learning in Students using **OpenAI API** features and **SpaCy**.
 7. Deep learning model for automated tuberculosis detection using **CNNs**.
 8. Automated Startup Summarisation Tool for Investor Insights Using **Language Models**
 9. Python Proficiency Assessment Bot for Evaluating Skills Using **Chainlit** and OpenAI API
-

SKILLS

Data Engineering

- Pandas - Data manipulation, analysis, and handling
- Cloud Computing (AWS) - S3, EC2, Lambda, Glue, Athena, CloudWatch, IAM
- Databases - SQL, MongoDB (Compass), AWS(DynamoDB, RDS)
- Big Data & Data Processing Tools - Hadoop, Airflow, PySpark, Prometheus, Grafana
- DevOps Tools - Docker, Cron Jobs, Git

Data Science

- Language models - Falcon, llama, OpenAi
 - Statistics - Standard Deviation, Covariance, Correlation, A/B Testing
 - Data Visualisation and analysis- Numpy, Matplotlib, Seaborn, Plotly
 - Machine Learning - Bayesian method, Decision Trees, Random Trees, XGBoost, Support Vector Machine(SVM), KNN, Principal Component Analysis (PCA), Clustering, NLP.
 - Deep learning - Convolutional Neural Network (CNN)
 - Recommender Systems - User-Based Filtering, Item-Based Filtering
 - AutoML - H2O, TPOT
 - Platform and Tools - HuggingFace, Pickle, Gradio, Pandas Profiling
-

EDUCATION

Master Data Engineering

GUVI, Chennai

Bachelor of Science Honors

Biomedical Sciences

Sri Ramachandra Institute of Higher Education and Research, Chennai

Higher Secondary Education

Kumararani Meena Muthiah Matric Higher Secondary School, Chennai

Secondary Education

Kendriya Vidyalaya CLRI, Chennai

CERTIFICATION

- Machine Learning, Data Science and Deep Learning with Python
- The Complete Data Structures and Algorithms Course in Python

LANGUAGES

- English - Full Professional Proficiency
- Hindi - Limited Working Proficiency
- Tamil - Native / Bilingual