(+91)-77080-02179
mailmeatrajagopal@gmail.com

# Rajagopal S

Data Scientist

linkedin.com/rajagopal-s
github.com/rajag0pal

## Experience

**Data Scientist** (Intern) **Jan 2023 — May 2023**
Microbiological Laboratory Research and Services (I) PVT LTD Coimbatore

- Led the development of an AI-based framework that utilizes **Machine learning and Deep learning** algorithms to interpret results from PCR run files, resulting in a 2x faster turnaround time for result delivery.
- Established a **feature store** (relational database), for efficient data management of features extracted from PCR run files.
- Spearheaded the implementation of an **GUI-based automation application** to streamline data extraction, **resulting in a 50% reduction** in turnaround time.
- Developed a web application, that performs data upload and retrieval from SQL database, with features such as **Visualization**, **Feature Extraction**, and **Report Generation**. **Reduced data analysis time from hours to minutes** by incorporating ML model for signal classification.

## Projects

**PyHRM (Open Source)** *pypi.org/PyHRM* **Apr 2023 — Present**

- Implemented a feature extraction module that will enable the classification of PCR products based on their melting curve shape, improving the accuracy rate by 30%.
- Engineered and released a Python library for processing PCR products' melting signals, resulting in 50% faster signal processing and feature extraction.
- Features like **melting temperatures**, **take-off** & **touch-down** points of melting signal (Temperature at which peak start rising and temperature at which peak falls down),**peak prominences**, and **area under the curve** are the outputs of PyHRM.

**An AI-based framework and data-driven methodology for post-PCR High-Resolution Melting Analysis (HRMA),**
M.Sc Thesis, Department of Computer Applications, Bharathiar University **Jan 2023 — May 2023**
*github.com/FEUSION*

- Conceptualized a framework for classifying pathogens with the HRM data extracted from PCR run files.
- Executed **data pre-processing** through a **vision-based** technique, to eliminate noisy signals from the analysis, using **CNN**, for classifying genuine and non-genuine peaks.
- **Generated image data** of melting signals from various PCR run files, and constructed a data set for **multi-class classification** (0- Noisy signal, 1- Signal with One peak, 2- Signal with Two peaks).
- Employed TensorFlow for building a dense neural network to learn patterns and features extracted from the signals using PyHRM, to classify positive **Meningitis** pathogens.

**ML-based web application for predicting Bank customer Churn using Deep Learning,**
M.Sc mini project, Department of Computer Applications, Bharathiar University **Aug 2022 — Dec 2022**

- Leveraged pre-existing datasets from Kaggle to train a deep neural network model in Tensorflow, due to the lack of original customer data in India.
- Scrutinized the data for analysis, including pre-processing and cleaning, and provided insights on the data through **Exploratory Data Analysis** and created an **interactive dashboard** using **Power BI** for visualization and data comparison.
- Integrated a web application using HTML, CSS, and **Flask** to predict customer churn based on user input. Achieved an accuracy of 85% in predicting customer churn through the developed application, that can be helpful to identify potential customer churn risks and improved decision-making for banking organizations.

**N-Gram Modelling for E-Commerce Product Reviews & Classification with Transformers pipeline,**
Self-interested project **Jul 2022 — Aug 2022**

- Generated a data set of reviews by scrapping over **5000 product reviews** from an e-commerce website using **Beautiful Soup (bs4)** and Selenium WebDriver in Python.
- Conducted data cleaning, feature engineering, and text pre-processing techniques such as **punctuation removal**, **stopwords removal**, **lemmatization**, **stemming**, and **word count**.
- Constructed a corpus, vocabulary, and **bigrams/trigrams**, utilizing Bayesian rules to predict the next probable word.
- Utilized **Transformers-pipeline** (Python) for sentiment analysis, enabling labeling of positive and negative reviews.

**Image to Hashtags,**
Tiny project on pre-trained models **May 2023 — May 2023**

- Developed an AI tool that identifies trending hashtags for images based on the content of the image.
- Employed **vit2-gpt2-image-captioning** pre-trained model from hugging face repository to caption images given by the user.
- Leveraged **openai's gpt-3.5-turbo** model to generate hashtags for the captions generated.
- Built an interface using gradio, and made users to upload any images or URL, to generate hashtags for the given image and deployed in hugging face spaces *huggingface.co/spaces/ImagetoHashtags*

## Education

**Master of Science, Data Analytics,** Bharathiar University, GPA: 8.34/10.00 **Oct 2021 — May 2023**

## Skills

| | |
|---|---|
| **Languages** | Python, R, C++, SQL |
| **Libraries & Tools** | Pandas, Numpy, Matplotlib, Seaborn, Plotly, Scipy, Tkinter, Customtkinter,Pyautogui, Sckitlearn, TensorFlow, Keras, NLTK, Flask, beautiful soup, sqlalchemy, PowerBI, Tableau, GIT, Github |
| **Data Wrangling** | Data generation, Data Extraction, Data Cleaning, Exploratory Data Analysis, Feature Engineering, Feature selection, Data visualization. |
| **ML Skills** | Data Modeling, Clustering & Classification, Quantitative Analysis, Regression, Predictive Modeling, Statistical Modeling, Model Validation, Model deployment, CNN & RNN, LSTM, BERT, Transformers. |

## Competitions

- Participated in Smart India Hackathon 2022, and submitted ideation for the problem stated by **Ministry of Micro, Small & Medium Enterprises (MSME) - RK1125 "Uncontrolled Growth of Water Hyacinth"**
  - Gathered real-time images of ponds with water hyacinth across the city and used them to train a CNN model that can detect such ponds.
  - Acquired satellite images of ponds where the growth of water hyacinth is highly visible and fed them into the model.
  - Examined various sensors to identify the attributes/qualities of water bodies where water hyacinth growth is likely to occur.

## Certifications

- Hacker Rank - Python (Basic)
- Hacker Rank - SQL (Basic)
- NPTEL - Machine Learning, ML
- UiPath - Introduction to RPA and Automation
- Coursera - What is Data Science, IBM
- Infosys - Soft Skills

## Blogs

medium.com/RajagopalS

## Protfolio

To explore more on my projects please visit: https://rajag0pal.github.io