Pranava Kailash Subramaniam Prema

Data Scientist / Data Engineer

Email: pranavakailashsp@gmail.com | Mobile: +447909807017 | linkedin.com/in/pranava-kailash | pranava-kailash.github.io

PERSONAL STATMENT

As an enthusiastic Data Scientist, I love solving complicated problems with data. With practical experience in **machine learning**, **web scraping, data analytics**, and **LLMs** I take every chance to make processes easier and automated for the sake of efficiency. Familiar with **Python, MySQL, NLP, TensorFlow, and AWS**, applying oneself in an energetic team is highly welcome since such a pro drives effective data-driven decisions.

SKILLS

- **Technical Skills**: Proficient in Python, MySQL, SQLite, TensorFlow, Scikit-learn, and AWS. Experienced in automation, data scraping (Selenium, Beautiful Soup), and version control (Git, GitHub)
- **Data Science & Development:** Expertise in data analysis, machine learning, NLP, and building ETL pipelines. Skilled in API development (Flask, FastAPI) and deploying solutions on cloud platforms (AWS, Docker)
- **Soft Skills:** Strong communication, teamwork, problem-solving, and time management skills. Adaptable and growth-oriented with leadership experience

WORK EXPERIENCE

Data Researcher | Fleet Street Research, UK | Remote, India **October 2022 - January 2024**

- Automation of **data extraction** with Python-based web scraping scripts, reducing manual entry time for this data by **50%**
- Improved data processing **workflow by optimizing internal Python libraries**, improving operational efficiency by up to **10%**
- Smoothened the process of **image importation** into DB, hence enhancing data quality
- Migrated key datasets using databases like **SQLite and MySQL**, without any record or incident of data loss and/or noticeable down times
- Ensured efficient database transition, with the **minimum generation of bugs** and ensured system performance

Data Science Intern | Yoshops.com | Remote, India **February 2022 - April 2022**

- Automated **data collection** through web scraping tools, increasing real-time pricing accuracy by **40%** and improving product performance insights
- Built machine learning models for price prediction, enhancing forecasting accuracy by **20%** through data analysis and market trends
- Developed a **90%** accurate machine learning model for osteoarthritis detection using **VGG16**, optimizing feature extraction in medical applications

Data Science Intern | Forsk Coding School| Remote, India **March 2021 - June 2021**

- Conducted in-depth **sentiment analysis** on customer reviews using **natural language processing (NLP)**, finding key areas for product improvement
- Built and deployed a scalable machine learning model on **AWS for sentiment analysis**, ensuring high availability and real-time insights from large datasets
- Leveraged statistical and machine learning techniques to extract **business insights** from unstructured datasets, providing valuable recommendations to stakeholders

EDUCATION

MSc in Data Science | University of Surrey, Guildford, London 2024 - 2025

BE in Computer Science and Engineering, First Class (9.5 CGPA) | SNS College of Technology, India 2019 - 2023

PROJECTS

CyNER2.0 – Cybersecurity Named Entity Recognizer | University of Surrey

- Developed a domain-specific **NER model** using **transformer-based models** (**BERT, DeBERTa, DarkBERT**) to identify critical cybersecurity entities, including malware, vulnerabilities, and threat actors.
- Enhanced entity extraction accuracy, achieving an **F1 score of 91.88%** by fine-tuning the **DeBERTa model** on augmented cybersecurity datasets.
- Deployed the **NER model** as an **API endpoint**, enabling seamless integration into existing cybersecurity tools for real-time threat detection.
- Pre-processed and **managed large datasets**, using data from platforms like **OpenCTI**, increasing the model's ability to detect complex threat patterns.
- Conducted rigorous statistical evaluations, using **Wilcoxon Signed-Rank** and **Bootstrap Resampling tests**, to ensure model **robustness** and **reliability** for cybersecurity applications.

Sentiment Analysis of Etsy Reviews| Forsk Coding School

- Developed a machine learning-based sentiment analysis model to classify customer feedback from Etsy's jewelry section into positive and negative sentiments.
- Scraped and processed customer reviews using custom Python scripts, automating data collection for sentiment classification.
- Built and implemented data visualization tools, including pie charts and word clouds, to showcase sentiment distribution and frequent terms in reviews.
- Designed an interactive user interface for manual sentiment checking, allowing users to test model predictions on new data.
- Ensured model persistence by saving the trained model for future use, improving efficiency for re-running sentiment analysis.

EXTRACURRICULAR ACTIVITIES

- **Organizer –** Led **40+** technical workshops and webinars focused on innovation in computer science and technology
- **Competitions – Won 2 Ideation** events, highlighting creative problem-solving and collaboration
- **Treasurer** – Managed funds **(~£500)** for college events, ensuring budget compliance and event success
- **Centre For Creativity Award -** Awarded for outstanding contributions to innovative thinking under design thinking framework
- **Founder, REGEX-CSE Club:** Set up and grew a student network to foster collaboration and innovation in computer science

LANGUAGES

- English (Professional Working Proficiency)
- Tamil (Native Proficiency)

CERTIFICATES

- Python MOOC course (With honours)
- Mathematics for Machine Learning