Pranava Kailash Subramaniam Prema

Data Scientist

E: pranavakailashsp@gmail.com | L: linkedin.com/in/pranava-kailash | W: pranava-kailash.github.io |G: github.com/Pranava-Kailash

PERSONAL SUMMARY

Data Scientist with experience and transferable skills in data analysis, predictive modelling, NLP, Machine Learning, Python, SQL, and Cloud platforms like AWS. I leverage data to uncover actionable insights, with a history of automating processes and improving operational workflows. I hold a Graduate visa until April 2027.

SKILL

- Technical Skills: Python, SQL, Selenium, Beautiful Soup, TensorFlow, Hugging Face Transformers, Git, GitHub, GitHub Actions, • Matplotlib, Seaborn, Docker, Astro, Apache Airflow, MLflow, PowerBI
- Data Science & Development: Data analysis, Data modelling, RAG, LLMs, ETL, API Development, Machine learning, Clustering, ٠ Regression, Performance monitoring, Agile, Data governance, AWS

WORK EXPERIENCE

Data Researcher/Engineer | Fleet Street Research UK | Freelance

- Automated data extraction pipelines with Python-based scripting, increasing data acquisition efficiency by 50% ٠
- Smoothened the process of image importation into database, hence enhancing data quality
- Migrated data pipelines using SQL Alchemy between systems like MySQL and SQLite, ensuring zero downtime or data loss
- Implemented Git version control for team collaboration

Data Scientist | Yoshops.com - Internship | India

- Automated data collection via web scraping tools increased real-time pricing accuracy by 40% and enhanced product performance insights
- Built predictive models improving decision-making accuracy, showcasing expertise in data integration and visual • presentation
- Developed a 90% accurate Deep learning model for osteoarthritis detection using VGG16, optimizing feature extraction in medical applications

Data Scientist | Forsk Coding School - Internship | India

- Conducted in-depth Sentiment analysis on customer reviews using Natural Language Processing (NLP), finding key areas for • product improvement
- Used **Flask** for **back-end integration** and webpage display
- Developed and deployed scalable machine learning models on cloud platforms (AWS), ensuring efficient data processing pipelines

EDUCATION

M.Sc. in Data Science University of Surrey, Guildford, England	February 2024 - February 2025
B.E in Computer Science and Engineering SNS College of Technology, India	August 2019 - April 2023

CERTIFICATIONS

•	Docker Fundamental Professional Certification Docker. Inc	May 2025
•	Academy Accreditation - Databricks Fundamentals Databricks	April 2025
•	Academy Accreditation - AWS Databricks Platform Architect Databricks	April 2025
•	SQL Certificate Hacker Rank	March 2025
•	Power BI Beginner to Pro Workshop Pragmatic Works	March 2025
•	Mathematics for Machine Learning Coursera	June 2021
•	Applied Machine Learning in Python Coursera	June 2021
٠	Retrieving, Processing, and Visualizing Data with Python (with Honors) Coursera	June 2021

October 2022 - January 2024

February 2022 - April 2022

March 2021 - June 2021

PROJECTS

Weather ETL Pipeline | Personal Project

- Designed and implemented a robust ETL pipeline using Astro and Airflow to collect, process, and store real-time weather data from Open Weather public APIs, ensuring high data availability and accuracy
- Automated **data ingestion** using **Airflow scheduled jobs**, normalizing and transforming raw weather data (temperature, humidity, wind speed) for downstream analytics and reporting
- Developed **data quality validation checks**, handling anomalies and **missing values** to maintain the integrity and consistency of the data warehouse

AI-Powered Job Search Assistant | Personal Project

- Developed a job search assistant tool leveraging Lang Chain, OpenAI, and Stream lit to automate discovery and analysis of job listings from LinkedIn based on user-defined roles and locations
- Engineered a custom web scraping solution to fetch and parse job listings, integrating with the Proxy Curl API to retrieve and display comprehensive job profiles and insights
- Utilized large language models to analyse job descriptions and generate actionable insights, including recommended skills, CV keywords, and role expectations, enhancing the job search experience
- Implemented secure environment variable management using python-dotenv for sensitive API keys, and incorporated third-party APIs (OpenAI, Proxy Curl, SerpAPI) for robust data retrieval and processing
- Built an intuitive, interactive UI using **Streamlit**, enabling users to easily enter search criteria and instantly view detailed job analytics and recommendations

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
- Used MLflow to maintain artifacts and experiments

Sentiment Analysis of Etsy Reviews | Forsk Coding School

- Developed a machine learning-based **sentiment analysis** model to classify **customer feedback** from **Etsy's jewellery 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 (Flask)** 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