YASH PATEL

Canada | P: +1 6474466971 | yashsanjaykumarpatel@gmail.com | Github | Portfolio

WORK EXPERIENCE

Data Scientist Intern — Toronto, ON
Enkiube Technologies Inc. — May 2023 – Sept. 2023
- Optimized data workflows by managing 5000+ image annotations with Label Studio and **validating** datasets, boosting model training **quality** by 60% and **accuracy** by 40%.
- Enhanced **ML** systems by building models, integrating **LLMs**, and developing 20+ Python features—improving app efficiency by 50% and driving smarter business decisions.

Data Engineer Intern — Mississauga, ON
Lambton College — May 2023 – Aug 2023
- Built an end-to-end **AWS data pipeline** to **extract**, **transform**, and **load** data, ensuring data quality through advanced cleaning techniques and meeting stakeholder requirements.
- Designed **Power BI** dashboards to visualize key insights and presented findings to stakeholders, demonstrating the impact of data engineering on decision-making.

Machine Learning Engineer Intern — India
Amkay Software Solutions — Sept 2020 – May 2021
- **Led a 5-member team** in the end-to-end development of a custom **CRM** system, managing project timelines, task delegation, and backend architecture aligned with complex requirements.
- **Implemented a four-tier user access system** within the CRM, significantly enhancing task distribution, user management, and tracking efficiency by 30%.

SKILLS

Languages & Frameworks: Python, R, JavaScript, TypeScript, PHP, Django, Flask, DAX

Analytic Tools: Microsoft Excel, Power BI, Tableau, QuickSight, Power Query

Database Technologies: SQL, MongoDB, PostgreSQL, Microservices

Cloud & DevOps: Docker, Render, AWS, Azure, Google Cloud Platform, Git, GitHub, CI/CD

Big Data Technologies (Working knowledge)**:** Hadoop, Hive, Spark

Supporting Tools and Technologies: Jira, Confluence, ChatGPT, Gemini, DeepSeek, HTML, CSS, REST APIs, OAuth2

PROJECTS

Reddit Sentiment Tracker (Python, ETL, Power BI, AWS [Athena, S3, Lambda], NLTK [VADER, punkt])
- Orchestrated **real-time Reddit Sentiment Tracker** for Canadian Politics using continuous data flow, **AWS Lambda**, **ETL**, and **NLTK**.
- Leveraged advanced technologies such as **AWS S3**, **Athena**, and **Power BI** for efficient data storage, analysis, and visualization.

Montagne Parfums (PRAW, BeautifulSoup, FastAPI, Cron,JSON, TailwindCSS)
- Engineered a **real-time fragrance deal tracker** by integrating the Reddit API to scrape user-submitted posts from fragrance-related subreddits with over 7000 users, enabling continuous monitoring of community-seller deals.
- Accomplished data fetching using **ChatGPT API** and **Prompt Engineering** to filter and convert unstructured post content into a structured JSON format with 6 key parameters, improving the accuracy and reliability of data extraction.
- **Automated** price comparison logic by scraping official websites to fetch **live pricing and availability**, dynamically showcasing the top 3 best-value options as new posts are published.

Netflix Trend Dashboard (Kaggle, Tableau)
- Analyzed 6,000+ rows of data using **Tableau**, visualizing movie and TV show trends by country, release year, genre, and rating.
- Enabled tailored viewing experiences by allowing users to explore specific movies and show characteristics such as release date, duration, and cast.

EDUCATION

Certified in Google Data Analytics Professional — Toronto, ON
Google — Nov 2023 - May 2024
Post Graduation Diploma of Big Data Analytics — Mississauga, ON
Lambton College — Jan 2022 - Aug 2023
Bachelor of Engineering in Information Technology — Ahmedabad, India
Gujarat Technological University — June 2017 - July 2021