**ARJUN KUMAR**

+447442220504 | arjunkumarwk1998@gmail.com | Birmingham, UK | [linkedin.com/in/arjun-kumar-w-k](https://linkedin.com/in/arjun-kumar-w-k) | [github.com/arjun1998](https://github.com/arjun1998)

**PROFESSIONAL SUMMARY**

Graduated in MSc Data Science at the University of Birmingham, UK. Ex Data Engineer with 3 years experience at Infosys Ltd in the automation of Test Data Management along with proven technical, organizational and communication skills. Aiming to expand my current skillset, learn new experimental methods and make a significant contribution as a Data Scientist / Analyst

# EDUCATION

# Masters in Data Science with Distinction, University of Birmingham 2022-2023

# Dissertations project- Science map embedding by dimensionality reductions of science journals.

* Developed Natural language processing (Hugging face modules, Neural Networks, AI and Data modelling techniques) and other text classification techniques
* Implemented Large Language Models, Analysed unstructured data collected using Selenium automation from 160 research papers to create datasets utilising (Data Collection, Data Processing, Scikit-learn, TensorFlow, Seaborn)
* Built, Tested and Deployed machine learning models over Dataset for Data Analysis, Clustering techniques (K-means, elbow method) and Statistics for Reporting, Data Visualisation. Principal component analysis for dimensionality reduction.
* Deployed machine learning techniques and frameworks, such as TensorFlow and PyTorch, Python libraries like NumPy, pandas, scikit-learn for natural language processing (NLP)
* Data Science Algorithms (Zero shot classifications, text analysis, Science Map, Keras, Neural Networks used in the project
* Inventory Management by Online shopping intention prediction
  + Machine Learning Algorithms used in Exploratory Data Analysis, Market Basket Analysis, Customer segmentation, Demand Prediction, Product Compatibility
  + Apriori model, FP growth model, Clustering, Random Forest, Support vector Machine, TF-IDF
  + Python development professional experience of 3 years used to utilize libraries Seaborn, pandas, numpy, Matplotlib for Building Data Models
* Storing and Managing Data- Implementation and Verification of a Data Vault
* Sql, PostgreSQl, Graph (Neo4j) MySql Python scripts for Data reading and loading, Data transformation
* Unix, JSONS and NoSQL utilized with Data Management Modules.
* Datawarehouse techniques and SQL expertise from professional experience applied.

**Work Experience**

**Systems Analyst** 05/2019 to 04/2022

Infosys LtdChennai, India

**Responsibilities:**

* Led an 8-member Master Data Management team in an agile environment for a data warehousing and test automation project. Managed data migration from a legacy on-prem DB to AWS RDS, involving data cleanup, batch processing testing (JSONs) via Python and shell scripting. Conducted exploratory, regression, stress testing, and UAT using database management data, reducing testing time from 8 weeks to 3 weeks.
* Automated API testing using **Behat with Gherkin syntax in PHP**, enhancing test precision and efficiency by applying behavior-driven development practices to streamline the testing process.
* **Test Lead-** Specialized in automated system processing rules testing, **Test Script Development** and data relocation testing by executing **SQL** queries in **Oracle** database using **Python**. Managed the JIRA board , Leading the Indian systems testing team.
* Created **Mapping** Documents and Developed **Selenium automated** elements mapping pipeline from XML, JSON and legacy stores to AWS S3 standard structures as per client’s requirement.
* Collection of test data from the database server, filter & analyze data as per given client requirements by executing SQLqueries in Oracledatabase using Python.
* Proficiency in Data visualization platforms such as **Power BI** and **Tableau** which helps to deliver test data efficiently according to client requirements through design of **UI**, effective communication and excellent presentation**.**
* Performed **Test Data Management** (TDM) and **ETL Data Testing** ensuring data consistency and accuracy of the organization's shared data by maintaining **ETL CI/CD pipelines**.
* Directed ETL framework workflow using Informatica and VBA tools, streamlining processes for 180,000 data files and system processing rules for the bank. Reduced timeline from 4 weeks to 3 days, earning a role upgrade.
* Maintained Less than 2 hours of Bug fixing time on Ataccama **REST API** services with Root cause Analysis by developing streamlined testing pipeline made by automating test suite using **SOAPUI** and **POSTMAN**,
* **CI/CD Maintenance of API** of citizens Bank’s Website and Version control using Git for 14months , Additionally, conducted thorough stress, performance, functional, and scalability testing using SOAPUI, POSTMAN, and Python on various platforms and devices (mobiles, Web apps)

**Accomplishments:**

* Won a role upgrade and “Employee of the Month” award at Infosys MCity Chennai DC for developing VBA macros to collect and compare test data from legacy servers to AWS.
* Created a Python-based data comparison automation tool, reducing testing time by 30% and saving Rs. 5,00,000.
* Automated the Master Data Management process, enabling real-time data access for the Data Testing team and eliminating the need for Functional Testing team resources, resulting in cost savings of approximately Rs. 3,00,000.

**AI Engineer, Software Developer Intern 03/2024 to 06/2024**

**Radical AI Remote**

* **Project Quiziffy**: Developed AI project utilising **Google cloud,** Embedding through **VertexAIEmbeddings** and Langchain, File ingestions and UI through Streamlit, Data transformations using ChromaDB. Document processing, split text chunks using CharacterTextSplitter, and created a Chroma collection in memory.
* Created the quiz generator using **Gemini-Pro model** and integrated with vectorstore for context-driven question generation to generate a quiz by looping through questions, ensuring uniqueness with a retry limit, and validating against the question bank. Processed documents using ChromaDB, Split text chunks using CharacterTextSplitter,
* Project Dynamo: Built a robust **FastAPI** app for YouTube video analysis, Utilized pydantic to validate input fields, create an endpoint to analyze videos, and test it locally with uvicorn. Retrieve the full transcript of the video using langchain-community and return specific data from the endpoint response
* Utilized **CORSMiddleware** to allow cross-origin requests and installing Axios to handle video analysis requests. Additionally, created a route to send links to the backend endpoint and handling responses to display data in the frontend. Tested the integration by enabling both servers and sending requests from the frontend

**Skills and Core Qualifications**

|  |  |
| --- | --- |
| * **Programming Languages**: C , C++, JAVA, Python, JavaScript, JSON, HTML, Shell scripting, Linux/Unix * Test Data Management (TDM) and Master Data Management(MDM) using CI/CD * **Databases**: SQL, NoSQL, Microsoft Office(Excel), Oracle, PostgreSQL, UNIX, MongoDB, Neo4j, Redshift, DataLake, Talend * **Tools**: Selenium, Mainframe, ETL Informatica, Jupyter, Spyder, Power BI, Tableau, Apache HIVE, Hadoop, Microsoft Azure, DevOps, AWS Cloud Practitioner | * **Data Warehousing:** ETL Data Testing, AWS Redshift, Automation testing, Talend API * **Data Science and Data Analytics**: Data Cleansing and Data Visualization, Predictive Analysis, Text Analysis, Statistical Analysis * **Machine Learning**: Regression, Classification, Clustering, Neural Networks and Deep Learning, PCA, Feature Selection and Association mining. * **Lifecycle models** : Waterfall, Agile and scrum methodologies, JIRA, HP ALM |

**Courses and Certifications**

* **Amazon Web Services** AWS Cloud Practitioner Essentials, AWS Big Data Essentials
* **Google Kaggle** Intermediate Machine learning
* **Mahanagar Telephone Mauritius Limited:** Industrial training in GSM Technologies(2G/3G/4G)

**Additional Information**

**Languages**: English IELTS Score 8, BEC Cambridge English Business Certification | Tamil | Hindi | Telugu