NITHIN MATHEW JOSEPH

✆ +44 7880111721 | nmj210199723@gmail.com | LinkedIn Profile

PROFILE SUMMARY

Seasoned Data Engineer with over 4 years of extensive experience across diverse industries. Skilled in Python for versatile applications, specializing in data manipulation. Proficient in SQL for efficient data management and visualization with PowerBI. Experienced in handling Json, XML, and CSV formats for data interchange and manipulation, with a knack for building intuitive GUIs. Possessing a deep understanding of AI technologies and NLP techniques. Proficient in developing ETL pipelines and adept at collaborating with stakeholders to tackle business challenges effectively.

CORE AREAS OF EXCELLENCE

Project Management | Statistics & Machine Learning | Data Modelling | Data Analysis & Visualization | Real-Time Data Delivery| Data Collection & Integration| Business Intelligence | Natural Language Programming| Deep Learning| Diagnostics Fault Analysis| Yield Analysis| Automation| Client Relationship Management| People Management

PROFESSIONAL EXPERIENCE

Since Jan 2024 | INRIX | Data Editor

- Providing real-time traffic incident and congestion data to various INRIX customers and products.
- Collecting incident data from diverse sources, department of transport , police, and social media.
- Leveraging cutting-edge GPS technology & custom-built mapping software to identify various factors affecting the flow of traffic.
- Communicating updates to commercial radio and TV stations in North America.
- Collaborating with the Traffic Data Journalist team at the Data Centre in Quinton, Birmingham.

Jun 2018 – Jun 2021 | Qualcomm | Data Engineer

- Led the initiation and successful completion of diverse technical projects by implementing agile methodologies..
- Executed tasks related to data wrangling, visualization, and statistical modeling, which comprised a significant portion - approximately 80% - of my workload.
- Operating relational databases like HeidiSQL entailed a range of responsibilities, including inserting, deleting, constructing interlinked details, formatting, and ultimately extracting essential information using SQL queries.
- Developed inhouse automations using diverse programming languages such as Python, Perl, C++, tickle and bash
- Utilizing modules such as openpyxl, xlsxwriter, numpy, pandas, scikit, matplotlib, tkinter, json, and basic standard libraries, the majority of Python automation scripts were constructed.
- Led the effective management of extensive datasets, including Standard Test Data Format (STDF) Files.
- Implemented automation scripts that utilize parallel executions for parsing STDF files exceeding 1GB size that lead to substantial reductions in overall resource and team effort expenses.
- The parsing scripts achieved an accuracy rate exceeding 90%, attributed to thorough research and meticulous brainstorming conducted during their development.
- Refined editor macro functions across various platforms including Vim, gVim, Notepad++, and Excel, facilitating data processing and showcasing critical thinking skills crucial for accomplishing pivotal tasks necessitating unique builds.
- Enhanced the upkeep of automation tools, ensuring optimal functionality.
- Supervised adherence to open socket validation by verifying test programs for scan diagnostics.
- Enabled precise correlation data analysis across various tester platforms, each holding substantial single wafer data ranging from 1 to 100 TB, through tailored Perl automation tools, achieving a remarkable 98% precision in generated reports.
- Coordinated collaborative efforts with various teams within the company, leading multiple campaigns.
- Addressed operational glitches related to scan diagnostics and memory diagnostics reported through the Jira platform.

Aug 2017- Jun 2021 | Anora Semiconductor Labs Pvt. Ltd.| Product Development Engineer

- Gained familiarity with integrated development environments (IDEs) for PyCharm and Eclipse.
- Trained on Python, Linux and Git version control systems.
- Educated on structured documentation practices following script release via the Oxygen XML Editor.
- Self tutored on Perl automation and graphical user interface (GUI) frameworks with in-house resources.
- Obtained instruction on in-house scripts utilized by test engineers to streamline and minimize redundant tasks.
- Contributed to the development of 40% of the scripts intended for in-house use within the company.
- In collaboration with teammates, developed multiple executable utilities while conducting thorough research on Python modules.
- Received training in Smartest 7 basics and fundamentals of electronics for test engineering.
- Coached on theoretical methods and third-party tools for identifying scan and memory diagnostics faults.

.

KEY PROJECTS HANDLED

Title: Temperature Sensor Data Extraction from Binary Files (STDF).

Description: Implemented an automation script in Python to extract temperature sensor data, stored it in a database format. Utilized a Python module for database creation. Extracted required information from the database using the Mysql module. Sorted and presented data analytically in Excel using the Pandas framework.

Tile: Heartstroke Prediction Model

Description: Focused on constructing a predictive model to evaluate the likelihood of stroke in individuals, this initiative entailed an in-depth analysis of patient data, incorporating a diverse range of variables, spanning from lab test results to observed symptoms.

- **Data Pre-processing and Visualization:** Conducted comprehensive data preprocessing on patient data, employing techniques to enhance the quality and relevance of information. Utilized advanced data visualization methods to discern patterns and relationships within diverse factors, providing valuable insights crucial for informing the predictive model.
- **Machine Learning Model Implementation:** Applied state-of-the-art, open-source machine learning models to address the central challenge of stroke risk prediction. Leveraged these models to perform accurate and reliable risk assessments based on individual patient information, showcasing proficiency in data analysis, machine learning, and data visualization. Advanced the field of stroke risk assessment and prevention, contributing to a more informed and effective approach.

Title: Accident Severity Model

Description: Created a predictor for accident severity utilizing data files from the UK Department of Transport on road safety incidents in the UK. The development process involved an initial exploratory data analysis and subsequent preprocessing steps, including data cleaning, normalization, and scaling. The resulting machine learning model demonstrates proficiency in predicting accident severity across different regions in the UK. It achieves this by utilizing accident-related information such as weather conditions, passenger details, and types of vehicles involved.

Title: Legal Query Engine

Description: The project focused on improving question-and-answer capabilities for interpreting legal text. The objective was to enable users to pose legal-related queries and obtain context-aware answers. The implementation of two crucial models was integral to this project.

- **Named Entity Recognizer (NER):** Generated a NER model using the open-source module 'spacy' to identify and classify entities within legal texts. This played a crucial role in understanding the context of user queries.
- **Question and Answer (Q&A) Transformer Model:** Developed a Q&A model utilizing data from the Hugging Face platform. This model empowered users to inquire about legal matters, generating comprehensive answers by contextual analysis. The successful execution of this project underscored expertise in natural language processing and seamless model integration.

TECHNICAL PURVIEW

- Programming Languages: Python, Perl, R, SQL, Shell scripting and Linux Syntax
- Version Control Systems: Git, SVN and RCS
- Software: Microsoft Office and Power BI
- Service: AWS
- Project Management Tool: Jira, Scrum

CERTIFICATIONS

- AWS Cloud Practitioner Essentials |Sep 2023
- Power BI Virtual Case Experience | Jul 2023
- Embedded Systems and Hardware Design | Aug 2016

EDUCATION

- **M.Sc. in Artificial Intelligence - Business Strategy |** Aston University, UK | 2022

Modules: Mathematics, Machine Learning, Deep Learning, Big Data for decision making (Data Mining), International Entrepreneurship and Business Strategy.

- **B.Tech. in Electronics & Communication |** Cochin University of Science & Technology, India | 2016

Modules: Computer Networking, Digital System Design, Embedded Systems & Wireless Communication