Furkhan Suhail

Data Scientist / Data Engineering
Chandler, Arizona, United States

Email: furkhansuhail@outlook.com
Phone: 334-497-2183
github.com/furkhansuhail
linkedin.com/in/furkhan-suhail/

PROFESSIONAL SUMMARY

Data Science Engineer with over 5 years of experience in utilizing AI/ML, model development, and analytics to drive data-driven decision-making through innovative solutions.

Proficient in Python, SQL, and ETL (Extract, Transform, Load) processes with a strong background in NLP, regression, predictive analytics and scalable AI solutions.

SKILLS AND COMPETENCIES

• *Programming Languages and Database Management: **Python, Embedded C, C++***
• *Machine Learning/Deep Learning: **scikit-learn, TensorFlow, XGBoost, Random Forest, Gensim, NLP, NLTK, Sentiment Analysis, Keras***
• *Transformer Models: **BERT, LSTM, Large Language Models (LLMs), OpenAI, LLaMA, Hugging Face***
• *Data Engineering & Analytics: **ETL, Pandas, SQL (PostgreSQL, MySQL, Snowflake), Data Mining***
• *Data Visualization and GUI: **Tableau, Power BI, Tkinter, Seaborn***
• *Software Development & Version Control: **Agile Methodologies, Scrum, DevOps, Object-Oriented Programming, Confluence, Git***
• *Data Science and AI Applications: **Artificial Intelligence, Deep Learning, Generative AI, Machine Learning Algorithms, Neural Networks, PySpark, PyTorch***
• *Cloud Deployment and Scheduling: **AWS (EC2, Lambda, SageMaker, Bedrock), PySpark, Kubernetes, CI/CD, Apache Kafka***

WORK EXPERIENCE

Microchip Technology | Data Science Engineer| Chandler, Arizona, United States

November 2022 – March 2025

- Ideated and developed a Bidirectional Encoder Representations from Transformers (**BERT**) machine learning model to classify customer queries into predefined categories, enhancing customer support efficiency. Built ETL pipelines for data preprocessing using Pandas and NumPy and applied **NLP** techniques with **NLTK**. Achieved a 22% reduction in case transfers among support engineers within the first six months, significantly improving case resolution efficiency.

- Developed XGBoost predictive models for Microchip device families to resolve operation challenges to enhance supply chain efficiency and production throughput. Integrated Kafka for real-time data streaming to improve data processing and decision-making speed. Leveraged advanced analytics to forecast demand, streamline operations, and drive data-driven decision-making. Improved supply chain reliability and reduced costs by delivering actionable business intelligence.

- Designed and deployed a [Retrieval-Augmented Generation (RAG)](#) system using **LangChain** and [Claude 3.5 Haiku LLM via Amazon Bedrock](#) to process and query historical **Microchip customer support data** (FAQs, tickets, resolutions). Implemented **semantic search** using **Amazon Titan Embeddings** and **vectorized document indexing**, enabling support agents to retrieve **context-aware** and **LLM-generated** responses in real time. The solution leveraged **modular prompt engineering**, **streamlined data ingestion pipelines**, and **chunking strategies** for document splitting. Achieved ~20% reduction in **mean time to resolution (MTTR)**, enhancing **case deflection**, **first-response accuracy**, and overall **customer satisfaction**.

- Owned product vision to address **latency issue** for focus accounts by developing a customer query ranking system to assess technical support performance using **statistical analysis** and data science techniques like **text analysis** (**NLTK, TextBlob, scikit-learn**), and **data mining** for insights. Designed and integrated a user interface using **Tkinter** to enable seamless interaction with the system. Managed background storage using **MySQL** and **PostgreSQL** databases. This initiative improved the case closure rate for focus customers by 18%, enhancing overall support responsiveness and efficiency.

- Developed and optimized a TensorFlow-based **multi-class classification model** to categorize microcontroller devices by name and family, mapping them to respective Business Units (BUs). Achieved **80% classification accuracy**, reducing manual effort, enabling engineers experienced with these technologies to easily take ownership of customer query cases, reducing customer query response times by an average of 6 hours.

- Spearheaded the product roadmap for an end-to-end **clustering pipeline (K-Means)** for account segmentation to classify the importance of accounts for Microchip Technology, including **data collection**, **preprocessing**, and **clustering model**

optimization. Automated dynamic cluster determination and tuned parameters for improved accuracy which was adopted by cross-functional teams globally for improved capacity allocation.

Microchip Technology | Support Engagement Engineer | Chandler, Arizona, United States

August 2019 – November 2022

- Implemented a [regression](#) model to forecast the inflow of technical queries, enabling proactive resource allocation and optimizing team workload distribution. Enhanced query handling efficiency by anticipating query volumes, reducing response times, and improving customer satisfaction. Utilized **Python**, **pandas**, and **scikit-learn** for **data analysis** and **model development** streamlining monitoring processes and supporting operational planning. The survey rating of the team improved by **8%**
- Architected and optimized a name-matching algorithm to standardize client accounts and eliminate alias discrepancies using string matching and **Naive Bayes** for classification. Integrated financial data using **yfinance API**, **Nasdaq Data Link API**, ChatGPT **API**, and internal chatbots for **data authentication**, improving **data integrity** and **validation** efficiency. This solution was adopted for **data analytics** and record cleanup, addressed prioritization and decision-making in pain point assessment, significantly improving data consistency across business units.
- Developed and deployed a **sentiment analysis** model to evaluate customer query responses. Applied **NLP** techniques, including **tokenization**, **stop-word removal**, **TF-IDF**, and **word embeddings**, to extract meaningful insights. Conducted **statistical analysis** on customer queries and responses, enabling data-driven resource allocation for handling irate customers, ultimately enhancing support team efficiency. Customer satisfaction score increased by an average of 16%
- Developed **CI/CD**-driven high-volume data pipelines for product analytics, automating statistical, diagnostic, and prescriptive analyses across multiple business units. Conducted interdependency and performance analyses to identify cross-functional impacts, driving improvements in operational efficiency and strategic decision-making. Defined key KPIs and presented insights to stakeholders, earning positive feedback from senior management.
- Conducted statistical analysis on customer and technical engineer response times to identify **behavioral patterns** and **reaction trends**. Leveraged **Python**, **pandas**, and **scipy** for **data analysis**, **hypothesis testing**, and **pattern detection**. Provided insights to optimize response strategies and improve communication efficiency.
- Experienced in resolving on call technical issues related to **I2C, SPI, EUSART** communication protocols, **Pulse Width Modulation** (PWM), **system clocks, and GPIO modules**. Performed bottleneck analysis to identify latency issues and optimized data throughput.
- Conceptualized, architected, and implemented a gesture-controlled bot using the TDK InvenSense MPU6050 (accelerometer & gyroscope) with an 8-bit PIC18F47Q10 microcontroller. Utilized I2C for sensor interfacing and EUSART for testing. Developed real-time gesture recognition firmware in Embedded C, optimizing performance with ADC, timers, and interruptions.
- Designed and optimized an embedded application for movement detection and distance measurement using IR sensors and Atmel AVR microcontrollers. Implemented real-time object tracking and optimized sensor data processing in Embedded C for accuracy and responsiveness in dynamic environments.

University of Texas at San Antonio | Research Assistant |San Antonio, Texas

January 2018 – April 2019

- Performance characterization of the SPEC CPU2006 benchmark suite on Intel i5 and i7 (5th generation Broadwell architecture).
- Determined the efficiency of the processors using **CINT2006** and **CFP2006** benchmark based on the outcome of hardware event sampling and memory access analysis using Intel VTune Profiler (formerly VTune Amplifier).

Education	
University of Texas at San Antonio| Master of Science in Electrical and Computer Engineering	*2016 – 2018*
Osmania University| Bachelor of Engineering in Information Technology	*2009 – 2013*

Honors and Awards

- **Recognized Lab Approver, MASTERS Conference – Microchip Technology** Served as a Lab Approver, providing technical assistance to attendees and ensuring smooth lab operations. Recognized by Microchip for outstanding support and technical expertise, including special recognition for developing an innovative crowd detection system.