

## Kin Wong

832-436-8892 | kinwong.ds@gmail.com | github.com/kinwong-ds

---

### PROFESSIONAL EXPERIENCE:

#### Optum:

##### Data Scientist II

Apr 2022 – Present

- Currently implementing a Large Language Model (LLM) for medical contract rate extraction to prevent rate mismatches.
- Created an AI/ML pipeline utilizing healthcare data to reduce 30% of claims in post-pay claim inventory with an increase of 20% true positive rate, reducing auditor needs by 2 across 3 platforms.
- Optimized claim audit prioritization with healthcare data and saved over \$144K annually by using PySpark for data cleaning, feature selection, and feature engineering in a in an AI/ML model pipeline.
- Supervised a production claim tiering pipeline in Apache Airflow, account for \$13.8M in annual savings by ensuring system performance though resolving Kubernetes errors, managing Docker image builds, and performing Spark version upgrades.
- Collaborated with engineering and operations teams to address system performance issues and engaged with business stakeholders to provide insights and answers.
- Developed a Streamlit dashboard for a claim tiering pipeline, leveraged by leadership to track KPIs and key metrics like feature drift and recovery amounts, enabling actionable insights, business decisions, and early issue detection.

#### HP:

##### NLP Data Scientist

Jul 2021 – Apr 2022

- Developed a deep learning MLOps pipeline to process over 2 million call logs weekly, providing insights into product issues and enabling targeted actions by the operational team.
- Fine-tuned BERT models to classify unstructured text data from helpdesk call logs, improving classification accuracy by 15% as part of a MLOps pipeline.
- Implemented FastText and mBART models on premise to identify languages of conversational text and translate non-English text, enabling analysis of product issues from overseas customers.
- Apply mBART model to translate non-English text on premise to address customer issues in multiple languages.
- Increased classification model accuracy by 20% by summarizing call logs using a GPT model.
- Created an OCR image to text pipeline using easyocr and cv2, saving 50 hours of manual data entry per month.

#### Calpine Corporation:

##### Data Scientist

Aug 2019 – Jul 2021

- Engineered features and developed time series models for machine learning projects, including anomaly detection to correct faulty turbine sensor outputs.
- Developed data pipelines and a classification system using ensemble methods to transform power plant data and accelerate turbine maintenance processes.
- Created an NLP model to categorize business reports, enhancing report classification efficiency.

##### Data Analyst

Jul 2017 – Dec 2018

- Streamlined system report for power plant users by developing a PowerApps application integrated with SharePoint.
- Designed, enhance, and test BIRT reports from EAM software Maximo utilizing SQL and JavaScript, delivering tailored solutions to meet business needs.
- Created on-demand SSRS reports using real-time data to improve power plant operational insights.

---

### CERTIFICATIONS & SKILLS:

- **Azure certifications:** DP-100, AZ-900, DP-900, AI-900
- **Programming Languages:** Python, R, JavaScript, SQL, Spark
- **Data Analytics/Data Science:** LLM, NLP, Transformer Model, Data Wrangling, Dashboard

---

### EDUCATION:

#### University of California Riverside

- MS in Engineering and Data Science
- Graduation date: June 2021

#### University of Houston

- BS in Computer Information Systems
- Graduation date: May 2018