Kaiwei Cao

Kaca4509@colorado.edu | 720-526-8671

Summary

Data Analyst / Business Intelligence Analyst with **2 years of experience** in **Data Science/Analysis, Machine Learning, and Power BI development.** Proficient in **Python** libraries and frameworks such as (**NumPy, Pandas, Scikit-learn, TensorFlow, Keras, Matplotlib, Seaborn**), **Power BI, Tableau, Looker, and Excel** for dashboard visualization**.** Experience working with **SQL** and **Oracle 12C** databases. Proficient in writing **SQL queries, sub-queries, and joins.** Solid background in **Machine Learning** with a broad understanding of **supervised and unsupervised learning** techniques and algorithms (e.g.: **Linear and Logistic Regression, Naive-Bayes, K-means, Decision Trees, Random Forest, XGBoost,** etc.). Experienced in deep learning technologies and frameworks like **Hugging Face, LLM, and GenAI** with **Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN),** and **Recurrent Neural Networks (RNN).**

Certifications

- Microsoft Certified Power BI Data Analyst Associate
- Oracle Database SQL Certified Associate
- PCAP-31-03 Certified Associate Python Programmer
- Microsoft Office Specialist: Excel Associate (Microsoft 365 Apps)
- IBM Data Science – Machine Learning with Python
- IBM Data Science – Deep Learning Fundamentals
- IBM Data Science – Deep Learning using TensorFlow
- IBM Data Science – Python for Data Science

Skills

Programming Languages:	Python, SQL, DAX, HTML
Libraries and Frameworks:	NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn (Sklearn), XGBoost, TensorFlow, Keras, PyTorch, ResNet, Hugging Face, GenAI, LLM
Database Management:	Oracle, MySQL, PLSQL, Microsoft SQL server, MongoDB
IDEs/Development Tools:	Jupyter Notebooks, PyCharm
Data Visualization Tools:	Power BI, Tableau, Looker Studio, Excel
Cloud Services:	Azure, Google Cloud Platform
Machine Learning Algorithms:	Linear and Logistic Regression, Decision Tree, Random Forest, SVM, Naive-Bayes, KNN, RNN, K-Means, Voting Classifier, AdaBoost, XGBoost

Employment History

SynergisticIT, Fremont - CA

Data Analyst / Business Intelligence Analyst

Aug 2022 – Present

Project – Rideshare Data Insights and Visualization using Looker Studio **Apr 2024 – Current**

The primary objective of this project is to perform comprehensive data analytics on Uber data to uncover valuable insights and trends. This will be achieved by leveraging **GCP Storage** for data storage, **Python** and **Jupyter Notebook** for data processing, Compute Instance for computational workloads, Mage **Data Pipeline** Tool for workflow automation, **BigQuery** for data analysis, and **Looker Studio** for data visualization. The project encompasses data ingestion, storage, processing, and visualization to extract meaningful insights and trends from data.

Roles and Responsibilities:

- Ingested Raw data into **GCP Storage** for secure and scalable storage.
- Employed **Python** scripts within **Jupyter Notebooks** to clean and transform the raw data.
- Split the cleaned data into fact and dimension tables for structured analysis.
- Utilized **Compute Instance** to efficiently run data processing tasks and handle computational workloads.
- Configured **Mage Data Pipeline** Tool to automate data processing workflows, ensuring seamless data flow and Integration.
- Executed advanced **SQL queries** in **BigQuery** to analyze the processed data and extract actionable insights.
- BigQuery is utilized to run **SQL queries** on the processed data to uncover insights.
- Developed interactive dashboards in **Looker Studio** to visualize key metrics, trends, and insights derived from the data.
- Used **Looker Studio** to create dynamic dashboards to visualize the analytical results, making the data insights actionable and easy to interpret.
- Regularly updated and maintained the dashboards to reflect the latest data and ensure ongoing relevance and accuracy of the visualizations.

Tools and Technologies used: Python, SQL, GCP Storage, BigQuery, Looker studio, Compute Instance, Jupyter Notebook, SQL

SynergisticIT, Fremont, CA
Data Analyst / BI Analyst **Dec 2023 – Mar 2024**

Project – Streaming Olympic data Insights with Azure and Power BI

This project focuses on analyzing Olympic data to derive meaningful insights and trends. This will be achieved by leveraging **Azure Data Factory** for data ingestion, **Azure Data Lake Gen 2** for data storage, **Azure Databricks** for data processing, Synapse Analytics for data analysis, and **Power BI** for data visualization. This project involves the processes of data ingestion, storage, processing, and visualization to provide a comprehensive understanding of the data.

Roles and Responsibilities:

- Ingested data into **Azure Data Lake Gen 2** using **Azure Data Factory** for scalable and secure data storage.
- Processed and transformed data with **Azure Databricks,** preparing it for analysis through collaborative data processing and advanced analytics.
- Executed **SQL queries** and performed in-depth analysis on processed data using **Synapse Analytics**.
- Developed dynamic and interactive dashboards using **Power BI** to present analytical results clearly and effectively.
- Implemented **DAX queries** in Power BI to perform complex data calculations and aggregations.
- Established automated data workflows to update and refresh dashboards regularly.
- Applied machine learning techniques in **Azure Databricks** to derive predictive insights and trends from the data.
- Utilized **Azure Data Lake Gen 2** as the scalable and secure storage solution for storing both raw and processed Olympic data.

- Synapse Analytics provides robust analytics and advanced querying capabilities on datasets.

- Coordinated with stakeholders to understand their analytical requirements and tailor the dashboards to meet their needs.

Tools and Technologies used: Azure, Power BI, DAX queries, Excel

---

SynergisticIT, Fremont, CA
Data Analyst / Power BI Developer **Aug 2023 – Nov 2023**

Project – Restaurant Data Insights: Leveraging SQL Server and Power BI
The objective of this project is to analyze data from Restaurants to uncover valuable insights and trends. Utilized **Microsoft SQL Server** for data storage and management, and **Power BI** for data visualization, incorporating **DAX expressions** for advanced analysis. Integrated data from diverse sources via **MS Excel**, enhancing the robustness of insights derived from the comprehensive analysis.

Roles and Responsibilities:

- Stored and Managed data stored in a **Microsoft SQL server** to ensure secure and efficient data handling.
- Utilized **SQL server** to perform data cleaning, ensuring the removal of inconsistencies and preparation of data for analysis.
- Conducted data transformation and normalization within the **SQL** server to streamline data structures for better analysis.
- Performed complex **SQL queries** to derive valuable insights and uncover trends in sales, customer behavior, and operational efficiency.
- Developed a dynamic and interactive dashboard in **Power BI** to visualize key performance indicators (**KPIs**) and analytical results.
- Employed **DAX expressions** in **Power BI** to create calculated columns and measures for enhanced data analysis.
- Integrated data from various sources using **MS Excel** for preliminary data exploration and verification before **SQL** server ingestion.
- Collaborated with the management team to understand their data need and tailor the **Power BI** dashboards to meet those requirements.
- Regularly updated and maintained the **Power BI** dashboards to reflect the latest data and ensure ongoing relevance and accuracy of visualization.
- Created dynamic and interactive dashboards using **Power BI** to present the analytical results in an accessible and comprehensible manner.

Tools and Technologies used: Microsoft SQL Server, Power BI, DAX expressions, MS Excel

---

SynergisticIT, Fremont, CA
Data Analyst / BI Developer **Apr 2023 – July 2023**
Project – Hospitality Data Analytics with Tableau Dashboard
The goal of this project is to develop a dashboard using **Tableau** for a call center that handles restaurant bookings across the UK. The collected raw data is transferred to **SQL** to organize, clean, and prepare for analysis, used **tableau** to create **dashboard visualizations** to turn the call center data into insightful visualizations that enhance operational efficiency and decision-making.

Roles and Responsibilities:

- Collected and aggregated raw data from the call center, focusing on restaurant bookings across the UK.
- Utilized **Microsoft SQL** to organize, clean, and prepare data for analysis.
- Employed **Power Query** in **Excel** for data transformation and integration tasks.
- Designed and developed interactive **dashboards** in **Tableau** to present key performance indicators and insights.
- Created **dashboard visualizations** to identify trends, analyze performance metrics, and highlight areas needing improvement.
- Conducted regular updates and refinements of the **Tableau dashboard** to ensure accuracy and relevance of the displayed information.
- Collaborated with call center management to interpret dashboard insights and recommend actionable strategies.
- Documented the data processing and visualization procedures to facilitate future updates and maintenance.

Tools and Technologies used: Microsoft SQL, Tableau, Excel, Power Query

---

SynergisticIT, Fremont, CA
Data Analyst / Data Scientist **Dec 2022 – Mar 2023**
Project – Stock Price Prediction using LSTM and RNN

This project focuses on utilizing **Recurrent Neural Network (RNN)** and **Long Short-Term Memory (LSTM)** models to predict stock prices. By applying deep learning concepts, specifically RNN and LSTM architectures, the project aims to provide accurate forecasts based on real-world financial data.

Roles and Responsibilities:

- Utilized **Python** for data preprocessing, cleaning, and manipulation to ensure historical stock price data is in a suitable format for training the models.
- Applied **Python-based** data analysis tools like **Pandas and NumPy** to handle large datasets, perform feature engineering, and prepare the data for model training.
- Real-world historical stock price data from Yahoo Finance is used for training and testing the predictive models.
- Implement **RNN** and **LSTM** architectures using **TensorFlow** deep learning frameworks.
- Leveraged **Hugging Face** libraries and **large language models (LLMs)** to enhance data preprocessing and feature extraction capabilities.
- Leveraged **GenAI** techniques with the **Hugging Face** ecosystem to augment data preprocessing and streamline model training, facilitating more robust predictions.
- **Recurrent Neural Networks** and **Long Short-Term Memory** networks are specialized deep learning architectures suitable for sequential data, making them ideal for time-series forecasting tasks.
- Train the models on the data to learn patterns and trends.
- Evaluate the trained models using relevant metrics such as **Mean Absolute Error** (MAE) and **Root Mean Squared** Error (**RMSE**).
- Understanding and applying concepts such as sequential data processing, feature engineering, model training, and evaluation in the context of stock price prediction.
- Analyze the models' performance on both training and validation datasets.
- Utilize the trained models to make predictions on unseen data for future stock prices.

Technologies used: Python, Pandas, NumPy, Deep Learning, TensorFlow, RNN, LSTM, Hugging Face, LLM, GenAI

SynergisticIT, Fremont, CA
Data Analyst / Data Scientist **Aug 2022 – Nov 2022**

Project – Image Classification Model using PyTorch and ResNet

The objective of this project was to develop an image classification model utilizing transfer learning with the **ResNet** pre-trained model in **PyTorch**. The model aimed to accurately categorize images into different classes, including social security cards, driving licenses, and others. The project involved setting up the **Google Colab** environment for efficient data processing and model training.

Roles and Responsibilities:

- Set up the **Google Colab** environment for seamless data processing and model training.
- Import the image dataset to **Google Colab** for preprocessing and modeling to convert the class labels to numerical values the model can understand.
- Resize and scale the images to a consistent size to ensure uniformity.
- Construct the **ResNet** architecture using **PyTorch'**s deep learning framework.
- Utilize a pre-trained **ResNet** model to suit the specific requirements of the image classifier task.
- Trained the **ResNet** model on the dataset to learn patterns and accurately classify images into different categories, i.e. social security cards, driving licenses, and others.
- Implemented data augmentation techniques to enhance the robustness and generalizability of the model.
- Utilized the pre-trained **ResNet** model as a starting point for the new image classification task.
- Applied transfer learning techniques to adapt the **ResNet** model for specific dataset requirements.
- Evaluate the model's performance on the training dataset using appropriate metrics such as accuracy, precision, recall, and **F1-score**.
- Validated the accuracy of the model's classifications by comparing predicted labels against the actual categories of the images.

Technologies used: Google Colab, PyTorch, ResNet, Transfer Learning, torchvision

Education

University of Colorado Boulder Boulder, CO

Master of Science, Applied Mathematics

Kaiwei Cao

For me, data science and data analysis have been my major passion and occupation for the last two years. I started my master's degree of applied science in University of Colorado Boulder, studying statistics and data science. And during the same time, I had also learned data analysis in a company where I utilized SQL, Excel, Power BI and some other BI visualization tools. After graduated from University of Colorado Boulder, I continued to working in the company where I wander between data analysis and data science, using all the methods and skills I have learned and learned more about data management and data structure.

And during this time, I have not only learned lots of skills about data science and data visualization but developed a passion for building visualization and machine learning models from various kinds of datasets. Like a treasure hunter, a data scientist or a data analyst is mainly responsible for discovering and gaining insights from an undiscovered territory. But for a data scientist, after discovering the first treasure, he or she can predict where the next treasure hunt would be.

So now, instead of wandering between data analysis and data science, I want to mainly focus on data science, which is the reason I want to join this amazing program where I can dive deeper in data science. Even though data analysis is interesting, a data science project feels like a more complete project for me where all kinds of skills are utilized, including data analysis and data visualization. And by participating in this program, I hope to become an expert in data science and help boost my career as a data scientist.