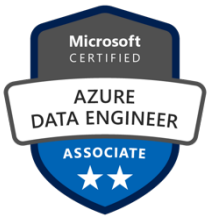

CHENFENG LI

Chicago, IL 60615 | (872)-215-0270 | cfli@chenfengli.com | https://ChenfengLi.com

Summary

Data Scientist / ML Engineer with 1-2 years of experience in Data Science, Machine Learning, and Power BI development. **MS in Statistics at the University of Chicago.** Proficient in **Python, R, SQL, Power BI, and Excel**. **Certified AWS Cloud Practitioner, Azure Data Engineer Associate, and Microsoft Certified Power BI Data Analyst Associate**. Demonstrated expertise in **Machine Learning and Deep Learning** including Facial Recognition, Neural Networks, Language Processing. Experienced in **Statistical Analysis** such as Generalized Linear Model, Bayesian Learning, Time Series. Strong foundation in **Programming Algorithms** including Data structures, Dynamic Programming, and Flow Algorithm. Proficient in **cloud computing with AWS and Azure** with setting up hybrid cloud architectures, managing databases, and conducting data processing and analysis using Azure Data Factory, Synapse Analytics, Databricks. Expert in creating **interactive dashboards and reports using Power BI**.

Skills and Certifications

- **Programming Languages:** Python, R, SQL, SAS, DAX, C, C++, HTML, CSS, JS, PHP
- **Libraries:** NumPy, SciPy, Pandas, Matplotlib, Seaborn, PyTorch, TensorFlow, Scikit-Learn, Hugging Face, LLM, GenAI
- **IDEs/Development Tools:** Jupyter Notebooks, RStudio, Visual Studio Code
- **Data Visualization Tools:** Power BI, Excel, Power Query, Python and R
- **Cloud Services:** AWS (Cloud watch, Cloud Formation, IAM, S3, SNS), Azure (Databricks, Data Factory, Synapse), GCP
- **ML Algorithms:** Generalized Linear Regression, Decision Tree, SVM, KNN, K-Means, Random Forest, Adaboost, XGBoost, Deep Neural Networks, CNN, NLP, LLM, etc.
- **Statistical Analysis:** Bayesian Inference, Markov Chain, Time Series, Non-parametric statistics

- **Power BI Data Analyst Associate** (Microsoft) – Data modeling, visualization, and interactive dashboard design
- **AWS Certified Cloud Practitioner** (AWS) - Cloud computing and infrastructure with AWS services
- **Azure Data Engineer Associate** (Microsoft) – Data integration, transformation, analysis with Azure services
- **Microsoft Office Specialist: Excel 2019 Associate** (Microsoft) – Data management and spreadsheet design
- **Google Advanced Data Analytics** (Google, Coursera) – Large datasets, data analytics, machine learning
- **Deep Learning Specialization** (DeepLearning.AI, Coursera) – Neural networks, Transformers and application in industry

Additional Strengths

- Strong communication, verbal and written with audiences or team members at all levels of technical expertise
- Leadership – coordinating tasks and communication within research teams
- Independent Research – Conducted in-depth 9-month master's thesis on politicians' facial expressions and media bias
- Website design and development
- Fluent in English, Mandarin Chinese and Cantonese

Employment History

SynergisticIT, Fremont, CA
Data Scientist
Project: Predictive Sales Analytics Platform

June 2024 – Present

Developed a machine learning model to predict total sales for each product and store for the upcoming month, using daily historical sales data. This involved data preprocessing, feature engineering, and applying models including **Ridge, XGBoost**, and **LightGBM**. The project optimized model accuracy and provided valuable forecasts for inventory management and strategic planning.

Roles and Responsibilities:

- Imported and merged table into **pandas DataFrames**, removed duplicates and imputed missing values.
- Implemented **TF-IDF** vectorizer to extract features from item and shop names, creating **text-based features**.
- Engineered lagged features and trend-based features for **time series analysis**.
- Conducted **Exploratory Data Analysis (EDA)**, including visualization of target distribution and time trends. Used multivariate heatmaps to analyze numerical and categorical pairings.
- Applied mean encoding on categorical features and matrix factorization for text features.
- Constructed and trained pipelines with the above transformations and **Ridge**, **XGBoost**, and **LightGBM** regressors.
- Performed feature selection using Recursive Feature Elimination with Cross-Validation (**RFECV**) and optimized hyperparameters using Bayesian optimization.
- Evaluated models through cross-validation, analyzing **Root Mean Square Error (RMSE).**
- Predicted future outcomes and compiled comprehensive reports.

Technologies Used: Python, Scikit Learn, Machine Learning Pipeline, NLP, TF-IDF, mean encoding, matrix factorization, Ridge Regressor, LightGBM Regressor, XGBoost Regressor, feature selection, hyperparameter optimization.

SynergisticIT, Fremont, CA
Data Scientist
Project: Social Media Sentiment Analysis and Reporting

March – May 2024

Partnered with a data scientist to develop an **Azure-based** system for sentiment analysis on 16,000 post-sale reviews for an online clothing store. Configured Azure Data Factory for data ingestion and used **SQL** in Synapse Analytics for cleaning and filtering. Implemented **NLP** techniques and a BERT-based model in Azure Databricks for sentiment analysis. Visualized results and generated comprehensive reports, with weekly data updates via scheduled triggers.

Roles and Responsibilities:

- Configured an **Azure Data Factory** pipeline to ingest data from on-premise database to **Azure Data Lake** Storage.
- Cleaned and filtered data using **SQL** scripts in **Azure Synapse Analytics**.
- Designed a notebook in **Azure Databricks** for data processing, analysis, and visualization.
- Implemented **NLP** models for text cleaning, lemmatization, stop-word removal, and perform sentiment intensity analysis with **BERT** based model from **Hugging Face** Transformers.
- Trained and evaluated a random forest classifier model to predict the sentiment of the apparel reviews.
- Visualized results by time and item and created comprehensive reports.
- Updated the dataset with a scheduled trigger in **Azure Data Factory** on a weekly basis.

Technologies Used: Python, Database management, Azure Data Factory, Synapse Analytics, Databricks, NLP, Hugging Face, BERT

SynergisticIT, Fremont, CA
Data Engineer / Data Scientist
Project: Hybrid Cloud Site-to-Site VPN Deployment for Secure Connectivity

December 2023 – February 2024

Designed and implemented a secure and scalable hybrid cloud architecture that connects an on-premises data center to an **AWS VPC** using a Site-to-Site VPN. The project involved setting up and configuring multiple **AWS services** to ensure secure, reliable, and monitored connectivity between the two environments.

Roles and Responsibilities:

- Set up **VPCs** in two regions with appropriate subnets, route tables, and internet gateways.
- Launched **Amazon Linux 2** instances in both regions with secure groups allowing SSH and ICMP traffic.
- Established Site-to-Site VPN connection using virtual private gateway with Openswan configuration.
- Utilized **Amazon S3** for critical data backup, enabling versioning and server-side encryption.
- Implemented **IAM** roles and policies for secure access management, ensuring least privilege principles.
- Monitored instance performance, network traffic and VPN connection using **AWS CloudWatch**.
- Configured **Amazon SNS** to send alerts for critical events.
- Automated infrastructure deployment with **AWS CloudFormation** for consistent, repeatable setups.

Technologies Used: AWS Instances, security groups, VPC, Site-to-Site VPN, IAM, CloudWatch, SNS, CloudFormation

SynergisticIT, Fremont, CA
Data Analyst / Business Intelligence Analyst

Project: Sport Corporation Sales Analysis

September – November 2023

Developed an advanced **Power BI** dashboard for an international sports corporation to analyze sales data, providing real-time insights into sales performance, discount analysis, and regional success. The dashboard was designed to facilitate data-driven decision-making by presenting key metrics in an interactive and user-friendly format.

Roles and Responsibilities:

- Created and filtered data using **SQL** on the on-premise dataset, leveraging a star schema data model with the Sales table at the center for optimized performance.
- Cleaned and transformed data using **Power Query**, ensuring data accuracy and consistency.
- Implemented advanced **DAX** formulas to create columns and measures for in-depth analysis, including update time display, discount calculations and fiscal year-specific insights.
- Designed a one-page interactive dashboard with key metrics, including total sales, customer counts, product sales, and discount analysis.
- Incorporated various **visualizations** for comprehensive sales insights, making the data easily interpretable.
- Enabled scheduled refresh to ensure real-time data updates, keeping the dashboard always up-to-date with the latest data.
- Published the dashboard to the **Power BI** service for broad access and distribution within the organization.
- Collaborated with stakeholders to understand business requirements and tailor the dashboard to meet their needs.
- Conducted user training sessions to ensure effective use and interpretation of the dashboard.

Technologies Used: Microsoft SQL Server, Power BI services, Power Query, DAX

SynergisticIT, Fremont, CA
Data Analyst / BI Analyst

Project: Retail Chain Transaction Analysis

June – August 2023

Developed a comprehensive **Power BI dashboard** to analyze retail chain transactions, providing actionable insights into product sales, customer behavior, seasonal trends, and the effectiveness of promotions. The dashboard facilitated data-driven decision-making with detailed visualizations and interactive features.

Roles and Responsibilities:

- Utilized **Power Query** to clean and transform data, including splitting and unpivoting the Product column.
- Developer advanced **DAX formulas** to create calculated columns and measures for in-depth analysis.
- Created detailed Product Analysis and Customer Analysis pages featuring key metrics, slicers, and interactive visuals such as **line charts, treemaps, and ribbon charts**.
- Ensured alignment and consistency across all dashboard pages for a cohesive user experience.
- Integrated product and customer analysis visuals into a combined Retail Analysis Page with interactive buttons for toggling between views using bookmarks.
- Designed and presented slides summarizing key insights and findings to stakeholders, facilitating informed decision-making.

- Conducted stakeholder meetings to gather requirements and incorporate feedback into the dashboard design.
- Provided training and support to users for effective utilization and interpretation of the dashboard.

Technologies Used: Power BI services, Power Query, DAX, Bookmarks, Slides

Department of Statistics, UChicago, Chicago, IL
Statistical Consultant
September – December 2022

Worked in a team of five consultants. Analyzed requirements from clients about data issue. Communicated with clients to verified details. Provided recommendation in data analysis and delivered consulting report.

- Suggested logistic regression application and method of grouping the patient data for a study from UChicago Medicine about the impact of a COVID medication on ventilation.
- Recommended a study from UChicago BSD about the effect of Home-based Community Services (HBCS) on Post-Acute Care (PAC) to use logistic regression without propensity score weighting.
- Advised a study from UChicago Hospital about significant of chest-to-left ventricle distance on CPR to drop highly correlated covariates. Helped determine the required sample size and linear regression models.

Skills Used: Teamwork, Client communication, Data analysis method review and feedback

Education

MS in Statistics | University of Chicago (GPA: 3.73/4) September 2022 – May 2024

Relevant Courses: Reinforcement Learning, Trustworthy Machine Learning, Deep Learning Systems, Algorithms

Scholarships: Tuition Scholarship of Statistics Master Program (2022, 2023)

BS in Mathematics | Chinese University of Hong Kong (CUHK) September 2018 – July 2022

Major Concentration: Computational Big Data Analytics; Minor: Statistics

Scholarships and Honors: BS degree with First Class Honor (2022), Undergraduate Mathematics Scholarship (2021), College Scholarships (2019, 2022)

Research Experience

Independent Researcher | Master's Thesis, Department of Statistics, UChicago
September 2023 - May 2024

Examining the Interplay Between Politicians' Facial Expressions in Media Images and News Corporation Bias.

- Constructed a name recognition model with politicians' images from Wikimedia to identify news photos.
- Analyzed facial expression logits of influential politicians from various media outlets.
- Implemented classification models and visualized results through dimensionality reduction on the logits.
- Concluded no significant effect of media orientation on facial expression selection.

Research Team Member | Department of Computer Science, UChicago
October - December 2023

Modified Attention with Non-Linear Kernels and its Impact on Few-Shot Learning.

- Collaborated within a three-fellows team.
- Trained GPT-2 models on nanoGPT using OpenWebText with replacing the dot product kernel in attention mechanism by Gaussian, polynomial and periodic kernels.
- Evaluated the models with MMLU, ARC and Translation tests. Determined that traditional dot product kernels performed best overall, with some non-linear kernels excelling in specific tests.
- Produced a poster, share results and communicated with other research teams and instructors at the final seminar.

Project Leader | Department of Statistics, UChicago
April - May 2023

Robustness to Spurious Correlations via Distributionally Robust Optimization (DRO).

- Led a team of three researchers, coordinating tasks and communication.
- Reviewed and analyzed the theory of DRO.
- Generated an MNIST dataset with spurious correlations. Applied DRO and empirical models with different levels of regularization on the dataset.
- Made comparison on performances and concluded DRO model effectively eliminates the influence of spurious correlations.