Curriculum Vitae

Yuna Liu

Profile

Yuna Liu is an experienced data engineer and data scientist, who in addition holds a PhD in Economics. She is proficient in working across the different stages of the data pipeline, building data-intensive dashboard, applying machine learning algorithms for finding complex data patterns that drive meaningful impact on the business.

She is now a consultant in data engineering and business intelligence engineering at Sales Dashboard team at Volvo Cars. Before that , she has gathered valid experiences from a project on predicting default probability based on old credit information at Collector Bank, and another project on building data pipelines to automate the deployment of battery cell data at Volvo Cars.

Worked as researcher and lecturer, she was well acquainted with statistical modeling and casual analysis. In this role she was also used to interact with an audience.

In addition to her PhD in Economics, she has completed a two-year program "Developer in Artificial Intelligence and Machine Learning".

Personally, she is structured, solution-oriented, enthusiastic and self-motivated.

Job
Data Analyst

Education

Developer in AI and Machine learning, IT University, Gothenburg, Sweden

Ph.D. in Economics, Umeå University, Umeå

Master in Economics and Management, Humboldt University of Berlin, Germany

Skills
Back-end development
Databricks Data Intelligence Platform
Enthusiastic
Exploratory data analysis
Front-end development
Interpersonal communication
Machine learning
Machine learning classification
Matillion ETL
Microsoft Azure
Microsoft Azure SQL Database
Microsoft Power Automate
Microsoft Power BI
Python
Self-motivation
Services
Snowflake
Solution-oriented
SQL
User support
AutoML

Capgemini

Professional Experience

Business Intelligence Engineer/Data Enginner | Volvo Cars
September 2023 – July 2024

Work in the area of application management of both backend and frontend for Sales Dashboard team in Volvo Cars. Take responsibility in user support in PowerBI report and Azure Analysis Services(Cube), maintenance and debug by using SQL in Snowflake Database, and development of data engineering pipelines within Azure, Snowflake and Matillion.

Applied Skills

| Application management | Microsoft Azure | Back-end development | Front-end development | Matillion ETL | Microsoft Azure Analysis Services | Microsoft Azure DevOps | Microsoft Business Intelligence | Microsoft Power Automate | Microsoft Power BI | Scrum | Services | Snowflake | SQL | User support |

Data Engineer | Volco Cars
January 2023 – May 2023

About the customer: Cell durability team, at Battery cell and module department under Traction Battery R&D at Volvo Cars in Torslanda. The team works on optimizing individual Li-Ion Battery systems.

Project: This project explores how to use data engineering techniques and cloud tools to design and implement data pipelines for cell durability data, and in the next step, automation of a dashboard in Power BI for the Battery team at Volvo Cars. The business value is to use data engineering techniques to improve the efficiency and effectiveness of data-driven decision making.

Roles and responsibilities: Data Engineer

I was responsible for designing and implementing data pipelines that automate the integration of data from local disk to Azure Blob Storage, and later to Azure SQL Database, and transform it into a usable format. This pipeline required me to utilize my knowledge of cloud-based data warehousing, ETL/ELT development and automation tool such as Microsoft Power Automate.

- Worked on cell test and simulation data at cell durability team;
- Built data pipelines using Azure to automate deployments of cell data in dashboard using Python, SQL, Azure technologies, Power BI and automation tools such as Microsoft Power automate;
- Developed and maintained ETL processes, and data pipelines to ensure data accuracy and completeness, which improve performance and reduce processing time by more than 70%;

Roles and responsibilities: Data Analyst

I was responsible for creating dashboard which connected to a regularly automatically refreshed dataset, in conveying insights to stakeholders. I created compelling visualizations using tools like Tableau and Power BI, which allowed me to communicate complex data insights in a clear and concise manner to cross-functional teams. Once the dashboard was created, I deployed it as an app so that end-users could easily access it without the need for either a Power BI account or Azure SQL Database credentials.

- Provided regular reports to stakeholders in Power BI and Tableau;
- Collaborated with cross-functional teams to gather requirements, prioritize tasks, and deliver high-quality solutions;
- Conducted workshops with senior cross-functional team and stakeholders to present findings and gather feedback for continuous improvement

Method and technology: Microsoft Azure SQL Database, Microsoft Azure SQL Server, Microsoft Azure Blob Storage, Microsoft Azure Data

Capgemini

Factory, Microsoft Power Automate, Microsoft Power BI, Tableau,
Presentation Skills, Communication & interpersonal skills

Applied Skills

| Microsoft Azure SQL Database | Microsoft Azure Blob Storage | Microsoft Azure Data Factory | Microsoft Azure SQL Server | Microsoft Power Automate | Microsoft Power BI | Python | Tableau |

Data Scientist/Machine Learning Engineer | Collector Bank
September 2022 – November 2022

About the customer: Collector Bank AB specializes in financing solutions for corporate and private individuals, with particular focus on small and medium-sized companies. The main business is conducted through four segments namely Corporate, Real estate, Consumer and Payments.

Project: This project is first to clean, transform, integrate and visualize credit information data which stored at Azure Data Lakehouse using Databricks data platform; and later on to predict default probability using old credit information bought from UC credit, to be more precisely, I used machine learning algorithms to estimate the default rate of a customer who was rejected previously and then make a next application.

The motivation is to save the cost on buying credit information from UC credit for customers applying for private loans repeatedly within short periods.

Roles and responsibilities: Data Scientist

- Focused on private loans' customer credit data which stored in Data Lakehouse.
- Predicted the default probability based on credit information.
- Conducted exploratory data analysis and statistical analysis;
- Developed machine learning models to solve default detection problem, using tools such as scikit-learn, TensorFlow, and pyTorch;
- Applyed Logistic Regression, Random Forest, XGBoost, MLP, SVM, Kmeans etc.

Roles and responsibilities: Data Analyst

- Data cleaning, data transformation and data integration of different datasets;
- Visualize raw data into insightful findings and visualizations using Python, Matplotlib, Seaborn, Plotly;

Method and technology: Microsoft Azure Databricks, PySpark, Python, SQL, Plotly, Matplotlib, Seaborn, Machine Learning (ML), XGBoost, Random forest, Regression Analysis, K-means, PyTorch, TensorFlow

Applied Skills

| Microsoft Azure | Data cleansing | Data integration | Exploratory data analysis | K-means clustering | Logistic regression | Machine learning | Matplotlib | Microsoft Azure Databricks | Plotly | PySpark | Python | PyTorch | Random forests | Regression analysis | Scikit-learn | Seaborn | SQL | Statistical analysis | TensorFlow | XGBoost |

Certificates

- Databricks Certified Machine Learning Associate – Databricks – 2024
- Databricks Certified Generative AI Engineer Associate – Databricks – 2024
- Microsoft Certified: Azure Data Engineer Associate – Microsoft – 2024
- Matillion ETL Foundations – Matillion – 2024
- SnowPro Core Certification – Snowflake – 2024
- Microsoft Certified: Azure Data Fundamentals – Microsoft – 2024
- Microsoft Certified: Azure Fundamentals – Microsoft – 2023

Additional skills

Capgemini

Technical: | Machine learning | Application management | K-means clustering | Logistic regression | Matplotlib | Pandas | Plotly | PySpark | PyTorch | Random forests | Regression analysis | Scikit-learn | Seaborn | Software development | Statistical analysis | SVM | TensorFlow | XGBoost | Data science |

Functional: | Back-end development | Exploratory data analysis | Front-end development | Machine learning classification | Services | User support | AutoML | Artificial intelligence | Corporate finance | Data analysis | Data cleansing | Data integration | Data pipeline | Data processing | Data transformation | Data-driven decision-making | Data engineering | Generative AI |

Behavioral: | Enthusiastic | Interpersonal communication | Self-motivation | Solution-oriented |

Business software: | Microsoft Power BI | Tableau |

Methods: | Microsoft Azure DevOps | Scrum |

Tools: | Matillion ETL | Microsoft Power Automate | Data Lakehouse |

Programming languages: | Python | SQL |

Frameworks: | Apache Spark | Spark ML |

Databases: | Microsoft Azure SQL Database | Snowflake | Microsoft Azure SQL Server | Microsoft SQL Server |

Platforms: | Databricks Data Intelligence Platform | Microsoft Azure | Microsoft Azure Analysis Services | Microsoft Azure Blob Storage | Microsoft Azure Data Factory | Microsoft Azure Databricks | Microsoft Business Intelligence | Azure Databricks | Databricks MLflow |

Languages

Chinese : Native
English : Native
Swedish : Fluent

Capgemini