

Yengkong Vang Sayaovong

414-739-1246 • Milwaukee, WI (Open to Relocation) • YSayaovong@gmail.com • <https://linkedin.com/in/ysayaovong>
<https://github.com/YSayaovong> • <https://ysayaovong.github.io/Portfolio/index.html>

Machine Learning / Data Engineer

Data Engineer with hands-on experience from a previous internship, specializing in Python, SQL, and data visualization. Skilled in analyzing complex datasets, automating data workflows, and generating actionable insights to support business decision-making. Proficient in data warehousing, ETL processes, and creating interactive dashboards that drive strategic growth. Actively seeking a machine learning or data engineer position to utilize my analytical skills and contribute to meaningful data-driven solutions.

Career Highlights

Data Engineer Intern | Refonte Infini – Remote 11/2024 – Present

- Implemented data ingestion with tools such as Sqoop, Flume, and Kafka.
- Optimized Hive queries using partitioning, bucketing, and indexing.
- Processed large datasets and real-time data using Apache Spark.
- Managed AWS services like RDS and EBS for secure and scalable solutions.
- Applied AWS deployment options tailored to business needs.
- Processed data on Amazon EMR using Hadoop tools.
- Built real-time data solutions with Amazon Kinesis and visualized data using QuickSight.
- Deployed Azure virtual machines, web apps, and databases.
- Integrated Azure Active Directory for secure identity management.
- Designed data solutions with Azure Synapse, Data Lake, and SQL Database.
- Developed batch and streaming data processes for performance optimization.

Mechanical Designer | Prolec-GE Waukesha – Waukesha, WI 10/2020 – Present

- Designed transformers and nameplates, ensuring compliance with industry standards using CREO 8.0.
- Led process optimization initiatives, reducing errors and improving design efficiency by 10%.
- Mentored 6 detailers, providing guidance to improve overall team performance.
- Worked cross-functionally to ensure timely and high-quality project delivery.
- Facilitated production and training sessions to enhance team knowledge and ensure efficient operational workflows.

Volunteer Experience

Board of Trustees Member | Hmong First Baptist Church | Milwaukee, WI 12/2023 – Present

- Participated in decision-making processes and maintained records for board meetings.

Pianist | Hmong First Baptist Church | Milwaukee, WI 1/2023 – Present

- Organized musical arrangements and collaborated with worship teams for weekly services.

Assistant Treasurer | Hmong First Baptist Church | Milwaukee, WI 11/2018 – 12/2021

- Managed financial records, budgeting, and payroll to ensure compliance with financial regulations.

Education

B.S., Information Technology; Minor in Music | Arizona State University, Expected Completion: 5/2025

A.I. & Machine Learning Engineer Career Path | Zero to Mastery, Expected Completion: 5/2025

A.S. in Mechanical Design Technology | Milwaukee Area Technical College, 5/2021

Technical Proficiencies

Languages:	<p>Python (Advanced): Skilled in object-oriented programming (OOP), data structures, algorithms, backend development (Flask, Django), and process automation.</p> <p>JavaScript (ES6+): Used in backend services, asynchronous operations, and event-driven architecture.</p> <p>C++ (Basic), C# (Basic), Java (Basic): Basic knowledge of syntax, programming logic, and foundational concepts.</p>
Machine Learning & Data Science:	<p>Scikit-learn, Pandas: Skilled in building, training, and fine-tuning machine learning models with scikit-learn, and proficient in Pandas for efficient data manipulation and preparation, enabling effective feature engineering and model performance.</p> <p>Feature Engineering & EDA: Strong ability to identify and create meaningful features, coupled with deep experience in exploratory data analysis to uncover patterns, trends, and insights from complex datasets.</p> <p>NLP: Hands-on experience with natural language processing, including text preprocessing and sentiment analysis, to derive insights from unstructured data sources.</p>
Data Engineering:	<p>Apache Kafka, Spark, Hadoop: Proficient in developing scalable, high-performance data pipelines with Kafka and Spark, and leveraging Hadoop for distributed processing of large datasets, enabling real-time and batch data ingestion.</p> <p>ETL Pipelines & Data Warehousing: Experienced in constructing end-to-end ETL pipelines that ensure data quality and accessibility, and skilled in designing data warehouses that support comprehensive business intelligence</p>
Cloud & DevOps:	<p>AWS Services: Experienced with EC2, S3, Lambda for scalable infrastructure, production deployment, and serverless computing.</p> <p>CI/CD & Containers: Proficient in Docker, Jenkins, and GitHub Actions for automated builds, testing, and deployments.</p> <p>Azure Synapse: Proficient in using Azure Synapse for unified analytics, integrating large-scale data solutions that support efficient data exploration, management, and analysis.</p>
Database	<p>PostgreSQL, SQL, NoSQL (MongoDB, Couchbase): Proficient in writing complex queries, managing relational and NoSQL databases, and executing ETL processes for data operations. Skilled in optimizing data storage and retrieval strategies to support large-scale data workflows.</p>
Tools:	<p>PostgreSQL, SQL, Apache Kafka, Hadoop, Spark, Redshift, Snowflake: Proficient in leveraging these tools to build, manage, and optimize data pipelines, ensuring high performance and reliability in data engineering environments.</p>

Projects

Predictive Analysis and Data Insights on the Titanic Dataset

November 2024

Description: Conducted a comprehensive analysis of the Titanic dataset to uncover factors influencing passenger survival and developed a predictive model to estimate outcomes.

- **Data Preprocessing:** Addressed missing values, removed irrelevant columns, and prepared the dataset for analysis.
- **Exploratory Data Analysis (EDA):** Utilized Python libraries (Seaborn, Matplotlib) to visualize survival trends across variables such as class and age.
- **Feature Engineering:** Selected and prepared key features for optimal model training.
- **Machine Learning Model:** Developed a RandomForestClassifier in scikit-learn, evaluating performance with metrics including confusion matrix and classification report.

Key Skills: Python, Data Analysis, Machine Learning, Data Visualization, EDA, scikit-learn, pandas, Seaborn, Matplotlib.