Suman Paudel

DATA SCIENTIST

dastonsuman1997@gmail.com | +9770 9866142700 | LINKEDIN | GITHUB

PROFILE SUMMARY

Highly qualified and motivated data scientist with over 2 years of data science, machine learning, and engineering roles, helping companies to meet and exceed projected expectations along with personal growth. Eager interest in developing data-intensive applications, LLMs, machine learning models, statistical modeling, and data science solutions.

WORK EXPERIENCE

Data Scientist | BitsKraft Apr 2021- Present

- Worked on Credit Risk Modelling using classical machine learning models for multiple banks, delivering accurate predictions and identifying relationships, trends, and factors that affected the results of the predictive model.
- Worked on Customer Segmentation by utilizing unsupervised learning in the finance and banking industries to uncover customer patterns and use insights for better business strategies using the CLV method.
- Query databases and data warehouses for dashboards and reporting using Azure SQL, Python, PostgreSQL, and Power BI which translated complex data insights into actionable strategies for stakeholders across the PMO office, HR Department, and Internal Data Analytics Team.
- A lifelong learner, stayed up-to-date with emerging analytical trends and technologies by reading scientific articles, conference papers, and online resources, and implementing them on live projects.

PROJECTS INVOLVED

- **Credit Risk Modelling**
  - Developed predictive models to assess customer risk using classical machine learning algorithms. My work involves gathering and preprocessing historical loan data, conducting exploratory analysis to uncover patterns, selecting relevant features, and training models on training data. I evaluate model performance using various metrics such as F1 score, precision, recall, accuracy, and confusion matrix.
  - Tools Used: SQL, Python, Machine Learning (XGBoost, Logistic Regression, Random Forest, Decision Tree, and SVM)

- **Customer Segmentation**
  - Developed a model focused on Customer Segmentation using Customer Lifetime Value (CLV). This project involved collecting and preprocessing customer transaction data, conducting exploratory analysis to identify patterns and correlations, and selecting relevant features for segmentation. Subsequently, I employed advanced techniques such as RFM (Recency, Frequency, Monetary) analysis and clustering algorithms to segment customers based on their CLV -- namely Low, Medium, and High. The segmentation results were evaluated for effectiveness using metrics such as silhouette score or within-cluster sum of squares (WSS).
  - Tools Used: SQL, Python, Machine Learning (Unsupervised Learning)

- **Analyzing Log Data Set**
  - Analyzed and Processed Huge Log Data sets of the security team for threat analysis. Log data was generated by Elastic Search and was dumped to the server (Daily > 10gb), then using Apache Spark loaded the file then I ingested and transformed the data. The transformed data was often the CSV file or loaded into an SQL table for further analysis. After the data was loaded, I used to make a dashboard to extract useful insights from the data as per the given business logic. (The solution was often rule-based rather than classical ML models.
  - Tools Used: Apache Spark, Python, MSSQL, and Power BI.

EDUCATION

Tribhuvan University — Balkhu, Kathmandu
Master in Data Science — Jan 2024 - Present

Chandigarh University — Punjab, India
Bachelor in Computer Science and Engineering — Aug 2017 - Jun 2021

CORE TECHNICAL SKILLS

- Programming Languages: Python, SQL, C++
- Machine Learning and Deep Learning Framework: Pandas, NumPy, Scikit Learn, Decision Tree, Random Forest, Linear Regression, Logistic Regression, TensorFlow, PyTorch, LLMs, Prompt Engineering
- Platform: Windows, Linux, Mac, AWS, Azure
- Theory: Mathematics, Statistics, and Linear Algebra
- Data Visualization: Tableau, Power BI, matplotlib

OTHER SKILLS

- Language: Nepali (Native), English (C1), Hindi (Fluent)
- Analytical Mindset and continuous learner.
- Leadership Skills
- Communication Skills

SIDE PROJECTS FOR RESEARCH/HANDS ON LEARNING

- **Document Bot using LLM with Rag**
  - Created a production grade document chat bot which can help you get the questions and answers from your own document related to the topic and the LLM answers to the prompt almost correctly. Also fully dockerized the application while also able to integrate with other vector databases as well. Tech Stack used:LlamaIndx, Qdrant Vector Database, OPENAI/GEMINI models, HuggingFace Embeddings, Docker, Python. You access the source code here.

- **Cash Flow Prediction Using Tensorflow**
  - Created a predictive algorithm that can predict the cash demand for the given bank. The model provides insights into the factors driving cash flow predictions, helping financial analysts and decision-makers understand the rationale behind the forecasts.

- **Music Genre Classification Using PyTorch (torch audio)**
  - Created a classification application that can classify the genre of music. The classification application leverages pre-trained models and transfer learning techniques, enabling faster development and better performance, especially with limited labeled data.