Tushti Vinod Verma

Data Scientist

+1 (646) 492-2693 || Email || LinkedIn || Github || New Jersey, USA

PROFESSIONAL SUMMARY

Data Scientist with **5+ years of experience in data science, machine learning, and software development**, specializing in **NLP, fine-tuning LLMs, and advanced ML/AI models** for impactful business solutions. Proficient in applying **LSTM models for time series forecasting,** enhancing predictive accuracy for financial and operational metrics. Demonstrated expertise in **regression modeling (Linear, Logistic, and Polynomial Regression)** to analyze and forecast business trends, supporting data-driven decision-making. Skilled in optimizing data pipelines and managing **ETL/ELT workflows across AWS and Azure**, with proven expertise in **YOLOv8 for object detection, GPT for text extraction, and RAG** for enhanced retrieval accuracy. Proficient in **Python, SQL**, and visualization tools like **Tableau and Power BI**, delivering actionable insights to drive business decisions and refine solutions based on feedback. Experienced in implementing robust **version control strategies using Git** and designing compliance workflows for policy enforcement, ensuring adherence to organizational standards and efficiency.

PROFESSIONAL EXPERIENCE

Data Science Engineer, ***Raven Risk AI, Inc*** **Dec 2023 - Present**

- Managed **ETL/ELT integrations across AWS S3** using Python and SQL establishing hybrid cloud workflows that improved data consistency.
- Worked on **RAG application** to enhance retrieval accuracy, **using semantic chunking** for context segmentation and **retrieval methods (MMR, hybrid, similarity-based)**. Integrated **reranking models with JINA and Cohere** to boost relevance and faithfulness of responses.
- Faced challenges with accurate extraction of tables and charts from complex documents. To resolve this, we trained a **YOLOv8 object detection model using Roboflow**, achieving precise detection of tables and charts with a **mAP of 95.2% and 96.7%.**
- Utilized **GPT & Claude models, OCR,** for advanced text extraction and data cleaning, leading to superior data quality.
- Integrated a **multi-thread executor** into the YOLOv8-based preprocessing pipeline, reducing preprocessing time by over **50% (1 hour to 5-10 mins)** and boosting operational efficiency.

Jr Data Scientist, ***Principal Financial Group, Inc, USA*** **May 2022 - Dec 2023**

- Cleaned and standardized data using **Python and Pandas**, **improving data quality by 30%** for accurate analysis.
- Conducted comprehensive **EDA using Python and R**, visualizing trends and insights in **Tableau and Power BI** to inform data-driven decisions and guide model selection.
- Designed and implemented **A/B testing strategies** to evaluate feature changes and optimize user engagement, ensuring statistical significance and actionable results.
- Performed **time series analysis and developed LSTM models** to forecast trends, **achieving a 12% improvement** in predictive accuracy for revenue projections.
- Measured model performance with **Weights & Biases to track experiments**, adjusting model parameters to enhance **accuracy by 15%** and **reducing computational costs by 20%**.
- Leveraged version control systems **Git and Bitbucket** for collaborative code development, maintaining accurate version history and minimizing conflicts in model and code updates, **resulting in a 25% reduction in code conflicts.**
- Presented actionable insights and final model outcomes to **senior stakeholders, incorporating their feedback** to refine results and align solutions with business goals.
- Enhanced compliance monitoring by leveraging third-party tracking tools to **identify and report on policy violations.**

Jr Data Scientist, ***EasyPack Softwares India Pvt Ltd*** **July 2020 - Aug 2021**

- Leveraged SQL for efficient data querying and manipulation, boosting segmentation **accuracy in K-means clustering by 25%**, leading to improved customer insights and targeted marketing strategies.
- Built a Random Forest multiclass classifier to profile **new customers with 86% accuracy**, enabling effective segmentation and enhancing personalized engagement.
- Conducted comprehensive data analysis on Azure Databricks to identify and analyze key customer behaviors, leading to an **11% increase in sales conversion rates**. The insights informed optimized customer journeys and targeted offers.

- Developed and deployed **regression models (Linear Regression, Polynomial Regression, etc.)** for sales forecasting, driving a 10% improvement in resource allocation.
- Refined machine learning models, **achieving a 20% improvement** in prediction accuracy and processing efficiency, which elevated model reliability and responsiveness in real-time applications.
- **Implemented and fine-tuned XGBoost models** to enhance classification tasks, achieving superior predictive performance and model robustness.

Software Developer Intern, *Vruksh Ecosystem Foundation* **Nov 2019 - May 2020**

- Conducted in-depth analysis on ETS (GRE) datasets, uncovering over **50 key data trends** that informed the foundation's sustainability initiatives. These insights led to targeted actions, **improving the efficiency of sustainability efforts by 20%.**
- Designed and deployed an interactive dashboard using Python and Tableau, providing real-time visualization and tracking of **10+ critical environmental metrics**.
- Built a **responsive website using HTML, CSS, and JavaScript, increasing user engagement by 25%** through improved navigation and accessibility. The website facilitated broader awareness and interaction with the organization's sustainability projects, **with a 40% increase in page visits** within the first month post-launch.

SKILLS

Methodologies:	Agile, Scrum, Kanban
Programming Languages	Python, SQL, R, Excel(VLOOKUP, SUM, AVERAGE, and SUMIF)
Cloud Technologies:	Apache Airflow, AWS, Microsoft Azure, GCP (Vertex AI), PySpark, Docker, Snowflake
Visualization Tools:	Power BI, Tableau
ML Frameworks and Tools:	Keras, TensorFlow, LangChain, Scikit-learn, spaCy, Fast API, Kafka, Weights & Biases, Roboflow
Databases:	MySQL, MS-SQL, MongoDB, Pinecone, PostgreSQL, Qdrant
Version Control:	Git, Bitbucket, GitHub, GitLab
Data Analysis/Reporting:	Excel functions: VLOOKUP, SUM, AVERAGE, SUMIF, Pivot Tables, and Macros for advanced data analysis and reporting.
ML/AI:	Neural Network Models, Supervised and Unsupervised Learning, Ensemble Models (Random Forest, Gradient Boosting), Regression Models, Classification Models, Clustering (K-means, DBSCAN), LLMs, Object Detection (YOLOv8), A/B Testing, Time Series Analysis

EDUCATION

Master of Science in Data Science**,** *New Jersey Institute of Technology, New Jersey*

- **Coursework**: Machine Learning, Deep Learning, Data Analysis in R, Natural Language Processing

Bachelor of Engineer in Computer Science & Engineering**,** *G.H. R.C.E, Nagpur*

- **Coursework:** Data Structures and Algorithms, System Design, Object Oriented Programming

CERTIFICATIONS

Deep Learning.AI - Supervised Machine Learning: Regression and Classification
Deep Learning.AI - Advanced Learning Algorithms