AWANTIKA SRIVASTAVA

Data Scientist | Statistics, SQL. Python

+91-8920482037 |sawantika81@gmail.com| LINKEDIN |Github

PROFILE SUMMARY

Data Scientist with **2+ years of experience** applying **statistical analysis, experimentation, and data science techniques** to solve real-world business problems. Strong foundation in **statistics and mathematics**, including hypothesis testing, confidence intervals, bias analysis, and model evaluation. Proficient in **SQL and Python** for end-to-end analytics—from messy data extraction to modeling and insight communication. Experienced in translating complex analytical results into **clear, actionable insights** for cross-functional stakeholders in regulated, high-impact environments.

CORE TECHNICAL SKILLS

- **Programming & Data Science:** Python (Pandas, NumPy, Scikit-learn), C++, Jupyter Notebook.
- **Statistics & Mathematics:** Statistical Modeling, Hypothesis Testing, Confidence Intervals, A/B Testing, Bias & Variance Analysis, Scenario Analysis.
- **Machine Learning:** Supervised & Unsupervised Learning, Regression, Classification, Clustering, Random Forest, Decision Trees, SVM, KNN, K-Means, XGBoost, Model Evaluation Metrics (Accuracy, Precision, Recall, F1-score, ROC-AUC).
- **Deep Learning :** Neural Networks, CNN, RNN, LSTM, Transformers (BERT), Transfer Learning, Model Fine-tuning, Computer Vision.
- **Deep Learning Frameworks:** TensorFlow, keras, Pytorch, TensorFlow Lite.
- **NLP:** Text preprocessing, Tokenization, Sentiment Analysis, Transformer-based models.
- **LLMs:** Transformers, HuggingFace, DistilBERT, Prompt Engineering, LangChain, RAG.
- **SQL & Data Wrangling:** Advanced SQL (Joins, CTEs, Window Functions), Data Cleaning, Handling Messy Data.
- **Data Visualization & Storytelling:** Exploratory Data Analysis (EDA), Business Insights, Stakeholder Communication.
- **Tools & Collaboration:** Git, GitHub, Peer Reviws, Documentation, Cross-functional Collaboration.

EXPERIENCE

Data Scientist / Machine Learning Engineer | PPS International Pvt. Ltd. **January 2024-Present**

- Analyzed **large, complex datasets** to uncover patterns, trends and insights supporting data-driven decision-making.
- Designed and executed **statistical analyses** including hypothesis testing and confidence interval estimation to validate business assumptions.
- Built and evaluated **machine learning models** (classification and regression) to support predictive and analytical use cases.
- Wrote **complex SQL queries** to **extract**, **transform**, and **analyze** data from relational and **NoSQL** data sources.
- Conducted **exploratory data analysis (EDA)** to identify data quality issues, bias, and anomalies prior to **modeling**.
- Partnered with product, compliance, and engineering teams to **translate analytical findings into clear business recommendations**.
- Designed and analyzed **A/B tests and controlled experiments** to evaluate changes in user workflows and system configurations, assessing impact using statistical significance and confidence intervals.
- Supported **peer reviews** and ensured analytical outputs met quality and documentation standards required in regulated environments.

PROJECTS

Railway Driver Assistance System (RDAS) | Enterprise ML Project

- **Designed and deployed a** real-time **computer vision–based ML system** for unsafe driver behavior detection using **CNN-based SSD MobileNet** model.
- Trained and optimized models on large-scale video datasets, achieving **20–25 FPS** real-time processing with **<150 ms inference latency**.
- Implemented **end-to-end ML pipelines** for data ingestion, preprocessing, model training, evaluation, and production inference.
- Deployed optimized models using **TensorFlow Lite** on edge/production environments for continuous monitoring.
- Built a **Flask-based web dashboard** to visualize detections and automatically record **30-second event clips**, reducing manual review effort.

Chatbot Using LLM & RAG | Applied ML Project

- Built a Lightweight **LLM-Powered chatbot** using **TinyLLaMA** to answer user queries over content.
- Implemented a **Retrieval-Augmented Generation (RAG) pipeline** to retrieve relevant resume sections for contextual question answering.
- Selected **TinyLLaMA** to ensure **low memory footprint and fast inference**, making the solution suitable for resource-constrained environments.
- Applied **prompt engineering techniques** to improve response relevance and consistency.
- Deployed the chatbot as an interactive **Streamlit web application** for real-time user interaction.

YouTube Comments Sentiment Analyzer | link - https://youtube-ai-analyzer-ndzqo6r2mepjrtsdjmwaxl.streamlit.app/

- Deployed transformer-based **NLP models** as **production-ready** services with **REST APIs**.
- **Fine-tuned** and served a **DistilBERT**-based **sentiment** classification model for large-scale text inference.
- **Built** and deployed an interactive **streamlit web application** to perform real-time **sentiment analysis** on YouTube comments.
- **Processed** high-volume text **data** with sub-second inference **latency** for real-time sentiment analysis.

CERTIFICATION

- IBM Data Science & AI Certification
- AWS Generative AI with Large Language Models
- OpenCV Computer Vision Certification

EDUCATION

IMS Engineering College, Ghaziabad **September - 2020**

Bachelor of Technology (Electrical and electronics engineering)