Upendra Tiwari

DATA SCIENTIST

CONTACT

📞 +91 790-534-9375

✉ upendra.tiwarisql@gmail.com

🐙 https://github.com/upenntiwari

💼 https://www.linkedin.com/in/
upendratiwariaiml33/

📍 27/8 Sector 45 Gurgaon
Haryana 122001

SKILLS

- Python, SQL, Git, Docker
- CNNs, RNNs, LSTMs, Transformers
- Data Processing & Feature Engineering
- Statistical Analysis & Predictive Modeling ,
- NLP & Text Analytics ETL &
- Data Pipeline
- Development: Apache Airflow, SQL, Pandas MLOps & Model Deployment
- Power BI, Tableau, Matplotlib, Seaborn A/B
- Testing & Experimentation AWS, GCP, Azure Business
- Tensorflow, Scikit Learn and Pytorch
- Algorithms
- NLTK, Hugging Face, Llama, Groq, Langchain
- Rag, LLMs
  Prompt Engineering

PROFILE SUMMARY

Experienced Data Scientist with expertise in machine learning, deep learning, NLP and Gen AI. Proficient in Python, SQL, and big data processing, with a strong foundation in data-driven decision-making. Skilled in designing and implementing AI/ML models to drive business insights and innovation. Adept at statistical modeling, NLP, and data pipeline development. Passionate about continuous learning and leveraging data science to solve real-world problems.

WORK EXPERIENCE

■ **TechnoSapphire Research**
Data Scientist
Feb 2024 – Present | Noida
- Designed and deployed machine learning models for large-scale data analysis and insights generation.
- Developed deep learning architectures (CNNs, RNNs) for NLP and computer vision applications.
- Built automated ETL pipelines for data preprocessing and feature engineering using Python.
- Optimized SQL queries for efficient database performance in predictive modeling.
- Created interactive dashboards to visualize model performance and business insights.
- Implemented MLOps best practices, including CI/CD pipelines for machine learning model deployment.
- Worked with cloud-based data lakes and warehouses (AWS, Azure) to manage large-scale datasets.
- Been part of several Gen AI and Agentic AI projects to enhance skills in the field of developing Gen AI field.

■ **HDFC Bank**
Data Scientist
Sep 2022 – Feb 2024 | Thane
- Collaborated with cross-functional teams to identify business challenges and translate them into data science solutions.
- Performed feature selection, dimensionality reduction, and model interpretability for better decision transparency.
- Designed and implemented custom evaluation metrics tailored to business goals, improving model relevance to stakeholders. Applied time series forecasting techniques (ARIMA, LSTMs) to predict demand trends and optimize inventory planning.
- Leveraged Spark and distributed computing to train models on large datasets efficiently.
- Worked on Rag Principles , Pinecone database and Langchain in projects

EDUCATION

JUL 2013 - Jun 2017
GLA UNIVERSITY

Bachelor of Technology

CERTIFICATIONS

Certification in Statistics
and Mathematics for Data
Science
Machine Learning, Data
Science, and Generative AI
with Python
Deep Learning
Specialization
Supervised Machine
Learning
Python and SQL for Data
Science

LANGUAGES

English (Fluent)
Hindi (Fluent)

Tata Capital Ltd. Data Analyst
Sep 2021 – Aug 2022 | Mumbai

- Conducted data cleaning, transformation, and statistical analysis on large financial datasets.
- Developed machine learning models using Scikit-Learn for risk assessment and fraud detection.
- Automated data extraction from multiple sources using Python and SQL.
- Designed deep learning architectures (LSTMs, RNNs) for NLP-based financial document processing.
- Built interactive dashboards in Tableau and Power BI to visualize financial trends and KPIs.
- Optimized data pipelines for real-time reporting and enhanced decision-making across business units.

Aditya Birla Finance
Data Analyst
Oct 2020 – Aug 2021 | Noida

- Analyze customer loan data to identify trends, risk patterns, and business opportunities.
- Deliver actionable insights to optimize lending strategies and portfolio performance.
- Develop and maintain interactive dashboards and regular MIS reports using SQL, Excel, and BI tools (Power BI/Tableau).
- Track KPIs like loan disbursal rates, delinquency, NPA trends, and customer acquisition costs.
- Collaborate with IT and operations teams to clean, validate, and enrich loan datasets.

Idfc First Bank
Sales Manager
Dec 2019– Aug 2020 | Delhi

Bajaj Finserv
Sales Manager
May 2019 – Dec 2019 | Delhi

Paisabazaar.com
Subject Matter Expert
Dec 2017 – Feb 2019 | Gurgaon

PROJECTS

✅ **Product Recommendation System using Cosine Similarity**
- Engineered a scalable content-based product recommendation system using cosine similarity, TF-IDF, and ML techniques to assess customer preferences and products metadata.
- Deployed using Docker on cloud platform with a front-end UI improving user experience through personalized product suggestions.

✅ **Sentiment Analysis on IMDB Reviews with LSTM**
- Created an LSTM-based deep learning model to classify sentiment on IMDB movie reviews. Preprocessed text data with tokenization, padding, and used GloVe word embeddings. Deployed the model via Streamlit,
- enabling real-time sentiment classification. Achieved 91%+ validation accuracy.

✅ **GPT-4o ChatBot using Streamlit**
- Built an interactive chatbot using OpenAI's GPT-4o with memory tracking, Streamlit for UI, and integrated token-based conversation history for context awareness.
- Demonstrated capabilities in generative AI, NLP, and model integration for user-friendly AI tools.

✅ **Customer Churn Prediction Using Machine Learning**
- Designed and evaluated churn prediction models using Decision Tree, Random Tree Classifier, and XGBoost Random Forest, improving forecast accuracy for customer retention.
- Utilized SMOTE for handling imbalanced data, performed robust feature engineering, and applied model interpretability tools to drive business decisions and enhance user engagement.

✅ **Deploying Machine Learning Model on Azure with Python**
- Deployed an end-to-end ML pipeline using Azure ML SDK, Python, and Scikit-learn for scalable cloud training and model versioning.
- Used Azure ML Studio, Azure Container Instances, and pipeline automation tools to deploy and monitor the model in a real-time production environment.

✅ **Built AI Agents using CrewAI, PydanticAI, HuggingFace, OpenAI and others.**

✅ **Spam Detection using Logistic Regression**
Developed a binary classification system to detect spam emails using Logistic Regression with TF-IDF vectorization, NLTK preprocessing, and data cleaning pipelines.
Achieved 96% accuracy, validated with cross-validation and ROC-AUC metrics, improving email security and filtering efficiency.