VISHWAJIT SEN
E-Mail: vishu1009@gmail.com
Contact No.: +91 8105445280, 8454954471
LinkedIn Profile: www.linkedin.com/in/vishwajit-sen-575ab031
Notice Period: Immediately Available
Available for Remote, Onsite, Freelance

Professional Summary

Data Science/AI Leader | Transforming Insights into Innovation

- ✓ I am a seasoned data science AI leader with over 15 years of experience, combining deep technical expertise with strategic vision to drive meaningful business outcomes. Currently, I am focused on building a cutting-edge conversation AI product utilizing technologies such as Python, PyTorch, Hugging Face, Large Language Models (LLMs), BERT, Dataiku, Sentence Transformers, word embedding models, matching algorithms, conversational AI, prompt engineering, and SBERT.

- ✓ Also, have experience in working with Seismic data **(. segy format)** for data analysis and identifying the bright spot in **Oil & gas sector**

- ✓ Throughout my career, I have successfully delivered high-impact solutions across diverse industries, including **manufacturing, healthcare, banking, pharmaceuticals, automotive, insurance, consulting and technology**. My passion lies in translating complex data into actionable insights that empower organizations to make informed decisions, foster innovation, and achieve sustainable growth.

- ✓ ✅Successfully completed an end-to-end MLOps training program designed for real-world AI project deployment. Gained deep, hands-on experience across the full machine learning lifecycle — from Linux-based environment setup on AWS EC2 to version control using Git/GitHub and DVC for collaborative model development.
  ✅ Developed strong expertise in experiment tracking using MLflow and DagsHub, containerization using Docker with custom image creation, and orchestrating production-grade ML pipelines.
  ✅ Implemented robust CI/CD workflows for scalable ML deployment using GitHub Actions, Jenkins, and CircleCI. Built and deployed real ML projects with integrated automation and monitoring pipelines.
  ✅ Acquired practical working knowledge of Kubernetes for container orchestration and used AWS SageMaker for scalable model training and deployment.
  ✅ Integrated Grafana for real-time model and infrastructure monitoring, ensuring transparency and reliability in production environments.
  ✅ Equipped with a strong foundation in MLOps principles, tools, and cloud infrastructure to lead cross-functional AI/ML initiatives, streamline deployment processes, and drive enterprise-scale AI adoption.

- ✓ **Technical Excellence:** Proficient in Python, R, SAS, SQL, cloud platforms (AWS, Azure), Big Data technologies, PySpark, deep learning frameworks (PyTorch, Dataiku, Transformers), Agentic AI, MLOps, data visualization tools (Power BI, Tableau), and advanced AI applications including Generative AI, LLMs, Retrieval-Augmented Generation (RAG), Computer vision problems like Image classification using Transfer Learning (VGG16, Resnet etc.), Object detection using Faster R-CNN, YOLO, Image segmentation, Langchain, Llama Index, Matching Algorithms using BERT, SBERT, voice cloning using XTTS_v2 and Hugging Face.

- ✓ **Strategic Leadership:** Experienced in building and leading high-performing data science teams, cultivating a culture of innovation, and aligning technical solutions with business goals through strong stakeholder collaboration.

- ✓ **Impact-Driven Delivery:** A results-oriented professional with a consistent record of delivering measurable business value, driving strategic initiatives, and exceeding expectations.

- ✓ As a lifelong learner, I am committed to staying at the forefront of the rapidly evolving data science landscape, continually seeking new challenges and opportunities to push the boundaries of what data and AI can achieve.

- ✓ Coached multiple folks to successfully build their career in the data science and problem-solving using ML/DL algorithms.

- ✓ Worked closely with the business heads / stakeholders for the Digital/Innovation / Analytics COE team in understanding the business problems and building the end-to-end problem solutions.
- ✓ A Six Sigma Lean certified and Six Sigma Green Belt Trained and Tested.
- ✓ Gained skills in undertaking analytics assignments for clients entailing collation and analysis of data and making recommendations based on available facts and conclusions.
- ✓ Excellent exposure of taking sessions on R, PYTHON and Machine Learning for the entire Analytics Team
- ✓ Strong theoretical understanding of ML/DL algorithms.
- ✓ Instrumental in building relations with upper-level decision makers, seizing control of critical problem areas and delivering on client commitments.
- ✓ Passionate about continuous learning and gaining new skills related to Statistics, Data Science, Leadership Principles and Project/ People Management.
- ✓ Also mentored students in the Springboard's Data Science Foundation workshop from Nov'15 onwards.

Key Projects

Voice-Based Conversational AI Assistant

Objective:

Built a real-time voice assistant AI that transcribes spoken input, retrieves intelligent responses using semantic similarity or an LLMs (Llama, Mistral, BlenderBot, Microsoft Phi2), and speaks the response back to the user.

Key Features:

- Audio recording and transcription using FFmpeg and Open AI Whisper.
- Semantic intent matching with Sentence-BERT for context-aware responses from a conversation database (Excel).
- Dynamic fallback to Mistral 7B LLM via llama-cpp for open-ended queries.
- Natural voice output via Edge TTS and PyDub playback.

Voice-Driven Semantic Matching System for Service Recommendations

Objective:

Developed a real-time voice assistant that identifies user intent related to service requests and matches it with a best-fit provider using semantic similarity and sentiment analysis.

Key Features:

- Captures voice input via sounddevice and transcribes speech using OpenAI Whisper.
- Uses Sentence-BERT for semantic similarity between user queries and a structured dataset (Excel).
- Adds a guardrail filter to validate relevant domain-specific intent.
- Performs sentiment analysis using a pretrained DistilBERT model to personalize responses.
- Generates natural spoken responses via Edge TTS and PyDub.

3D Seismic Geobody Extraction and SEG-Y Export Tool

Objective:

Developed a Python tool to identify bright spot geobodies from seismic data and export them as industry-standard SEG-Y files, preserving spatial headers for direct integration with interpretation software like OpendTect.

Key Features:

- Loaded and structured 3D seismic volumes using SEG-Y IO (segyio).
- Isolated high-amplitude geobodies via thresholding and masking using NumPy.

- Exported cleaned SEG-Y files with correct inline, crossline, and spatial (X/Y) metadata in IEEE float (Format 5).
- Visualized the geobody structure in 3D using Mayavi for QC and interpretation.

Credit Risk Prediction System Using Machine Learning Models

Customer Churn Forecasting in Life Insurance Using XGBoost and SHAP Explainability

Fraud Detection in Real-Time Banking Transactions with Anomaly Detection Algorithms

Built My Own Agentic AI Voice Assistant — Runs Daily Like Clockwork!

Professional Experience

Delphi Consulting Middle East (Remote)

Principal Consultant
Jan 2024 - Apr 2024

- Led the entire data science practice team and worked for the clients in UAE Market
- **Business Use Case:** Predicting Insurance Claim Denials (Detailed Explanation Below)
- **Business Use Case:** Forecasting Financial Metrics in Healthcare
- **Objective:**
  To forecast key financial metrics—claim amount, paid amount, and unremittance amount—for a UAE-based healthcare client. The goal was to enhance financial planning and improve revenue cycle management by predicting future cash flows with accuracy.
- **Data Sources:**

  Historical financial data from the client, including claim submissions, payment records, and unremitted amounts

  Patient demographic information and treatment details

  External economic indicators relevant to the healthcare sector
- **Features:**

  Time-series data of claim amounts, paid amounts, and unremittance amounts

  Lagged variables to capture temporal dependencies

  Seasonality components (monthly, quarterly) to account for cyclical patterns

  External factors such as policy changes or economic conditions
- **Models Used:**

  ARIMA (Autoregressive Integrated Moving Average): Applied for its strength in modelling linear time-series data and capturing trends and seasonality

  LSTM (Long Short-Term Memory): Leveraged for its ability to model complex temporal dependencies and nonlinear patterns in financial data

  Prophet: Utilized for its flexibility in handling seasonality and holidays, making it well-suited for healthcare

financial forecasting

- **ROI and Business Impact:**

  Improved accuracy in financial forecasting by up to 20%, leading to better budget planning and resource allocation

  Enabled proactive decision-making by predicting cash flow challenges, allowing the client to mitigate risks

  Enhanced overall financial stability, contributing to more predictable and sustainable operations

Mahindra & Mahindra

Senior Manager: Data Science
Aug 2022 - Oct 2023

- Led a team of 10 members to solve business problems and improve customer experience.
- Developed predictive ML models and sales forecasting.
- Worked for Service strategy team on solving multiple business problems

L&T InfoTech

Senior Specialist: Data Science
Jun 2021 - Jul 2022

- Led the data science team for POCs in banking, commodity, and cab aggregator sectors.

RPG Enterprises, Mumbai

Manager: Digital& Innovation
Jul 2018 - Nov 2020

- Led multiple analytics projects across business units like CEAT and Raychem.

ALLSTATE, Bangalore

Lead Consultant/Data Scientist
Aug 2015 - Jan 2018

- Led the data science team to develop ML models for insurance-related projects.

IHS Advanced Analytics, Bangalore

Sr. Information Analyst
May 2013 - Apr 2015

- Worked on automotive sales forecasting and customer loyalty analytics projects.

Genpact Analytics, Bangalore

Business Analyst
July 2011 – May 2013

- Worked on SAS, SQL, R, and **pharmaceutical analytics** Projects, Sales planning and analytics.
  Dashboard preparation for pharmaceutical products performance and process, ad hoc sales.

Datacore India Pvt. Ltd., Kolkata

Quality Analyst
May 2010 - Oct 2010

- Identifying the key clinical information to be recorded in NextGen from the source document. (worked with different modules- Medication, Allergy, Past Medical History, Chronic problems, Lab modules, etc).

Sun Knowledge, Kolkata

Assistant Clinical Analyst
June 2009 - Feb 2010

- Served as the first line resource for Clinical Customer Issues with apt resolution instructions and escalating the same to the physicians.
  Medicare Part D
  Accountable for Claims processing, handling denials & appeals, medication therapy management, step therapy, suggesting alternative medication in the formulary and tracking drug interaction.

Education

Manipal College of Pharmaceutical Sciences
Pharmacy (8.0 GPA)
2009

D.A.V. Hehal, Ranchi, CBSE Board
12th Grade (80%)
2004

Don Bosco Academy, ICSE Board
10th Grade (86%)
2002

Awards and Certifications

- Presented the White Paper on “Application of Latent Class Analysis in Insurance Industry” at IISc Business
- Analytics Conference, Bangalore.
- Got the Best Team Award in Genpact for the year 2011 and recognised as an efficient employee.
- Received the “Star of the Month Award” at ALLSTATE (June 2016)

- Got the Certificate of Excellence award from the client in recognition of achieving exceptional results by demonstrating outstanding behaviors embodied in the tenets of Business Analytics.
- Six Sigma LEAN Trained, Tested and Certified from Genpact
- Six Sigma Green Belt Trained
- Online Computing for Data Analysis using R from Coursera
- MINITAB Quality training from MINITAB
- AWS Machine Learning Speciality course from UDEMY
- Data Mining with WEKA from University of Waikato.
- Free Online Management for A competitive Edge from ICMS
- The Data Scientist's Toolbox from Coursera
- Introduction to Cloud Computing from ALISON
- An Introduction to Programming with Python from ALISON
- Introduction to Time Management from ALISON
- Machine Learning from COURSERA.

- Principles of Project Management from Polytechnique West.
- Exploratory Data Analysis from COURSERA
- Introduction to R Programming from data camp
- Analytics Case Studies from Udemy
- SQL Overview from Udemy
- Guide to diploma in Project Management from Udemy
- Insurance Essentials from The Institues
- Introduction to Analytics and Language of SAS from Udemy
- Data Science and Machine Learning Essentials from edx
- Predictive Modelling Fundamentals from BigdataUniversity
- Editing and Debugging SAS Programs from SAS
- Statistics for Business-ii from IIM- B offered by EDX
- Data Science and Machine Learning Essentials from Microsoft
- Scraping and Data Mining for beginners from Udemy
- Logit Regression in Python from Udemy
- Linear Regression in R from Udemy
- Data Science Project course from Dataquest
- Digital Analytics for Marketing Professionals from Coursera.
- Quality Assurance from HIS
- Statistics Foundation 1 from LinkedIn learning