

# SUSHANT K [REDACTED] L

## PROFESSIONAL SUMMARY

Data Scientist with 3 years of expertise in machine learning, deep learning, model development, and data analysis. Proficient in Python, R, and SQL, with a track record of deploying predictive models to drive business value. Strong communicator and adept at translating technical insights into actionable strategies. Seeking opportunities to leverage skills in a dynamic environment.

## WORK HISTORY

### Data Science Research Assistant, 07/2023 - Current

Indiana University Bloomington, Bloomington, IN

Computer vision:

- Developed an advanced Sign Language Detection model by combining diverse CNN architectures, ensemble methods, and the integration of pre-trained VGG16, achieving an 85% accuracy on the test dataset.
- Implemented OpenCV image processing for 20% image noise reduction through Thresholding and Contour Detection.
- Managed and analyzed 3 diverse datasets, totaling 120,000+ images of hand gestures totaling a size of approximately 4GB.
- Prioritized model explainability, achieving 15% improvement by visualizing Convolution, Pooling, and Filter operations.
- Established a literature review on Sign Language Detection in Explainable AI, synthesizing insights from 15+ papers for research

NLP/LLMs:

- Implemented an LLM-based news research tool for equity research analysts using LangChain and OpenAI
- Leveraged FAISS Index vector database to store the word embeddings from the chunks of text extracted from the news articles
- Utilized Streamlit UI for creating the interface for the users to work with the model with their respective news articles

### Data Analyst, 06/2022 - 09/2022

SBI General Insurance, Mumbai, India

- Designed and developed interactive dashboards using Dash and Plotly to visualize insurance KPIs such as policy renewal rates
- Conducted ABC analysis of 1000+ insurance policies based on premium amounts, claim frequencies, and policy types optimize resource allocation and prioritize customer retention efforts
- Trained a model using RandomForest and XgBoost to predict high-propensity customers who are likely to renew insurance policies thereby achieving 70% accuracy

 Atlanta, GA 30316



## SKILLS

- Analytical mindset
- Process Analysis
- Data Gathering
- Client Requirements Assessment
- Reporting capabilities
- Processes and procedures
- Research and Development
- Project Management
- Data Analytics
- Cloud Computing

## TECHNICAL SKILLS

- **Programming Languages:** Python, R, SQL, HTML, CSS
- **Databases & Version Control:** MySQL, PostgreSQL, SQLite, Neo4j, Oracle SQL, Git, GitHub
- **Libraries & Frameworks:** Numpy, Pandas, Seaborn, Matplotlib, Scikit-learn, TensorFlow, Keras, Plotly, Dash, NLTK
- **Data Science & Machine Learning Techniques:** Regression, Classification, Clustering, Neural Networks-CNN, RNN(LSTM, GRU), Time; Series Forecasting (SARIMA, ARIMA, SARIMAX), NLP/LLM models
- **Statistical Techniques:** Hypothesis Testing, A/B Testing, T-tests, Chi-squared test, Predictive Modelling
- **Tools:** Tableau, Power BI, MS Excel, Looker Studio, Docker, Azure Machine Learning, AWS (Athena, Sagemaker, Glue, EC2, S3)

## CERTIFICATIONS

- Microsoft Certified Azure Data Scientist Associate,
- TensorFlow Developer Certificate

## LANGUAGES

### **Senior System Engineer (Data Science Analyst), 07/2020 - 07/2021**

**Infosys Ltd, Pune, India**

- Developed a propensity model for predicting customer responses across 3 different campaigns, achieving an F1-score of 69%
- Leveraged BigQuery to extract approximately 50GB of customer demographics and past campaign data for model building
- Employed Decision Trees, K-NN, Random Forest, and Gradient Boosting for model development and training
- Achieved a 3-fold increase in the click-rate, reducing the messaging costs by 50%

### **System Engineer, 10/2018 - 06/2020**

**Infosys Ltd, Pune, India**

- Automated the document generation process by implementing XML parser using Java Struts into JSON objects which resulted in a 35% reduction in man-hours
- Enhanced the search and filter functionality by optimizing complex SQL queries for efficient storage and retrieval of data, in turn reducing API response time by 40%

## **EDUCATION**

---

### **Master of Science, Data Science, 05/2023**

**Indiana University Bloomington - Bloomington, IN**

- Relevant Coursework: Applied Machine Learning, Deep Learning, Data Visualization, Introduction to Statistics, Data Analysis
- Capstone Project: Geolocate (Feb 2023 - May 2023). Conducted sentiment analysis on 250k+ tweets related to European football clubs using Twitter API (Tweepy). Orchestrated data preprocessing workflows to extract tweet data by setting up pipelines in Apache Airflow. Created an interactive Looker Studio dashboard to visualize tweet sentiments and identify potential club fan bases globally
- Project: Readify (Apr 2022 - Jul 2022). Led a team of 4 to build a recommendation engine platform for users to explore books, create book lists, and get suggestions. Built a content-based recommendation system using the book title, genre, author, and rating; utilized cosine similarity metric and CountVectorizer to build the recommendation model for 30k+ books. Resolved the cold start problem with the Bray-Curtis distance metric to calculate book similarity based on user's top 3 genres

### **Bachelor of Engineering, Computer Engineering, 05/2018**

**University of Mumbai - Mumbai, India**

English

Full Professional

Hindi

Native or Bilingual