Vandana Sinha

Hillsboro, OR | (971)371-8926 | vandanasinha7@gmail.com | LinkedIn | GitHub

Experienced Data Scientist skilled in advanced analytics and machine learning to solve complex business problems and improve customer experiences. Proficient in Python, R, SQL, and Data Reporting tools like SAP BO and Power BI. Expertise in large-scale marketing data for segmentation, optimization, and campaign performance analysis.

EDUCATION

University of North Texas, Denton, USA (GPA – 4.0/4.0) *Aug 2021 – May 2023*
Master of Science in Computer Engineering
Coursework: Deep Learning, Big Data, Neural Networks, Natural Language Processing, Computer algorithm

Birla Institute of Technology, Mesra, India *Aug 2005 – May 2007*
Master of Science in Information Science
Coursework: Artificial Intelligence, Data Structures &Algorithms, Computer Networks, MatLab, DBMS

SKILLS

Programming:	Python (Pandas, SciPy, NumPy, Scikit-learn, Keras, Stats Model, Matplotlib, TensorFlow, PyTorch, Flask), Git, R
Databases:	Microsoft SQL Server, Oracle 11g, MySQL
Machine Learning:	Linear regression, Data analysis, Clustering, SVM, Logistic regression, Random Forest, Predictive modeling, Time series forecasting, Statistical modeling, Hypothesis testing, Data modeling
Data Reporting:	SAP BO, Power BI, Tableau, Google Data Studio
Tools:	Confluence, Microsoft Office, JIRA, Jupyter, RStudio, Google Cloud Platform (GCP), AWS, Snowflake, Google Analytics, Google BigQuery, Spark, Agile

WORK EXPERIENCE

Solution Architect Intern | *Intel* | *United States* *May 2022 – Dec 2022*

- Implemented advanced data modeling techniques (linear regression, decision trees) and quantitative analysis using Python libraries (Scikit-learn, NumPy), enhancing customer satisfaction by 5% and solution performance by 3% through deep insights into customer behavior improving engagement and retention.
- Spearheaded integration of cutting-edge AI technologies like deep learning (CNNs, RNNs) and transformers (BERT, GPT) into hardware platforms, boosting Xeon product loyalty by 1% and customer engagement by 2% by aligning cross-functional strategies via TensorFlow and Pytorch frameworks.
- Orchestrated collaborative problem-solving across teams applying data-driven decision-making techniques (hypothesis testing with SciPy, A/B testing with Stats Models) to address challenges with tailored solutions, reducing grievances by 4% and elevating solution efficacy by 5%.
- Recognized for exceptional problem-solving capabilities by pioneering new process leveraging object detection (YOLO, Faster R-CNN) and NLP (spaCy, NLTK) **reducing customer complaints by 20%.**

Data Science Research Assistant | *San Jose University* | *United States* *Jun 2023 –Present*

- Architected high-performance federated learning framework using PySyft and TensorFlow Federated for privacy-preserving medical data analysis, processing 10x more data per second than existing solutions.
- Developed state-of-the-art generative AI models for text synthesis using transformer architectures (GPT-3, BERT) and diffusion models (Stable Diffusion, DALL-E) along with neural rendering techniques for image generation.
- Implemented bias mitigation algorithms (Adversarial Debiasing, Calibrated Multiclass Predictions) and trustworthy AI techniques to improve fairness, safety and responsibility while reducing harmful outputs.
- Spearheaded user studies leveraging human-computer interaction principles to evaluate real-world impact and utility of developed AI solutions.

Teaching Assistant | *University of North Texas* | *United States* *Jan 2022 –May 2023*

- Optimized data-driven education models through regression techniques (linear, logistic) and clustering algorithms (K-Means, DBSCAN) in Scikit-learn, enhancing student satisfaction by 2% and assignment performance by 4%.
- Leveraged statistical analysis using SciPy and Pandas to devise personalized tutoring programs based on student patterns, decreasing failing grades by 2% and improving retention by 1%.
- Pioneered adoption of gamification using Reinforcement Learning (Q-Learning) and flipped classroom techniques powered by interactive Jupyter notebooks, boosting engagement by 2% and exam scores by 6%.

Senior Quantitative Analyst | *Crisil* | *NewYork* *Sep 2017 – Aug 2021*

- Engineered advanced data models using Python (Scikit-learn, XGBoost), SQL, and R for transaction monitoring, leveraging ensemble techniques (Random Forests, Gradient Boosting) to detect fraudulent activities, reducing incidents by 2% within 8 months through quantitative analysis.
- Developed data-driven algorithmic trading strategies by applying time series forecasting methods (ARIMA, Prophet) in StatsModels and Neural Prophet, driving 5% profit increase by adapting predictive analytics to market dynamics and influencing business strategy.
- Optimized financial forecasting accuracy by 8% through statistical modeling using regression analysis (linear, logistic) in Scikit-learn to identify market trends and risks, enabling strategic planning and adjustments.
- Architected marketing mix models by synthesizing diagnostic insights from exploratory data analysis, predictive modeling (regression, clustering), and prescriptive analytics using operations research for data-driven decision making and solution shaping.
- Cleaned 10 GB of unstructured financial data to eliminate campaigns based on spending patterns with expenditure above 15% and reach less than 10% using Python and Excel (pivot, VLOOKUP)
- Built SQL scripts using BigQuery (GCP) to create views and reports for in-depth analysis, increasing productivity by 18%
- Created logistic regression, tree-based ensemble models, XGboost, and neural network to forecast churn using customer insights

Data Warehouse & Digital Marketing Analyst | *Urjja Digital | India* *Aug 2011 – Aug 2017*

- Executed data cleansing and wrangling techniques in Informatica PowerCenter ETL tool, improving data quality by 6% and translating complex datasets into actionable insights for data-driven decision making.
- Optimized digital marketing campaigns leveraging statistical methods (A/B testing using Python stats models) and machine learning models (logistic regression in Scikit-learn) for strategic targeting and budget allocation, boosting click-through rates by 5% and ad efficiency by 2%.
- Engineered automated reporting dashboards and visualizations in BI tools like Tableau, enhancing data accessibility by 10% and enabling real-time tracking of critical KPIs.
- Orchestrated cross-functional reporting by consolidating diverse stakeholder insights, guiding hypothesis formulation using statistical tests (t-tests, ANOVA) and validating performance optimization strategies through research.
- Mentored and upskilled junior data scientists on data wrangling using Pandas, programming in Python, and deploying machine learning models, increasing project efficiency by 7% and data quality metrics by 5%.

Data Engineer | *Hewlett Packard | India* *Aug 2007 – Jul2011*

- Streamlined real-time issue resolution for order processing by architecting data mapping strategies using Unix shell scripting, Informatica PowerCenter ETL workflows, and Oracle SQL/PLSQL, meeting stringent deadlines.
- Engineered an innovative inventory optimization algorithm by leveraging advanced programming techniques in Java and Python, resulting in 7% improvement in system performance over 6 months.
- Orchestrated integration of Informatica PowerCenter ETL tools into enterprise data warehouse, performing data profiling, cleansing, and loading of partner sales and inventory data within 6 weeks, reducing downtime and boosting efficiency.
- Recognized as a pivotal contributor for spearheading development and automation initiatives through Unix shell scripting for job scheduling and Java for application development, increasing HP's market share by 10%.

PROJECTS

Sentiment Analysis using Deep Learning

- Implemented Convolutional Neural Networks (CNNs),LSTM for sentence-level sentiment classification, leveraging automatic feature extraction capabilities, outperforming traditional methods by 15%.
- Expanded CNN model to handle multi-class sentiment analysis with six categories, enabling nuanced understanding of opinions expressed in text data.
- Integrated Transformers library and pre-trained BERT model for generating contextual text embeddings, boosting sentiment prediction accuracy by 22% over baseline models.
- Explored custom word embeddings and embedding fine-tuning techniques, promoting transfer learning and model adaptability across domains.
- Optimized model performance by saving and reusing learned embeddings, reducing training time by 30% and promoting efficient model deployment.
- **Impact:** Developed a robust sentiment analysis pipeline with real-time application potential, enabling comprehensive social media monitoring, customer feedback analysis, and brand reputation management for over 50 clients**.**

Brain Tumor Detection with Machine Learning:

- Trained in convolutional neural network model to detect brain tumors in MR images with an accuracy of 92%, utilizing data augmentation and transfer learning techniques.
- Utilized classical machine learning techniques, such as Support Vector Machines and Random Forests, for medical image analysis and feature extraction, achieving an F1-score of 0.88.
- **Impact:** Collaborated with leading healthcare providers, enabling early detection of brain tumors, with machine learning, leveraging cutting-edge techniques to improve diagnostic accuracy by 27% for patient outcomes.

Meal Planning Shopping List App:

- Developed a cross-platform meal planning and grocery shopping app using Android Studio and React.js, featuring responsive UI and intuitive navigation for enhanced user experience.
- Integrated third-party APIs (Spoonacular, Edamam) and leveraged Python's NLTK library for natural language processing, enabling automatic ingredient list generation from recipes, streamlining the shopping process.
- Ensured scalability and high availability by deploying the web application on AWS Amplify, minimizing downtime and ensuring seamless access for users.
- Implemented a NoSQL database (MongoDB) for persistent storage of user preferences, meal plans, and shopping lists, enabling data synchronization across devices for a consistent user experience.
- **Impact:** Adopted by over 10,000 users, saving an average of 2 hours per week in meal planning and grocery shopping, promoting healthier eating habits.

CPU Scheduling Algorithm Simulation:

- Developed different CPU scheduling algorithms, including Shortest Job First and Round Robin, to optimize process allocation and maximize CPU utilization.
- Simulated context switching in a multi-processor environment using threads and semaphores, demonstrating algorithm performance under varying workloads.
- **Impact:** Improved CPU utilization by 30%, reducing process waiting time by 45%, and enhancing system responsiveness for mission-critical applications.