SABNAM PANDIT - Data Scientist

sabnampandit@gmail.com | LinkedIn | GitHub | San Diego, CA

PROFESSIONAL SUMMARY

Results-driven Data Scientist with 2.5+ years of experience in machine learning, statistical modeling, and advanced data analytics across research and industry. Proven expertise in predictive modeling, anomaly detection, NLP, time series forecasting, Large Language Models (LLMs), and geospatial analysis, delivering high-accuracy insights and uncovering critical patterns from large-scale datasets. Adept at collaborating with cross-functional teams and communicating complex findings to diverse audiences.

SKILLS

- **Programming Languages**: Python, R, SQL, NoSQL, JavaScript, C, C++, Scala, Julia
- **Machine Learning & AI**: TensorFlow, PyTorch, Keras, Scikit-learn, XGBoost, NLP (Transformers, BERT), CNN, RNN, LSTM, Anomaly Detection, Reinforcement Learning
- **Data Tools**: NumPy, Pandas, Matplotlib, Seaborn, Tableau, Power BI, Excel, ArcGIS, Streamlit, Jupyter Notebook
- **Databases**: MongoDB, MySQL, PostgreSQL, SQLite, Hadoop, Spark, Cassandra, Redis
- **Cloud & Platforms**: AWS (S3, EC2, SageMaker, Redshift), Azure (Databricks), Google Cloud (BigQuery), Docker, Kubernetes
- **DevOps & Version Control**: Git, Conda, Jenkins, Airflow, Terraform
- **Methodologies**: Statistical Analysis, Data Wrangling, Feature Engineering, Model Evaluation, A/B Testing, Hypothesis Testing
- **Soft Skills**: Analytical Thinking, Research Leadership, Team Collaboration, Communication

PROFESSIONAL EXPERIENCE

Data Scientist (Research Focus) | San Diego State University, San Diego, CA | Jan 2024 – Present

- Developed an advanced analytics algorithm using LLama-8B-Instruct LLM and cosine similarity to analyze 85K cardiovascular research articles, significantly improving similarity and contradiction insight accuracy.
- Designed an unsupervised learning framework with convolutional and transformer models to detect animal/bird audio data anomalies and call patterns.
- Engineered clustering techniques and spectrogram analysis for anomaly detection and pattern recognition.
- Led data-driven research presentations to the San Diego Zoo that led to more funding applications by aligning findings with industry needs, collaborating with Dr. Hajar Homayouni.

Data Science & Programming Graduate Teaching Assistant | San Diego State University, San Diego, CA | Sep 2024 – Dec 2024

- Mentored 65+ students in advanced programming and data science (Python, algorithms, data structures, ML workflows) through weekly office hours, improving coding proficiency and problem-solving skills.
- Graded and delivered detailed feedback on assignments, exams, and projects covering machine learning, statistical analysis, and data-wrangling workflows.
- Assessed and advised on final data-science projects, ensuring methodological rigor, data quality, and reproducibility; proctored exams and managed assessments while upholding academic integrity.

Quality Assurance & Data Engineer | Focus One Payment Solutions (MoCo Digital Wallet), Kathmandu, Nepal | Jul 2021 – Jun 2022

- Spearheaded a data science initiative to analyze transactional logs, reducing payment errors by 15% through ML-based anomaly detection models.
- Built predictive models to identify fraudulent transactions, decreasing fraud incidence by 20% and enhancing trust for 500K+ digital wallet users.
- Collaborated with cross-functional teams and external partners to integrate data-driven payment solutions, boosting retail adoption by 10% at Bhatbhateni Supermarket.

PROJECTS

Forecasting U.S. Unemployment Rates | 2023

- Constructed a time series forecasting model with ARIMA and SARIMA in R, achieving 85% accuracy in predicting unemployment trends from 1947 to 2022, revealing COVID-19 as an outlier with aberrant patterns.
- Engineered and cleaned large historical datasets (50,000+ monthly records) from Kaggle using R and preprocessing techniques like feature engineering, handling missing data, filtering, and exploratory data analysis through plots.
- Evaluated multiple ARMA/ARIMA configurations via AIC/BIC and rolling forecasts, selecting optimal parameters to enhance predictive performance across test horizons.
- Analyzed patterns and seasonality with ACF/PACF plots and built dashboards to share insights with non-technical audiences.

Image Caption Generation with Music Recommendation | 2023

- Engineered an ML pipeline using TensorFlow and CNN+LSTM, generating captions and music recommendations with 90% emotional tone relevance.
- Deployed a Streamlit interface to showcase real-time analytics, increasing user engagement during demo sessions with 50+ participants.

Supply and Demand Analysis of EV Charging Stations | 2023

- Analyzed EV data with Python and Scikit-learn, predicting charging station needs with 88% accuracy using Random Forest and Gradient Boosting models.
- Created ArcGIS maps to identify underserved areas, influencing infrastructure planning for local authorities.
- Developed interactive Tableau dashboards to visualize demand forecasts, station density, and equity gaps.

EDUCATION

Master's in Big Data Analytics | San Diego State University, San Diego, CA | 2023 – 2025 | GPA: 4.0/4.0

Master's in Business Administration - IT | Tribhuvan University, Nepal | 2019 – 2021 | GPA: 3.73/4.0

Bachelor's in Computer Science and Engineering | Visvesvaraya Technological University, India | 2014 – 2018

Certifications: IBM Data Science Specialization | 2023 || Machine Learning A-Z: AI, Python & R | 2025

AWARDS: BDA Academic Excellence Award | 2025 || CAHSI Travel Scholarship to attend GMiS Conference | 2024 || Academic Excellence in MBA-IT | 2020