SABNAM PANDIT

sabnampandit@gmail.com ◇ LinkedIn ◇ GitHub ◇ Website

PROFESSIONAL SUMMARY

Curiosity-fueled Data Scientist with dual Master's degrees (Big Data Analytics & MBA-IT), combining expertise in predictive analytics, statistical modeling, and machine learning with hands-on experience in NLP and LLMs.

EDUCATION

Masters in Big Data Analytics, San Diego State University, USA **GPA: 4.0/4.0** 2023 - 2025
Masters in Business Administration - IT, Tribhuvan University, Nepal **GPA: 3.73/4.0** 2019 - 2021
Bachelors in Computer Science and Engineering, Visvesvaraya Technological University, India 2014 - 2018

PROFESSIONAL EXPERIENCE

Research Data Analyst Jan. 2024 - Present
San Diego State University. *Advisor: Dr. Hajar Homayouni*, *San Diego, CA*

- Analyzed 85K+ PubMed abstracts with topic modeling and LLMs (Llama-8B, Sentence Transformers) to extract themes and quantify alignment/divergence across 15 research topics using cosine similarity analysis, producing visualizations.
- Processed 192+ minutes of wildlife audio with transformer-based embeddings (Animal2Vec), clustering (K-Means, HDBSCAN, UMAP), and spectrogram, histogram (time-of-day) analyses to reveal behavioral patterns in birds/animals audio.
- Built an interactive Streamlit dashboard integrating clustering results, visualizations, and audio playback

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

- Mentored 65+ students in *'Advanced Programming Language'* and *'Principles and Techniques of Data Science'* (Python, algorithms, ML workflows), evaluating assignments, exams, and final project proposals and reports and providing feedback.

Quality Assurance Engineer & Data Analyst July 2021 - Jun 2022
Focus One Payment Solutions (MoCo Digital Wallet) *Kathmandu, Nepal*

- Led QA testing of digital payment platforms through APIs/UI automation improving test coverage and reducing execution time; developed predictive models to identify fraudulent transactions enhancing trust for over 500K digital wallet users.
- Collaborated with cross-functional teams to integrate data-driven payment solutions, streamlining data processes and boosting retail adoption by 10%. Project Coordinator for smooth thrid party installation/integration of payment solutions.

TECHNICAL SKILLS

Data Analytics & Visualization: Python (Pandas, NumPy, Matplotlib, Seaborn), R, SQL, NoSQL, JavaScript, ArcGIS, Tableau, Power BI, Excel, Streamlit, Data Wrangling, A/B Testing, Hypothesis Testing, KPI Tracking, EDA, Feature Eng.
ML & Data Science: NLP, CNN, RNN, Classification, Time Series Analysis, Scikit-learn, Pytorch, Keras, Predictive Modeling
Cloud Platforms: AWS (S3, EC2, SageMaker, Redshift), Azure (Databricks), Google Cloud (BigQuery), Docker, Kubernetes
Databases & ETL: MongoDB, MySQL, PostgreSQL, SQLite, ETL workflows

PROJECTS

Book Genre Prediction using Language Model based on Description

- Web-scraped and cleaned book data from books.toscrape.com and Kaggle, consolidating genres and removing duplicates for better prediction quality. Fine-tuned BERT, RoBERTa, and DistilBERT models (from HuggingFace), and evaluated performance on book descriptions. The models achieved accuracy around 78%.

Image Caption generation with custom music recommendation:

- Trained CNN+LSTM model end to end in Tensorflow in Flickr dataset to generate captions based on image. The captions were used to create engaging, Instagram- friendly text and recommend music using text analytics to match the emotional tone in image. Engineered ML pipeline and deployed an interactive Streamlit-based user interface for local demonstration.

Supply and Demand of the Electric Vehicles (EV) charging stations in relation to EV population

- Predicted EV charging station demand with 88% accuracy using Random Forest and Gradient Boosting models in Python, integrating results with ArcGIS to map underserved areas and support infrastructure planning.Built interactive Tableau dashboards to visualize demand forecasts, station density, and equity gaps.

Forecasting U.S. Unemployment Rates Using Time Series Models

- Built ARIMA and SARIMA forecasting models in R, optimizing parameters with AIC/BIC and rolling forecasts, achieving 85% accuracy in predicting U.S. unemployment trends (1947–2022) and identifying COVID-19 as an outlier.
- Processed and engineered 50K+ monthly records from Kaggle, conducting EDA with ACF/PACF plots and developing dashboards to communicate seasonality and trend insights to non-technical audiences.

AWARDS & CERTIFICATIONS

BDA Excellence Award 2025
CAHSI Travel Scholarship – GMiS Conference 2024
Academic Excellence Award – MBA-IT 2020
IBM Data Science Specialization 2023
Machine Learning A-Z: AI, Python & R 2025