

SUMMARY

Dynamic Data Science professional with over 4 years of experience delivering insights and scalable AI solutions across different domains. Skilled in machine learning, NLP, deep learning, and data engineering, with a strong focus on transforming raw data into strategic business value. Adept at leveraging Artificial Intelligence to solve real-world challenges and drive data-driven decision-making.

WORK EXPERIENCE

Data Scientist | *University of Houston, Houston, USA*

Jun 2024 – Present

- Led large-scale web scraping efforts, extracting and preprocessing thousands of political speeches and articles for advanced text analytics.
- Conducted **sentiment analysis, Latent Semantic Scaling (LSS), emotional analysis, and topic modeling** to uncover insights in political discourse.
- Applied **cosine similarity metrics to evaluate document similarity** across technical content, future outlook, negation detection, readability, and thematic relevance.
- **Built and deployed DistilBERT-based classification models with 80% accuracy**, identifying patterns in political rhetoric and public opinion.
- Delivered key insights through **dynamic visualizations** including word clouds and detailed analytical reports.

Data Science Intern | *University of Houston, Houston, USA*

May 2024 – Jun 2024

- Applied **statistical analysis, time series modeling, and neural networks** to geothermal and climate datasets, enhancing model precision.
- Conducted policy effectiveness studies using **advanced statistical methods and developed visualizations** to communicate results to stakeholders.

Data Scientist | *Bharti Airtel Ltd, Pune, India*

Aug 2022 – Aug 2023

- **Reduced customer churn by 20%** by developing predictive models for churn prediction and segmentation, driving **data-driven strategies**.
- Built executive **Tableau dashboards** to visualize customer behavior and business KPIs, enhancing **stakeholder communication**.
- Collaborated with data engineering teams to design scalable **ETL pipelines using AWS Glue, EMR, and S3**, improving **data processing efficiency by 25%**.
- **Streamlined ELT workflows in Amazon Redshift** and maintained **ML Ops pipelines using Amazon SageMaker**.

Junior Data Scientist | *Tata Consultancy Services, Hyderabad, India*

Sep 2020 – Aug 2022

- Built and optimized **ETL pipelines with PySpark and SQL on Apache Spark**, enabling faster **big data processing** for telecom datasets.
- Performed customer analytics and risk segmentation with Python and R, **applying statistical methods** to support **data-driven decision-making**.
- Created **interactive Power BI dashboards** to track KPIs such as churn rate, data usage, and customer lifetime value.
- Led **A/B testing** initiatives and collaborated in developing **clustering and recommendation models using scikit-learn and Databricks**, increasing customer engagement by personalizing offers and service recommendations.
- Contributed to a **15% improvement in network performance** by delivering actionable data insights.

EDUCATION

University of Houston, Houston, Texas, USA

Master of Science in Engineering Data Science | GPA: 3.96/4.0

Aug 2023 – May 2025

SKILLS

Programming Languages: Python, R, Java, C, HTML, SQL, JavaScript, MATLAB

Data Science & ML: Data Mining, EDA, Feature Engineering, Predictive Modeling, Deep Learning, Natural Language Processing, Statistical Modeling, A/B Testing

Frameworks: TensorFlow, PyTorch, Scikit-learn, Keras, XGBoost, Pandas, Numpy, Matplotlib, Seaborn, Plotly, NLTK, BERT, spaCy, Hugging Face, Gensim, Flask

Databases: MySQL, Microsoft SQL Server, MongoDB, AWS DynamoDB, PostgreSQL, Amazon S3

Visualization Tools: Power BI, Tableau, SAS Visual Analytics, Excel (Pivot Tables, Macros, VBA)

Cloud & Big Data: AWS, GCP, Databricks, PySpark, Kafka, Hadoop, HDFS, Spark

DevOps: Git, Docker

Generative AI: LLMs, Prompt Engineering, Retrieval-Augmented Generation (RAG), LangChain

ACADEMIC PROJECTS

Text Summarization Using BiLSTM, Attention Mechanism, and Pegasus Transformer

Aug 2024 – Dec 2024

- Developed abstractive summarization models for U.S. legislative texts, implementing BiLSTM with GloVe embeddings and Attention Mechanism.
- Fine-tuned Pegasus Transformer to generate high-quality summaries; evaluated models using BERTScore and human assessments.

Café Chatbot Using NLP and PyTorch

Jan 2024 – May 2024

- Developed an interactive café chatbot using NLTK and Flask, deployed on Google Cloud Platform (GCP).
- Designed conversation flows with an intents file and processed user inputs using tokenization, lemmatization, and TF-IDF.
- Trained PyTorch models to handle customer queries and manage café orders, enhancing customer service experiences.

Plant Species Classification Using Leaf Images

Aug 2023 – Dec 2023

- Preprocessed leaf images and built classifiers using CNN models (MobileNetV2, InceptionV3, ResNet50).
- Fine-tuned models using transfer learning and achieved high precision, recall, and F1-scores.

Health Insurance Data Analysis

Aug 2023 – Dec 2023

- Built SAS Visual Analytics dashboards to identify claim anomalies and visualize health insurance trends.
- Leveraged SAS PROC SQL and Enterprise Guide to preprocess and analyze data, producing detailed statistical reports.

CERTIFICATIONS

[AWS Certified Cloud Practitioner](#)