

# AJITESH GOSWAMI

7575 Frankford Road, Dallas, Tx 75252

📞 945-274-8150    ✉ [ajiteshgswami@hotmail.com](mailto:ajiteshgswami@hotmail.com)    [in linkedin.com/in/ajitesh-goswami](https://www.linkedin.com/in/ajitesh-goswami)    [github.com/ajiteshgswami](https://github.com/ajiteshgswami)

## Education

**The University of Texas, Dallas**

*Master of Science in Data Science*

**Aug 2023 – Dec 2024**

*Dallas, Texas*

**Rajasthan Technical University, India**

*Bachelor of Technology in Computer Science*

**Aug 2016 – Sep 2020**

*Rajasthan, India*

## Technical Skills

**Languages:** Python (NumPy, Pandas, Matplotlib, PyArrow, Tensor Flow), PySpark, R, MATLAB, XML, JSON, SQLite

**Analytical:** SQL, Power BI, Excel, Hypothesis Testing, Mongo DB, Regression, Time Series Analysis, A/B testing

**Machine Learning:** KNN, Regression, Neural Network, Clustering, PCA, CNN, RNN, Transformers, GPT

## Relevant Coursework

- Advance Statistics
- Machine Learning
- Deep Learning
- Predictive Analytics
- Prescriptive Analytics
- Web Analytics
- DBMS
- Business Analytics

## Experience

**Principal Global Services,**

*Data Scientist*

**Nov 2020 – July 2023**

*Pune, India*

- Built a Customized Targeting Machine Learning model, optimized for incremental propensity to convert conditioned on chat with an agent, the model delivered 12% growth in conversions.
- Developed a Machine learning model for Trade surveillance to monitor market abuse regulations which avoided a potential \$5 Million fine on the company if found non-compliant in regulator's audits.
- Executed AWS data pipeline: Automated 100% data ingestion from 20+ sources, applied transformations, and organized files for seamless querying resulting in enhanced data accessibility.
- Implemented data pipelines for processing data received from multiple company-owned boutiques and generated reports on AWS Quicksight, resulting in a 180-hour reduction in manual workload.
- Engineered a master data management system to reduce manual work by 70%, reducing the risk of data duplication and error on AWS Aurora DB. Created APIs for performing CRUD operations

**Samyak Innovations**

*Digital Marketing and Web Development Intern*

**Dec 2018 – Jan 2019**

*Mumbai, India*

- Designed and developed 2 portfolio websites and landing pages. Analyzed user engagement on the website through Google Analytics and increased it by 20%, leading to a better customer acquisition rate.
- Created Google Ads search campaigns for the Whiteboard animation product of the company. Increased user traffic on the website by 35% coming in from normal searches.

## Projects

**YouTube Q&A** | *Python, OpenAI API's, Whisper, LangChain, Vector Database, GenAI, NLP*

**Apr 2024**

- Built a robust Retrieval Augmented Generation system for answering questions about any YouTube video. Enabled the system to understand and interpret the content of YouTube videos accurately using OpenAI API's.
- Used Whisper to transcribe YouTube Video. Leveraged OpenAI Embeddings to convert transcribed text into multiple embedded documents and systematically stored them in a Vector Database, facilitating efficient retrieval and analysis.
- Used LangChain to build a system for fetching most relevant documents from the Vector Database and pass those documents as context to OpenAI API for answering questions asked by the user. Also added conversational memory.

**Churn Prediction for Subscription Service** | *Python, LightGBM, Pandas, PySpark*

**Mar 2024**

- Developed Machine Learning Model to Predict Customer Churn for KKBOX Music Streaming Service. Performed Exploratory Data Analysis (EDA) on user subscription data to understand churn patterns.
- Identified key factors influencing churn and provided data-driven recommendations to the KKBOX business team for customer retention strategies for different segments of customers.
- Built a LightGBM machine learning model to predict user churn with an F1-score of 0.92, exceeding industry benchmarks.

**GPT - Shakespeare** | *Python, PyTorch, Pandas, Transformer model*

**Dec 2023**

- Built neural networks from scratch in Python to understand backpropagation, language modeling, and deep learning fundamentals like Batch Normalization and hyperparameter tuning.
- Implemented deep learning concepts in various architectures using PyTorch for efficient tensor manipulation.
- Trained language models to grasp core deep learning algorithms and their application in language processing. Implemented a GPT-like model based on the "Attention is All You Need" paper to generate Shakespeare text