

# VAISHNAVI RAJPUT

(332) 207-9599 - [vrjpt11@gmail.com](mailto:vrjpt11@gmail.com) - [linkedin.com/in/vaishnavirjpt10](https://www.linkedin.com/in/vaishnavirjpt10) - [github.com/vrjpt10](https://github.com/vrjpt10) - [vrjpt10.github.io/](https://vrjpt10.github.io/)

## WORK EXPERIENCE

**Development Data Analysis Intern** - Block Convey, New York Jun 2023 - Present

- Managed a team of 5 interns and collaborated with cross-functional teams to develop a website to implement an advanced machine learning system integrated with blockchain, leading to a remarkable 30% reduction in data breaches
- Optimized and deployed an application with a Deep Learning Forecasting model for brain tumor detection with 93% accuracy
- Led extensive research and analysis on the strategic integration of AI/ML and Blockchain technologies in the market, resulting in actionable insights that drove substantial sales growth

**Graduate Assistant** - NYU, New York Oct 2022 - May 2023

- Coordinated and facilitated doubt-clearing sessions and exams for undergraduate courses - Data Structures and Algorithms and Introduction to Programming, supporting the administration of over 300 students

**Data Science Intern (Remote)** - British Airways, New York Feb 2023 - Mar 2023

- Developed an ETL pipeline using AirFlow and a statistical model which enhanced personalized marketing strategies by 85%
- Employed web scraping to extract 70 GB of review data, and performed sentiment analysis leveraging NLP techniques to identify 4 key areas of concern

**Software Engineer** - YM Grad, India Jun 2020 - Jul 2021

- Created an administration system for service selection, with CRUD operations, and upgraded SEO ranking by 29%

## TECHNICAL PROJECTS

**Recommender System (Python, Apache Spark, Hadoop)** Mar 2023 - May 2023

- Built a music recommendation system on the ListenBrainz dataset by fabricating a popularity baseline model and collaborative filtering techniques; improved efficiency by 10% reduction in errors
- Incorporated light FM and fast search to compare single-machine execution with cluster-based execution achieving a 20.6% reduction in model fitting time and a 30.1% reduction in accuracy calculation time
- Deployed and fine-tuned alternating least square method in Spark to learn latent factor representation of users to increase recommendation precision by 5.7%

**Churn Prediction (Python, Apache Spark, Hadoop, Flask)** Apr 2023 - May 2023

- Performed EDA and feature engineering on telecommunication data to identify 8 factors influencing customer churn
- Leveraged decision tree, random forest classifier, and PCA to construct a prediction model with 94% accuracy
- Designed a Flask application to predict customer churn confidence level and enable real-time access to the churn prediction model

**Mu-GAN: Generating Adversarial Modified Music (Python, TensorFlow, Keras)** Aug 2022 - Dec 2022

- Formulated a DC-GAN extension architecture to generate novel music of a particular genre or modify existing music to a new genre using 7000 songs from GTZAN, FMA, and GACMIS datasets
- Produced adversarial modified genre transferred clips of duration more than 30 seconds that can bypass copyright claims and fine-tuned synthesized music for 7 distinct genres employing Librosa library

**ETL System Development for Data Warehousing (SQL, MySQL, PostgreSQL, Tableau)** Aug 2022 - Dec 2022

- Constructed a web application for an inventory management system and achieved efficiency by reducing the turnaround time by 12% using 3NF Normalized queries and Query Optimization
- Implemented ETL pipeline to automate data extraction, transformation, and loading in the PostgreSQL data warehouse, reducing data processing time by 50% and used Tableau for data visualization

## TECHNICAL SKILLS

**Languages:** Python, SQL, Java, HTML, CSS, C++

**Tools and Technologies:** AWS, Google Cloud Platform, Apache Spark, Hadoop, MapReduce, AirFlow, BigQuery, Python (Matplotlib, Plotly, Seaborn, ggplot, Geoplotlib, Keras, TensorFlow, PyTorch, Sklearn (Scikit-learn), NLTK, Numpy, Pandas, Scipy, OpenCV), Tableau, Microsoft Excel, Git, DataProc, MATLAB, MySQL, MongoDB, PostgreSQL, PL/SQL, Data Warehousing, Data Analytics, Time Series Analysis, Data Modeling, ETL Pipelining, Data Visualization, Database Querying, Recommendation Engines, Customer Segmentation, Big Data Processing

## EDUCATION

**Master of Science in Computer Engineering**

New York University - New York, USA

Sep 2021 - May 2023