

# MOHIT BHOIR

Willimantic, CT 06226 | 959-995-0104 | mohitbhoir789@gmail.com | linkedin | github

## Technical Skills

**Programming Languages:** Python, Java, JavaScript, C++, R, SQL

**Data Science & ML Frameworks:** TensorFlow, PyTorch, Scikit-Learn, Keras, Pandas, NumPy, Matplotlib, SciPy; Machine Learning (Classification, Regression, Clustering, NLP, Time-Series, Data Mining, Statistics)

**Databases & MLOps:** PostgreSQL, MySQL, MongoDB; Docker, Airflow, MLflow, CI/CD Pipelines

**Cloud Platforms & Tools:** AWS, GitHub, Jupyter, Power BI, Tableau, Hadoop, ETL Tools, PyCharm, MCP, AI Agents

## Work Experience

### Community Dreams Foundation

Sep 2025 – Dec 2025

Data Science Intern

Remote

- Worked on end-to-end data science applications involving data collection, preprocessing, modeling, and deployment for social impact initiatives.
- Processed and analyzed large-scale structured and unstructured datasets (**1M+ records**) using Python, Pandas, NumPy, and SQL to derive actionable insights.
- Built and evaluated machine learning models including **Logistic Regression, Random Forest, XGBoost, and Gradient Boosting** for prediction and classification tasks, improving model accuracy by up to **22%**.
- Developed data pipelines for cleaning, feature engineering, and transformation, reducing data inconsistencies by **30%**.
- Applied statistical analysis and hypothesis testing to identify trends, correlations, and key performance drivers across datasets.
- Implemented model evaluation techniques such as cross-validation, ROC-AUC, precision-recall, and confusion matrices to ensure robustness and reliability.
- Collaborated with cross-functional teams to translate business requirements into scalable data science solutions and dashboards.
- Deployed trained models using Python-based workflows and versioned experiments using MLflow and Git for reproducibility.

### Beats by Dre

Jun 2025 – Aug 2025

Data Science & Consumer Insights Extern (Remote)

- Performed exploratory data analysis and sentiment analysis on over 5,000 Amazon reviews using Python, BeautifulSoup, TextBlob, Pandas, and Seaborn to uncover product sentiment, preferences, and brand positioning.
- Segmented customer survey data using Pandas and NumPy to define user personas based on generation, price sensitivity, and feature prioritization.
- Extracted key demand drivers such as bass-forward sound, battery life, and design aesthetics through polarity scores, frequency distributions, and word cloud visualizations.
- Automated the data scraping pipeline using BeautifulSoup and OxyLabs, reducing manual data collection time by 90% and improving dataset scale and quality.
- Translated technical findings into data-backed launch recommendations, including product specifications, pricing, and go-to-market strategy.
- Applied skills in data scraping, text preprocessing, EDA, sentiment analysis, and survey segmentation to support product strategy and consumer insight generation.
- Utilized tools such as Python, Pandas, NumPy, Matplotlib, Seaborn, BeautifulSoup, TextBlob, and WordCloud throughout the analysis workflow.

### Amdocs, India

Jul 2021 – Jul 2024

Software Development Engineer

Pune, India

- Automated regression test suites using Selenium and Robot Framework, reducing release cycles by **30%** and boosting deployment efficiency.
- Engineered scalable ServiceNow workflows that eliminated **40%** of repetitive manual tasks, streamlining request processing time by **25%**.
- Built real-time dashboards in Power BI, enhancing reporting efficiency and enabling data-driven decision-making for key stakeholders.
- Collaborated with **4+** cross-functional teams to design AI-based automation tools, improving accuracy and team productivity by **20%**.
- Integrated regression testing pipelines within development cycles, achieving **95% test accuracy** and accelerating QA feedback loops.
- Developed scripts to test APIs for Order Management System (OMS) and Customer Service Provisioning (CSP), ensuring robust backend integration and improving reliability.

## Projects

### Movie Recommendation Chatbot | Python, RAG, Hugging Face, Pinecone, PostgreSQL, TMDb API

- Developed a semantic-search-based chatbot using a dataset of **343K+ IMDB movies (2000–2024)**.
- Used Hugging Face embeddings + Pinecone vector DB to enable real-time recommendation retrieval with **under 1-second latency**.
- Managed metadata for **200K+ unique movie entries** via PostgreSQL; enriched data using TMDb API to improve content coverage by **30%**.
- Achieved over **92% accuracy** in matching user query intents to relevant movie descriptions through embedding tuning.

### Statistical Analysis of Corporate Takeovers | Python, Scikit-learn, Regression, XGBoost

- Analyzed takeover data from **126 U.S. firms over 8 years** to identify predictors of acquisition likelihood.
- Built classification models (Poisson regression, Random Forest, XGBoost); logistic regression model achieved **AUC = 0.78** and **77% accuracy**.
- Applied scaling and feature selection to reduce model variance by **25%** and improve interpretability.

### Cricket Analysis Dashboard | SQL, Tableau, LOD, KPI Metrics

- Created an interactive Tableau dashboard using **1.2M+ ODI ball-by-ball records (2002–2023)** to analyze player and team performance.
- Performed SQL-based ETL and implemented LOD calculations to generate **20+ KPIs** for match summaries, venue stats, and team trends.
- Improved data cleanliness and consistency by **35%** through custom data wrangling scripts.

## Education

**University of Connecticut, Storrs, CT**  
*Master of Science in Data Science*

**Aug 2024 – Present**  
**GPA:** 3.79/4.0

**Relevant Coursework:** Statistics, Machine Learning, NLP, Deep Learning, Algorithms, Data Mining, Gen-AI.

**University of Mumbai, Mumbai, India**  
*Bachelor of Engineering in Electronics Engineering*

**Aug 2017 – Jul 2021**  
**GPA:** 7.78/10

**Relevant Coursework:** Operating Systems, Python, Cryptography & System Security, DBMS