MOHIT BHOIR

📍 Willimantic, CT 06226 📞 959-995-0104 ✉ mohitbhoir789@gmail.com in linkedin.com/in/mohitbhoir789 ⬛ github.com/mohitbhoir789

Objective

Graduate student in Data Science with hands-on experience in AI/ML, data pipelines, algorithm design, and dashboarding. Known for strong attention to detail and analytical thinking. Passionate about improving data quality and operational efficiency through machine learning, cloud computing, and statistics. Excellent communicator and team collaborator, seeking to leverage my background to support impactful, data-driven decisions.

Education

University of Connecticut, *Storrs, CT*   **Aug 2024 – Present**
Master of Science in Data Science   **GPA:** 3.78/4.0
Relevant Coursework: Statistics, Machine Learning, NLP, Deep Learning, Algorithms, Data Mining, Gen-AI.

University of Mumbai, *Mumbai, India*   **Aug 2017 – Jul 2021**
Bachelor of Engineering in Electronics Engineering   **GPA:** 7.78/10
Relevant Coursework: Operating Systems, Python, Cryptography & System Security, DBMS

Technical Skills

Languages: Python, R, SQL, Java, JavaScript

Libraries/Frameworks: TensorFlow, PyTorch, Scikit-Learn, Keras, Pandas, Matplotlib, SciPy, Flask, BeautifulSoup

Databases: PostgreSQL, MySQL, MongoDB, SQLite3

Tools/Technologies: GitHub, Jupyter, AWS, Power BI, Tableau, Hadoop, PyCharm, Microsoft Excel, ETL Tools, MCP

AI/ML: Classification, Regression, Clustering, Semantic Search, Time-Series Forecasting, Data Mining, Analytics, Statistics

Experience

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 **5+** 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.

Doordarshan, Mumbai   **Dec 2019 – Jan 2020**
Media Analytics and Network Intern   *Mumbai, India*
- Analyzed broadcast and signal data to identify viewership trends across **50+ programs**.
- Maintained electronics and network systems for **3 channels**, ensuring **99.9% uptime**.
- Supported diagnostics that reduced data and signal errors by **15%**.

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.