

Harshit Khatta

harshitkhatta7@gmail.com · +91 62807 33152 · LinkedIn · GitHub: HarshitKhatta

EXPERIENCE

Data Analyst

Dec 2024 - Feb 2025

Ernst and Young (EY)

Noida, India

- Automated data pipelines using **Python** for data cleaning, transformation, and validation, improving data quality and reducing processing time by **30%**
- Developed **Alteryx ETL workflows**, reducing manual data handling time by **40%** and improving weekly reporting efficiency
- Optimized **SQL queries** on datasets exceeding **5M+ records**, accelerating dashboard refresh rates by **25%** and enabling faster insights

Data Scientist Trainee

Jan 2024 - Jun 2024

Sunfocus Solutions Pvt. Ltd.

Mohali, India

- Automated data workflows with **Python**, processing over **500,000 records** using **Pandas** and **NumPy**, improving runtime efficiency by **20%**
- Performed comprehensive **EDA** and built **50+ interactive visualizations**, enhancing internal decision-making by **30%**
- Developed ML pipelines for **news classification** and **sentiment analysis** using **TF-IDF** and **Logistic Regression**, achieving **85%+ accuracy**

PROJECTS

Sleeping Disorder Analysis | *Pandas, NumPy, Scikit-learn, Seaborn, Streamlit* [Link]

Mar 2024 - May 2024

- Analyzed **10,000+ medical records** to identify key predictors of sleep disorders using **ML** and statistical techniques, including **50+ visualizations** demonstrating lifestyle impacts
- Built and tuned **Random Forest** and **SVM models**, improving detection accuracy by **15%** and model generalization by **20%**
- Deployed a lightweight predictive app using **Pickle** and **Streamlit** for early diagnosis

LLM-Powered Resume Evaluator (AI/ML, RAG, Generative AI) | *Streamlit, LangChain, Hugging Face, Ollama* [Link]

- Built an AI-powered resume screening app leveraging **LangChain**, **Hugging Face**, **Ollama**, and **OpenAI** to provide real-time scoring and improvement suggestions
- Integrated **RAG-based Knowledge Systems** to enhance document retrieval and deliver accurate, AI-driven resume insights
- Developed an interactive **Streamlit interface** for accessible, recruiter-friendly resume feedback

SKILLS

Languages: Python (OOps, scripting), SQL (MySQL, PostgreSQL), C/C++

Libraries & Frameworks: Pandas, NumPy, Matplotlib, Seaborn, Plotly, Scikit-learn, TensorFlow, Keras, Streamlit, Selenium, Hugging Face, LangChain

Machine Learning & AI: Linear and Logistic Regression, Decision Trees, Random Forest, SVM, KNN, Naive Bayes, Clustering (K-Means, Hierarchical), PCA, NLP, **Generative AI**, **Natural Language Processing (NLP)**

LLM Tools & Platforms: OpenAI, Ollama, LangChain, Hugging Face, **RAG-based Knowledge Systems**, VS Code, PyCharm, IntelliJ, Alteryx, Tableau, MS Excel (Advanced), TensorBoard, RESTful APIs

Databases: MongoDB, SQLite, Oracle

Concepts: EDA, Feature Engineering, Data Cleaning, Model Evaluation, ETL, LLM Deployment, Prompt Engineering, Data Pipelines, Data Extraction

Version Control & Utilities: Git, GitHub, Pickle, Scikeras

EDUCATION

Punjabi University

Patiala, India

Master of Computer Applications | 8.3 CGPA

Sep 2022 - May 2024

Relevant Coursework: Machine Learning, Data Science, Deep Learning, Natural Language Processing, Database Management

Arya College (Affiliated to P.U.)

Ludhiana, India

Bachelor of Computer Applications | 8.2 CGPA

Aug 2019 - Apr 2022