Kushaagra Verma

+91-6264873679 kushaagrav10@gmail.com /kushaagra-verma /KushaagraV

Education

Vellore Institute of Technology, Bhopal Aug 2022 – Present
Integrated M.Tech in Artificial Intelligence *Bhopal, Madhya Pradesh*

Technical Skills

Languages: Python, SQL, Java
Frameworks: Numpy, Pandas, Scikit-learn, Matplotlib, Seaborn, TensorFlow, NLTK
Developer Tools:: Google Colab, VS Code, Jupyter Notebook, PyCharm, GitHub

Experience

InfoAxon Technologies Oct 2025– Dec 2025
AI Intern (On-site) *Noida, Uttar Pradesh*

- Built a production-grade NLP pipeline processing thousands of unstructured insurance documents into structured datasets using Python and BeautifulSoup.
- Fine-tuned a DistilBERT transformer model for text classification using HuggingFace and PyTorch, performing hyperparameter optimization to achieve 80% validation accuracy and improved generalization across document types.
- Deployed scalable inference APIs to serve real-time and batch predictions, automating insurance document classification and reducing manual processing effort by 60–70%.
- Implemented end-to-end ML lifecycle practices including data preprocessing, model evaluation, modular pipeline design, and production integration, enabling scalable, reproducible, and maintainable ML systems.

Projects

MovieMatch | *Python, NLP, NLTK, Streamlit* Jun 2025

- Developed a content-based movie recommendation system using the TMDB 5000 dataset, enabling personalized movie suggestions based on user input.
- Applied NLP techniques including tokenization and stemming with NLTK to pre-process genres, keywords and cast metadata.
- Implemented cosine similarity on vectorized movie features to identify and rank the most relevant movie recommendations.
- Designed an efficient recommendation function to retrieve top-N similar movies and deployed the application using Streamlit for interactive user experience

Zomato Predictive Analytics & EDA | *Python, Pandas, Scikit-learn, XGBoost* Jan 2025

- Performed extensive Exploratory Data Analysis (EDA) on a large-scale restaurant dataset, identifying key correlations between location, cuisine type, and customer ratings.
- Built predictive regression models using XGBoost and Linear Regression, achieving an R^2 score of 86.4% to forecast restaurant success metrics.
- Mitigated multicollinearity through Feature Selection and statistical analysis, optimizing model complexity for better interpretability.
- Implemented data preprocessing pipelines including missing value treatment, encoding, and scaling, improving model stability and generalization across unseen data.

Forest Fire Prediction | *Python, Pandas, Scikit-learn, XGBoost, Jupyter Notebook* Sep 2024

- Built a model to predict burned area from meteorological and temporal features using the UCI Forest Fires dataset, achieving $R^2 = 0.80$ on held-out test data.
- Benchmarked XGBoost, Random Forest, LightGBM and Linear Regression; selected XGBoost after RandomizedSearchCV tuning, reducing test RMSE by 22% versus baseline.
- Engineered temporal features, applied log-transform, validated with time-aware CV, and generated SHAP explanations.
- Established a validated ML pipeline with cross-validation and error analysis, improving model stability and reliability across temporal splits.

Achievements & Extracurricular Activities

- Leadership: Led PR and sponsorship initiatives for a student club, organizing multiple successful events and driving consistent sponsor engagement.
- Completed 50 days of Code under G.Viswanathan challenge.
- Completed *Applied Machine Learning in Python* (University of Michigan, Coursera).
- Completed *Cloud Computing* (IIT Kharagpur, NPTEL).