Sumit Singh

Rudrapur, Udham Singh Nagar, Uttarakhand, India
8439684589 sumitsin712@gmail.com sumit-singh sumit-singh-99

Education

Graphic Era Hill University **September 2022 – July 2026**
B.tech Computer Science Engineering *Dehradun, Uttarakhand*
Grade: 7.1/10

St. Mary's Senior Secondary School **April 2019 – May 2020**
Class XII *Rudrapur, Uttarakhand*
Grade: 72.5/100

St. Mary's Senior Secondary School **April 2019 – May 2020**
Class X *Rudrapur, Uttarakhand*
Grade: 90.2/100

Technical Skills

Programming Languages: Python, C++, Java, SQL

Data Analytics Tools: MS Excel, Tableau, Jupyter Notebook

Machine Learning: Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn

Developer Tools: VS Code, PyCharm, MySQL Workbench

Relevant Coursework: Machine Learning, Data Science, Database Management

Projects

EduTube Recommender | *Python, BeautifulSoup, YouTube API, Pandas, Scikit-learn, Flask* **Feb 2025 - Present**

- Developing an AI-driven YouTube playlist recommendation system that suggests top educational playlists based on user-input subjects and courses.
- Extracting video metadata and user comments using YouTube API and BeautifulSoup for data collection.
- Performing sentiment analysis (VADER/BERT) on user comments and ranking playlists using a machine learning model (Random Forest, Decision Tree) based on engagement and sentiment metrics.
- Deploying a Flask-based web app with a user-friendly search interface for real-time playlist recommendations.

Movie Recommender System | *Python, NumPy, Pandas, Scikit-learn, Streamlit* **15 Nov – 15 Dec 2024**

- Developed a content-based movie recommender using datasets on movie details and credits with (4,803 rows).
- Using genres, keywords, and cast details, metadata were leveraged in computing similarity scores with the cosine similarity.
- Built a Streamlit app that gives movie recommendations personalized according to user preferences.

Spam Detection Classifier | *Python, Pandas, Scikit-learn, Streamlit* **1 Nov – 10 Nov 2024**

- Developed a spam classifier with an accuracy of 97% and precision of 100% using Python and Scikit-learn.
- Preprocessed over 5,000 email entries with the application of TF-IDF and Count Vectorization for text cleaning.
- Trained and tested models, Naive Bayes and Decision Tree, with an 80-20 train-test split.
- Streamlit is deployed with this solution to provide real-time email classification predictions.

Achievements and Leaderships

- Secured 3rd rank in a hackathon organized by Graphic Era University as a member of a 4-member team, where we developed a search engine.
- Secured first position in the school-level Math Olympiad in 9th grade and qualified for the zonal level competition.

Professional Summary

Aspiring data analyst with strong skills in Python, data visualization, and machine learning. Passionate about solving real-world problems using data and statistical analysis. Seeking opportunities to contribute to impactful projects through innovative solutions.