

Saanvi Grover

+91 7303552946 | Email | GitHub | LinkedIn | Portfolio

Skills

- Programming:** Python, SQL
- Machine Learning:** Regression, Classification, Clustering (K-Means), Time Series Forecasting, XGBoost, Feature Engineering, Model Evaluation, Hyperparameter Tuning
- Generative AI / LLM:** Retrieval-Augmented Generation (RAG), Natural Language→SQL, LLM Integration, Prompt Engineering
- Data Analytics:** Data Cleaning, Data Wrangling, EDA, Statistical Analysis, Business Analytics
- Databases, Backend & Deployment:** PostgreSQL, PostGIS, FastAPI, Flask, Streamlit
- Visualization:** Power BI, Tableau, Excel, Matplotlib, Plotly
- Libraries:** Pandas, NumPy, Scikit-learn **Tools:** Git, GitHub, Jupyter Notebook, Google Colab

Experience

Persistent Systems

May 2026 – Jun 2026

Mentorship Program – Data Analytics & Machine Learning Project

GitHub | Live Demo

- Built **SkillLens / HSEP**, a workforce intelligence and skill forecasting platform during a 2-month mentorship at Persistent Systems.
- Analyzed 3M+ observations (120,000+ job postings, 213,000+ job-skill relationships, 40,000+ salary records, 2.7M+ skill mentions).
- Expanded the skill universe from 35 to 114 curated skills and reconstructed 1,209 skill-year observations (2013–2025).
- Applied K-Means clustering and hybrid forecasting to classify skills into Future-Proof, Growing, Stable, and High-Risk categories.

Projects

HeatFluxAI: Urban Heat Island Estimation System

Apr 2026 – May 2026

GitHub | Live Demo

- Developed an ML system to estimate urban heat island intensity across 25 global cities from environmental data.
- Engineered 27 environmental features from ERA5 weather data and MODIS satellite imagery across 4,500+ observations.
- Evaluated eight regression models and achieved the best performance with XGBoost, reaching an RMSE of approximately 2.14°C.

HR TalentScope: Employee Attrition & Workforce Analytics

Mar 2026 – Apr 2026

GitHub | Live Demo

- Developed an HR analytics platform using 19,000+ employee records to predict employee attrition risk and workforce trends.
- Performed feature engineering and EDA across 16 workforce variables, compared five ML models, and built an analytics dashboard with 10+ workforce insight modules.

FloatChat: Natural-Language Querying of Ocean Data (RAG + NL→SQL)

Aug 2026

Finalist – ANALYTICA 2026 National-Level Data Science Hackathon (Team Entropy)

- Built the AI backend for a system answering plain-language questions over 12,442 real ARGO ocean-float profiles (7.9M measurements) in PostgreSQL + PostGIS, across 5 languages.
- Implemented a RAG-over-schema NL→SQL pipeline (LLM generation with an offline rule-engine fallback) and grounded every answer against the query result set, exposing the SQL and source float IDs for auditability.
- Engineered a query safety layer (sqlglot parsing, single-SELECT enforcement, table allow-list, read-only role, statement timeout); 36/36 safety and accuracy tests passing.

Certifications & Achievements

- Certifications:** Google Advanced Data Analytics, Google AI Essentials; HackerRank SQL Basic/Intermediate/Advanced; HackerRank Python Basic
- Achievements:** Finalist, ANALYTICA 2026 National-Level Data Science Hackathon (VESASC, Mumbai); Finalist, Statistella BASH 8.0 Data Analytics Competition (IIT-BHU); Rank 1, Tableau Dashboard Design Round 1

Education

Vishlesan i-Hub, IIT Patna / Masai

Feb 2026 – Aug 2026

Certification Program in Artificial Intelligence & Machine Learning (Hindi)
3 Certificates of Excellence: Eval 1 · Eval 2 · Eval 3

CGPA: 9.16

Jaypee Institute of Information Technology (JIIT)

2024 – 2029

Integrated M.Tech in Computer Science & Engineering

CGPA: 8.3

Delhi Public School, Ghaziabad Meerut Road 10+2 PCM

2024 92.2%