Hrishik Guha

Data Scientist

Contact
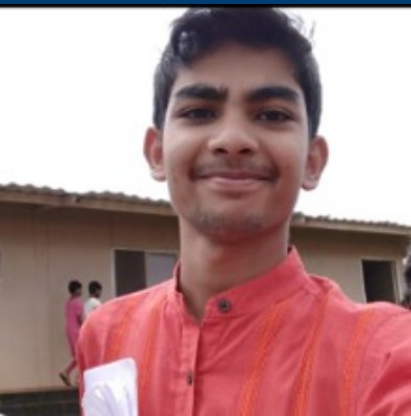
Address
New Delhi India

Phone
9871663902

E-mail
hrishikguha@gmail.com

Portfolios, Profiles

- hrishikg.github.io
- https://
  www.linkedin.com/in/
  hrishik-guha-695280371

Skills

- Programming and query: Python, Stata, SQL, MongoDB

- Data preprocessing and cleaning: NumPy, Pandas, Scikit-Learn etc.

- ML & AI: TensorFlow, PyTorch, Scikit-learn (supervised and unsupervised learning algorithms) , XGBoost, time-series forecasting

- Data visualization: Matplotlib, Seaborn, Plotly express, Cufflinks, Geoplot, GeoPandas

- Statistical analysis: SciPy, Statsmodels

- ETL: Apache Airflow, Talend

- Native fluency in 3 languages (English, Hindi and Bengali). Spoken and written proficiency in French

- Trained athlete (extremely passionate about track and field)

- Trained guitarist (Western Classical and Hindustani Classical)

I am a data scientist, and a graduate in economics, building end-to-end machine learning solutions, across industries (finance, telecom, energy etc.), that survive outside notebooks. My portfolio demonstrates technical rigor, reproducibility, and clear business impact. With a strong foundation in data cleaning, machine learning pipelines, and statistical testing, using Python, I'm eager to learn from experienced professionals and contribute to a dynamic team.

Education

2025-06 — **Certificate: Applied Data Science**

WorldQuant University
What I've learnt:
- Cleaned and preprocessed data
- Built machine learning pipelines- linear regression models, logistic regression models, decision tree models, random forest classifiers and predictors, K-means clusterings models, principal component analysis,
- Auto-regression models for time-series forecasting
- Predictive models to optimize marketing campaigns
- Analyzed customer behavior to identify opportunities for revenue growth
- ETL pipelines by coding custom python classes
- Stock volatility forecasting using a GARCH model and deployed the model using FastAPI
- Chi-square tests and AB testing
- Created interactive dashboards, visualizations for communication and much more

2022-05 — **Bachelor Of Arts: Economics**

O.P. Jindal Global University - Sonipat
Important Courses:Principles of Economics, Econometrics(I and II),Applied Econometrics, Macroeconomics (I and II), Microeconomics (I and II), Political Economy (Capitalism and Marxism), Game Theory (I and II), Development Economics, Public Finance, Public Economics, International Trade, Statistics (I and II), Applied Economic Analysis, Global Political Economy of the 21st Century, History and Evolution of Economics

2019-04 — **All India Senior School Certificate Examination**

The Mother's International School - New Delhi

Work History

2025-08 - Current — **Freelance Data Scientist**

Neev Energy, New Delhi
- Operational analytics (using python) on current & voltage imbalance, neutral current ratio, and phase load share to flag asset risk, quantify losses, and prioritize maintenance. Directly leading to cost-saving efforts.
- Energy consumption forecasting for LED street-lighting and time-series modeling with engineering diagnostics
- Framing deliverables as action boards for operations + finance (what to fix, where, and estimated impact).
- Cleaned and drew insights from 3000000 rows of IoT data

2025-02 - 2025-05 — **Media Analyst**

Kanalytics , New Delhi
- Analyzed and created daily reports on current events and international news coverage for The Ministry of External Affairs as the client.
- Analyzed media presence for various clients.
- Tracked and analyzed India's image in neighbouring countries
- Tracked and analyzed coverage and overall image in international publications (eg.French publications) during international visits.
- Employed monitoring tools like TalkWalker and Solr effectively to oversee client image.
- Utilized social listening tools effectively to monitor online conversations about brands and their competitors

Projects portfolio

- Financial news sentiment → price-movement modeling: Article-level pipelines (NewsAPI ingestion, FinBERT scoring, leakage-safe temporal splits). Data collection, sentiment scoring, EDA and price direction modeling.
- Credit-card fraud detection (Kaggle, European cardholders): Exploratory analysis, deep-learning pipeline with class-imbalance handling; AUC ≈ 0.98, test accuracy ≈ 0.98; clear thresholding
- DCRNN to predict future traffic patterns based on historical speed data from the METR-LA dataset. The model combines spatial dependencies (via diffusion convolution over road networks) and temporal dynamics (via GRU-style recurrent units) to forecast traffic congestion.
- Customer churn prediction: Feature engineering + tree-based models; interpretable outputs (feature importance, LIME) to guide retention actions.
- Explored the relationship between fundamental financial indicators and 1-year forward stock returns. Companies are clustered based on financial fundamentals(e.g., ROE, liabilities, liquidity, leverage), using unsupervised learning techniques and the resulting groups are analyzed for return characteristics and business profiles.