

HARSH GUPTA

hgpt999@gmail.com | +91 987 065 0985
<https://www.linkedin.com/in/harshgupta007/>
<https://github.com/harshgupta-007>

EDUCATION

NIT HAMIRPUR

MSc MATHEMATICS & COMPUTING

Grad. June 2022 | Hamirpur,

Himachal Pradesh

CGPA: 9.21 / 10.00

UNIVERSITY OF DELHI

BSc MATHEMATICS HONOURS

Grad. June 2020 | Delhi,

CGPA: 8.31 / 10.00

SKILLS

PROGRAMMING LANGUAGES

Proficient

Python

Familiar

• C/C++ • MATLAB

DATABASE TECHNOLOGIES

Proficient

• SQL Server • MongoDB Compass

ML LIBRARIES

Proficient

• SciKit-Learn • Pandas • Plotly

• NumPy • Selenium • Pycaret

VISUALIZATION SOFTWARE

Proficient

• Microsoft PowerBI

Familiar

• DOMO • Tableau

POR

CORE TEAM

MEMBER | DU 2018, 2019.

Coordinated with a team of 15

Members as a Bursar for the Conducting Mathematical Fest having footfall of 450+ students.

CERTIFICATIONS

- Power BI for Business Intelligence (Udemy)
- Google Data Analytics (Coursera)
- SQL for Data Science (Coursera)
- MongoDB Basics (Coursera)

EXTRA CURRICULAR

ACTIVITIES

- Participated in KV National Cricket Tournament (School)
- Secured **2nd position** in Inter Year Cricket Tournament (MSc)

WORK EXPERIENCE

ANALYST MERCADOS ENERGY MARKET INDIA | JUN'22- PRESENT

- Collected, cleaned, and processed large datasets from multiple energy sector websites to extract valuable insights.
- Responsible for the daily data insertion into the database and dashboard.
- Handled the database and fulfilled client data requirements.
- Leveraged **parallel processing** techniques, resulting in a **70% reduction** in process execution time.
- **Scheduled** and update multiple cron jobs on server infrastructure to automate data collection processes.
- Developed **robust web scraping** code to extract data from multiple energy sector websites using Python and web scraping libraries.

DATA SCIENCE & BUSINESS ANALYSIS INTERN | JUN'21- JUL'21

- Employed Linear Regression to forecast student academic scores based on study hours with **94.2% accuracy** and **4.87% MAE**.
- Visualized the relationship between study hours and academic scores using the Best Fit Line.
- Employed the **Decision Tree Algorithm** to predict the type of iris flower using the features provided in the iris dataset.

PROJECTS

ELECTRICITY LOAD FORECASTING | CLIENT | JUL'22- DEC'22

- Removed dirty value and inputted **8 years** demand data using statistical methods. Identified temperature, rainfall, and wind gust as key influencers of demand.
- Conducted data analysis to identify **seasonal trends** and **anomalies** in historical electricity demand patterns.
- Applied multiple regression techniques and identified **gradient boosting** by hyperparameter tuning have least MAPE percentage.
- Post process forecast result in order to make custom ensemble model using **random forest regressor** which reduced overall test MAPE by **approx. 3 %**.

AUTOMATED METER READING USING YOLO | JAN'22- MAY'22

- Constructed a cascading model architecture to achieve accurate object detection using private dataset of **7000+ meter images**.
- Made model to automatically crop meter area and save it in a folder.
- Utilized cropped meter images to annotate **10 digits** and trained digit detection using **YOLOv5**.
- Achieved **99.7% accuracy** in number area identification and **87.8% accuracy** in digit detection.

OPINION MINING ON AIRLINE TWEETS | OCT'21- NOV'21

- **Preprocessed** data by removing stop words, special characters, hashtags, and tokenizing the text using text cleaning libraries for further analysis.
- Conducted sentiment analysis on **14000** Virgin Airline tweets and identified that **60%** of the tweets were negative.
- Resampled data for balance positive and negative tweets for model training.
- Applied multiple classification algorithm to fit train and found out that logistics regression performs better with **93% accuracy**.