# Harikesh Kushwaha

Location: New Delhi, Delhi

LinkedIn | Portfolio | GitHub | Kaggle    Email: harikeshkumar0926@gmail.com | Mobile: +919838422934

## Data Scientist

As a recent graduate with a strong foundation in **statistics** and data science, I have worked on several personal projects with real datasets using SQL and Python. In my projects, I have showcased my skills in **data cleaning**, **data visualizations**, and **modeling**. With a passion for solving complex problems and a drive to constantly learn and improve, I am excited to take on new challenges in the field of Data Science.

## Technical Skills

| | | |
|---|---|---|
| **Languages** | : | Python (Proficient), SQL, C++ |
| **Frameworks** | : | Scikit-learn, TensorFlow, Keras, Django, Streamlit |
| **Libraries** | : | matplotlib, plotly, pandas, NumPy, Seaborn, BeautifulSoup, Selenium, Statsmodels |
| **Databases** | : | MySQL, MongoDB |
| **Dev Tools** | : | VS Code, Tableau, Cognos, Git, GitHub, Jupyter Notebook, Anaconda, AWS, S3 |
| **Soft Skills** | : | Analytical and Problem-Solving Skills, Good Communication Skills, Statistics |

## Education

**Indian Institute of Technology Delhi**                    New Delhi, India
*Master of Science in Physics, (8.6 GPA)*          *July 2021 – May 2023 (Expected)*

**Banaras Hindu University**                    Varanasi, Uttar Pradesh India
*Bachelor of Science in Physics, (8.4 GPA)*                    *July 2018 – May 2021*

## Projects

**House Prices Prediction**          *Python, pandas, scikit-learn, kaggle, Matplotlib, Seaborn*          Source Code

- Analyzed over **80** features to predict house prices using machine learning.
- Performed **Exploratory Data Analysis** and **feature engineering** to get insight from data.
- Trained **multiple models** using scikit-learn and selected the best one by applying **grid search** and **cross-validation**. Used ensemble of the top performing models to achieve a **top 10%** ranking on Kaggle.

**Credit Risk Assesment**          *Python, pandas, scikit-learn, kaggle*          Source Code

- Developed a credit textbfrisk assessment model by analyzing various customer features, performing data cleaning, feature engineering, and exploratory data analysis.
- Established a basic model for initial experimentation, and trained advanced models such as **LR**, **SVM**, **XGBoost**, **Catboost**. Top performing model, achieved a test AUC-ROC score of **0.97** and precision of **0.96**.

**IBM Data Analytics Capstone Project**          *Python, pandas, Matplotlib, Web Scraping, API, Dashboard*

- Gathered and analyzed data from various sources, including **API** and **web scraping**. Conducted **exploratory data analysis** and **wrangling** to prepare the data for further analysis.
- Built a **dynamic dashboard** to extract valuable insights from the collected data, and effectively **communicated** the findings to others through an **engaging presentation**.

**Tableau Dashboards**          *Tableau, Web Scraping, Web API, BeautifulSoup*          Music Books

- Created an interactive **Tableau viz** showcasing my **Spotify streaming history** over several years, using **data blending** and **calculated fields** to present key insights.
- Utilized **web scraping** techniques to extract my book reading history from **Goodreads** and created an interactive **Tableau** dashboard to analyze and visualize the data.

## Certifications

- Deep Learning Specialization (DeepLearning.AI) Certificate
- Machine Learning Specialization (DeepLearning.AI) Certificate
- IBM Data Analyst Capstone Project (IBM) Certificate