

# HARSHIT GAUR

gaurharshit04@gmail.com | +16173960129 | **Portfolio:** <https://harshitgaur6155.github.io/>  
[www.linkedin.com/in/harshitgaur/](https://www.linkedin.com/in/harshitgaur/) | [public.tableau.com/app/profile/harshitgaur](https://public.tableau.com/app/profile/harshitgaur) | [www.github.com/harshitgaur6155](https://www.github.com/harshitgaur6155)

## EDUCATION

**Master of Science in Computer Science, Machine Learning and Data Analytics** September 2021 - March 2023  
**Northeastern University, Boston** **GPA: 4.0**

- Coursework: Intermediate Analytics, Big Data, Machine Learning, Statistics

**Bachelor of Technology, Information Technology** August 2013 - July 2017  
**Guru Gobind Singh Indraprastha University** **GPA: 3.71**

- Awards: *magna cum laude*; Silver Medalist in Single's & Double's Badminton Cup

## WORK EXPERIENCE

**Mobikasa** December 2018 - August 2021

### Data Scientist (Software Engineer)

- Utilized **Python**, **Azure Cognitive Service** to implement unsupervised and deep learning technique on 4TB unstructured data to achieve Speech-to-Text Translation of debrief-timeout audio sessions of surgeries and operations with 67% accuracy.
- Wrangled 11TB user data in **Hadoop** distributed file system to segment customers' shopping behaviors using machine learning.
- Designed and created database using Structured Query Language (SQL) and utilized Python, PHP, Java, Shopify to develop microservice applications: **AnytownUSA**, **1-800-Flowers.com**, **DigiNurse (BSSNY Hospital)**, driving \$3.45 billion annual revenue.
- Achieved accurate prediction of 30% increased sales and better customer traction on website using analysis performed on Tableau and machine learning libraries (Pandas, Seaborn, Scikit-Learn).
- Enriched data availability, security, and scalability with **GCP** and **Amazon S3** storages in applications serving 2 million consumers.
- **Achievements:** Awarded 'Top Performer' for delivering innovative solutions for clients.

**Appster** July 2017 - December 2018

### Software Engineer

- Created RESTful APIs in Python, PHP, Microsoft Azure with fetching data from MySQL using SQL.
- Analyzed user stories using JIRA, Excel and enhanced project roadmap by 7% reduced deadline.
- Developed mobile-applications (MyHero, Bazar, COEY) catering clients' business requirements with 350,000 plus downloads version controlled using GIT.
- Incorporated cryptography for 2-way security in applications, and payment gateways for online payment processing.
- **Achievements:** Received 'Going the Extra Mile' award for exceptional contribution and 'Best Performer of the Month' twice.

## PROJECTS

### Airbnb Price Estimator (Recommendation System)

- Created dynamic Airbnb price estimator, deployed on Streamlit, using **Predictive Analytics** taking in user input to calculate price.
- Predicted 80% accurate house pricing using **Random Forest**, **XGBoost** algorithms, after dimensionality reduction using **PCA**.

### H&M Retail - Customers' Next Purchase Predictor

- Built classification model using **Ensemble Methods** to predict customers likely to purchase in next 90 days from last purchases.
- Identified different buyer groups using **Clustering Method** for H&M marketing teams to focus on offers to increase revenue.

### Prediction of readmittance of diabetic patients using analysis of HbA1c value

- Analyzed that 32.7% diabetic patients got readmission when HbA1c value is above 7 at time of primary diagnosis.
- Found SVM prediction of readmissions at 86% accuracy to be better than **k-NN**, **Random Forest Classification**, **Decision Tree**.

### New York Tree Census 2015 Analysis Dashboard using Tableau

- Reported causal factors of tree problems on **Tableau**, examined by tree survival rates and health attributes of dominant 5 species.
- Detected root and trunk problems being major in trees, measured by analyzing 3 species covering 34% plantation across city.

## SKILLS

- **Programming Languages:** Python, SQL, R, PHP, Java, NoSQL, Hive, Impala
- **Databases:** MySQL, MongoDB, HBase, Databricks, MSSQL
- **Machine Learning & Statistics:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, XGBoost, SciPy, AutoML, Clustering (K-Means), Gradient Boosting, Deep Learning (PyTorch), Hypothesis Testing, Regression Analysis, A/B Testing, Model Evaluation, NLP
- **Data Visualization Tools:** Tableau, Microsoft Power BI, Microsoft Excel
- **Cloud Technologies & Services:** Microsoft Azure, Amazon Web Services (AWS), GCP, Spark, Hadoop, GIT, JIRA, Docker