

# HARSHIT GAUR

gaurharshit04@gmail.com | 617-396-0129 | Greenville, SC, USA

Portfolio: [harshitgaur6155.github.io](https://github.com/harshitgaur6155) | [linkedin.com/in/harshitgaur-](https://linkedin.com/in/harshitgaur-) | [github.com/harshitgaur6155](https://github.com/harshitgaur6155)

## WORK EXPERIENCE

### Greatpoint HR Inc.

Boston, MA, USA

#### Data Scientist / Lead Data Analyst

May 2023 - Present

- Engineered **NLP**-driven automated job categorization with **Large Language Models** using **BERT** and **Sentence Transformers**, achieving 88% accuracy and 40% reduction in manual HR processing workload.
- Architected and managed scalable **ETL** pipeline using **Snowflake**, **SQL**, and **Python**, processing **2.3+ billion** records with **99.9%** data integrity and optimizing HR analytics workflows.
- Applied predictive modeling & causal inference to analyze compensation trends, increasing candidate acceptance rates by 25% and reducing recruitment acquisition costs through data-driven insights.
- Led and collaborated with cross-functional teams to develop **AI-powered Recommendation System** for salary structures & grades generation, resulting in 20% growth in hiring KPIs.
- Designed interactive **Tableau** dashboards presenting recruitment metrics, enhancing stakeholder communication and boosting hiring manager satisfaction scores by 20%.
- Performed hypothesis-testing driven **root cause analysis (RCA)** to uncover anomalies in financial compensation data, providing actionable insights that improved talent acquisition and business outcomes by 15%.

### Mobikasa

New Delhi, India

#### Data Scientist (Software Engineer)

December 2018 - August 2021

- Employed **Azure Cognitive Service** to implement **supervised learning** techniques on 4TB healthcare unstructured data, attaining 87% accuracy in Speech-to-Text Translation of complex surgical debrief and timeout audio data.
- Analyzed e-commerce customer behaviors with **Power BI** and **clustering** techniques for customer segmentation, uncovering key insights that boosted targeted marketing strategies and transformed ROI by 35%.
- Led and mentored team of 4 through collaboration, delivering solutions for business objectives, optimizing operational efficiency.
- Designed and orchestrated **SQL** databases and leveraged **Python**, **PHP**, and **Shopify** to develop applications: [1-800-Flowers.com](https://1-800-Flowers.com), *AnytownUSA*, *DigiNurse (BSSNY Hospital)*, driving 40% growth in annual business revenue.
- Drove 30% improved website traffic and conversions with focused strategic analytics, increasing sales and customer engagement.

### Appster

Gurugram, India

#### Software Engineer

July 2017 - December 2018

- Created RESTful APIs using Python, PHP, Java, MySQL and version control (**GIT**), ensuring seamless integration in applications.
- Developed mobile applications achieving 350K downloads, implementing transaction payment gateways & cryptographic security.
- Optimized agile project management with **JIRA** & **Excel**, reducing deadlines by 7% via data-driven planning & problem-solving.
- Performed data analysis on app performance metrics, identifying trends and optimizing user engagement strategies, leading to a 15% increase in retention.

## EDUCATION

### Northeastern University, Boston

GPA: 4.0

#### Master of Science in Computer Science, Machine Intelligence and Data Analytics

September 2021 - April 2023

### Guru Gobind Singh Indraprastha University

GPA: 3.71

#### Bachelor of Technology, Information Technology

August 2013 - July 2017

## PERSONAL PROJECTS

### Airbnb Price Estimator: Recommendation System ([click to view](#))

- Built **Predictive Model** using **Random Forest**, **XGBoost** & **PCA**, deployed on Streamlit, with 80% accurate price estimation.

## SKILLS

- Programming & Databases:** Python, SQL, Snowflake, R, PHP, MySQL, PostgreSQL, NoSQL (MongoDB)
- Data Science Libraries:** Pandas, NumPy, Scikit-learn, SciPy, PyTorch, Matplotlib, Seaborn, AutoML
- Machine Learning & AI:** Regression, Classification, Clustering, Regularization, Dimensionality Reduction (PCA), Large Language Models (LLMs), Natural Language Processing (NLP), Feature Engineering
- Business Intelligence Tools:** Tableau, Microsoft Power BI, Microsoft Excel
- Statistical Modeling:** Exploratory Data Analysis (EDA), Root Cause Analysis, Hypothesis Testing, Causal Inference, A/B Testing, ETL (Extract, Transform, Load)
- Cloud Platforms & Big Data:** Microsoft Azure, Amazon Web Services (AWS), GCP, Spark, Hadoop, Agile (JIRA)