

# HARSHIT GAUR

gaurharshit04@gmail.com | 617-396-0129 | Boston, MA, 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

### Data Scientist

May 2023 - November 2024

Greatpoint HR Inc.

Boston, MA, USA

- Engineered and launched NLP-driven automated record categorization with fine-tuned Large Language Models (LLMs) using BERT and Sentence Transformers saving 15 hours/week of manual work, which contributed to a reduction in operational costs by \$50K annually.
- Architected and managed scalable ETL pipeline with Snowflake, SQL, and Python, processing 2.3+ billion records with 99.9% data integrity and achieving 80% reduction in data processing time.
- Led and collaborated with cross-functional teams to develop AI-powered Recommendation System for salary structures & grades generation, improving talent acquisition KPIs (time-to-hire, offer acceptance rate) by 20%.
- Developed Generative AI (GenAI) - GPT model to auto-generate customized compensation packages using market data, profile, and role specifications, reducing offer generation time by 40%.
- Applied predictive modeling (XGBoost) & causal inference (Propensity Score) to analyze client compensation trends, reducing recruitment acquisition costs by 18%.
- Automated model training, tracking, and deployment using GitHub Actions, MLflow, and Kubeflow, enabling streamlined CI/CD pipelines.
- Performed hypothesis testing driven root cause analysis (RCA) to uncover anomalies in financial data, providing data-driven actionable insights via interactive Tableau dashboards to Co-Founder and CTO.

### Data Scientist (Software Engineer)

December 2018 - August 2021

Mobikasa

New Delhi, India

- Achieved Speech-to-Text Translation of complex surgical debrief and timeout audio data using supervised learning techniques with Microsoft Azure AI Services on 4TB of healthcare unstructured data, streamlining clinical workflows and reducing documentation time by 50%.
- Analyzed e-commerce customer behaviors with Clustering techniques (K-Means++, DBSCAN) for customer segmentation, uncovering key insights using Power BI that boosted targeted marketing strategies and increased ROI by 35%.
- Led & 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://www.1-800-flowers.com), *AnytownUSA*, *DigiNurse (BSSNY Hospital)*, driving 40% growth in annual business revenue.
- Drove 30% increase in website traffic and conversions with focused strategic analytics, increasing sales & customer engagement.

### Software Engineer

July 2017 - December 2018

Appster

Gurugram, India

- Created RESTful APIs using Python, PHP, Java, MySQL and version control (GIT), ensuring seamless applications integration.
- Developed mobile applications achieving 350K downloads, integrating 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 & user engagement, leading to 15% increase in retention.

## SKILLS

- Programming & Databases:** Python, SQL, Snowflake, R, PHP, MySQL, PostgreSQL, NoSQL (MongoDB)
- Data Science Libraries:** Pandas, NumPy, Scikit-learn, SciPy, PyTorch, Matplotlib, Seaborn, Hugging Face, OpenAI
- Machine Learning & AI:** Regression, Classification, Clustering, Regularization, Dimensionality Reduction (PCA), Large Language Models (LLMs), Natural Language Processing (NLP), Generative AI (GenAI)
- 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, Feature Engineering, ETL (Extract, Transform, Load)
- Cloud Platforms, Big Data:** Microsoft Azure, Amazon Web Services (AWS), GCP, Spark, Hadoop, JIRA (Agile), MLflow

## 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.

## EDUCATION

### Northeastern University, Boston

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

September 2021 - April 2023

GPA: 4.0

### Guru Gobind Singh Indraprastha University

Bachelor of Technology, *Information Technology*

August 2013 - July 2017

GPA: 3.71