HARSHIT GAUR

gaurharshit04@gmail.com | 617-396-0129 | **Portfolio**: harshitgaur6155.github.io/
linkedin.com/in/harshitgaur-/ | public.tableau.com/app/profile/harshitgaur | github.com/harshitgaur6155

WORK EXPERIENCE

Greatpoint HR Inc.   Boston, MA
Lead Data Analyst   *May 2023 - Present*

- Built and deployed machine learning model with **Large Language Models (LLMs)** as **BERT** and **Sentence Transformers** for automated job categorization, achieving 88% accuracy and boosting HR analytical efficiency by 40%.
- Architected and managed high-volume scalable **ETL** data pipeline for multi-source ingestion using **Snowflake**, **SQL**, and **Python**, processing over **2.3 billion** records.
- Performed predictive modeling & hypothesis testing in statistical analysis to identify key drivers of compensation trends, resulting in job price optimization model that enhanced hiring strategies & augmented candidate acceptance by 25%.
- Led and collaborated with cross-functional teams to develop market responsive **Recommendation System** that automated salary structures & grades generation, resulting in 20% growth in KPIs.
- Designed interactive **Tableau** dashboards that presented recruitment metrics, enhanced stakeholders communication, and delivered data-driven insights, reporting 20% rise in hiring manager satisfaction scores regarding new hires' performance.
- Analyzed financial compensation data to uncover trends & anomalies, yielding actionable insights that guided executive leadership in salary planning, budgeting & decision-making, accomplishing 15% improved talent acquisition & business outcomes.

Mobikasa   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.
- Designed and orchestrated **SQL** databases and leveraged **Python**, **PHP**, and **Shopify** to develop applications: *1-800-Flowers.com, AnytownUSA, DigiNurse (BSSNY Hospital)*, driving 40% growth in annual business revenue.
- Led and mentored a team of 4, cultivating collaboration to deliver data-driven solutions aligned with strategic business objectives, optimizing operational efficiency.
- Analyzed e-commerce customer behaviors with **Power BI** and **clustering** techniques, uncovering key insights that boosted targeted marketing strategies and transformed ROI by 35%.
- Drove 30% sales increase and elevated customer engagement, improving website traffic and conversions with focused strategies.

Appster   Gurugram, India
Software Engineer   *July 2017 - December 2018*

- Created REST APIs using Python, PHP, Java, MySQL and version control (GIT), ensuring seamless integration across application.
- Optimized project roadmap with **JIRA** and **Excel**, reducing deadlines by 7% through data-driven planning and problem-solving.
- Developed mobile applications (*MyHero, Bazar, COEY*) to meet clients business requirements, achieving over 350,000 downloads.
- Implemented cryptography for 2-way security and streamlined online transactions with secure payment gateways integration.

PROJECTS

Airbnb Price Estimator: Recommendation System *(click to view)*

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

EDUCATION

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

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

SKILLS

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