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

Data Scientist / Lead Data Analyst *May 2023 - Present*

- Researched, implemented & deployed innovative machine learning model (ML) with **Large Language Models (LLMs)** as **BERT** and **Sentence Transformers** to automate job category assignments, achieving 88% accuracy, enhancing HR analytical precision.
- Architected and managed high-volume **ETL** data pipeline for multi-source ingestion, transformation, processing over 2.3 billion records with **Snowflake, SQL**, and **Python**, facilitating **data analysis** and **machine learning** model development.
- Led and collaborated with cross-functional teams to develop **Recommendation System** that analyze job posting, market trends and geographical factors to deliver automated salary structures & grades, with 20% increase in key performance indicators (KPIs).
- Analyzed financial compensation data to uncover trends and anomalies, providing actionable insights guiding executive leadership in salary planning, budgeting and decision-making, achieving 15% improved talent acquisition and strategic business outcomes.
- Designed interactive **Tableau** reporting dashboards, presenting data-driven insights to business stakeholders & clients, enhancing communication of key findings and driving strategies for job price optimization and performance metrics.
- Performed statistical analysis and predictive modeling to identify key drivers of compensation trends, leveraging hypothesis testing to provide recommendations to refine hiring strategies, enhance employee retention, and improve budget forecasting for clients.

Mobikasa Delhi, India

Data Scientist (Software Engineer) *December 2018 - August 2021*

- Utilized **Azure Cognitive Service** to implement **unsupervised** and **deep learning** techniques on 4TB unstructured data, achieving 87% accuracy in Speech-to-Text Translation of complex surgical debrief and timeout audio sessions.
- Designed and orchestrated databases using **SQL** and utilized **Python, PHP, 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, enhancing technical collaboration and teamwork, driving project success to deliver data-driven solutions that align with strategic business objectives and optimization of operational efficiency.
- Achieved a 30% increase in sales and enhanced customer engagement by categorizing customer behaviors using **Power BI** and **cluster analysis**, enabling focused marketing strategies that boosted website traffic and conversions.

Appster Gurugram, India

Software Engineer *July 2017 - December 2018*

- Created REST APIs in Python, PHP, and Java for applications, utilizing MySQL and maintaining version control with GIT.
- Analyzed user stories via **JIRA** and **Excel**, optimizing project roadmap, reducing deadlines by 7% with data-informed planning.
- Developed mobile applications (*MyHero, Bazar, COEY*) catering clients business requirements with 350,000 plus downloads.
- Incorporated cryptography for 2-way security in applications. Integrated payment gateways to streamline online transactions.

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
- **Other Skills**: Problem-solving abilities, Attention to detail, Presentation and Communication skills for engaging stakeholders