HARSHIT GAUR

Boston, MA | gaurharshit04@gmail.com | +16173960129 | **Portfolio**: https://harshitgaur6155.github.io/
www.linkedin.com/in/harshitgaur-/ | public.tableau.com/app/profile/harshitgaur | www.github.com/harshitgaur6155

SKILLS

- **Programming Languages**: Python, SQL, NoSQL, PHP, Java, R, HTML, CSS, JavaScript, Bootstrap, Hive, Impala
- **Databases**: MySQL, MSSQL, Oracle Database, PostgreSQL, MongoDB, Apache HBase, Databricks
- **Machine Learning Libraries**: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, XGBoost, Fastai, SciPy, AutoML
- **Data Visualization Tools**: Tableau, Microsoft Power BI, Microsoft Excel, R Shiny, Google Data Studio
- **Cloud Technologies and Tool**: Amazon Web Services (AWS), Microsoft Azure, GCP, Spark, Hadoop, GIT, JIRA, Shopify

EDUCATION

Master of Science in Computer Science, *Machine Learning and Data Analytics* — September 2021 - March 2023
Northeastern University — **Boston, MA**
- Coursework: *Intermediate Analytics, Machine Learning, Data Mining, Big Data, Statistics, Fundamentals of AI* — **GPA: 4.0**

Bachelor of Technology, *Information Technology* — August 2013 - July 2017
Guru Gobind Singh Indraprastha University — **Delhi, INDIA**
- Awards: *magna cum laude*; Silver Medalist in Single's & Double's Badminton Cup — **GPA: 3.71**

WORK EXPERIENCE

Software Development Engineer, *Mobikasa Private Limited* — December 2018 - August 2021
- Led the design and development of multiple enterprise-level microservice applications and websites: ***1-800-Flowers.com, AnytownUSA, DigiNurse (BSSNY Hospital)***, driving $3.45 billion in revenue annually, using AWS, PHP, Python, SQL, Java, Shopify.
- Achieved translation of audio to lingual using ***Microsoft Azure Cognitive Services*** with 67% accuracy within the project of an established US-based hospital chain (BSSNY) to record & track debrief-timeout audio sessions of surgeries and operations.
- Developed e-commerce marketplace website ***AnytownUSA***, deployed on AWS EC2 instance, which served 125,000 US citizens.
- Accomplished improved customer-oriented sections on the website with 30% increased sales by integrating Google Analytics to monitor traction and spearheading data mining techniques to investigate customers' shopping behaviors and patterns.
- ***Achievements****: Got recognized by CEO for technical prowess and adaptability to cater to complex problems and requirements.*

Software Engineer, *Appster LLP* — July 2017 - December 2018
- Created RESTful APIs with MVC architecture implementation using scrum methodology. Assisted in User Story Analysis.
- Developed enterprise mobile-applications (MyHero, Bazar, COEY) catering client's business logics with 350,000 plus downloads.
- Designed database architectures. Incorporated cryptography for 2-way security in applications, and payment gateways.
- ***Achievements****: Got considered for the role of Business Analyst. Commended with Best Performer of the Month twice.*

PROJECTS

Airbnb Price Estimator
- Created dynamic Airbnb price estimator, deployed on Streamlit, using predictive analytics taking in user input to calculate price.
- Predicted price using **Decision Tree**, **Random Forest**, **XGBoost** algorithms, and **LIME**, **SHAP** for interpretability.

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 customer groups for H&M marketing teams to focus for the next promotional offers to bolster revenue.

Prediction of readmittance of diabetic patients using analysis of HbA1c value
- Uncovered that readmission of diabetic patients depends on HbA1c value above 7 and primary diagnosis.
- Achieved readmission predictions of patients using **k-NN**, **Random Forest Classification**, **Decision Tree**, and **SVM** algorithms.

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.

Business Solution to identify inaccuracy in ADCS prediction for company - DocDigitizer
- Examined that 21% times Automated Data Capture System (ADCS) of DocDigitizer extracts incorrect information. Implemented Natural Language Processing (NLP) to collect textual data in documents based on the business logic of the company.
- Built predictive model (Decision Tree) identifying incorrect extracted information & factors responsible for inaccuracy of ADCS.