



RAHUL GUPTA

DOB: 09/08/1995

PASSPORT : Z5733220

AADHAR : 952452161704

EMAIL: rahul.gupta2608@yahoo.com

WEBSITE: www.rahulg.info

NATIONALITY: INDIA

MARITAL STATUS: SINGLE

+91 9911666741

14/84 GF, VIKRAM VIHAR,
LAJPAT NAGAR 4,
NEW DELHI, DELHI - 110024



Summary

Development Engineer with 4+ years of experience in machine learning and artificial intelligence, specialising in building **end-to-end ML solutions from exploratory data analysis to model deployment**. M.Tech. in Computer Science from San Diego State University. Strong expertise in the fundamentals of machine learning and **artificial intelligence**, with proven ability to implement the complete **ML lifecycle from EDA to model deployment**. Proficient in **data analysis tools** including **Jupyter, NumPy, Matplotlib**, and **Pandas** for comprehensive data exploration and visualisation. Advanced programming skills in **C/C++** and **Python**, consistently delivering reliable, maintainable, secure, and **performance-optimised** code across multiple platforms. Experienced in **data synthesis and processing for model training**, with particular focus on creating **robust datasets** for computer vision applications. Extensive hands-on experience with deep learning frameworks including **TensorFlow, PyTorch, Keras**, and **Torch**, implementing **state-of-the-art models for image recognition** and processing. Demonstrated expertise in **computer vision and graphics**, having developed **feature extraction algorithms** and image enhancement solutions for creative software applications. Experience in collaborating with cross-functional teams to integrate **ML capabilities into production-ready software**.

Skills

- **Programming Languages:** Python, SQL, R, DAX, C, C++
- **Database/Data Warehouse:** MySQL, PostgreSQL, Microsoft SQL Server, AWS RDS, Snowflake, Azure Synapse Analytics, Google BigQuery
- **Cloud Platform:** AWS (S3, Glue, RDS, DMS, CloudWatch, CloudFormation, IAM, SNS), Azure (Data Factory, Databricks, Synapse Analytics), GCP (BigQuery, Cloud Storage)
- **Data Processing/Orchestration:** Apache Spark, PySpark, Apache Airflow, Apache Kafka, AWS Glue, Azure Data Factory
- **Data Visualization/Report:** Power BI, Tableau, Excel, Power Query, DAX, Python (Matplotlib, Seaborn), R
- **Libraries:** NumPy, Pandas, SciPy, Matplotlib, Seaborn, PySpark, Boto3, Scikit-Learn, PyTorch, TensorFlow, Hugging Face, NLP toolkits, GenAI APIs
- **ML/AI Techniques:** Regression (Linear, Logistic, Ridge), Decision Tree, Random Forest, KNN, SVM, KMeans, Ensemble Model (XGBoost, LightGBM), Neural Network (DNN, CNN), NLP (text preprocessing, sentiment analysis), Feature Engineering
- **Statistical Analysis:** GLM, Bayesian Inference, Time Series, Non-Parametric statistics
- **Development/Version Control:** Jupyter Notebooks, RStudio, Visual Studio Code, Git, GitHub

Employment History

San Diego State University, San Diego, CA

Independent Researcher

Project: Neural Compression, Texture Synthesis, Image Processing

June 2024 - Present

Led comprehensive research and implementation of a groundbreaking approach to texture set compression that seamlessly integrates traditional **GPU texture representation** methodologies with advanced Neural Image Compression techniques. By leveraging multiple levels of redundancy across channels, resolutions, and spatial dimensions, this research establishes a **new state-of-the-art benchmark** for texture compression in photorealistic rendering environments, significantly outperforming both **conventional compression methods** and **competitive neural approaches** in comprehensive evaluations.

Roles and Responsibilities:

- Developed an asymmetric **auto-encoder framework** with global transformer, **grid constructor**, **grid sampler**, and texture synthesizer components.
- Implemented **multi-resolution support** for mip levels essential for texture filtering in real-time rendering.
- Analyzed compression performance across different texture types, **achieving 240x more** efficiency compared to traditional methods.
- Optimized **positional encoding** in the texture synthesizer to enable **high-quality reconstruction**.
- Evaluated model performance using **PSNR** and **SSIM** metrics, demonstrating **88% BD-Rate savings versus ASTC**
- Created visualizations of grid features showing dual-bank representation capturing different frequency information

Technologies Used: PyTorch, Neural Networks, Computer Graphics, Texture Compression, Positional Encoding, Autoencoder Architectures, PSNR/SSIM Analysis

WalmartLabs, Sunnyvale, CA

ML ENGINEER

Project: PredictiveSales Analytics Platform

January 2019 – May 2022

Developed a **machine learning forecasting model** to predict total sales for each product and store for the upcoming months using daily historical sales data. This involved data preprocessing, feature engineering, and applying models including **Ridge**, **XGBoost**, and **LightGBM**. The project optimized model accuracy and provided valuable forecasts for inventory management and strategic planning.

Roles and Responsibilities:

- Imported and merged multiple datasets into **pandas** DataFrames, removed duplicates and imputed missing value.
- Engineered numerical, categorical, and text-based features, including **TF-IDF** embeddings and **matrix factorization** of product and store names, lag features, and trend-based **time series indicators**.
- Conducted **Exploratory Data Analysis (EDA)**, including visualisation of target distribution and time trends. Used multivariate heat-maps to analyse numerical and categorical pairings.
- Applied **mean encoding** for categorical variables and constructed ML pipelines with **Ridge**, **XGBoost**, and **LightGBM** regressors.
- Performed **feature selection** using **Recursive Feature Elimination with Cross-Validation (RFECV)** and optimized hyper-parameters using Bayesian optimisation to minimise **Root Mean Square Error (RMSE)**.
- Evaluated models with cross-validation and deployed the best-performing pipeline for ongoing forecasting. Predicted future outcomes and compiled comprehensive reports. Improved forecasting accuracy by 20% over baseline, reducing inventory mismatches and supporting proactive stocking decisions across multiple stores.

Technologies Used: Python, Scikit-Learn, Machine Learning Pipeline, NLP, TF-IDF, mean encoding, matrix factorization, Ridge Regressor, LightGBM, XGBoost, feature selection, hyperparameter optimization.

SynergisticIT, Fremont, CA
Data Analyst / Data Engineer

Project: Social Media Sentiment Analysis and Reporting

June 2018 – June 2019

Partnered with a data scientist to develop an Azure-based NLP system for sentiment analysis on 16,000 post-sale reviews for an online clothing store. Built an **Azure Data Factory** ingestion pipeline and implemented an **Azure Databricks (PySpark)** workflow to clean, process, and classify review text using a **BERT-based model**. Visualized sentiment trends and generated automated weekly reports to support product feedback analysis and marketing decisions.

Roles and Responsibilities:

- Configured an **Azure Data Factory** pipeline to ingest data from an on-premise database into **Azure Data Lake Storage**, scheduled for weekly updates.
- Designed an **Azure Databricks** notebook using **PySpark** to perform data cleaning, deduplication and null handling.
- Implemented **NLP** models for text cleaning, lemmatisation, stop-word removal, and perform sentiment intensity analysis with **BERT** based model from **Hugging Face Transformers**. Trained a **Random Forest classifier** to predict sentiment labels.
- Evaluated model performance and optimised preprocessing parameters to improve classification accuracy.
- Visualised results by time and item, producing comprehensive sentiment trend reports. Concluded an overall satisfactory rate of 96% and trend upward over time.
- Automated the entire workflow from ingestion to reporting using **ADF triggers** and Databricks job scheduling

Technologies Used: Python, PySpark, Azure Data Factory, Azure Databricks, Azure Data Lake Storage, Hugging Face Transformers, BERT, Random Forest, NLP

Education

MS in Computer Science San Diego State University (GPA: 3.47/4)	September 2022 – June 2024
BS in Computer Science San Diego State University (GPA: 3.73/4)	September 2014 – July 2018