ROHIT KUMAR

Bengaluru, IN 560073 ◆ +91-8144185441 ◆ rohitgitam014@gmail.com ◆ LinkedIn

PROFESSIONAL SUMMARY

Detail-oriented Data Scientist with hands-on experience in machine learning, predictive modelling, NLP, and data visualization. Passionate about leveraging AI-driven solutions to extract actionable insights and drive business growth. Strong in building end-to-end data science workflows, including data preprocessing, feature engineering, model development, and evaluation. Proficient in Python, SQL, Power BI, and Tableau, with a solid foundation in real-time data processing, analytics, statistics, and business problem-solving, along with foundational exposure to NLP-based chatbots and prompt-driven Generative AI workflows.

SKILLS & TOOLS

- Programming & Data Science: Python, SQL, Machine Learning, NLP, Deep Learning, Time Series Forecasting, Feature Engineering, Dimensionality Reduction (PCA), Data Preprocessing (StandardScaler, OneHotEncoder, LabelEncoder), Feature Selection, Model Evaluation & Selection (Cross-validation, GridSearchCV, Stratified K-Fold), Imbalanced Data Handling (SMOTE, class weights)
- Data Analytics & Visualization: Tableau, Power BI, EDA, Predictive Modeling, Statistical Modeling
- Business & Communication: Client Presentations, Business Understanding, Report Generation
- Frameworks & Tools: Scikit-learn, Seaborn, Pandas, NumPy, Hadoop
- Math & Stats: Linear Algebra, Probability & Statistics, Gradient Descent, Hypothesis Testing, Regularization, Bayes Theorem, Central Limit Theorem, Statistical Inference, Correlation & Covariance, Information Theory (Entropy).

PROFESSIONAL EXPERIENCE

Data Scientist

Palette – Artful Craft, Colorful Tech (The AI-Powered Creative Solutionists)(Aug 2025 – Feb 2026)

- Applied Python-based data science techniques to analyze the *Hazzys Fashion* retail dataset and develop machine learning models for Total Sales prediction.

- Conducted comprehensive Exploratory Data Analysis (EDA) to understand customer purchase behavior, product-level sales distribution, pricing impact, and key drivers influencing sales performance.
- Performed data preprocessing and feature engineering, including missing value treatment, categorical encoding, and numerical feature scaling, to prepare high-quality structured datasets.
- Built and optimized a Random Forest Regressor to model market trends and customer purchase patterns, achieving accurate sales predictions across multiple product categories.

- Evaluated model performance using R^2, RMSE, MSE, MAE, and MAPE, ensuring reliable, interpretable, and business-ready predictions.

- Translated analytical findings and model outputs into actionable business recommendations, supporting inventory planning, demand forecasting, and revenue optimization.
- Extended traditional data science workflows to NLP and Generative AI–style use cases, including chatbot development and prompt-based response experimentation using pre-trained language models.

Data Scientist

***Ahana Systems and Solutions* (Jan 2025 - July 2025)**

- Developed and delivered predictive models projects focused on predictive modeling and machine learning to address real-world business challenges across finance and operations.
- Performed data preprocessing, feature engineering, feature selection, and model optimization using Python libraries such as Pandas, Scikit-learn, and NumPy, boosting model accuracy and performance.
- Developed AI-powered solutions to uncover patterns and generate actionable business insights supporting strategic decisions.
- Built and deployed a stock price prediction model, using Last Traded Price (LTP) as the target variable and engineered features from open, close, high, low, and volume data.

- Applied machine learning algorithms such as Random Forest to predict LTP movement, enabling more informed and timely investment risk decisions.
- Validated model performance using metrics like R-squared, MAE, and MSE, and visualized key results using Seaborn and Matplotlib for stakeholder communication.

Data Science Intern

***BDreamz Global Solutions Private Limited* (Jun 2023 - Aug 2024)**

- Built a classification model using Python to solve a real-world business prediction problem with practical implications.
- Conducted Exploratory Data Analysis (EDA) to uncover meaningful patterns and relationships in structured data.
- Created clear and informative data visualizations using Seaborn and Matplotlib to aid in feature understanding and presentation.
- Performed data preprocessing, including handling of missing values, encoding categorical variables, and feature scaling using StandardScaler.
- Engineered new features to improve model performance and eliminate redundant ones using correlation analysis and feature importance techniques.
- Addressed imbalanced class distribution using SMOTE to enhance the model's ability to detect minority class outcomes.
- Validated key regression assumptions such as linearity, independence, and homoscedasticity to ensure robustness and interpretability.
- Trained and evaluated models using Logistic Regression, achieving 87% recall score, significantly improving positive case detection.
- Assessed model effectiveness using metrics such as Precision, Recall, F1-score, Confusion Matrix, and ROC-AUC.
- Collaborated with mentors and peers to document findings, present results, and refine models based on business feedback.

Data Science Capstone Project

***UpGrad | Uber India Systems Private Limited (Remote)* (Nov 2023 - Feb 2024)**

- Solved a real-world business problem using Python, machine learning, and end-to-end data science methodologies.
- Followed the CRISP-DM framework to structure the project lifecycle from problem understanding to model evaluation.
- Conducted detailed Exploratory Data Analysis (EDA) using Pandas, Seaborn, and Matplotlib to identify trends, correlations, and outliers.
- Performed data preprocessing including missing value imputation, encoding, and feature scaling using Scikit-learn tools.

- Applied feature selection techniques (correlation analysis, model-based selection) to improve model performance and reduce overfitting.
- Trained multiple regression models and evaluated their performance using R-squared, MAE, and RMSE.
- Achieved 89% accuracy (R^2 score) with Linear Regression, outperforming baseline metrics.
- Presented final results with data visualizations and actionable insights tailored to stakeholders for business decision-making.

PROJECTS

Stock Market Price Prediction | GitHub Repository

- Analyzed historical stock price data (Open, High, Low, Close, Volume) to identify market patterns and investor behavior.
- Built a regression-based machine learning model using Random Forest to predict Last Traded Price (LTP).
- Performed comprehensive EDA and feature engineering, including outlier treatment, correlation analysis, and scaling to prepare the dataset.
- Applied feature selection techniques to eliminate redundant variables and enhance model interpretability.
- Evaluated model performance using R-squared, Mean Absolute Error (MAE), and Mean Squared Error (MSE) to assess prediction accuracy.
- Delivered risk management insights by identifying price volatility patterns and helping inform trading decisions.
- Created clear data visualizations using Seaborn and Matplotlib to support analysis and findings for stakeholders

AI Chatbot using NLP | GitHub Repository

- Developed an AI-powered chatbot capable of handling user queries and simulating natural conversation using Natural Language Processing (NLP) techniques.
- Implemented text preprocessing workflows including tokenization, normalization, and lemmatization using NLTK for cleaner input handling.

- Designed an intent recognition pipeline to classify user queries and extract relevant entities for accurate response routing.
- Combined rule-based logic with pre-trained language models to improve response precision and fallback handling.
- Created a structured dialogue flow to manage multi-turn conversations and context-aware responses.
- Tested chatbot functionality across various scenarios and intents to improve accuracy, relevance, and user satisfaction.
- Visualized token-level analysis and word frequency trends to debug and enhance language understanding capabilities.

Fake URL Detection using LSTM

- Built a phishing detection system using deep learning (LSTM) to classify malicious and legitimate URLs with high precision.
- Collected and processed real-world URL datasets from PhishTank, performing cleaning, labeling, and transformation into a suitable format for sequence modeling.
- Converted textual URL data into numerical sequences using tokenization and embedding techniques, enabling LSTM to learn URL patterns effectively.
- Designed and trained a Long Short-Term Memory (LSTM) neural network to capture sequential dependencies in character-level URL structures.
- Achieved high detection accuracy, significantly improving model performance in identifying phishing attempts and enhancing cybersecurity awareness.
- Evaluated model performance using confusion matrix, precision, recall, F1-score, and ROC-AUC to ensure robustness across test scenarios.

EDUCATION & CERTIFICATIONS

- Certified Data Scientist | University of Texas
- Post Graduate in Data Science & Business Analytics | Great Learning
- Graduate Certificate in Data Science | UpGrad
- B.Tech in Computer Science & Engineering | GITAM Institute of Science and Technology