

Shivam Kumar

itshivam96@gmail.com | +91 8920012521 | linkedin.com/in/itshivam96 | github.com/itshivam96 | Gurugram, India (Open to Relocate)

hackerrank.com/profile/shivamji9953 | shivamk-eta.vercel.app

SUMMARY

Proficient in Python, SQL, Advanced Excel and Power BI with expertise in statistical analysis and predictive modeling. Successfully developed ML models achieving 90%+ accuracy and optimized database performance by 40%. Skilled in creating interactive dashboards and translating complex data into actionable business insights.

EXPERIENCE

- Capgemini & AISECT, Data Analyst Trainee**11/2025 – 03/2026 | Gurugram, India
 - Completed a 6-month industry-focused Data Analytics training, building hands-on skills in Python, SQL, Excel, and Power BI for real-world business use cases.
 - Analyzed 10,000+ customer records using Python (Pandas), SQL, and Excel to identify trends and improve data interpretation accuracy by ~20%.
 - Built interactive Power BI dashboards to track KPIs, improving reporting efficiency and supporting faster data-driven decision-making (~25%).
 - Performed data cleaning, transformation, and exploratory data analysis (EDA), reducing inconsistencies by ~30% and generating actionable insights.
- Honeywell & ICT Academy, Cybersecurity Trainee**01/2025 – 02/2025 | Gurugram, India
 - Achieved 95% threat detection accuracy by systematically analyzing 1,000+ security log entries to identify and document critical patterns.
 - Identified and documented 15+ critical vulnerabilities through comprehensive network analysis and risk assessment reporting.
 - Reduced potential attack vectors by 40% through rigorous testing across 3 system architectures to enhance security protocols.
 - Improved team incident response efficiency by 25% by coordinating with cross-functional teams to implement a new security framework.

EDUCATION

- Bachelor of Technology in Data Science, DPG Institute of Technology and Management**2021 – 2025 | Gurugram, India
CGPA: 8/10
- Class 12th (PCM), Government Senior Secondary School**2020 – 2021 | Gurugram, India
Percentage: 92.6%

SKILLS

Programming Languages Python, R, SQL, Java (Basic)	Libraries/Frameworks Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn	Databases MySQL, PostgreSQL, Advanced SQL (Joins, CTEs, Window Functions, Query Optimization)	Machine Learning & Statistics Regression, Classification, Clustering, A/B Testing, Hypothesis Testing
Data Visualization Dashboards, DAX, Power Query, Data Storytelling, KPI Tracking	Spreadsheet Tools Microsoft Excel (Pivot Tables, XLOOKUP/VLOOKUP, INDEX-MATCH)	Tools & Platforms Git / GitHub, Jupyter Notebook, Kaggle, Google Colab	Professional Skills Team Collaboration, Project Management

PROJECTS

- Customer Shopping Behaviour Analysis** 
Tools Used: Python, Pandas, Scikit-learn, SQL, Power BI
 - Analyzed 99,000+ retail transactions across customer demographics, product categories, time periods, and regions, revealing that customers aged 25–34 purchasing electronics and clothing generated 62% of total revenue.
 - Created Tableau dashboards to track purchase patterns, peak shopping hours, and basket sizes, while using SQL-based RFM segmentation to support 5 targeted marketing recommendations.
- Medical Insurance Cost Prediction** 
Tools Used: Python, Scikit-learn, Pandas, Machine Learning
 - Developed an end-to-end machine learning pipeline to predict medical insurance charges using statistical analysis, feature engineering, and regression modeling on 1,338 records, ensuring clean data preprocessing and structured workflow execution.
 - Built and evaluated multiple models including Linear Regression and Random Forest Regressor, improving performance from $R^2 = 0.78$ to $R^2 = 0.865$ and reducing error to $RMSE \approx 4,578$, demonstrating effective model selection and optimization.
 - Performed exploratory data analysis and feature importance analysis to generate actionable insights, identifying smoking status (61% influence), BMI, and age as key cost drivers for risk and cost optimization strategies.
- Customer Churn Prediction System** 
Tools Used: Python, Pandas, Scikit-learn, SQL, Power BI
 - Created a predictive model using Random Forest that achieved 90% accuracy in identifying at-risk customers, enabling targeted retention efforts for the top 20% most likely to churn.
 - Executed comprehensive feature engineering and analysis that identified key churn drivers (support ticket volume, usage frequency), supporting a projected 5% reduction in monthly customer attrition.
- Medical Imaging Analysis System** 
Tools Used: Python, Scikit-learn, Pandas, Matplotlib, Seaborn
 - Designed and trained a deep learning model (CNN) using TensorFlow to classify chest X-rays, streamlining a previously manual diagnostic process and achieving 92% model accuracy.
 - Executed the full data science pipeline: cleaned and pre-processed 5,000+ raw images with OpenCV, streamlined data augmentation to balance classes, and validated results using a confusion matrix and classification report.

CERTIFICATIONS

Python for Data Science  Issued by IBM	SQL (Intermediate)  Issued by HackerRank	Data Analyst Training  Issued by Capgemini	Cybersecurity Training  Issued by Honeywell
---	---	---	--