ANKIT DEY

ankit.dey.pc@gmail.com | +91 7319240891 | LinkedIn | N.S.B Road Bye Lane, Raniganj, West Bengal, 713347

PROFESSIONAL SUMMARY

Data Analytics & **AI/ML** Intern skilled in **Python**, **SQL**, and **Scikit-learn**, with **hands-on experience** in **building predictive models**, **automating data cleaning workflows**, and **creating visual dashboards**. Passionate about leveraging data modeling and machine learning to drive product intelligence and decision-making in fintech environments.

EDUCATION

Haldia Institute of Technology
Applied Electronics and Instrumentation Engineering
June, 2026
Haldia, IN
8.13 CGPA

CBSE, XII - HEM SHEELA MODEL SCHOOL
March, 2020
Core Science – 74.2%

TECHNICAL SKILLS

- Programming & Data: Python, SQL, Pandas, NumPy, Scikit-learn, TensorFlow (basic)

- Data Extraction: PDF → Excel, HTML Tables, Web Scraping (BeautifulSoup), OCR (Tesseract)

- Excel & Automation: Excel Dashboards, Pivot Tables, VLOOKUP/XLOOKUP, Power Query, VBA (Macros)

- Data Visualization: Power BI, Tableau (Basics), Matplotlib, Seaborn

- Statistical Analysis: Descriptive Stats, Correlation, Hypothesis Testing

- Data Engineering: ETL Pipelines, API Integration, Data Quality Checks, Validation Audits

- Version Control: Git, GitHub

PROJECTS

Credit Card Fraud Detection (AI/ML MODEL) | *REPORT*   *Python, Scikit-learn, Data Preprocessing*

- Designed a fraud detection model using Logistic Regression, Decision Tree, and Random Forest with Scikit-learn.
- Handled **data cleaning, feature scaling, and class imbalance correction (SMOTE)** on a **284K-record Kaggle dataset**.
- Improved **precision from 28% → 33%** while maintaining **99.6% accuracy**, strengthening detection reliability.
- Built a **modular pipeline architecture** supporting multi-model integration for scalability.
- Evaluated models using **accuracy, recall, and F1-score**, outlining use in **real-time fintech fraud systems**.

House Price Prediction using Linear Regression | *REPO*   *Python, Scikit-learn, Pandas, Numpy, Matplotlib Seaborn*

- Developed a **Linear Regression model** to estimate housing prices based on numerical and categorical property features.
- Conducted EDA, handled missing values, and evaluated model performance using **Mean Squared Error** and **R² score** metrics.
- Model achieved **R² ≈ 0.87** on test data, indicating strong predictive accuracy.

🍷 Wine Quality Prediction | *REPO*   *Python, Scikit Learn, Pandas, Numpy*

- Built predictive models (**Random Forest, SVC, SGD Classifier**) to classify wine quality from physicochemical features such as **acidity** and **alcohol content**.
- Performed data cleaning, feature correlation, and performance evaluation using accuracy, precision, and recall metrics.
- Result: Achieved **91% accuracy** with **Random Forest**; visualized key relationships via **heatmaps**.

Project: Project Samarth – Intelligent Q&A System (Bharat Digital Fellowship Challenge) | *WEB APP*

- Designed and deployed an intelligent data-driven Q&A system using Streamlit and Python, integrating real datasets from the Ministry of Agriculture and IMD to analyze India's agricultural and climate patterns.
- Implemented data cleaning, merging, and real-time querying pipelines to answer complex, natural language questions with accurate, source-cited insights.
- Deployed a live, interactive web app on Streamlit Cloud for public access and policy-oriented data exploration.

PROFESSIONAL SKILLS

- Clear written and verbal communication for stakeholder collaboration.
- Strong analytical thinking with data-driven problem solving, critical thinking, and ability to navigate ambiguous tasks independently or in teams.
- Independent contributor with cross-functional teamwork and adaptability.
- Efficient time management and system optimization through project execution.
- Strong attention to detail with experience performing data accuracy checks and quality audits.

INTERNSHIPS

Data Analytics & AI/ML Intern | Oasis Infobyte | Oct 2025 – Nov 2025   OFFER LETTER

- Clean, transform, and model structured and semi-structured datasets using Python, Pandas, and NumPy to support AI/ML initiatives.
- Build and maintain ETL pipelines and prepare model-ready datasets for segmentation, prediction, and scoring.
- Assist in developing machine learning models (classification, regression, clustering) and feature engineering for product insights.
- Create dashboards and visualizations to deliver actionable insights for business and product stakeholders alongwith conducting data quality checks and validation audits to ensure accuracy of extracted and transformed datasets

Key Tools & Skills: Python, SQL, Pandas, NumPy, scikit-learn, TensorFlow (familiar), ETL/Pipelines, Dashboards, ML Models.

Data Analyst – I (Trainee Intern) | Intrainz Innovation | Jul 2025 – Sep 2025   *CERTIFICATE*

- Designed interactive dashboards for healthcare data, enhancing insights and decision-making.
- Processed patient datasets, applied data preprocessing and anomaly detection.
- Collaborated on data-driven product improvements—aligned with fintech data initiatives.
- Ensured data consistency and usability through SQL-based preprocessing and performance optimization.
- Performed data extraction from PDFs, web pages, and semi-structured text sources and converted them into structured Excel datasets.

AI and Machine Learning Industrial Internship | *CERTIFICATE*   *Ardent Computech*
Sep 2024 – Nov 2024

- Gained hands-on exposure to **AI model deployment, data security, and anomaly detection**.
- Applied **ML algorithms** for identifying irregularities in large datasets, improving model reliability.

CERTIFICATIONS AND RECOGNITION

- Certified by Oracle Corporation in Dat Science | *Certificate*
- Certified by HackerRank in Advanced SQL | *Certificate*
- Certified in Data Science & Analytics (HP LIFE) with hands-on knowledge of leading practices, methodologies, and tools for driving data-driven business decisions. | *Certificate*
- Obtained the "Getting Started with Tableau" certificate from Analytics Vidhya, in interactive dashboard development, data visualization techniques, and translating data into actionable insights. | *Certificate*
- Certified in Data Analysis with Apache Hive – gained hands-on expertise in big data querying, data warehousing, and analytical processing. | *Certificate*