📘Final Project Report: Nifty50 Stock Return Prediction

🕑Problem Statement

The objective of this project is to analyze two decades of historical data from the Nifty 50 index within the Indian Stock Market. The goal is to develop data-driven strategies for constructing a stock portfolio capable of delivering optimal future returns.

📊Dataset Overview

Source: Kaggle - Nifty50 Stock Market Data
Composition: 50 individual company CSV files
Total Records: 235,192 rows covering the period from 2000 to 2021

Key Attributes:

- Date, Open, High, Low, Close, Volume, Turnover, Deliverable Volume

---

🔍Methodology

🕰️Step 1: Data Consolidation & Cleaning

- Integrated 50 separate stock CSV files into one cohesive dataset
- Standardized date formatting and ensured chronological consistency
- Removed duplicates and handled missing values systematically

🕰️Step 2: Exploratory Data Analysis (EDA)

- Analyzed trends in daily returns and rolling volatility
- Assessed variability in performance across different stocks

🕰️Step 3: Feature Engineering

- Generated time-based financial features including:
- Daily Return
- 3-day, 5-day, and 7-day Forward Return
- 7-day Rolling Volatility
- Encoded stock identifiers for use in model training

🕰️Step 4: Strategy Development

- Computed per-stock return and volatility statistics
- Ranked and shortlisted stocks with high average returns and stable volatility
- Focused on consistent performers to construct a robust portfolio foundation

🕰️Step 5: Predictive Modeling

- Defined a binary classification target: whether a stock's 7-day forward return exceeds 1%
- Trained a Random Forest Classifier
- Addressed class imbalance using SMOTE oversampling
- Evaluated model using classification metrics on the test dataset

📈Model Evaluation Metrics:

Metric	Value
Accuracy	52%
Precision (1)	54%
Recall (1)	63%
F1 Score (1)	0.58

🗒️Technologies & Libraries Used

- Programming Language: Python
- Data Manipulation: Pandas, NumPy
- Machine Learning: Scikit-learn, Imbalanced-learn
- Visualization: Matplotlib, Seaborn

💻Author

Saurabh Tanwer
Capstone Project — Executive Post Graduate Certification in Data Science & Artificial Intelligence

GitHub: @mrsaurabhtanwer

End of Assignment 2 Report