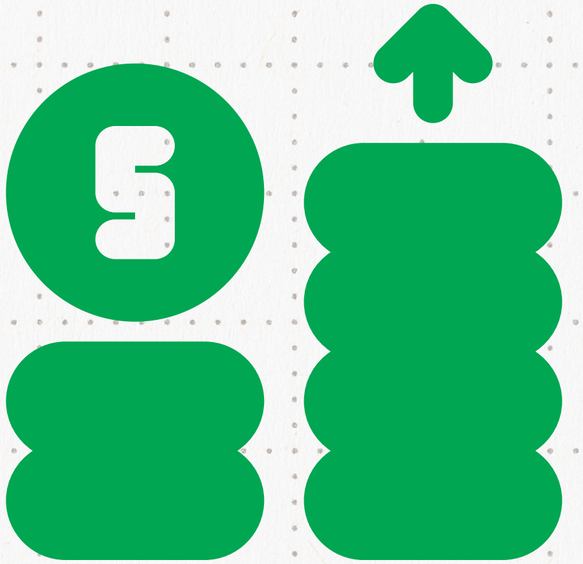
Predicting NIFTY50

USING PRINCIPAL COMPONENT ANALYSIS

Bhumika Mittal

What is NIFTY50?



NIFTY50 Meaning

The NIFTY 50 is a benchmark Indian stock market index representing the weighted average of 50 of the largest Indian companies listed on the National Stock Exchange.

Why NIFTY50 matters?

There is always the hunt for high-performing stocks that offers greater returns and helps in portfolio diversification. And this is where Nifty enters the picture.

Stock Selection Process

HDFC Bank, RIL, HDFC, Infosys, ICICI Bank, TCS, Kotak Bank, HUL

01 These are high-weighted stocks - contribute ~50% to NIFTY50 index

02 These are from different industries, hence contributing to the diverse portfolio

03 Most of them are large/mid-cap stocks

Why PCA?

What is PCA

Principal component analysis (PCA) is a popular technique for analyzing large datasets containing many dimensions per observation while preserving the maximum amount of information. Formally, PCA is a statistical technique for reducing the dimensionality of a dataset.

PCA & NIFTY50

Each stock acts as one dimension for the data. Hence we have 8 dimensions for our data set. Using the covariance matrix and eigenvalue, we can observe that 4 dimensions can present more than 80% of the information. Therefore, we can apply PCA to reduce the dimensions from 8 to 4.

Visualisation of Data

PCA emphasize variation and bring out strong patterns in a dataset. It's often used to make data easy to explore and visualize.

How is PCA used in predicting NIFTY50?

01

Import libraries & Data and clean data

Import all the required libraries. Then load the data from csv using pandas. Since not all columns are required, we remove unnecessary columns and assign x and y.

02

Standardize data and obtain covariance matrix

Since the data range varies from dimension to dimension, we must standardise the data before working with it. Then find covariance and eigen values for the same.

03

Find the cumulative variance and train the model

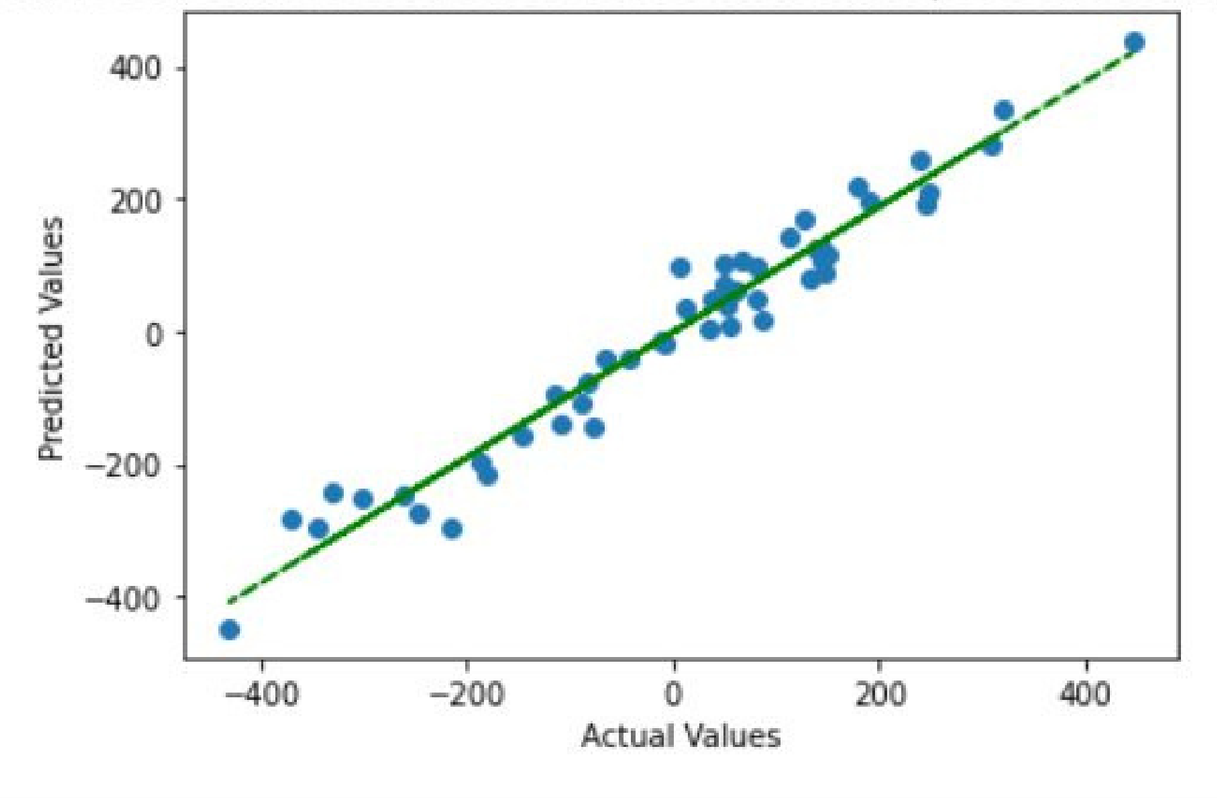
Observe that only the first 4 components contribute to more than 80% of the information. Hence, the rest can be reduced in dimension reduction. Divide the data into Training and Test with an 80:20 ratio.

04

Check accuracy by plotting the scatter plot

Since most of the data points lies on the trend line, we can say that the prediction model predicts the accurate value of NIFTY50.

Result of the model



RSME value ~ 40

The RSME value of the model is ~40, which means the prediction accuracy is very high, and the NIFTY50 returns are predicted correctly using 4 components instead of 8, as these 4 contribute to 84% of the information. Using PCA, we apply the method of dimension reduction.

Pros and Cons of this prediction technique

PROS

PCA help in improving performance at a very low cost of model accuracy. It also helps in the reduction of noise in the data, feature selection (to a certain extent), and the ability to produce independent, uncorrelated features of the data.

CONS

Although dimensionality reduction is useful, it comes at a cost. Information loss is a necessary part of PCA. Balancing the trade-off between dimensionality reduction and information loss is unfortunately a necessary compromise that we have to make when using PCA.

THANK YOU