1. **Problem**

The article addresses the problem of accurate forecasting in the Indian Stock Exchange. Accurate forecasting is a relevant issue prevalent in many branches of the financial and business spheres. What's more, forecasts can be used for the development of a business strategy. The authors outline three main types of forecasting: short-term, medium-term, and long-term forecasts, distinguished by the time period considered for forecasting, ranging from weeks to many years.

Generally, there are five main steps in a forecasting task. Initially, the aim, along with the demand for forecasts, is identified. The next step involves gathering the necessary, appropriate data for making the forecast. Thereafter, the analysis of data identifies beneficial or less useful information. Once this step is complete, a model with the best prediction result is chosen and used for making forecasts. Lastly, the model is put into practice with cyclical evaluation and tuning.

The solution proposed by the authors is an ARIMA (Auto-Regressive Integrated Moving Average) model, which was employed to predict stock market movements.

2. **Actuality**

Forecasting is an important problem with vital importance in all areas of the real world like business, medicine, social science, politics, finance, government, economics, environmental sciences and others.

Financial forecasting involves predicting future business conditions that are expected to impact a company, organisation, or an economy. It provides a forecast regarding the forthcoming financial status of a company. The article mentions advantages in an efficient financial forecast, which include driving the business in an accurate direction, specifying a point of reference, and identifying probable threats.

India is a rapidly developing country and offers opportunities for investment in two of the largest Indian stock markets: NSE (National Stock Exchange) and BSE (Bombay Stock Exchange).

Research regarding modelling time series has witnessed significant attention in recent years. Time series have wide applications in business, economics, science, engineering, banking, and finance.

3. **The state of the point of study**

The work is complete, and the authors of the article do not explicitly state future prospects. However, the paper suggests that the ARIMA method can be applied to many other fields where time series prediction is needed. The most promising areas for its application include business, social and environmental sectors, finance, and medicine.

4. **Forecasting models**

AR(p) MODEL (AutoRegressive Model of Order p)

In an autoregressive model, the future behaviour of a variable is predicted by considering a linear combination of its past values. Essentially, this means that the current state of the variable y(t) is determined as follows:

$y(t) = C + \Sigma(i=1 \text{ to } p)\ \varphi(i) * y(t-i) + \varepsilon(t)$ or $=f(y_{t-1}, y_{t-2}, \ldots, \varepsilon(t))$,

where "p" is a parameter that can vary in value, "C" represents a constant, φ(i) are coefficients that determine the influence of the previous p values of y(t), and ε(t) is the error term at time t.

MA(q) MODEL (Moving Average Model of Order q)

The moving average model is based on a function of error terms and uses early errors terms for prediction. It is expressed as:

$y(t) = \mu + \Sigma(i=1 \text{ to } q)\ \theta(i) * \varepsilon(t-i) + \varepsilon(t)$,

where y(t) represents the observed data point at time t, "μ" is the mean of the time series, “θ(i)” are coefficients that determine the influence of past q error terms ε(t-i), and ε(t) is the error term at time t. Fitting the MA model can be more challenging compared to the AR model due to the need to estimate the coefficients θ(i).
This model is harder than AR model to fit the time series because the errors are not foreseeable.

ARMA(p, q) MODEL (AutoRegressive Moving Average Model)

The ARMA model combines both autoregressive (AR) and moving average (MA) components. It is expressed as:

$y(t) = c + \Sigma(i=1 \text{ to } p)\ \varphi(i) * y(t-i) + \Sigma(j=1 \text{ to } q)\ \theta(j) * \varepsilon(t-j) + \varepsilon(t)$

In this model, $y(t)$ is the observed data point at time t, "c" is a constant, $\varphi(i)$ are autoregressive coefficients, $\theta(j)$ are moving average coefficients, and $\varepsilon(t)$ represents the error term at time t. The ARMA model allows for capturing both the autocorrelation of the time series (AR) and the influence of past error terms (MA).

These models are essential tools in time series analysis and forecasting, helping analysts understand and predict patterns in time-dependent data.

ARIMA MODEL (AutoRegressive Integrated Moving Average)

Compared to ARMA, the ARIMA model is suitable for handling non-stationary time series behaviour. This model transforms non-stationary data into a stationary format through differencing once or more times.

The ARIMA(p, d, q) model is an AutoRegressive Integrated Moving Average model, where "yt" is modelled as:

$yt = c + \varphi 1yt\text{-}1 + \varphi 2yt\text{-}2 + ... + \varphi pyt\text{-}p + \theta 1\varepsilon t\text{-}1 + \theta 2\varepsilon t\text{-}2 + \ldots + \theta q\varepsilon t + \varepsilon t$

$\hat{y}_t=\mu+\phi_1 y_{t-1}+\ldots+\phi_p y_{t-p}-\theta_1 e_{t-1}-\ldots-\theta_q e_{t-q}$

Here, "yt" represents a different time series.

The ARIMA model involves three parameters: "p" for the order of the AutoRegressive (AR) part, "d" for the degree of differencing, and "q" for the order of the Moving Average (MA) part.

The model combines three fundamental methods:

AR (AutoRegressive) Part: This part regresses "yt" with its own lagged values.

Differencing Part: Differencing transforms the non-stationary time series into a stationary model. For example, first-order differencing is expressed as D(i) = Data(i) - Data(i-1). The parameter "d" represents the number of times the differencing method was applied.

MA (Moving Average) Part: The MA part considers the number of lagged variables used in the model to account for the influence of past error terms.

In summary, ARIMA is a versatile model that combines these three components to handle non-stationary time series data effectively. The "Integrated" component (denoted as "I" for differencing) makes ARIMA suitable for transforming non-stationary data into a stationary format before applying ARMA modelling, especially when dealing with data exhibiting trends or seasonality. ARMA, in contrast, assumes that the data is already stationary and does not include the differencing component.

A. LSTM (The authors haven’t considered it).

Methodology

1. Stationarize the time-series → Use tests to check for stationarity → Use differencing
2. Plot ACF and PCF
3. ARIMA model (plot the data -> decomposition of the time series data into its essential constituents - > logarithmic format -> plot ACF and PCF -> Dickey Fuller Test -> plot ACF and PCF for differenced
4. Use the ARIMA(0,1,0) model. Auto arima function to calculate trace for two datas.
5. After finding the best model predict series for Nifty and Sensex -> L-Jung-Box test -> Get results that model is free from autocorrelation

5. **What was done?**

This work serves as an introduction to the concepts of forecasting and time series analysis, using the Indian economy's stock market as an example. The paper aims to apply the ARIMA (AutoRegressive Integrated Moving Average) model to efficiently predict the Nifty and Sensex indexes, which are Indian analogues of the S&P 500. The data for the analysis comprises publicly available time series data from the Indian stock market.
A comparison with actual results reveals an average deviation of 5% in mean percentage error for the analysed indexes. To validate these results, various tests can be employed. In this study, the authors use the Augmented Dickey-Fuller (ADF) test and the Ljung-Box tests for validation.

This paper introduces the concept of time series analysis and forecasting in the context of the Indian economy. The recent significant depreciation of the Indian rupee underscores the critical need for accurate stock market predictions to safeguard the interests of investors. This paper aims to build an efficient ARIMA model for predicting Indian stock market volatility. Publicly available time series data from the Indian stock market is utilised for this study. The predicted time series is compared with the actual time series, revealing an average deviation of roughly 5% in mean percentage error for both Nifty and Sensex.

Various tests can be employed for validating the predicted time series. However, in this study, we have utilised the 'ADF test' and the 'Ljung-Box tests' for validation purposes. We suggest that the ARIMA approach is sufficiently robust for handling time series data and can be highly valuable in diverse sectors, including healthcare, education, finance, and policymaking, for predictive purposes.

6. **Replicate the results**

The ARIMA model relies on the stationarity of the time series. To be suitable for ARIMA modelling, a time series should lack trend and seasonality, exhibiting constant variance and a constant mean over a given period. Various tests are employed to assess whether a time series is stationary or not. In this study, the authors used the 'Dickey-Fuller test' to verify the stationarity of the time series.

The AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) are statistical methods used to determine the parameters of the 'AR' (AutoRegressive) and 'MA' (Moving Average) terms in an ARIMA model. These methods help in understanding the relationships between observations in a time series.

The general form of ACF is as:

$$\frac{Covariance\,(Xt, Xt-h)}{Std.dev(Xt).Std.dev(Xt-h)} = \frac{Covariance\,(Xt, Xt-h)}{Variance(Xt)}$$

5 stages of prediction framework:

1. Time series analysis
2. Identify if time series is stationary or not
3. Estimation - plot ACF and PACF charts to look for parameters
4. ARIMA model
5. Forecast

L-Jung-Box test is for the validations that the residuals of the time series are following the random pattern, or if there is a certain degree