

# PREVISIONE SERIE STORICHE

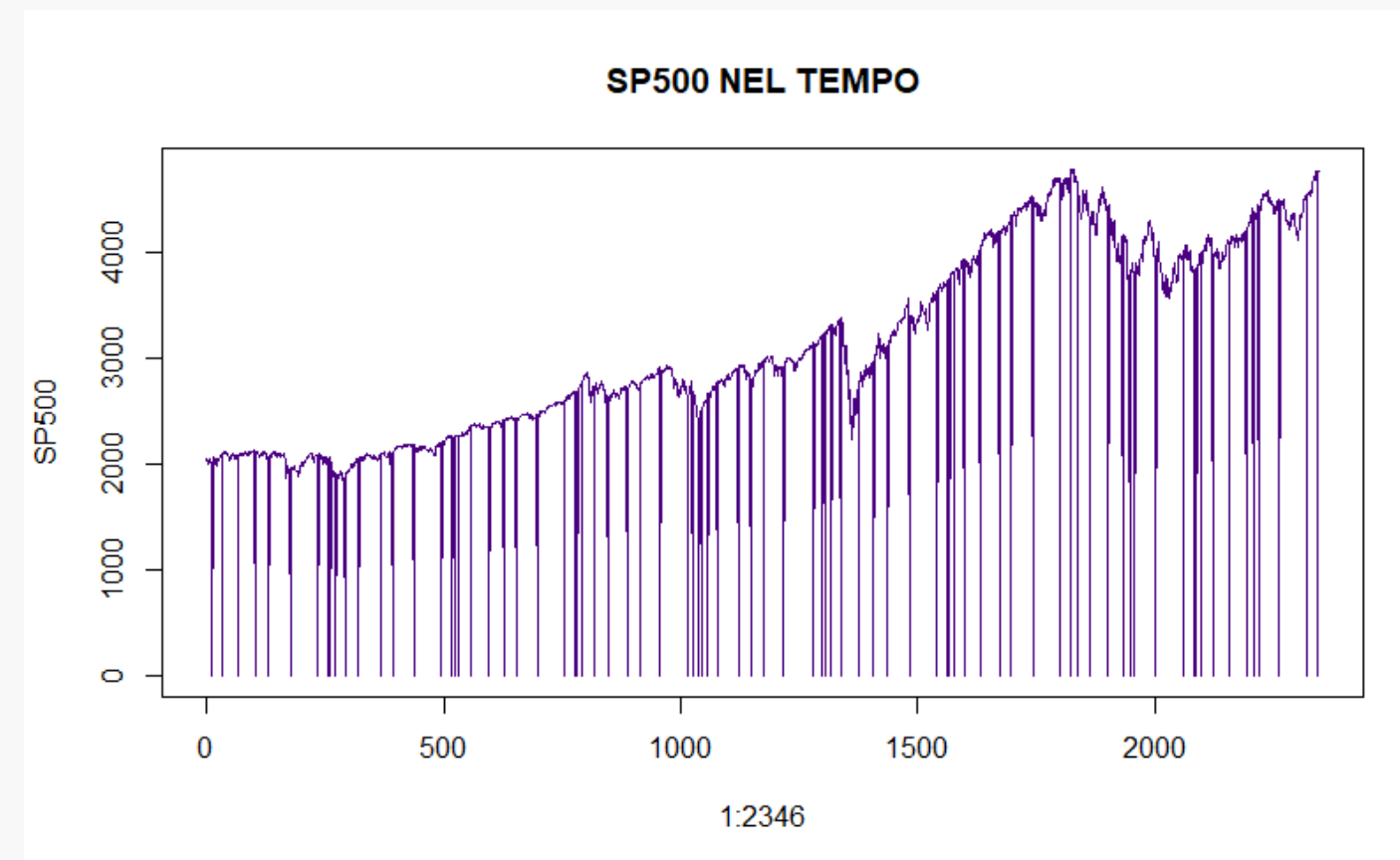
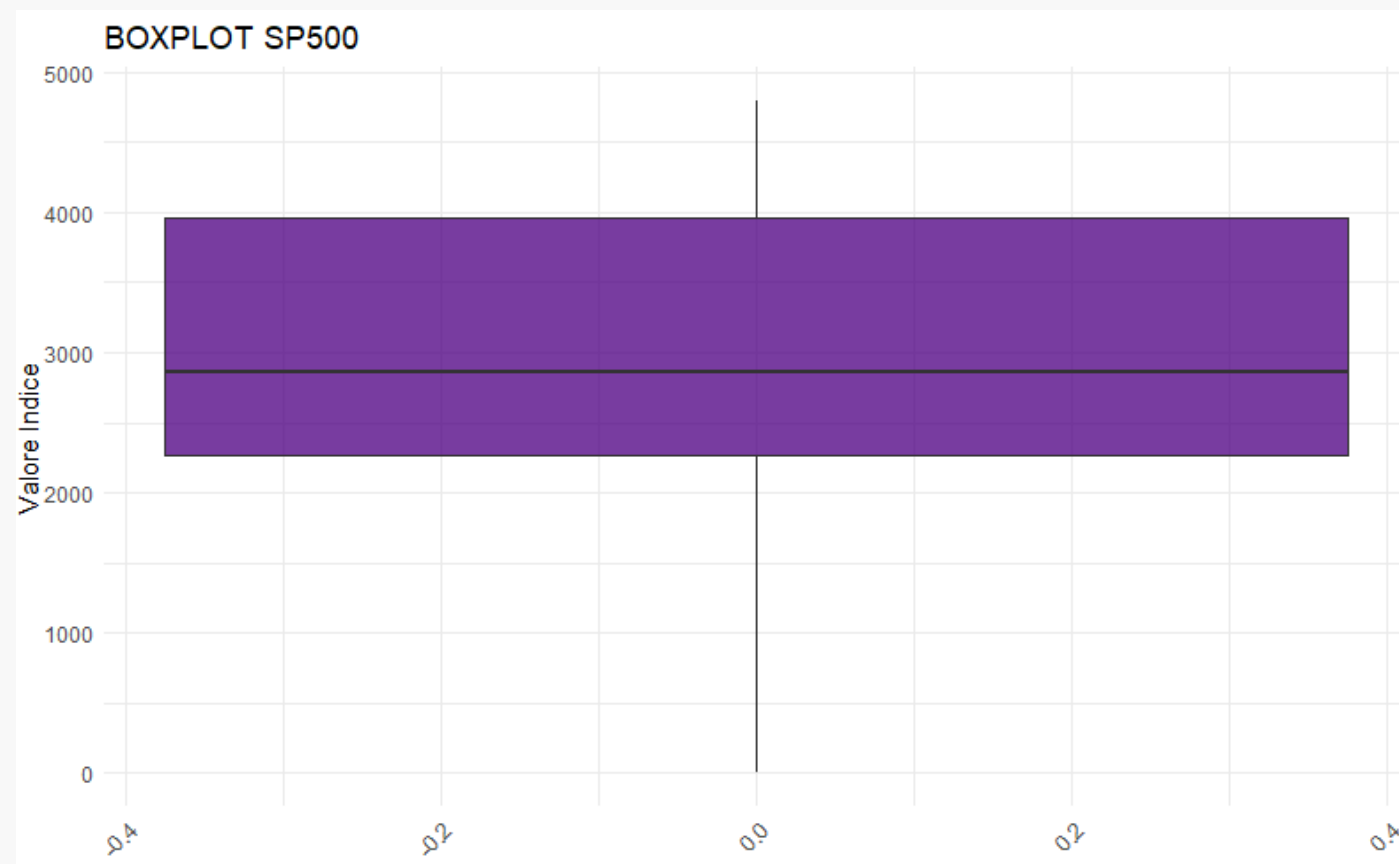


# OBIETTIVO

## Confronto Random Forest e AutoArima

- Applicazione **Random Forest** e **AutoArima** sulla stessa serie storica
- Validazione risultati attraverso **Walk-Forward Validation**
- Confronto **MSE** e **MAE**

# SERIE STORICA ANALIZZATA: S&P500



La serie  
rappresenta i  
massimi giornalieri  
dell'**S&P 500** dal  
**02/01/2015** al  
**29/12/2023**

# RIORGANIZZAZIONE DATI

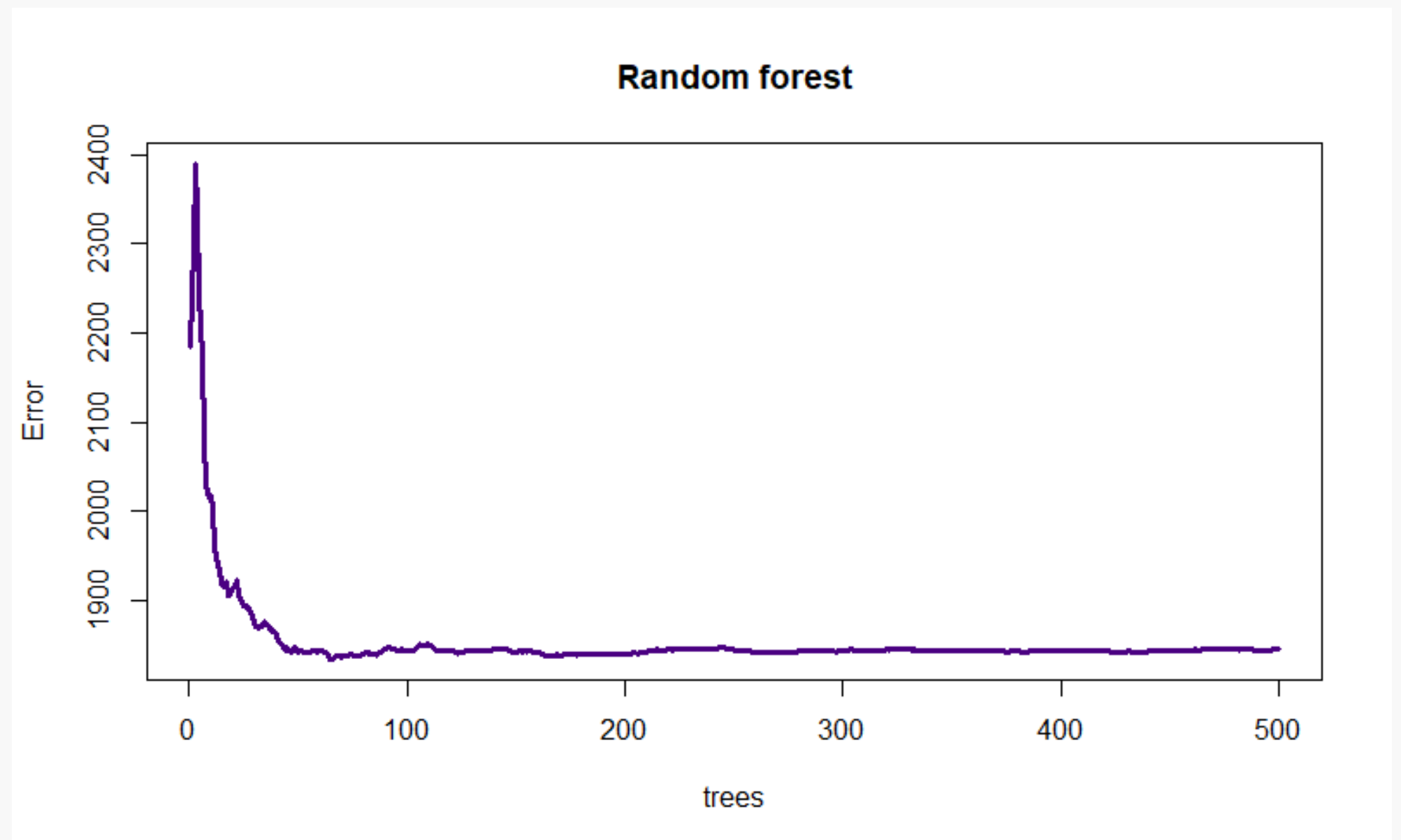
X	Y
2058.20	2020.58
2020.58	2002.61
2002.61	2025.90
2025.90	2062.14
2062.14	2044.81
2044.81	2028.26

Per essere adatti all'apprendimento supervisionato, riorganizzo i dati in un dataframe di due colonne: la prima colonna rappresenta la **serie storica con lag=1**, la seconda la **serie storica con lag=0**.

Cancellando le righe con valori NA, abbiamo un dataframe dove **nella riga i abbiamo il valore della serie al tempo i ed al tempo i+1**

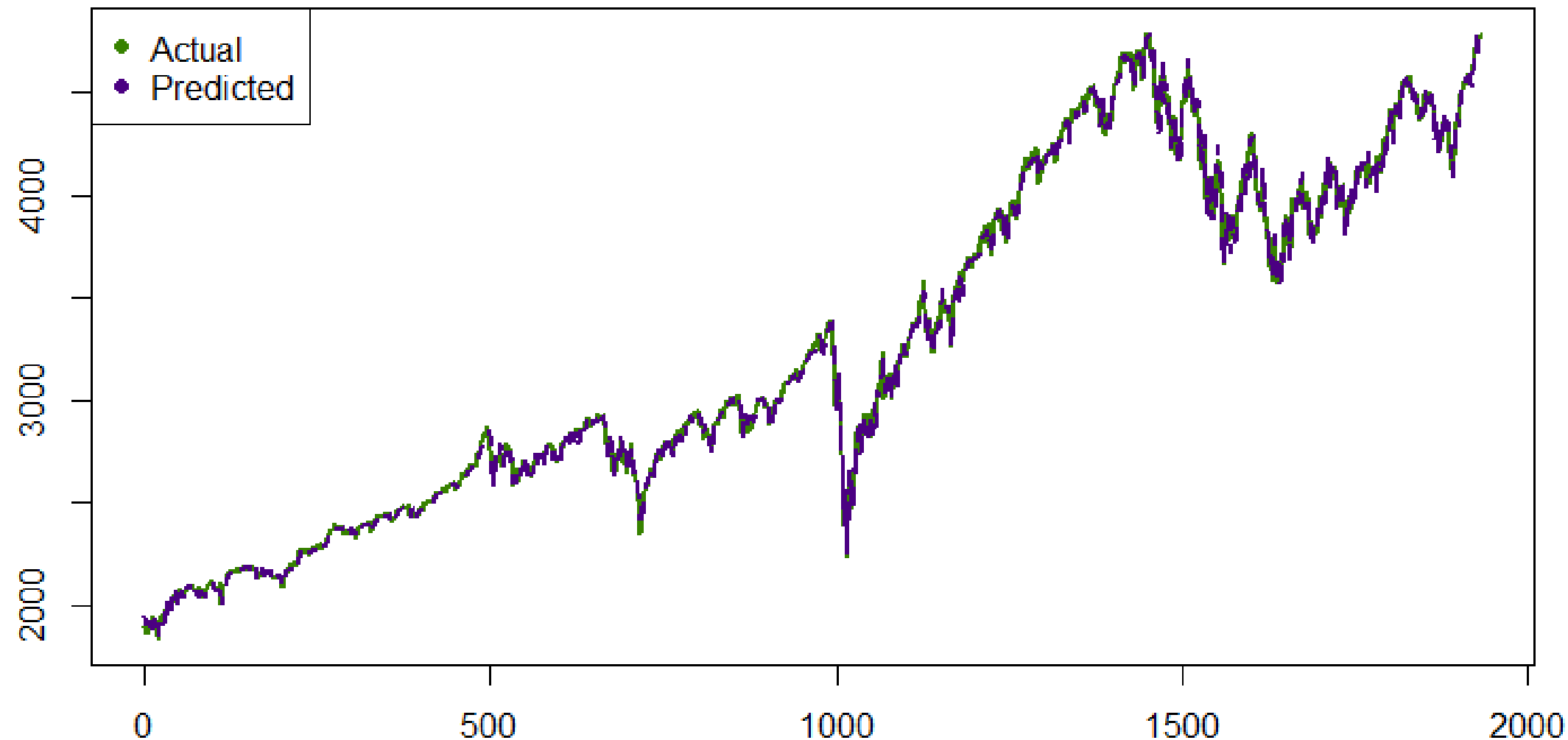
# RANDOM FOREST: SCELTA NTREE

Ho splittato la serie prendendo le prime **1930** osservazioni come **Train** e le ultime **251** come **Test**. Pare che da **ntree** uguale a **120** i valori dell'errore iniziano a **stabilizzarsi**.



# PREVISIONE E WALK FORWARD VALIDATION

Previsione con RF



## PERCHÈ NON LA CROSS VALIDATION?

L'ordine delle osservazioni in una serie storica è molto importante: prendere campioni casuali **non è adatto** in questo caso

## ALTERNATIVA: WALK FORWARD VALIDATION

Si sceglie una **finestra di train** minima (in questo caso di **249** osservazioni). Ad ogni passo si addestra con i valori della finestra, **si predice il passo appena successivo** alla finestra (che poi si confronterà con il valore originale) e, infine, **si aumenta la dimensione della finestra di 1**

## RISULTATI PREVISIONE

Si denota che la RF ha predetto **in maniera efficace** l'andamento della serie storica, nonostante delle leggere differenze

# MODELLI ARIMA: CENNI TEORICI

$$\phi(B)\nabla^d x_t = c + \theta(B)w_t$$

$$\nabla^d = (1 - B)^d$$

$$Bx_t = x_{t-1}$$

ARIMA è l'acronimo di **AutoRegressive Integrated Moving Average**:

## AR

Utilizza le relazioni tra le osservazioni passate per prevedere i valori futuri

$$x_t = \sum_{j=1}^p \phi_j x_{t-j} + w_t$$

## MA

Utilizza gli errori di previsione passati per migliorare le previsioni future.

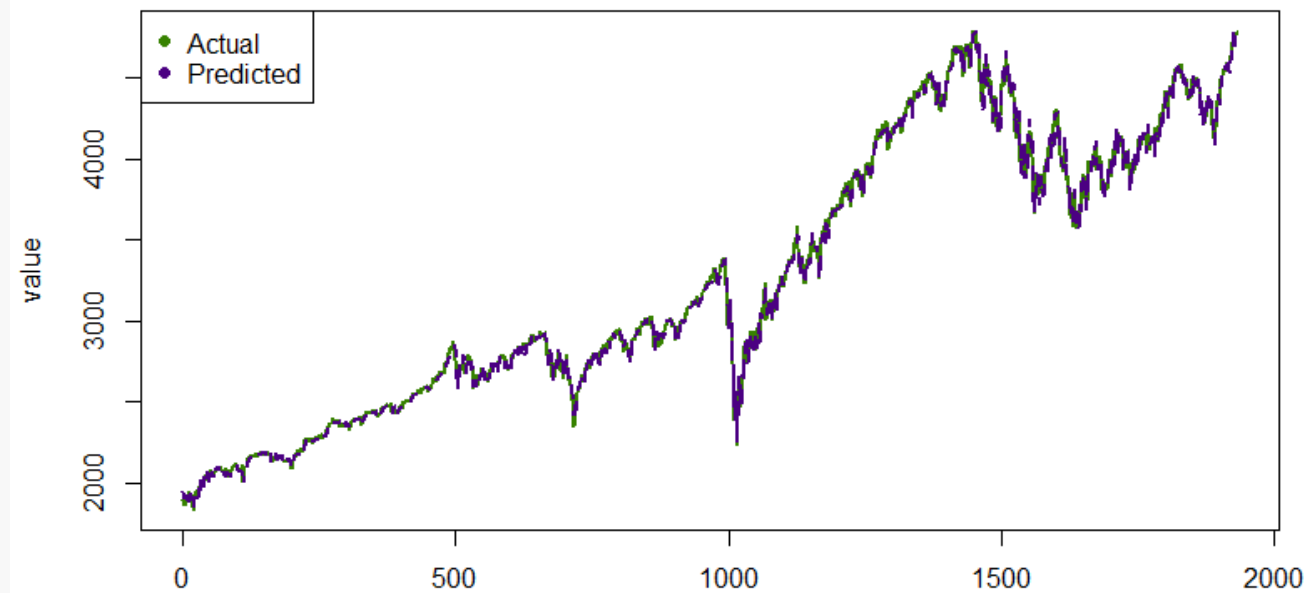
$$x_t = w_t + \sum_{j=1}^q \theta_j w_{t-j}$$

## I

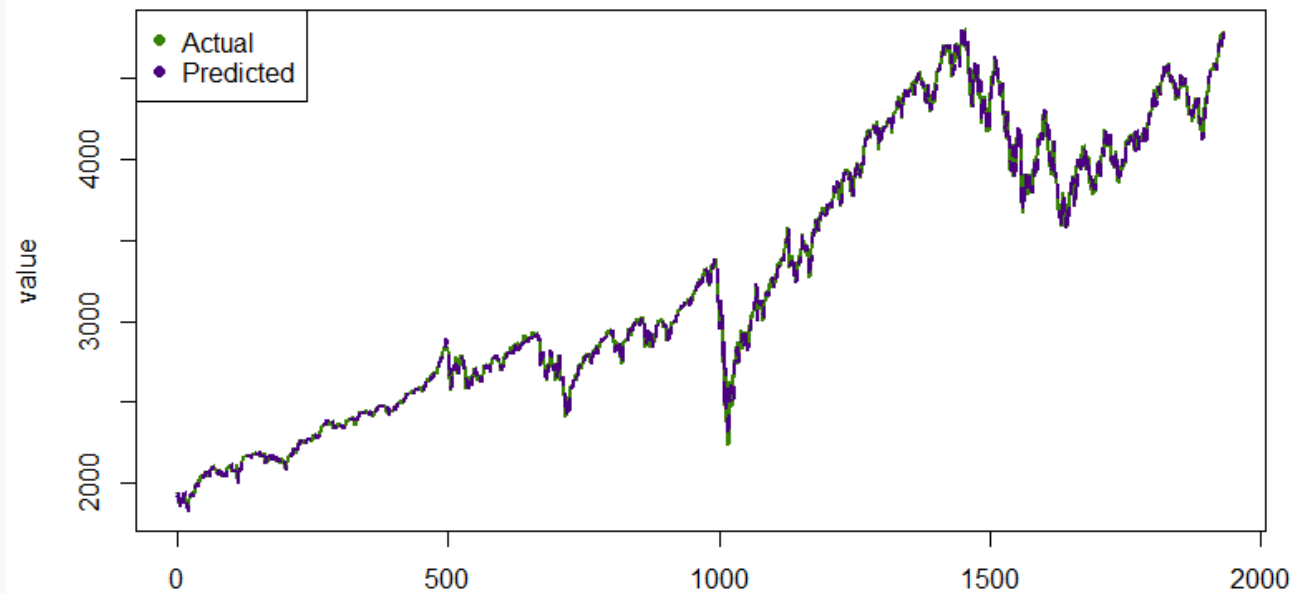
Involve la differenziazione dei dati per renderli stazionari, ovvero per stabilizzare la media della serie temporale.

# PREVISIONE CON AUTO-ARIMA E CONFRONTO RISULTATI

Previsione con RF



Previsione con Auto-Arima



Model	MSE	MAE
AutoARIMA	1533.206	25.57525
Random Forest	1868.673	29.18911



# FONTI:

[HTTPS://WWW.STAT.BERKELEY.EDU/~RYANTIBS/TIMESERIES-F23/LECTURES/ARIMA.PDF](https://www.stat.berkeley.edu/~ryantibs/timeseries-f23/lectures/arima.pdf)

[HTTPS://MACHINELEARNINGMASTERY.COM/RANDOM-FOREST-FOR-TIME-SERIES-FORECASTING/](https://machinelearningmastery.com/random-forest-for-time-series-forecasting/)

[HTTPS://MACHINELEARNINGMASTERY.COM/TIME-SERIES-FORECASTING-SUPERVISED-LEARNING/](https://machinelearningmastery.com/time-series-forecasting-supervised-learning/)

[HTTPS://MACHINELEARNINGMASTERY.COM/CONVERT-TIME-SERIES-SUPERVISED-LEARNING-PROBLEM-PYTHON/](https://machinelearningmastery.com/convert-time-series-supervised-learning-problem-python/)

[HTTPS://MACHINELEARNINGMASTERY.COM/BACKTEST-MACHINE-LEARNING-MODELS-TIME-SERIES-FORECASTING/](https://machinelearningmastery.com/backtest-machine-learning-models-time-series-forecasting/)

[HTTPS://WWW.MDPI.COM/1999-4893/10/4/114](https://www.mdpi.com/1999-4893/10/4/114)

