UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Scienze Fisiche, Informatiche e Matematiche
Corso di Laurea in Informatica

Tesi di Laurea Triennale

Applicazione di Modelli Previsionali Basati su Machine Learning Legati all'Ambito Finanziario

RELATORE
Prof. Riccardo Martoglia

LAUREANDO
Mattia Savoia

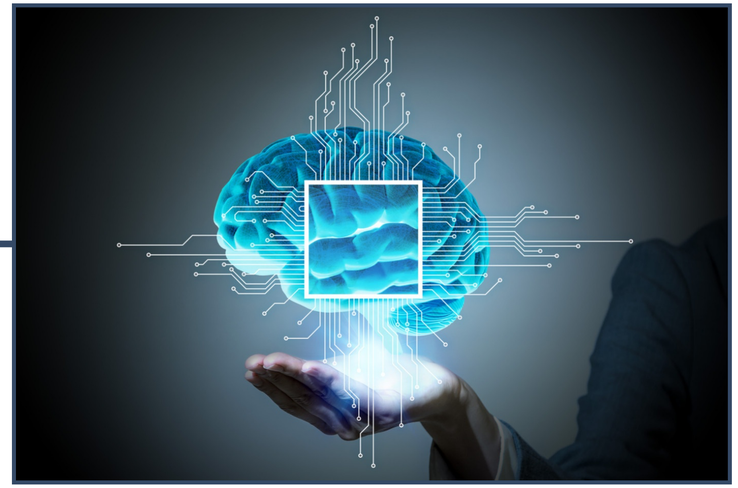
Anno Accademico 2019/2020

AMBITO DI RICERCA

Introduzione

Il trading in borsa ha guadagnato un’enorme popolarità a livello globale

Molte società stanno studiando metodi per prevedere l’andamento dei mercati



PREVISIONE DEL MERCATO FINANZIARIO

Introduzione

Sono a disposizione **grandi quantità di dati** finanziari che vengono generati ogni giorno

Problemi:

- Le **informazioni** relative ad un mercato sono **complesse e incerte**
- Il mercato è **instabile** e **imprevedibile**

Molto **complicato** eseguire previsioni accurate

OBIETTIVI DEL PROGETTO

Introduzione

1. Acquisizione dello **storico dei dati** degli indici finanziari

STANDARD &POOR'S 500 Nasdaq DOW JONES

2. Progettazione e implementazione di modelli di ML per la **previsione** di serie temporali finanziarie

3. **Valutazione** dei modelli previsionali

4. **Confronto** dei modelli previsionali

Indice del progetto

1 Tecnologie utilizzate

2 Progettazione

3 Valutazione dei modelli

4 Confronto dei modelli

5 Conclusioni e sviluppi futuri

TECNOLOGIE

Tecnologie Utilizzate

Linguaggio

python™

DataFrame

yahoo! finance

pandas

plotly

matplotlib

Modelli

statsmodels

Keras

pmdarima

NumPy

TensorFlow

scikit learn

Indice del progetto

1 Tecnologie utilizzate

2 Progettazione

3 Valutazione dei modelli

4 Confronto dei modelli

5 Conclusioni e sviluppi futuri

CREAZIONE E VISUALIZZAZIONE DEL DATAFRAME

Progettazione

Sorgente dei dati

yahoo! finance

Pandas

S&P 500

Date	Open	High	Low	Close	Adj_close	Volume
1990-01-02	353.3999938964844	359.69000244140625	351.9800109863281	359.69000244140625	359.69000244140625	162070000
1990-01-03	359.69000244140625	360.5899963378906	357.8900146484375	358.760009765625	358.760009765625	192330000
1990-01-04	358.760009765625	358.760009765625	352.8900146484375	355.6700134277344	355.6700134277344	177000000
1990-01-05	355.6700134277344	355.6700134277344	351.3500061035156	352.20001220703125	352.20001220703125	158530000
1990-01-08	352.20001220703125	354.239990234375	350.5400085449219	353.7900085449219	353.7900085449219	140110000

Visualizzazione

MODELLI DI PREVISIONE

Progettazione

Obiettivo: calcolare previsioni dell'andamento di mercati finanziari

Modelli utilizzati:

- Auto-Regressive Integrated Moving Average (**ARIMA**)
- Long Short-Term Memory (**LSTM**)

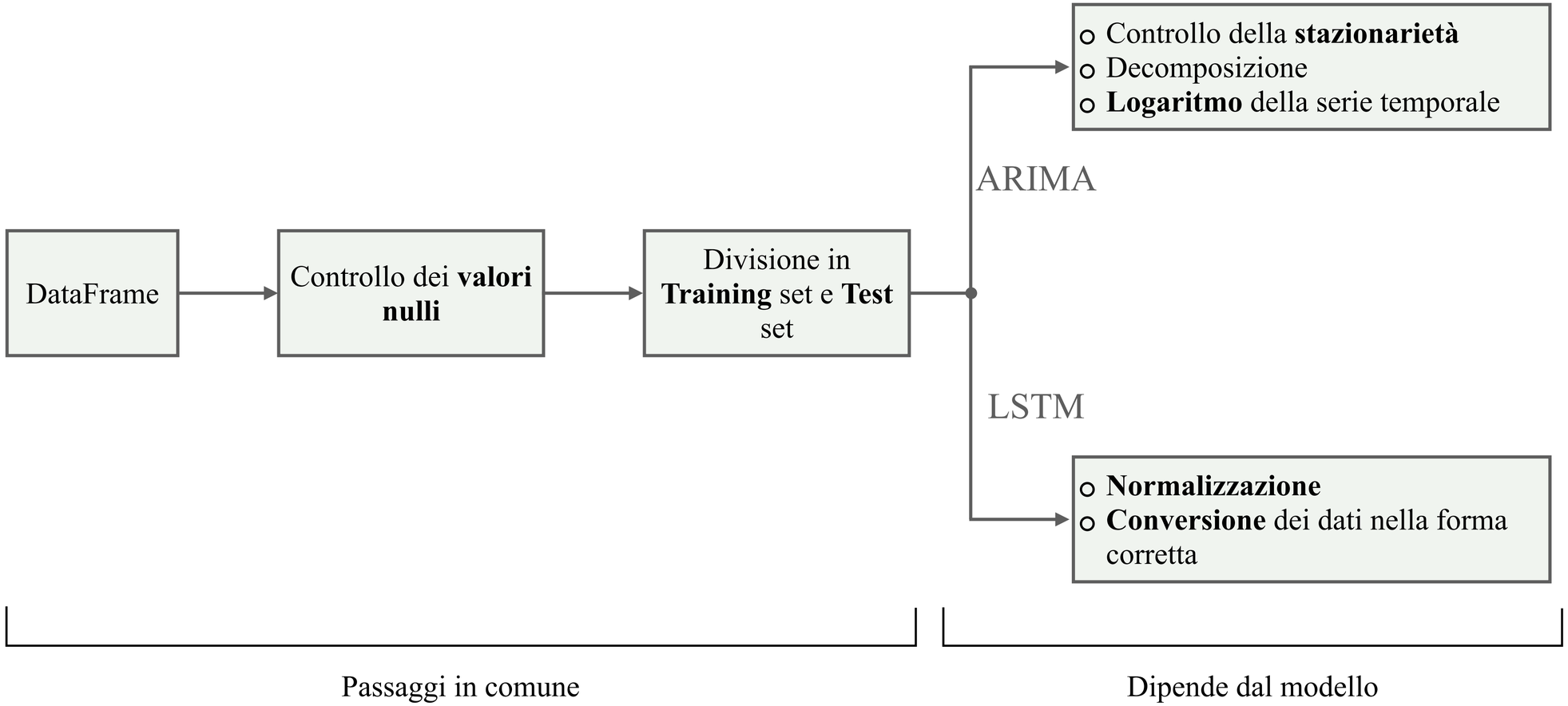
Motivi:

- Due tecniche **molto utilizzate** nel campo economico e finanziario
- In generale raggiungono una **buona accuratezza di previsione**

PREPROCESSING DEI DATI

Progettazione

Tecnica **fondamentale** di preparazione, organizzazione e pulizia dei dati per renderli adatti alla costruzione e all'utilizzo dei modelli.



CREAZIONE DEI MODELLI

Progettazione

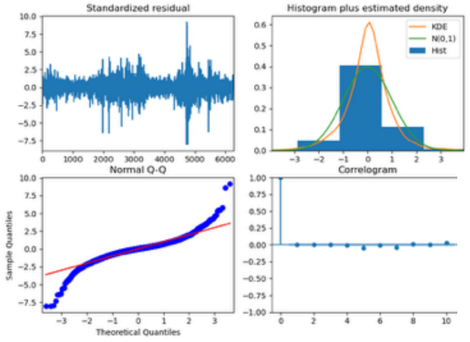
1 Impostazione degli **iperparametri**

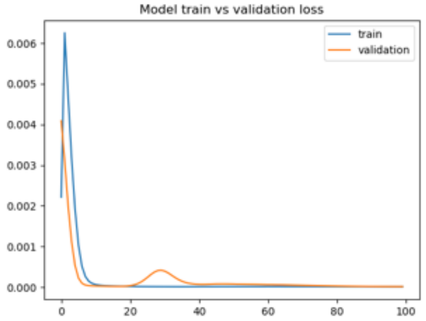
2 **Adattamento (fit)** del modello

```
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=-38295.696, Time=1.28 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=-38314.562, Time=0.89 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=-38316.474, Time=2.05 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=-38293.995, Time=0.48 sec
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=-38324.577, Time=2.02 sec
 ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=-38323.149, Time=1.37 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=-38323.305, Time=2.08 sec
                                    : AIC=-38325.314, Time=1.74 sec
                                    : AIC=-38323.272, Time=2.58 sec
                                    : AIC=-38321.188, Time=1.86 sec
                                    : AIC=-38322.703, Time=0.56 sec

                                    ercept
```

```
Model: "RNN_model"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm (LSTM)                  (None, 100)               40800
_________________________________________________________________
dropout (Dropout)            (None, 100)               0
_________________________________________________________________
dense (Dense)                (None, 1)                 101
=================================================================
Total params: 40,901
Trainable params: 40,901
Non-trainable params: 0
_________________________________________________________________
```





3 **Goodness-of-fit**

4 **Previsione**

Indice del progetto

1 Tecnologie utilizzate

2 Progettazione

3 Valutazione dei modelli

4 Confronto dei modelli

5 Conclusioni e sviluppi futuri

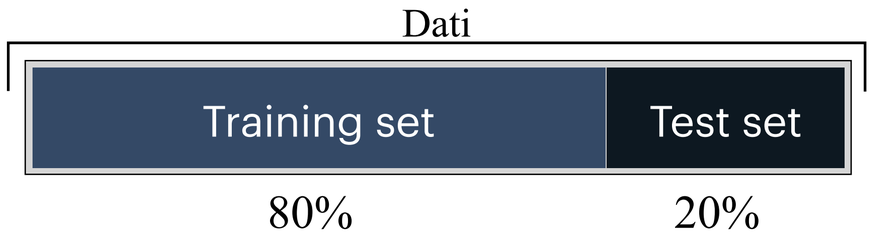
PREVISIONE

Valutazione dei modelli

L'obiettivo è quello di prevedere il **prezzo di chiusura del giorno successivo** in riferimento ai vari indici considerati

Metodologia di valutazione:

- Allenamento del modello sul training set
- Valutazione della previsione **one-step** sul test set



ARIMA

Valutazione dei modelli

Reale

Previsione

S&P 500

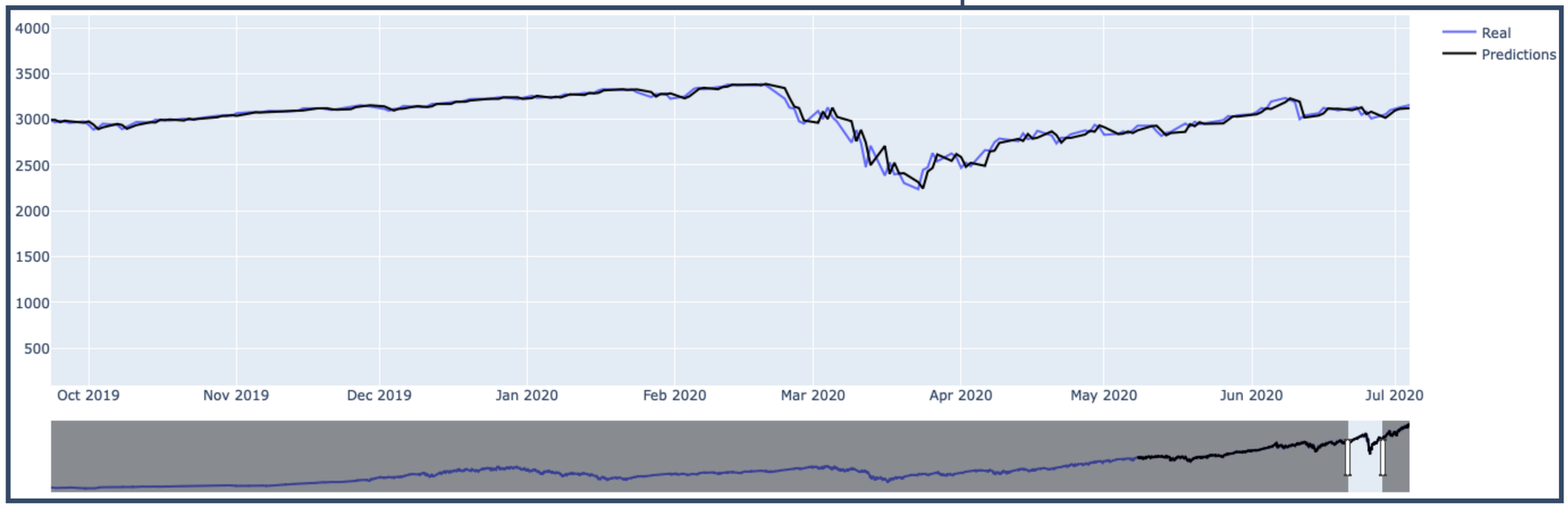
Dow Jones Industrial

NASDAQ

ARIMA

Valutazione dei modelli

S&P 500

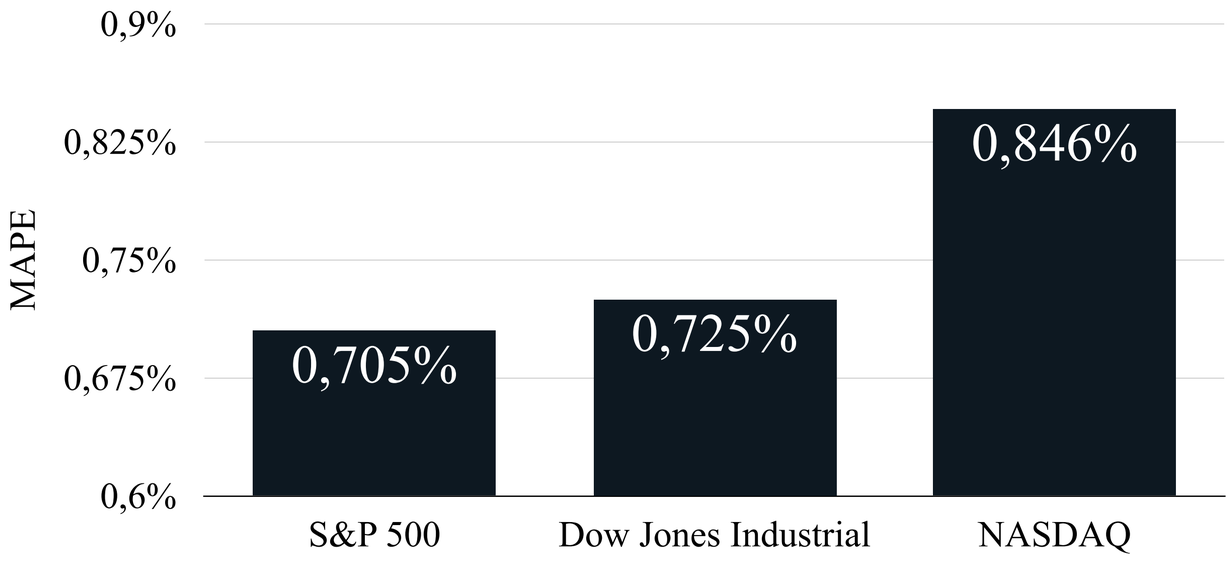


ARIMA

Valutazione dei modelli

Valutazione ARIMA:

- Metrica MAPE per calcolare l'errore medio in percentuale dei tre indici

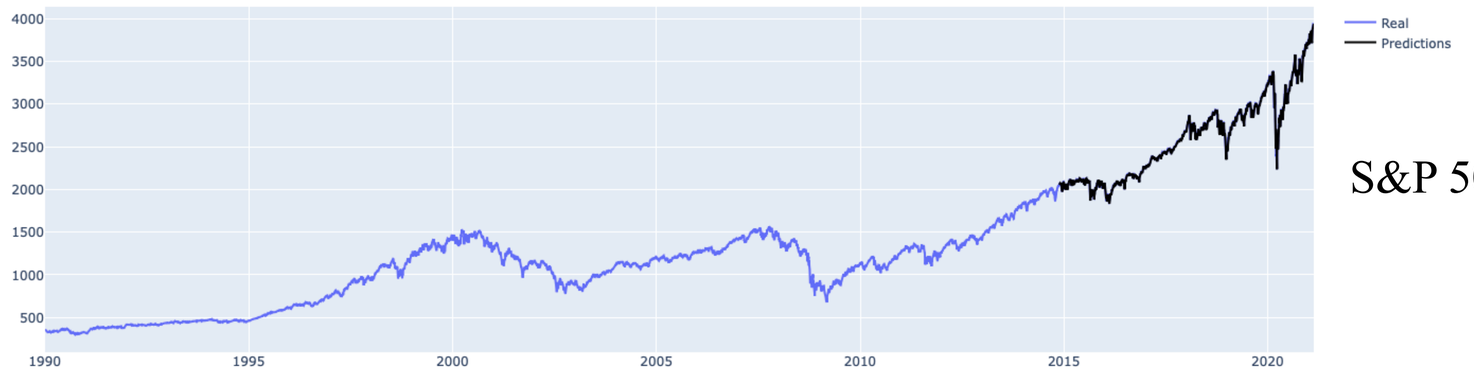


LSTM

Valutazione dei modelli

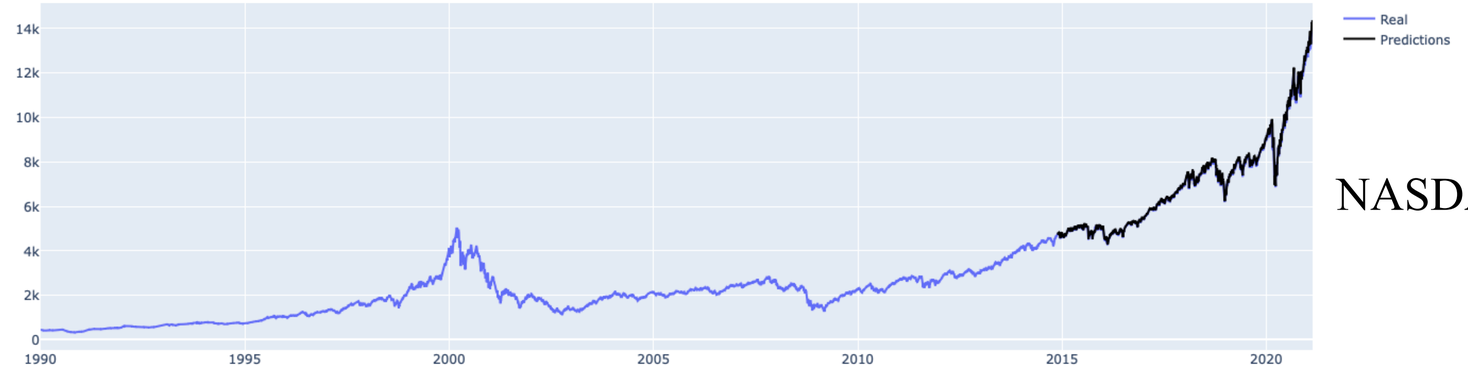
Reale

Previsione



S&P 500

Dow Jones Industrial



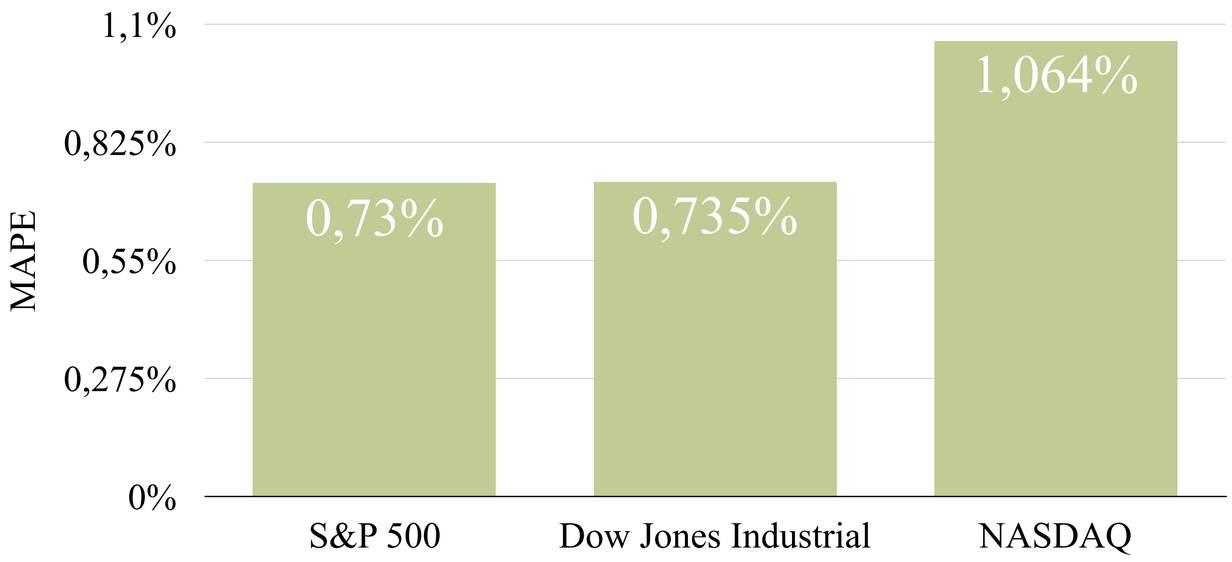
NASDAQ

LSTM

Valutazione dei modelli

Valutazione LSTM:

- Metrica MAPE per calcolare l'errore medio in percentuale dei tre indici



Indice del progetto

1 Tecnologie utilizzate

2 Progettazione

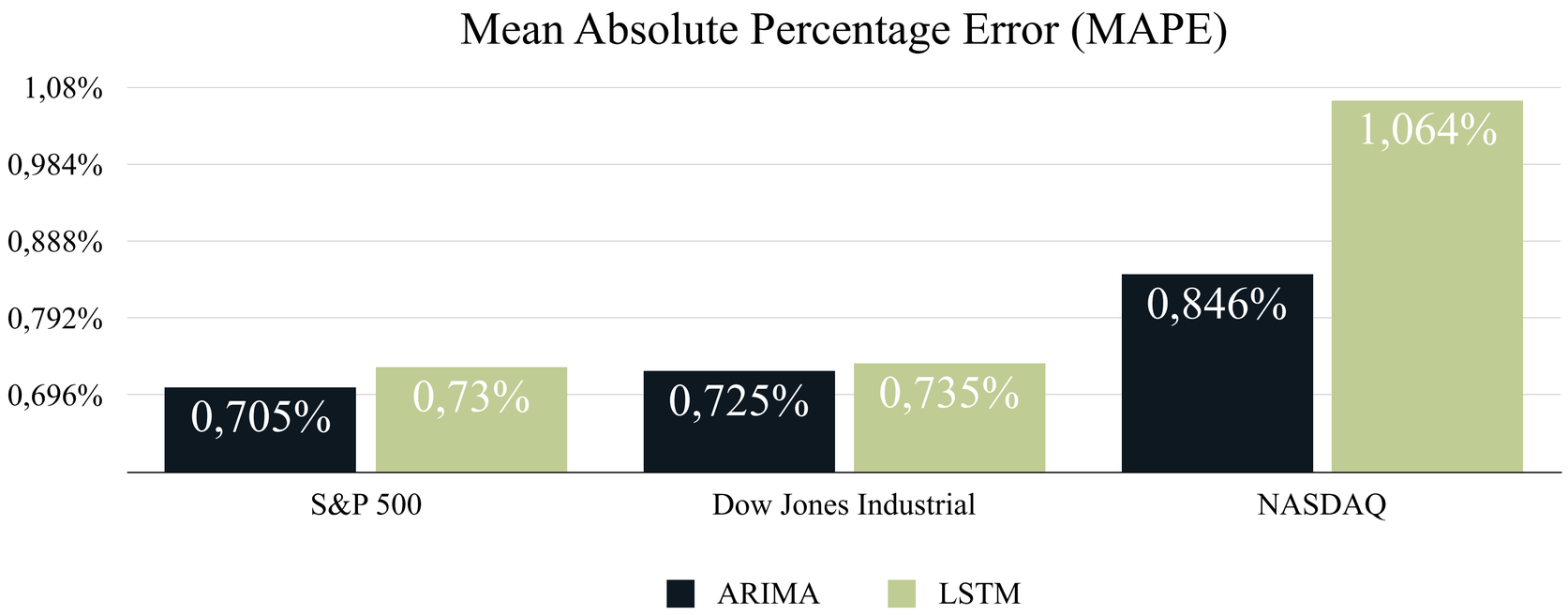
3 Valutazione dei modelli

4 Confronto dei modelli

5 Conclusioni e sviluppi futuri

CONFRONTO

Confronto dei modelli



Considerazioni:

- **Confronto con altre ricerche**: buona accuratezza di previsione
- Non è facile riuscire a confrontare i risultati ottenuti con risultati derivanti da altre prove sperimentali

Indice del progetto

1 Tecnologie utilizzate

2 Progettazione

3 Valutazione dei modelli

4 Confronto dei modelli

5 Conclusioni e sviluppi futuri

CONCLUSIONI

Conclusioni e sviluppi futuri

Risultati raggiunti:

- Le prestazioni di ARIMA sono migliori rispetto a quelle di LSTM
- La differenza di errore tra i due modelli non è alta
- I risultati sono in linea con quelli dello stato dell'arte

Sviluppi futuri:

- Miglioramento dei modelli
- Sviluppo di una "Sentiment Analysis"

Grazie per l’attenzione