

# Data Science Lab: Ristoranti

Federico Bidone 892054

30 giugno 2023

## Sommario

L'Obiettivo del progetto è fare una previsione del fatturato di 6 ristoranti con serie storiche di diversa lunghezza, ripulendo i dati, arricchendoli, modellandoli e aggregandoli per migliorare la previsione. Fare una previsione annuale del fatturato può essere molto utile per un ristorante per diversi motivi:

- **Pianificazione finanziaria:** Conoscere la previsione del fatturato per l'anno successivo può aiutare il ristorante a pianificare le sue finanze, ad esempio per stabilire un budget per le spese, per investimenti o per il pagamento delle tasse.
- **Gestione delle scorte:** Una previsione accurata del fatturato può aiutare il ristorante a gestire meglio le sue scorte, ad esempio per evitare sprechi o carenze di cibo e bevande.
- **Pianificazione del personale:** Conoscere la previsione del fatturato può aiutare il ristorante a pianificare il suo personale, ad esempio per assumere più personale durante i periodi di maggiore affluenza o per ridurre il personale durante i periodi di minore affluenza.
- **Marketing e promozioni:** Una previsione accurata del fatturato può aiutare il ristorante a pianificare le sue attività di marketing e promozioni, ad esempio per offrire sconti o promozioni durante i periodi di minore affluenza per attirare più clienti.

In sintesi, fare una previsione annuale del fatturato può fornire al ristorante informazioni preziose che possono aiutare a prendere decisioni informate e a gestire meglio il suo business.

# Indice

<b>Indice</b>	<b>2</b>
0.1 Introduzione . . . . .	3
0.2 Obiettivo . . . . .	3
0.3 Aspetti metodologici . . . . .	3
0.4 Lettura e preparazione dei dati . . . . .	5
0.5 Analisi . . . . .	5
0.6 Conclusioni . . . . .	11

## 0.1 Introduzione

Il dataset "ristoranti" consiste nei dati giornalieri del fatturato e del numero di scontrini di 6 ristoranti, dal 01/01/2018 al 31/05/2023. Sono stati aggiunti 4 dataset "meteo" ricavati dai dati forniti dall'ARPA (Agenzia regionale per la protezione ambientale) contenenti i dati giornalieri di precipitazioni e temperature medie giornaliere, ed il dataset "SerieA" fornito da <https://www.football-data.co.uk/> contenente il numero di partite giornaliere di Serie A. Questi dataset sono stati usati come regressori esterni per cercare di migliorare la previsione. Sono stati testati due modelli, SARIMAX e Prophet, con un approccio di tipo grid search per la ricerca degli iperparametri ottimali, mentre è stato preso l'RMSE normalizzato come metrica dell'errore.

### Punti Principali

- SARIMAX: un modello di serie temporali che utilizza la regressione con media mobile integrata autoregressiva stagionale (Seasonal Autoregressive Integrated Moving Average with eXogenous regressors).
- Prophet: un modello di serie temporali sviluppato da Facebook che utilizza i regressori esogeni per fare previsioni.
- Previsione: le previsioni effettuate dai modelli sono caratterizzate da un ruolo primario all'interno del progetto.

## 0.2 Obiettivo

L'obiettivo del progetto è quello di utilizzare i modelli SARIMAX e Prophet per prevedere le vendite dei ristoranti. Conoscere la previsione relativa alle vendite potrebbe essere utile ai gestori dei ristoranti per organizzare la disponibilità di magazzino e di spedizione.

## 0.3 Aspetti metodologici

Il codice utilizza due modelli di serie temporali per fare previsioni sulle vendite dei ristoranti: SARIMAX e Prophet.

### SARIMAX

Il modello SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous regressors) è un'estensione del modello ARIMA che include la stagionalità e i regressori esogeni. Un modello ARIMA è un modello di serie temporali che utilizza la media mobile integrata autoregressiva (ARIMA) per fare previsioni.

Un modello SARIMAX ha tre componenti principali: la componente autoregressiva (AR), la componente di integrazione (I) e la componente a media mobile (MA).

La componente AR modella la correlazione tra un'osservazione e un certo numero di osservazioni precedenti. In altre parole, questa componente cerca di catturare l'effetto che le osservazioni passate hanno sull'osservazione corrente.

L'ordine della componente AR, indicato con il parametro  $p$ , rappresenta il numero di osservazioni passate che vengono utilizzate per prevedere l'osservazione corrente.

La componente I differenzia i dati per rendere la serie temporale stazionaria. La stazionarietà è una proprietà importante delle serie temporali perché molti modelli di previsione assumono che la serie temporale sia stazionaria. La differenziazione consiste nel calcolare la differenza tra un'osservazione e l'osservazione precedente. L'ordine della differenziazione, indicato con il parametro  $d$ , rappresenta il numero di volte che i dati vengono differenziati.

La componente MA modella l'errore della previsione come una combinazione lineare degli errori di previsione passati. In altre parole, questa componente cerca di catturare l'effetto che gli errori di previsione passati hanno sull'errore di previsione corrente. L'ordine della componente MA, indicato con il parametro  $q$ , rappresenta il numero di errori di previsione passati che vengono utilizzati per prevedere l'errore di previsione corrente.

Il modello SARIMAX estende il modello ARIMA includendo una componente stagionale e i regressori esogeni. La componente stagionale modella la correlazione tra un'osservazione e le osservazioni precedenti a intervalli regolari, come ad esempio ogni anno o ogni mese. Questa componente cerca di catturare gli effetti stagionali presenti nella serie temporale. I regressori esogeni sono variabili esterne alla serie temporale che possono influenzare le previsioni, come

ad esempio le condizioni meteorologiche o le festività.

Un modello SARIMAX ha sette parametri:  $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ ,  $Q$  e  $m$ . I primi tre parametri rappresentano l'ordine della componente AR, l'ordine della componente I e l'ordine della componente MA. I parametri  $P$ ,  $D$  e  $Q$  rappresentano l'ordine della componente stagionale AR, l'ordine della componente stagionale I e l'ordine della componente stagionale MA. Il parametro  $m$  rappresenta il periodo della stagionalità, ad esempio 12 per i dati mensili o 4 per i dati trimestrali.

## Prophet

Il modello Prophet è un modello di serie temporali sviluppato da Facebook per fare previsioni su serie storiche con forti componenti stagionali. Prophet è una libreria open-source per la previsione di serie temporali univariate (una sola variabile). Implementa quello che viene definito come un modello di previsione di serie temporali additivo e l'implementazione supporta tendenze, stagionalità e festività.

Prophet può essere considerato un modello di regressione non lineare, dove la funzione di previsione è composta da quattro componenti principali: una curva di tendenza lineare a tratti o logistica, una componente stagionale annuale modellata utilizzando una serie di Fourier, una componente stagionale settimanale e una componente per catturare gli effetti delle festività.

Il modello è progettato per essere facile da usare e completamente automa-

tico, ovvero può essere utilizzato per fare previsioni su una serie temporale senza dover specificare manualmente i parametri del modello. In questo modo, è possibile utilizzare Prophet per fare previsioni su dati interni all'azienda, come ad esempio le vendite o la capacità produttiva.

La curva di tendenza lineare a tratti o logistica cattura la tendenza generale della serie temporale. Prophet rileva automaticamente i cambiamenti nella tendenza selezionando i punti di cambiamento dai dati. La componente stagionale annuale modella i cambiamenti stagionali che si verificano ogni anno, mentre la componente stagionale settimanale modella i cambiamenti stagionali che si verificano ogni settimana. La componente per le festività cattura gli effetti delle festività sulla serie temporale.

rappresentano il fatturato giornaliero di ciascun ristorante. Si nota che mancano alcuni dati e che le serie storiche hanno lunghezze diverse. Inoltre, all'inizio della serie ci sono alcune mensilità con solo valori 0 tranne un giorno con un valore anomalo, il che indica che i dati sono stati campionati a livello mensile. Si può quindi dividere il dataframe in 6 dataframe diversi, uno per ogni ristorante, riempire i valori mancanti con 0 e unire ogni dataframe dei ristoranti con il corrispondente dataframe meteo e con il dataframe SerieA. Si ottengono così 6 dataframe distinti con la data, l'ID del ristorante (con i dati sul fatturato) e le colonne: 'partite giornaliere', 'precipitazioni(mm)', 'temperatura(°C)'. Si può considerare la lunghezza massima di ogni serie storica in cui sono presenti i dati su tutte e 4 le colonne e ricampionare i dati a livello mensile.

## 0.4 Lettura e preparazione dei dati

Il dataset dei ristoranti è un file CSV in cui le date sono rappresentate come numeri interi nel formato "YYYYMMDD". Per far sì che pandas le interpreti correttamente, è necessario convertirle in date. Durante l'esplorazione dei dati, si nota che ogni data è associata a più ristoranti e quindi ripetuta più volte. Per risolvere questo problema, si può fare un pivot della tabella eliminando anche la colonna "scontrini". Si ottiene così un dataframe con una colonna data e 6 colonne (R000,R001,R002,R003,R004,R005) che

## 0.5 Analisi

Nel codice vengono utilizzati due modelli di serie temporali, SARIMAX e Prophet, per fare previsioni sulle vendite dei ristoranti. Entrambi i modelli vengono addestrati sui dati di training e valutati sui dati di test. Si definiscono dei range di iperparametri e il codice prova ogni combinazione di iperparametri nel range per trovare quella ottimale. Ogni serie temporale viene divisa in due sezioni, "train" e "test", corrispondenti rispettivamente al primo 80% e al restante 20% del dataset. Il codice testa ogni combinazione di iperparametri per addestrare i modelli sulla parte di train, poi fa una

previsione nel futuro corrispondente alla parte di test. A questo punto viene calcolato l'RMSE per ogni combinazione e viene stabilita quella migliore (con l'RMSE più basso). Una volta stabilito il modello migliore, viene riaddestrata tutta la serie storica e viene fatta una previsione nel futuro di un anno.

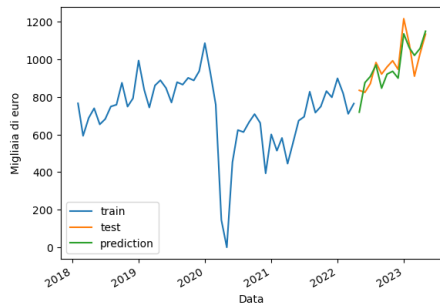


Figura 0.1: Confronto tra il forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R000

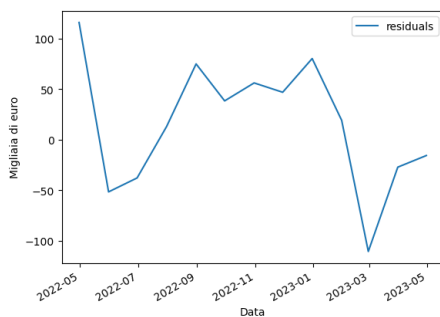


Figura 0.2: Plot dei residui del forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R000

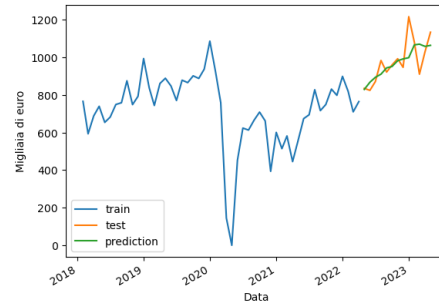


Figura 0.3: Confronto tra il forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R000

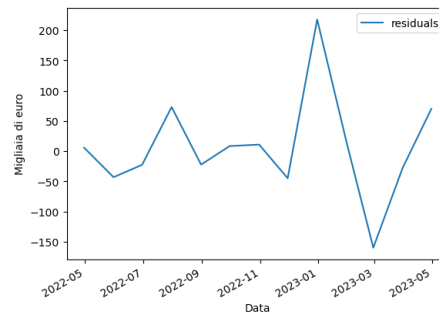


Figura 0.4: Plot dei residui del forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R000

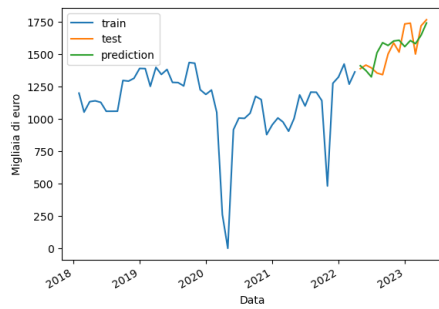


Figura 0.5: Confronto tra il forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R001

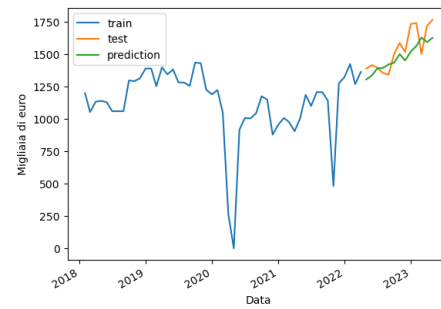


Figura 0.7: Confronto tra il forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R001



Figura 0.6: Plot dei residui del forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R001

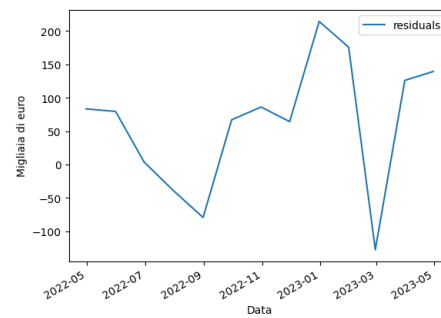


Figura 0.8: Plot dei residui del forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R001

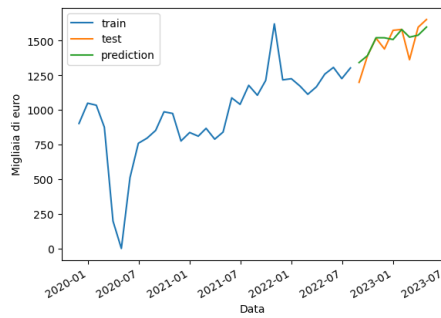


Figura 0.9: Confronto tra il forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R002

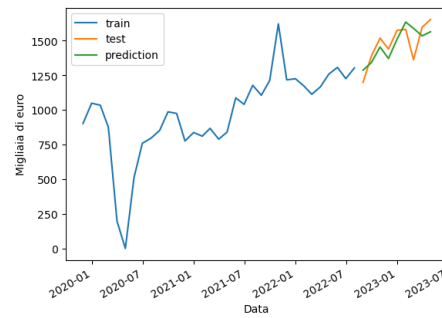


Figura 0.11: Confronto tra il forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R002

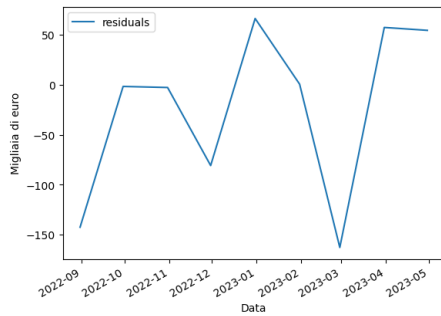


Figura 0.10: Plot dei residui del forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R002

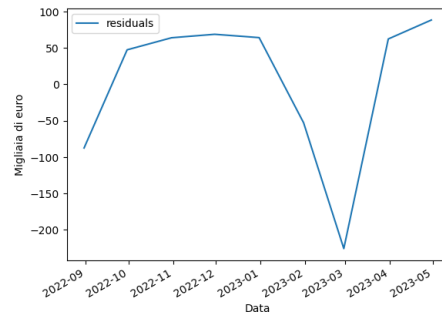


Figura 0.12: Plot dei residui del forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R002



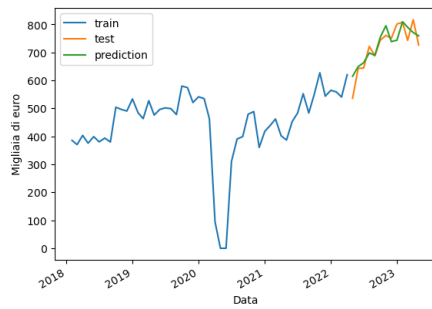


Figura 0.13: Confronto tra il forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R003

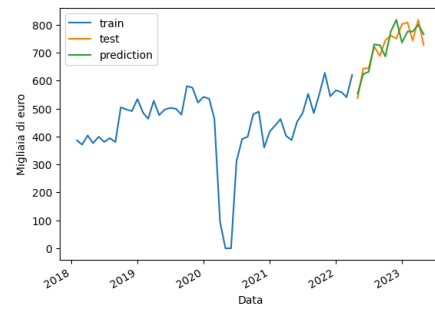


Figura 0.15: Confronto tra il forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R003

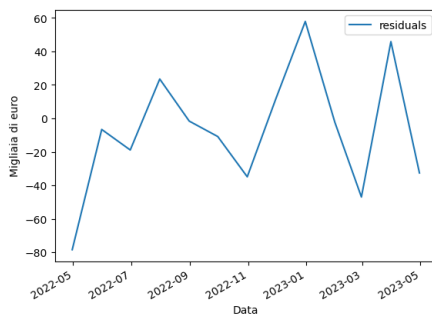


Figura 0.14: Plot dei residui del forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R003

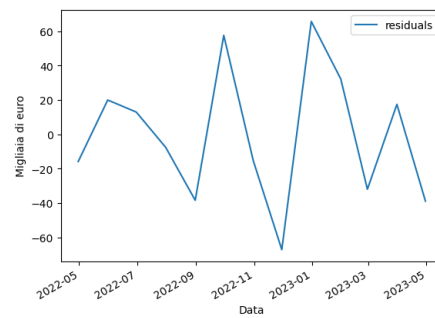


Figura 0.16: Plot dei residui del forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R003

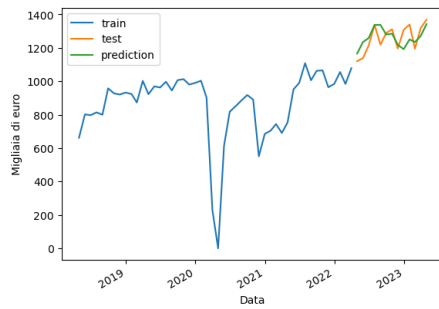


Figura 0.17: Confronto tra il forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R004

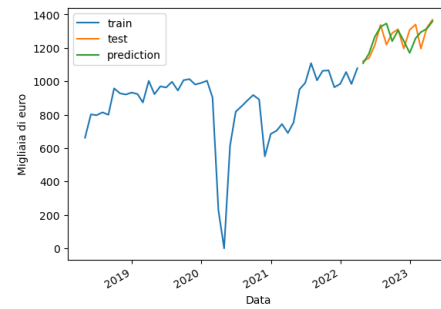


Figura 0.19: Confronto tra il forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R004

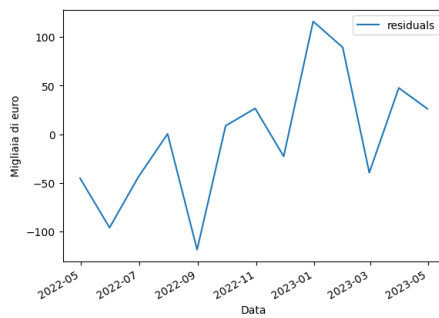


Figura 0.18: Plot dei residui del forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R004

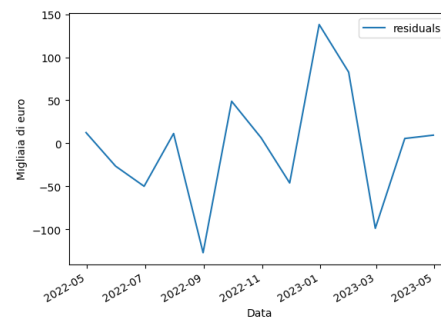


Figura 0.20: Plot dei residui del forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R004

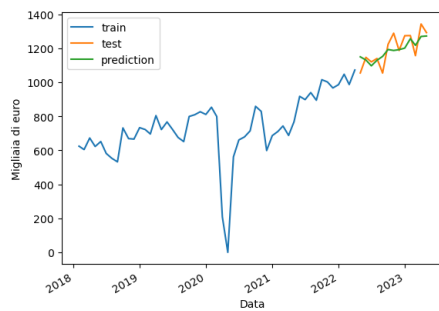


Figura 0.21: Confronto tra il forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R005

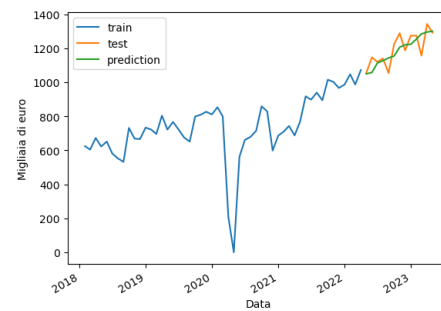


Figura 0.23: Confronto tra il forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R005

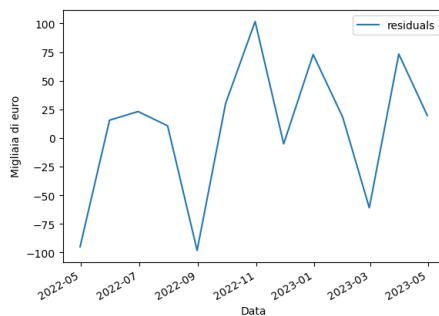


Figura 0.22: Plot dei residui del forecast addestrato sul dataset di train, con il modello SARIMAX, ed il dataset di test, del ristorante R005

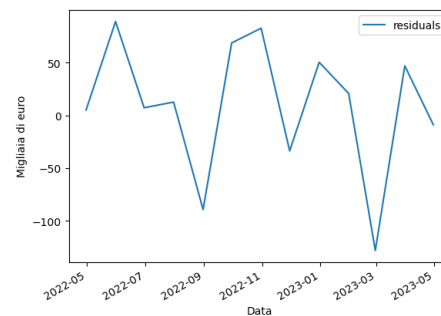


Figura 0.24: Plot dei residui del forecast addestrato sul dataset di train, con il modello Prophet, ed il dataset di test, del ristorante R005

## 0.6 Conclusioni

L'approccio grid search si è dimostrato essenziale per trovare gli iperparametri migliori.

Ristorante	Prophet	SARIMAX
R000	(1.5, 0.01, 0.001, multiplicative)	(2, 5, 6, 1, 1, 2)
R001	(8.3, 0.01, 0.001, multiplicative)	(7, 2, 5, 0, 0, 1)
R002	(1.7, 0.01, 0.01, additive)	(5, 2, 2, 1, 0, 0)
R003	(3.0, 0.1, 0.001, additive)	(0, 2, 4, 0, 1, 2)
R004	(3.8, 0.01, 1, multiplicative)	(1, 2, 1, 1, 0, 1)
R005	(0.6, 0.01, 0.001, multiplicative)	(7, 2, 6, 1, 0, 0)

Come mostra la tabella e l'istogramma, i risultati ottenuti dal codice mostrano

che entrambi i modelli sono in grado di fare previsioni accurate sulle vendite dei ristoranti. Tuttavia, il modello SARIMAX sembra avere prestazioni leggermente migliori rispetto al modello Prophet. I due modelli hanno mostrato in livello di errore molto simile, con SARIMAX che presenta un NRMSE mediamente più basso del 8,74%.

Modello	NRMSE
SARIMAX R000	0,158713012
Prophet R000	0,211500839
SARIMAX R001	0,266835606
Prophet R001	0,266006329
SARIMAX R002	0,186094034
Prophet R002	0,218986048
SARIMAX R003	0,130022089
Prophet R003	0,134819672
SARIMAX R004	0,260912711
Prophet R004	0,273414772
SARIMAX R005	0,205963518
Prophet R005	0,215949011

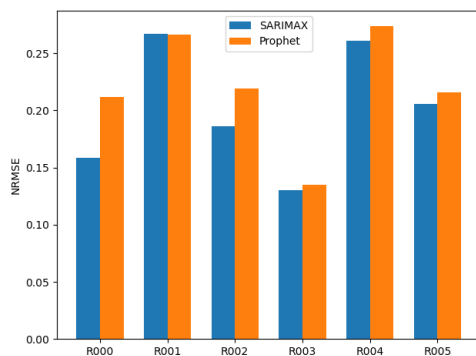


Figura 0.25: Questa figura mostra l’NRMSE del modello Prophet e del modello SARIMAX per ogni ristorante

Le seguenti tabelle mostrano le previsioni a 12 mesi per ogni ristorante, il forecast è molto simile, SARIMAX presen-

ta valori mediamente più alti del 0,226%.

Tabella ristorante R000:

Mese	Prophet (migliaia di euro)	SARIMAX (migliaia di euro)
2023-05	1078	1169
2023-06	1107	1182
2023-07	1126	1164
2023-08	1160	1221
2023-09	1166	1256
2023-10	1193	1233
2023-11	1206	1306
2023-12	1212	1341
2024-01	1288	1297
2024-02	1280	1323
2024-03	1259	1309
2024-04	1268	1327

Tabella ristorante R001:

Mese	Prophet (migliaia di euro)	SARIMAX (migliaia di euro)
2023-05	1689	1710
2023-06	1765	1638
2023-07	1779	1811
2023-08	1806	1840
2023-09	1829	1921
2023-10	1891	1942
2023-11	1884	1996
2023-12	1927	1908
2024-01	1983	1939
2024-02	2032	1864
2024-03	2001	1904
2024-04	2026	1912

Tabella ristorante R002:

Mese	Prophet (migliaia di euro)	SARIMAX (migliaia di euro)
2023-05	1560	1611
2023-06	1607	1699
2023-07	1623	1638
2023-08	1655	1553
2023-09	1658	1648
2023-10	1703	1765
2023-11	1764	1751
2023-12	1735	1782
2024-01	1786	1888
2024-02	1810	1753
2024-03	1804	1803
2024-04	1821	1913

Tabella ristorante R003:

Mese	Prophet (migliaia di euro)	SARIMAX (migliaia di euro)
2023-05	786	746
2023-06	821	771
2023-07	844	824
2023-08	867	829
2023-09	865	848
2023-10	901	826
2023-11	920	827
2023-12	914	851
2024-01	944	861
2024-02	958	873
2024-03	960	918
2024-04	976	889

Tabella ristorante R004:

Mese	Prophet (migliaia di euro)	SARIMAX (migliaia di euro)
2023-05	1340	1409
2023-06	1402	1446
2023-07	1431	1420
2023-08	1467	1481
2023-09	1460	1527
2023-10	1502	1517
2023-11	1523	1589
2023-12	1503	1609
2024-01	1548	1606
2024-02	1584	1661
2024-03	1577	1647
2024-04	1583	1668

Tabella ristorante R005:

Mese	Prophet (migliaia di euro)	SARIMAX (migliaia di euro)
2023-05	1293	1300
2023-06	1369	1318
2023-07	1376	1329
2023-08	1390	1351
2023-09	1386	1415
2023-10	1463	1372
2023-11	1480	1435
2023-12	1465	1431
2024-01	1505	1424
2024-02	1536	1449
2024-03	1529	1487
2024-04	1549	1461

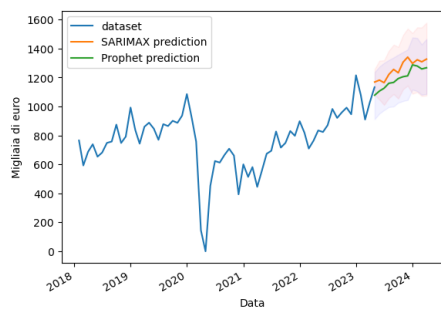


Figura 0.26: In questa figura vediamo la proiezione a 12 mesi del fatturato mensile dei modelli Prophet e SARIMAX del ristorante R000

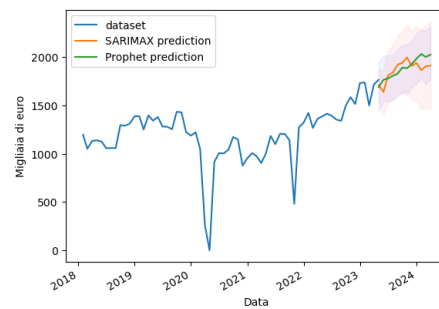


Figura 0.27: In questa figura vediamo la proiezione a 12 mesi del fatturato mensile dei modelli Prophet e SARIMAX del ristorante R001

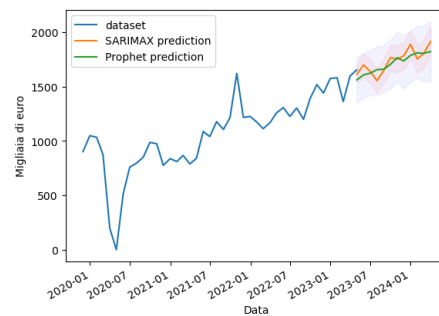


Figura 0.28: In questa figura vediamo la proiezione a 12 mesi del fatturato mensile dei modelli Prophet e SARIMAX del ristorante R002

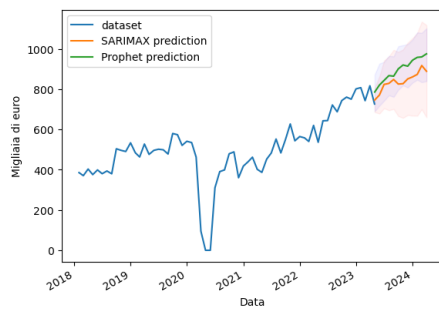


Figura 0.29: In questa figura vediamo la proiezione a 12 mesi del fatturato mensile dei modelli Prophet e SARIMAX del ristorante R003

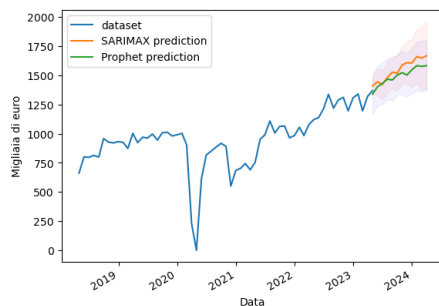


Figura 0.30: In questa figura vediamo la proiezione a 12 mesi del fatturato mensile dei modelli Prophet e SARIMAX del ristorante R004

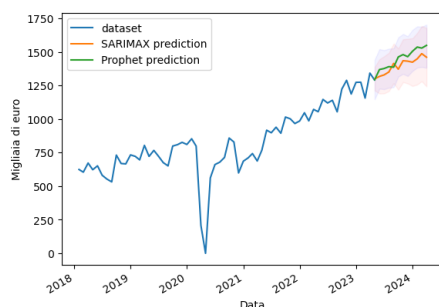


Figura 0.31: In questa figura vediamo la proiezione a 12 mesi del fatturato mensile dei modelli Prophet e SARIMAX del ristorante R005

## Limitazioni e ulteriori ricerche

I modelli utilizzati in questo studio, SARIMAX e Prophet, hanno dimostrato di essere efficaci nel fare previsioni sulle vendite dei ristoranti. Tuttavia, come tutti i modelli di previsione, presentano alcune limitazioni.

Una limitazione di entrambi i modelli è che si basano su ipotesi riguardo alla distribuzione dei dati e alla relazione tra le variabili. Se queste ipotesi non sono soddisfatte, i modelli potrebbero non fornire previsioni accurate. Inoltre, i modelli potrebbero non essere in grado di catturare eventi imprevedibili o cambiamenti improvvisi nella serie temporale.

Un'altra limitazione è che i modelli utilizzati in questo studio sono stati addestrati su un dataset specifico e potrebbero non generalizzare bene ad altri dataset o contesti. Inoltre, i modelli potrebbero richiedere una calibrazione regolare per mantenere la loro precisione nel tempo.

Per superare queste limitazioni, ulteriori ricerche potrebbero esplorare l'utilizzo di altri modelli di previsione o l'integrazione di tecniche di apprendimento automatico per migliorare la precisione delle previsioni. Inoltre, potrebbe essere utile esplorare l'utilizzo di dati aggiuntivi, come informazioni sulle tendenze del mercato o sui comportamenti dei consumatori, per migliorare la precisione delle previsioni.