

Ristorazione: serie storiche e previsioni future.

Gabriele Strano, *g.strano@campus.unimib.it*
Laura Nembrini, *l.nembrini@campus.unimib.it*
Antonio Lombardo, *a.lombardo33@campus.unimib.it*
Valentina Camagni, *v.camagni1@campus.unimib.it*
Gabriele Pagani, *g.pagani10@campus.unimib.it*

Università degli Studi di Milano-Bicocca, CdLM Data Science

27 Agosto 2021

Sommario

A seguito dell'avvento del Covid e delle relative restrizioni, le vendite delle attività appartenenti al settore della ristorazione hanno subito grosse depressioni e forte instabilità.

In un periodo storico come questo, per quanto tale instabilità le renda meno efficaci, le tecniche di previsione delle serie storiche assumono, specialmente in questo settore, vitale importanza.

Avere la possibilità di poter regolare la produzione in tutti i suoi aspetti può aiutare a ridurre sensibilmente i costi di gestione superflui, permettendo così di limitare i danni derivanti dai minori introiti.

In questo elaborato si è fatto ricorso a modelli di previsione delle serie storiche, come ARIMA e TBATS.

Soprattutto il secondo modello, con la possibilità di inserire due diverse stagionalità, ha permesso di generare una rappresentazione abbastanza affidabile del periodo futuro, che possa servire da guida alle decisioni del management.

Tramite la funzione di regressione sono state, inoltre, considerate anche variabili esterne, come la temperatura ed il peso del fine settimana e le influenze che essi generano sulle vendite complessive.

Sono state ancora analizzate le prestazioni di tutti i ristoranti, in termini di variazioni percentuali rispetto agli anni precedenti, per riuscire a capire quale modello di management stia pagando tra quelli adottati.

Parole chiave: Serie storiche; Self-Organizing Map; Arima; TBATS; Vendite.

Indice	3	Aspetti metodologici	3
1 Introduzione	2	4 I dati	4
2 Obiettivo/problema affrontato	3	5 Analisi/Processo di trattamento dei	

dati	6
5.1 Self-organizing map	6
5.2 Decomposizione	8
5.3 Modello ARIMA	8
5.4 Modello TBATS	9
5.5 Analisi di regressione	10
5.6 Variazioni percentuali delle vendite	11
6 Risultati	11
7 Conclusione e possibili sviluppi	12

1 Introduzione

Il mondo della ristorazione è sicuramente uno dei settori economici più colpiti dall'emergenza sanitaria da Covid-19 che ha interessato l'Italia e il resto del mondo. L'impossibilità di evitare qualsiasi tipo di contatto sociale ha reso i ristoranti, bar e pub i primi luoghi di aggregazione sottoposti a chiusura totale.

Ora, grazie all'introduzione di protocolli di protezione che garantiscono il distanziamento sociale, la maggior parte delle attività sono ripartite.

Nonostante la riapertura, senza ombra di dubbio, la pandemia ha cambiato il "modo di fare ristorazione". Infatti, durante il secondo lockdown i ristoranti che hanno proposto il servizio di asporto hanno potuto parzialmente riprendere l'attività; quelli che invece non lo comprendevano, in molti casi, hanno deciso di introdurlo per evitare il fallimento. Il food delivery, dunque, è stato un elemento importante che ha riguardato il settore della ristorazione durante l'emergenza Coronavirus.

Un altro aspetto fondamentale conseguente alla pandemia è stato l'attenzione posta al problema dell'igiene. La necessità di introdurre mascherine, guanti, dispenser di gel igienizzante insieme a procedure di sanificazione dei locali hanno comportato dei costi non indifferenti per attività già colpite. Tuttavia, tutto ciò non ha portato soltanto i gestori ad adattarsi a questo nuovo modo di fare ristorazione ma li ha visti anche costretti a riorganizzare gli ambienti interni in modo da garantire la sicurezza dei clienti. I proprietari, infatti, hanno provveduto a sistemare i tavoli ad una certa distanza l'uno dall'altro, an-

che fornendosi di pannelli divisorii, diminuendo significativamente il numero di posti disponibili. In questo contesto, diventa cruciale individuare le fonti di costi non essenziali per arginare i danni economici causati dalla pandemia.

Nei paragrafi successivi andremo a spiegare brevemente i passaggi effettuati.

In particolare, nel paragrafo 2 verrà brevemente illustrato l'obiettivo dell'analisi.

In Aspetti metodologici, verranno, invece, spiegati i modelli di previsione utilizzati (ARIMA, TBATS) e la Self-Organizing Map. Infatti, la multidimensionalità dei dati ha reso necessario ricorrere alle SOM, mappe auto-organizzanti che permettono la rappresentazione discretizzata dello spazio di input. Questo ci ha permesso di visualizzare il dataset in una maniera più semplice ed intuitiva individuando i periodi più o meno simili.

Successivamente, è stata fornita una descrizione dei dati a nostra disposizione. In particolare, essi riguardavano 6 ristoranti localizzati in Emilia-Romagna le cui vendite e i cui scontrini sono stati registrati per un arco temporale di 4 anni, dal 1° gennaio 2017 al 12 aprile 2021. Di conseguenza, avendo a disposizione solamente due indicatori di bilancio è stato necessario ricercare ulteriori informazioni riguardo al periodo studiato in modo da arricchire il nostro dataset.

Le vendite e gli scontrini di un ristorante, infatti, possono essere correlate a diversi fattori quali il meteo, il giorno della settimana (ci aspettiamo una maggiore affluenza di clientela nel weekend) e indubbiamente, nel periodo di pandemia, la presenza o meno di norme restrittive. Tali variabili esterne sono poi state inserite in un modello di regressione in modo da testare la significatività di ognuna nello spiegare le vendite.

Nel paragrafo 5 abbiamo racchiuso i vari procedimenti effettuati per la visualizzazione dei dati tramite la SOM e per la previsione delle serie storiche.

L'andamento complessivo delle vendite di ogni ristorante è stato analizzato nel tempo così da far emergere la presenza di stagionalità nelle serie storiche, che insieme alle altre componenti ha permesso di realizzare previsioni future ab-

bastanza affidabili.

A livello informativo è stato utile analizzare la variazione percentuale delle vendite rispetto all'anno precedente, al fine di comprendere se la gestione attuale operata dal management sia stata virtuosa e sostenibile nel tempo.

Infine, in un contesto storico come quello che stiamo vivendo, si è ritenuto interessante studiare da una parte l'impatto del Covid sui ristoranti e dall'altra la loro capacità nel gestire una fase così complessa.

In conclusione, verranno presentati brevemente i risultati ottenuti e i possibili sviluppi futuri.

2 Obiettivo/problema affrontato

Il lavoro è volto a far emergere nuove informazioni ed eventuali ciclicità dai dati in nostro possesso relativi alle sei attività appartenenti al mondo della ristorazione.

Attraverso l'analisi delle serie storiche e delle loro componenti, soprattutto della stagionalità, è possibile realizzare previsioni future delle vendite abbastanza affidabili.

Queste previsioni, in settori come la ristorazione e il retail, sono di cruciale importanza in quanto vengono spesso utilizzate per fissare obiettivi e conseguenti premi di produzione volti ad incentivarne il raggiungimento. Fungono, inoltre, da guida per determinare come comporre il magazzino e la forza vendita, in sintesi per regolare la capacità di produzione.

Si è cercato di far emergere similitudini tra i diversi periodi dell'anno, tramite l'utilizzo delle SOM, e di analizzare l'andamento delle diverse attività tenendo conto non tanto delle vendite in termini assoluti quanto piuttosto delle variazioni percentuali, in modo tale da poter comprendere quanto le scelte del management abbiano pagato nel tempo.

3 Aspetti metodologici

Al fine di produrre una rappresentazione dello spazio dei dati in input si è ricorsi alle Self-Organizing Maps (SOM).

L'obiettivo è quello di rappresentare la distribu-

zione dei dati in input organizzandoli spazialmente su una mappa bidimensionale.

In particolare, le SOM sono molto utili per visualizzare i dati, analizzarne la tendenza di clustering e studiare le relazioni esistenti tra le variabili del dataset. Dunque, permettono un'esplorazione intuitiva dei dati in analisi.

Nel nostro studio tale metodo di indagine sarà utilizzato a puro scopo descrittivo e non a finalità di clustering poiché riteniamo esigie le informazioni che disponiamo a riguardo dei 6 ristoranti.

Gli unici indicatori per ogni ristorante riguardano le vendite e gli scontrini effettuati giornalmente. Questo limita le nostre possibilità di classificarli in diverse tipologie di ristoranti.

In questa fase del lavoro si è usufruito degli strumenti offerti dai pacchetti *aweSOM* e *kohonen* di R. Nello specifico, il primo introduce grafici interattivi che facilitano l'analisi delle SOM create, per cui a questo elaborato verranno allegati i file html per la visualizzazione interattiva dei grafici.

Poiché tra i 6 ristoranti il primo sembra mostrare l'andamento medio fra tutti, si è deciso di condurre questa analisi preliminare sui dati relativi a questo ristorante.

E evidente che il numero e la tipologia degli indicatori selezionati per il ristorante influenzano i risultati della SOM, per cui la selezione è stata effettuata tenendo conto dei risultati che si vogliono conseguire.

Poiché le SOMs hanno una elevata capacità di elaborare grosse quantità di dati e di trovare relazioni multidimensionali tra le variabili, è possibile utilizzare un elevato numero di indicatori in input; tuttavia, più numeroso è il vettore di dati, più difficile sarà poi comprendere e commentare i risultati conseguiti. Nel caso in esame, oltre alle vendite sono stati scelti gli indicatori che maggiormente si prestano a profilare le vendite stesse del ristorante.

Per valutare la significatività delle variabili, è stato inizialmente condotto un test ANOVA che ha evidenziato come 5 indicatori siano particolarmente rilevanti nello spiegare le vendite.

Inoltre, sempre allo scopo di rendere la mappa il più esplicativa possibile si è deciso di considerare i dati aggregati mensilmente così che ogni

profilo v_i passato alla mappa rappresenti un mese dell'arco temporale considerato.

In primo luogo si creerà una SOM Hitmap in cui ogni neurone (o tassello) della mappa visualizzerà il numero di istanze che sono assegnate a quel neurone. Lo scopo sarà quello di osservare se le istanze di input siano o meno equamente distribuite.

In secondo luogo si sono plottate sulla mappa le variabili numeriche identificate come indicatori del ristorante, dunque verrà visualizzato il vettore rappresentativo di ogni neurone della mappa. All'interno di ogni neurone, viene mostrata la media di ciascuna variabile. In questo modo, si può osservare dove si addensano i valori più grandi o più piccoli di ciascuna variabile. Infine, per testare quanto bene la mappa rappresenti i nostri dati è stata creata la U-matrix, la quale associa ad ogni neurone della mappa la distanza euclidea media (in R la somma delle distanze) fra il proprio codebook e i codebook dei neuroni primi vicini. In questo modo si può valutare se la mappa contiene "fratture" o è uniforme. In aggiunta a questo si è valutato l'errore di quantizzazione, ovvero la distanza euclidea media dei dati in input mappati in un neurone dal codebook di quel neurone.

Per effettuare la previsione della serie storica sono stati utilizzati due diversi modelli entrambi contenuti all'interno del pacchetto "*Forecast*" di R: il modello ARIMA, con aggiunta della componente stagionale, ed il modello TBATS. Preliminarmente, sono state effettuate analisi esplorative sulle serie storiche; andava verificata la struttura della serie come somma delle sue componenti.

Per fare ciò è stato necessario decomporre la serie storica con il comando *decompose*. Per ovviare al problema della non stazionarietà è stato, invece, utilizzato il comando *diff()*, che consiste nel calcolare le differenze tra le osservazioni consecutive.

Il modello ARIMA è un modello risultante dall'unione delle componenti autoregressiva (AR) e media mobile (MA).

All'unione di queste due componenti si aggiunge una parte integrata, che permette di gestire la differenziazione di una serie, ossia la differenza

di un'osservazione da un'osservazione nella fase temporale precedente.

Nel nostro caso specifico abbiamo dovuto includere anche la componente stagionale, in quanto presente nei nostri dati. Abbiamo applicato questo modello ARIMA stagionale dopo aver aggregato i dati settimanalmente, perché questi modelli non possono tenere conto di due diverse stagionalità.

Per avere invece una previsione abbastanza affidabile della serie temporale, senza dover ridurre la granularità dei nostri dati, abbiamo impiegato un modello TBATS che permette di considerare la doppia stagionalità al proprio interno. I due diversi periodi impostati per la stagionalità sono 7 giorni e 365.25 giorni in modo da considerare sia la stagionalità settimanale che quella annuale.

4 I dati

I dati che ci sono stati forniti riguardano 6 ristoranti localizzati in Emilia-Romagna e comprendono uno spazio temporale di poco più di 4 anni, dal 1° gennaio 2017 fino al 12 aprile 2021. Le variabili a nostra disposizione riguardano il numero di scontrini giornalieri effettuati dal singolo ristorante e le vendite effettuate, anch'esse giornaliere. Nello specifico, abbiamo quindi 12 variabili che sono:

1. Vendite_1: vendite effettuate dal primo ristorante;
2. Scontrini_1: numero di scontrini emessi dal primo ristorante;
3. Vendite_2: vendite effettuate dal secondo ristorante;
4. Scontrini_2: numero di scontrini emessi dal secondo ristorante;
5. Vendite_3: vendite effettuate dal terzo ristorante;
6. Scontrini_3: numero di scontrini emessi dal terzo ristorante;
7. Vendite_4: vendite effettuate dal quarto ristorante;

8. Scontrini_4: numero di scontrini emessi dal quarto ristorante;
9. Vendite_5: vendite effettuate dal quinto ristorante;
10. Scontrini_5: numero di scontrini emessi dal quinto ristorante;
11. Vendite_6: vendite effettuate dal sesto ristorante;
12. Scontrini_6: numero di scontrini emessi dal sesto ristorante.

Sono tutte variabili quantitative; in particolare quelle che riguardano gli scontrini effettuati sono una variabile di tipo intero, mentre quelle che riguardano le vendite sono una variabile di tipo float.

Le caratteristiche dei ristoranti sono a noi ignote, ovvero non ci sono state fornite informazioni sulla tipologia del ristorante, sulla sua planimetria, sulla sua precisa localizzazione ecc.

Per poter raggiungere gli obiettivi prefissati, si è deciso di provvedere ad una integrazione dei dati a nostra disposizione.

Sono stati integrati inizialmente i dati che riguardavano l'aspetto meteorologico ottenuti dalla fonte <https://www.ilmeteo.it/>, che è uno fra i siti meteo più affidabili.

Essendo a noi ignote le caratteristiche dei ristoranti, si è scelto di utilizzare i dati meteo relativi alla città di Bologna, capoluogo dell'Emilia-Romagna.

I dati, anch'essi con granularità giornaliera, sono composti dalle seguenti caratteristiche:

1. Temperatura media;
2. Temperatura minima;
3. Temperatura massima;
4. Punto rugiada;
5. Umidità (%);
6. Visibilità (km);
7. Vento medio (km/h);
8. Vento max (km/h);

9. Raffica (Km/h);
10. PressioneSML (mb);
11. Pressione media (mb);
12. Pioggia (mm);
13. Fenomeni.

Dovendo effettuare un'analisi in ambito ristorazione, ci è sembrato utile e rilevante tenere in considerazione se quello specifico giorno è stato un giorno di pioggia o meno. Per questo motivo, a partire dalla variabile fenomeni, che è di tipo qualitativo, è stata generata un'ulteriore variabile denominata 'Maltempo'.

Quest'ultima è stata creata come variabile binaria e presenta 1 nel momento in cui in quella giornata si sono presentati fenomeni atmosferici come 'pioggia', 'neve', 'temporale', 0 altrimenti. L'arco temporale dei dati a nostra disposizione comprende anche il periodo di pandemia da Covid-19, di cui si è iniziato a parlare già da dicembre 2019.

Essa ha colpito moltissimi settori a causa delle numerose chiusure e l'ambito ristorazione è stato proprio uno dei settori a risentirne maggiormente. Per questo motivo è stata generata una variabile denominata 'Covid'. Essa è una variabile binaria che presenta il valore 1, in caso in cui erano presenti restrizioni all'attività lavorativa, in termini di chiusura parziale o totale, 0 altrimenti. Nello specifico, il periodo di chiusure, parziali o totali, va dal 9 marzo 2020 fino al 12 aprile 2021.

Infine, i dati sono stati integrati con un'ulteriore variabile binaria generata da noi, ovvero 'Weekend'.

Infatti, generalmente, in ambito ristorazione, nel weekend c'è una maggiore affluenza di persone e per questo ci è sembrato utile evidenziarlo. Essa presenta il valore 1 nel momento in cui il giorno ricade di Sabato oppure di Domenica, 0 altrimenti.

5 Analisi/Processo di trattamento dei dati

Per analizzare l'andamento complessivo di ogni locale nel tempo, si è deciso di aggregare i dati mensilmente in modo da facilitarne l'utilizzo nonostante la perdita di granularità e di dettaglio.

Dalla Fig. 1 si evince come i ricavi mensili di ogni ristorante abbiano lo stesso andamento con drastiche perdite nel periodo di lockdown. Si nota, come la curva in verde, ristorante 3, e quella in viola, ristorante 6, non partano dal Gennaio 2017 ma da periodi successivi. Si suppone che queste attività siano state avviate in una data successiva o che i dati non siano stati raccolti.

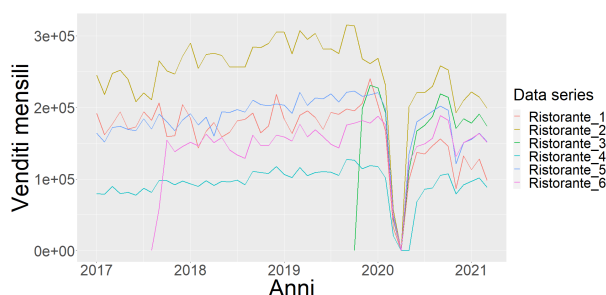


Figura 1: Andamento vendite mensili dei sei ristoranti.

Per le successive analisi si è deciso di analizzare solamente il primo dei ristoranti in quanto sembra rappresentare l'andamento medio fra tutti. Analizzando, quindi, le vendite registrate del primo ristorante per i diversi anni si vede, in Fig. 2, come l'andamento annuale sia sostanzialmente invariato con valori minimi nel mese di Febbraio e con valori crescenti per i mesi di Ottobre, Novembre e Dicembre.

Questo ci lascia ipotizzare la presenza di stagionalità annuale che più avanti andremo meglio a verificare.

Valutando l'anno 2020, in cui è presente la variabile causata dal Covid-19, si vedono vendite mensili inferiori con una netta discesa nel mese di Marzo causata dalla chiusura totale delle attività.

L'andamento viene confermato anche andando a confrontare l'andamento fra gli stessi mesi in

anni differenti, Fig. 3. Qui possiamo notare come il 2020 sia diventato l'anno peggiore, nonostante fosse iniziato con vendite superiore al triennio precedente.

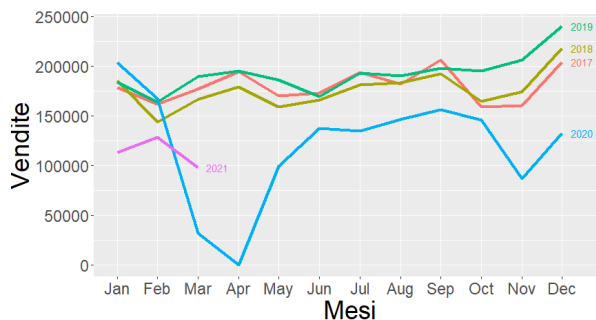


Figura 2: Andamento vendite mensili negli anni del ristorante 1.

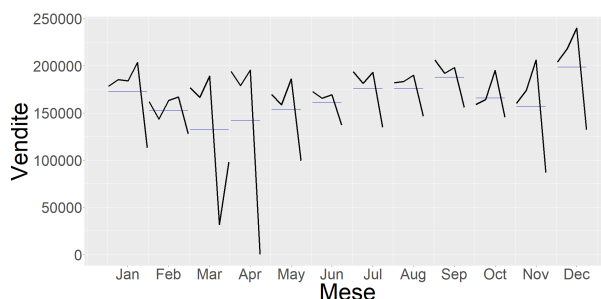


Figura 3: Confronto vendite per lo stesso mese nei diversi anni del ristorante 1.

5.1 Self-organizing map

Scopo di una SOM è di proiettare dati appartenenti ad uno spazio multidimensionale in uno spazio a dimensione ridotta, nel caso in esame bidimensionale R^2 .

L'algoritmo implementato si compone di una fase di inizializzazione e una di apprendimento automatico da parte dell'algoritmo inizializzato.

Nella fase iniziale si definiscono:

- Le dimensioni della mappa: n righe e m colonne che definiscono la risoluzione della mappa.

In questo lavoro si è scelto a posteriori di creare una griglia di 4x4 tasselli attraverso la creazione della SOM di tipo Hit-map in Fig.4 che ha mostrato un neurone

con nessuna osservazione. Lì dove vi sono più neuroni ai quali non vi sono attribuite osservazioni sarebbe stato necessario diminuire le dimensioni della griglia;

- La forma dei tasselli: esagonale;
 - La topologia di tipo euclideo;
 - La dimensione del campione di training che si osserva. Nel caso in esame, poiché lo scopo è di visualizzazione dei dati in input, si è deciso di utilizzare l'intero dataset disponibile.
- In termini di indicatori del ristorante, sulla base del test ANOVA, si sono considerate 5 variabili: *SCONTRINI_1*, *TMEDIA* (*temperatura media*), *Covid*, *WEEKEND*, *MALTEMPO*.
- I parametri del processo di apprendimento, dunque la legge con cui il diametro dell'intorno di influenza diminuisce, è stata mantenuta l'impostazione predefinita di diminuire linearmente l'intorno di influenza da 0,05 a 0,01;

Dopo aver selezionato i dati di addestramento della SOM (ridimensionando ogni variabile alla varianza unitaria), la funzione *somInit* inizializza i prototipi della mappa.

Per impostazione predefinita, questa operazione viene eseguita utilizzando un metodo basato su PCA. I dati di addestramento, i prototipi iniziali e altri parametri (*rlen=100*, *radius=c(2.65,-2.65)*, *dist.fcts="sumofsquares"*) vengono passati alla funzione *kohonen::som* per l'addestramento.

AweSOMplot crea una varietà di visualizzazioni interattive della SOM. Utilizzando l'argomento *type* della funzione si è creata la Hitmap, Fig.4, che rappresenta la SOM da $4 \times 4 = 16$ neuroni e da essa si evince chiaramente come le istanze di input non siano equamente distribuite, ma si addensino in alcuni neuroni, mentre altri contengono poche o nessuna istanza.

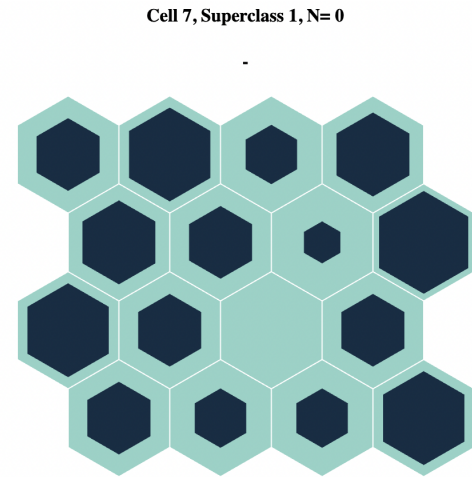


Figura 4: SOM Hitmap.

Si può notare come in basso a sinistra vi sono neuroni che contengono un numero simile di istanze.

Tracciata la Hitmap, si è deciso di rappresentare nel dettaglio il vettore delle medie di ciascuna variabile che definisce il singolo neurone attraverso il *type* "Circular".

Si può osservare in Fig.5 come i valori più alti della variabile Covid si concentrino in basso a sinistra. Grazie all'interattività della mappa si può riscontrare, come atteso, che i primi mesi investiti dalla pandemia Covid, dunque Aprile 2020 e Maggio 2020, rientrano nello stesso neurone (cella 2) che si caratterizza per l'assenza totale di vendite e, dunque, di scontrini.

Inoltre, a testimonianza di una buona funzionalità della mappa, i successivi mesi di lockdown (in cui i ristoranti svolgevano attività ma limitatamente per l'asporto - per cui la variabile Covid è stata settata a 1) si localizzano nella stessa area, dunque in neuroni vicini.

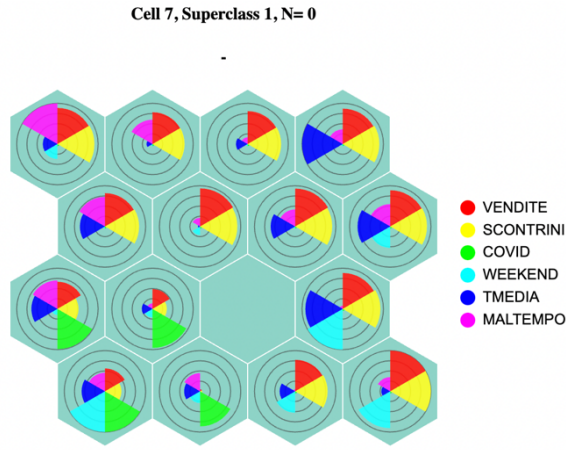


Figura 5: SOM Circular.

Per stabilire la qualità della mappa, dunque quanto bene la mappa rappresenti i dati, in ingresso si è utilizzato lo stesso comando con *type* “*U-matrix*” in Fig. 6. Questo è un modo per esplorare la topografia della mappa. La mappa mostra la distanza media tra ogni neurone e i prototipi dei suoi vicini su una scala di colori. I neuroni più scuri sono attigui ai loro vicini mentre quelli più luminosi sono più distanti dai loro vicini.

La funzione *somQuality* restituisce l’errore di quantizzazione della mappa e la percentuale di varianza spiegata, rispettivamente, pari a 0.779 e 86,41%.

La U-matrix conferma, anche visivamente, la distanza tra i primi mesi di pandemia, e dunque di chiusura totale, dagli altri.

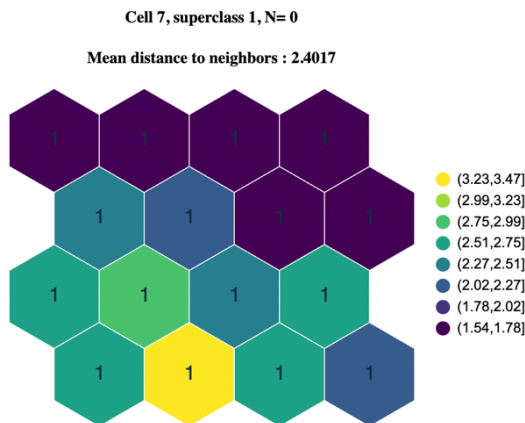


Figura 6: SOM U-matrix.

5.2 Decomposizione

Prima di effettuare previsioni con i modelli ARIMA o TBATS è opportuno effettuare una decomposizione della serie storica per visualizzare le varie componenti presenti al suo interno.

Questa operazione ci permette di far emergere le componenti di trend, le stagionalità e l’innovazione.

Nei dati relativi alle vendite del ristorante 1, Fig. 7, emergono due differenti periodicità, è presente, quindi, una doppia stagionalità all’interno dei dati.

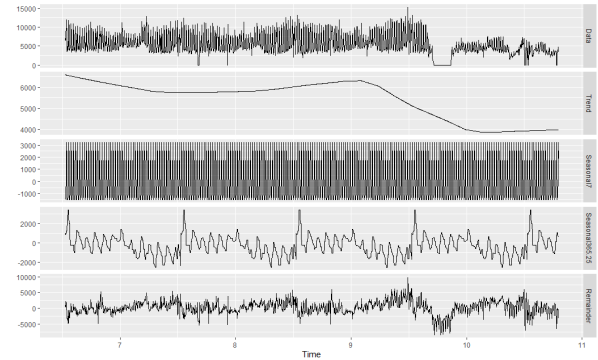


Figura 7: Componenti serie storica decomposta.

Quello che emerge da questo grafico, oltre alla già citata presenza di due stagionalità, è che queste, a differenza della serie storica nel suo complesso, sembrano godere delle proprietà di stazionarietà.

Completata questa ennesima analisi esplorativa sui dati, si può adesso procedere alla fase di previsione della serie storica.

5.3 Modello ARIMA

Come precedentemente anticipato per lo sviluppo della previsione con il modello ARIMA, è stato necessario aggregare i dati settimanalmente, perdendo, quindi, granularità nei dati.

Questo, però, è stato fatto perché i modelli ARIMA non permettono di considerare contemporaneamente più stagionalità ed non sono stati adatti per previsioni troppo lunghe.

Per la ricerca dei coefficienti da considerare per inquadrare al meglio i dati all’interno di un modello ARIMA, sono stati scelti come punto di partenza gli output della funzione *auto.arima*

del pacchetto *forecast* di R e dei risultati dell'analisi dei grafici ACF e PACF.

Tra i vari modelli ARIMA testati quello che ha aderito meglio ai dati è stato un modello ARIMA (0,1,1) (1,0,0). Come si può notare il secondo coefficiente della parte non stagionale indica che la serie storica ha avuto bisogno di essere differenziata in quanto non era stazionaria ¹.

Successivamente è stato deciso di dividere la serie storica in train set e test set per poter in seguito verificare che il modello fosse abbastanza affidabile nelle sue previsioni.

Per la fase di validazione vanno analizzati i residui del modello.

Nelle Fig. 8 e 9 si nota come nei grafici ACF e PACF non si rilevino correlazioni anomale e la distribuzione risulti normale, come possiamo osservare grazie al Q-Q plot ed alla forma simmetrica dell'istogramma.

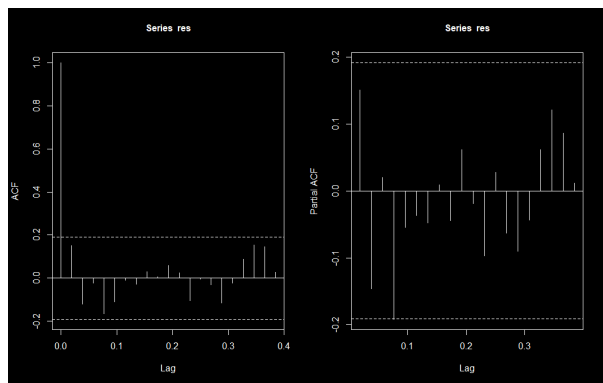


Figura 8: ACF E PACF.

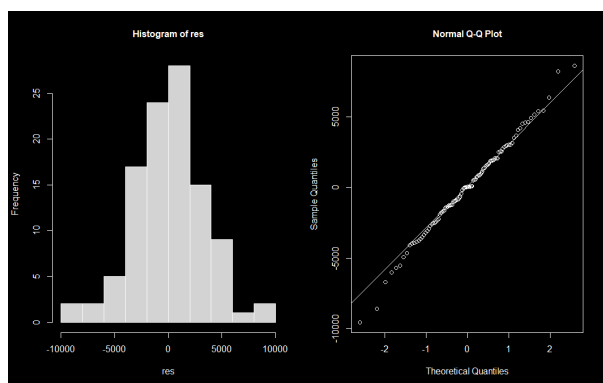


Figura 9: Istogramma dei reidui e Q-Q plot.

¹Un processo si definisce stazionario quando la media e la varianza non cambiano nel tempo.

A questo punto è possibile andare a prevedere come si evolverà la serie storica analizzata. In Fig. 10, si vede la previsione, comparata con ciò che è stato realmente registrato nell'anno 2019, dato che avevamo diviso i nostri dati in train set e test set.

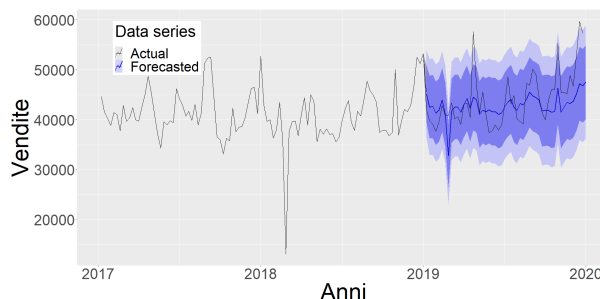


Figura 10: Previsione serie storica con dati aggregati settimanalmente.

La previsione sviluppata con questo modello ARIMA stagionale (0,1,1) (1,0,0) si è rivelata abbastanza precisa, le due bande di colore turchese presenti intorno alla previsione della serie, sono gli intervalli di confidenza, quello più grande è al 95 % mentre il più piccolo al 80 %.

Nonostante questa previsione si sia rivelata abbastanza affidabile, non verrà utilizzata per le previsioni future, in quanto si è scelto di puntare sul modello TBATS, che verrà illustrato nel successivo paragrafo.

5.4 Modello TBATS

TBATS è un acronimo ed ognuna delle lettere rappresenta una delle seguenti peculiarità del modello: Trigonometric, Seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal.

I modelli TBATS, rispetto ai modelli ARIMA, hanno il vantaggio di poter tenere conto di differenti stagionalità al loro interno; di contro risultano computazionalmente più pesanti rispetto a questi ultimi.

Come il precedentemente modello ARIMA, anche il modello TBATS è stato prima addestrato sul train set e successivamente testato sul test set. In questo caso però, grazie alla capacità di gestire più periodi stagionali, non è stato necessario ridurre la granularità dei dati che sono rimasti dati giornalieri.

In Fig. 11 vengono mostrate le previsioni ed i dati reali relativi agli ultimi 3 mesi del 2018 ed i primi 3 mesi del 2019.

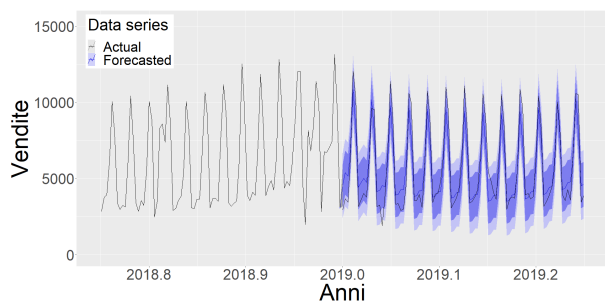


Figura 11: Previsione serie storica con dati giornalieri.

Dal grafico viene confermata la bontà delle previsioni effettuate con il modello TBATS; l'errore percentuale medio assoluto del modello (MAPE) è, infatti di 15.15.

Anche nel grafico di questo modello, come nel precedente (ARIMA), sono visibili due diversi intervalli di confidenza, nuovamente al 95 % ed al 80 %.

A questo punto, fatte le dovute verifiche circa la bontà dei modelli, è possibile utilizzare i modelli presi in considerazione per prevedere l'andamento futuro della serie storica relativa alle vendite del ristorante1, Fig. 12.

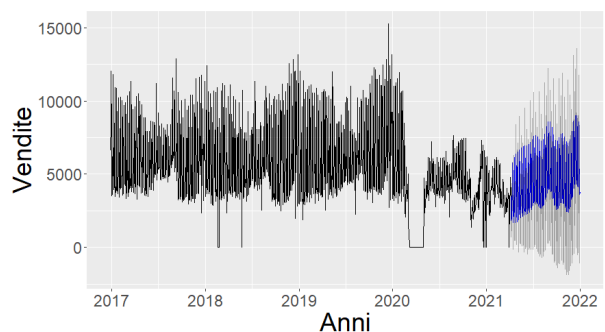


Figura 12: Previsione serie storica per l'anno 2021.

Questa previsione futura della serie storica relativa alle vendite può essere utilizzata come base di partenza per decisioni strategiche ed operative di management (la scelta degli obiettivi da raggiungere per la forza vendita ed eventuali incentivi per raggiungere tali obiettivi, per

regolare la numerosità del personale, gestire il magazzino ed evitare sprechi che vadano ad aumentare i costi di gestione).

5.5 Analisi di regressione

Per andare a scoprire quali fattori hanno un'influenza maggiore sui valori delle nostre vendite sono state sperimentati diversi modelli di regressione.

Alcune variabili, che erano state inizialmente considerate, sono state rimosse, in quanto toglievano significatività al modello poiché ridondanti. Alcune dimostrazioni: la variabile fenomeni con la variabile temperatura media, la variabile colore con la variabile dicotomica Covid e le altre variabili rappresentanti le condizioni climatiche.

Si è, quindi, trovato un modello che considerasse solamente variabili tra loro non correlate, evitando così di andare a falsare i risultati finali. Il modello che più si adattava ai dati trattati in questa analisi conteneva, oltre alla variabile dipendente, vendite del ristorante 1, tre variabili esplicative: temperatura media, weekend (variabile dicotomica dove 1 indica che la giornata appartiene al fine settimana) e Covid (variabile dicotomica dove 1 indica che in quel giorno vi erano chiusure o limitazioni dovute alla pandemia). Nella Fig. 13, è possibile osservare l'output del modello contenente tutti i coefficienti relativi all'analisi di regressione.

```

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5237.161    95.802   54.666 < 2e-16 ***
data$TMEDIA   -27.214     5.212   -5.221 2.03e-07 ***
data$WEEKEND  3681.625    93.083   39.552 < 2e-16 ***
data$COVID   -2364.653   139.160  -16.992 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1627 on 1488 degrees of freedom
(71 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared:  0.5548,    Adjusted R-squared:  0.5539
F-statistic: 618.2 on 3 and 1488 DF,  p-value: < 2.2e-16

```

Figura 13: Analisi di regressione

Da questi risultati si evince come il modello risulti nel complesso significativo, così come ognuna delle variabili esplicative considerate. Il valore dell' R^2 è pari a 0.555 e ciò indica che la parte spiegata dal modello è all'incirca del 55 %.

Guardando, invece, ai coefficienti delle singole variabili si nota come contro intuitivamente la temperatura media influenzi negativamente le vendite totali.

Quelli che, invece, confermano le aspettative iniziali sono i coefficienti delle variabili weekend e Covid. Come ci si poteva immaginare, la prima influenza positivamente le vendite, mentre la seconda negativamente.

Restando ferme le altre variabili, la presenza delle restrizioni legate al Covid fa diminuire in media di 2364 euro il totale delle vendite del ristorante 1.

Le giornate del fine settimana, sempre a parità delle altre variabili, fanno registrare vendite mediamente superiori di circa 3681 euro rispetto alle altre giornate.

5.6 Variazioni percentuali delle vendite

Analizzare la variazione delle vendite in percentuale rispetto all'anno precedente, risulta molto più informativo rispetto all'analisi in termini assoluti. La variazione percentuale permette, a prescindere dalla grandezza del fenomeno, di capire se la gestione attuale operata dal management sia virtuosa e sostenibile nel tempo.

Nel contesto storico analizzato, oltre all'andamento degli anni precedenti, ci permette anche di stimare in maniera approssimativa (con più dati a disposizione tutte le analisi risulterebbero più precise) l'impatto del Covid sui singoli ristoranti e la resilienza e la capacità con cui il management ha gestito una fase tanto delicata. Nella Fig. 14 è possibile osservare il grafico rappresentante le variazioni percentuali delle vendite dei singoli ristoranti rispetto ai risultati degli anni precedenti.

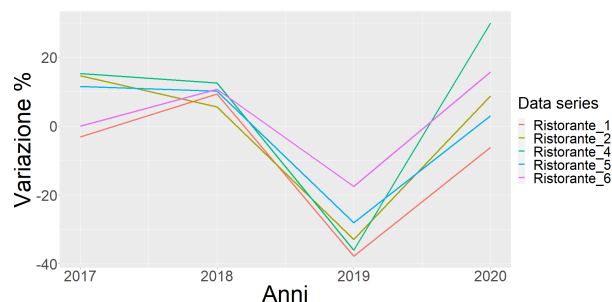


Figura 14: Variazione % rispetto all'anno precedente.

In questa analisi è stato volutamente omissivo il ristorante numero 3 in quanto i primi dati registrati partivano dal periodo di restrizioni conseguenti la pandemia Covid-19.

Il ristorante che sembra aver arginato meglio degli altri la riduzione dei ricavi sembra essere il ristorante numero 6.

Stranamente il ristorante numero 4 malgrado in periodo pre-pandemico risultasse in crescita costante, è stato quello che ha gestito peggio le perdite derivanti dalla pandemia parimenti al ristorante numero 1.

6 Risultati

Tramite le SOMs, ovvero le Self-Organizing Maps, si è effettuato un'analisi puramente descrittiva del fenomeno considerato.

Considerando i dati ottenuti in questa rappresentazione, si sono analizzati principalmente i neuroni che racchiudevano i dati riguardanti il periodo caratterizzato dalla presenza di Covid-19. I risultati dimostrano come a causa delle numerose restrizioni imposte, le vendite dei ristoranti siano state nettamente inferiori, o assenti, rispetto ad altri periodi. Inoltre, è stato possibile notare come la maggior parte dei neuroni con maggior concentrazione di istanze non riguardano i periodi caratterizzati da Covid. Questo aspetto è di facile interpretazione poichè dati a nostra disposizione coprono un periodo temporale di 4 anni, di cui solamente l'ultimo anno e mezzo si contraddistingue per la presenza del virus.

Dalle analisi di previsione effettuate, invece, è emersa una componente stagionale stazionaria quindi, facilmente prevedibile, composta da due

diverse periodicità (settimanale e annuale).

Sono state ottenute buone stime, sia con il modello ARIMA stagionale (0,1,1) (1,0,0) che ha un valore del MAPE pari a 9.92, che con il modello TBATS con doppia stagionalità che presenta un valore del MAPE di 15.15.

Si è preferito, tuttavia, affidarsi al secondo modello per le stime, in quanto, nonostante un valore di errore leggermente superiore, garantisce la possibilità di lasciare i dati giornalieri non riducendone così la granularità.

Tramite l'analisi di regressione sono emerse quali variabili più utili per spiegare l'andamento delle vendite dei ristoranti: la temperatura media, la variabile dicotomica weekend e la variabile dicotomica Covid.

Il modello di regressione scelto è risultato nettamente significativo e presentava un valore di R^2 di circa 0.55.

Tutte le variabili all'interno di questo modello risultavano significative; come ci si poteva immaginare, la variabile weekend influenzava molto forte e positivamente il valore delle vendite, mentre la variabile Covid aveva anch'essa un'influenza forte ma negativa.

Sorprendentemente è emerso che restando immutate le altre variabili, ad un aumento unitario della temperatura media corrispondeva una diminuzione di 27 euro in media dei ricavi.

7 Conclusione e possibili sviluppi

Partendo da una precisa analisi descrittiva dei dati forniti, abbiamo implementato un modello predittivo in grado di processare le serie storiche presenti nel dataset per effettuare previsioni, prevalentemente di carattere economico, con una discreta accuratezza.

Questo studio permette di fornire una solida struttura gestionale per intraprendere decisioni strategiche di investimento ed operative.

Queste, in ambito ristorativo e non, riguardano principalmente la scelta degli obiettivi aziendali da raggiungere, passando da strategie di investimento, fino ad arrivare alla gestione di magazzino per evitare sprechi che aumentino i costi di

gestione.

Queste decisioni sono di fondamentale importanza nel settore della ristorazione che, in questo periodo storico, assumono un valore sempre più crescente.

Infatti, in una situazione di pandemia caratterizzata da continue restrizioni e chiusure per ristoranti e bar, una scelta strategica errata può comportare, non solo una perdita di fatturato, ma la chiusura definitiva dell'attività.

In questa situazione, il modello predittivo oggetto di studio, può assumere un'importanza vitale in uno dei settori più colpiti dalla pandemia: "In 14 mesi sono stati bruciati 514 mila posti di lavoro, il doppio di quelli creati tra il 2013 e il 2019, e per oltre 6 ristoratori su 10 il calo di fatturato ha superato il 50 % del volume d'affari dell'anno precedente. Il Rapporto annuale sulla ristorazione in Italia per il 2020 di Fipe-Confcommercio appare come "un bollettino di guerra" di un settore a "pezzi", in cui il 97,5 % delle imprese ha perso fatturato nel 2020." ([Sole24ore](#)).

In questo scenario disastroso c'è la necessità di delineare degli sviluppi futuri in grado di incrementare l'accuratezza e la globalità del modello predittivo in modo tale da ampliare il raggio d'azione per ridurre i rischi per gli imprenditori. Questo è attuabile se si hanno a disposizione dati maggiormente esaustivi e che comprendano diverse sfere gestionali dell'attività ristorativa, al fine di estendere l'analisi ad aspetti commerciali tralasciati dal modello predittivo.

Riferimenti bibliografici

- [1] <https://www.reply.com/machine-learning-reply/it/essentials-self-organizing-maps>
- [2] <https://cran.r-project.org/web/packages/aweSOM/vignettes/aweSOM.html>
- [3] https://clarkdatalabs.github.io/soms/SOM_NBA
- [4] <https://rpubs.com/AlgoritmaAcademy/som>
- [5] https://eric.univ-lyon2.fr/~ricco/cours/slides/en/kohonen_som.pdf
- [6] <https://towardsdatascience.com/multiple-time-series-forecast>
- [7] <https://link.medium.com/2pdpWK4T8ib>
- [8] <https://towardsdatascience.com/a-guide-to-forecasting-in-r-6b0c9638c261>
- [9] <https://robjhyndman.com> › TBATS model - Forecasting: principles and practice.
- [10] <https://robjhyndman.com> › hyndsight. Forecasting with long seasonal periods | Rob J Hyndman.