

# Previsione dei consumi elettrici U1-U6 Bicocca

Autore 1: Lorenzo Bellotti, l.bellotti5@campus.unimib.it, 795192

Autore 2: Giovanni Tosi, g.tosi12@campus.unimib.it, 873263

L'obiettivo principale del nostro progetto è quello di individuare il miglior modello predittivo per quanto riguarda le stime dei consumi energetici relativi agli edifici U1 e U6 dell'Università Statale Di Milano-Bicocca per il periodo compreso tra il 5 ottobre 2020 e il 20 dicembre 2020. L'aspetto di maggior valore riguarda la possibilità da parte degli organi tecnici preposti dall'università di poter sapere con anticipo quelli che sono i possibili consumi elettrici degli edifici sottoposti ad analisi con l'opportunità, da parte degli stessi, di poter prevedere quello che sarà l'importo dei costi relativi al consumo di energia elettrica. Si è cercato di perseguire l'obiettivo dell'analisi aggregando i dati sia con frequenza giornaliera che con frequenza settimanale in modo tale da poter valutare le performance dei modelli utilizzati. Di seguito si dimostrerà che la granularità maggiormente efficace per gli scopi del progetto è quella giornaliera che garantisce un miglior risultato dal punto di vista della precisione della previsione. Il seguente progetto è stato svolto solamente analizzando due edifici di tutto il campus universitario, tuttavia, non vi sono limiti all'applicabilità dei modelli considerati anche ad altri edifici per avere un panorama completo su quelli che potrebbero essere i futuri fabbisogni elettrici dell'intero campus universitario.

Indice	<u>Parole chiave</u>
Parole chiave.....1	L'obiettivo dell'analisi si può riassumere in tre parole chiave: Spesa Energetica, Consumo Energetico, Previsione.
1 Introduzione.....1	
2 Obiettivo.....2	
3 Aspetti Metodologici.....2	
4 Dati.....4	
5 Trattamento dati.....5	
6 Analisi U1.....6	
7 Analisi U6.....8	
8 Prezzi energia elettrica.....10	
9 Risultati.....11	
10 Dashboard.....11	
11 Conclusioni.....11	
12 Fonti.....12	

	<u>1 Introduzione</u>
	Il focus di questo lavoro è quello di stabilire quale sia il miglior modello predittivo, tra quelli testati, per definire un budget preventivo per quanto riguarda la spesa energetica degli edifici U1 e U6.

A tale scopo per ciascun edificio si è proceduto come segue: la prima operazione svolta è stata il preprocessing dei dati al fine di renderli utilizzabili dai modelli predittivi. In seguito, si è provveduto a testare tali modelli sfruttando diverse granularità (giornaliera e settimanale). In particolare, per verificare le prestazioni dei modelli predittivi utilizzati, si è deciso di verificarne l'accuratezza nel periodo compreso tra 5 ottobre 2020 e il 20 dicembre 2020.

La scelta di tale periodo non è casuale, ma coincide con i primi tre mesi di lezioni prima della pausa natalizia, così facendo i modelli hanno avuto la possibilità di addestrarsi anche sul periodo COVID 19 migliorando la loro capacità predittiva in quanto, nel periodo utilizzato come test, la didattica si è tenuta principalmente a distanza.

Una volta ottenuta la previsione dei consumi energetici si è pensato di stimare i prezzi dell'energia elettrica per lo stesso periodo utilizzando i medesimi modelli predittivi precedentemente applicati.

Infine, si è deciso di sviluppare una dashboard che permetta all'utente di osservare l'andamento della spesa energetica reale e prevista per gli edifici oggetto di indagine.

In questo modo la direzione amministrativa dell'università potrebbe avere a disposizione uno strumento in grado di fornire un'indicazione per quanto riguarda l'allocatione preventiva di fondi per la copertura delle spese riguardanti il fabbisogno energetico degli edifici U1 e U6 potenzialmente estendibile a tutti gli altri edifici dell'Università.

## 2 Obiettivo

L'obiettivo del progetto è quello di trovare il miglior modello predittivo finalizzato alla creazione di uno strumento utile all'amministrazione dell'università in fase di pianificazione della spesa energetica degli edifici considerati. Un'ottimizzazione in fase di allocatione preventiva dei budget garantirebbe una migliore gestione dei fondi permettendo un notevole aumento dell'efficienza nella distribuzione delle risorse.

Per raggiungere questo scopo si è deciso di selezionare un periodo temporale di circa tre mesi. Questo periodo non è casuale ma risulta il compromesso ottimale tra lunghezza del range temporale previsto e conservazione della capacità predittiva del modello.

Per rendere maggiormente fruibile all'utente i risultati del modello si è sviluppata un'apposita dashboard finalizzata al monitoraggio delle spese energetiche di U1 e U6.

In particolare, l'utente finale al quale si rivolge il nostro progetto è la Commissione Energia di Ateneo che affianca la governance per effettuare una valutazione critica dei consumi di energia e delle modalità di acquisto e gestione della stessa, identificando strategie di miglioramento e riduzione dei consumi e dei costi.

## 3 Aspetti metodologici

Per conseguire gli obiettivi del progetto si è scelto di sfruttare i seguenti modelli: ARIMA, TBATS e PROPHET. Mentre per lo sviluppo della dashboard si è scelto di avvalersi delle funzionalità di Tableau. Di seguito si andranno ad approfondire i modelli utilizzati.

### 3.1 ARIMA

Il modello ARIMA è uno dei modelli maggiormente utilizzati per quanto riguarda la previsione su serie storiche; nasce dalla combinazione dei modelli: autoregressivo (AR), a media mobile (MA) con l'aggiunta dell'integrazione (I).

In particolare, il modello ARIMA è caratterizzato dalle componenti:

- P: ordine della componente autoregressiva;
- D: grado della differenziazione;
- Q: ordine della componente a media mobile;

il modello ARIMA può essere formulato come segue:

$$\varphi_b(B)\Delta dZ_t = \theta_0 + \theta q(B)a_t : ARIMA(p,d,q)$$

Dove:

- $\varphi_b(B)\Delta dZ_t$  : componente AR di ordine p;
- $\Delta dZ_t$  : parte integrata del modello con d numero di differenziazioni;
- $\theta_0$  : costante;
- $\theta q(B)a_t$  : componente a media mobile di ordine q.

### 3.2 TBATS

Oltre al modello ARIMA si è deciso di utilizzare il modello TBATS in grado di prendere in considerazione stagionalità multiple e complesse.

Il modello può essere formalizzato come segue:

$$y_t^{(\lambda)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = \phi b_{t-1} + \beta d_t$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i e_{t-i} + e_t$$

Dove:

$y_t^{(\lambda)}$  - time series at moment  $t$  (Box-Cox transformed)

$s_t^{(i)}$  -  $i$ th seasonal component

$l_t$  - local level

$b_t$  - trend with damping

$d_t$  - ARMA( $p,q$ ) process for residuals

$e_t$  - Gaussian white noise

### 3.3 PROPHET

PROPHET (presente nel pacchetto "prophet" di R) è una procedura per la previsione di serie storiche basata su modelli additivi, dove i trend non lineari sono analizzati con diverse stagionalità (annuale, settimanale e giornaliera). In particolare, il modello decompone la serie storica in trend, stagionalità e festività.

La previsione è ottenuta per mezzo della somma di trend e stagionalità usando come regressore il tempo analogamente ai modelli generalizzati (GAM).

$$y(t) = g(t) + s(t) + h(t) + e(t)$$

dove:

- $g(t)$ : rappresenta i cambiamenti non periodici (trend);
- $s(t)$ : rappresenta le componenti periodiche (stagionalità settimanali, mensili e/o annuali);
- $h(t)$ : rappresenta le feste ricorrenti ma che non ricadono sempre nello stesso giorno o periodo (es. Pasqua);
- $e(t)$ : rappresenta l'errore che il modello non riesce a spiegare.

Il modello gestisce sia trend non lineari costanti che lineari non costanti. Per modellizzare la funzione  $g(t)$  si identificano dei punti di cambiamento dove il tasso di crescita  $\kappa$  è soggetto ad una variazione che viene opportunamente compensata attraverso il parametro compensativo  $m$ .

Considerando la funzione di capacità  $C(t)$  si ottengono le seguenti formulazioni della funzione  $g(t)$ :

- trend non lineari ma costanti

$$g(t) = \frac{C(t)}{1 + e^{-(\kappa + a(t)^T \delta)(t - (m + a(t)^T \gamma))}}$$

- trend lineari ma non costanti

$$g(t) = (\kappa + a(t)^T \delta)t + (m + a(t)^T \gamma)$$

Dove:

$\kappa$ : rappresenta il tasso di crescita;

$\delta$ : rappresenta le regolazioni del tasso di crescita;

$m$ : rappresenta il parametro di compensazione;

infine, per rendere la funzione continua,  $y_j$  è impostata come:

$$-s_j \delta_j$$

La predizione dei trend futuri è influenzata dall'incertezza che viene modellizzata come segue: il tasso di cambiamenti in un intervallo T del futuro avrà la stessa frequenza e numerosità media di un analogo periodo T nel passato. Aumentando la flessibilità del modello sui dati del passato l'errore nel training set diminuisce in maniera consistente, tuttavia, questa diminuzione di errore si paga nella previsione futura che sarà caratterizzata da un ampio intervallo temporale.

L'incertezza legata a tale intervallo risulta essere un buon indicatore per quanto riguarda il problema dell'overfitting.

Al fine di stabilire il valore delle componenti stagionali si ricorre a funzioni periodiche. Considerando una serie di Fourier standard si stimano i 2N parametri:

$$\beta = [a_1, b_1, \dots, a_N, b_N]^T$$

attraverso la costruzione di una matrice  $X(t)$  di vettori stagionali per ogni istante t per i dati passati e futuri. Supponendo che  $\beta$  sia distribuito normalmente si ottiene:

$$s(t) = X(t)\beta$$

#### 4 Dati

Le fonti dei dati selezionate per il progetto sono le seguenti:

- Università degli studi di Milano-Bicocca
- [www.mercatoelettrico.org](http://www.mercatoelettrico.org)

I dati dei consumi elettrici provenienti dall'ateneo di Milano-Bicocca riguardano gli edifici U1 e U6 e presentano la seguente struttura: Per ciascun edificio sono presenti le rilevazioni riguardanti l'energia assorbita con un intervallo di tempo di 15 minuti. Tali rilevazioni sono raccolte in diversi file

excel, all'interno dei quali sono presenti le osservazioni relative a ciascun mese. I dati a disposizione coprono il periodo compreso tra 1 gennaio 2018 e 31 dicembre 2020.

Gli attributi che caratterizzano i dati sono:

- POD: identificativo del contatore;
- DATA: anno, mese, giorno in cui viene effettuata rilevazione;
- ORA: orario in cui viene effettuata la rilevazione (le rilevazioni sono effettuate ogni 15 minuti a partire dalla mezzanotte del primo giorno del mese);
- FL\_ORA\_LEGALE: indica se le rilevazioni sono State raccolte in un periodo di ora legale oppure solare;
- CONSUMO\_ATTIVA\_PRELEVATA: consumo energetico (espresso in kW) nei 15 minuti considerati;
- CONSUMO\_REATTIVA\_INDUTTIVA\_PRELEVATA: consumo energetico prelevato ma non effettivamente impiegato per produrre lavoro utile;
- POTENZA\_MASSIMA: limite massimo di potenza prelevabile dal contatore (espresso in kW) dalla rete elettrica;
- TIPO\_DATO: identifica la tipologia dei dati.

In seguito ad opportune operazioni di pulizia e di unificazione dei dati si è ottenuto un file excel per ciascun edificio contenente per U1 105405 righe, mentre per i dati relativi all'edificio U6 si sono ottenute 105309 righe entrambi dotati delle 8 colonne precedentemente descritte.

I dati del consumo energetico sono stati integrati con i dati relativi al prezzo dell'energia elettrica. Tale valore è stato ricavato dalle rilevazioni presenti sul sito: [www.mercatoelettrico.org](http://www.mercatoelettrico.org). Le osservazioni sono organizzate come segue: tre file distinti per ciascun anno di rilevazione contenenti 24 colonne e 8761 righe. Ai fini del progetto sono stati selezionati solo i seguenti attributi:

- DATA: anno, mese, giorno in cui viene effettuata rilevazione del prezzo;
- ORA: orario in cui viene effettuata la rilevazione (le rilevazioni sono effettuate ogni ora);
- PUN: prezzo unico nazionale, è il prezzo di riferimento dell'energia elettrica che

viene acquistata sul mercato espresso in MWh.

Da una prima osservazione macroscopica dei dati si nota la perfetta sovrapposizione delle serie storiche relative ai consumi energetici dei due edifici considerati per quanto riguarda il mese di giugno 2020.

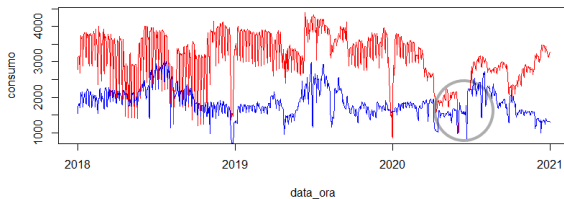


Figura 1: plot serie storiche di U1(blu) e U6(rosso)

Approfondendo il problema si è scoperto che i dati relativi al mese di giugno 2020 di U6 sono stati erroneamente rimpiazzati con i dati di U1 relativi al medesimo periodo.

Proseguendo con l'analisi preliminare dei dati, un'altra problematica emersa è stata quella relativa alla presenza di missing values e di valori duplicati. Tali criticità hanno riguardato le colonne: "CONSUMO\_ATTIVA\_PRELEVATA", "DATA" e "ORA".

Tra i punti di forza delle osservazioni vi è il fatto di avere a disposizione dati con una granularità pari a 15 minuti che ha permesso diversi tipi di aggregazione con la possibilità di fare analisi predittive con intervalli temporali differenti.

## 5 Trattamento dati

### 5.1 Creazione dei due dataframe U1 e U6

La prima fase per quanto riguarda il preprocessing è stata dedicata alla creazione di due dataset contenenti tutte le rilevazioni relative ai due edifici per i tre anni considerati.

Si è operato sui singoli file excel che raccoglievano le osservazioni per ciascun mese effettuando le seguenti operazioni:

- Eliminazione colonne superflue;
- Uniformazione nomi colonne;
- Unificazione di excel diversi che raccoglievano i dati relativi ad un singolo mese;

Svolte le seguenti operazioni si è proceduto alla creazione dei due file excel U1 e U6 sfruttando la funzione rbind disponibile in R.

### 5.2 Preprocessing dati U1

Il primo problema affrontato è stato quello relativo all'uniformazione del numero di cifre all'interno della colonna "ORA". Per risolvere tale problematica è stata creata una funzione ad hoc in grado di leggere il numero di cifre per ogni riga della colonna "ORA" ed in caso di valori minori di 6 cifre provvedesse a completare il campo con le cifre mancanti aggiungendo il valore 0.

Inoltre, per poter trasformare le colonne "DATA" e "ORA" in un formato datetime si è sviluppata una funzione in grado di aggiungere dei caratteri separatori opportuni tra le cifre delle colonne sopra citate.

Conclusa questa operazione si è deciso di unificare in un'unica colonna ("data\_ora") le colonne "DATA" e "ORA".

In seguito, sono state selezionate le colonne di maggior interesse per lo sviluppo del progetto, ovvero: "CONSUMO\_ATTIVA\_PRELEVATA" e "data\_ora".

Successivamente, anche la colonna "CONSUMO\_ATTIVA\_PRELEVATA" è stata oggetto di modifiche finalizzate a renderla omogenea (si è opportunamente arrotondato i valori al primo decimale sfruttando la funzione round di R), inoltre, si è deciso di rinominare la stessa per una maggior interpretabilità.

Grazie alla funzione summary si è riscontrata la presenza di valori pari a 0.0 per quanto riguardava la colonna "CONSUMO\_ATTIVA\_PRELEVATA". Risulta chiaro che questi valori siano privi di senso e che vadano opportunamente trattati. Infatti, si è deciso prima di sostituire i valori pari a 0.0 con NA

e in secondo luogo si è provveduto a gestire tali valori con un'opportuna funzione in grado di sostituire quei valori con la media dell'osservazione precedente e successiva.

Per quanto riguarda la presenza di eventuali duplicati dovuti al passaggio da ora solare a legale sono stati eliminati conservando le rilevazioni corrette.

Conclusa la parte di data cleaning ci si è concentrati sull'aggregazione a livello giornaliero per quanto riguarda i valori relativi alla colonna "CONSUMO\_ATTIVA\_PRELEVATA" opportunamente rinominata: "consumo". Lo scopo dell'aggregazione è stato quello di ottenere per ciascun giorno il consumo elettrico totale espresso in KWh. Per ottenere questo risultato si è deciso di fare la media su base giornaliera e successivamente di moltiplicarla per 24.

Infine, per uniformare il numero di rilevazioni per ogni anno si è deciso di eliminare la rilevazione relativa al 29/02/2020.

### 5.3 Preprocessing dati U6

Per quanto riguarda la parte di data cleaning e di preprocessing le operazioni effettuate sui dati relativi ad U6 sono del tutto analoghe a quelle svolte per U1 eccetto per l'aspetto relativo ai dati del mese di giugno 2020.

Infatti, come precedentemente detto i dati per il mese di giugno dei due dataframe sono identici. Sulla base di indagini svolte si è stabilito che le osservazioni corrette erano quelle appartenenti ad U1.

Per sostituire al meglio le osservazioni relative al mese oggetto di analisi si è scelto di applicare un opportuno modello previsionale in quanto il metodo precedentemente utilizzato nella sostituzione dei valori mancanti sarebbe stato piuttosto impreciso. Ottenuti tali valori predetti, si è provveduto a sostituire i vecchi valori con i nuovi.

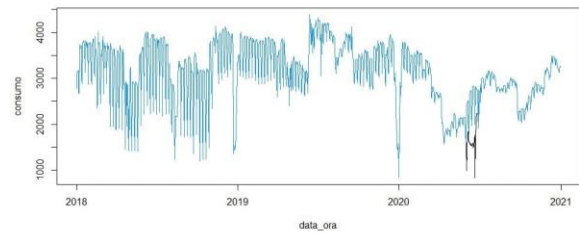


Figura 2: confronto valori predetti con valori originali

## 6 Analisi U1

### 6.1 Analisi preventive

Inizialmente si è deciso di adottare un'aggregazione su base settimanale per quanto riguarda i consumi, in questo modo è stato possibile svolgere delle analisi preventive riguardanti trend e stagionalità dei consumi elettrici dei due edifici dell'ateneo.

Tuttavia, questo tipo di aggregazione non è risultato funzionale al raggiungimento dell'obiettivo preposto.

In seguito a questo "esperimento" si è optato per un'aggregazione su base giornaliera alla quale è stata applicata la funzione decompose implementata in R.

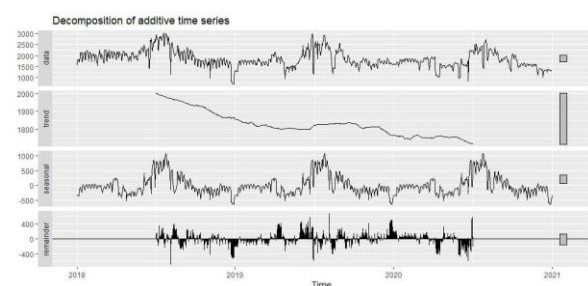


Figura 3: decomposizione trend e stagionalità

Dal grafico soprastante si evince che il trend sia decrescente per gli anni considerati. Per quanto riguarda la stagionalità, risulta evidente che in tutti e tre gli anni vi sia un fisiologico aumento nei consumi energetici per quanto riguarda i mesi estivi dovuto principalmente all'uso dell'aria condizionata.

Inoltre, per verificare la stazionarietà della serie storica si è utilizzato il test Dickey-Fuller che evidenzia la non stazionarietà della serie.

Ad ulteriore riprova del fatto che la serie storica non sia stazionaria si sono osservati i grafici di acf e pacf che hanno confermato la presenza di non stazionarietà e di stagionalità. In particolare, l'andamento degli spike dell'acf mostra una ricorrenza dei picchi ogni 6 lag.

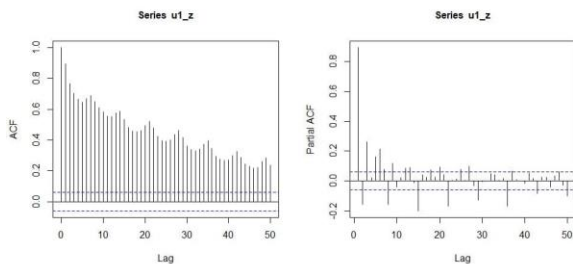


Figura 4: acf e pacf relativi ad U1

## 6.2 Previsioni

Per conseguire l'obiettivo prefissato, si è scelto di separare la serie storica iniziale in due parti distinte:

- Training set: porzione della serie storica compresa tra 1/1/2018 e il 4/10/2020;
- Test set: porzione della serie storica compresa tra il 5/10/2020 e il 20/12/2020.

### ARIMA

Il primo modello previsionale che si è scelto di adottare è il modello ARIMA, il quale privilegia il modello con AIC minore. Nella fattispecie, il miglior modello selezionato da auto.arima (disponibile nel pacchetto forecast), è il seguente:  $ARIMA(3,1,3)(0,1,0)$ .

La non stazionarietà prima citata, è stata correttamente trattata dall'algoritmo auto.arima che ha provveduto ad effettuare una differenziazione per rendere maggiormente stazionario l'andamento del grafico delle rilevazioni. Di seguito il grafico della previsione effettuata per mezzo del modello ARIMA che mostra una leggera sovrastima dei consumi per quanto riguarda le osservazioni del mese di dicembre.

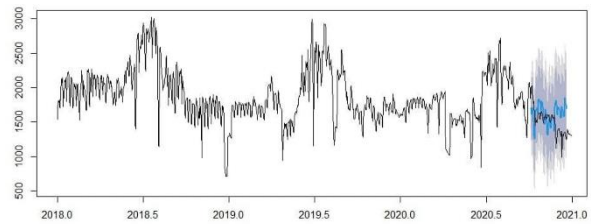


Figura 5: previsione ARIMA U1

Per quanto riguarda l'analisi dei residui, dai grafici è possibile desumere la non normalità degli stessi dovuta alla presenza di valori influenti o outliers presenti nel periodo delle rilevazioni condizionato dal COVID.

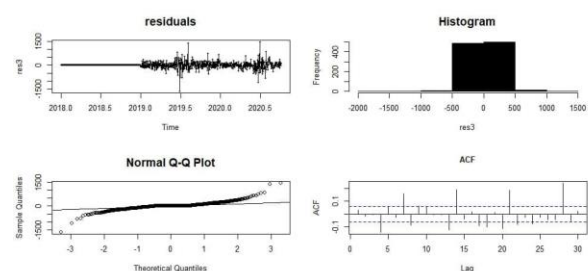


Figura 6: analisi residui ARIMA U1

### TBATS

Il secondo modello previsionale utilizzato è TBATS appositamente scelto per la sua capacità di trattare le stagionalità.

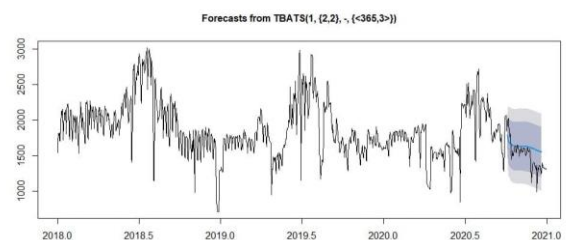


Figura 7: previsione TBATS U1

Dal grafico della previsione si denota come il modello riesca ad interpretare in maniera più precisa, rispetto ad ARIMA, l'andamento del trend della serie storica. Riesce ad intuire con minor efficacia i picchi nei consumi elettrici a causa delle caratteristiche del modello stesso.



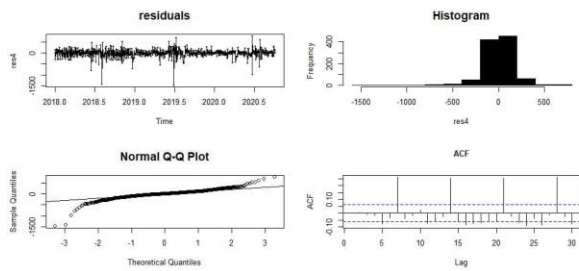


Figura 8: analisi residui TBATS U1

Come per i residui del modello ARIMA anche in questo caso il Q-Q plot indica una non normalità per i residui, in particolare per la coda sinistra. La ragione di tale andamento, come precedentemente detto, sta nella presenza di outliers che condizionano la precisione della previsione.

## PROPHET

Il terzo ed ultimo modello previsionale applicato è PROPHET. La scelta è ricaduta su tale modello poiché particolarmente adatto a trattare dati con frequenza giornaliera. Di seguito è possibile visualizzare l'andamento della previsione effettuata.

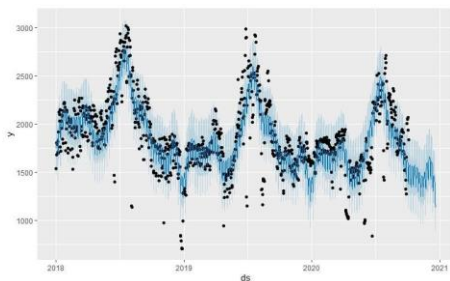


Figura 9: previsione PROPHET U1

La previsione fornita dal modello risulta particolarmente abile nell'individuare i picchi dovuti alla stagionalità. Inoltre, PROPHET interpreta in maniera attendibile il trend della serie storica fornendo una previsione affidabile.

Un'altra funzionalità del modello riguarda la possibilità di sfruttare il comando "prophet\_plot\_components" per mettere in evidenza le tre componenti della serie storica (trend, stagionalità settimanale e stagionalità annuale).

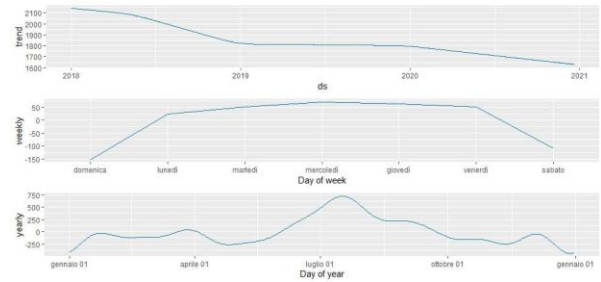


Figura 10: decomposizione trend e stagionalità

Il grafico mostra in maniera evidente quanto già ipotizzato in precedenza, ovvero, l'andamento decrescente nel trend dei consumi dovuto verosimilmente all'effetto COVID-19 e l'aumento degli stessi durante i mesi estivi per far fronte al raffrescamento dell'edificio. Infine, come prevedibile, i consumi elettrici a livello settimanale risultano stabili nei giorni feriali per diminuire sensibilmente nei giorni festivi.

## 7 Analisi U6

### 7.1 Analisi preventive

Analogamente a quanto già visto per le analisi preventive dell'edificio U1 si è scelto di visualizzare attraverso un apposito grafico la scomposizione in trend e stagionalità per quanto riguarda i consumi elettrici dell'edificio U6.

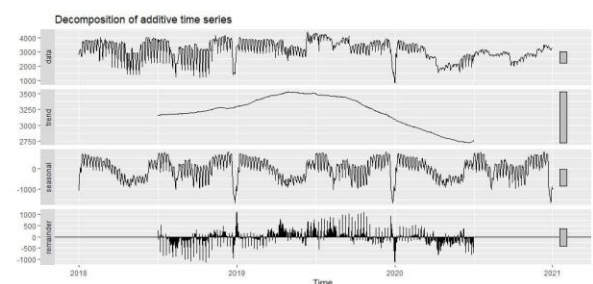


Figura 11: decomposizione trend e stagionalità

Il trend anche in questo caso è decrescente, con una diminuzione nei consumi elettrici soprattutto per l'anno 2020, in particolare per il periodo successivo all'entrata in vigore del lockdown.

Per quanto riguarda la stagionalità si può notare un comportamento meno deciso nell'andamento dei consumi elettrici rispetto alla serie storica relativa ai dati dell'edificio U1 durante il periodo



estivo. Questo aspetto è imputabile al fatto che l'edificio U6 non utilizza energia elettrica per il raffrescamento al contrario di U1.

Anche in questo caso, si è provveduto al controllo della stazionarietà della serie storica per mezzo del test Dickey-Fuller il quale, come per il caso precedente, ha evidenziato la non stazionarietà della serie.

L'analisi di acf e pacf conferma la non stazionarietà dei consumi elettrici e la ricorrenza nei picchi dell'acf ogni 6 lag.

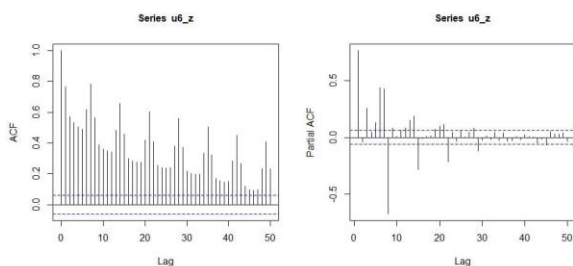


Figura 12: acf e pacf relativi ad U6

## 7.2 Previsioni

Analogamente a quanto svolto in precedenza per l'edificio U1 si è scelto di separare la serie storica iniziale in due parti distinte:

- Training set: porzione della serie storica compresa tra 1/1/2018 e il 4/10/2020;
- Test set: porzione della serie storica compresa tra il 5/10/2020 e il 20/12/2020.

### ARIMA

In questo caso il miglior modello selezionato è il seguente: ARIMA (2,1,3) (0,1,0). La previsione che segue è effettuata sulla base del modello selezionato da auto.arima.

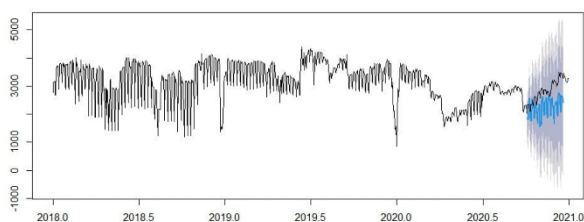


Figura 13: previsione ARIMA U6

A differenza di quanto visto in U1 in questo caso il modello tende a sottostimare i consumi energetici cogliendo però il trend crescente degli stessi.

L'analisi dei residui non è positiva in quanto la normalità non è verificata da Q-Q plot.

### TBATS

Il secondo modello applicato, anche in questo caso è TBATS. Tuttavia, i risultati, come è possibile notare dal grafico sottostante, non sono affidabili.

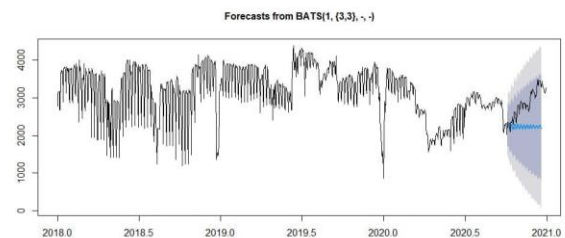


Figura 14: previsione TBATS U6

Nonostante lo studio dei residui sia più positivo rispetto a quelli individuati da ARIMA rimane comunque evidente il problema legato alla normalità degli stessi.

### PROPHET

Il terzo modello adottato è quello che fornisce i risultati migliori sotto il punto di vista della precisione della previsione. Questo aspetto non sorprende vista l'abilità di PROPHET a prevedere dati giornalieri.

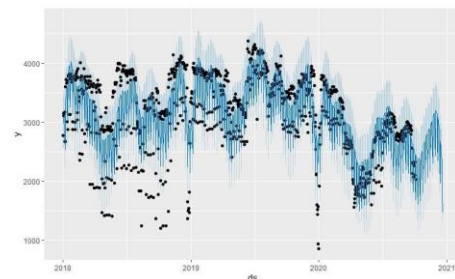


Figura 15: previsione PROPHET U6

Inoltre, il modello non subisce particolarmente il condizionamento degli outliers che sono presenti in maniera importante all'interno delle osservazioni di U6.

Per quanto riguarda la decomposizione l'aspetto maggiormente interessante che emerge è quello relativo all'aumento dei consumi nei mesi estivi che non è così accentuato come per l'edificio U1. La causa di questa discrepanza sta nella diversità delle fonti energetiche utilizzate per il raffrescamento degli edifici.

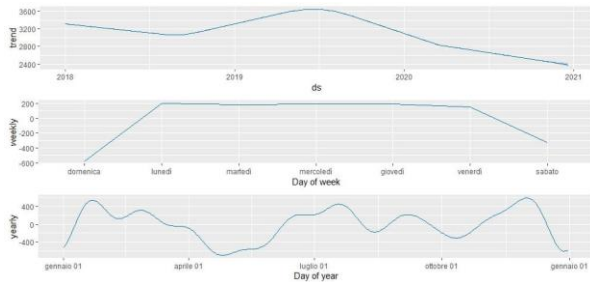


Figura 16: decomposizione trend e stagionalità

## 8 Prezzi energia elettrica

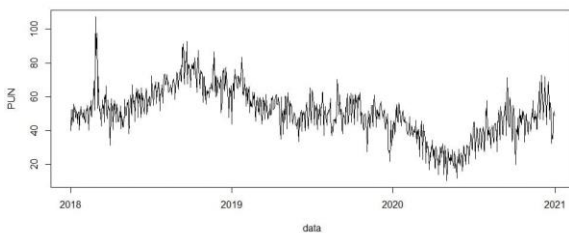


Figura 17: serie storica prezzo energia elettrica

### 8.1 Preprocessing

Per quanto riguarda il preprocessing relativo ai prezzi dell'energia elettrica si è deciso di unire i dati provenienti dai file excel relativi ai singoli anni in un unico file.

Da questo file sono state eliminate tutte le colonne ad eccezione di quelle relative alla data dell'osservazione e del PUN (prezzo unico nazionale).

Per favorire l'integrabilità con i dataset relativi ai consumi energetici si è deciso di creare una funzione con l'obiettivo di adattare il formato della data a quello dei dataset relativi agli edifici U1 e U6.

### 8.2 Analisi

Anche in questo caso il primo aspetto considerato è stato quello relativo alla stazionarietà della serie storica. Prevedibilmente i prezzi non hanno un andamento stazionario come dimostrato dal risultato del test Dickey-Fuller. Acf e pacf confermano l'andamento non stazionario della serie.

### 8.3 Previsioni

L'iter seguito per la scelta di un modello previsionale per i prezzi dell'energia elettrica è stato il medesimo di quello adottato per U1 ed U6.

Alla luce dei risultati ottenuti il modello migliore nelle previsioni è PROPHET.

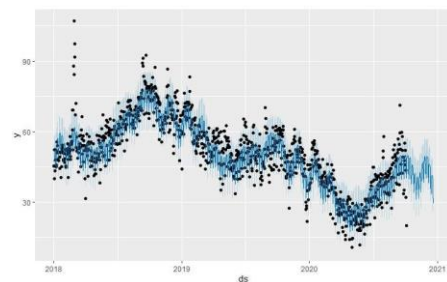


Figura 18: previsione PROPHET prezzi energia

Ancora una volta tale modello si conferma poco sensibile alla presenza di outliers garantendo una stima maggiormente accurata rispetto ai modelli ARIMA E TBATS.

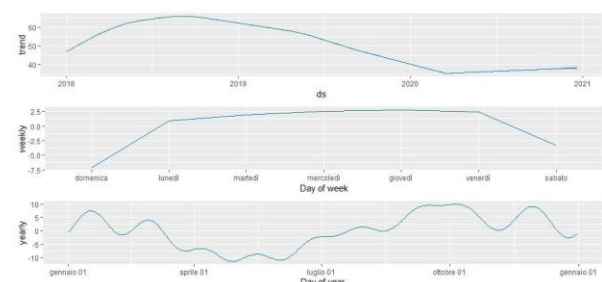


Figura 19: decomposizione trend e stagionalità

Per concludere, è interessante notare come l'andamento dei prezzi dell'energia subisca una notevole flessione in corrispondenza dell'inizio del periodo pandemico.

## 9 Risultati

Per poter selezionare il modello più opportuno per le finalità del progetto si è utilizzata una metrica in grado di tener conto dell'errore di previsione.

La scelta è ricaduta sul MAPE (Errore percentuale medio assoluto):

$$MAPE = \frac{\sum \frac{|A-F|}{A} \times 100}{N}$$

- A: valori reali
- F: valori predetti
- N: numero delle osservazioni

Si tratta della media aritmetica dei rapporti tra il valore assoluto degli errori di previsione e la domanda effettivamente verificatasi. Di seguito si illustrerà il MAPE relativo a ciascun modello per ogni edificio.

	U1	U6
ARIMA	16.12%	19.78%
TBATS	11.82%	19.91%
PROPHET	10.40%	10.41%

Alla luce dei risultati appena esposti si evince che il miglior modello predittivo è PROPHET che garantisce un errore previsionale nettamente inferiore rispetto agli altri modelli per quanto riguarda la previsione di valori con frequenza giornaliera.

Per quanto riguarda la previsione del prezzo dell'energia elettrica l'unico modello utilizzabile risulta essere PROPHET in quanto ARIMA e TBATS non forniscono stime accurate. In particolare, il valore del MAPE per PROPHET è di 15.93%.

## 10 Dashboard

Per rendere maggiormente fruibile il risultato dell'analisi agli organi amministrativi dell'università si è deciso di sviluppare un'opportuna dashboard sfruttando le potenzialità del software Tableau.

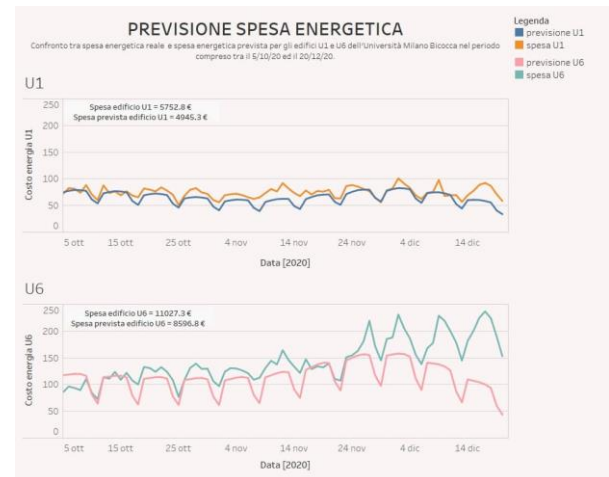


Figura 20: dashboard spesa energetica

Per ottenere il valore della spesa energetica prevista per il periodo oggetto di studio si è moltiplicato il valore delle previsioni dei consumi per il valore del prezzo dell'energia elettrica previsto.

Per fornire un'indicazione complessiva riguardante la spesa energetica si è aggiunto un riquadro contenente il totale della spesa reale con il totale della spesa prevista per il periodo considerato.

## 11 Conclusioni

L'obiettivo di questo progetto è stato quello di individuare un budget da allocare per la spesa energetica futura degli edifici U1 e U6. I risultati ottenuti sono soddisfacenti, in particolar modo per U1, considerando l'ampiezza dell'intervallo temporale di previsione.

L'intento del progetto è quello di fornire una stima per il valore complessivo della spesa prevista che sia indicativa per gli organi amministrativi preposti alla gestione delle spese dell'ateneo.

È da vedersi in quest'ottica la scelta di un periodo di previsione così ampio, in grado di garantire informazioni su una finestra temporale di medio-lungo termine a discapito della precisione della stima.

Questo aspetto emerge in particolar modo per quanto riguarda il mese di dicembre 2020 relativo all'edificio U6.

La risoluzione di alcune criticità potrebbe garantire migliori risultati dal punto di vista della precisione della previsione. In particolare, la riduzione dell'intervallo temporale di previsione garantirebbe una miglior accuratezza dei modelli, sarebbe inoltre necessario interfacciarsi con la ragioneria dell'università per comprendere le loro esigenze; Infine, risulterebbe utile la possibilità di essere affiancati da esperti di dominio per la gestione degli outliers e per conoscere la quantità e il tipo di utenze presenti in ciascun edificio.

Potrebbe essere interessante estendere il progetto all'intero patrimonio immobiliare dell'ateneo per avere una visione complessiva delle spese relative ai consumi energetici dell'Università di Milano-Bicocca.

## 11 Fonti

[www.mercatoelettrico.org](http://www.mercatoelettrico.org)

<https://www.unimib.it/ateneo/bicocca-sostenibile/aree-lavoro-e-progetti/energy>

<https://otexts.com/fpp2/arima.html>

<https://medium.com/future-vision/the-math-of-prophet-46864fa9c55a>

[https://elearning.unimib.it/pluginfile.php/1010112/mod\\_resource/content/1/Fundamentals%20of%20Time%20Series%20Analysis%20for%20the%20Working%20Data%20Scientist%20-%20Draft.pdf](https://elearning.unimib.it/pluginfile.php/1010112/mod_resource/content/1/Fundamentals%20of%20Time%20Series%20Analysis%20for%20the%20Working%20Data%20Scientist%20-%20Draft.pdf)

<https://ibf.org/knowledge/glossary/mape-mean-absolute-percentage-error-174>

[https://public.tableau.com/app/profile/giovanni4739/viz/Dashboard\\_Previsione\\_Spesa\\_Energetica/Dashboard1?publish=yes](https://public.tableau.com/app/profile/giovanni4739/viz/Dashboard_Previsione_Spesa_Energetica/Dashboard1?publish=yes)