Journal of Cleaner Production 435 (2024) 140265

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

Machine learning models for short-term demand forecasting in food catering services: A solution to reduce food waste

Miguel Rodrigues [a,*], Vera Miguéis [a], Susana Freitas [b], Telmo Machado [b]

[a] Faculdade de Engenharia da Universidade do Porto, INESC TEC, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
[b] LIPOR, R. Morena 805, 4435-996 Baguim do Monte, Portugal

ARTICLE INFO

Handling Editor: Cecilia Maria Villas Bôas de Almeida

Keywords:
Food waste
Machine learning
Food catering services
Sustainability
Demand forecast
Waste reduction

ABSTRACT

Food waste is responsible for severe environmental, social, and economic issues and therefore it is imperative to prevent or at least minimize its generation. The main cause of food waste is poor demand forecasting and so it is essential to improve the accuracy of the tools tasked with these forecasts. The present work proposes four models meant to help food catering services predict food demand accurately and thus avoid overproducing or underproducing. Each model is based on a different machine learning technique. Two baseline models are also proposed to mimic how food catering services estimate future demand and to infer the added value of employing machine learning in this context. To verify the impact of the proposed models, they were tested on data from the three different canteens chosen as case studies. The results show that the models based on the random forest algorithm and the long short-term memory neural network produced the best forecasts, which would lead to a 14% to 52% reduction in the number of wasted meals. Furthermore, by basing their decisions on these forecasts, the food catering services would be able to reduce unmet demand by 3% to 16% when compared with the forecasts of the baseline models. Thus, employing machine learning to forecast future demand can be very beneficial to food catering services. These forecasts can increase the service level of food services and reduce food waste, mitigating its environmental, social, and economic consequences.

1. Introduction

Food loss and waste have severe environmental, social, and economic impacts, such as the production of greenhouse gases, global food security problems, and related costs, respectively (Girotto et al., 2015; Kamble et al., 2020; Martin-Rios et al., 2018). Considering that, in 2011, it was estimated that one-third of all food produced is either lost or wasted, which translates into more than a billion tonnes of food, the impact of food loss and waste is immeasurable (FAO, 2011).

Food loss and waste occur throughout the entire food supply chain. Food loss (FL) happens at an early stage of the food supply chain (agricultural production), and food waste (FW) happens at the final consumer stage (Priefer et al., 2016; Messner et al., 2020). FW has the worst consequences of the two, since at this stage, the food has already been processed, packaged, stored, and possibly distributed (San Martin et al., 2021; Martin-Rios et al., 2021). However, food service companies are not actively innovating in the waste domain (Martin-Rios et al., 2018), despite their owners, chefs, and kitchen managers demonstrating a general willingness to reduce FW (Hennchen, 2019). It is, therefore, both urgent and imperative to find ways to reduce FW and to extract value from FW already generated, and then provide this information to such stakeholders.

There are several options when it comes to extracting value from FW. For example, this can be done by motivating and providing the means for customers to take what remains of their meals home (DRAAF, 2014). This can also be achieved by transforming FW into a valuable resource, e.g., biofuel, bioenergy, and biomaterials (Filimonau and De Coteau, 2019; Girotto et al., 2015; Priefer et al., 2016). The most popular techniques to achieve this are biogas production (De Clercq et al., 2019; Curry and Pillay, 2012) and anaerobic digestion (Dai et al., 2013; Li et al., 2018).

Although being able to extract value from FW is essential, preventing its generation is still the most desirable approach to minimize its consequences (Papargyropoulou et al., 2014; Messner et al., 2020). The literature presents different approaches to prevent FW, such as through the use of packaging technologies (Brennan et al., 2021) and data gathering technologies (Martin-Rios et al., 2021). Nevertheless, FW generation is usually caused by food overproduction, which in turn is mostly due to poor demand forecasts (Filimonau and De Coteau,

* Corresponding author.
E-mail addresses: up202003371@up.pt (M. Rodrigues), vera.migueis@fe.up.pt (V. Miguéis), susana.freitas@lipor.pt (S. Freitas), telmo.machado@lipor.pt (T. Machado).

https://doi.org/10.1016/j.jclepro.2023.140265
Received 3 August 2023; Received in revised form 24 November 2023; Accepted 15 December 2023
Available online 20 December 2023

2019; Parfitt et al., 2010; Pirani and Arafat, 2016). Most companies in the hospitality sector do not trust the accuracy of their forecasts, and therefore cook 10% more food than their estimated needs in order to avoid stock-outs (Pirani and Arafat, 2016). Therefore, a major effort should be made to develop more accurate and reliable demand forecasting models.

Food demand forecasting models based on machine learning algorithms (such as the ones proposed in the present paper) have the potential to improve waste management in the HoReCa (Hotel, Restaurant, and Café) sector with their technological innovation (Martin-Rios et al., 2021), providing businesses with a tool capable of contributing to the minimization of FW generation while also preventing underproduction. By knowing the quantities of food they need to produce/serve in advance, businesses can minimize overproduction and thus FW.

The present paper proposes four food demand forecasting models that can be divided into two categories. The first category comprises two casual models, one based on a random forest (RF) algorithm and the other based on the light gradient-boosting machine (LightGBM) algorithm. The second category comprises two time series models, one based on a long short-term memory (LSTM) neural network and the other based on a transformer neural network. Given the topic explored in the present study, the casual models should outperform the time series models.

Causal models adopt a set of independent variables in order to better understand the environment at hand and increase their forecasting capabilities. Within the context of the present paper, these variables could be the menu, the number of reservations, and date-related variables such as the day of the week, day of the month, and month of the year. On the other hand, most time series models rely only on the previous values of a target variable and its patterns, trends, and cyclicality to forecast future values. Some examples of these models are the ones based on transformer neural networks or the Holt–Winters algorithm. There are however some exceptions that also take advantage of external variables to help forecast future values, such as models based on LSTM neural networks and autoregressive integrated moving average with exogenous variables (ARIMAX) models.

The performance of food demand forecasting models is usually measured and analyzed in relation to the quality of their forecasts. In theory, the closer the forecast demand values are to the real demand, the better the model. It is important to note that in most cases, these models not only seek to minimize FW but also avoid underproduction so as to help the catering services meet their financial objectives (Messner et al., 2020).

This study is structured as follows. First, a review of the literature on FW is presented, exploring both the quantitative and predictive branches. Second, the methods adopted in the present study are described, and the data collected from each FCS is presented. Then, using data from each FCS, the machine learning-based forecasting models are compared with the baseline models to determine how much FW could have been prevented. The key conclusions from the results are then pointed out and discussed, providing insights and guidelines for future work.

2. Literature review

The literature on FW can be divided according to two types of approach: qualitative and quantitative. As the names suggest, the former focuses on quantifying FW, while the latter highlights the drivers of FW, how to prevent it, and how to allocate it when it occurs.

The qualitative branch is the most popular of the two in terms of the number of studies, which have remained consistent throughout the years (Hebrok and Boks, 2017; Dhir et al., 2020). In 2010, Parfitt et al. (2010) reported a lack of descriptive analysis studies and pointed out that the few that existed used data mainly from the 1980s. In terms of quantitative studies, Schneider (2013) identified four (focusing on the hospitality sector) in 2013, Filimonau and De Coteau (2019) identified 47 in 2019, Dhir et al. (2020) identified 41 in 2020, and Rodrigues et al. (2023) identified 65 in 2022.

Although it is hard to quantify the exact number of existing works, the increasing popularity of the quantitative approach to FW is clear. Rodrigues et al. (2023) studied the methods and data used in the studies published in the field and classified the studies as descriptive, predictive, or prescriptive. Of these three types of analysis, predictive analysis is particularly important when it comes to preventing FW generation. Accurate forecasts mean that FCSs can predict future demand and adapt their production accordingly. When poor demand forecasts are produced, they motivate overproduction, leading to FW (Filimonau and De Coteau, 2019; Parfitt et al., 2010; Pirani and Arafat, 2016).

Demand forecasting can be considered one of many means to prevent FW. Martin-Rios et al. (2021) explored FW management innovations in the food service industry and highlights the focus on providing faster responses to market/customer demands. Innovation theory is just one of the angles from which to tackle FW management, be it by preventing, reducing, or recycling it (Martin-Rios et al., 2018). Another approach to minimizing FW is practice theory. By focusing on the knowledge/competence component, it is possible to identify FW prevention related issues and try to derive alternatives to reduce FW (Hennchen, 2019). Messner et al. (2020) explored the issue from a different angle and presented a paradoxical idea of FW prevention. While the most promising option to reduce the impact of FW is its prevention (Papargyropoulou et al., 2014), FW prevention actions taken by industry and governments can sometimes be distorted to mean management rather than prevention, thus promoting structures that manage rather than reduce food waste. Messner et al. (2020).

Most of the studies tackling future demand forecasts adopt time series models. These models identify seasonalities and trends to estimate future demand. The algorithms most commonly used in these models are the autoregressive integrated moving average (ARIMA) algorithm and its variations (Møller Christensen et al., 2021; Sakoda et al., 2019), the trigonometric seasonality, Box–Cox transformation, ARMA errors, trend and seasonal components (TBATS) algorithm (Dharmawardane et al., 2021; Posch et al., 2021), naïve algorithms (Møller Christensen et al., 2021; Malefors et al., 2021b), and the Prophet algorithm (Posch et al., 2021; Malefors et al., 2021b). These kinds of models usually benefit from long data collection periods, such as the ones available to Arunraj and Ahrens (2015) and Dharmawardane et al. (2021), which ranged from four to ten years.

In addition to studies using time series models, some studies employ casual models that use external variables to better understand the environment at hand and, in theory, produce better forecasts. Different algorithms are used in this context, with neural networks being the most popular (Lee et al., 2020; Faezirad et al., 2021; Sakoda et al., 2019). Other, less popular algorithms also adopted in this field are the RF algorithm (Malefors et al., 2021a) and the extreme learning machine algorithm (Gružauskas et al., 2019). Regarding the external variables adopted in the causal models, they mostly concern the number of reservations and time-related variables (day, month, year, and holidays). Other, less common external variables include the number of Covid-19 infections (Malefors et al., 2021a), weather-related variables (Arunraj and Ahrens, 2015), and the menu served on a given day (Faezirad et al., 2021). It should be noted that food demand forecasting studies generally estimate demand daily, with most performing a next-day forecast (Sakoda et al., 2019; Malefors et al., 2021b; Møller Christensen et al., 2021).

Most of the studies available in the field do not evaluate how the proposed forecasting models impact FW reduction. In fact, to the best of the authors' knowledge, only Dharmawardane et al. (2021), Lee et al. (2020) and Arunraj and Ahrens (2015) verified the added value of the proposed models by comparing them with a baseline model. The first study compared a baseline model based on the Holt–Winters algorithm with the developed TBATS model (Dharmawardane et al., 2021) using data from two different settings. The developed model beat the baseline model in the first setting but lost in the second. The second study adopted an LSTM-based model that achieved a higher F1 score than a baseline model based on the collaborative filtering technique (Lee et al., 2020). Lastly, the third study employed four models. Three ARIMA variants (SARIMA-MLR, SARIMA-QR, and SARIMA) and a multi-layer perceptron (MLP) neural network that produced a lower root mean square error (RMSE) and mean average percentage error (MAPE) than a baseline model based on a seasonal naïve algorithm (Arunraj and Ahrens, 2015). Although there are other studies that mention improvements over a baseline model (Malefors et al., 2021a,b; Gružauskas et al., 2019; Faezirad et al., 2021), even quantifying these improvements (albeit not always via error measures), they do not mention the techniques used to produce these baseline models.

In short, the use of forecasting models (such as those based on machine learning algorithms) to reduce FW is a topic that is still in an early stage of development. There is a need for further studies on this topic, particularly with a focus on causal models that include more diverse variables and on exploring the potential of advanced machine learning algorithms. Moreover, more studies should evaluate how the proposed models impact FW reduction in different settings. The present paper proposes to do exactly that. It contributes to the literature by proposing a set of food demand forecasting models that include a diverse set of variables and are based on advanced machine learning algorithms that have yet to be explored in this context, and it innovates by testing the added value of the proposed models on data from three different FCSs.

3. Methodology and data

The present study seeks to quantify how beneficial food demand forecasting can be for the food catering sector. To that end, four food demand forecasting models were developed, i.e. two causal models and two time series models. Each model was based on a different machine learning algorithm, and all models were designed to predict demand in the short future (next-day forecasts). The forecasts produced by these models should aid managers in determining the quantity of food to prepare and the resources that should be available/allocated to meet the estimated demand.

To validate the applicability and potential of the proposed food demand forecasting models, these were tested on data from three food catering services (FCSs): two student canteens and a company canteen. Since the three units have different inherent features, the goal is also to explore how the different algorithms may impact the quality of the forecasts. The data gathered from these FCSs allow the use of an entire year to test the models' ability to forecast the food demand and consequently reduce FW. To quantify the theoretical FW reduction provided by the proposed models and thus evaluate their added value, they were compared with two baseline models designed to mimic how FCSs typically estimate future demand.

In the case of the causal models and the model based on the LSTM neural network, external variables were specified. These variables should be able to represent the environment of the different FCSs and the day for which the demand is being forecast. Some of these external variables were collected by the FCSs directly, e.g., the number of reservations and the menu for the day. In contrast, other variables such as weather data were extracted from alternative databases. The variables considered for each FCS are detailed in the following section. As for the model based on the transformer neural network, it was not necessary to specify any variables other than the demand.

The hyperparameters of the different employed algorithms were optimized through nested cross-validation. Additionally, by comparing the predicted demand with the actual demand recorded by the FCSs, it was possible to compute different metrics that quantify the forecasting performance of the proposed models. Then, by analyzing these metrics, it was possible to determine the model with the best performance, i.e., that can lead to the highest reduction in FW.

Nevertheless, to assess the impact and infer the added value of the proposed models, the produced demand forecasts should be compared with the demand estimated by the decision-makers at the FCSs used as case studies. However, since these decision-makers did not make such registers, two baseline models designed to mimic how they might have forecast the demand were used instead.

The following subsections provide a detailed description of the methodology and data used. First, a more in-depth description of the data available for each FCS is provided, indicating the time horizon and the variables adopted in the models. Next, the models proposed are introduced, the implementation of these models is explained, and the hyperparameter tuning process is presented. Lastly, the baseline models proposed are described in detail.

3.1. Data available

Each FCS provided data related to a different set of factors that were unique to their environments. It was also possible to extract and generate other variables that were not directly related to the FCS in question. Some examples of the latter include variables related to the date for which the demand is being forecast, such as the day of the week, the month, whether it is a holiday, or whether it is during the Queima week (a week-long traditional academic festival that marks the end of the academic year for university students and during which there are no classes), and variables related to the weather predicted for the same day, such as temperature and precipitation. The weather-related variables were extracted from the Visual Crossing database (Crossing, 2003).

3.1.1. Food catering service 1 - Student canteen

One of the FCSs used as case studies operates a student canteen and will henceforward be referred to as FCS1. This canteen serves an engineering school with over 8000 students. Since this school is quite large, besides the canteen operated by FCS1, two other canteens serve the students to ensure their needs are met. Nonetheless, FCS1 still records an average of 400 daily meals served.

The data available for FCS1 includes historical demand by dish typology (such as meat, fish, and vegetarian) and the daily number of students who have classes only in the morning, only in the afternoon, or in both periods. This data refers to the period from January 2017 to December 2019. The number of daily meals served by dish typology is displayed in Fig. 1. Looking at the y-axis, it is apparent that the popularity of the three dishes is very different.

However, their distribution throughout the year seems to be relatively similar and constant. By plotting the sum of the demand of the three dishes, as seen in Fig. 2, this distribution becomes clearer. First, a constant number of meals served throughout the year is depicted via the blue line, as the demand was fitted to it, showing almost no slope. Second, there is a clear drop in the number of meals served at the end of June, only to return to normal levels in October. This is due to students no longer having classes at the end of May, even though some students still have exams and therefore still attend the university. Additionally, this is also a time when both students and university staff are on vacation, which further drives the demand down. Although students do not have classes at the end of the year, this period is much shorter and its effects are not felt as much, except for the last two weeks of December due to the Christmas and New Year's weeks. Finally, the drop in demand around the first week of May is due to this being the Queima week.

Regarding the number of daily meals served, these are better displayed in Fig. 3, where the boxplots for each dish are plotted side by side. Although there are many types of dishes, it is apparent that the diet, extras, and salad dishes are not as popular as the rest. In fact, there were days when no such dishes were served at all. On the other

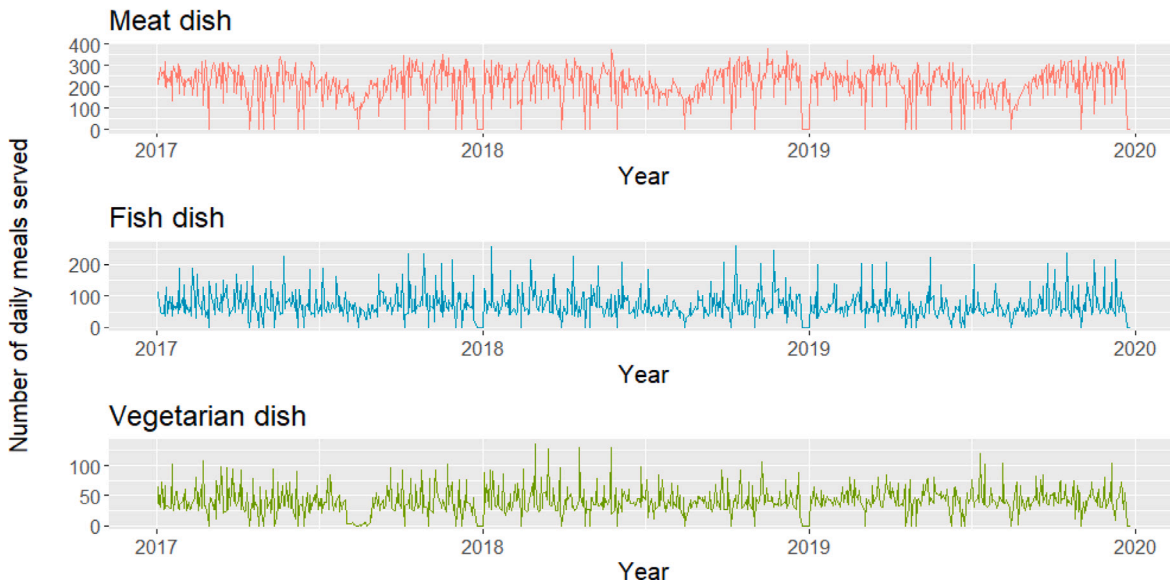

Fig. 1. Number of daily meals served by FCS1 separated by dish.

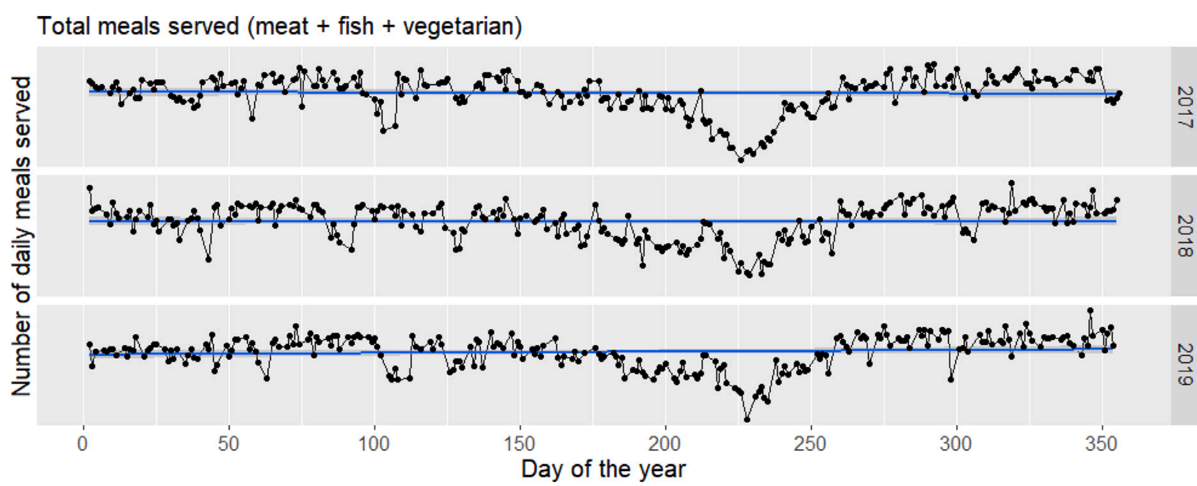

Fig. 2. Number of daily meals served by FCS1.

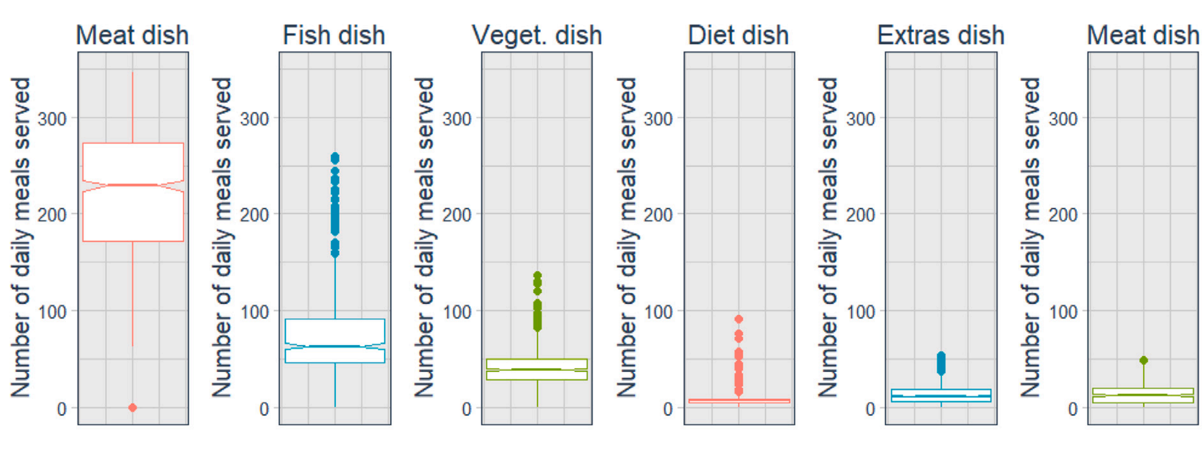

Fig. 3. Number of daily meals served by FCS1 separated by dish.

hand, meat, fish, and vegetarian dishes are significantly more popular. In particular, the meat dish was served more times than the fish and vegetarian dishes combined. While the meat dish only has bottom outliers, i.e., values lower than 1.5 times the Interquartile Range (IQR)

Table 1
Summary of the variables available for FCS1.

Variable name	Type	Description
Dish_meat; Dish_fish; Dish_vegetarian; Dish_diet; Dish_extra; Dish_total; Dish_soups	Continuous	Daily number of meals served for the meat, fish, vegetarian, diet, extras, and soup dishes, respectively
Students_Morning; Students_Afternoon; Students_AllDay	Continuous	Daily number of students who have classes in the morning, afternoon, or all day
Day	Discrete	Day of the month
Week_Monday; Week_Tuesday; Week_Wednesday; Week_Thursday	Binary	Day of the week
Month_January; Month_February; …; Month_October; Month_November	Binary	Month of the year
Holiday	Binary	Holiday or not
HolidayAfter1; HolidayAfter2; HolidayAfter3	Binary	Holiday in the next day, in two days, or in three days, respectively
HolidayBefore1; HolidayBefore2; HolidayBefore3	Binary	Holiday in the previous day, two days ago, or three days ago, respectively
Temperature	Continuous	Average temperature for the day, measured in Celsius
Precipitation	Continuous	Average precipitation for the day, measured in millimeters
Weather_Condition_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Rain_Overcast; Weather_Condition_Rain_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Wind	Binary	Weather condition

Time horizon: January 2017 to December 2019.

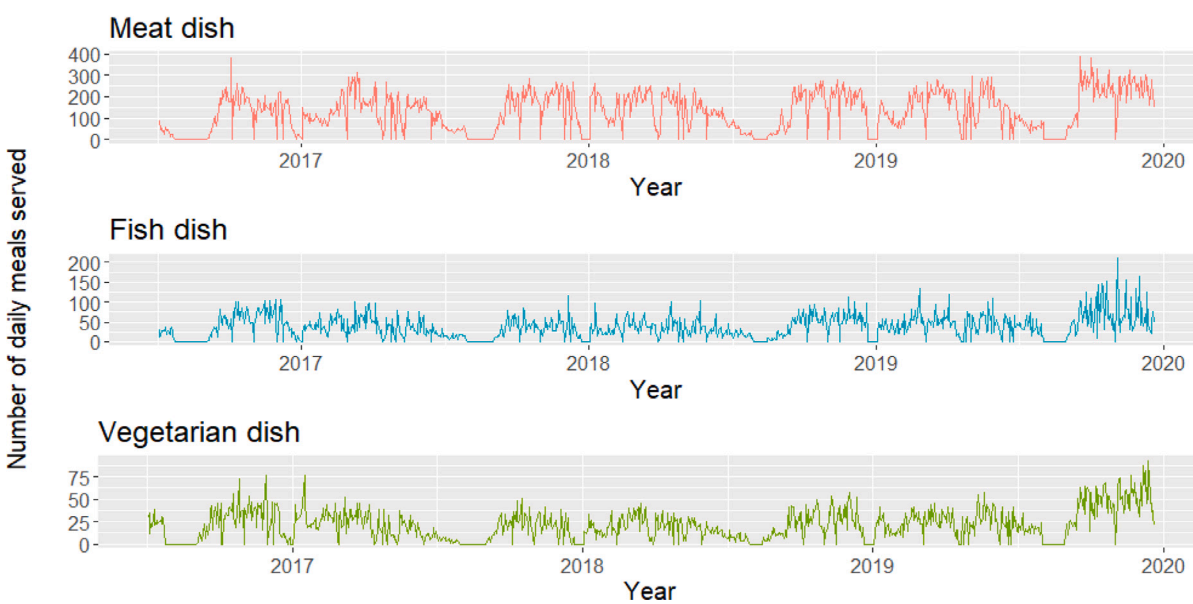

Fig. 4. Number of daily meals served by FCS2 separated by dish.

from the bottom hinge (the 25th percentile), the fish and vegetarian dishes have many upper outliers, i.e., values higher than 1.5 * IQR from the top hinge (the 75th percentile), alluding to a more irregular distribution of the demand for the latter.

In addition to the historical demand by dish typology and the daily number of students with classes, date and weather-related variables were also included in the models proposed for FCS1. In the case of holidays, not only was it noted whether the day was a holiday, but also if the next or previous three days were holidays as well. Table 1 provides more details about the available data.

3.1.2. Food catering service 2 - Student canteen

The second FCS (FCS2) also operates a student canteen but in a different setting. In particular, it serves around 4000 students from a medical school. Although it is the only canteen in its university, there are other FCSs nearby. This explains why FCS2 only serves around 225 daily meals, a small portion of the total population it could serve. The competition around FCS2 includes the canteen from a neighboring hospital and several restaurants (including fast food restaurants) in a nearby shopping mall.

The data provided by FCS2 includes the historical demand by dish typology (meat, fish, or vegetarian) and the menu for each day between July 2016 and December 2019. The menus were subsequently categorized according to the ingredients used, e.g., type of meat and type of fish, and how they were cooked, e.g., grilled and stewed.

The daily number of meals served varies according to the dish typology, as displayed in Fig. 4. The meat dish is also the most popular in the context of FCS2, although with a volatile distribution. When plotting the sum of the demand for the three dishes, the trend that is drawn is similar to that of FCS1's. However, the drop in demand from June to October is more accentuated (see Fig. 5). As with FCS1, this drop is likely due to students not having classes during this period and also because it is a vacation period for the university.

The meat dish represents more than half of all dishes served. However, this FCS operates on a smaller scale than the previous one. This,

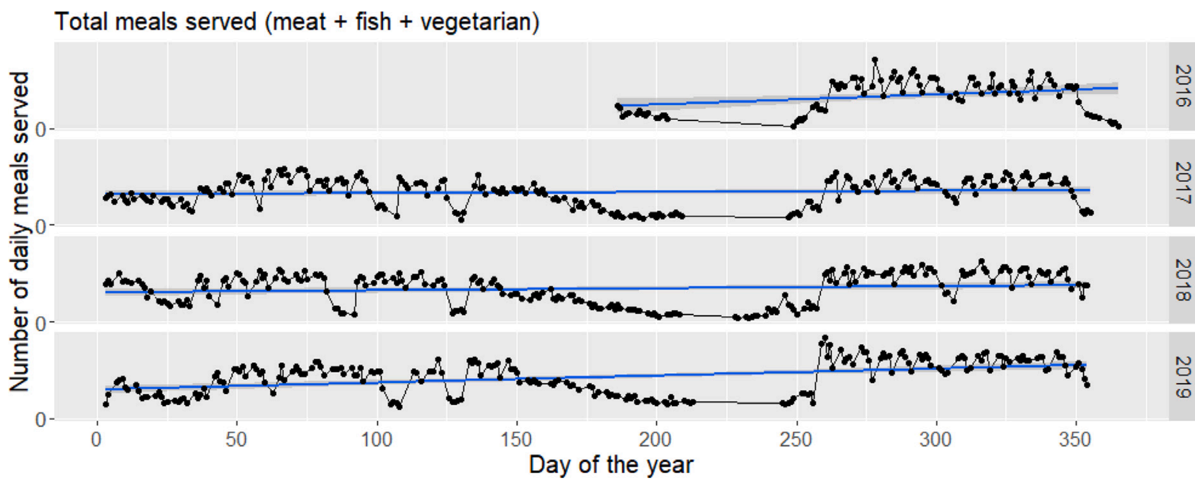

Fig. 5. Number of daily meals served by FCS2.

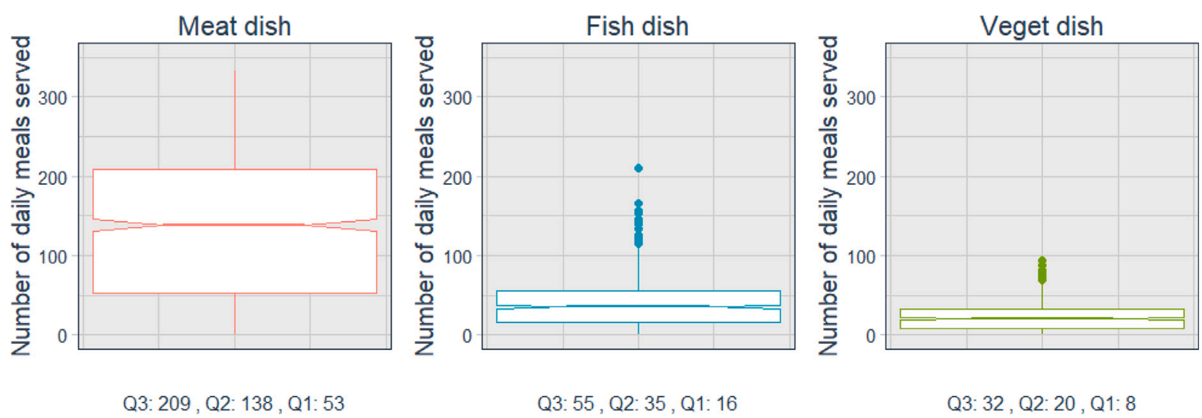

Fig. 6. Number of daily meals served by FCS2 separated by dish.

allied with a more volatile distribution, leads to the number of meat meals served not having any lower outliers when plotting its values in a boxplot (Fig. 6). As for the fish and vegetarian dishes, these still have upper outliers.

Other information taken into account includes the day of the month, the day of the week, and the month of the day for each recorded observation. Additionally, it was noted whether that day and the three next or previous days were holidays. The temperature and precipitation were also considered to help forecast future demand values. Table 2 provides a more comprehensive overview of the available variables.

3.1.3. Food catering service 3 - Company canteen

The final FCS (FCS3) used as case study is different from the other two because it serves the workers of a company instead of students. This is an engineering and technology company with around 3700 employees in its headquarters. Unlike the other FCSs, this one operates a canteen that serves a significant portion of the company's workers, handling around 1900 daily meals.

The data provided by FCS3 contains the number of meals served by type of dish (meat, fish, or diet), as well as reservations for each of the dishes, which are announced two days in advance. The available records refer to the period between September 2011 and December 2019. Therefore, this is the FCS with the most prolonged data collection horizon.

Additionally, the average number of meals served daily by FCS3 is significantly higher than the two other FCSs, as seen in Fig. 7. In this figure, it is possible to see that there is a growing trend in the number of meat meals served throughout the years, while the same is not as obvious for the fish and diet dishes.

The overall demand trend of FCS3 becomes more evident when plotting the total number of meals served, as presented in Fig. 8. In general, demand remained constant throughout the years, with some dips presumably due to holidays or other extraordinary occurrences. As with the other two school canteens, demand was lower in the summer. Although less accentuated, there were also days with no demand whatsoever, which may symbolize that the company, canteen, or both were closed for vacations.

Just like in the other two canteens, the meat dish was the most consumed, being served significantly more than the other dishes. Possibly due to its popularity, there is only a single outlier in the number of meat dishes served. This can be visualized in Fig. 9. While the fish and diet dishes were served less often than the meat dish, there were days when demand for the former peaked very high. In the case of salad dishes, the median values of meals served were higher than the previous two dishes, but the demand for salads was more constant and did not reach such high peaks. Finally, the vegetarian option seems to be the least popular in this company.

As with the previous FCSs, date and weather-related variables were also considered for FCS3. Additionally, the number of reservations for a given day was available in the case of FCS3 (the employees of this company need to show interest in the dish they want two days in advance, although they can still attend the canteen without a reservation). A summary of the variables available for FCS3 is displayed in Table 3

The available data relates to the period from September 2011 to December 2019.

Table 2
Summary of the variables available for FCS2.

Variable's name	Type	Description
Dish_meat; Dish_fish; Dish_vegetarian	Continuous	Daily number of meals served for the meat, fish, and vegetarian dishes, respectively
Meat_pork; Meat_beef; Meat_chicken; Meat_roast; Meat_grilled; Meat_stew; Meat_boiled; Meat_fried; Meat_others	Binary	Meat menu classification
Fish_forkbeard; Fish_tuna; Fish_codfish; Fish_horsemackerel; Fish_seafood; Fish_perch; Fish_hake; Fish_octopus_squid; Fish_sardine; Fish_salmon; Fish_flounder; Fish_redfish; Fish_swordfish; Fish_whiting; Fish_dogfish	Binary	Fish menu classification
Vegetarian_soya; Vegetarian_tofu; Vegetarian_seitan; Vegetarian_vegetables	Binary	Vegetarian menu classification
Day	Discrete	Day of the month
Week_Monday; Week_Tuesday; Week_Wednesday; Week_Thursday	Binary	Day of the week
Month_January; Month_February; …; Month_October; Month_November	Binary	Month of the year
Holiday	Binary	Holiday or not
HolidayAfter1; HolidayAfter2; HolidayAfter3	Binary	Holiday in the next day, in two days, or in three days, respectively
HolidayBefore1; HolidayBefore2; HolidayBefore3	Binary	Holiday in the previous day, two days ago, or three days ago, respectively
Temperature	Continuous	Average temperature forecast for the day, measured in Celsius
Precipitation	Continuous	Average precipitation forecast for the day, measured in millimeters
Weather_Condition_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Rain_Overcast; Weather_Condition_Rain_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Wind	Binary	Weather condition

Time horizon: July 2016 to December 2019.

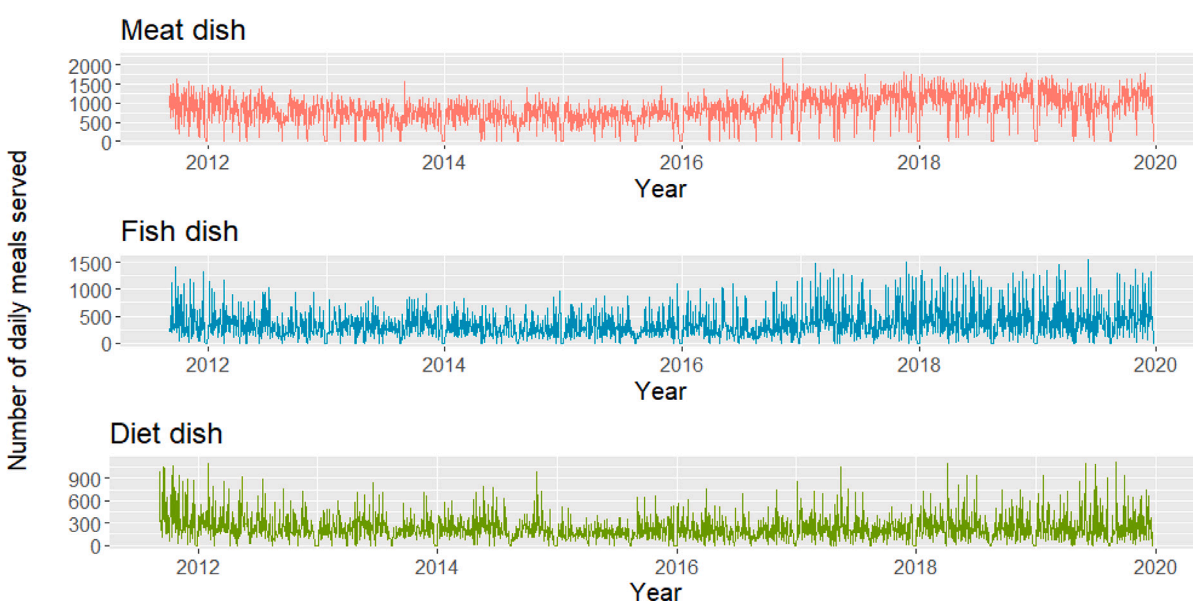

Fig. 7. Number of daily meals served by FCS3 separated by dish.

3.2. Models proposed

Four food demand forecasting models were developed and then refined (to enhance their predictive capabilities). These models were designed to improve the operations of FCSs by outputting reliable next-day forecasts. In other words, considering today to be moment t, i.e., the moment when the forecast is produced, the aim is to forecast demand for $t + 1$, i.e., the next day. To better illustrate the potential of the developed models, the present study focused on forecasting the demand for the meat dish, as this was the type of dish served most often by all three FCSs. This means that the dependent variable of the developed models is the daily number of meat dishes served.

To further test the duality and performance of short-term demand forecasting, two types of forecasting models were employed and tested: time series and casual models. Time series models start by analyzing a time series and identifying patterns, trends, and cyclicalities. Then, they replicate the results to predict future values that are aligned with the setting in question. On the other hand, casual models employ exogenous variables to help categorize the environment in question and use these variables to predict a target variable. Both have their

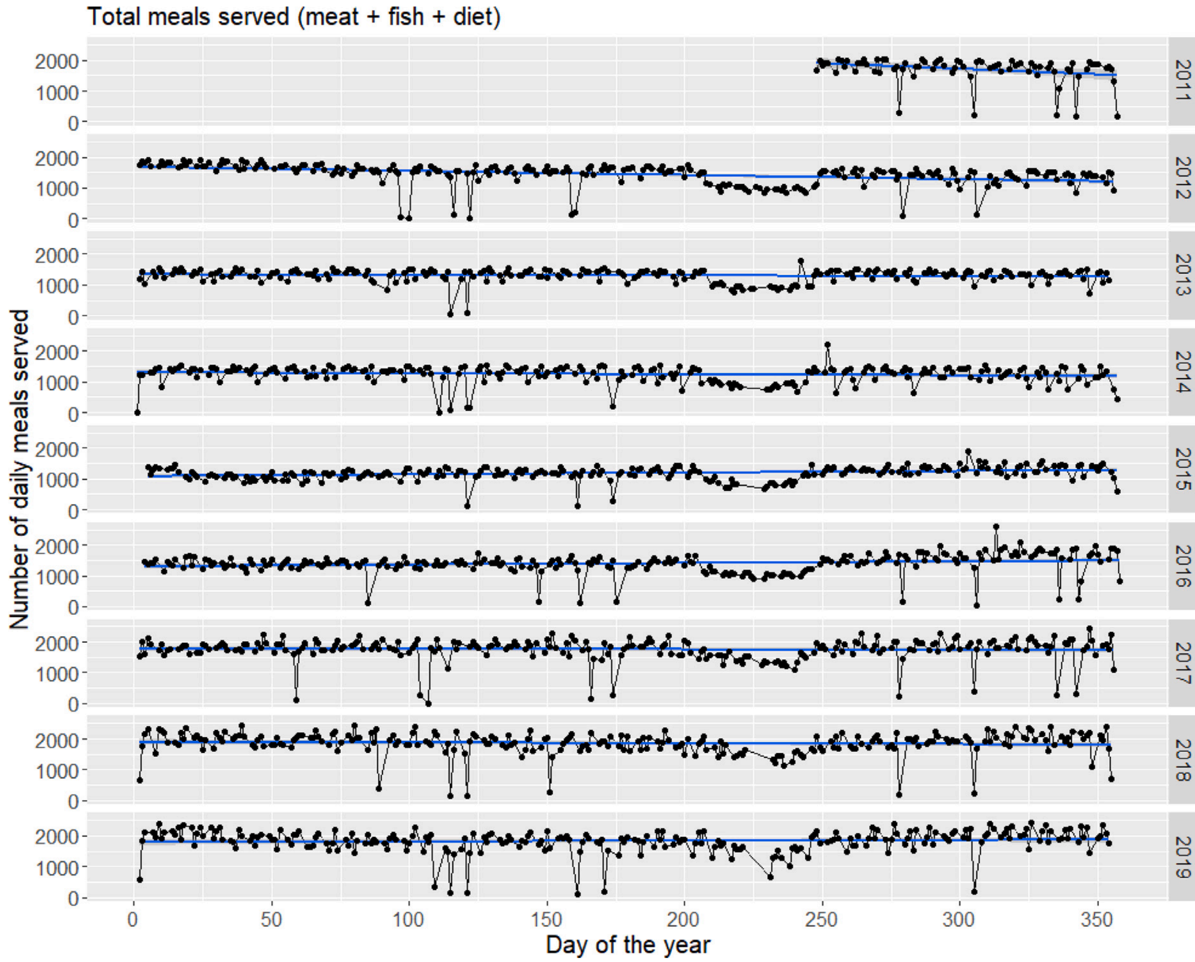

Fig. 8. Number of daily meals served by FCS3.

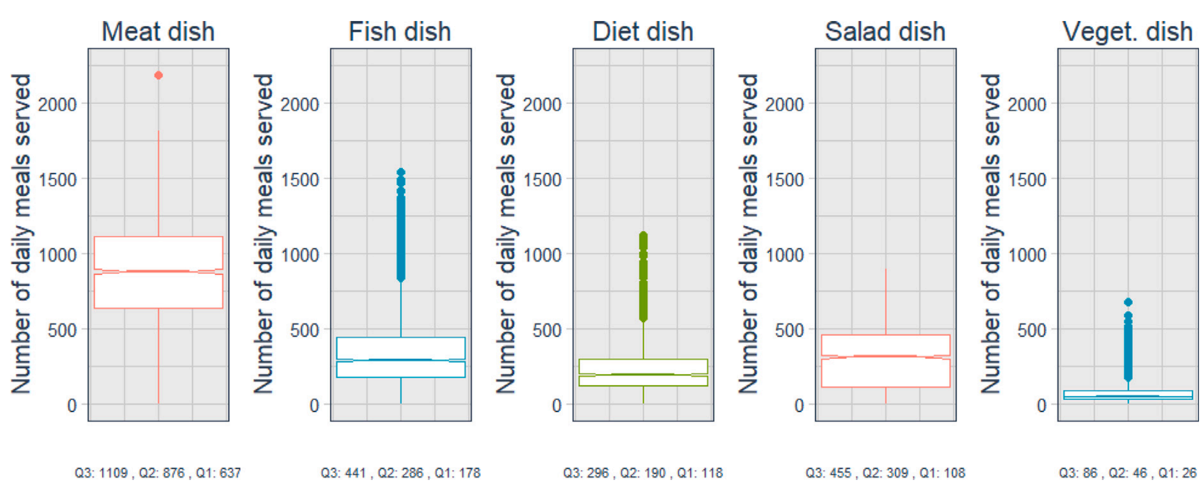

Fig. 9. Number of daily meals served by FCS3 separated by dish.

advantages and disadvantages and their performance depends on the setting at hand.

The causal models were based on the RF and LighGBM algorithms, while the time series models were based on LSTM and transformer neural networks. Considering the time horizon and the number of variables available for each FCS, the casual models should outperform the time series models in the case of FCS1 and FCS2. On the other hand, the model based on the LSTM neural network should outperform the casual models in the case of FCS3, since the time horizon is significantly wider. Lastly, the regression side of transformer neural networks has not been widely adopted in the literature yet. Hence, this is an opportunity to assess its performance in a real-life scenario considering both a small and a large time horizon.

Casual models rely on external factors known as independent variables. In the present study, the variables considered were the number of meals served, the menu, date-related features, weather-related features, the number of students expected to be attending classes, and the number of reservations. These variables were used in both casual

Table 3
Summary of the variables available for FCS3.

Variable name	Type	Description
Dish_meat; Dish_fish; Dish_diet; Dish_salads; Dish_vegetarian	Continuous	Daily number of meals served for the meat, fish, diet, salad, and vegetarian dishes, respectively
Reservation_meat; Reservation_fish; Reservation_diet; Reservation_salads; Reservation_vegetarian	Continuous	Daily number of reservations for the meat, fish, diet, salad, and vegetarian dishes, respectively
Day	Discrete	Day of the month
Week_Monday; Week_Tuesday; Week_Wednesday; Week_Thursday	Binary	Day of the week
Month_January; Month_February; …; Month_October; Month_November	Binary	Month of the year
Holiday	Binary	Holiday or not
HolidayAfter1; HolidayAfter2; HolidayAfter3	Binary	Holiday in the next day, in two days, or in three days, respectively
HolidayBefore1; HolidayBefore2; HolidayBefore3	Binary	Holiday in the previous day, two days ago, or three days ago, respectively
Temperature	Continuous	Average temperature for the day, measured in Celsius
Precipitation	Continuous	Average precipitation for the day, measured in millimeters
Weather_Condition_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Rain_Overcast; Weather_Condition_Rain_Partially_Cloudy; Weather_Condition_Rain; Weather_Condition_Wind	Binary	Weather condition

Time horizon: September 2011 to December 2019.

Table 4
Independent variables used in moment t to forecast the demand for moment t + 1.

Type	Variable	Corresponding days		Variable type
		From	To	
No. of meals served	Dish_meat; Dish_fish; Dish_vegetarian	t	t − 4	Continuous
Menu classification	Meat_pork; Meat_beef; ...; Vegetarian_seitan; Vegetarian_vegetables	t + 1	t − 3	Binary
Date	Day	t + 1	t − 3	Discrete
	Week_Monday; ...; Week_Thursday; Month_January; …; Month_November; Holiday; HolidayAfter1; …; HolidayBefore3	t + 1	t − 3	Binary
Weather	Temperature; Precipitation	t + 1	t − 3	Continuous
	Weather_Condition_Partially_Cloudy; ...; Weather_Condition_Wind	t + 1	t − 3	Binary

models and in the model based on the LSTM neural network. A more comprehensive description of all the independent variables used in the present study is available in Tables 1–3.

There are some variables whose values can be known before the day for which demand is being forecast, e.g., the number of students expected to attend classes on a given day, and others whose values can only be known after their respective occurrences, e.g., the number of meals served for a given dish. In other words, considering today to be moment t and supposing that the forecast is for the next day (moment t+1), the number of students expected to attend classes on t+1 and the number of meals served in moment t are used as independent variables. Moreover, for variables with a known $t+1$ value, the values for moments $t+1, t, t-1, t-2$, and $t-3$ are used as input. For example, the number of students expected to attend classes on the next day and the number of students who attended classes today and in the three previous days are used to forecast tomorrow's demand. On the other hand, for variables with a value that is only known in moment t, the values for moments $t, t-1, t-2, t-3$, and $t-4$ are used as input. This was the case for the independent variables, such as the number of meals of the three dishes that were served. The list of time-dependent variables adopted and their corresponding time horizon coverage is displayed in Table 4.

3.3. Hyperparameters and model validation

Several iterations of the developed models were put to the test (performing next-day forecasts), wherein each iteration had a different combination of hyperparameters. The platform used for these tests was Google Colab, which employs Python. The packages responsible for running the different algorithms were "tensorflow" (LSTM), "lightgbm" (LightGBM), "sklearn" (RF), and "darts" (transformer).

Different hyperparameters were tuned to maximize the algorithms' predictive power for a specific data set. For the RF, the tuned hyperparameters were the percentage of variables to use and number of trees. For the LightGBM algorithm, they were the percentage of variables to use, the learning rate, and the type of boosting. For the LSTM neural network, they were the number of neurons in the first and second layers. And finally, for the transformer neural network, they were the number of expected features in the encoder/decoder inputs and the number of encoder/decoder layers. Different combinations of hyperparameter values were defined for each algorithm and validated before doing another test on new data. The specific values tested for each hyperparameter are shown in Table 5.

The technique used in the development and testing of the machine learning models was a nested cross-validation with a rolling window, as depicted in Fig. 10. This technique employs a sequential rolling window approach, initially dividing the dataset into training and test

Table 5
Hyperparameters.

Algorithm	Hyperparameter	Values	Number of combinations
Random forest	% of features used Number of trees	0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1 100; 200; 300; 400; 500; 600; 700; 800; 900; 1000	100
LightGBM	Boosting Learning rate % of features used	GBDT, DART, GOSS 0,05; 0,10; 0,15; 0,20; 0,25; 0,30; 0,35; 0,40; 0,45; 0,50 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1	300
LSTM	Number of neurons in the first layer Number of neurons in the second layer	10; 50; 100 10; 50; 100	9
Transformer	Number of expected features in the encoder/decoder inputs Number of encoder/decoder layers	16; 32; 64 2; 4; 6	9

sets through an outer loop. Subsequently, within this training set, a secondary rolling window method is applied to partition it into training and validation subsets, constituting an inner loop. One of the advantages of this cross-validation technique is that it can be applied to both casual and time series models.

In the present study, the test period (outer loop) for the FCSs was considered to be the last year for which the data was available. Coincidentally, it was the year 2019 for all FCSs. Therefore, the test period corresponded to the 52 weeks/260 weekdays of 2019. As in a normal rolling window, the test period (2019) was divided into 52 folds with five observations each, i.e., each fold corresponded to a week. In other words, the models forecast future values in groups of weeks.

The validation period was part of the inner loop, meaning it was indexed to the test period. More specifically, the validation period corresponded to the 20 weekdays before the testing period. Since the test period was divided into folds, the validation period also moved with the fold considered at any given moment. For instance, for the first fold of the test period, i.e., the first week of 2019, the validation period corresponded to the last four weeks of 2018; and for the second fold of the test period, i.e., the second week of 2019, the validation period corresponded to the first week of 2019 and the last three weeks of 2018.

The rolling window technique was also applied to the inner loop. Then, the validation period, i.e., the last 20 weekdays before the test period fold, was divided into four folds, each composed of five weekdays. Thus, as with the outer loop, demand was forecast in groups of weeks, but this time for those 20 weekdays. A visual representation of the inner loop is displayed in Fig. 10.

Considering the last weekday of 2018 to be moment t, and the first weekday of 2019 to be moment $t+1$, the first fold of the test period would be from moment $t+1$ to moment $t+5$, while the validation period in the inner loop for this specific fold would be from moment $t-19$ to moment t. However, the validation period would be further divided into four folds, e.g., from moment $t-19$ to moment $t-15$, from moment $t-14$ to moment $t-10$, from moment $t-9$ to moment $t-5$, and from moment $t-4$ to moment t.

The validation period was used to estimate the forecast performance of different hyperparameters throughout a set of 20 weekdays. The hyperparameters that achieved the best forecasts in that validation period were then used to forecast the values in the fold of the test period that the validation period was indexed to. Thus, each fold of the test period can have different hyperparameters. Nevertheless, the overall forecast performance of the whole test period was considered to be the potential performance of the model.

The forecast performance in the validation and test periods was assessed via error measures. Although each error measure has its benefits and demerits, the one chosen to determine the best set of hyperparameters in the present study was the RMSE. The RMSE was chosen over the commonly adopted MAPE since the former penalizes outliers more heavily than the latter. This is crucial in the context of the present study, where it is more important to be consistently close to the real demand than to be spot on some of the time and wildly inaccurate the rest of the time. Furthermore, FCSs can adapt their operations on the spot, and the more negligible the deviation between the forecast and the actual demand, the easier it is for them to do that adaptation. It should be noted that minimizing the RMSE can help prevent not only the overproduction of meals but also the underproduction of meals, which is an important factor for catering services, which are of course driven by economic goals.

Nevertheless, several other performance metrics were examined. Firstly, the difference between the number of meat meals estimated and the actual demand was calculated. This resulted in three values: the number of meals wasted, i.e., the forecast above the actual demand; the unmet demand, i.e., the forecast below the actual demand; and the absolute difference between the forecast and the actual values. These values allowed quantifying the deviation between the actual demand and the produced forecast. At the same time, error measures such as the RMSE, MAPE, and the coefficient of determination (R2) were computed and analyzed in order to obtain a metric that could be used to easily compare the different models not only when looking at the same FCSs, but also when looking at different FCSs.

3.4. Baseline models

To quantify the benefits provided by the proposed models, the estimates from the models should be compared with the estimates from the decision-makers at the FCSs. However, these FCSs did not have such records, and a baseline model had to be created instead. Ergo, two preliminary baseline models were designed to mimic how food demand estimates would have been made. One of the preliminary baseline models was based on a naïve approach and the other was based on a moving average (MA) approach.

The naïve approach estimates demand by considering that the demand will be the same as the demand observed in a previous period. For this approach, different look-back periods were considered: 1, 5, 10, 15, 20, 25, and 30 days. Thus, for instance, when forecasting the demand for $t+1$, the demand could be considered to be the same as in the previous day, i.e., moment t, or the same as 30 days ago, i.e., moment $t-29$. Then, the look-back period that produces the best results is chosen.

Concerning the MA approach, the demand estimated for a given day will be equal to the average demand of a set of previous days. The set of days considered corresponded to the 5, 10, 15, 20, 25, and 30 previous days, which is similar to the number of days contemplated for the naïve approach. This means that, for instance, when forecasting demand for $t+1$, the demand could be considered to be the same as the mean of the last five days, i.e., the average demand between moments $t-4$ and t, or the mean of the last 30 known days, i.e., the average demand between moment $t-29$ and t. Then, the period that bears the best results is chosen.

4. Results

The forecast results demonstrate the benefits of developing models that predict future demand in the context of preventing FW. Most of the proposed models were able to beat the baseline models responsible

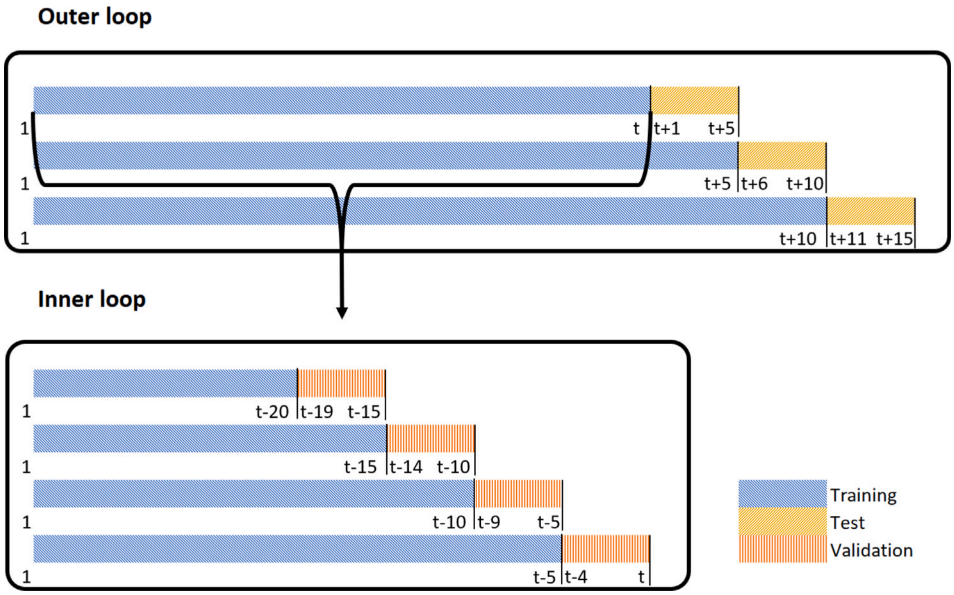

Fig. 10. Nested cross-validation with a rolling window.

Table 6
Results of the forecasting models for FCS1.

Model	Error measures			Forecast differences		
	RMSE	MAPE	R2	Wasted meals	Unmet demand	Total absolute difference
RF	51.57	0.2214	0.2805	4671	5276	9947
LightGBM	57.12	0.2400	0.1819	4867	6322	11,189
LSTM	65.70	0.2638	0.1397	5901	6518	12,419
Transformer	57.60	0.2367	0.1702	3869	6577	10,446
Moving average (15 days)	55.65	0.2460	0.1726	5453	5471	10,924
Naïve (5 days)	75.17	0.3053	0.0059	7496	7189	14,685

Note: Around 56,000 meat meals were served in 2019, which translates into roughly 215 daily meals.

Table 7
Results of the forecasting models for FCS2.

Model	Error measures			Forecast differences		
	RMSE	MAPE	R2	Wasted meals	Unmet demand	Total absolute difference
RF	47.01	0.2314	0.6806	3268	4391	7659
LightGBM	51.06	0.2400	0.6344	3200	5047	8257
LSTM	60.90	0.3061	0.5066	3734	6292	10,026
Transformer	76.53	0.3543	0.3614	2521	10,487	13,008
Moving average (5 days)	56.22	0.2765	0.5510	4478	4642	9120
Naïve (1 day)	57.71	0.2533	0.5710	4627	4655	9282

Note: Around 40,000 meat meals were served in 2019, which translates into roughly 150 daily meals.

Table 8
Results of the forecasting models for FCS3.

Model	Error measures			Forecast differences		
	RMSE	MAPE	R2	Wasted meals	Unmet demand	Total absolute difference
RF	227.25	0.2062	0.5477	19,026	25,111	44,137
LightGBM	213.22	0.1941	0.5902	15,153	25,004	40,157
LSTM	173.36	0.1432	0.7325	15,151	17,356	32,507
Transformer	358.88	0.3813	0.0254	21,955	47,476	69,431
Moving average (30 days)	323.83	0.3996	0.0381	31,782	30,059	61,841
Naïve (5 days)	432.13	0.4427	0.0259	39,685	43,767	83,452

Note: Around 270,000 meat meals were served in 2019, which translates into roughly 1050 daily meals.

for mimicking how managers do their forecasts. This means that by employing more sophisticated forecasting techniques, it is possible to better predict future demand. Thus, these forecasting models constitute a valuable asset for managers to make better decisions and ultimately minimize FW generation and maximize the number of meals served.

The forecast performance (measured through forecast differences and error measures) of the best model for each algorithm is detailed in Tables 6–8. The model based on the RF algorithm produced the best forecasts for the first two FCSs, while the LSTM algorithm had the better results for FCS3, i.e., its results had the lowest total absolute difference

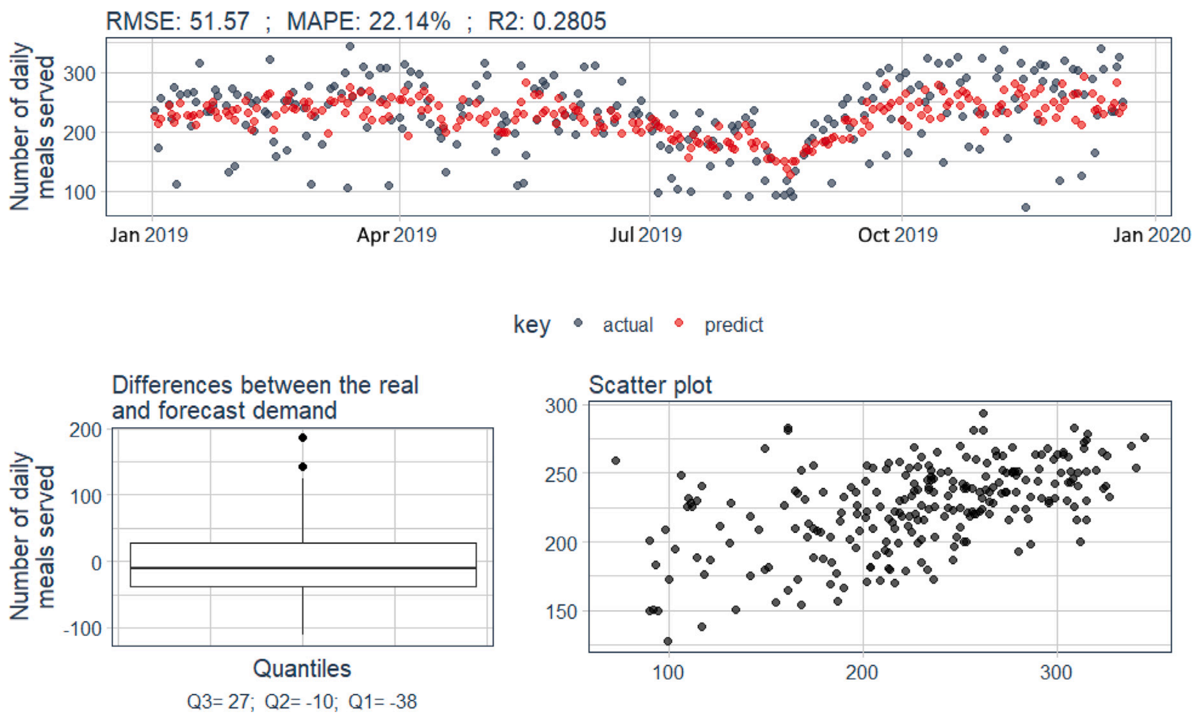

Fig. 11. Visual representation of the real and forecast demand for FCS1.

between the forecast and actual values, which in turn means the overall lowest RMSE. In the case of the baseline models, the one based on the MA approach was always better than the one based on the simple naïve approach.

4.1. Food catering service 1 - Student canteen

For the first FCS, the error metrics and the absolute difference between the forecast and actual values are displayed in Table 6. The model based on the RF algorithm was the best out of all developed models in terms of error metrics, achieving a RMSE of 51.57. As for the baseline models, the one based on the MA approach was the best when considering the average of the last 15 days.

By basing their decisions on the forecasts provided by the RF algorithm instead of an MA of the last 15 known days, this FCS would be able to reduce the number of meals wasted in 2019 by 782, which translates into a 14% reduction in the FW generated. At the same time, it would be able to reduce the unmet demand in the same period by 195 dishes. A visual representation of the actual demand and the forecast demand produced by the model based on the RF algorithm is displayed in Fig. 11.

Although the other developed forecasting models performed worse than the model based on the MA approach, they were still better than the model based on the naïve approach when considering the quality of the predictions as a whole. However, the forecasts produced by the models based on the LightGBM algorithm and the transformer neural network would lead to fewer wasted meals when compared with the forecasts produced by the model based on the MA approach. When considering the model based on the naïve approach instead, every developed model would lead to fewer wasted meals. However, every developed model except the one based on the RF algorithm would lead to fewer meals being served (i.e., a higher unmet demand) when compared with the one based on the MA approach, and the inverse when compared with the one based on the naïve approach. In short, the best forecasting model is the one based on the RF algorithm, followed by the one based on the MA approach, the other developed forecasting models, and, lastly, the model based on the naïve approach.

However, if the goal was solely to minimize the number of wasted meals, the forecasts produced by the model based on the transformer neural network would be the most desirable. Otherwise, the forecasts produced by the model based on the RF algorithm are superior to every other prediction.

4.2. Food catering service 2 - Student canteen

For the second FCS, only the models based on the RF and LightGBM algorithms were able to beat the best baseline model, as highlighted by their lower RMSE, MAPE, and total absolute difference, and their higher R2. The performance of all forecasting models is presented in Table 7. Only the forecasts produced by the model based on the RF algorithm resulted in a lower amount of wasted meals and unmet demand when compared with the best baseline model. The forecasts of the model based on the RF algorithm would result in 3268 meals being wasted in 2019, given the actual demand.

Although this is a considerable number of wasted meals, it is still 1210 fewer wasted meals (or minus 27% meals in relative terms) when compared with the baseline model based on the MA of the last five known days. Additionally, while the model based on the RF algorithm would lead the FCS to sell 4391 fewer meals in 2019 (i.e., a higher unmet demand), this is still 251 more meals served when compared with the baseline model based on the MA approach. An overall view of the forecasts generated by the model based on the RF algorithm for FCS2 is provided in Fig. 12.

The forecasts produced model based on the LightGBM algorithm are also noteworthy. Although these would lead to a lower number of meals served (i.e., a higher unmet demand) when compared with the forecasts of the models based on the RF algorithm and the baseline models, it would reduce the amount of food wasted more significantly than the other two. While the model based on the RF algorithm and the baseline models would lead to 3268 and 4478 wasted meals, respectively, the model based on the LightGBM algorithm would lead to only 3200 wasted meals.

The model based on the LSTM neural network would lead to more wasted meals than the models based on the RF and LightGBM algorithms. When compared with the baseline models, the model based

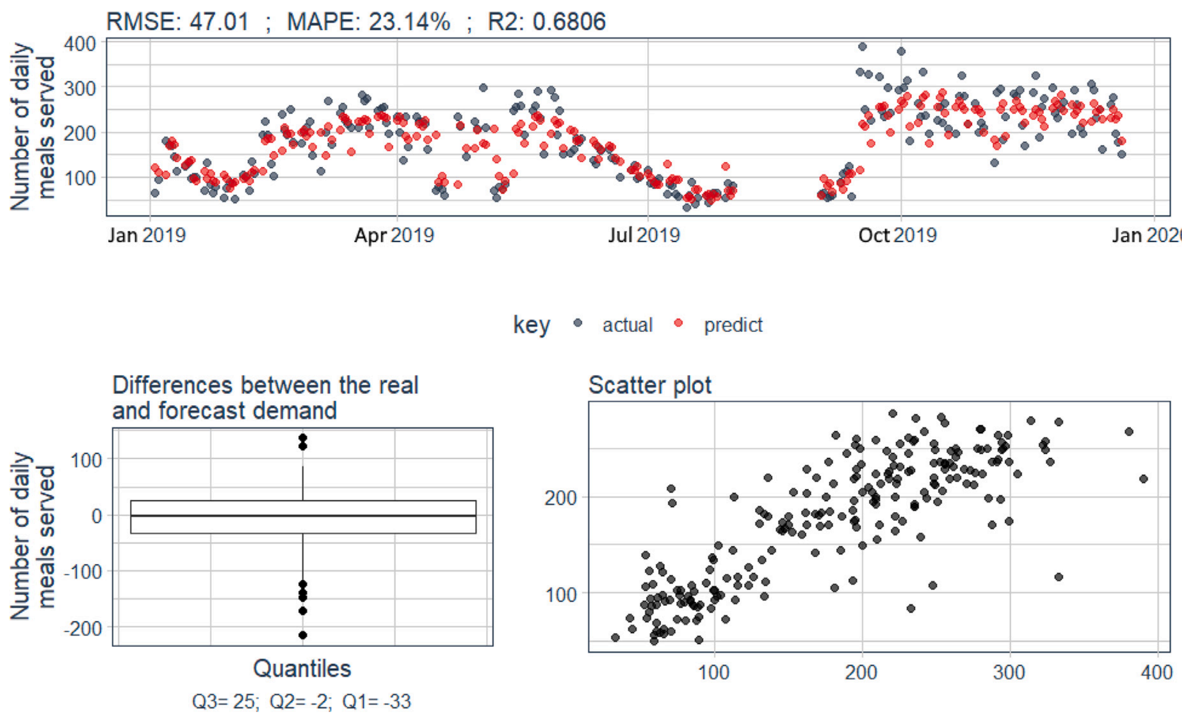

Fig. 12. Visual representation of the real and forecast demand for FCS2.

on the LSTM neural network would lead to fewer wasted meals, but a considerably higher unmet demand. Ultimately, its error measures and total absolute difference metric are worse than those of both baseline models.

The model based on the transformer neural network produced the worst forecasts out of all the models in the context of FCS2. Although it would lead to the lowest number of wasted meals (2521), this would be at the cost of a very high unmet demand.

In brief, the model based on the RF algorithm would grant FCS2 more reliable forecasts in comparison with the baseline models. The forecasts produced by the model based on the RF algorithm would allow FCS2 to generate considerably less FW while slightly increasing the number of meals served, i.e., the unmet demand would be lowered. However, if this FCS wanted to waste the lowest number of meals possible, regardless of the number of meals served, the model based on the transformer neural network would be a better choice than the one based on the RF algorithm.

4.3. Food catering service 3 - Company canteen

This FCS operates on a much larger scale than the other two other FCSs, having a considerably larger number of meals served in 2019. Still, even in this setting, the model based on the RF algorithm outperformed the best baseline model. Moreover, for FCS3, the model based on the LSTM neural network provided the most accurate forecasts, winning on every metric adopted, as seen in Table 8.

By employing the model based on the LSTM neural network, it would be possible to waste 16,631 fewer meals in 2019 than when compared with the model based on the MA of the last 30 known days, which would lead to 31,782 wasted meals in the same period. This difference is considerable, translating into 52% less FW in favor of the model based on the LSTM neural network. Not only that, but the forecasts provided by the model based on the LSTM neural network would also manage to minimize the amount of unmet demand, i.e., the number of meals not served due to underproduction. Although it would cause demand for 17,356 meals to go unmet in 2019, this model would still lead to 12,703 more meals being served than the baseline model based on the MA approach. A visual representation of the forecasts provided by the model based on the LSTM neural network and the real demand is drawn in Fig. 13.

The models based on the RF and LightGBM algorithms also outperformed the baseline model based on the MA approach. These two forecasts would lead to fewer wasted meals and less unmet demand by a considerable margin, especially considering the number of wasted meals. In terms of unmet demand, these two models performed similarly, but the one based on the LightGBM algorithm would produce significantly fewer wasted meals, essentially matching the same number produced by the model based on the LSTM neural network.

The demand forecast produced by the model based on the transformer neural network would lead to the most wasted meals out of all the predictions, but fewer wasted meals than both baseline models. However, to achieve this, the model would lead to the fewest meals being served out of any model, causing the FCS to not meet the demand for 47,476 meals in 2019.

Contrary to the other two FCSs, for FCS3, the model based on the LSTM neural network produced the best forecasts. Not only that, but it would also minimize the number of wasted meals while providing better demand predictions to maximize the number of meals served. Thus, for FCS3, this model would be the preferred choice.

4.4. Overall results

As expected, each FCS had unique properties and, thereby, each developed model performed differently depending on the FCS. Nevertheless, it is clear that the proposed models would enable all the studied FCSs to reduce their FW.

The RMSE, being an absolute metric, was significantly higher for FCS3 than FCS1 and FCS2. On the other hand, relative measures like the MAPE and R2, which should not be impacted by the value range of the target variable, were also different for each FCS. At first glance, and given that a higher relative error measure would indicate a lower predictive power of the forecasting model, it is possible to verify that the developed models were the most accurate for FCS2, while the food demand in FCS1 was the hardest to forecast.

More specifically, the model based on the RF algorithm provided the best results for FCS1 and FCS2, i.e., the school canteens, especially

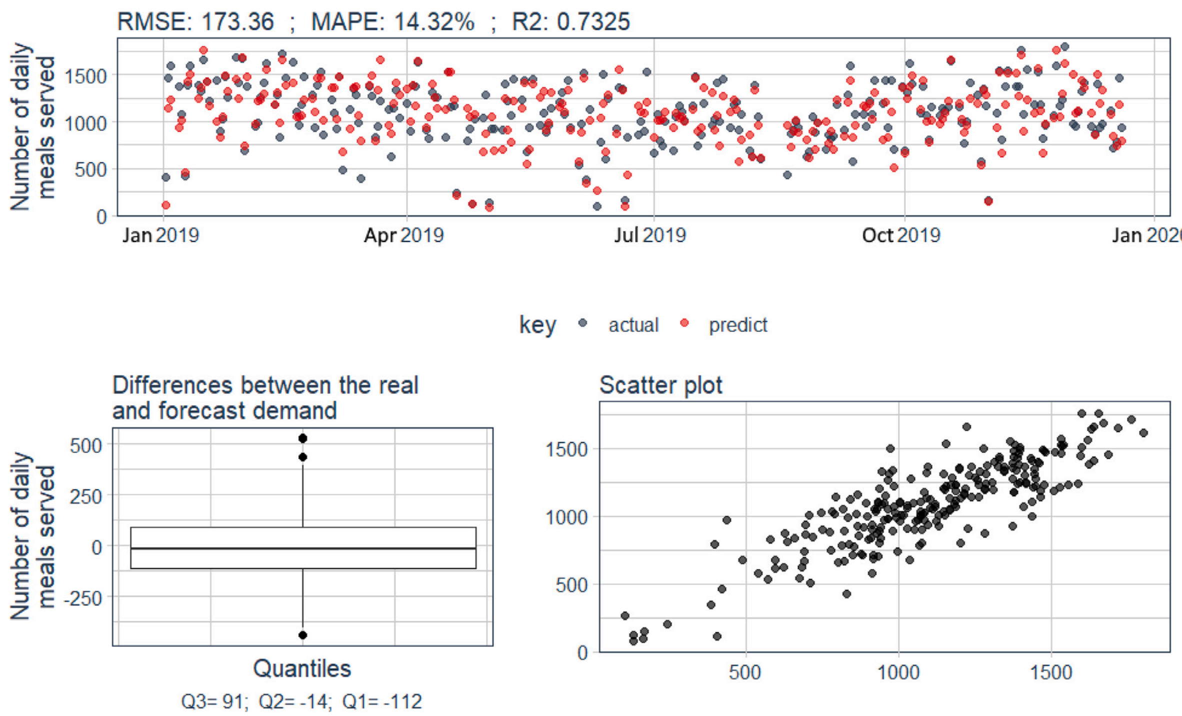

Fig. 13. Visual representation of the real and forecast demand for FCS3.

in terms of minimizing the number of wasted meals in 2019. The predictions of the RF model were better than those of the baseline models in terms of both the number of meals wasted and the number of meals served. Although the model based on the transformer neural network provided forecasts that would minimize the number of meals wasted, this would happen due to an underestimation, and therefore at the cost of increasing the number of meals not provided. Thus, for FCS1 and FCS2, adopting the model based on the RF algorithm would be the best option. This is also reinforced by the RMSE, MAPE, and R2 scores attained.

For FCS3, which operates in a company setting, the best model was instead the one powered by the LSTM neural network, which performed significantly better than the rest. This model would be able to minimize the number of meals wasted while also minimizing the unmet demand. This model's superiority is also evident from its attained error measures, which were significantly better than the other models'. Nevertheless, the models based on the RF and LightGBM algorithms also produced excellent results, beating the baseline model based on the MA of the last 30 days by a significant amount. Like in the previous FCSs, the model based on the transformer neural network would lead to the least provisioned meals, but in this case, it would also cause the most wasted meals out of all the developed models. Additionally, the error measures of the model based on the transformer neural network were worse than the ones of the baseline model based on the MA approach, but better than the ones of the baseline model based on the naïve approach.

5. Discussion

Out of the four proposed food demand forecasting models, only two outperformed the others in a given setting. More specifically, the model based on the RF algorithm outperformed the rest in the school settings (FCS1 and FCS2), and the model based on the LSTM neural network outperformed the rest in the company setting (FCS3). While the setting may partially explain these results, it is also important to analyze the time horizon and the levels of demand of each FCS. The model based on the LSTM neural network excelled in a context with an extensive time horizon (nine years) and high demand (around 1900 daily meals), while the model based on the RF algorithm stood out in two contexts with short time horizons (between three and four years) and a more modest number of daily meals served (between 200 and 400). Nonetheless, more experiments should be conducted before claiming any such correlations. Some other noteworthy performances are those of the model based on the LightGBM algorithm in the context of FCS2 and FCS3 and the model based on the transformer neural network in the context of FCS1.

Out of the two baseline models, the one based on the MA approach provided the best results in all studied FCSs. Furthermore, this model always managed to outperform one or more of the proposed forecasting models. The model based on the naïve approach produced the worst forecasts in the context of FCS1 and FCS3 but managed to outperform the model based on the LSTM and transformer neural networks in the context of FCS2.

The predictions produced by the developed models were not perfect. However, they were still significantly better than the best forecasts generated by the baseline models. For FCS1, which served 56,000 meat meals in 2019, the adoption of the model based on the RF algorithm would lead to 4671 meals being wasted. Furthermore, FCS1 would not be able to meet the demand for 5276 meals due to underproduction on some days. Although these numbers amount to a significant portion of the total number of meals served, they are still lower than what possibly happened (as measured by the baseline model). More specifically, in the test performed for 2019, adopting the predictions of the model based on the RF algorithm would lead to a 14% reduction in the number of wasted meals, or 782 meals, while also serving 195 more meals.

The tests performed for the other FCSs had similar outcomes. In the case of FCS2, which served 56,000 meat meals in 2019, following the forecasts of the model based on the RF algorithm would lead to a 27% decrease in FW (1210 meals) over the moving average baseline model, while also being able to serve 251 more meals. As for FCS3, the differences were significantly larger, possibly due to the size of the canteen and the larger time horizon of the data used to train the models. Namely, in this setting, it would be possible to reduce FW by more than half, or 31,782 meals, whilst serving 12,703 more meals in the same period.

However, simply stating that applying these forecasting models throughout the test period (2019) would prevent 18.623 fewer meals from being wasted is disingenuous. Doing so is assuming that the FCSs (1) do their forecasts exactly as the moving average baseline model; (2) follow the forecasts blindly; and (3) have no measures in place to allow for some variance between the forecast and actual demand, e.g., produce more meals if the demand is above the one that was forecast, and recall meals produced or halt production if the demand is below the one that was forecast. Considering that FCSs can compensate for deviations between the forecast and actual demand, choosing the model with the lowest RMSE should be beneficial. The more consistently close the forecast values are to the actual demand, the easier it is for FCSs to perform minor adjustments on the day, depending on how it unfolds.

The present study has some further limitations. First, the three studied settings may not be representative of how most school or company canteens operate. Thus, the results obtained should not be generalized. Furthermore, since the data collected corresponds to the number of dishes served and not the quantity served, the demand estimates provided by the developed models may not be the most appropriate in terms of suggesting the amount of food to prepare. This is due to the fact that the amount served on each plate may vary slightly. Additionally, it is important to emphasize that although this study focuses on the waste generated by FCSs, the amount of waste generated by customers who do not finish their meals (i.e., plate waste) is very significant.

A further limitation is that the evaluation of the proposed models was supported by baseline models that were created to mimic the forecasts of the studied FCSs. It was necessary to design these baseline models because the data collected did not include the estimates made by these services, possibly because they were never registered. Nevertheless, the adopted baseline models may not accurately reflect the thought process of the managers of these FCSs and other approaches may be preferred. It is also important to highlight that the results obtained might have been affected by some data quality issues, namely the accuracy and accessibility of the data. The proposed models were limited by a lack of relevant information that could have been used to infer the potential demand for a particular dish, such as daily menu information (with the exception of FCS2) and the preferences of individual customers. However, even when daily menu information was recorded (FCS2), the relatively few days during which a specific menu was served during the analysis period implied a broad classification of the menus.

6. Conclusion and future research

Mass amounts of FW are produced annually, leading to immeasurable environmental, social, and economic consequences. One of the main reasons for FW generation is overproduction, which comes mainly from poor demand forecasts. Thus, the present work proposes and compares food demand forecasting models based on different algorithms to support the decision-makers at FCSs. By having more accurate and reliable forecasts, these decision-makers can make sure that their services are better prepared to meet customer demand.

To test the performance of the proposed models, data was collected from three different FCSs: two operating school canteens and one operating a company canteen. Among the tested models, the ones based on the RF algorithm and LSTM neural network provided the best results. Comparing the forecasts provided by these models with the forecasts of the baseline models, the former managed to reduce the amount of generated FW by between 14% and 52%.

The forecasts provided by the model based on the RF algorithm were more reliable in the context of the two school canteens, while the model based on the LSTM neural network proved to be better in the context of the company canteen. The model based on the RF algorithm produced better results in settings with three to four years of historic data and an average of 200 to 400 meals served daily, while the model based on the LSTM neural network outperformed the rest when the data available had a time horizon of more than nine years and around 1900 meals served daily.

Overall, the LSTM models produce the greatest theoretical amount of FW reduction, both in absolute and relative terms. This fruitful performance comes mainly from the results of FCS3, where the results may be a consequence of the longer time horizon of the data, which allowed the prediction models to better understand and adapt to the setting.

The forecasts produced with the support of machine learning techniques proved to be more accurate and reliable than those that FCSs usually perform on their own. Even the best baseline model was always beaten by the models based on the RF algorithm or LSTM neural network, by a significant margin. That said, FCSs that do not rely on machine learning forecasts are better off relying on an MA approach, i.e., the average of the last known days, over a naïve approach, i.e., the same demand of a previous day.

Even though the machine learning forecasts beat the baseline models, enabling a theoretical reduction of FW, it is important to understand that in the setting at hand, there is no one-size-fits-all algorithm, with different algorithms capitalizing on different features and being more adequate to different settings. Additionally, this study only focused on forecasting demand for a single dish. Although this dish was the most popular in the case of all three studied FCSs, full menu predictions could also be beneficial for planning purposes. Moreover, it would be useful for future forecast works to explore different research avenues, such as understanding individual preferences (in cases where reservations are made) and/or suggesting future menus in order to motivate demand and minimize FW. Furthermore, it would be relevant to confirm whether larger time horizons of data in the field of FCSs are indexed to better forecasting results. Lastly, it would be very interesting to design a system that can be trained on data from different settings simultaneously in order to leverage pattern recognition.

CRediT authorship contribution statement

Miguel Rodrigues: Conceptualization, Investigation, Methodology, Software, Validation, Writing – original draft. **Vera Miguéis:** Conceptualization, Investigation, Methodology, Resources, Supervision, Validation, Writing – review & editing. **Susana Freitas:** Conceptualization, Resources. **Telmo Machado:** Conceptualization, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

Arunraj, N., Ahrens, D., 2015. A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. Int. J. Prod. Econ. 170, 321–335. http://dx.doi.org/10.1016/j.ijpe.2015.09.039.

Brennan, L., Langley, S., Verghese, K., Lockrey, S., Ryder, M., Francis, C., Phan-Le, N.T., Hill, A., 2021. The role of packaging in fighting food waste: A systematised review of consumer perceptions of packaging. J. Clean. Prod. 281, 125276. http://dx.doi.org/10.1016/j.jclepro.2020.125276.

Crossing, V., 2003. Weather condition databse. https://www.visualcrossing.com/weather-data.

Curry, N., Pillay, P., 2012. Biogas prediction and design of a food waste to energy system for the urban environment. Renew. Energy 41, 200–209. http://dx.doi.org/10.1016/j.renene.2011.10.019.

Dai, X., Duan, N., Dong, B., Dai, L., 2013. High-solids anaerobic co-digestion of sewage sludge and food waste in comparison with mono digestions: Stability and performance. Waste Manage. 33 (2), 308–316. http://dx.doi.org/10.1016/j.wasman.2012.10.018.

De Clercq, D., Jalota, D., Shang, R., Ni, K., Zhang, Z., Khan, A., Wen, Z., Caicedo, L., Yuan, K., 2019. Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data. J. Clean. Prod. 218, 390–399. http://dx.doi.org/10.1016/j.jclepro.2019.01.031.

Dharmawardane, C., Sillanpää, V., Holmström, J., 2021. High-frequency forecasting for grocery point-of-sales: intervention in practice and theoretical implications for operational design. Oper. Manage. Res. 14 (1–2), 38–60. http://dx.doi.org/10.1007/s12063-020-00176-7.

Dhir, A., Talwar, S., Kaur, P., Malibari, A., 2020. Food waste in hospitality and food services: A systematic literature review and framework development approach. J. Clean. Prod. 270, 122861. http://dx.doi.org/10.1016/j.jclepro.2020.122861.

DRAAF, 2014. Proposer une boîte à emporter. Note réglementaire.

Faezirad, M., Pooya, A., Naji-Azimi, Z., Amir Haeri, M., 2021. Preventing food waste in subsidy-based university dining systems: An artificial neural network-aided model under uncertainty. Waste Manage. Res. 39 (8), 1027–1038. http://dx.doi.org/10.1177/0734242X211017974.

FAO, 2011. Global Food Losses and Food Waste - Extent, Causes and Prevention. Rome.

Filimonau, V., De Coteau, D.A., 2019. Food waste management in hospitality operations: A critical review. Tour. Manag. 71 (October 2018), 234–245. http://dx.doi.org/10.1016/j.tourman.2018.10.009.

Girotto, F., Alibardi, L., Cossu, R., 2015. Food waste generation and industrial uses: A review. Waste Manage. 45, 32–41. http://dx.doi.org/10.1016/j.wasman.2015.06.008.

Gružauskas, V., Gimžauskienė, E., Navickas, V., 2019. Forecasting accuracy influence on logistics clusters activities: The case of the food industry. J. Clean. Prod. 240, http://dx.doi.org/10.1016/j.jclepro.2019.118225.

Hebrok, M., Boks, C., 2017. Household food waste: Drivers and potential intervention points for design – An extensive review. J. Clean. Prod. 151, 380–392. http://dx.doi.org/10.1016/j.jclepro.2017.03.069.

Hennchen, B., 2019. Knowing the kitchen: Applying practice theory to issues of food waste in the food service sector. J. Clean. Prod. http://dx.doi.org/10.1016/j.jclepro.2019.03.293.

Kamble, S., Gunasekaran, A., Gawankar, S., 2020. Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. Int. J. Prod. Econ. 219, 179–194. http://dx.doi.org/10.1016/j.ijpe.2019.05.022.

Lee, H., Choi, I., Moon, H., Kim, J., 2020. A multi-period product recommender system in online food market based on recurrent neural networks. Sustainability (Switz.) 12 (3), http://dx.doi.org/10.3390/su12030969.

Li, Y., Jin, Y., Li, H., Borrion, A., Yu, Z., Li, J., 2018. Kinetic studies on organic degradation and its impacts on improving methane production during anaerobic digestion of food waste. Appl. Energy 213, 136–147. http://dx.doi.org/10.1016/j.apenergy.2018.01.033.

Malefors, C., Secondi, L., Marchetti, S., Eriksson, M., 2021a. Food waste reduction and economic savings in times of crisis: The potential of machine learning methods to plan guest attendance in Swedish public catering during the Covid-19 pandemic. Socio-Econ. Plann. Sci. http://dx.doi.org/10.1016/j.seps.2021.101041.

Malefors, C., Strid, I., Hansson, P.-A., Eriksson, M., 2021b. Potential for using guest attendance forecasting in Swedish public catering to reduce overcatering. Sustain. Prod. Consum. 25, 162–172. http://dx.doi.org/10.1016/j.spc.2020.08.008.

Martin-Rios, C., Demen-Meier, C., Gössling, S., Cornuz, C., 2018. Food waste management innovations in the foodservice industry. Waste Manage. http://dx.doi.org/10.1016/j.wasman.2018.07.033.

Martin-Rios, C., Hofmann, A., Mackenzie, N., 2021. Sustainability-oriented innovations in food waste management technology. Sustainability (Switz.) http://dx.doi.org/10.3390/su13010210.

Messner, R., Richards, C., Johnson, H., 2020. The "prevention paradox": food waste prevention and the quandary of systemic surplus production. Agricult. Hum. Values http://dx.doi.org/10.1007/s10460-019-10014-7.

Møller Christensen, F., Solheim-Bojer, C., Dukovska-Popovska, I., Steger-Jensen, K., 2021. Developing new forecasting accuracy measure considering product's shelf life: Effect on availability and waste. J. Clean. Prod. 288, http://dx.doi.org/10.1016/j.jclepro.2020.125594.

Papargyropoulou, E., Lozano, R., K. Steinberger, J., Wright, N., Ujang, Z.B., 2014. The food waste hierarchy as a framework for the management of food surplus and food waste. J. Clean. Prod. 76, 106–115. http://dx.doi.org/10.1016/j.jclepro.2014.04.020.

Parfitt, J., Barthel, M., MacNaughton, S., 2010. Food waste within food supply chains: Quantification and potential for change to 2050. Philos. Trans. R. Soc. B 365 (1554), 3065–3081. http://dx.doi.org/10.1098/rstb.2010.0126.

Pirani, S., Arafat, H., 2016. Reduction of food waste generation in the hospitality industry. J. Clean. Prod. 132, 129–145. http://dx.doi.org/10.1016/j.jclepro.2015.07.146.

Posch, K., Truden, C., Hungerländer, P., Pilz, J., 2021. A Bayesian approach for predicting food and beverage sales in staff canteens and restaurants. Int. J. Forecast. http://dx.doi.org/10.1016/j.ijforecast.2021.06.001.

Priefer, C., Jörissen, J., Bräutigam, K.R., 2016. Food waste prevention in Europe - A cause-driven approach to identify the most relevant leverage points for action. Resour. Conserv. Recy. 109, 155–165. http://dx.doi.org/10.1016/j.resconrec.2016.03.004.

Rodrigues, M., Miguéis, V., Freitas, S., 2023. A literature review on the quantitative approaches to food waste: descriptive, predictive, and prescriptive analyses. Environ. Sci. Pollut. Res..

Sakoda, G., Takayasu, H., Takayasu, M., 2019. Data science solutions for retail strategy to reduce waste keeping high profit. Sustainability (Switz.) 11 (13), http://dx.doi.org/10.3390/su11133589.

San Martin, D., Orive, M., Martínez, E., Iñarra, B., Ramos, S., González, N., Guinea de Salas, A., Vázquez, L., Zufía, J., 2021. Multi-criteria assessment of the viability of valorising vegetable by-products from the distribution as secondary raw material for animal feed. Environ. Sci. Pollut. Res. 28 (13), 15716–15730. http://dx.doi.org/10.1007/s11356-020-11752-6.

Schneider, F., 2013. Review of food waste prevention on an international level. Proc. Inst. Civ. Eng.: Waste Resour. Manage. 166 (4), 187–203. http://dx.doi.org/10.1680/warm.13.00016.