Do Forecasting Algorithms Need a Crisis-Mode? Machine Learning Based Sales Forecasting in Times of COVID-19

Tobias Fahse[1]

[1] University of St. Gallen, Institute of Information Management, St. Gallen, Switzerland,
tobias.fahse@unisg.ch

Abstract.
On average, algorithmic forecasts outperform human forecasts by 10%. This disparity increases further thanks to cutting-edge machine learning (ML) algorithms. As business decisions based on sales forecasting are regarded as particularly important and a variety of other activities rely on them, accurate sales forecasting is critical to companies' profitability. At the same time, being able to predict the next day's sales more accurately can significantly reduce food waste and help fulfilling sustainability. Thus, sales forecasting is one of the primary value propositions of artificial intelligence (AI). However, it is crucial for the acceptance and adoption of ML-based sales forecasting algorithms to perform reliably during pandemics such as the covid-19 pandemic. Although governments' containment measures highly impact the sales of a bakery's products, no study has yet scrutinized incorporating the stringency of containment measures as an input variable for sales forecasting. Hence, this paper examines the performance of a ML sales forecasting system for baked goods in times of covid-19 and proposes incorporating a covid containment measurement stringency index as an additional input variable to increase forecast accuracy in times of pandemics. This way, prediction accuracy increases by 4.61% on average. Consequently, a containment measures stringency variable should be used to increase accuracy in future pandemics. By simulating an upcoming pandemic, it is further demonstrated how learnings from the covid-19 pandemic could be meaningfully transferred. For this study, real data is used: A Swiss bakery chain provides real sales data covering 5 years including 2 years of the covid-19 pandemic.

Keywords: Sales Forecasting, Machine Learning, COVID-19, containment measures

1 Introduction

In this paper, the usefulness of a covid containment measures stringency index to increase prediction accuracy of the next day's sales of baked goods is examined based on real sales data that is provided by a Swiss bakery chain.

On average, algorithmic forecasts outperform human forecasts by 10% [1]. This disparity increased further due to cutting-edge machine learning (ML) algorithms [2]. Business decisions based on sales forecasting are regarded as particularly important and a variety of other activities rely on them, accurate sales forecasting is critical to companies' profitability [3, 4]. At the same time, food waste is an increasingly significant challenge for society and is recognized by the United Nations as a sustainable development goal. Being able to predict the next day's sales more accurately can significantly reduce food waste and help to fulfill sustainability goals when applied in the food industry or in retail. According to the Swiss Federal Office for the Environment (BAFU), 2.8 million tons of food are wasted in Switzerland per year, which equals 330 kg per person per year, or about a quarter of the environmental impact caused by our food. Among all food categories, baked goods have the greatest environmental impact due to food waste [5].

The opportunity to increase sales by avoiding shortage and to reduce costs and food waste by avoiding overproduction calls for more accurate sales forecasting. Consequently, sales forecasting is one of the primary value propositions of artificial intelligence (AI) [6]. However, it is crucial for the acceptance and adoption of ML-based sales forecasting algorithms to perform reliably during pandemics such as the recent covid-19 pandemic. Restrictions like quarantines, mandatory home office or mask obligations highly impact the sales of a typical bakery's products and negatively affects the performance of ML models [7].

To the best of the author's knowledge, there are currently no studies that incorporate the severity of covid-19 containment measures into ML sales forecasting algorithms as an input variable. Several studies investigate how the covid-19 pandemic has affected sales forecasting and suggest novel algorithms to deal with the distortions [8-11]. For instance, the covid-19 pandemic affected the importance of feature variables: Comparing pre- vs. post-pandemic feature importance, features that negatively influence sales increased in importance in a study in the context of an online marketplace [12]. To deal with the distortions of the pandemic, one study incorporates expert opinions into a ML forecast algorithm [13], while another study adds flags to their data to indicate the different stages of the covid-19 pandemic (panic state, relief period after a wave, etc.), thus improving the prediction of monthly car sales [14]. However, pandemic containment measures were not considered in these studies.

To tackle the outlined research gap, this study examines the effectiveness of including a covid containment measures stringency index into a ML sales forecasting system to increase forecast accuracy in times of covid-19. For the elaborations, real data is used: A Swiss bakery chain provides transactional sales data covering approximately 5 years including 2 years of the covid-19 pandemic. The usefulness of the covid stringency index is evaluated based on the 15 best selling products of the largest branch of the bakery chain. First, for 13 of 15 products, the onset of the covid pandemic represents a structural break in the data. Second, it is shown that in case of a future pandemic, a restriction stringency variable should be used, as it increases the prediction accuracy by 4.61%. Third, situations in which a "pandemic variable" is beneficial or detrimental to the accuracy of forecasts are examined by simulating the start of an upcoming pandemic and comparing the ML system with and without considering the covid stringency index as an input variable. Fourth, this study shows how learnings from the covid-19 pandemic could be meaningfully transferred to an upcoming pandemic. These analyses are executed to answer the following research questions (RQ):

RQ1: *Does the COVID-19 stringency index improve one-day-ahead forecasts in sales forecasting in the bakery domain?*

RQ2: *Can a pre-trained pandemic variable improve one-day-ahead forecasts in times of pandemic?*

By answering these research questions, this paper contributes to practice by showing the usefulness of variables that measure the strictness of the government's containment measures as a reaction to pandemics in ML-based sales forecasting algorithms. This study contributes to theory by evaluating the degree to which ML based sales forecasting systems can be relied on in times of great societal change like pandemics. The robustness of such a system is critical for their acceptance and adoption.

The results indicate that a crisis mode is necessary to enable the algorithm to adapt to changing consumption behavior in times of pandemics but there is no need for practitioners to switch to full "manual mode" when a pandemic occurs.

2 Conceptual and Theoretical Background

In times of a pandemic, forecast accuracy declines due to changing importance of predictor variables and changing behavior of customers [7, 12]. This study investigates the effectiveness of a covid-19 containment measures stringency index increase forecast accuracy of ML sales forecasting models during the covid-19 pandemic. Consequently, three main topics are relevant: (1) Machine Learning for Sales Forecasting, (2) Data Drift and (3) the covid-19 stringency index.

2.1 Machine Learning for Sales Forecasting

Sales forecasting is the process of predicting future sales values using a model of historical data. Forecasting is critical for corporate strategy because many organizational choices are based on it [4]. As a result, effectively anticipating future sales based on historical data is critical to a company's success [3].

Most of a retailer's essential business functions, including manufacturing and procurement, supply chain optimization, marketing, and staff planning, require high precision sales projections [15]. It is difficult to manually handle procurement for all items in all locations since merchants routinely manage assortments with many distinct products. As a result, merchants rely on automated procurement systems, which rely on sales forecasts in turn [15].

During the last few decades, significant work has gone into developing and refining advanced information systems and the underlying mathematical models and algorithms for retail sales forecasting. In the past, specialized exponential smoothing models and ARIMA models were commonly used for sales forecasting, and they still perform reasonably well with minimum computing effort [16]. When the surrounding macroeconomic conditions are generally steady, classic exponential smoothing and ARIMA models are likely to be precise [17].

When consumer behavior is rapidly changing, however, recent research suggest that nonlinear models can outperform standard techniques in terms of prediction accuracy [15]. Different forms of neural networks [18], multilayer functional link networks [19], support vector machines [20], and decision trees are just a few examples [21]. Model training is computationally costly; hence such techniques generally have unfeasible resource needs [22].

The focus of this study is sales forecasting of baked goods in a medium-sized Swiss bakery. Whereas in retail sales forecasts might be needed on an hourly basis, for this project, predictions are needed at a higher aggregation level: the next business day's sales are most important for decision makers as they need to place an order for the next day at the production department at the end of each day. This leads to relatively lower amount of data points per product and per branch. In this context, with the provided data, a conventional time series approach like ARIMA performs almost equally well. More advanced ML models like neural networks can leverage their strengths especially with higher-frequency data. Against this backdrop, an autoregressive OLS model is chosen for the elaborations in this paper, as it requires few computational resources, and its interpretability allows for analyses which would not be possible with black-box models like neural networks or boosted tree models like XGBoost [23].

In addition, food waste is increasingly recognized as a social challenge. According to the Swiss Federal Office for the Environment, 2.8 million tons of food are wasted in Switzerland every year, which equals 330 kg per person per year, or about a quarter of the environmental impact caused by our food. The food category with the greatest environmental impact due to food waste is baked goods and overproduction is a main driver of food waste in the baked goods category [5].

2.2 Data Drift and Structural Breaks

The covid-19 pandemic has had great impact on the sales of the bakery chain at hand. Prior to the pandemic's start in March 2020, the branch sold on average 22130 baked goods items per month. During the pandemic, meaning from March 2020 until and including January 2022, the sales per month were 28052 on average. This constitutes a statistically significant change of sales and promotes the assumption that the pandemic must be accounted for in a ML-based forecasting system. Figure 1 shows the total sales per month of the store and 15 products in focus of the paper. The start of the covid-19 pandemic is denoted by the vertical red line. The black line is the regression line for the pre-pandemic sales per month and the dashed line depicts the regression line for the post-pandemic sales.

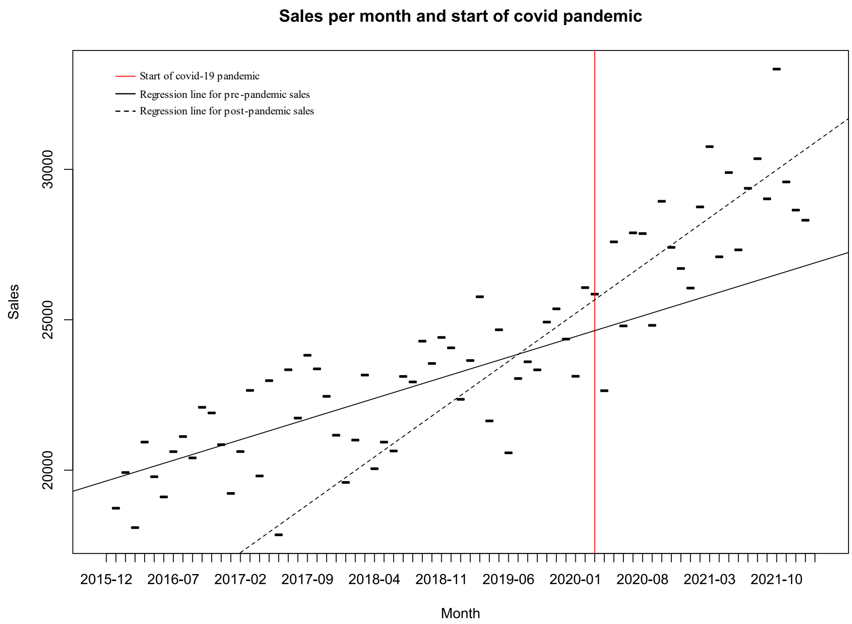


Fig. 1. Monthly sales and regression lines for pre- and post-pandemic sales.

Statistically significantly different slopes of the two regression lines indicate a structural break in the data and consequently the necessity to account for this break.

To test for a structural break in the data at the start of the pandemic in March 2020, the chow test performs a linear regression over all data points and each one over the data points prior and after the structural break point to be tested [24]. If the regression parameters of the joint regression equal the parameters of the two subsections of the time series, no structural break is detected, and the null hypothesis is accepted. Table 1 shows the results for the chow test performed on all 15 products in the data set.

Table 1. Chow test for all 15 products and respective p-values (*:p<.05; **:p<.01; ***: p<0.001).

Article Name	p-value	Structural break?	Significance level
Bread rolls	2.62E-14	Yes	***
Curd braid	3.02E-12	Yes	***
Butter braid	0	Yes	***
Pretzel croissant	0	Yes	***
Croissant	0	Yes	***
Corn croissant	2.66E-15	Yes	***
Pretzel bun	1.11E-16	Yes	***
Il pollo forte	0	Yes	***
Cream slice	0	Yes	***

Hazelnut croissant	5.69E-13	Yes	***
Roll	0.1562	No	
Bun	0.0180	Yes	*
Raspberry doughnut	0.0003	Yes	***
Sunflower croissant	5.77E-15	Yes	***
Almond croissant	0	Yes	***

As shown in Table 1, a structural break can be assumed for 13 out of 15 products on significance level of 0.001, for 1 product the significance level is 0.05. These results indicate that it is senseful to account for the start of the pandemic in March 2020 in a ML-based sales forecasting model.

2.3 COVID-19 Stringency Index

In the context of this study, "crisis" is defined as a pandemic that leads to changes in the sub-indicators listed below. Here, the covid-19 containment measures in Switzerland are quantified using the KOF StringencyPlus Index (KSI+) [25]. The index is calculated based on different sub-indicators and ranges from 0 to 100, with 0 indicating no measures and 100 indicating full lockdown. Concretely, the ten sub-indicators used to derive the KSI+ index are: school closing, workplace closing, cancellation of public events, restrictions on gatherings, closure of public transport, stay-at-home requirements, restrictions on internal movement, international travel controls, public info campaigns, and facial coverings. The sub-indicators are coded based on the Oxford Stringency Index coding methodology [26].

The KSI+ index is not specific to the COVID-19 pandemic, which is a requirement to use it in sales forecasting for a potential next pandemic. The index could hence be derived again and used as an input variable in a forecasting model. Fig. 2 depicts the course of the covid pandemic from January 2020 to February 2022. In the course of this paper, the KSI+ index is called "pandemic variable".

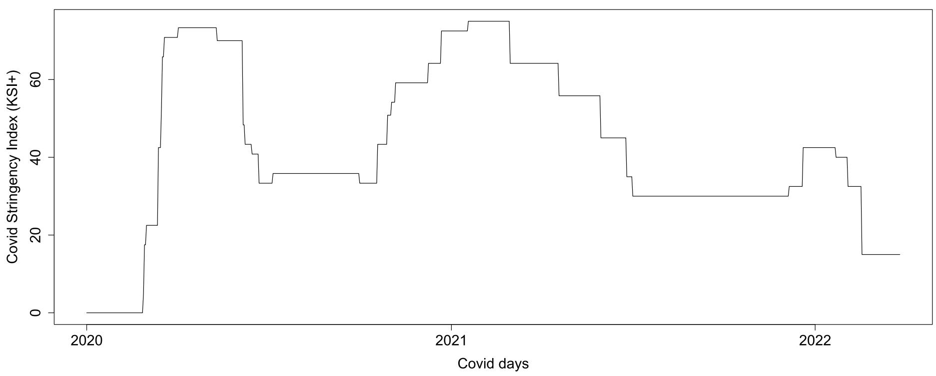


Fig. 2. Progression of covid-19 pandemic over time.

3 Methodology

In total, three analysis/ experiments are run in this paper. First, the forecast accuracy of two different models is compared. One model uses the pandemic variable as input, the other does not. Apart from that, both models are the same. In this analysis the training set consists of approximately 1500 non-covid and 300 covid days. The test set consists of around 400 covid days. Second, a simulation experiment where one-day-ahead predictions are made followed by continuous retraining is performed. Here, the training set consists of all non-covid days and the test set of all covid days (~700). This experiment is conducted to simulate a situation in which a pandemic such as the covid pandemic is spreads and ML forecasting models need to adapt. Again, two models are used: One that uses the pandemic variable and one that does not. Third, the simulation is replicated with one change: the regression coefficient corresponding to the pandemic variable is set, not learned. This simulation is performed to test the possibility to transfer learnings from the COVID-19 pandemic to upcoming pandemics. The prediction accuracy of the model that uses the pre-trained pandemic variable is compared against a model that uses the pandemic variable as-is and a model that uses no pandemic variable.

4 Data and Artifact

In the following, the data set and the used algorithm are described in detail.

4.1 Data Set

The data set used in this paper consists of transactional point-of-sale data over 5 years and is provided by a Swiss bakery chain. It covers a total of 901565 transactions over a timespan starting in December 2015 until February 2022 (2182 selling days). To confirm that the findings of this paper are valid and impactful for a typical bakery, the data set contains the 15 top-selling products. Included variables are a timestamp, the type of product, the sold amount, and the price. Based on the timestamp, the generated features weekday and month are included. External data was used to add weather indicators, specifically daily averages of radiation, rainfall, and temperature, as well as national holidays, (school) vacations and the pandemic variable. Performing extensive autocorrelation analysis, lagged features (1-day-lag, 2-day-lag, and 7-day-lag) as well as a rolling average over 7 days were included. Table 2 shows all products that are included in the analyses and the respective amount of selling (covid) days for each product. The amount of selling days is relevant as one-day-ahead predictions are made so that the training set size per product is the amount of selling days per product.

Table 2. Analyzed products and the respective amount of selling (covid) days.

Article Name	ID	#Selling days	#Selling covid days
Bread rolls	1	2182	698
Curd braid	2	2027	698

Butter braid	3	1994	697
Pretzel croissant	4	2182	698
Croissant	5	2182	698
Corn croissant	6	2182	698
Pretzel bun	7	2182	698
Il pollo forte	8	1870	600
Cream slice	9	2176	696
Hazelnut croissant	10	2182	698
Roll	11	2182	698
Bun	12	2176	698
Raspberry doughnut	13	1845	604
Sunflower croissant	14	2182	698
Almond croissant	15	2182	698

4.2 Algorithm: OLS model

Based on this data and these features, an ML algorithm predicts sales of the next day for each product. Specifically, an Ordinary Least Squares (OLS) regression model makes one-step ahead forecasts, that is, the sales for the next day.

Extensive hyperparameter tuning leads to slightly better predictions for XGBoost [23] compared to OLS regression. However, due to its low computational effort and the interpretability of its parameters, OLS regression is used for the analysis in this paper. In addition, there are no indications that the findings of this paper would not apply in the case of an XGBoost model. On contrary, due to its higher complexity and its ability to find non-linear relationships between the different input variables, it is most likely that the positive influence of the pandemic variable on the prediction accuracy is even higher when a XGBoost model is used.

5 Results

To answer RQ1, the data is split into a training set and a test set by a 75/25 percent split (no randomization, as the order of data is relevant in time series forecasting). Thus, the training set contains both times of pandemic and times of no pandemic. The test set contains solely times of pandemic, as the pandemic makes up more than a quarter of the total available data. For each product, two models are trained using the training set: One model that uses the pandemic variable (model-1) and one that does not (model-2).

Model-1:

```
Sold ~ weekday + temperature + rain + sun + lag1 + lag2 +
lag_week + holiday + mean_week + school + covid
```

Model-2 uses the same input variables but lacks the variable `covid`. In model-1, the variable `covid` is highly significant (p-value < 0.001) for eight products, significant (p-value < 0.01) for three products, and weakly significant (p-value < 0.05) for one product. For one product, the p-value is < 0.1. Subsequently, for both models, forecasts for each product are made based on the test set. The forecast accuracy is measured in terms of RMSE (root mean squared error). Table 3 depicts the results of this experiment. In the third column, the deviation between both models is shown for each product.

Table 3. Different products and respective prediction accuracy with and without pandemic variable.

Article Name	Prediction w/ pandemic variable (RMSE)	Prediction w/o pandemic variable (RMSE)	Deviation (RMSE)
Bread rolls	14.63	15.31	+4%
Curd braid	7.24	7.20	-1%
Butter braid	16.32	19.69	+17%
Pretzel croissant	16.20	17.24	+6%
Croissant	50.17	52.84	+5%
Corn croissant	15.95	16.69	+4%
Pretzel bun	12.81	13.63	+6%
Il pollo forte	3.53	3.83	+8%
Cream slice	24.38	26.19	+7%
Hazelnut croissant	12.95	13.24	+2%
Roll	10.70	10.85	+1%
Bun	11.96	12.07	+1%
Raspberry doughnut	26.79	27.19	+1%
Sunflower croissant	16.16	16.87	+4%
Almond croissant	11.48	11.64	+1%
Mean	16.75	17.63	+4.61%

Averaged over all products, model-1 outperforms model-2 by 4.61%. The outperformance differs per product, as some products are affected by the pandemic containment measures more strongly than others. In case of curd braid, including the pandemic variable even led to less accurate forecasts on the test set, possibly due to overfitting on the training set.

To further test the beneficial influence of the pandemic variable on the forecast accuracy, a simulation is run to simulate the start of the covid pandemic. During the iterative simulation, similar to a real-world situation, a model is trained based on all available historic data $[t_{-n}: t_0]$, t_0 being the last non-covid day, and provides predictions for each product for the next day (t_1). The training set hence contains all non-pandemic days. In the next iteration (it_2), the previously predicted day becomes part of the training set (now $[t_{-n}: t_1]$) and the next day (t_2) is predicted. These iterations are continued until t_{m-1} is the last day in the training set to predict t_m (left illustration in Figure 3).

The second simulation represents an attempt to use the information content about the response of sales to changes in the covid index already at the beginning of a pandemic. Since there has not yet been an onset of a new pandemic in which the regression coefficient from the covid-19 pandemic could be used, the second simulation is intended to help provide an estimate of whether the use of the learned regression coefficient from the covid-19 pandemic may be feasible in a next pandemic. To achieve this, the same procedure is used as in the first simulation, but the coefficient of the pandemic variable is determined based on all available data before and after the day to be predicted ($train_1 + train_2$). Subsequently, the pre-trained covid coefficient is cached and the model is re-trained based solely on the available historic data $[t_{-n}: t_0]$, t_0 being the last non-covid day ($train_1$). The right illustration in Figure 3 shows $train_1$ and $train_2$ and the forecasting horizons for each iteration ($it_1, it_2, ..$). However, the covid coefficient of the model is set to the value of the pre-trained cached covid coefficient. Consequently, the forecasting model has no access to future data (after t_0) as it would be the case in a real scenario.

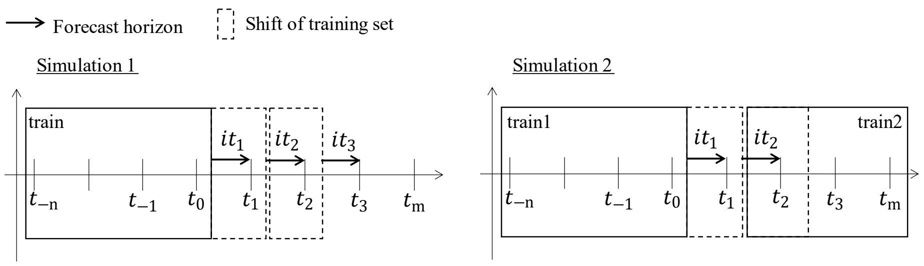


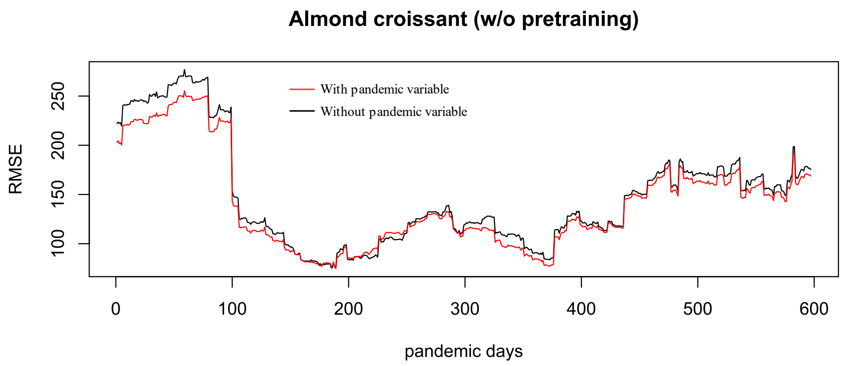
Fig. 3. Sales per month, regression lines for pre- and post-pandemic sales and start of covid pandemic.

The results of Simulation 1 and 2 indicate that over all covid-19 pandemic days, including the pandemic variable (model-1) increases the forecast accuracy by 2% (averaged over all products). If the pandemic variable is pre-trained (model-3), the deviation to the model without pandemic variable (model-2) is even higher (2.28%). However, there are differences between different time periods: During the first 100 days of the covid-19 pandemic, model-1 outperforms model-2 only slightly (0.54%). In this period, model-1 and model-3 differ the most, as model-3 outperforms model-2 by 2.4%. After 300 covid days, model-1 and model-3 have almost aligned with each other, as the value of the pandemic variable in both models gradually become similar to each other.

Table 4. Comparison of forecast accuracy (RMSE) with vs. without pandemic variable for different periods during the covid pandemic for Simulation 1 (without pre-trained pandemic variable) and Simulation 2 (with pre-trained pandemic variable).

Article ID	All days (Sim. 1)	All days (Sim. 2)	< 100 days (Sim. 1)	< 100 days (Sim. 2)	> 300 days (Sim. 1)	> 300 days (Sim. 2)
1	0.29%	0.56%	-1.91%	0.80%	0.98%	0.78%
2	0.40%	0.08%	0.45%	4.58%	-0.85%	0.09%
3	9.55%	9.81%	13.35%	14.43%	8.06%	7.73%
4	2.54%	2.98%	0.95%	1.63%	2.12%	2.10%
5	1.15%	1.50%	-1.89%	-1.87%	1.55%	1.43%
6	1.31%	1.70%	-1.21%	0.89%	1.80%	1.67%
7	2.79%	3.05%	-0.26%	-0.60%	3.28%	3.15%
8	1.94%	1.71%	1.47%	2.41%	0.75%	0.86%
9	1.13%	1.26%	-0.70%	0.72%	2.11%	2.18%
10	2.63%	3.12%	3.82%	7.23%	1.56%	1.86%
11	0.76%	1.23%	-3.52%	-1.60%	2.15%	1.84%
12	-0.16%	0.16%	-4.92%	-3.30%	1.10%	0.41%
13	0.98%	1.52%	-3.01%	1.44%	1.04%	1.01%
14	2.12%	2.67%	-1.36%	-0.46%	3.10%	2.64%
15	2.49%	2.86%	6.81%	9.88%	1.83%	1.99%
Mean	**2.00%**	**2.28%**	**0.54%**	**2.41%**	**2.04%**	**1.99%**

These effects can be exemplarily seen in Figure 4, which shows two the forecast accuracy in terms of RMSE for model-1 (no pandemic variable) vs. model-2 (with pandemic variable) in the upper illustration and for model-2 (with pandemic variable, but no pretraining) vs. model-3 (with pre-trained pandemic variable) in the lower illustration.



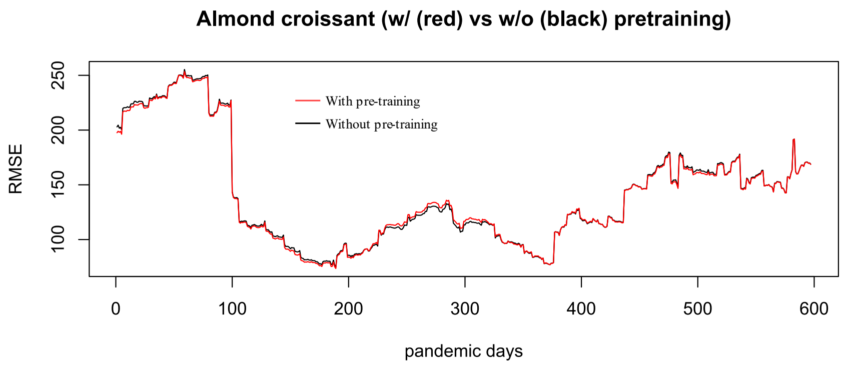


Fig. 4. Comparison between model-1 and model-2 (upper illustration) and model-2 and model-3 (lower illustration) in terms of forecast accuracy (RMSE) during covid-19 pandemic.

The lower illustration of Figure 4 depicts the difference between pre-training and no pre-training. After around 300 days of pandemic, both models do not differ much anymore. Table 5 shows average forecast accuracy in terms of RMSE for the beginning of the covid-19 pandemic (first 200 days) vs. the following 200 days (mid-pandemic). For all three groups (no pandemic variable, covid-variable, pre-trained covid-variable), the forecast accuracy is >10% lower (i.e., higher RMSE) in the beginning of the pandemic compared to mid-pandemic. However, including the pandemic variable into the model decreases the average RMSE in the beginning of the pandemic by 9.3 and including a pre-trained pandemic variable decreases the average RMSE by 15.7.

Table 5. Forecast accuracy (RMSE) with and without pre-trained and not pre-trained pandemic variable for different periods during the covid pandemic (deviation between first 200 days vs. day 200 – 400).

Scenario	Mean (RMSE)	Deviation
days < 200 (pandemic variable, no pre-train)	404.6	
200 > days <400 (pandemic variable, no pre-train)	353	11.80%

days < 200 (no pandemic variable)	413.9	
200 > days < 400 (no pandemic variable)	374.1	14.50%
days < 200 (pre-trained pandemic variable)	398.2	
200 > days < 400 (pre-trained pandemic variable)	353.2	14.20%

To summarize, RQ1 can be answered: Including the COVID-19 stringency index (pandemic variable) improves one-day-ahead forecasts in sales forecasting in the bakery domain. Both the bulk prediction (Table 3) and the iterative predict-retrain-predict process (Table 4 and Table 5) show that the pandemic variable is beneficial for the chosen ML forecast model in the bakery domain. RQ2 can be partly answered: In our case, pre-training the pandemic variable based on all available sales data during the covid pandemic has increased forecast accuracy more than using a non-pre-trained pandemic variable. However, there is no data for the same products during another pandemic. Consequently, RQ2 cannot be fully answered as it would require training the pandemic variable based on one pandemic and applying it during the next pandemic.

6 Discussion and Conclusion

In this study, the usefulness of a variable ("pandemic variable") that operationalizes the stringency of covid-19 containment measures in Switzerland for ML-based sales forecasting algorithms was tested. For the operationalization, the KOF StringencyPlus Index (KSI+) [25] was used. It was shown that covid-19 containment measures have a significant impact on the sales of a medium-sized Swiss bakery chain. Thus, including the pandemic variable improves accuracy in ML-based sales forecasting of baked goods.

Three main experiments, for which real data from a medium-sized Swiss bakery was used, were conducted. As a basis, an autoregressive OLS model was used. First, the influence of the pandemic variable on the forecast accuracy in a bulk forecast of ~500 days during the covid-19 pandemic was evaluated. On average, including the pandemic variable increased forecast accuracy by 4.6%. Second, a simulation was run. During this simulation, the algorithm made one-step-ahead sales forecast and was retrained after making the forecasts and comparing it to the actual value. The actual sales are known because the simulation was set in the past. This simulation yielded promising results: the forecast accuracy of the model that uses the pandemic variable was 2% higher over all 15 products and days. Third, the same simulation was run, but the pandemic variable was pre-trained based on the full data set. The pre-trained pandemic variable was used in the iterative setting as described above. In this case, the forecast accuracy was 2.3% higher compared to the model without a pandemic variable.

The containment measures stringency index is generic, meaning that the factors influencing the index value are not specific to the covid-19 pandemic. This way, the stringency index could be determined and applied in ML-based sales forecasting in future pandemics. Findings of related work confirm that the covid-19 pandemic poses a challenge to ML-based forecasting [7, 12]. Several studies deal with the challenge by optimizing the forecasting algorithm itself without including additional feature variables

[8-11]. One study [13] aims to increase forecast accuracy by considering expert opinions as additional input variables in the model. Another study [14] adds a feature variable that indicates the different stages of the covid-19 pandemic to increase forecast accuracy. In opposition to the study at hand, a temporal perspective is used, focusing on the phases of the pandemic, rather than a perspective that considers the actual constraints imposed on the population.

An advantage of the chosen methodology is the use of real sales data. Since also the progression of the pandemic variable has not been changed in hindsight, there is no doubt that the pandemic variable would have really improved forecasts. In addition, the pandemic variable is deterministic and can be incorporated without a lag. While other variables like weather or Google Trends Index need to be forecasted themselves (weather) or included with a lag of one day (Google Trends Index), the value of the pandemic variable is known around one week early.

However, this work does not come without limitations and potential for future research. Decision makers who are responsible for production orders in bakeries have three options in case of future pandemics. First, they can decide to use no pandemic variable at all. In Simulation 1, the test set contains all non-covid days. One-step-ahead forecasts followed by consecutive retraining was done using the pandemic variable. Although the pandemic variable needs some time to adapt, even during the first 200 days of covid-19, the model that uses the pandemic variable outperformed the model that did not. These results indicate that the pandemic variable should be used, which is the second option. Third, decision makers can use the pandemic variable trained based on the covid-19 pandemic and use it their forecasting model in the beginning of the next pandemic. It is difficult to validate the effectiveness of option 3 because the dynamics of a future pandemic are not known yet. It is not certain that the relationship between containment measures and sales remain the same. Customers possibly change their behavior, for example due to accustoming to pandemics in general. To evaluate this, further experiments with data from different pandemics, would need to be done. Although the results in this paper are encouraging, more sophisticated transfer learning methods [27] might be useful to optimize forecasts in the future. In addition, future research could repeat the same experiments using XGBoost or neural networks to compare the findings of this paper with different algorithms.

In summary, the results of this study indicate that a pandemic variable is useful to improve forecasts accuracy in times of pandemics. In this sense, a crisis mode is necessary to enable the algorithm to adapt to changing consumption behavior in times of pandemics.

This study contributes to practice by showing the usefulness of variables that measure the strictness of governments' containment measures as a reaction to pandemics in ML-based sales forecasting algorithms. This study contributes to theory by evaluating the degree to which ML based sales forecasting systems can be relied on in times of great societal change like pandemics. The robustness of such a system is critical for their acceptance and adoption.

References

1. Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., Nelson, C.: Clinical versus mechanical prediction: a meta-analysis. Psychol. Assess. 12, 19–30 (2000).
2. Blohm, I., Antretter, T., Sirén, C., Grichnik, D., Wincent, J.: It's a Peoples Game, Isn't It?! A Comparison Between the Investment Returns of Business Angels and Machine Learning Algorithms. Entrepreneurship Theory and Practice. (2020).
3. Sun, Z.L., Choi, T.M., Au, K.F., Yu, Y.: Sales forecasting using extreme learning machine with applications in fashion retailing. Decision support systems. 46, 411–419 (2008).
4. Mentzer, J.T., Bienstock, C.C.: Sales Forecasting Management: Understanding the Techniques, Systems and Management of the Sales Forecasting Process. SAGE Publications (1998).
5. Beretta, C., Hellweg, S.: Lebensmittelverluste in der Schweiz: Mengen und Umweltbelastung. Wissenschaftlicher Schlussbericht. ETH Zürich (2019).
6. Cam, A., Chui, M., Hall, B.: Global AI Survey: AI proves its worth, but few scale impact. McKinsey Global Institute. (2019).
7. Heaven, W.D.: Our weird behavior during the pandemic is messing with AI models, https://www.technologyreview.com/2020/05/11/1001563/covid-pandemic-broken-ai-machine-learning-amazon-retail-fraud-humans-in-the-loop/?utm_source=nextdraft&utm_medium=email, last accessed 2022/08/25.
8. Liu, P.: Intermittent demand forecasting for medical consumables with short life cycle using a dynamic neural network during the COVID-19 epidemic. Health Informatics J. 26, 3106–3122 (2020).
9. Lakhal, D.A., Hadj Kacem, S.B., Tagina, M., Amara, M.A.: Prediction of psychiatric drugs sale during COVID-19. In: 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE). pp. 1–6 (2021).
10. Nigam, B., Shukla, A.C.: Sales forecasting using Box Jenkins method based Arima model considering effect of covid -19 pandemic situation. International Journal of Engineering Applied Sciences and Technology. (2021).
11. Nirmala, V.W., Harjadi, D., Awaluddin, R.: Forecasting by Using Exponential Smoothing Method and Trend Method to Optimize Product Sales in PT. Zamrud Bumi Indonesia During the Covid-19 Pandemic. Int. J. Eng. (2021).
12. Sindihebura, T.T., Pu, X., Chen, J.: Predictors of Sales and the Covid-19 Disruption: Evidence from an Online Marketplace. In: PACIS 2022 Proceedings (2022).
13. Zhang, H., Song, H., Wen, L., Liu, C.: Forecasting tourism recovery amid COVID-19. Ann. Touris. Res. 87, 103149 (2021).
14. Leenawong, C., Chaikajonwat, T.: Event Forecasting for Thailand's Car Sales during the COVID-19 Pandemic. Brown Univ. Dig. Addict. Theory Appl. 7, (2022).
15. Fildes, R., Ma, S., and Kolassa, S.: Retail Forecasting: Research and Practice. International Journal of Forecasting (2019)
16. Ramos, P., Santos, N., and Rebelo, R.: Performance of State Space and Arima Models for Consumer Retail Sales Forecasting. Robotics and Computer-Integrated Manufacturing (34), 151-163 (2015).
17. Zhang, G. P., and Qi, M.: Neural Network Forecasting for Seasonal and Trend Time Series. European Journal of Operational Research 160(2), 501-514 (2005).
18. Veiga, C. P. D., Veiga, C. R. P. D., Puchalski, W., Coelho, L. D. S., and Tortato, U.: Demand Forecasting Based on Natural Computing Approaches Applied to the Foodstuff Retail Segment. Journal of Retailing and Consumer Services (31), 174-181 (2016).

19. Chen, F. L., and Ou, T. Y.: Sales Forecasting System Based on Gray Extreme Learning Machine with Taguchi Method in Retail Industry. Expert Systems with Applications 38(3), 1336-1345 (2011).
20. Di Pillo, G., Latorre, V., Lucidi, S., and Procacci, E.: An Application of Support Vector Machines to Sales Forecasting under Promotions. Quarterly Journal of Operations Research 14(3), 309-325 (2016)
21. Ali, Ö. G., Sayın, S., Van Woensel, T., and Fransoo, J.: Sku Demand Forecasting in the Presence of Promotions. Expert Systems with Applications 36(10), 12340-12348 (2009).
22. Gür Ali, Ö., and Yaman, K.: Selecting Rows and Columns for Training Support Vector Regression Models with Large Retail Datasets. European Journal of Operational Research 226(3), 471-480 (2013).
23. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System, http://arxiv.org/abs/1603.02754, (2016).
24. Chow, G. C.: Tests of Equality Between Sets of Coefficients in Two Linear Regressions. Econometrica: Journal of the Econometric Society 28(3), 591–605 (1960).
25. Pleninger, R., Streicher, S., and Sturm, J.-E.: Do COVID-19 Containment Measures Work? Evidence from Switzerland. Swiss Journal of Economics and Statistics 158(1), 5 (2022).
26. Hale, T., Angrist, N., Cameron-Blake, E., Hallas, L., Kira, B., Majumdar, S., Petherick, A., Phillips, T., Tatlow, H., & Webster, S.: Variation in government responses to COVID-19. BSG Working Paper Series. Blavatnik School of Government. University of Oxford (2020).
27. Obst, D., Ghattas, B., Cugliari, J., Oppenheim, G., Claudel, S., and Goude, Y.: Transfer Learning for Linear Regression: A Statistical Test of Gain. In: arXiv [math.ST] (2021).