

---

# Predicting seasonal influenza using supermarket retail records

Leonardo Vona

Data Mining  
December 9, 2022



# Seasonal influenza

Seasonal influenza is a major burden to the health care systems, with 3 - 5 millions infected worldwide every year.

The study explores whether the inclusion of retail records in a predictive model improves seasonal influenza forecasting.



# Baskets as proxy for seasonal flu



## Retail data as proxy

Items purchased in a shopping cart are a good proxy of consumers' behavioral changes.

Idea: exploit the retail data to capture the spread of seasonal flu

## Baskets instead of products

Instead of using single items, consider sentinel baskets (products bought together) as a proxy for the actual seasonal flu.

# Obtaining sentinel baskets

## Products

Identify a set of sentinel products whose volume of purchase is historically correlated with the previous flu season.

## Customers

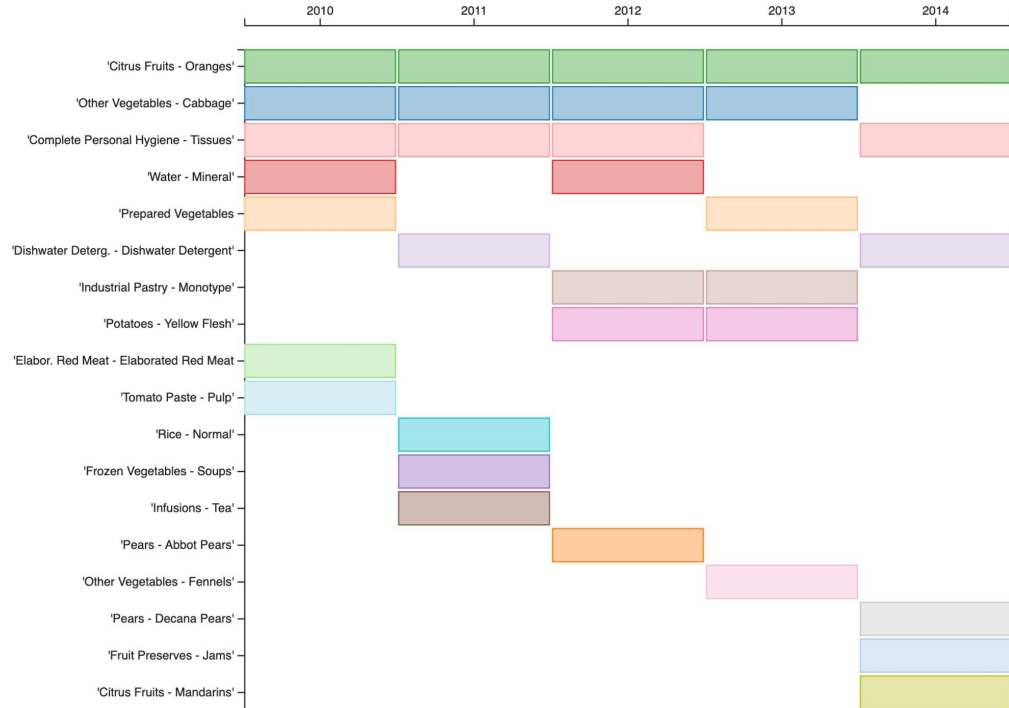
Consider the whole purchase history of customers buying sentinel products.

## Baskets

Using an Apriori algorithm, identify sentinel baskets.

# What's in sentinel baskets?

Many products in sentinel baskets are rather common to prepare a healthy meal that could benefit a sick person.



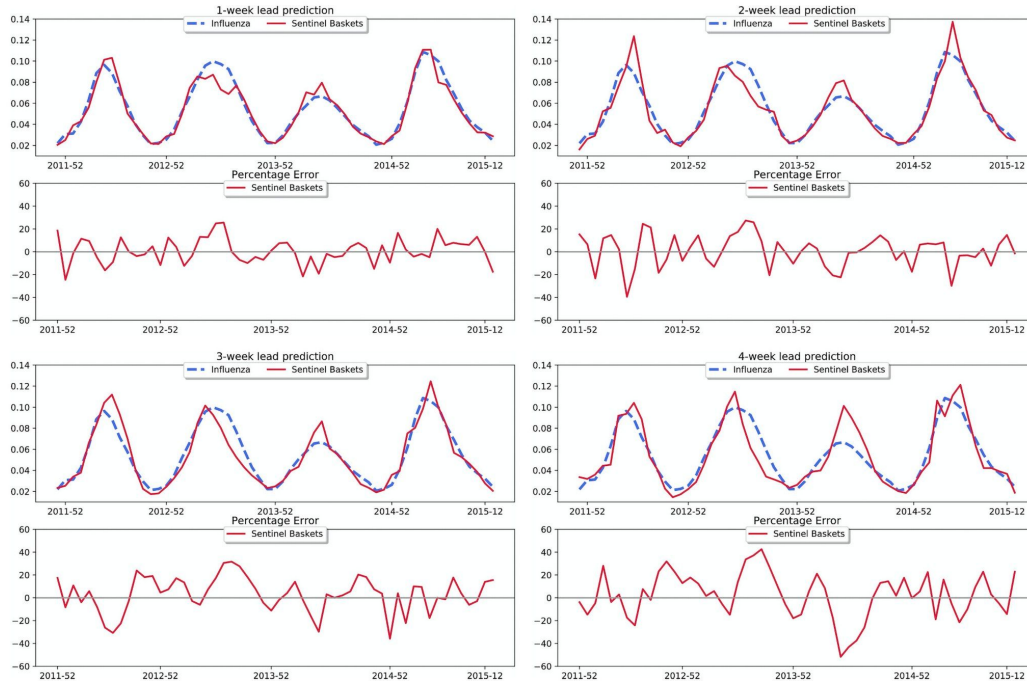
# Regression model

By using sentinel baskets purchases, develop a nowcasting and forecasting algorithm that provides seasonal flu incidence in Italy estimates up to 4 weeks ahead.

- ▷ SVR (Support Vector Regression) model used to produce the predictions.

$$I_{t^s+k} = \alpha^k + \sum_{i=0}^{h-1} a_i^k I_{t^s-i} + \sum_{n=1}^{N_S} \sum_{j=-1}^{h-1} b_j^{kn} S_{t^s-j}^n$$

# Predictions vs ground truth



The predictions accuracy show that the sentinel baskets track the influenza activity level very accurately.

# Assessing the results

	Pearson correlation				MAPE				RMSE			
	1 week ahead	2 weeks ahead	3 weeks ahead	4 weeks ahead	1 week ahead	2 weeks ahead	3 weeks ahead	4 weeks ahead	1 week ahead	2 weeks ahead	3 weeks ahead	4 weeks ahead
<i>autoreg</i>	0.95	0.82	0.76	0.77	9.79	19.65	24.15	27.79	0.79	1.53	1.81	1.77
<i>Product-5</i>	0.60	0.49	0.28	0.01	41.47	41.80	44.22	51.07	2.88	3.07	3.42	3.76
<i>Basket-1</i>	0.96	0.94	0.94	0.91	8.77	11.48	12.29	16.65	0.74	0.99	0.97	1.24
<i>Basket-5</i>	0.96	0.94	0.93	0.87	11.80	13.48	14.77	17.62	0.75	0.95	1.02	1.35

## THE ACCURACY IMPROVES

The results indicate that the seasonal influenza forecast accuracy improves with the use of retail records.

## BASKETS VS PRODUCTS

Forecasts obtained by using sentinel baskets are significantly more accurate than those obtained using single products' time series.

## EVALUATION METRICS

The forecast performance is evaluated using Pearson correlation, mean absolute percent error and root mean square error.



# Thanks!

## References:

Miliou I, Xiong X, Rinzivillo S, Zhang Q, Rossetti G, Giannotti F, et al. (2021) Predicting seasonal influenza using supermarket retail records. PLoS Comput Biol 17(7): e1009087

<https://doi.org/10.1371/journal.pcbi.1009087>