A knowledge-intensive methodology for explainable sales prediction

Matteo Cristani[a,*], Luca Pasetto[a], Claudio Tomazzoli[a]

[a]Dipartimento di Informatica, University of Verona, Italy

Abstract

Sales prediction in food market is a complex issue that has been addressed in the recent past with machine learning techniques. Although some promising results, an experimental work that we describe in this paper shows some drawbacks of the above mentioned data-driven method and habilitates the definition of a novel methodology, strongly involving *a piori* knowledge.

1. Introduction

Predicting sales is a quite complex issue in machine learning. During the pandemic of COVID-19 some specific commercial sectors have been strongly affected, in different directions, by the severe lockdown that has been spread out in the world. We had the chance to observe the performance of a prediction system in use in the field of *food production* during the pandemic, and could identify some general issues of the prediction approaches currently used in literature for the above mentioned sector.

We found that, essentially, *data-driven* methods of prediction are less effective than *knowlede-driven* ones, determine the reasons for this, and devise a general-purpose methodology that reduces the risks that are associated with the data-driven prediction approach.

To illustrate the nature of the problem from a general viewpoint, we report an example of the problems that we may encounter in the development of the methodology.

* CMatteo Cristani
E-mail address: matteo.cristani@univr.it

Example 1. *A dataset formed by past promotional sales (STORI) that incorporates hundreds of customers, each of which has hundreds of promotions for hundreds of products is used as the base to predict the behaviour of promotions that are sheduled in the future (PREVI). The purpose of a prediction process is twofold:*

- *Determine the total amount of that particular product to be sold in the promotional period, by means of determining the daily average sellout (DAS) that is used, along with the promotion duration to determine the Total Promotional Sellout (TPS). This problem is named* Sale Forecast Problem *(SFP);*
- *Split the TPS onto Daily Promotion Sellout (DPS) associated with each day of the promotion. This problem is named the* Sale Split Problem*.*

In this paper we shall only focus on the SFP problem. The Key Performance Indicator used to test the effectiveness of the methods we apply is the accuracy of the forecast (ACC), measured as the absolute value of the difference between the forecast value (PREVI) and the actual measure on the runset, provided on the actual data of the sellout.

The benchmark of the KPI is the value of ACC of the model as compared to ACC of the forecast made by the sales persons.

The paper is organized as follows. Section 2 discusses and formalises the problem, Section 3 shows where the effort made by Machine Learning methods fail, and Section 4 provides a description of the methodology we devised. Section 5 introduces related work, while Section 6 sketches further work and takes some conclusions.

2. Forecasting sales: the issue

In this paper we use the following standard terminology for statistical learning methods:

- **Dataset** is a table with a set of structured data. We do not make any assumptions on the nature of this, but when a dataset is a proper database table with a primary key, we indifferently name it a *table*.

- **Training set** is a dataset used to train a classification model.

- **KPI** or *key performance indicators* are measures of the quality of a machine learning technology.

- **Predictors** are those functional values in a dataset that are measured in a KPI.

- **Milestone value** is a threshodl value for KPI that is used to measure the performance of predictors. A predictor that is validated on a KPI against a milestone value is named *good*.

- **Testset** is a dataset used to measure the KPI of a model, by comparing the forecast with the actual results. When a model generates a forecasting, it issues it on a *validation set*, namely the testset without the forecast columns, that are generated by the model. The testset is then measured in comparison to the forecasting added to the validation test by means of the KPI. A testset is used to validate and tune the model, and is formed by objects of forecasting that have already been computed in the real world.

- **Runset** is a live forecasting set on which we add the forecasting.

- **Statistical anomaly** is a record in a dataset that behave in a way that differs in a significant way from the behavior of the other values in the dataset itself. Tests for statistical anomaly are for instance Guttman's G index or or U3 index. A dataset that does not contain statistical anomalies is named *cleaned* and a dataset that has not yet been cleaned up is said to be *rough*.

- **Big** datasets and **Small** datasets are determined by thresholds of anomaly on size. In particular, it is absolutely small a dataset containing no or one record, and typically also some datasets containing only two. In general for small datasets results impossible to determine correctly the unbiased standard deviation, and this generally means that we do not have a reasonable way to distinguish statistical anomalies. Small sets are generally rough,

therefore.

Consider the following problem: determine the total sellout of a promotional sale for a large retailer, detailed on *point of sale* and single *product*. The resulting data are henceforth named *promotional forecast*, and consist in a table in which for every point of sale (henceforth named a *customer*) and each product we compute (or decide not to compute) the forecast.

Usually we can assume, without global risks, that in regular time periods, the behaviour of *similar promotional sales* in the past may be a good predictor of the forecast, based on the following assumptions that are justified by common sense and in some specific cases, explicitly indicated, are also related to general axioms of descriptive statistics.

The approach based on rules and founded on the above assumptions is named in the current literature *explainable AI*, as suggested in the specific case of the domain of interest in [16].

Assumption 1. *Mean and potentially other* position indices *are good predictors for reasonably low* variability indices.

The above assumption is the logical base of classification methods, including in particular those based on the *discrimination indices* such as Gini Index. When we can run a machine learning method based on classification approaches, that is generated by, for instance, random forest methods, or neural networks, or again, deep learning systems, the assumption allows to distinguish several types of statistical anomalies, that may need to be considered. This is the concern of Assumption 2. In the following assumptions we name a set of data

Assumption 2. *Every training set of a prediction model for sales that is cleaned up by statistical anomalies has a better performance on forecast than a roguh one.*

The ground for usage of good predictors is the ability of a training set to issue adequate forecasting. This lies on the assumpton 3.

Assumption 3. *Small training sets, or small groupings in training sets generate bad predictors.*

Given the above mentioned assumptions, we are now able to devise the specific problem of promotional forecasting in food production for big retailers. First of all, let us formalise the specialised problem.

We are concerned with two large (but not very large) datasets: the continuous customer table (CCT) and the promotional customer table (PCT). The CCT and PCT contain the selling data of every point of sale of one specific retailer, for every day of one year (may 2018-may 2019) for every product of one specific food provider. The food provision regards some categories of *fresh food*, that have different expiry times (variable from a few days towards a few weeks) and some categories of *frozen food* with longer expiry dates. It is important to observe that, obviously, shortage of a specific item in a point of sale during sales promotion is economically undesirable, and analogously unsold items constitute a loss, for they might be either sold at lower price or sent back to the distibution.

We also have the validation test for promotional customer table, that is performed on an excpert of the PCT on the period September-December 2019. This dataset is provided in two versions: without the forecast column, that is the *average daily sellout* (ADS) during the promotional period for that specific product, and with the column. In the first version is the *testset* for the models, and in the second version is the validation model. We also have tests run on the live data of March and April. The KPI we value in this paper is only one: the percent of the difference between ADS in the forecast and in the actual tested dataset, against the total sellout in that period at detail of point of sale and product. Milestone forecast is provided and therefore the quality of a model, and therefore, if possibly determined, the quality of predictors is settled to the comparison to the threshold. The milestone is the forecast error in prediction operated by *sales units* while employing empirical forecasting methods.

3. Machine learning: evidence of a struggle and an explanation

The run of the experiments has been threefold:

- On the first run, we have tested many different models, including Neural Networks, Associative Rules, ILP, Random Forest. On the first run it neatly emerged that Random Forest was the unique acceptable candidate. Preliminary results obtained in this phase are omitted for the sake of space.

- On the second run we systematically analyse the performance of Random Forest with different refinements. Due to some intrinsic limits that we determined and used to envision a change in the methodology, we issues a specific *bayes model*, that we name the SFBN (sales forecasting belief network) that outperformes Random Forest.
- On the third and last run we refine the SFBN with many different combinations of rules and finally obtain the SBNF^{+} approach that makes use of a set of rules and is generated by a specific analytical methodology.

In this section we illustrate the experimental results. We have executed the below tests, following the dynamics of differential evaluation of each model towards the KPI on the milestone values, as referred to the last two above mentioned phases.

Potential predictors are in the STORI table are:

CC Client code; that is the detail of the promotional classification;

PC Product code; that is also the detail of the promotional classification;

PrC Promotional code;

PD Duration of the promotion;

WD Weekday of the promotion start;

DS Daily Sale;

D_clust cluster based on duration of the promotion;

S_clust cluster based on size of customer ;

ADS_SDev standard deviation of Average Daily sales;

ADS_sclust_avg Average Daily sales of cluster based on size of customer;

ADS_sclust_sdev standard deviation of Average Daily sales of cluster based on size of customer;

ADS_dclust_avg Average Daily sales of cluster based on duration of the promotion;

ADS_dclust_sdev standard deviation of Average Daily sales of cluster based on duration of the promotion;

The forecast processes are listed below, in tables representing the values of ACC(Previ, Actual) and ACC(Sales Persons, Actual). Depth of the Random Forest is a parameter that establishes the maximum depth of the trees generated by the algorithm. In the Tables below we list the following run tests:

RF30-Full Random Forest with 30 depth using all the predictors CC, PC, PD, WD, PD, D_clust, S_clust, ADS_SDev, ADS_sclust_avg, ADS_sclust_sdev, ADS_dclust_avg, ADS_dclust_sdev ;

RF30-02 Random Forest with 30 depth using predictors CC, PC, PD, WD, PD, D_clust, S_clust, ADS_SDev, ADS_sclust_sdev, ADS_dclust_avg, ADS_dclust_sdev;

RF30-03 Random Forest with 30 depth using predictors CC, PC, PD, WD, PD, D_clust, S_clust, ADS_SDev, ADS_sclust_sdev, ADS_dclust_sdev;

RF30-04 Random Forest with 30 depth using predictors CC, PC, PD, WD, PD, D_clust, S_clust, ADS_sclust_sdev;

RF30-05 Random Forest with 30 depth using predictors CC, PC, PD, WD, PD, D_clust, S_clust, ADS_sclust_sdev ;

RF30-06 Random Forest with 30 depth using predictors CC, PC, PD, WD, PD, D_clust, S_clust ;

RF30-07 Random Forest with 30 depth using predictors CC, PC, PD, WD, PD, D_clust, S_clust, ADS_dclust_sdev;

RF30-08 Random Forest with 30 depth using predictors CC, PC, PD, WD, PD, D_clust ;

RF30-09 Random Forest with 30 depth using predictors CC, PC, PD, WD, PD, S_clust ;

RF30-10 Random Forest with 30 depth using predictors CC, PC, PD, D_clust, S_clust ;

RF30-11 Random Forest with 30 depth using predictors CC, PC, PD, ADS_SDev, ADS_sclust_avg, ADS_sclust_sdev, ADS_dclust_avg, ADS_dclust_sdev;

The accuracy of Sales prediction, provided in Table 1 is benchmarked by the accuracy in the same test period for sales persons, that is measured in 35.8%, being therefore the distance measured on the column DIFFACC(Sales Person).

Table 1: Comparison table for random forest run tests.

Model	ACC(Previ, Actual)	DIFFACC(Sales Person)
RF30-Full	46.0%	-10.20%
RF30-02	45.7%	-9.90%
RF30-03	45.7%	-9.90%
RF30-04	46.7%	-10.90%
RF30-05	48.1%	-12.30%
RF30-06	51.4%	-15.60%
RF30-07	47.9%	-12.10%
RF30-08	60.0%	-24.20%
RF30-09	51.5%	-15.70%
RF30-10	51.7%	-15.90%
RF30-11	45.8%	-10.00%

4. Knowledge Intensive Sales Prediction (KISP)

After testing on the test sets the random forest methods, because the quality of the results is definitely insufficient on a practical base, we identify a sequence of rules that apply to the different circumstances of the configurations, as exposed in this section.

A general structure of the execution flow of the aforementioned rules is presented in Figure 1.

The rules are, specifically, the structure presented in Algorithm 1. The expression *baseline* is intended to identify the pre-computed value of the average of daily total sellout at the detail client/product. Pivot is indeed the average of daily sellout at the details of the single rule levels. Duration segment, as introduced above, is the cluster of duration to which the duration of the required forecast belongs (namely similar durations[1]). Analogously cluster of dimensions for the total sellout of clients represents the group of clients of similar dimension in terms of total daily sellout.

---

[1] We omit numerical details of the segmentations for the sake of space.
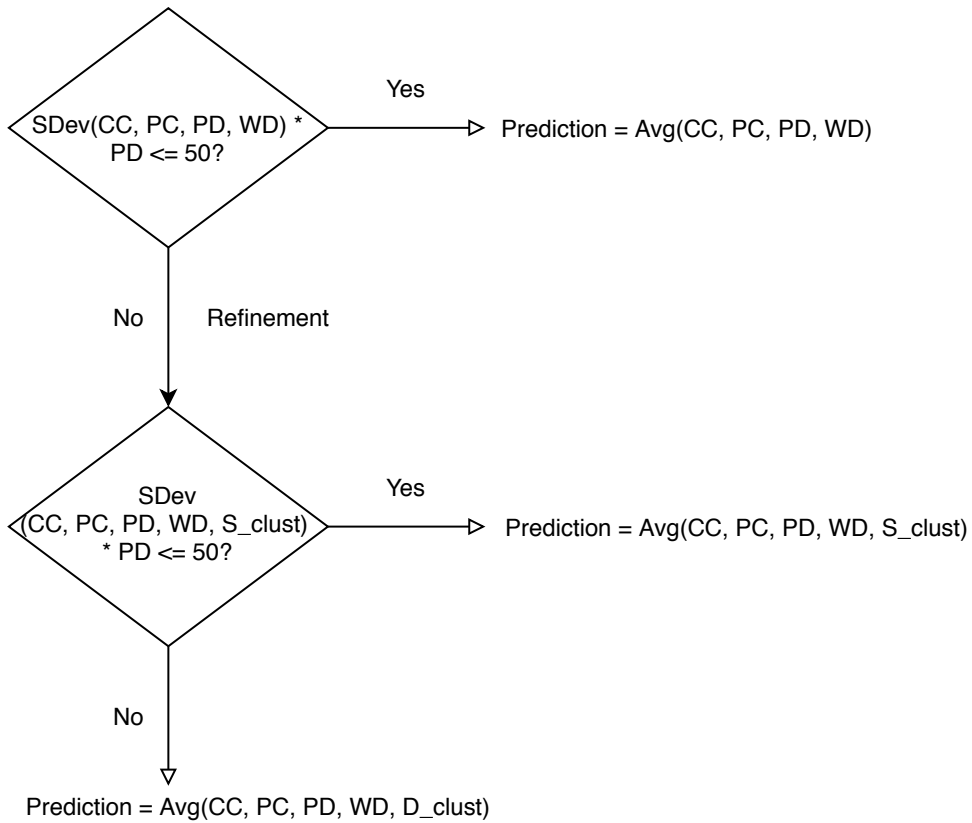
Fig. 1: The processing pipeline of analysis of sales data.

- At the detail of client/product/duration/weekday;
- At the detail of client/product/duration segment;
- At the detail of client segment/product/duration;

Accuracy of the method provided in Algorithm 1 is 32.6%, that outperforms sales persons forecast accuracy of +3.1%. The implementation technique chosen for the tests is based on the IBM technology *SPSS*. The SPSS flow of the technique is presented in Figure 2.
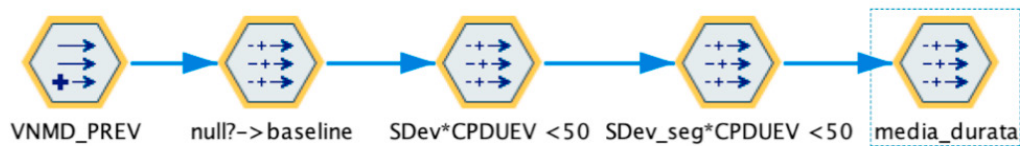

Fig. 2: The SPSS flow implementing Algorithm 1.

5. Related work

Sales forecasting is an iconic research field in Knowledge Engineering [12, 11, 2]. Several techniques, especially based on neural networks [17], However, food sales prediction has received much attention only recently. A specific

---

Algorithm 1 Calculate PREVI value of Average Daily Sellout

Require: STORI values of promotional sales in one year
Ensure: PREVI value is settled
  if average(pivot) exists **then**
    Forecast=average(pivot)
  else
    Forecast=average(daily_sellout)
  end if
  if stddev(pivot,duration,weekday) $\times$ duration ≤ 50 **then**
    Forecast=average(pivot,duration,weekday)
  else
    if stddev(pivot,duration_cluster,weekday) $\times$ duration ≤ 50 **then**
      Forecast=average(daily_sellout)
    else
      if stddev(size_*cluster*, product,duration) $\times$ duration ≤ 50 **then**
        Forecast=average(daily_sellout)
      end if
    end if
  end if

---

analysis of sales prediction peculiarities has been extensively conducted in [24], and some technical cases have been also documented, as in a recent referential survey [10] analyses the techniques examined so far by scholars of the field.

The approach we adopted here is essentially inspired by BAyesian prediction models, as in [8], and also by methods of prediction employed in presence of promotions [1]. A line of research that has proven to be very effective is the one based on Data Mining [20, 21].

An approach to rule-based classification in forecasting contexts has been attempted by some of the authors of this study [9, 4, 3], and while forecasting models in risk analysis [6, 5]. Moreover there has been a scholarly effort to devise effects of the forecasting processes on the working model of supply chain [7] that has shown the nature and impact of these behaviours.

6. Conclusions and further work

In this paper we show that a Knowledge Intensive Methodology for Sales Prediction in the food market can be more effective (and specifically more accurate) in predicting promotional sales than other techniques, and in particular the best performing ones in the machine learning field. The investigation is still in course of action and we are going to define a new method for *splitting* the forecast during the promotional period to forecast.

During this investigation we have introcuced a limited group of tests that have shown to perform better than many others we have explored and not reported here for the sake of space. It is specifically conclusive that:

- Pure machine learning techniques, in particular *Neural Networks*, *Random Forest*, *Kohonen networks*, and *Blend methods* taken from the aforementioned domain are outperfomed by *Naive Bayes Rule Systems* and also by *Associative Rules*;
- Inductive Logic Programming methods have shown to underperform with respect to Random Forest methods;
- Knowledge-Intensive approaches are more explainable, more effective, more accurate and also faster than Pure Machine Learning methods;
- Deep Learning tests have shown analogous drawbacks for the accuracy aspects, though they result quite effective in terms of coverage.

In conclusion Knowledge Engineering results essential for unbiased sales prediction in the food market, at least as related to promotions.

Future work is essentially based on the idea of using predictors to determine *geographic* and *social relationships* as a base for behaviour of final consumers determination [23, 19] among customers, made available to perform a prediction that incorporates errors of the current model. In particular we are interested in determining when a customer suddenly appears on the customer table and simultaneously someone else disappear. We shall use the researches previously performed by the same research group regarding spatial reasoning [13] and social network analysis [14, 15].

Acknowledgments

References

[1] O.G. Ali, S. Sayin, T. van Woensel, and J. Fransoo. Sku demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36(10):12340–12348, 2009.

[2] Z.-Y. Chen, Z.-P. Fan, and M. Sun. A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, 223(2):461–472, 2012.

[3] M. Cristani, E. Karafili, and C. Tomazzoli. Energy saving by ambient intelligence techniques. pages 157–164, 2014.

[4] M. Cristani, E. Karafili, and C. Tomazzoli. Improving energy saving techniques by ambient intelligence scheduling. volume 2015-April, pages 324–331, 2015.

[5] M. Cristani, E. Karafili, and L. Vigan. Towards a logical framework for reasoning about risk. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7465 LNCS:609–623, 2012.

[6] M. Cristani, E. Karafili, and L. Vigan. Tableau systems for reasoning about risk. *Journal of Ambient Intelligence and Humanized Computing*, 5(2):215–247, 2014.

[7] J. Gou, G. Shen, and R. Chai. Model of service-oriented catering supply chain performance evaluation. *Journal of Industrial Engineering and Management*, 6(1 LISS 2012):215–226, 2013.

[8] S.L. Scott and H.R. Varian. Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1-2):4–23, 2014.

[9] C. Tomazzoli, M. Cristani, E. Karafili, and F. Olivieri. Non-monotonic reasoning rules for energy efficiency. *Journal of Ambient Intelligence and Smart Environments*, 9(3):345–360, 2017.

[10] G. Tsoumakas. A survey of machine learning techniques for food sales prediction. *Artificial Intelligence Review*, 52(1):441–447, 2019.

[11] W.K. Wong and Z.X. Guo. A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*, 128(2):614–624, 2010.

[12] Y. Xie, X. Li, E.W.T. Ngai, and W. Ying. Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3 PART 1):5445–5449, 2009.

[13] M. Cristani and N. Gabrielli. Practical issues of description logics for spatial reasoning. volume SS-09-02, pages 5–10, 2009.

[14] M. Cristani, D. Fogoroasi, and C. Tomazzoli. Measuring homophily. volume 1748, 2016.

[15] M. Cristani, C. Tomazzoli, and F. Olivieri. Semantic social network analysis foresees message flows. volume 1, pages 296–303, 2016.

[16] M. Bohanec, M. Kljaji BorÅtnar, and M. Robnik-Åikonja. Explaining machine learning models in sales predictions. *Expert Systems with Applications*, 71:416–428, 2017.

[17] T. Chen, H. Yin, H. Chen, L. Wu, H. Wang, X. Zhou, and X. Li. Tada: Trend alignment with dual-attention multi-task recurrent neural networks for sales prediction. volume 2018-November, pages 49–58, 2018.

[18] D. DovÅan, V. Logar, and I. Åkrjanc. Solving the sales prediction problem with fuzzy evolving methods. 2012.

[19] T. Geva, G. Oestreicher-Singer, N. Efron, and Y. Shimshoni. Using forum and search data for sales prediction of high-involvement projects. *MIS Quarterly: Management Information Systems*, 41(1):65–82, 2017.

[20] W. Huang, Q. Zhang, W. Xu, H. Fu, M. Wang, and X. Liang. A novel trigger model for sales prediction with data mining techniques. *Data Science Journal*, 14, 2015.

[21] N. Khalil Zadeh, M.M. Sepehri, and H. Farvaresh. Intelligent sales prediction for pharmaceutical distribution companies: A data mining based approach. *Mathematical Problems in Engineering*, 2014, 2014.

[22] Y. Yang, R. Fuli, C. Huiyou, and X. Zhijiao. Svr mathematical model and methods for sale prediction. *Journal of Systems Engineering and Electronics*, 18(4):769–773, 2007.

[23] X. Yu, Y. Liu, X. Huang, and A. An. A quality-aware model for sales prediction using reviews. pages 1217–1218, 2010.

[24] I. Åliobaite, J. Bakker, and M. Pechenizkiy. Towards context aware food sales prediction. pages 94–99, 2009.