

Gradient Boosting Application in Forecasting of Performance Indicators Values for Measuring the Efficiency of Promotions in FMCG Retail

Joanna Henzel*, Marek Sikora^{||}

Department of Computer Networks and System
Faculty of Automatic Control, Electronics and Computer Science
Silesian University of Technology
ul. Akademicka 16, 44-100 Gliwice, Poland
Email: *joanna.henzel@polsl.pl, ^{||}marek.sikora@polsl.pl

Abstract—In the paper, a problem of forecasting promotion efficiency is raised. The authors propose a new approach, using the gradient boosting method for this task. Six performance indicators are introduced to capture the promotion effect. For each of them, within predefined groups of products, a model was trained. A description of using these models for forecasting and optimising promotion efficiency is provided. Data preparation and hyperparameters tuning processes are also described. The experiments were performed for three groups of products from a large grocery company.

I. INTRODUCTION

FOOD retailing is an industry that most people have contact with. It provides products which are necessary for everyday life. Mostly, food is bought on an ongoing basis and, because of this, precise planning of logistics, chain supplies and sales is very important. Because of the characteristics of sale of these products, they are often called *fast-moving consumer goods* (FMCG).

On the market, many retailers offering FMCG products are available, therefore it is crucial to remain competitive. One way to do this is to offer products in promotion. The importance of creating promotions in the FMCG sector can be proven by seeing the amount of money that are spent on this purpose – in 2014 it was \$1 trillion every year as it was mentioned in [1]. Therefore, it is necessary to forecast the promotion effect and plan them with equal importance as a regular sale.

In some cases promotions are planned based on *judgmental forecasting* or using simple baseline statistical forecast with a judgmental adjustment [2]. It means that the promotion planning process is often done manually. However, studies have shown that using only these kinds of forecasting methods may bring bias [3]. A better idea may be to use more advanced methods that rely mostly on knowledge that comes from historic data. Very little has been written about using Machine Learning (ML) methods for the problem of promotion optimisation and forecasting promotion effect.

The objective of this paper is to propose a new way of forecasting promotion effect using the gradient boosting

method. Six different indicators are presented in order to capture the efficiency of promotions. The paper describes an advanced data preparation process. Among three groups of products, a model for each indicator was trained, examined and the optimisation of hyperparameters was conducted. The paper also describes how to use the created models in order to perform optimisation of promotions to get better outcome of the forecast. The paper is organised as follows: the next section provides the review of literature and related works, section III describes problem statement and presents proposed indicators. Afterwards, the data preparation process is presented, followed by the experiments explanation. The paper ends with some conclusions and discussion of the results.

II. RELATED WORKS

Sales forecasting plays an important part in planning and managing many commercial enterprises, including those connected with the retail sector.

Traditionally, forecasting was made using statistical methods, for example: exponential smoothing [4], moving average and the Auto-regressive Integrated Moving Average (ARIMA) model. Well known and widely used is SARIMA – seasonal auto-regressive integrated moving average. Some improvements of this method were proposed regarding the problem of sales forecasting in the papers [5] and [6].

Over time, more complex methods were used and evaluated in the field of sales forecasting. In [7] a comparison of various linear and non-linear models for this task was conducted. The best obtained model was the neural network built on deseasonalized time series data. The results suggested that non-linear models should be highly considered when dealing with modelling retail sales. Another neural network algorithm regarding forecasting retail sales which was used for this task was back-propagation neural network (BPNN) [8]. Evolutionary neural networks (ENN) were also considered in [9]. The use of the extreme learning machine (ELM) algorithm was also investigated in this area, for example in the papers [10], [11] and [12]. Also, a successful proposition of adding linguistic knowledge in the forecasting process using linear regression

has been proposed in [13]. In the paper [14] an interesting forecast technique was presented. The authors combined pre-purchase online search data with economic variables to predict monthly car sales.

An important part of retail forecasting is making sales forecasts for short shelf-life food products, which are very often referred to as Fast-Moving Consumer Goods (FMCG). It is an even more complex task, because the additional products, whose sales may be overestimated, cannot be stored for a very long time in the shop. In the paper [15] a radial basis function (RBF) neural network and a designed genetic algorithm were successfully used for forecasting the sales of fresh milk. In the aspect of FMCG, the authors of [16] showed benefits of applying Machine Learning methods in creating demand forecasting models. The use of the Autoregressive Distributed Lag model was presented in the paper [17]. The authors of [18] proposed using the Dynamic Artificial Neural Network for food sales forecasting for one of multiplexes in India. In the paper [19], different classifiers were analysed and a proposition of combining various forecasting models using neural network was presented in order to improve results for forecasting demands of warehouses. Experiments were performed on real sales data of a national dried fruits and nuts company from Turkey.

Decision and regression tree-based methods were also taken into consideration regarding the sales forecasting. A hybrid method of k-means algorithm and C4.5 algorithm (decision tree classifier) was shown in [20]. In the paper [21] a comparison of different Machine Learning Techniques was conducted regarding sales-forecasting of retail stores. The authors concluded that boosting algorithms gave better results than the regular regression ones. For them, the best results were obtained for the GradientBoost algorithm and the XGBoost implementation has been used in order to increase the accuracy.

Forecasting sales during promotions is a very challenging task as it was mentioned in [2]. In this paper authors pointed out that usually the promotional effect was estimated by combining simple statistical forecasting methods and adding judgmental adjustment, which could lead to miscalculations.

The research about effectiveness of promotions has been conducted for a long time, mostly in the marketing research area and it is described in the practitioner literature. This problem was raised in [22] and [23]. The authors of [1] proposed a new formula for the promotion optimisation problem in the FMCG industry. Although these works concerned estimating the effectiveness of promotions, all of them focus on domain knowledge and do not use machine learning techniques for this task.

Multiple models for forecasting the demand during promotion periods were tested in the paper [24]. The use of PCA and pooled regression was presented in the paper [25] in order to predict sales in the presence of promotions. In the case of direct marketing, machine learning methods were compared and tested in the paper [26]. Interesting findings are presented in [27]. The authors showed that simple statistical

methods performed very well for data without promotions. For periods with promotions more advanced methods had to be used. In this paper, regression trees were used for grocery sales forecasting.

To the best of our knowledge, the tree boosting algorithm, especially the extreme gradient boosting (XGBoost) algorithm, has not been used to forecast the effect of promotions and to optimise the promotion itself. XGBoost was introduced in [28]. It is a well known fact that XGBoost is highly effective for a vast range of classification and regression problems. It was, for example, used in the following areas: medicine [29], fault detection [30], finances [31], accident detection [32], and many others.

XGBoost implementation has a wide array of hyper-parameters. In order to obtain the best results, optimisation of those parameters can be performed. The most commonly used methods are random search (RS) and Bayesian Tree Parzen Estimator (TPE). These methods were used in [33] and [34]. Hyper-parameters optimisation was done using Bayesian optimisation, random search, grid search, and manual search in the paper [35].

III. PROBLEM STATEMENT

In different industries, promotions may have various characteristics. For example, in fashion retail it is noticeable that promotions take place mostly in specific periods during the year – at the end of the fashion seasons. The situation is different in grocery retail business. Multiple promotions can be observed at the same time and they are changing very rapidly. Also, alongside the regular promotions, we can distinguish promotions related to holidays and special days (e.g. Christmas, Easter or St. Valentine's Day) and discounts that are caused by upcoming expiration date.

The purpose of the promotions may be not so obvious. They should give a company bigger profit, but it is not equivalent to the willingness to sell as much as possible of a promoted product. Of course, selling is one of the components of a successful promotion but not the only one. For example, a grocery retail company that set up a promotion does not want customers to buy only the promoted product but wants clients to buy also multiple different products alongside that may be in their regular prices.

In order to capture the effectiveness of each promotion, six different indicators are proposed:

- AVERAGE NUMBER OF SOLD UNITS OR KILOGRAMS EACH DAY (shortcut: AVG. AMOUNT) – This indicator shows how many units or kilograms of the promoted product, on average, were sold during the promotion each day.
- AVERAGE NUMBER OF RECEIPTS WITH THE PROMOTED PRODUCT (shortcut: AVG. NB. RECEIPTS) – The indicator explains in how many baskets the promoted product appeared, on average, each day during the promotion. It can be treated as an indicator of how many customers bought the product each day.

- AVERAGE VALUE OF A BASKET CONTAINING THE PROMOTED PRODUCT (shortcut: AVG. BASKET) – This indicator says what an average value of a basket was where the promoted product appeared. Assuming that customers went for shopping with the will to buy the specific product in promotion, the indicator says how much money they spent in total. The higher the indicator, the more products were bought or the more expensive products were chosen.
- AVERAGE VALUE OF A BASKET CONTAINING THE PROMOTED PRODUCT BUT DISREGARDING THE VALUE OF THE PROMOTED PRODUCT (shortcut: AVG. BASKET WITHOUT ITEM) – This indicator is very similar to the previous one. It shows what an average value of a basket was where the promoted product appeared but the value of the promoted product was not taken into account. It means that this indicator is equal to 0 if the customer buys only the promoted product.
- AVERAGE NUMBER OF UNIQUE PRODUCTS IN THE BASKET (shortcut: AVG. NB. UNIQUE ITEMS) – It says how varied the basket is. The higher the value of the indicator, the better – it means that the customer not only bought a specific product but also many others.
- AVERAGE NUMBER OF THE BASKETS (shortcut: AVG. NB. CLIENTS) – The indicator shows how many, on average, transactions were performed each day during the promotion. It does not matter if the customer bought a promoted product or not.

The reason for choosing the following indicators is that the information they carry is of interest to a company operating a large international retail shop chain and with which we collaborated during the research process.

The values of indicators are calculated per promotion. It means that each promotion can be described by the 6 proposed indicators.

These indicators may seem very similar, because the differences between them are very subtle. In order to show their utility, some examples are introduced:

- 1) 100 kg of apples were sold during the promotion. The indicator AVERAGE NUMBER OF SOLD UNITS OR KILOGRAMS EACH DAY tells us about it, but it does not give an information if this amount was bought by one person or by 50 people who bought 2 kg on average. This information will be provided by the AVERAGE NUMBER OF RECEIPTS WITH THE PROMOTED PRODUCT.
- 2) The average value of the basket, with a product that was in promotion, was 50\$. It is the value of the indicator AVERAGE VALUE OF A BASKET CONTAINING PROMOTED PRODUCT. Now we may want to know if the rest of the products were a big part of the basket (e.g. 80 %) or only an addition to the promoted product (e.g. 10 % of the total value). The AVERAGE VALUE OF A BASKET CONTAINING THE PROMOTED PRODUCT BUT DISREGARDING THE VALUE OF THE PROMOTED PRODUCT gives this information. We also might want

to know if the customers, on average, bought 2 unique products, that gave the value of 50 \$, or they bought 25 unique products – the indicator AVERAGE NUMBER OF UNIQUE PRODUCTS IN THE BASKET is proposed in order to capture this.

Each of the proposed indicators are *gain measures*. It means that the higher the value, the better is the promotion. They can be inversely correlated – for example, if the price is very low, clients may buy a lot of the specific product but the diversity of products inside the basket may be very poor.

The proposed indicators describe each promotion very precisely. Knowing the value of each of them, the evaluation of the promotions can be performed. What is even more interesting, is the evaluation of future promotions so it is connected with the promotions planning. By setting up the features of the future promotion, it is possible to determine whether the predicted effect will be satisfying.

The forecasting of the promotion effect can be done for every product separately. Having the history of the promotions and their effects, we can model the characteristics of the promotion for the specific product and it is possible to predict what the effect in the future will be. Unfortunately, a number of past promotions for many products is small, so there are not many examples for training a model. Additionally, a question has been raised how to predict the promotion effect for a new product or an item that has never been in promotion. One solution may be to find similar products that have similar characteristic of sales. The problem is that it is difficult to assure that this will translate to similar characteristics of promotion effect. Another idea would be to create, based on domain knowledge, groups of products that act the same during the promotions. Then a model would be built for each of these groups. This issue, however, is out of scope of our paper.

The problem of forecasting indicators for unknown and rarely promoted products was solved by the authors – the products were grouped by the predefined categories, e.g. vegetables, fruits, dairy products or meat. It is assumed that the products within the group will act similarly during the promotion because they are akin to each other. Therefore, it is expected that the characteristics of the indicators describing the promotion effect will be similar for products within the group.

To summarize: a new approach to the problem of forecasting the promotion effect is to calculate a model for each of the 6 proposed indicators for each predefined category (group) of products.

IV. DATA PREPARATION

In developing models for promotions indicators and in experiments, data from a large grocery retail company were used (more than 500 stores). The data from groups: vegetables, fruits and dairy products were taken into account. Only regular promotions were investigated, therefore the promotions that happened before or during holidays were not included. Additionally, promotions that applied only when:

TABLE I
EXAMPLE OF RECORD DESCRIBING PROMOTION BEFORE PREPARATION

store ID	product	start date	end date	conditional attributes	value of indicator
10	pears	2018-01-22	2018-01-25	...	123.56

- multiple units were bought (type “buy 2 pay for 1”),
- minimum weight condition was met (type “buy minimum 5 kg and get 15 % off”),
- when combination of products was bought

were not taken into consideration. The same goes for products that had reduced prices because of the approaching best-before date. The reason for choosing only regular promotions was that they were the majority of all promotions and we were advised that non-regular promotions have a different characteristic that may bring a bias to the model. Also, in the examined data there were no promotions longer than 7 days. Promotions from the years 2015 to 2018 were used. Data for 2015 and part of 2016 were not completed, so there was a visibly smaller number of promotions at that period.

One record of data described one promotion in one store. Therefore, for example, if there would be a promotion on pears in the store with ID 10 from 2018-01-22 to 2018-01-25, the record, before preparation, would look like in table I.

A. Attributes

In the research, extended numbers of conditional attributes were taken into consideration when preparing data sets. A few main categories of the attributes can be distinguished:

- connected with price,
- connected with the time and duration of the promotion,
- describing the advertisement media (promotion channels),
- describing the store and its surroundings,
- describing the impact of other promotions.

In the first category, only 2 attributes were included: the price of a product and a change of the price.

Time attributes connected with the promotion were:

- number of days of the promotion,
- weekday of the first day of the promotion,
- attributes created based on the date of the first day of the promotion: year number, month number, day number, week number, number of a day in the year, and the season.

Considering information about promotion channels, binary attributes were added. They described if the promotion was advertised on TV, on the radio, on the Internet or in a different way.

Additionally, new variables describing combinations of the promotion channels were added to the data sets. For each combination, new attributes were created as a result of binary operations AND, OR and XOR (only when combination consisted of 2 elements). For example, if the undermentioned statements, were true, then a new variable got value 1, otherwise – 0.

- Promotion was on TV or on the radio. (OR operation)

- Promotion was on TV or on the radio or on the Internet. (OR operation)
- Promotion was on the TV and on the radio. (AND operation)
- Promotion was either on the Internet or on the radio. (XOR operation)

We can assume that promotions in similar stores (for example in small villages or in big cities) can have similar characteristics. For example, the customers in a rich city buy more expensive products in general, therefore the value of the basket is automatically higher than in other stores. The exemplary attributes that were used in order to capture these characteristics were:

- number of inhabitants within 1 km,
- number of inhabitants per 1 square km,
- number of inhabitants within a 5-minute driving range,
- unemployment rate,
- number of cars per 1,000 inhabitants,
- average monthly salary,
- tourism ratio, etc.

The last but not least, attributes connected with the impact of other promotions were added. As it was mentioned in the section III, promotions rarely ever take place one at a time. It is a possible situation, that a client that bought the considered product came to the store because of another promotion. It is impossible to capture clients' intentions fully, but it can be assumed that the more promotions in the shop, the more clients will come. Because of this, the following attributes were added to the data:

- Number of all promotions in a store.
- Number of all promotions that were advertised on TV, radio or internet.
- Number of all promotions that were advertised on TV, radio, internet or in a different way.

B. Matching periods without promotions

In order to capture the characteristics of products in the group, matching records without promotions were found for most of the records in the data set. The matching period had to meet the following conditions:

- It considered the same product as the promotion.
- It considered the same store.
- It had to last as many days as the considered promotion.
- It had to start on the same weekday as the promotion.
- The considered product was not in promotion on any given day.
- The period without promotion could occur maximum 4 weeks and minimum 1 week before the promotion.

The matching period was not found for all promotions because of the lack of meeting the requirements.

The illustration of finding the matching periods was shown in figure 1.

In the final data sets, records connected with periods without promotions were distinguished from promotions by having 0 value in an attribute describing the change of a price.

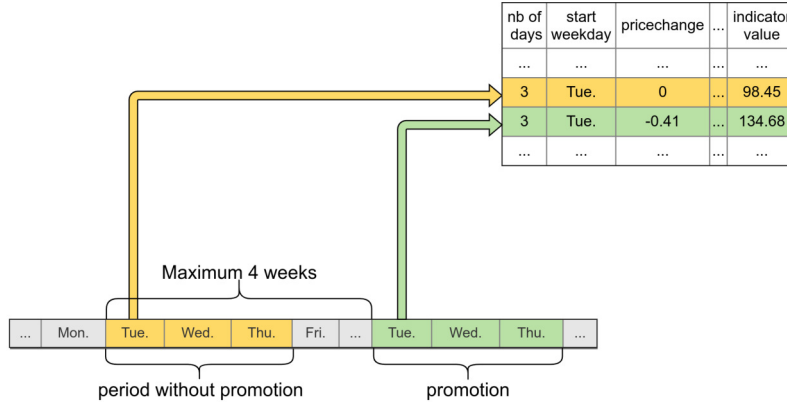


Fig. 1. Finding matching record without promotion

C. Standardisation of the indicators

The standardisation of two proposed indicators was performed. These were:

- AVERAGE NUMBER OF SOLD UNITS OR KILOGRAMS EACH DAY and
- AVERAGE NUMBER OF RECEIPTS WITH THE PROMOTED PRODUCT.

The z-score standardisation was used, but for each product and each store separately. The reason for using standardisation for those indicators was that they were referring to the specific values connected with the sale characteristics of a considered product. For example, it is predictable that during promotions with 20% reduction, apples will be sold more than pomelos, because apples are cheaper and they are bought more often in general. The values of the indicator AVERAGE NUMBER OF SOLD UNITS OR KILOGRAMS EACH DAY will be from a different range for those products. This does not mean, however, that the impact of the 20% reduction does not affect in the same way the increase of sold units of apples and pomelos. In order to capture the general characteristics of products in a group, the standardisation of those indicators was performed.

V. EXPERIMENTS

The experiments of the proposed solution for problem of forecasting the promotion effect were conducted for the following categories of products: fruits, vegetables and dairy products. For each category and each proposed indicator, a forecasting model was constructed. In training data sets, records from 2015-2017 describing promotions and matching periods without promotions were included. In test data sets, records with promotions from 2018 were used. For all indicators within one group of products, conditional attributes in data were the same (described in subsection IV-A). The decision attributes were the values of the considered indicators.

When testing models, cross-validation was not performed. The reason for this is the fact that although the data sets were not typical time-series data, the records could be set in chronological order. Using cross-validation, the testing of

a model might be performed on records preceding the training data.

XGBoost (eXtreme Gradient Boosting) [28] from the R package `xgboost` [36] implementation was used for training forecasting models. This gradient boosting framework was chosen because it is a well-known method, which get very good results when working with table-structured data. For example, among the 29 challenges winning solutions posted on a machine learning competition site named Kaggle in 2015, 17 solutions used XGBoost [28]. The experiments described in this paper were also based on tabular data, therefore using XGBoost was a justified idea. Additionally, the paper [21] showed that this algorithm has given the best results for sales-forecasting of retail stores in their experiments, so it was very likely to give good results also for the problem of forecasting the promotion effect in retail sector. In order to evaluate the models efficiency, the following error measures were used:

- *Mean Absolute Error (MAE)*:

$$MAE = \frac{\sum_{i=1}^n |F_i - A_i|}{n}$$

- *Root Mean Square Error (RMSE)*:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (F_i - A_i)^2}{n}}$$

- *Mean Absolute Percentage Error (MAPE)*:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right|$$

- *Weighted Mean Absolute Percentage Error (WMAPE)*:

$$WMAPE = \frac{\sum_{i=1}^n |A_i - F_i|}{\sum_{i=1}^n A_i}$$

where A_i is the actual value and F_i is the forecast value.

The mean absolute percentage error (MAPE) is very intuitive and easy to interpret, however it is meaningful only when the values are large. If the actual value is close to 0, the value of MAPE is approaching infinity and it gives uninterpreted results. In order to bypass these disadvantages,

TABLE II
RESULTS OF MODELS EFFECTIVENESS USING DEFAULT
HYPERPARAMETERS

category	indicator	MAE	RMSE	MAPE	WMAPE
dairy products	AVG. AMOUNT	12.35	19.49	0.51	0.38
dairy products	AVG. NB. RECEIPTS	5.75	8.15	0.44	0.33
dairy products	AVG. BASKET	14.95	21.53	0.19	0.18
dairy products	AVG. BASKET WITHOUT ITEM	14.41	20.97	0.18	0.19
dairy products	AVG. NB. UNIQUE ITEMS	2.12	2.86	0.14	0.14
dairy products	AVG. NB. CLIENTS	165.67	247.54	0.10	0.10
fruits	AVG. AMOUNT	44.57	84.33	1.18	0.51
fruits	AVG. NB. RECEIPTS	27.92	45.22	0.87	0.39
fruits	AVG. BASKET	18.79	26.64	0.19	0.20
fruits	AVG. BASKET WITHOUT ITEM	17.40	25.09	0.19	0.20
fruits	AVG. NB. UNIQUE ITEMS	2.26	3.17	0.13	0.14
fruits	AVG. NB. CLIENTS	135.50	178.62	0.08	0.08
vegetables	AVG. AMOUNT	24.37	44.89	0.48	0.35
vegetables	AVG. NB. RECEIPTS	21.04	37.49	0.42	0.33
vegetables	AVG. BASKET	18.31	26.29	0.18	0.19
vegetables	AVG. BASKET WITHOUT ITEM	18.10	25.53	0.19	0.20
vegetables	AVG. NB. UNIQUE ITEMS	2.34	3.24	0.13	0.14
vegetables	AVG. NB. CLIENTS	171.61	229.00	0.10	0.10

TABLE III
RESULTS OF MODELS EFFECTIVENESS AFTER HYPERPARAMETERS
OPTIMISATION

category	indicator	MAE	RMSE	MAPE	WMAPE
dairy products	AVG. AMOUNT	12.31	18.71	0.53	0.38
dairy products	AVG. NB. RECEIPTS	5.75	8.08	0.45	0.33
dairy products	AVG. BASKET	13.93	20.14	0.17	0.17
dairy products	AVG. BASKET WITHOUT ITEM	14.26	20.32	0.19	0.18
dairy products	AVG. NB. UNIQUE ITEMS	2.04	2.72	0.14	0.13
dairy products	AVG. NB. CLIENTS	129.75	177.15	0.08	0.08
fruits	AVG. AMOUNT	39.72	74.62	1.11	0.45
fruits	AVG. NB. RECEIPTS	24.87	39.39	0.85	0.35
fruits	AVG. BASKET	15.29	22.44	0.16	0.16
fruits	AVG. BASKET WITHOUT ITEM	14.73	21.78	0.17	0.17
fruits	AVG. NB. UNIQUE ITEMS	1.84	2.60	0.12	0.11
fruits	AVG. NB. CLIENTS	125.15	164.49	0.07	0.07
vegetables	AVG. AMOUNT	22.97	42.42	0.47	0.33
vegetables	AVG. NB. RECEIPTS	19.52	34.78	0.41	0.31
vegetables	AVG. BASKET	14.39	21.56	0.15	0.15
vegetables	AVG. BASKET WITHOUT ITEM	14.63	21.65	0.16	0.16
vegetables	AVG. NB. UNIQUE ITEMS	1.89	2.71	0.11	0.11
vegetables	AVG. NB. CLIENTS	135.57	178.51	0.08	0.08

a similar measure – WMAPE – was used. It is the sum of absolute errors divided by the sum of the actual values and it works well with smaller numbers. It is widely used in the retail sector.

Firstly, the XGBoost method was used with default hyperparameters. The results, obtained for test data sets, are presented in table II. For two indicators that were standardised (see subsection IV-C), error measures were calculated after changing forecasted, standardised values to the real values.

A. Optimisation

The optimisation of hyperparameters was performed for each created model. A *grid search* method was used. Six hyperparameters were optimised:

- *nrounds* – maximum number of boosting iterations; range: $[1, \infty)$.
- *base_score* – the initial prediction score of all instances; range: $(-\infty, \infty)$.
- *eta* – boosting learning rate; range: $[0, 1]$.
- *gamma* – minimum loss reduction required to make a further partition on a leaf node of the tree; range: $[0, \infty)$.
- *max_depth* – maximum depth of a tree; range: $[1, \infty)$.
- *subsample* – subsample ratio of the training instance; range: $(0, 1]$.

TABLE IV
VALUES OF HYPERPARAMETERS USED IN OPTIMISATION PROCESS

hyperparameter	tested values
<i>nrounds</i>	1, 21, 41, 61, 81, 101, 121, 141, 161, 181, 201
<i>base_score</i>	Depending on indicator values. Calculated as 11 quantiles from indicator values with the following probabilities: 0.0, 0.1, 0.2, ..., 0.9, 1.0.
<i>eta</i>	0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0
<i>gamma</i>	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
<i>max_depth</i>	1, 4, 7, 10, 13
<i>subsample</i>	0.0001, 0.1001, 0.2001, ..., 0.9001

TABLE V
RMSE IMPROVEMENT AFTER HYPERPARAMETER OPTIMISATION.
 $RMSE_{diff}$ IS A DIFFERENCE OF RMSE BEFORE OPTIMISATION
(TABLE II) AND RMSE AFTER OPTIMISATION (TABLE III).

indicator	$RMSE_{diff}$		
	dairy products	fruits	vegetables
AVG. AMOUNT	0.78	9.72	2.47
AVG. NB. RECEIPTS	0.07	5.83	2.71
AVG. BASKET	1.38	4.19	4.74
AVG. BASKET WITHOUT ITEM	0.65	3.31	3.88
AVG. NB. UNIQUE ITEMS	0.13	0.56	0.53
AVG. NB. CLIENTS	70.39	14.13	50.49

A detailed description of the above parameters can be found in [36].

In the beginning, all possible sequences in which the hyperparameters could be optimised were determined. Six parameters were used, so 720 permutations were obtained. For example, the first permutation was *eta*, *base_score*, *gamma*, *max_depth*, *nrounds*, *subsample* – it means that at first the *eta* hyperparameter was optimised, then *base_score*, afterwards *gamma* and so on. In each permutation, each hyperparameter was changed several times in order to find the best value. The table IV shows values that were used in this process. After iterating through each hyperparameter, the best set of the hyperparameters values of the specific permutation was obtained. Having results for 720 permutations, the best among them was chosen. After this step, the best order of optimising the parameters and the best values for them were determined. In the end, the neighbourhood of the examined hyperparameters values were searched. It was performed in the order determined in the previous step (the order of the best permutation). The optimisation was performed using the validation set that was extracted from the training data set. The flowchart of the described optimisation process is shown in figure 2. The RMSE measure was used as the optimisation criterion.

The results of models efficiency, calculated for the test data sets after hyperparameters optimisation, were shown in table III. It can be observed that for most of the models metrics, the optimisation has given better results than for default models. The details can be seen by comparing table II and table III. The optimisation was carried out based on the RMSE measure. The improvement of this metric was observed for every examined case. The table V shows the exact results.

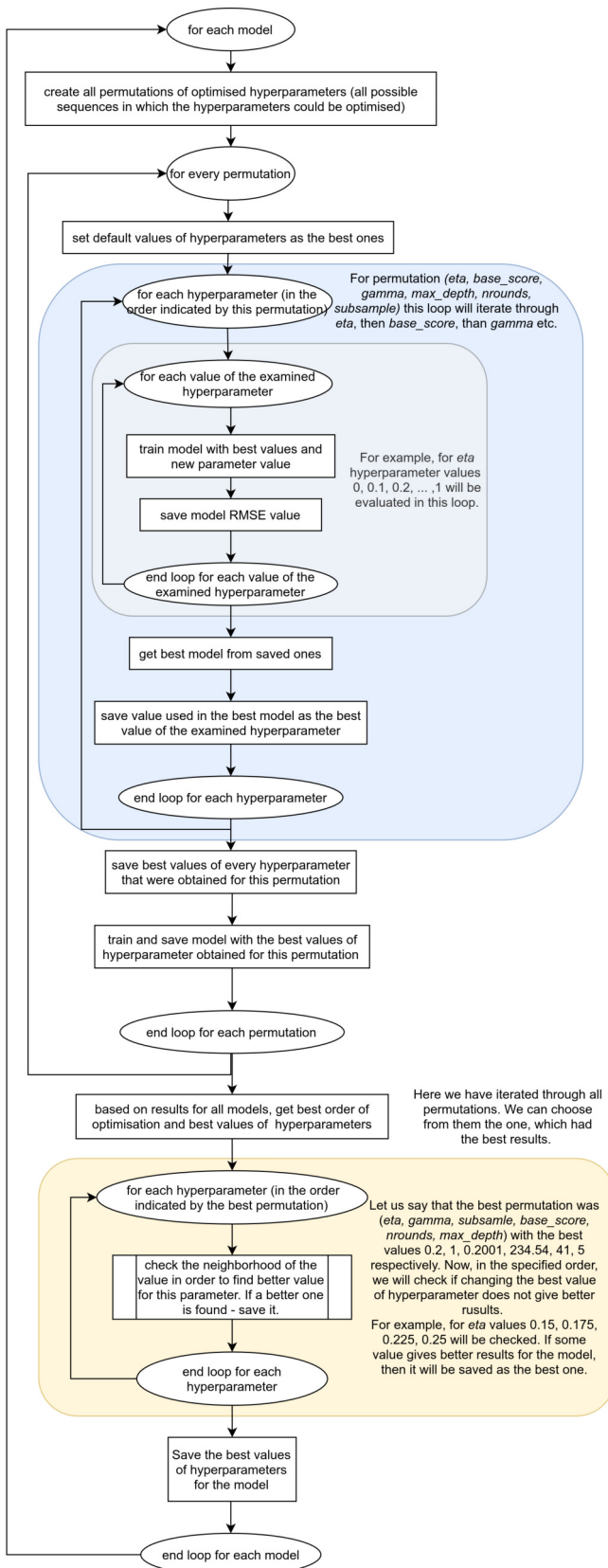


Fig. 2. Flowchart of the hyperparameter optimisation process.

VI. CONCLUSION AND DISCUSSION

Promotions play an important role in the retail sector. When performed suitably, they can give a company bigger profit and bring in more clients to the store.

This study has attempted to introduce a new method of planning and forecasting future promotions using the XGBoost algorithm. Six unique indicators that measure the promotion efficiency were proposed in this paper. These indicators not only describe the sale of a specific product, but characterise the promotions in a much more profound way. Being able to forecast the value of each of them, promotions can be better planned. Indicators forecasts give information if the future promotion, with the given characteristics, like change of price or the weekday when it should start, is likely to be performed satisfactorily. If not, better attributes can be chosen.

In the paper the authors described the data sets preparation process with the use of extended and precisely chosen attributes that could be not so obvious to use. The authors also proposed a solution for forecasting the promotion effect for new, unknown products or products with a small number of past promotions. The models were developed for groups of products and not for each product separately. The experiments were performed for 3 groups: vegetables, dairy products and fruits. A model using XGBoost was developed for each indicator and each group of products. Additionally, the hyperparameters optimisation was performed in order to obtain better models accuracy. It is worth emphasizing that such optimisation can be carried out for any error measure.

The created models provide also a description of the features importance. Figure 3 shows a plot of 10 most important attributes of the model trained for indicator AVG. AMOUNT and dairy products. It can be observed that the change of a price and the price itself are the most important features that influence the amount of sold units during the promotions for this model. In the process of planning promotions, when the results of forecast are not satisfactory, one can tune, starting from these 2 attributes, the promotions characteristics in order to get better results. After making changes in the planned promotions, the predictions can be performed again. If the results are still not satisfying, the previous steps can be repeated. This way the process of optimising future promotions can be performed.

Five most important features for each indicator are presented below. The order, in which the attributes are listed below, was obtained by calculating average importance score of each feature taking into account the results of each group of products:

- AVG. AMOUNT: change of a price; day number (in the year); price; number of all promotions that are happening in the store and are advertised on TV, radio or Internet; day number (in the month).
- AVG. NB. RECEIPTS: change of a price; number of competitors; number of inhabitants within a 10-minute driving range; number of inhabitants within 1 km; number of inhabitants within 500 m.

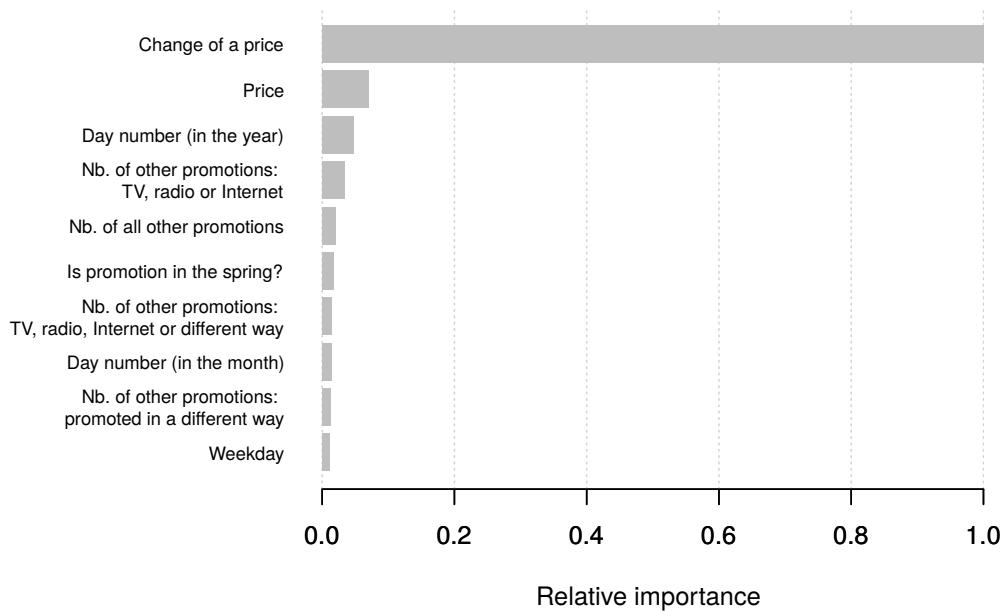


Fig. 3. Plot of feature importance for the model of the indicator AVG. AMOUNT for dairy products. The most important features are shown and the important values are represented as relative to the highest ranked feature.

- AVG. BASKET: price; number of inhabitants within 500 m; change of a price; day number (in the year); weekday.
- AVG. BASKET WITHOUT ITEM: price; number of inhabitants within 500 m; change of a price; day number (in the year); weekday.
- AVG. NB. UNIQUE ITEMS: number of inhabitants within 500 m; price; change of a price; weekday; distance from a competitor.
- AVG. NB. CLIENTS: number of inhabitants within 500 m; number of inhabitants within 1 km; number of inhabitants within a 5-minute driving range; purchasing rate; tourism ratio.

As it can be observed, not all features are possible to change in the process of the promotions planning. However, the ranking may suggest the order in which attribute values should be tuned to get better forecasting results. The most important features for AVG. NB. CLIENTS are not connected with promotions, so the conclusion can be drawn that this indicator is little affected by them.

Summarising the practical aspect of the research: using the presented methodology it is possible to train models for forecasting promotion efficiency. At the input of the models, the features of the future promotion are placed, including change of a price, promotion channels, store attributes and a number of days of the promotion. At the output of the models, the values of the indicators are obtained. They give information on whether the promotion will be successful.

The challenge for future research will be to investigate the efficiency of multi-target prediction methods for the problem of forecasting all six proposed indicators.

Also, there are a few possible additional applications that could benefit from a proposed method. Firstly, the models could be created not for predefined groups of products but for the products that have similar characteristic of a regular sale. For example, products that are bought in general much more often on Saturdays than during different weekdays could be in one group, the products bought steadily through all year long could be in a second group and another group would be products very popular during summer. It is possible that the character of a group might be not so obvious to define for a human. The clustering algorithms for time-series of historical sales could help in finding new groups of products. Another idea is to make models for each special kind of promotions, for example for promotions where multiple units of a product had to be bought. Some modifications of a presented method would need to be proposed because the definition of a matching period without promotion would need to be changed. Lastly, the forecast of an indicator could be obtained as a combination of forecasts from multiple models created for different groups in which this product belongs. The models would be trained in this same manner as described in this paper, only the definition of a group would change. These topics, however, require a great deal of further research.

In conclusion, this paper has shown a new way of planning

and forecasting promotions using Machine Learning techniques. This, to our knowledge, is the first study to examine the utility of the Gradient Boosting method in the problem of forecasting the future promotion effect.

ACKNOWLEDGMENT

This work was partially supported by the European Union through the European Social Fund (grant POWR.03.05.00-00-Z305). The work was carried out in part within the project co-financed by European Funds entitled "Decision Support and Knowledge Management System for the Retail Trade Industry (SensAI)" (POIR.01.01.01-00-0871/17-00).

REFERENCES

- [1] M. C. Cohen, N. H. Z. Leung, K. Panchamgam, G. Perakis, and A. Smith, "The impact of linear optimization on promotion planning," *Operations Research*, vol. 65, no. 2, pp. 446–468, 2017. doi: 10.1287/opr.2016.1573
- [2] R. Fildes, P. Goodwin, and D. Önkal, "Use and misuse of information in supply chain forecasting of promotion effects," *International Journal of Forecasting*, vol. 35, no. 1, pp. 144–156, jan 2019. doi: 10.1016/j.ijforecast.2017.12.006
- [3] S. Makridakis, "The art and science of forecasting An assessment and future directions," *International Journal of Forecasting*, vol. 2, no. 1, pp. 15–39, 1986. doi: 10.1016/0169-2070(86)90028-2
- [4] E. S. Gardner Jr., "Exponential smoothing: The state of the art," vol. 4, no. October 1983, pp. 1–28, 1985.
- [5] T.-M. Choi, Y. Yu, and K.-F. Au, "A hybrid SARIMA wavelet transform method for sales forecasting," *Decision Support Systems*, vol. 51, no. 1, pp. 130–140, apr 2011. doi: 10.1016/j.dss.2010.12.002
- [6] N. S. Arunraj and D. Ahrens, "A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting," *International Journal of Production Economics*, vol. 170, pp. 321–335, dec 2015. doi: 10.1016/j.ijpe.2015.09.039
- [7] C. W. Chu and G. P. Zhang, "A comparative study of linear and nonlinear models for aggregate retail sales forecasting," *International Journal of Production Economics*, vol. 86, no. 3, pp. 217–231, dec 2003. doi: 10.1016/S0925-5273(03)00068-9
- [8] C. Y. Chen, W. I. Lee, H. M. Kuo, C. W. Chen, and K. H. Chen, "The study of a forecasting sales model for fresh food," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7696–7702, dec 2010. doi: 10.1016/j.eswa.2010.04.072
- [9] K.-F. Au, T.-M. Choi, and Y. Yu, "Fashion retail forecasting by evolutionary neural networks," *International Journal of Production Economics*, vol. 114, no. 2, pp. 615 – 630, 2008. doi: 10.1016/j.ijpe.2007.06.013
- [10] Z.-L. Sun, T.-M. Choi, K.-F. Au, and Y. Yu, "Sales forecasting using extreme learning machine with applications in fashion retailing," *Decision Support Systems*, vol. 46, no. 1, pp. 411–419, 2008. doi: 10.1016/j.dss.2008.07.009
- [11] M. Xia, Y. Zhang, L. Weng, and X. Ye, "Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs," *Knowledge-Based Systems*, vol. 36, pp. 253–259, dec 2012. doi: 10.1016/j.knsys.2012.07.002
- [12] Y. Yu, T.-M. Choi, and C.-L. Hui, "An intelligent fast sales forecasting model for fashion products," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7373–7379, jun 2011. doi: 10.1016/j.eswa.2010.12.089
- [13] K. Kaczmarek and O. Hryniewicz, "Linguistic knowledge about temporal data in bayesian linear regression model to support forecasting of time series," in *2013 Federated Conference on Computer Science and Information Systems, FedCSIS 2013*, 2013. ISBN 9781467344715 pp. 651–654.
- [14] P. Wachter, T. Widmer, and A. Klein, "Predicting automotive sales using pre-purchase online search data," in *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems*, 2019. doi: 10.15439/2019F239 pp. 569–577.
- [15] P. Doganis, A. Alexandridis, P. Patrinos, and H. Sarimveis, "Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing," *Journal of Food Engineering*, vol. 75, no. 2, pp. 196–204, jul 2006. doi: 10.1016/j.jfoodeng.2005.03.056
- [16] E. Tarallo, G. K. Akabane, C. I. Shimabukuro, J. Mello, and D. Amancio, "Machine learning in predicting demand for fast-moving consumer goods: An exploratory research," *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 737–742, 2019. doi: 10.1016/j.ifacol.2019.11.203
- [17] T. Huang, R. Fildes, and D. Soopramanien, "The value of competitive information in forecasting FMCG retail product sales and the variable selection problem," *European Journal of Operational Research*, vol. 237, no. 2, pp. 738–748, sep 2014. doi: 10.1016/j.ejor.2014.02.022
- [18] V. Adithya Ganesan, S. Divi, N. B. Moudhgalya, U. Sriharsha, and V. Vijayaraghavan, "Forecasting food sales in a multiplex using dynamic artificial neural networks," in *Advances in Intelligent Systems and Computing*, vol. 944. Springer Verlag, 2020. doi: 10.1007/978-3-030-17798-0_8. ISBN 9783030177973. ISSN 21945365 pp. 69–80.
- [19] I. Islek and S. Gunduz Oguducu, "A decision support system for demand forecasting based on classifier ensemble," in *Communication Papers of the 2017 Federated Conference on Computer Science and Information Systems*, 2017. doi: 10.15439/2017F224 pp. 35–41.
- [20] S. Thomassey and A. Fiordaliso, "A hybrid sales forecasting system based on clustering and decision trees," *Decision Support Systems*, vol. 42, no. 1, pp. 408–421, oct 2006. doi: 10.1016/j.dss.2005.01.008
- [21] A. Krishna, V. Akhilesh, A. Aich, and C. Hegde, "Sales-forecasting of retail stores using machine learning techniques," in *Sales-forecasting of Retail Stores using Machine Learning Techniques*. IEEE, 2018. doi: 10.1109/CSITSS.2018.8768765. ISBN 9781538660782 pp. 160–166.
- [22] R. C. Blattberg and A. Levin, "Modelling the effectiveness and profitability of trade promotions," *Marketing Science*, vol. 6, no. 2, pp. 124–146, 1987. doi: 10.1287/mksc.6.2.124
- [23] J. Zhang and M. Wedel, "The effectiveness of customized promotions in online and offline stores," *Journal of Marketing Research*, vol. 46, no. 2, pp. 190–206, 2009. doi: 10.1509/jmkr.46.2.190
- [24] K. H. Van Donselaar, J. Peters, A. De Jong, and R. Broekmeulen, "Analysis and forecasting of demand during promotions for perishable items," *International Journal of Production Economics*, vol. 172, pp. 65–75, feb 2016. doi: 10.1016/j.ijpe.2015.10.022
- [25] J. R. Trapero, N. Kourentzes, and R. Fildes, "On the identification of sales forecasting models in the presence of promotions," *Journal of the Operational Research Society*, vol. 66, no. 2, pp. 299–307, 2015. doi: 10.1057/jors.2013.174
- [26] G. Cui, M. L. Wong, and H. K. Lui, "Machine learning for direct marketing response models: Bayesian networks with evolutionary programming," *Management Science*, vol. 52, no. 4, pp. 597–612, 2006. doi: 10.1287/mnsc.1060.0514
- [27] Ö. G. Ali, S. Sayin, T. van Woensel, and J. Fransoo, "SKU demand forecasting in the presence of promotions," *Expert Systems with Applications*, vol. 36, no. 10, pp. 12340–12348, dec 2009. doi: 10.1016/j.eswa.2009.04.052
- [28] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. KDD '16. ACM, 2016. doi: 10.1145/2939672.2939785. ISBN 9781450342322 pp. 785–794. [Online]. Available: <http://doi.acm.org/10.1145/2939672.2939785>
- [29] L. Torlay, M. Perrone-Bertolotti, E. Thomas, and M. Baciú, "Machine learning–XGBoost analysis of language networks to classify patients with epilepsy," *Brain Informatics*, vol. 4, no. 3, pp. 159–169, sep 2017. doi: 10.1007/s40708-017-0065-7
- [30] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *IEEE Access*, vol. 6, pp. 21020–21031, mar 2018. doi: 10.1109/ACCESS.2018.2818678
- [31] J. Nobre and R. F. Neves, "Combining principal component analysis, discrete wavelet transform and XGBoost to trade in the financial markets," *Expert Systems with Applications*, vol. 125, pp. 181–194, jul 2019. doi: 10.1016/j.eswa.2019.01.083
- [32] A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, and A. K. Mohammadian, "Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis," *Accident Analysis and Prevention*, vol. 136, p. 105405, mar 2020. doi: 10.1016/j.aap.2019.105405

- [33] Y. Wang and X. S. Ni, "A XGBoost risk model via feature selection and Bayesian hyper-parameter optimization," *International Journal of Database Management Systems*, vol. 11, no. 1, pp. 1–17, jan 2019. [Online]. Available: <http://arxiv.org/abs/1901.08433>
- [34] M. Nishio, M. Nishizawa, O. Sugiyama, R. Kojima, M. Yakami, T. Kuroda, and K. Togashi, "Computer-aided diagnosis of lung nodule using gradient tree boosting and Bayesian optimization," *PLoS ONE*, vol. 13, no. 4, apr 2018. doi: 10.1371/journal.pone.0195875
- [35] Y. Xia, C. Liu, Y. Y. Li, and N. Liu, "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring," *Expert Systems with Applications*, vol. 78, pp. 225–241, jul 2017. doi: 10.1016/j.eswa.2017.02.017
- [36] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano, T. Zhou, M. Li, J. Xie, M. Lin, Y. Geng, and Y. Li, *xgboost: Extreme Gradient Boosting*, 2019, r package version 0.90.0.2. [Online]. Available: <https://CRAN.R-project.org/package=xgboost>