Predicting the Optimal Date and Time to Send Personalized Marketing Messages to Repeat Buyers

Alexandros Deligiannis[1], Charalampos Argyriou[2], Dimitrios Kourtesis[3]

Research & Development Department

Apifon S.A.

Greece

Abstract—Most of today's digital marketing campaigns which are sent through email and mobile messaging are bulk campaigns which deliver the same message at the same time to all customers, regardless of their needs and preferences. The outcomes are bad customer experience, low engagement and low conversion rates. Modern marketing automation tools aim to facilitate personalized communications, such as scheduling of individual marketing messages based on each individual subscriber's profile. This research focuses on the problem of automatically deciding on the optimal date and time for sending consent-based personalized marketing messages. We specifically focus on the case of repeat consumers of consumer packaged goods (CPG) which require regular replacement or replenishment. The objective is to timely anticipate the needs of consumers in order to increase their level of engagement as well as the rate at which they repurchase products. The proposed solution is based on a regression model trained with transactional data and instant messaging metadata. We describe the way such a model can be created and deployed to a scalable high-performance environment and provide pilot evaluation results that suggest a significant improvement in marketing effectiveness.

Keywords—*Personalized marketing automation; customer relationship management; conversion rate optimization; customer engagement; machine learning; XGBoost regression; cloud computing; data privacy*

I. Introduction

The objective of marketing communications is to reach and engage customers through effective marketing campaigns. Effectiveness is predicated on several different properties of a marketing communication activity. Two highly critical properties are how relevant the marketing message is to the needs and interests of each customer and how well it is timed [1]. The ability of an organization to deliver engaging marketing communication experiences that achieve high conversion rates requires solutions for optimizing both content and timing.

Most of today's digital marketing campaigns sent through email and mobile messaging are bulk campaigns which deliver the same message at the same time to all customers, regardless of their needs and preferences. The typical consequences are bad customer experience, low engagement and low conversion rates [2]. Marketing professionals attempt to improve outcomes through audience segmentation, e.g. by dividing recipients into smaller groups. In this way, they can differentiate the type of message and time of delivery for different groups [3]. In an ideal setting, these customer segments would be defined based on customers' past purchase history and engagement behaviour relative to similar past campaigns. However, such data segmentation is rarely available or practical to analyse.

Modern organizations seek to adapt the message and timing of marketing communications to the needs and preferences of individual customers (i.e. marketing to segments of one) in order to achieve the highest marketing effectiveness [4]. Although personalization at scale remains a hard problem for marketing communication professionals, such marketing automation technology will allow organizations to leverage unique customer profiles derived from past transaction and engagement data to deliver the right message at the optimal date and time to each individual customer [5].

A growing number of marketing technology providers are investing research and development resources to address the challenge of personalized marketing at scale [6]. Taking into account recent research advances in machine learning and predictive analytics [7], this paper presents some of the research results obtained from project PRIME - an R&D effort by business messaging technology provider Apifon [8].

PRIME focuses on developing new services for consent-based personalization of business-to-consumer mobile communications, utilizing data from past purchase transactions and past exchanges of instant messages between the business and the customer. In this paper, we focus on the part of this research that relates to optimizing the time that marketing messages are scheduled for delivery to repeat buyers of consumer packaged goods (CPG) via instant messaging. Specifically, we look at the case of products that require frequent replenishment - such as food & beverages, cosmetics or household products.

Our proposed solution utilizes predictive models which are trained with data from previous purchase transactions as well as past message exchanges. The objective is to improve the timing of messages relative to subscribers' predicted needs and to increase the rate of follow-on purchases.

Section II outlines some of the published research relating to this work, while Section III presents the scope and focus of our research. Section IV provides a walk-through of the required data processing approach and Section V discusses the date and time optimization model in detail, from how predictive models can be created based on a scalable machine learning approach to how they can be deployed to a cloud infrastructure. Section VI presents the evaluation method and Section VII discusses the results of a real world pilot case study. The results suggest a measurable improvement in marketing effectiveness as measured by customer engagement and conversion rates (product repurchases).

II. Related Work

A. Time-Aware Recommender Systems

Much of the published research results in the field of personalized recommender systems integrate complementary contextual information aiming to improve the efficacy of recommendations, notably the timing and frequency of interactions between consumers and content or products. Ding and Li observed that Collaborative Filtering methods are not sensitive to changes in user purchase interests over time [9]. They suggested that any product or item which was recently rated by a user should have a bigger impact on the prediction of the particular user's future behaviour compared to an item that was rated by the same user a long time ago. They proposed an algorithm to compute time weights for different products/items rated by a user, in a manner that assigns a decreasing weight to old data, thereby introducing a personalized interest decay factor.

Koren also observed that customer inclinations and tastes are evolving and that capturing time drifting patterns in user behavior is essential to improving the accuracy of recommenders [10]. He however argued that classical time-window or instance decay approaches are limited, as they lose too much signal when discarding data instances. He proposed a more sensitive model which can better distinguish between transient effects and long term patterns in customer interests and evaluated the model on a large movie rating dataset by Netflix, showing increased prediction accuracy.

Baltrunas and Amatriain introduced a time-aware recommender system aiming to accurately predict a user's music tastes given the time of day, week or year [11]. Their approach assumes that user preferences change with time but have temporal repetition. The main idea of the approach is to partition a user profile into smaller, time-dependent profiles and use these micro-profiles for the prediction task, instead of a single profile.

B. Repeat Purchase Prediction

Wang and Zhang worked on augmenting recommender systems to promote repetitive purchases [12]. They adapted the proportional hazards modeling approach in survival analysis to the recommendation research field and proposed a new opportunity model to explicitly incorporate time in an e-commerce recommender system. The model can estimate the joint probability of a user making a follow-up purchase of a particular product at a particular time.

A common obstacle in predicting repetitive purchase events is the "noise" introduced by customers' irregular purchases. Dey et al. addressed the problem by building a Poisson-Gamma model to estimate the average purchase frequency, along with a Dirichlet model for estimating the purchase probability of a product category [13]. The approach builds on the observation that regular purchases in e-commerce websites are a reliable indicator of customer satisfaction and loyalty. Their model was shown to predict a user's average repeat purchase frequency and dominant product category with satisfactory accuracy.

Many researchers formulate the problem of predicting when a customer will make the next purchase as a typical classification task. The procedure of training a predictive model for this kind of task does not significantly differ from other classification tasks. Feature engineering, however, is likely to be more complex for prediction tasks in e-commerce compared to other classification problems. This is because a large number of attributes is needed in order to capture customer preferences, behaviors and interactions. Liu et al. have shown that ensemble techniques which blend multiple classifiers together result in superior performance in repeat buyer prediction [14].

Chamberlain et al. presented how ASOS.com, a global online fashion retailer, distinguishes between loyal and non-loyal customers to predict Customer Lifetime Value (CLTV) and enable personalized shopping experiences [15]. The predictive models used by ASOS are trained with an unsupervised learning approach on user session data (i.e. sequences of products viewed by each customer) and afford high accuracy in predicting repetitive orders.

C. Demand Forecasting

Being able to accurately predict product purchases and repurchases enables enterprises to produce accurate revenue forecasts, manage their stock levels, optimize their pricing and adapt their operations. This is the realm of demand forecasting, which is another stream of research that can inform the design of solutions like the one presented in this paper.

Traditional forecasting methods are based on time series analysis. They are therefore applied under the hypothesis that past demand can provide a statistical estimation of future demand. This works well for simple environments but fails to work in complex domains where demand is affected by multiple parameters. Other kinds of forecasting methods such as causal modeling can tackle complex domains beyond the limits of time series models. Machine learning methods could actually be viewed as approaches to causal modeling since they can incorporate time series, categorical variables, fuzzy variables, text analysis, etc. [16].

Ferreira et al. presented the way that an online retailer, Rue La La, is using transactional data and machine learning to enable demand forecasting and optimize pricing decisions on a daily basis [17]. The researchers used machine learning techniques on top of historically lost sales to predict the future demand of new products, and developed an algorithm to dynamically suggest optimized prices.

D. Email Campaign Optimization

One last stream of research related to the work presented in this paper is marketing campaign optimization - and more specifically, optimization of email campaigns. Several studies over the past decade have addressed the question of computing the optimal time to send out an email campaign [18, 19]. For instance, Paralič et al. have recently proposed a method for email campaign planning that is optimized to increase open rates, taking into account the marketing campaign content and the type of customer [20]. The researchers applied a combination of classification models which group together similar customers and specify the optimal time to send the campaign message to each group, resulting in significantly increased open rates (i.e. percentage of customers opening the messages).

III. Research Scope and Goals

A. Project PRIME

The research results presented in this paper are part of an ongoing R&D project carried out at Apifon, a European company providing marketing communication technology and telecommunication services to the global market (www.apifon.com).

The project is titled PRIME (Predictive Personalization of Conversational Customer Communications with Data Protection by Design) and focuses on the development of a new services platform which enables personalization of business-to-consumer mobile communications through the use of predictive analytics and machine learning technologies. The motivation is to make communication: (a) more personalized and relevant to each customer's preferences; (b) more direct, interactive and content-rich; and (c) safer for both sides, by protecting consumers' personal data and ensuring GDPR compliance for businesses.

B. PRIME Platform Capabilities

The capabilities offered by the PRIME platform are the following:

- Date and Time Optimization: The goal of any business that promotes its products through direct marketing campaigns is to find the best time to send out the campaign content to its subscribers, so as to achieve the highest possible conversion rate. The same is true for the special case of messages which are not part of a mass campaign but are addressed to individual consumers as reminders for them to reorder a consumable or disposable product that they had purchased in the past. The PRIME platform offers services to automatically determine the optimal date range, day of the week and time of day to send out a specific marketing message to each customer individually, based on his/her unique profile.

- Segment Recommendation: A great challenge faced by many marketing professionals during their daily work has to do with how to choose the recipient list for a particular campaign message. Marketers often lack the tools to be able to specify a highly relevant target audience for a campaign. Messages are often sent out to large lists of non-relevant customers, resulting in low engagement rates and poor customer experience. Through automated segmentation, the PRIME platform enables marketers to reach those subscribers who are most likely to find the content of a specific campaign relevant and compelling enough to engage.

- Keyword Recommendation: The topic of a marketing communication activity (concepts, terms) and the written language used (style, expression) are essential for personalization. The PRIME platform offers personalized content enrichment for the message of an upcoming campaign through automated keyword recommendations.

- Click-Through Rate Estimation: One of the challenges that marketers face in their everyday work is how to estimate the ROI (Return on Investment) of a marketing communication activity at the planning stage. Marketers often have little information to help them estimate how effective a campaign can be before executing it. The PRIME platform aims to produce a reliable estimation of a messaging campaign's Click-Through Rate before it is actually sent [8]. This is tied to predicting who of the recipients will successfully receive the message and will engage with the content (e.g., by opening a link in the message and following a call-to-action).

- Risk Factor Estimation: Another issue marketers face is knowing when a customer is highly likely to cease purchasing. Estimating the risk of losing a customer is about computing a reliable indicator of the client's risk of leaving at any given time. The PRIME platform can dynamically organize customers into groups, based on their risk factor. This enables marketers to take preventive measures, like sending out special offers.

C. Focus and Contributions of this Research

The solution approach and evaluation results that we discuss in this paper relate to the first category of the platform capabilities which we outlined above: Date and Time Optimization.

We present a solution to the problem of scheduling the delivery of consent-based personalized marketing messages to individuals who are existing customers of a retail enterprise. Specifically, we focus on the case of marketing communications which are addressed to repeat buyers of consumer packaged goods (CPG) products which require routine replacement or replenishment, such as food & beverages, cosmetics and household consumables and disposables. The objective is to improve the timing of messages relative to the subscribers' predicted needs, leading to an improved customer experience and a higher rate of follow-on purchases.

The problem is tackled by training a regression model that is able to accurately predict the number of days between the last product purchase made by a particular customer and their next purchase. Two types of data are being used to train the model: (i) historical data describing purchase transaction details (e.g. purchased products, transaction dates and value), and (ii) historical data describing the customer's engagement with mobile message communications received from the retail enterprise in the past (message delivery dates, message content, delivery status, open/click actions, engagement frequency and dates, etc.).

Combining these two types of data allows us to build a model that can accurately determine the most likely date that a specific product would need to be repurchased by an individual customer. The date predicted by the model is subsequently processed by a domain-specific personalization algorithm which fine tunes the timing to the optimal day of the week and optimal time of the day to actually deliver the message. This finally allows the PRIME platform to automatically schedule the delivery of the repurchase reminder messages for each consumer.

We evaluated this solution using online (live) testing with actual customers of a European retail brand in the market of

baby products and collected data which suggests a significant improvement in marketing effectiveness, compared to the standard method of scheduling product repurchase reminders - which uses fixed time intervals determined by the last date of a product's purchase. As discussed later in the paper, our personalized scheduling approach resulted in a significantly higher rate of both customer interaction and conversion from reminder message to repurchase transaction.

IV. Data Processing

A. Collection of Message Exchange and Engagement Data

Messaging service providers, like Apifon, function as intermediaries in the communication between different companies and their customers. As such, they are in position to collect and analyze the data generated during this messaging exchange. This may include events regarding if and when a message was delivered to the customer's mobile device, if, when and how many times a message was opened, and whether or not the customer followed hyperlinks included in the message.

Provided that the consumer has given consent, the messaging services provider is in position to collect a host of historical data describing the consumer's engagement with mobile message communications received from the retail enterprise, throughout the lifetime of the customer relationship.

B. Integration of External Transactional Data

Message exchange and engagement data are one of the two types of data being utilized in our approach. The second type of data is purchase transaction details. This data needs to be sourced from the information systems used by the retail enterprise.

For this purpose, we developed a secure data exchange infrastructure to facilitate data collection. We used a message broker on top of the RabbitMQ open-source software [21], utilizing the Advanced Message Queuing Protocol (AMQP). This protocol provides a platform for both sending and receiving messages which remain secure until they reach their destination [22]. Table I provides an example of a potential purchase transaction entry.

C. Compliance with Data Privacy Policies

Compliance with the General Data Protection Regulation (GDPR) is a fundamental requirement in the design of the PRIME platform. Any type of data processing taking place in the platform is only possible with the data subject's consent. The way that this consent is being obtained depends on the data privacy and marketing communication policies of the retail enterprise that maintains the direct relationship with the customer (i.e. the data owner).

Inside the platform, there is a series of data management services that are built to support GDPR-specific requirements, such as ensuring the data subject's right to access and right to be forgotten. These terms describe the right of the customer to receive a copy of all the personal information stored by the platform, as well as to remove this information permanently.

Additional measures that ensure GDPR compliance of the solution presented in this paper include data encryption

(for instance, hashed phone numbers) and data segregation. Segregation is an architecture design property which ensures that no single data processing node can have read access to the full set of data relating to any individual. Customer data which includes personally identifying information are kept separate from the computed user profile data generated by the predictive model and are maintained only for the time period required for the necessary calculations [23, 24].

D. Data Parsing and Feature Engineering

Once the data are imported into the PRIME platform they are taken through a pre-processing pipeline to extract and transform primary features into a format suitable for the predictive models, but also to derive new, secondary features based on a set of business domain rules.

The features used by the predictive model presented in this paper are shown in Table II. These features are acquired either by performing simple calculations on the incoming data, or by applying more complex data transformations. The "Value Cluster" and "Gender" features are an exception because they require two additional computation processes. The values for the Value Cluster feature are derived with a "K-Means" unsupervised learning algorithm based on the "Sum Order Value" feature of each customer (Table II). The "K-Means" algorithm determines the optimal number of clusters to be used using a Silhouette score [25]. The Gender feature values are derived with a special classification model which has been previously trained with a dataset of male and female names labelled by gender.

V. Date and Time Optimization

A. Classification of Regular / Repeat Customer

The first step in the execution of our optimized message scheduling solution is answering the question of whether someone is a regular customer or not. The answer determines if automated date and time optimization will be attempted. In the case of a new or irregular customer, a different execution path could be followed. Newly acquired customers for whom there is little data available do not present the same opportunity for optimization that repeat and regular customers provide.

The answer to the question of what makes a repeat and regular customer is very much dependent on the business domain. The definition of regular customer for a food retail business would probably be quite different from that of a cosmetics retailer. Our pilot test bed was a European retail brand in the market of consumable and disposable baby products. Anonymized customer purchase transaction data were sourced from the information systems used by the company and analyzed, focusing on a specific product.

The definition of repeat and regular customer that we eventually arrived at, for the business domain of this particular pilot company, is customers who: (i) made at least 2 purchases of a specific product in the last 2 months, and (ii) also had less than 20 days of variance between their purchases of that product, ever since the first time they bought the product.

To arrive at this definition we tested the accuracy of the regression model that will be introduced in the next section against six sets of customers of the pilot company who were

TABLE I. TYPICAL INCOMING TRANSACTION ENTRY EXAMPLE.

Key	Explanation	Type
Account Id	Unique identifier of the company (retail enterprise)	Numerical
Transaction Date	Date of the transaction	Date
Customer Code	Unique identifier of the customer who made the transaction	Numerical
Transaction Code	Unique identifier of the transaction	Numerical
Order Source	How the customer made the transaction	Categorical
Products	List of products purchased in the transaction	List of Categorical
Total Price	Total amount of money paid for the transaction	List of Categorical
Payment Method	How the customer paid for the transaction	Categorical
Order Status	Status of the payment	Categorical

TABLE II. FEATURES USED TO FEED THE PREDICTIVE MODEL.

Feature	Explanation	Type
Last Order	Date of the most recent transaction of each customer	Date
Minimum Consumption Duration	Minimum no. of days between two consecutive transactions per product unit	Numerical
Average Consumption Duration	Average no. of days between two consecutive transactions per product unit	Numerical
Maximum Consumption Duration	Maximum no. of days between two consecutive transactions per product unit	Numerical
Average Fluctuation	Standard deviation in no. of days between two consecutive transactions per product unit	Numerical
Days from First	Days lapsed since the first purchase of each customer	Numerical
Days from Last	Days lapsed since from the last purchase of each customer	Numerical
Last Product Amount	Number of products included in the last purchase per subscription	Numerical
Transaction Day	Preferred purchasing day of week per customer	Categorical
Transaction Hour	Preferred purchasing time of day per customer	Categorical
Mean Order Value	Average money spent per purchase per customer	Numerical
Sum Order Value	Total money spent per customer to date	Numerical
Value Cluster	Customer's cluster assignment based on total money spent to date	Categorical
Gender	Customer's gender prediction based on gender classification model	Categorical

segmented with respect to different combinations of values for: (i) total number of product purchases done by the customer, (ii) number of recent repurchases, and (iii) consistency in the time span recorded between the customer's repurchases. The values for these three metrics can be seen in Table III under "Transactions", "Recent purchases" and "Average Fluctuation", respectively.

TABLE III. PREDICTIVE PERFORMANCE ANALYSIS WITH ALTERNATIVE DEFINITIONS OF REPEAT/REGULAR CUSTOMER.

Transactions	Recent Purchases	Average Fluctuation	MAE	R^2
≥ 2	2	< 15	4.48	0.37
≥ 2	2	< 20	4.41	0.45
≥ 3	2	< 15	5.16	0.39
≥ 3	2	< 20	4.88	0.43
≥ 3	3	< 15	5.05	0.40
≥ 3	3	< 20	4.91	0.42

We measured the predictive accuracy of the regression model on the aforementioned customer data set in terms of the MAE and the R^2 score [26]:

- Mean Absolute Error (MAE): The average of the absolute errors (1). The MAE units are the same as the predicted target, which is useful for understanding whether the size of the error is of concern or not being robust to outliers. The smaller the MAE is, the better the algorithm's performance [27]. In (1), N is the total number of errors and $|x_i - x|$ equals the absolute errors.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |x_i - x| \tag{1}$$

- R^2 Score: A statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable in a regression model. The R^2 value varies between 0 and 1, where 0 represents no correlation between the predicted and actual value and 1 represents complete correlation [28].

As displayed in Table III, there was one particular set of customers for whom the regression model produced predictions with the smallest MAE and simultaneously the highest R^2 score. This was the set of customers who had made at least 2 purchases of a specific product in the 2 prior months and who had less than 20 days of variance between their purchases of that product.

B. Prediction of Repurchase Date using Regression

The next step in the solution is estimating the future point in time when a specific repeat/regular customer will need to repurchase a consumable or disposable product that they have been using in the past.

This estimation problem can be mathematically formulated as a regression problem of trying to predict a continuous value, i.e. the number of days from the customer's previous purchase to the next one. The target feature that our regression model is expected to produce is "Days to Next Purchase".

In search of the best approach to develop our regression model we tested a variety of machine learning algorithms including Logistic Regression [29], Random Forests [30], K Nearest Neighbors [31], CatBoost [32], LightGBM [33]. The algorithm that we found most balanced in terms of accuracy and processing performance for our goals was the XGBoost regression algorithm [34].

XGBoost belongs to the ensemble learning methods which are based on the need to rely upon the results of more than one machine learning models giving their aggregated output. Bagging and boosting are two widely used ensemble learner models. These two techniques can be used with several statistical models and the most predominant usage was achieved with decision trees [35]. XGBoost is a scalable machine learning algorithm that is based on boosting approach and incorporates several useful features, such as parallel processing, high flexibility, automated handling of incomplete values, inherent tree pruning and built-in cross-validation [36].

A pre-processing phase comes before the training of the main predictive model. This phase starts with the elimination of undue data features that appear to have a negative impact to the algorithm's predictive accuracy. After that a data transformation process takes place in order to convert all the categorical features into numerical representations. All the available data are then subdivided into training and testing subsets, following a fixed distribution of 80% and 20%, respectively. The algorithm is then ready to get trained using the training data and having multiple variable parameters, which have to be adjusted in the next step. The trained model is eventually able to make predictions on the "Days to Next" feature, which is then getting rounded in order to express number of days.

Configuring the XGBoost regression algorithm has to do not only with the parameters inside the algorithm itself, but also with the number of folds to be used during the cross-validation process which is responsible for the main tuning. Cross-validation is generally used to assess a machine learning model's ability to meet first-time data using a limited sample each time (i.e. data not used during the training phase). The number of different restricted samples is determined by the number of folds [37].

The described procedures utilize the so-called Grid-SearchCV function provided by the scikit-learn Python library [38]. It is possible to set the number of kernels to be used by the algorithm to the maximum permissible number by the system. In our configuration, we set the number of folds to 5 and selected the algorithm evaluation metric to be the R^2 score. The percentage shrinkage rate used in each update of the algorithm to avoid overfitting (i.e. the inability of the algorithm to adapt to new data despite its good performance in the test data) was set to alternate between the values of 0.01, 0.02, 0.05, 0.1, and 0.3. The maximum number of features to be used by the algorithm was declared to alternate between 3, 4, and 7, while the maximum depth of node representation of the algorithm tested up to levels 3, 4, and 5. XGBoost was let to randomly test 50%, 60%, and 70% of the training data before developing nodes, preventing the overfitting phenomenon. We also limited the number of repetitions during the training of the algorithm to not exceed 20 in order to prevent overfitting (Table IV).

C. Fine Tuning of the Message Delivery Date and Time with a Domain-Specific Personalization Algorithm

The last step in the execution of the solution is to run a domain-specific personalization algorithm. The algorithm accepts as input the repurchase date that was predicted by the regression model and applies domain-specific business rules to extract the optimal day of the week and time of the day to deliver a reminder message to the customer.

Note that the actual date and time that the reminder message should be delivered to the customer is not the same as the repurchase date produced by the regression model. The latter date is when the transaction is predicted to take place, whereas the message needs to be come earlier and serve as a timely reminder.

The day and time that the reminder message is sent to each customer should be chosen based on their individual profile and optimized to maximize conversion rate. Looking at past purchase transaction data it is easy to determine if some day of the week had a higher number of transactions associated with it, which could potentially suggest that making a purchase is more convenient for that customer on that specific day. Conversely, looking at past message exchange & engagement analytics data it is easy to determine if the customer is more likely to open, read or engage with a mobile message during a specific period within the day.

The algorithm analyzes the distribution of the transactions of a customer on a daily basis throughout the week, and then adds a weight to each day in order to reveal the one with the maximum weight, i.e. the dominant day. To calculate the dominant day, it is important to distinguish between the case where there is only one day with the maximum weight and the case where the customer has performed an equal number of purchases in more than one days during a week.

In the first case, the candidate dominant days are the ones having at least as many transactions as the average transactions of the whole week. In the second case, the candidate dominant days are the ones having the maximum number of transactions. For each one of those days, the weight is given by formula (2), where the days of week are numbered from 0 to 6 representing the days from Monday to Sunday:

$$Weight = |\# \ of \ daily \ transactions - |day \ of \ week - predicted \ day \ of \ week| \ | \qquad (2)$$

After calculating the weights, the algorithm selects the day with the maximum score (i.e. the customer's most active day) which is as close as possible to the day predicted by the regression model. It also checks whether the predicted day is the same as the day with the maximum weight (i.e. if the adapted day is the same as the one originally predicted as repurchase date). If this is not the case, the algorithm shifts the predicted day of week to the closest day having the maximum weight. At the end of the process, there is a personalized product reminder scheduled for each customer, which is based on their personal purchase and interaction preferences (i.e. "Adapted Predicted Days" feature).

Table V shows an example of applying the algorithm to transaction history data to get the distribution of the customer's

TABLE IV. SELECTED XGBOOST HYPERPARAMETER TUNING VALUES.

Parameter	Description	Value
Learning rate	Step size shrinkage used to prevent overfitting.	0.01
Max depth	How deeply each tree is allowed to grow during any boosting round.	3
Max features	The number of features to consider when looking for the best split.	10
Min child weight	Minimum sum of instance weight (hessian) needed in a child.	3
Subsample	Percentage of samples used per tree.	0.6
N jobs	The number of jobs to run in parallel for both fit and predict.	-1
Evaluation metric	Metric function to evaluate the predictions on the test set.	MAE
Early stopping rounds	Finishes training of the model early if the hold-out metric does not improve for a given number of rounds.	20

purchases over the days of the week. As shown in the table, the customer's most active day of week is Wednesday, with a total of 8 transactions.

TABLE V. EXAMPLE DISTRIBUTION OF A CUSTOMER'S PURCHASES.

Day	Purchases
Monday	3
Tuesday	6
Wednesday	8
Thursday	0
Friday	1
Saturday	1
Sunday	0

It so happens that there is a unique maximum value in the distribution, so the algorithm calculates the weights of days that are above the average transactions (i.e. in this case the average of Monday, Tuesday and Wednesday is equal to 2.7). Based on the respective aforementioned formula, their respective weights will be 3, 1 and 4. In this particular customer instance the predicted repurchase day was Sunday but the algorithm eventually set the message delivery date to Wednesday.

Our algorithm aims to move the message delivery date earlier than the customer's most active (dominant) day of the week, rather than just close to it. The algorithm checks whether the difference between the predicted repurchase date and the dominant day is less than 5 days and then shifts the message delivery to the immediately preceding dominant day. If the difference is at least 5 days, the algorithm shifts the message delivery date to the nearest dominant day, as previously. If the personalized message delivery date is in the past, the algorithm finds the same dominant day of week in the next week, and therefore moves the reminder day accordingly (i.e. producing a "Days to Remind" feature).

The optimal time to send a personalized message during the day is generated as a feature during the pre-processing phase (i.e. "Transaction Hour" feature in Table II). We also paid attention to enabling configuration of 'Do not Disturb' hours, i.e. time periods when sending message notifications to customers should not be allowed. The algorithm moves all recommendations between 9 p.m. and 3 a.m. to 9 p.m., and all recommendations between 3 a.m. and 9 a.m. to 9 a.m.

The combination of message delivery date and time which are determined by the algorithm produce the exact date and time that the personalized reminder will be scheduled.

D. Deployment Infrastructure

Regarding the deployment infrastructure, the solution has been implemented on IBM Cloud as a scheduled Jupyter Notebook with an Apache Spark configuration (i.e. on a server cluster) using the capabilities of the IBM Watson Studio suite [39, 40]. We used a regularly executing Python notebook that acts like an updater to the respective database which contains the mapping of date and time to send the next reminder message and unique encrypted customer code. The deployment model is depicted in Fig. 1.
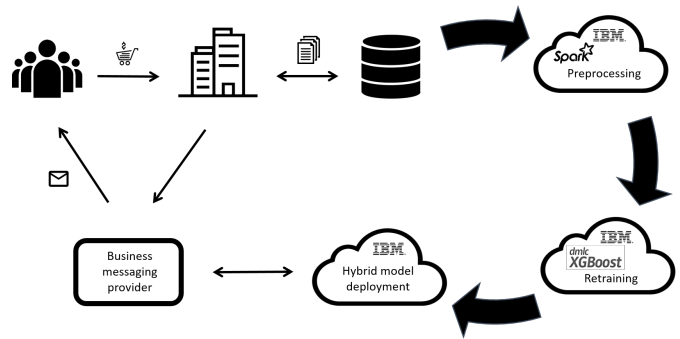

Fig. 1. Overview of the adaptive model deployment architecture.

VI. EVALUATION

A. Pilot Test Context

As highlighted in the sections above, the research presented in this paper addresses the problem of scheduling the delivery of marketing messages to repeat buyers of CPG products which require routine replenishment or replacement - such as food & beverages, cosmetics or household products. The objective is to increase both the customer engagement and the conversion rate. An indicator of customer engagement is the rate of messages that are read by the recipients (i.e. seen rate), as well as the percentage of customers who click on the hyperlink found inside the message text (i.e. click-through rate). The ultimate goal is to optimize the date and time that each individual consumer will receive a message that prompts them to order the product again, such that the rate of conversion from reminder message to product repurchase can be increased.

To evaluate the effectiveness of our approach we conducted a pilot test with one of Apifon's clients, a European CPG

brand in the market of baby products. This is a company that sells direct to consumers. Its revenues and profitability depend on its customers placing regular product reorders. The company has been using mobile text messages to regularly remind its customers to reorder consumable and disposable products that they purchased in the past, following a fixed scheduling approach. Whenever a customer orders a product of a specific type, a reminder message is scheduled to be sent *n* days later (the preset number of days varies per product type). The message is scheduled for delivery provided that the customer has given explicit consent to receive such reminders.

B. Experiment Design

To evaluate our proposed method, we analyzed and compared its effectiveness against the company's standard practice of scheduling reminders at fixed time intervals. The evaluation metrics that were used to compare the two different methods are the message seen and click-through rates and the customer repurchase rate that are achieved by each approach.

Our system was deployed live for a period of one week during September 2019 and managed the message scheduling process for an experiment cohort of 40 repeat customers who met the following criteria: (i) they made at least 2 purchases of a specific product in the last 2 months, and (ii) they also had less than 20 days of variance between their purchases of that product, ever since the first time they bought the product.

We measured the message seen and click-through rates, as well as the conversion rate for this cohort of 40 customers after 1, 2, 12 and 24 hours since the time that the reminder message was delivered to each of them, to determine the percentage of customers who received a reminder and actually went ahead and reordered the product.

As a next step we collected historical data from September 2018 (one year earlier) and created a benchmark cohort of 40 repeat customers who met the same criteria as the experiment cohort (i.e. identical number of recent product purchases and same consistency in the time span recorded between repurchases). Following, we measured the message seen and click-through rates, along with the conversion rate for these 40 customers after 1, 2, 12 and 24 hours since the time they had received the purchase reminder message (under a fixed scheduling approach).

This setup allows us to compare the impact of the scheduling method on each cohort: (i) the benchmark cohort consisting of 40 customers, who were sent product repurchase reminders using the company's standard method of fixed time intervals, and (ii) the experiment cohort consisting of 40 customers, who were sent reminders using the optimized scheduling method.

The hypothesis was that customers in the optimized scheduling group would not only interact with the marketing message they received (as measured by seen and click-through events), but also would repurchase products at a rate higher than that of customers in the other group, at a level of statistical significance not lower than the conventionally accepted 95% (p-value 0.05).

VII. Evaluation Results

The results of the pilot test are displayed in Tables VI, VII and VIII. They suggest that the new method of timing the reminder messages using the approach proposed in this paper outperforms the approach of scheduling messages at fixed time intervals. This is evident by both the message seen and click-through rates, and the customer repurchase rates within 1, 2, 12 and 24 hours from sending the message.

The message seen rate and click-through rate of the experiment cohort surpassed those of the benchmark cohort in all time span measurements. The results are statistically significant, with confidence levels over 95% across cases, as can be seen from the p-values in Tables VI and VII.

Similarly, the repurchase rate for the experiment cohort was higher than that of the benchmark cohort across all measurements (Table VIII). The measured increase in repurchase rate recorded after 1, 2 and 24 hours of sending the message has a confidence level over 95%. This is not the case for repurchases at the 12 hour time span. Fig. 2 illustrates how the repurchase rates of the experiment cohort (i.e. date and time optimized reminders) evolve over time compared to those of the benchmark cohort (i.e. the standard method).

VIII. Conclusion

Modern marketing automation technology is evolving to allow enterprises to achieve personalized communications with their customers at scale - delivering the right message at the optimal date and time to each individual subscriber. In the past five years we have seen an increasing number of research studies on the topic of email marketing personalization and data-driven models for repeat purchase prediction. Some of the most interesting work comes from research teams inside online retailers like Amazon.com and ASOS.com, who have direct access to large scale data.

Nevertheless, we have found limited existing research focusing on the following topics: (i) personalization of marketing communications via instant messaging platforms; (ii) personalization of marketing communications for repeat buyers of consumer packaged goods (CPG); (iii) personalization of instant messaging marketing with data protection by design.

The research discussed in this paper focuses on the intersection of the aforementioned areas. Specifically, this paper presents a solution to the problem of scheduling the delivery of consent-based, personalized instant-messaging reminders to repeat buyers of frequently replenished CPG products.

A notable contribution of the personalization approach presented in this paper is the use of predictive models trained on two distinct types of data: historical purchase transaction data augmented with instant messaging engagement data. The system is able to predict the next product repurchase date for each customer, which is then taken through a domain-specific personalization algorithm to determine the optimal day of the week and hour of the day when a repurchase reminder message should be sent to that customer.

The aim is to improve the timing of the repurchase recommendations sent to a subscriber such that they better reflect their needs and preferences, leading to a better customer experience and resulting in a higher rate of follow-on purchases.

The evaluation of our solution took place using online (live) testing with actual customers of a European retail brand in

TABLE VI. COMPARATIVE MESSAGE SEEN RATES FROM THE PILOT TEST.

Time lapsed since message sent	Benchmark cohort seen rates	Experiment cohort seen rates	Seen rate diff	p-value
1 hour	10/40 (25%)	17/40 (42.5%)	+ 17.5%	0.0461
2 hours	13/40 (32.5%)	21/40 (52.5%)	+ 20%	0.0323
12 hours	17/40 (42.5%)	25/40 (62.5%)	+ 20%	0.0338
24 hours	21/40 (52.5%)	29/40 (72.5%)	+ 20%	0.0295

TABLE VII. COMPARATIVE MESSAGE CLICK-THROUGH RATES FROM THE PILOT TEST.

Time lapsed since message sent	Benchmark cohort click-through rates	Experiment cohort click-through rates	Click-through rate diff	p-value
1 hour	0/40 (0%)	6/40 (15%)	+ 15%	0.0039
2 hours	1/40 (2.5%)	6/40 (15%)	+ 12.5%	0.0212
12 hours	4/40 (10%)	10/40 (25%)	+ 15%	0.0359
24 hours	5/40 (12.5%)	12/40 (30%)	+ 17.5%	0.0251

TABLE VIII. COMPARATIVE CUSTOMER CONVERSION RATES FROM THE PILOT TEST.

Time lapsed since message sent	Benchmark cohort repurchases	Experiment cohort repurchases	Repurchase rate diff	p-value
1 hour	0/40 (0%)	3/40 (7.5%)	+ 7.5%	0.0359
2 hours	0/40 (0%)	3/40 (7.5%)	+ 7.5%	0.0359
12 hours	2/40 (5%)	5/40 (12.5%)	+ 7.5%	0.1155
24 hours	2/40 (5%)	7/40 (17.5%)	+ 12.5%	0.0356

the market of baby products. The results suggest a significant improvement in marketing effectiveness, compared to the standard method of scheduling product repurchase reminders at fixed time intervals. Our personalized scheduling approach resulted in a significantly higher rate of customer interactions and conversions within 1, 2, 12 and 24 hours from sending the personalized reminder. There was a notable increase in the ratio of customers who engaged with the message after receiving it (up to 17.5% higher), as well as an increase in the ratio of customers who completed a repurchase transaction (up to 12.5% higher). The results of the pilot test are encouraging although the experiment is limited in two ways. Firstly, the total sample size is not very large. Secondly, we used a benchmark comparison approach rather than a controlled experiment setup. It is understood that these factors could potentially be contributors to sampling error, however the research team has tried to mitigate this by careful sample design and using data from the same period of the year to neutralize the seasonality effect. As part of future work we plan to design a larger-scale experiment and generate additional insights into the effectiveness of our proposed method.

ACKNOWLEDGMENT

REFERENCES

[1] Chen, P. and Hsieh, H. (2012). Personalized mobile advertising: Its key attributes, trends, and social impact. Technological Forecasting and Social Change, 79(3), pp.543-557.

[2] Koch, O. and Benlian, A. (2015). Promotional Tactics for Online Viral Marketing Campaigns: How Scarcity and Personalization Affect Seed Stage Referrals. Journal of Interactive Marketing, 32, pp.37-52.

[3] Heimbach, I., Kostyra, D. and Hinz, O. (2015). Marketing Automation. Business & Information Systems Engineering, 57(2), pp.129-133.

[4] Awadzi, D., Calloway, D. and Awadzi, D. (2018). Increasing Conversions through Behavioral Retargeting. International Academic Journal of Social Sciences, 05(01), pp.23-27.

[5] Deligiannis, A., Argyriou, C. & Kourtesis, D. (2019). Predictive personalization of conversational customer communications with data protection by design. IEEE/WIC/ACM International Conference on Web Intelligence on - WI '19 Companion.

[6] Dangi, H., & Malik, A. (2017). Personalisation in marketing: an exploratory study. International Journal Of Internet Marketing And Advertising, 11(2), 124. doi: 10.1504/ijima.2017.084079

[7] Abakouy, R., En-naimi, E., Haddadi, A., & Lotfi, E. (2019). Data-driven marketing. Proceedings Of The 4Th International Conference On Smart City Applications - SCA '19. doi: 10.1145/3368756.3369024

[8] Deligiannis, A., Argyriou, C., & Kourtesis, D. (2020). Building a Cloud-based Regression Model to Predict Click-through Rate in Business Messaging Campaigns. International Journal Of Modeling And Optimization, 10(1), 26-31. doi: 10.7763/IJMO.2020.V10.742

[9] Ding, Y., & Li, X. (2005). Time weight collaborative filtering. In Proceedings of the 14th ACM international conference on Information and knowledge management (pp. 485-492).

[10] Koren, Y. (2009). Collaborative filtering with temporal dynamics. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 447-456).

[11] Baltrunas, L., & Amatriain, X. (2009). Towards time-dependant recommendation based on implicit feedback. In Workshop on context-aware recommender systems (CARS'09) (pp. 25-30).

[12] Wang, J. and Zhang, Y. (2013). Opportunity model for e-commerce recommendation: Right product; right time. Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval - SIGIR '13.

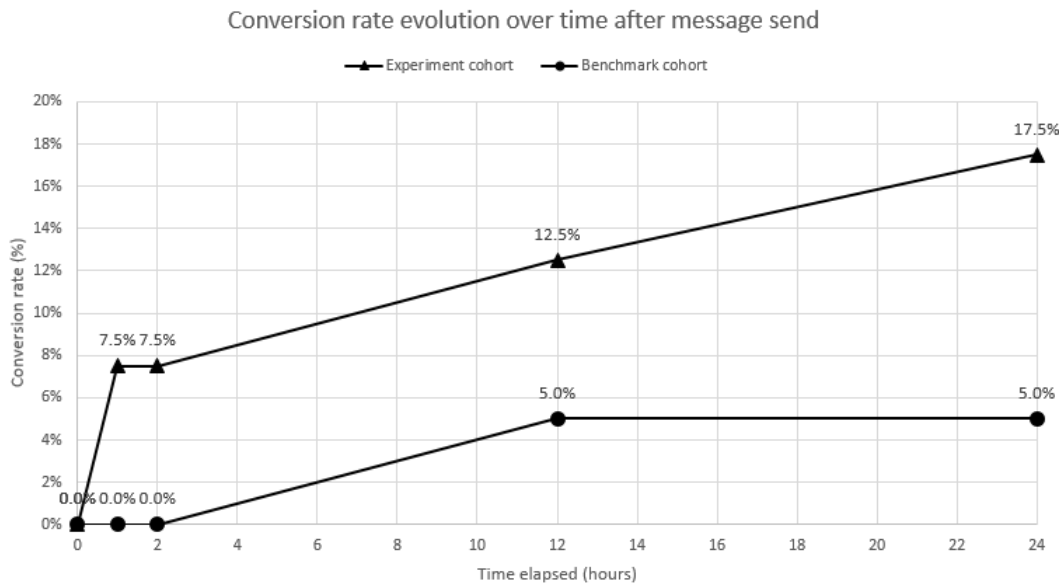[13] Dey, S., Mitra, P. and Gupta, K. (2016). Recommending Repeat Pur-

Fig. 2. Comparing the repurchase rates of benchmark cohort and experiment cohort at 1, 2, 12 and 24 hour intervals.

chases using Product Segment Statistics. Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16.

[14] Liu, G., Nguyen, T., Zhao, G., Zha, W., Yang, J., Cao, J., Wu, M., Zhao, P. and Chen, W. (2016). Repeat Buyer Prediction for E-Commerce. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16.

[15] Chamberlain, B., Cardoso, Â., Liu, C., Pagliari, R. and Deisenroth, M. (2017). Customer Lifetime Value Prediction Using Embeddings. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17.

[16] Cadavid, J. P. U., Lamouri, S., & Grabot, B. (2018). Trends in Machine Learning Applied to Demand & Sales Forecasting: A Review.

[17] Ferreira, K., Lee, B. and Simchi-Levi, D. (2016). Analytics for an Online Retailer: Demand Forecasting and Price Optimization. Manufacturing & Service Operations Management, 18(1), pp.69-88.

[18] Bonfrer, A., Drèze, X. (2009). Real-Time Evaluation of E-mail Campaign Performance, Marketing Science 28(2), 251-263.

[19] Zhang, X., Kumar, V. and Cosguner, K. (2017). Dynamically Managing a Profitable Email Marketing Program. Journal of Marketing Research, 54(6), pp.851-866.

[20] Paralič, J., Kaszoni, T. and Mačina, J. (2019). Predicting Suitable Time for Sending Marketing Emails. Advances in Intelligent Systems and Computing, pp.189-196.

[21] Naik, N. (2017). Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP. 2017 IEEE International Systems Engineering Symposium (ISSE).

[22] Dobbelaere, Philippe & Sheykh Esmaili, Kyumars. (2017). Kafka versus RabbitMQ

[23] George, D., Reutimann, K. and Tamò-Larrieux, A. (2018). GDPR Bypass by Design? Transient Processing of Data Under the GDPR. SSRN Electronic Journal, pp.6-17.

[24] Lachaud, E. (2019). Adhering to GDPR Codes of Conduct: A Possible Option for SMEs to GDPR Certification. SSRN Electronic Journal.

[25] Sai, L., Shreya, M., Subudhi, A., Lakshmi, B. and Madhuri, K. (2017). Optimal K-Means Clustering Method Using Silhouette Coefficient. International Journal of Applied Research on Information Technology and Computing, 8(3), p.335.

[26] Botchkarev A. (2019). A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. Interdisciplinary Journal of Information, Knowledge, and Management, 14, pp.045-076.

[27] Willmott, C. and Matsuura, K. (2005). Advantages of the mean absolute

error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, 30, pp.79-82.

[28] Alexander, D., Tropsha, A. and Winkler, D. (2015). Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. Journal of Chemical Information and Modeling, 55(7), pp.1316-1322.

[29] Steyerberg, E. W., Harrell Jr, F. E., Borsboom, G. J., Eijkemans, M. J. C., Vergouwe, Y., & Habbema, J. D. F. (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. Journal of clinical epidemiology, 54(8), 774-781.

[30] Evans, J. S., & Cushman, S. A. (2009). Gradient modeling of conifer species using random forests. Landscape ecology, 24(5), 673-683.

[31] Batista, G. E., & Monard, M. C. (2002). A Study of K-Nearest Neighbour as an Imputation Method. HIS, 87(251-260), 48.

[32] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

[33] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems (pp. 3146-3154).

[34] Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). ACM.

[35] Sumana, B. and Santhanam, T. (2015). Optimizing the Prediction of Bagging and Boosting. Indian Journal of Science and Technology, 8(35).

[36] Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, pp.785-794.

[37] Zhang, L., & Zhan, C. (2017, May). Machine learning in rock facies classification: an application of XGBoost. In International Geophysical Conference, Qingdao, China, 17-20 April 2017 (pp. 1371-1374). Society of Exploration Geophysicists and Chinese Petroleum Society.

[38] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.

[39] Palmer, T. (2019). Predict and Optimize Business Outcomes with IBM Decision Optimization for Watson Studio and IBM Cloud Pak for Data. The Enterprise Strategy Group, Inc., pp.3-14.

[40] Bozhinov, I. (2019). AI and big data on IBM Power Systems servers. U.S.A.: IBM Corporation.