IEEE *Access*
Multidisciplinary : Rapid Review : Open Access Journal

Forecasting Conversion Rate for Real Time CPC Bidding with Target ROAS

SEMIH BULUT[1] and EMIN AVCI[2] and AHMET BULUT[3]

[1]Accounting and Finance PhD Program, Social Sciences Institute, Marmara University, İstanbul, Türkiye; semihblt@googlemail.com
[2]Department of Business Administration, Marmara University, İstanbul, Türkiye; eavci@marmara.edu.tr
[3]Department of Computer Science, Acıbadem University, İstanbul, Türkiye; ahmet.bulut@acibadem.edu.tr

Corresponding author: Ahmet Bulut (e-mail: ahmet.bulut@acibadem.edu.tr).

Dr. Bulut was with Carbon Health https://carbonhealth.com/ during the course of the study.

ABSTRACT For bidding in real time, the rate of customer conversion needs to be predicted in real time. Using the rate prediction and the target return on ad spend, a competitive CPC bid can be computed. In our study, we built two models, i.e., MoM and MCI namely, for forecasting the rate of conversion. The results we obtained by applying our models on the marketing campaigns of two startups were promising. Both MoM and MCI run in constant $O(1)$ time, and require $O(n)$ space for n observations. Furthermore, both models can be updated with fresh data in $O(1)$ time; hence, they are suitable for a data streaming application where new data arrives continuously in an online manner.

INDEX TERMS real time bidding, conversion rate modeling, return on ad spend

I. INTRODUCTION

THE advances in electronic commerce and information technologies have enriched customer acquisition practices. Companies reach their customers through Internet, and offer their products and services to a wider audience [1]. Online marketing campaigns have become the primary way of acquiring new customers. Digital ad spend worldwide is expected to reach $601.84 billion this year, up 9.5% from $549.51 billion in 2022, according to eMarketer forecast [1]. Collecting right data, monitoring operational metrics, and relating them to customer acquisition are essential in a competitive environment [2]. The data collected includes the number of users visiting a website, the amount of time they spend on the site, and whether or not they purchase goods and services. It is now possible to connect all the way through the customer journey, which starts with an ad click and which ends with a purchase. Since the main goal of online campaigning is to bring customers in and have them purchase goods and services, the rate of "buying" customers is a key success factor.

Advertisers use search engines for marketing their products to online users. A search marketing campaign consists of search keywords. When a user query matches up with a search keyword in a given campaign, then one of the ads in the campaign is displayed on the search results page (SERP) to the query. This page contains an ordered list of webpages and ads, which are relevant to the query. To determine which ads to display, the search engine holds a real time bid auction among all the eligible ads. There are a limited number of ad slots on any given page, and the actual number of slots determines the number of ads to display along with the organic search results. When a user clicks on an ad displayed on the SERP, the owner of the ad gets charged by the search engine. This cost per click (CPC) is the actual amount paid by the advertiser for the ad click.

In order to measure the success of an ad campaign, a ratio of how often users click on the ads they view is used. This ratio is called click-through rate (CTR). When an ad is shown to a user on a search results page, it counts as one impression. When the user clicks on the ad, it counts as one click. CTR is the ratio of user clicks to user impressions:

$$CTR = \frac{Clicks}{Impressions} \times 100 \qquad (1)$$

If an ad gets $1,000$ impressions and 85 clicks; its CTR is $\frac{85}{1000} \times 100$, i.e., 8.5%. When a customer clicks on the ad and purchases something post click, it is called a conversion. The percentage of "converted" customers is a commonly used performance metric called conversion rate (CR). For a company selling sunglasses, the percentage of people, who buy a

pair of sunglasses post-click, may be a suitable measurement of CR. Formally,

$$CR = \frac{Conversions}{Clicks} \qquad (2)$$

If $1,000$ people click on the sunglasses ad and only 20 of them end up buying a pair, then $20/1000 = 2\%$ is the CR of the ad.

Suppose that a search campaign consists of a single keyword in order to sell a product online. The product sells for \$50.00 per unit. Each time a potential customer clicks on the ad, the company will pay a price per click. Assume that 50 users clicked through the ad, and only 5 of them bought the product. The company's revenue will be $\$50.00 \times 5 = \$250.00$. In order to have this revenue, the company purchased 50 user clicks. In this case, the return on ad spend[2] (ROAS) of this campaign is equal to:

$$ROAS = \frac{Gain}{Cost} \qquad (3)$$

where the gain is the total revenue made out of all conversions, and the cost is the price paid for all user clicks. It is vital for a company to lower the cost and increase ROAS. The above formula can also be re-written as:

$$Cost = \frac{Gain}{Target\ ROAS} \qquad (4)$$

which implies that the company can adjust the cost per campaign according to the projected earnings and the target ROAS. If the target ROAS is 200%, then the overall cost cannot exceed $\$250.00/2 = \$125.00$. Hence, the cost per click[3] must be at most $\$125.00/50$, i.e., \$2.50.

The marketer determines a bid price she is willing to pay per click. With a competitive bid, a good quality ad has a better chance of a placement on a given SERP. The bid based on CPC is computed as follows:

$$Clicks \times CPC = \frac{Conversions \times Conversion\ Value}{Target\ ROAS} \qquad (5)$$

$$CPC\ bid = \frac{Conversions \times Conversion\ Value}{Clicks \times Target\ ROAS} \qquad (6)$$

The conversion value is fixed. The target ROAS is an input parameter. For real time bidding, we need to forecast the rate of conversion, i.e., $Conversions\ /\ Clicks$ in real time. Using the target ROAS and the $\widehat{CR}$ forecast, we can compute the CPC bid as follows:

$$CPC\ bid = \widehat{CR} \times \frac{Conversion\ Value}{Target\ ROAS} \qquad (7)$$

This CPC bid would be used in a real time bid auction. Henceforth, our objective boils down to having an accurate CR forecast in real time for CPC bidding with the ultimate goal of hitting the target ROAS according to Equation 7.

---

[2]https://support.google.com/google-ads/answer/6268637
[3]Note that there were a total of 50 user clicks purchased.

A. THE ORGANIZATION OF THE PAPER

In the next section, the related works are discussed in detail. In the methodology section, our theoretical framework including our models and algorithms are presented. In the experimental evaluation section, we present our empirical findings and discuss their implications. In the final section, we put our contributions into perspective, and conclude with key takeaways from our study.

II. LITERATURE REVIEW

Learning user tendencies from historical data is a well-studied problem. There exist efficient methods for this purpose [3], [4], but the training of such models require extensive time, which is not suitable in real time applications. On the other hand, simpler numerical methods are faster to train. They are real time and could adapt fast to the changes in the market because of their speedy re-training on new and fresh data.

In real time bid auctions, each advertiser submits a bid. To compute the bid, the rate of return needs to be predicted. The submitted bid may turn out to be optimistic or pessimistic depending on the nature of prediction. An overly pessimistic approach runs the risk of getting no return. An overly optimistic approach however may increase the cost of advertising prematurely. Naturally, there exists a rich spectrum between the two extremes. A flexible mechanism to explore this rate spectrum would help discover a working trade-off between optimism and pessimism [5].

The ideal is to find a balance between exploiting rates, which are known with certainty to deliver good returns, and exploring the range of rates where there is uncertainty about their possible returns. Various algorithms have been proposed to find a balance between exploration and exploitation. A well-known method is called upper confidence bound [6], [7]. The Gittins' bayesian optimal approach maximizes the expected cumulative returns over a given prior distribution [8]. There is a heuristic called Thompson sampling [9], [10]. It was used in revenue management [11], web site optimization [12], in online advertising [13], in selective data acquisition for improving machine learning models [14], and more interestingly in designing multi-armed bandits (MABs) [15]. The MABs were shown to be useful in a wide range of applications, which require continuous improvement, ranging from online service economy [16], to portfolio selection in finance [17], [18], and to real time bid prediction in online advertising [19].

Sampling a "prospective" rate of conversion from a beta distribution, which models the historical rate data, enables flexibility and experimentation between the exploitative and the exploratory [20]. Furthermore, a confidence interval provided by the underlying rate distribution presents itself as yet another flexible inference tool for experimenting with possible futures.
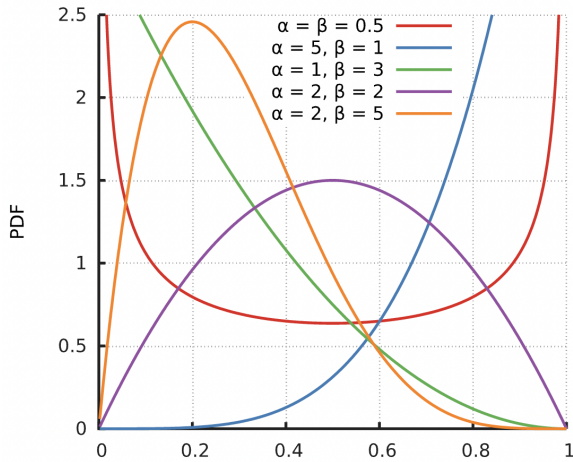
IEEE *Access*

FIGURE 1. The PDF of the beta distribution for varying values of α and β (source: Wikipedia).

III. METHODOLOGY

Our first method is to use the median of the medians of multiple beta distributions for a multi-resolution view over the observed CR data; our second method is to combine multiple confidence intervals over the expected CR into a single CR estimate. The building block for both of our methods is the beta distribution, the properties of which we present next.

A. BETA DISTRIBUTION

A beta distribution is a continuous probability distribution to model "probabilistic" outcomes. One could use it to model the conversion rate of a campaign, the click-through rate of customers visiting a website, the 10-year survival chance of patients with aortic aneurysm, and etc. It models the probability of success where the probability is a random variable.

The beta distribution has two shape parameters $\alpha, \beta > 0$. One can think of α as the number of trials with some positive payoff, i.e., successes, and β as the number of trials with zero payoff, i.e., failures. Suppose that the number of successes are 95 out of 100 trials. Then, α is 95 and β is 5.

For $\alpha \gg \beta$, the distribution will be right-skewed; whereas for $\beta \gg \alpha$, it will be left-skewed. As α and β values get larger, the spread of the distribution gets narrower and it becomes more concentrated. For a given $\alpha, \beta > 0$, the probability density function (PDF) of the beta distribution is computed as follows:

$$f(X|\alpha,\beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} X^{\alpha-1}(1 - X)^{\beta-1} \quad (8)$$

where Γ is the gamma function. The PDF for varying values of α and β are shown in Figure 1.

When there is complete information over a given phenomenon, we could make the best decision. In practice, an iterative approach is used in order to have confidence while making decisions. In this regard, a model that could be

FIGURE 2. Multiple time windows are constructed at different time resolutions over the observations.

updated easily with the information accrued over time is a practical decision making tool (with respect to the Bayesian theory) [21]. The beta distribution has an additive structure, which allows the gradual integration of newly observed data into the model. With δ new successes and δ new failures, the update procedure is completed in $O(1)$ time as follows:

$$\alpha_{new} = \alpha_{old} + \delta_{success} \quad (9)$$
$$\beta_{new} = \beta_{old} + \delta_{failure} \quad (10)$$

The new beta model with α_{new} and β_{new} as its shape parameters could be used in making future predictions.

A beta distribution has a mean and a median, which could easily be used as a CR forecast. The mean and the median of a beta distribution are computed in $O(1)$ time using α and β values as follows [22]:

$$E(X|\alpha,\beta) = \frac{\alpha}{\alpha + \beta} \quad (11)$$

$$Median(X|\alpha,\beta) \approx \frac{\alpha - 1/3}{\alpha + \beta - 2/3} \text{ for } \alpha,\beta > 1 \quad (12)$$

The distribution also provides a confidence interval over the observed CR. A point estimate for CR may be error-prone. To increase confidence in forecast, one could use a numerical range rather than a numerical value, and then expect that the true CR may fall within range. This range is called the confidence interval [23].

We use multiple beta distributions modeling the observed CR data at multiple time resolutions, i.e., time windows. Each model is used to compute a separate CR forecast, which is either a single point value or a numerical range. Then, we synthesize these multiple forecasts into a single CR forecast. We propose two methods for the synthesis: (1) the median of medians, and (2) the merge of confidence intervals.

B. MEDIAN OF MEDIANS (MOM)

The median is the value at the center position when all observations are ordered naturally [24]. The median divides the frequency distribution into two halves [25]. It is used for quantitative data, it is easy to compute, and it is not distorted by the data outliers.

Consider the example shown in Figure 2. The time windows w_is of length 1, 2, 4, 8, and 16 days are used to summarize the observations made over the last 1 day, last 2

days, last 4 days, last 8 days, and the last 16 days respectively. For a given time window w_i, the number of clicks received and the number of conversions made are known. In order to model these observations, a beta distribution with $\alpha_i = conversions_i$ and $\beta_i = clicks_i - conversions_i$ is computed. Note that the conversions are being treated as successes while the clicks without conversions are being treated as failures. Using the distribution, multiple statistics can be computed regarding the observations made, including mean, median, mode, $90-95\%$ confidence intervals, and etc. In the running example, five medians are computed, i.e., one median for each of the five windows w_0, w_1, w_2, w_3, and w_4. The set of medians is a series itself, which will have its own median. In fact, this median of medians (MoM) becomes our final CR forecast.

Algorithm 1 The MoM algorithm for computing the median of multiple beta medians as the CR forecast.

Require: α_i and β_i values of all windows $\{..., w_i, ...\}$;
1:
2: $medians = []$ // initialize the median series.
3:
4: **for** $w_i \in \{..., w_i, ...\}$ **do**
5:     $median_i = beta.median(\alpha_i, \beta_i)$
6:     $medians.append(median_i)$ // append to the series.
7: **end for**
8:
9: $CR = median(medians)$ // find median of medians.

The MoM method is parametric. The number of time windows used and the resolution (length) of these windows could be adjusted. In the running example, only five time windows (exponentially growing in length) were used. The method scales linearly with the number of windows used as shown in Algorithm 1.

1) Time and space complexity

Suppose that the size of the largest window is n. In this case, MoM needs to store n-many past CR values. Therefore, its space complexity is $O(n)$. Let k denote the number of windows used, which also corresponds to the number of medians to compute. We can compute each median in constant $O(1)$ time using Approximation 12. Hence, the runtime complexity of MoM is $k \times O(1) = O(k)$, which is linear in the number of windows used [26]. The ideal value of k can be found using grid search in the parameter space. The time complexity of the search would be $O(\log n)$ when the sizes of the windows used are powers of 2, i.e., $\{2^0, 2^1, 2^2, \ldots, 2^k\}$ where $2^k = n$.

C. MERGE OF CONFIDENCE INTERVALS (MCI)

The confidence interval is calculated by computing equal areas around the median. The confidence level is provided as input, and the output is confidence interval at that level. Suppose that there are 5 successes and 20 failures in 25 trials. For the time window covering all data points (all time window), an interval with a lower limit of 0.096 denoted by
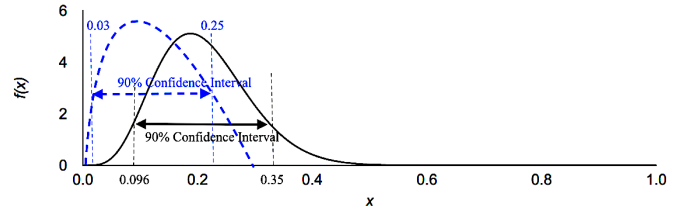

FIGURE 3. The confidence intervals illustrated for a hypothetical example.

L and an upper limit of 0.35 denoted by U is obtained at a confidence level of 90%. For an alternative time window of the last 30 days (recent time window), a new interval with a lower limit of 0.03 denoted by l and an upper limit of 0.25 denoted by u is obtained as shown in Figure 3.

Algorithm 2 The MCI algorithm for merging the two confidence intervals to a final CR forecast.

Require: A and B for **all time window**;
Require: α and β for **recent time window**;
1:
2: // Obtain 90% CI limits & mean for **all time window**.
3: $L, U = beta.interval(90\%, A, B)$
4: $M = beta.mean(A, B)$
5:
6: // Obtain 90% CI limits & mean for **recent time window**.

7: $l, u = beta.interval(90\%, \alpha, \beta)$
8: $\mu = beta.mean(\alpha, \beta)$
9:
10: // Combine intervals into a CR forecast: see Figure 4.
11: **if** $l \leq L$ **then**
12:     $CR = min(u, M)$
13: **else if** $L < l$ and $l < U$ **then**
14:     $CR = U$
15: **else if** $U \leq l$ **then**
16:     $CR = \mu$
17: **else**
18:     $CR = M$
19: **end if**

A confidence interval represents a certain belief over the observed CR. Given two different beliefs –one computed over all data points and the other computed over only the most recent data points– the question becomes how to merge these two different beliefs into a single belief such that the new information present in more recent data weighs more than less recent data. The hypothesis is that the concept drifts in data, e.g., consumer trends changing over time, could be more pronounced within the recent time window. The algorithm for merging two confidence intervals into one belief is outlined in Algorithm 2. Note that the confidence level is fixed at 90%; however, it could be parameterized.

In order to clarify how the merge operation works, all applicable notions and use cases are illustrated in Figure 4.
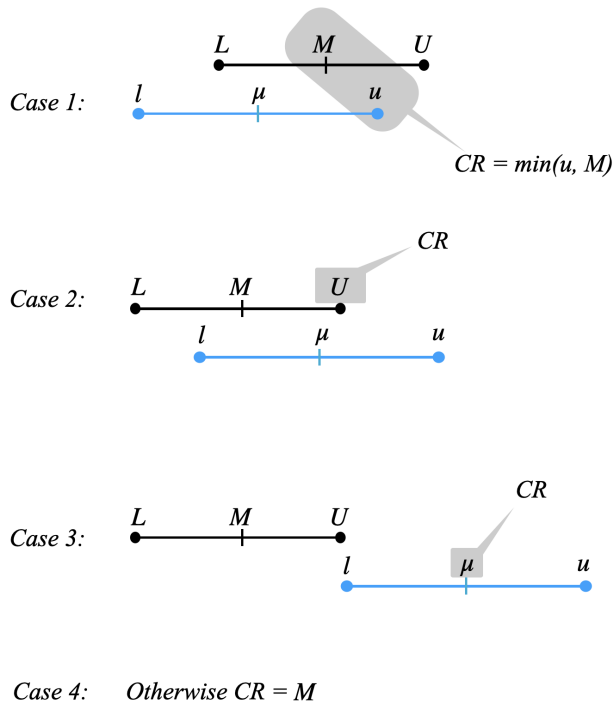
4

IEEE *Access*

FIGURE 4. The merging of two confidence intervals into a CR forecast.

When $l \leq L$, this would mean that the CR may be on a downward trend recently; therefore, the smaller value between u and the mean M of the all time window is chosen as the CR forecast. This is a pessimistic choice. When l falls within $[L, U]$, this would mean that the CR may be on an upward trend recently; therefore, the upper limit U is chosen as the CR forecast. This is an optimistic choice. If l exceeds U, the CR trend indicates that the future could be brighter. Therefore, the mean $\mu \geq l$ of the recent time window is used as the CR forecast. This is an overly optimistic choice. In all other cases, the CR forecast is set to M as a neutral choice.

1) Time and space complexity

The MCI algorithm requires the computation of two confidence intervals in steps 3 and 7. The computation of an interval takes $O(1)$ time because there exist numerical methods to estimate a binomial proportion [27]. The computation of the means in steps 4 and 8 takes $O(1)$ time as well using Equation 11. The algorithm maintains two counters as the number of successes and the number of failures for each window. The counters can be maintained in $O(1)$ time by subtracting the oldest observations and adding in the newest observations in a sliding window. A naive MCI would need to store all observations within "all" time window of length n. Overall, the time complexity of MCI is $O(1)$ while its space complexity is $O(n)$. The ideal window combination can be determined in $O(\log n)$ time using binary search in the parameter space.

IV. EXPERIMENTAL EVALUATION

We used the datasets of Doost and Grou.ps. Doost is a web-based roadside assistance services provider. Doost ran online campaigns between May 2019 and June 2022. Its dataset contains daily performance data. A summary of the dataset is shown in Table 1. Grou.ps is an online platform for building online communities [28]. Its dataset contains weekly performance data for the marketing campaigns, which ran from June 2012 to June 2013. A summary of the dataset[4] is shown in Table 2.

TABLE 1. A subset of data belonging to Doost's generic services campaign.

Day	Clicks	Conversions
2021-04-26	16	4
2021-04-27	6	1
2021-04-28	8	2
2021-04-29	16	6
2021-04-30	6	0
2021-05-01	8	0
2021-05-02	23	3
2021-05-03	15	2

TABLE 2. A subset of data for one of the Grou.ps campaigns.

Week	Impressions	Clicks	Conversions	Cost
2012-06-18	2436	392	3	$761.14
2012-06-25	2862	332	2	$629.99
2012-07-02	2527	226	2	$629.17
2012-07-09	2121	241	2	$618.88
2012-07-16	3022	213	4	$563.87

Doost was a digital-first initiative by the Netherlands-based Achmea group and Eureko Sigorta of Turkey. As the first entrant to the roadside assistance market in Turkey, its goal was to ease the pain associated with a car breakdown. Doost deployed 64 campaigns in total, two of which spent most of the marketing budget. The first of these two campaigns called *Generic Services* had a spend of $255K$ Turkish liras. It received $152,164$ customer clicks and $23,807$ conversions. The other campaign called *Roadside Assistance* had a spend $123K$ Turkish liras. It received $70,738$ customer clicks and $14,822$ conversions.

Established in 2005, Grou.ps allows individuals and groups to come together and create interactive communities around a shared interest. It provided service to online gaming groups, e-learning classes, fan communities, charity organizations, alumni communities, and event planning companies. Its website traffic soared from zero in 2008 to 8 million monthly unique visitors in 2011.

A. EVALUATION METRICS

We evaluated the performance of each model in predicting the rate of conversion. The CR predictions were compared with the actual CRs in order to compute the prediction

---

[4]The campaigns are anonymized due to proprietary reasons.

accuracy. There are two widely used metrics for measuring the accuracy of numerical predictions:

1) Root Mean Square Error (RMSE): For an evaluation window of n predictions, the errors made in each prediction are first squared; then, all these square errors are summed up. Finally, the square root of the square-sum divided by n is computed. The result is the RMSE:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{n}} \qquad (13)$$

where y_i is the actual CR value and $\hat{y_i}$ is the CR prediction. The difference between the two is the prediction error.

2) Mean Absolute Error (MAE): For an evaluation window of n predictions, the errors made in each prediction are first summed up. Then, the final sum is divided by n. The result is the MAE:

$$MAE = \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})}{n} \qquad (14)$$

MAE and RMSE are commonly used together to diagnose the variation in the errors made in a given set of predictions. Since the errors are squared before they are averaged, the RMSE is always greater than or equal to the MAE. The greater is the difference between them, the greater is the variance in the individual errors made.

B. SELF EVALUATION
We performed a fine-tuning study in order to identify the best parameter combination for each of our models. We present the empirical results obtained here.

1) MoM's prediction accuracy vs. window size
The MoM's performance was tested on Doost's Generic Services campaign. Table 3 shows the results when a single window was used. In this experiment, the prediction accuracy of MoM was measured while varying the window size. The window size varied from the last 2 days to the last 64 days. The results indicate that as the window size increased, the error decreased until a certain time period, i.e., 8 days in this particular case. After 8 days, the error increased with increasing window size as shown in Figure 5. In summary, a week long window sufficed to capture the data dynamics well.

TABLE 3. The impact of the window size on MoM's prediction accuracy. Doost's Generic Services campaign was used in the experiment with a single window.

Window size	2	4	**8**	16	32	64
MAE	0.030	0.029	**0.026**	0.026	0.028	0.030
RMSE	0.046	0.046	**0.035**	0.036	0.038	0.040

The impact of the window size was analyzed on Doost's Roadside Assistance campaign using the same setup. A similar trend was observed: as the window size increased, the error decreased until a certain time period. Two weeks long
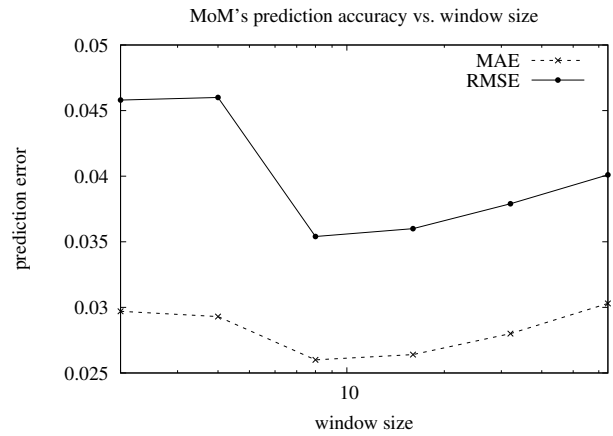

FIGURE 5. The MoM's prediction accuracy vs. window size. Doost's Generic Services campaign was used in the experiment with a single window.
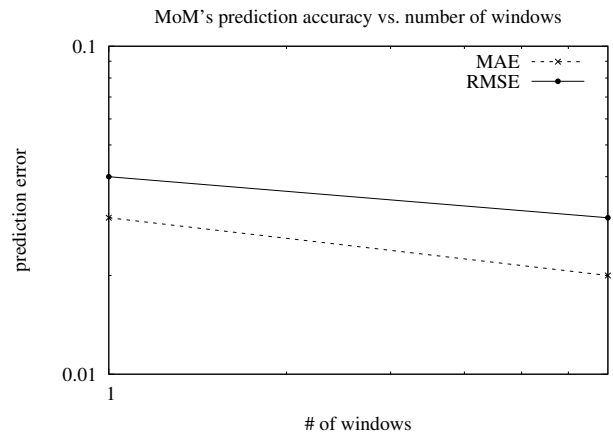

FIGURE 6. The impact of the number of windows on MoM's prediction accuracy. Doost's Generic Services campaign was used in the experiment.

to a month long window worked well in practice as shown in Table 4.

TABLE 4. The impact of the window size on MoM's prediction accuracy. Doost's Roadside Assistance campaign was used in the experiment with a single window.

Window size	2	4	8	**16**	32	64
MAE	0.028	0.027	0.026	**0.025**	0.025	0.028
RMSE	0.038	0.037	0.036	**0.036**	0.036	0.039

2) MoM's prediction accuracy vs. number of windows
The MoM's performance was evaluated while varying the number of windows on Doost's Generic Services campaign. The number of windows varied between 1 and 7. The results show that the error decreased as the number of windows increased as shown in Figure 6, which provides empirical evidence for the median of multiple medians elevating model performance.

In summary, the MoM model responded well to both the number and the size of time windows. As the window size
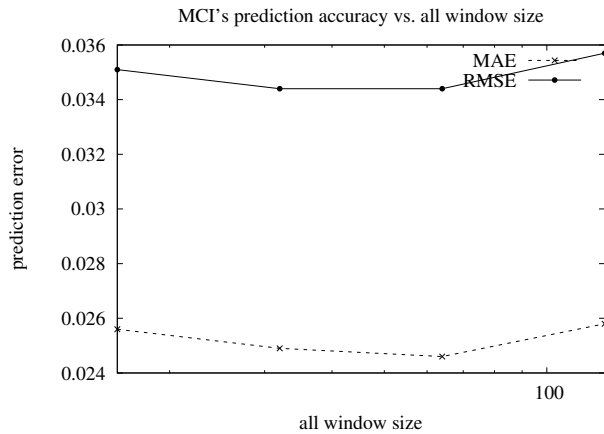
IEEE *Access*

FIGURE 7. The MCI's prediction accuracy vs. all window size. Doost's Generic Services campaign was used in the experiment; the recent window size was fixed at 8 days.

increased, the prediction error decreased until a threshold. We observed that shorter or longer time windows than this threshold adversely effected the prediction accuracy whereas the ideal window size captured the market dynamics well.

3) MCI's prediction accuracy vs. all window size

The MCI's accuracy was evaluated while varying all window size on Doost's Generic Services campaign. The longer (all) window size varied from 16 to 128. The shorter (recent) window size was fixed at 8. Increasing the longer window size decreased the error till a saddle point, i.e., 64 days. Beyond the saddle point, the error trend reversed as shown in Figure 7. The MCI method worked well with a combination of windows that captured the market dynamics well. In our experimental case, the best performing $(all, recent)$ window combination was $(64, 8)$.

A similar experiment was performed on Grou.ps dataset as well. The results are shown in Table 5. The saddle dynamics was also observed in this case. The best performing $(all, recent)$ window combination was $(12, 2)$. Both of these results indicate that a non-linear combination of window sizes had the most impact. When the windows are linearly correlated, the sudden and rapid changes in the market dynamics might be harder to detect.

TABLE 5. The MCI's prediction accuracy vs. all window size. The second campaign of Grou.ps was used in the experiment.

All window size	Recent window size	MAE	RMSE
4	2	0.015	0.020
8	2	0.011	0.014
12	**2**	**0.011**	**0.013**
16	2	0.013	0.014

4) MCI's prediction accuracy vs. confidence level

The MCI's performance was evaluated while varying the confidence level on Doost's Generic Services campaign. The

$(all, recent)$ window combination was set to $(128, 8)$. The priors were set to $(\alpha, \beta) = (1, 1)$. Increasing the model's certainty level adversely affected the model's performance as shown in Table 6. Since the MCI method is based on confidence intervals, the wider the confidence intervals are the better is the adaptation of the model to varying data dynamics.

TABLE 6. The impact of the confidence level on MCI's prediction accuracy. Doost's Generic Services campaign was used in the experiment. The settings are $(all, recent) = (128, 8)$ and $(\alpha, \beta) = (1, 1)$.

Confidence level	MAE	RMSE
80%	**0.025**	**0.035**
95%	0.026	0.036

In summary, the results on MCI revealed that the model behaved well with a longer window size. The prediction error decreased till a saddle point. As was the case in the MoM model, the combination of multiple windows captured the market dynamics well. The effect was more visible when we combined a 64-days-long window for the purpose of capturing deeper trends, with a 8-days-long window for the purpose of capturing recent changes in the market. We found evidence that increasing the model certainty aggressively affected the model's performance adversely. Wider confidence intervals with lower levels of confidence performed better.

C. COMPARATIVE EVALUATION

We compared MoM and MCI with the Holt-Winters method, which forecasts the future values of a given time series [29]. Exponential smoothing refers to the use of an exponentially weighted moving average (EWMA) to smooth out the individual values of the series. The Holt-Winters method has three such EWMA based "smoothing" components: (i) one for level, i.e., a typical value in the series, (ii) one for trend over time, and (iii) one for seasonality, i.e., a cyclical repeating pattern, the series may exhibit. Hence, the method is also known as triple exponential smoothing. The model and its parameters are learnt using the sequential quadratic programming optimization algorithm called SLSQP[5] [30], which requires $O(n^2)$ space and $O(n^3)$ time. Since the time and space complexity of the optimization procedure are prohibitive for a real time application, the Holt-Winters model should be re-trained only periodically in order to amortize the cost of optimization. Hence, it becomes important to study the asymptotic performance for an arbitrarily long forecast window. When the accuracy of the model were to deteriorate below a tolerable level, the model would be rebuilt from scratch.

1) Prediction accuracy vs. size of forecast window

The performance of all three methods were measured while varying the size of forecast window. The window size determines the number of days to look ahead via forecast. It

---

[5] Modern software packages including the Python package we used in this study provide this optimizer as a default setting.
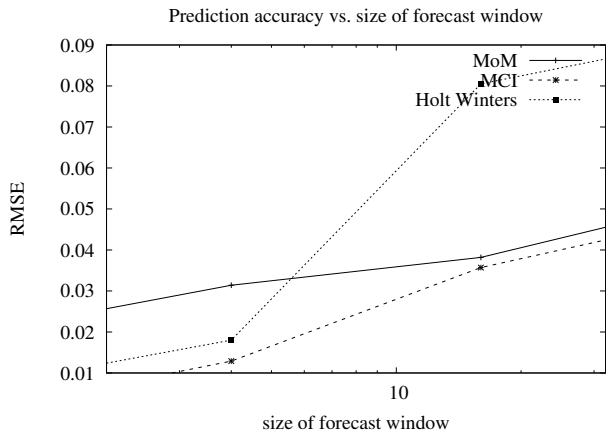
IEEE *Access*

FIGURE 8. Prediction accuracy in terms of RMSE vs. size of forecast window. Doost's Generic Services campaign was used in the experiment.
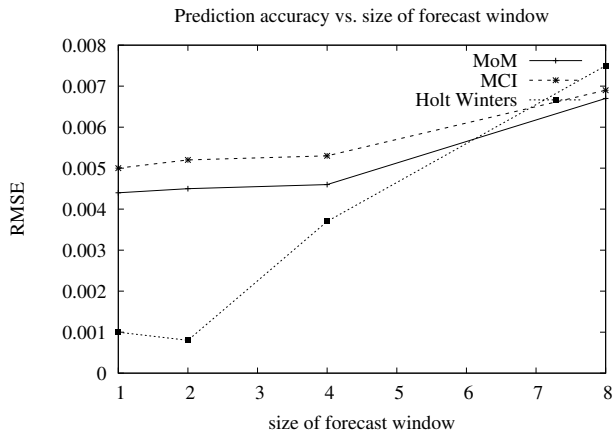

FIGURE 9. Prediction accuracy in terms of RMSE vs. size of forecast window. The third campaign of Grou.ps was used in the experiment.

varied from 1 to 64. The asymptotic performance of both MoM and MCI is better than that of Holt Winters as shown in Figure 8. The MCI method was the clear winner. A similar experiment was conducted on the third campaign of Grou.ps. Since this campaign contains weekly data and there are only 50 weeks of data available, the window size was capped at 8. The results are shown in Figure 9. Holt Winters performed better for smaller windows. For larger windows to look ahead into, MoM or MCI should be the method of choice.

V. CONCLUSION

We devised novel methods to predict the rate of online customer acquisition for real time bidding. We created two models for this purpose, i.e., MoM and MCI. The results we obtained by the application of our models on the marketing campaigns of two startups were promising. Both MoM and MCI require $O(n)$ space for n observations, and operate in constant time, i.e., $O(1)$. Since both MoM and MCI can be updated in $O(1)$ time with new data, they are suitable for data streaming applications where new data arrives continuously in an online manner.

In the near future, we plan to develop new models using other statistical measures besides median. Additionally, the merging of more confidence intervals than two needs further exploration. During our empirical evaluation, we worked with daily and weekly campaign data because of the granularity of the data available at our disposal. However, it is worthwhile to investigate how finer grain data, e.g., hourly data, could be used in real time decision making.

REFERENCES

[1] A. Bulut and A. Mahmoud, "Generating campaign ads & keywords for programmatic advertising," *IEEE Access*, vol. 11, pp. 43 557–43 565, 2023.

[2] J. A. Petersen, L. McAlister, D. J. Reibstein, R. S. Winer, V. Kumar, and G. Atkinson, "Choosing the right metrics to maximize profitability and shareholder value," *Journal of Retailing*, vol. 85, no. 1, pp. 95–111, 2009, enhancing the Retail Customer Experience. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022435908000894

[3] A. Bulut, B. Kocak, S. K. Kose, O. Sagoglu, and H. Tilbec, "An empirical analysis of the state of the art in response prediction," *Informatika*, vol. 45, pp. 27–32, 2017.

[4] M. Almasharawi and A. Bulut, "Estimating user response rate using locality sensitive hashing in search marketing," *Electronic Commerce Research*, vol. 22, no. 1, pp. 37–51, 2022. [Online]. Available: https://doi.org/10.1007/s10660-021-09472-1

[5] J. Hartman and D. Kopič, "Exploration with Model Uncertainty at Extreme Scale in Real-Time Bidding," in *Proceedings of the 16th ACM Conference on Recommender Systems*, 2022, p. 469–471.

[6] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, pp. 4–22, 1985.

[7] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine learning*, vol. 47, no. 2, pp. 235–256, 2002.

[8] J. C. Gittins, *Multi-armed Bandit Allocation Indices*. Wiley, Chichester, NY, 1989.

[9] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, no. 3/4, pp. 285–294, 1933. [Online]. Available: http://www.jstor.org/stable/2332286

[10] D. Russo, B. V. Roy, A. Kazerouni, and I. Osband, "A tutorial on thompson sampling," *CoRR*, vol. abs/1707.02038, 2017. [Online]. Available: http://arxiv.org/abs/1707.02038

[11] K. J. Ferreira, D. Simchi-Levi, and H. Wang, "Online network revenue management using thompson sampling." *Operations Research*, vol. 66, no. 6, pp. 1586–1602, 2018. [Online]. Available: http://dblp.uni-trier.de/db/journals/ior/ior66.html#FerreiraSW18

[12] D. N. Hill, H. Nassif, Y. Liu, A. Iyer, and S. V. N. Vishwanathan, "An efficient bandit algorithm for realtime multivariate optimization," *CoRR*, vol. abs/1810.09558, 2018. [Online]. Available: http://arxiv.org/abs/1810.09558

[13] E. M. Schwartz, E. T. Bradlow, and P. S. Fader, "Customer acquisition via display advertising using multi-armed bandit experiments." *Marketing Science*, vol. 36, no. 4, pp. 500–522, 2017. [Online]. Available: http://dblp.uni-trier.de/db/journals/mktsci/mktsci36.html#SchwartzBF17

[14] Y. Li, X. Yu, and N. Koudas, "Data Acquisition for Improving Machine Learning Models," *Proceedings of the VLDB Endowment*, vol. 14, no. 10, pp. 1832–1844, June 2021.

[15] A. Slivkins, "Introduction to Multi-Armed Bandits," *Foundations and Trends in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019.

[16] S. L. Scott, "Multi-armed bandit experiments in the online service economy," *Applied Stochastic Models in Business and Industry*, vol. 31, no. 1, pp. 37–45, 2015. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asmb.2104

[17] W. Shen, J. Wang, Y. Jiang, and H. Zha, "Portfolio Choices with Orthogonal Bandit Learning," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*. AAAI Press, 2015, p. 974. [Online]. Available: http://ijcai.org/Abstract/15/142

[18] X. Huo and F. Fu, "Risk-aware Multi-armed Bandit Problem with Application to Portfolio Selection," *Royal Society of Open Science*, vol. 4, no. 171377, 2017.

[19] E. Ikonomovska, S. Jafarpour, and A. Dasdan, "Real-Time Bid Prediction Using Thompson Sampling-Based Expert Selection," in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2015, p. 1869–1878.

[20] O. Chapelle and L. Li, "An Empirical Evaluation of Thompson Sampling," in *NIPS*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, Eds., 2011, pp. 2249–2257. [Online]. Available: http://dblp.uni-trier.de/db/conf/nips/nips2011.html#ChapelleL11

[21] P. Lee, *Bayesian Statistics: An Introduction*, ser. A Charles Griffin book. Oxford University Press, 1989. [Online]. Available: https://books.google.com.tr/books?id=_hXvAAAAMAAJ

[22] J. Kerman, "A closed-form approximation for the median of the beta distribution," 2011.

[23] M. Bland, *An Introduction to Medical Statistics*. Oxford University Press, 1987.

[24] M. J. Campbell and T. D. V. Swinscow, *Statistics at Square One*. Publisher Wiley & Sons, Incorporated, John, 2009.

[25] F. Gravetter and L. Wallnau, *Statistics for the Behavioral Sciences*, ser. Available Titles Aplia Series. Wadsworth, 2009. [Online]. Available: https://books.google.si/books?id=wWFmkwxSUfUC

[26] S. Dasgupta, C. Papadimitriou, and U. Vazirani, *Algorithms*, 1st ed. McGraw-Hill Education, 2006.

[27] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval Estimation for a Binomial Proportion," *Statistical Science*, vol. 16, no. 2, pp. 101 – 133, 2001.

[28] E. Sokullu, *Grou.ps Hikayesi*. İkinci Adam Yayınları, 2023.

[29] R. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. Australia: OTexts, 2018.

[30] D. Kraft, "Algorithm 733: TOMP–Fortran Modules for Optimal Control Calculations," *ACM Transactions on Mathematical Software*, vol. 20, no. 3, pp. 262–281, 1994.

AHMET BULUT received his B.S. degree in Computer Engineering from Bilkent University, Türkiye in 2000, and his PhD degree in Computer Science from University of California at Santa Barbara in 2005. He is a professor of Computer Science at Acıbadem University, Türkiye. He has 15+ years of field experience in search marketing acquired at multiple companies including Udemy and most recently Carbon Health.

SEMIH BULUT received his BS degree in Actuarial Sciences in 2008 from Marmara University and M.A. degree in International Trade Management in 2017 from Boğaziçi University, Türkiye. He is currently pursuing his PhD degree in Accounting and Finance doctoral program at Marmara University, Türkiye.

EMIN AVCI received his BS degree in Business Administration in 1998, and PhD degree in Accounting-Finance in 2006 from Marmara University, Türkiye. He is a professor of Finance in Marmara University, Türkiye.