From Bids to Bliss:
Achieving Campaign Excellence with ML and Data Insights

Robin Schneider

Dissertation written under the supervision of professor
Nuno Filipe Paiva

Dissertation submitted in partial fulfilment of requirements for the MSc in Business Analytics, at the Universidade Católica Portuguesa, 03.04.2024.

From Bids to Bliss:
Achieving Campaign Excellence with ML and Data Insights

Robin Schneider
April 03, 2024

Abstract

Telecommunication companies encounter challenges, such as regulatory requirements, fast-paced technological changes, high competition, and customer expectations. One of their biggest challenges is the efficient allocation of their marketing budget. This study delves into three modeling approaches for optimizing budget allocation across four digital marketing channels to increase total sales. Our goal is to increase sales and transform digital marketing practices in the telecommunications sector by implementing AI-powered budget allocation methods.

The three modeling approaches are a baseline (Random Forest) machine learning model leveraging the automatic creation of multivariate time-series data, as well as two Marketing Mix Models (MMM): one pipeline that implements the ideas from Meta's Robyn framework and the other using Google's Lightweight-MMM framework. We aim to gain maximum data insights from models offering diverse perspectives and approaches. We evaluate the models based on their ability to predict total sales, analyzing different performance metrics such as RMSE and R^2. Additionally, we assess the models for their practical relevance regarding their utilization for the marketing budget allocation.

One important discovery is the significant impact of Channel 0 on total sales. However, the models only capture some of the dynamics and interactions between the marketing channels. The LightweightMMM model achieved the best performance, with a $R2$ of 0.666 on unseen data. Therefore, we recommend a further development of the the MMM approaches in this study. Also, because a MMM provides detailed data insights for long-term budget allocation and captures the crucial carryover and saturation effect evident in digital marketing campaigns.

Keywords: Marketing Budget Allocation, Machine Learning, Machine Learning Algorithms, Marketing Mix Modeling (MMM), Sales Optimization

Resumo

As empresas de telecomunicações enfrentam desafios, como mudanças tecnológicas aceleradas, pressão regulatória e concorrencial além das expectativas Cliente em usufruir de melhores condições e experiência. Um dos maiores desafios é alocar com eficiência o orçamento de marketing digital para atrair novos clientes em um mercado saturado.

Este estudo investiga três abordagens de modelação para otimizar a alocação de orçamento em quatro canais de marketing digital para aumentar as vendas de produtos no segmento de Consumidor. O nosso objetivo é aumentar as vendas, mas também transformar as práticas de marketing digital no setor de telecomunicações, implementando métodos de alocação orçamental baseados em IA.

Explorámos três abordagens de modelação: um modelo de machine learning (Random Forest) com criação automática de variáveis a partir séries temporais multivariadas, um modelo machine-learning incluindo features típicas deste tipo de problemas – efeitos de saturação e carry-over e um modelo Marketing Modeling Mix (MMM) open-source da Google (Lightweight).

Avaliamos os modelos com base na capacidade de prever vendas totais com investimento em plataformas distintas – Google e Meta - analisando diferentes métricas de desempenho como RMSE e R^2.

Descobrimos o impacto significativo de um dos canais nas vendas totais bem como quantificámos as interações entre os diferentes canais de marketing.

O modelo LightweightMM obteve o melhor desempenho, com $R2$ de 0,666 em dados que não são utilizados no processo de treino. Face a estes resultados, recomendamos aprofundar ainda mais as abordagens MMM no contexto estudado.

Keywords: Alocação de orçamento de marketing, aprendizado de máquina, algoritmos de aprendizado de máquina, Marketing Mix Modeling (MMM), otimização de vendas

Contents

List of Figures

List of Tables

1 Introduction

"Half the money I spend on advertising is wasted; the trouble is I don't know which half." John Wanamaker's famous quote still holds in the digital era today. With companies generating vast amounts of data, it's essential to analyze the collected data to understand customer behavior, preferences, and engagement to maximize growth. One way to achieve this is by optimizing marketing budget allocation. But tracking and gathering data has become increasingly challenging due to regulations on tracking and the abundance of user-level data. Over the years, various approaches have been introduced, including time series analysis, machine learning, predictive analytics, and Marketing Mix Modeling (MMM). The latter is exciting as it produces valuable results based on aggregate rather than individual user data. By applying these models in a data-rich and customer-focused environment, we can find valuable solutions to complex challenges.

This thesis tackles the challenge of optimal digital marketing budget allocation in the telecommunication sector by developing and applying advanced machine learning models. The impact of marketing spend directly influences business growth and profitability. The thesis aims to unveil a strategy that maximizes leads and sales and ensures an optimal return on investment. We lay the foundation by developing a valuable machine-learning model that can be used in future work to find the optimal allocation strategy.

Our ultimate goal is to solve the problem of MMM [Pandey et al., 2021], using statistical frameworks like Robyn and LightweightMMM, as well as machine learning models. Our work aims to offer a practical solution to a common issue faced by many companies operating in the digital environment. Additionally, we incorporate existing literature while addressing a gap in the current research - implementing the actual effect of this knowledge with real-world data.

This paper introduces the research and methods in the literature review to help the reader understand the intricate challenges and solution approaches of market budget allocation. We then introduce the methodology that delves deeper into the business and data understanding and the exploratory data analysis of the available data we use for modeling. This part also includes a thorough breakdown of the two main models we develop. Finally, we compare the results of these models and recommend future work while discussing the implementation of such a model. The thesis aims to explore the prediction and explanation of sales using digital marketing data to answer the following research question:

- **What is the most effective machine learning model that lays the groundwork for determining the optimal budget allocation strategy for multi-channel digital marketing campaigns, and how can data insights and machine learning help in future budget allocation?**

2 Literature Review

We followed a comprehensive and systematic approach to identify and analyze the most relevant and high-quality sources for the literature review. Initially, we established specific criteria for including and excluding sources. We reviewed only English papers published between 2015 and 2023 and identified relevant databases such as Scopus, Google Scholar, and Arvix to search for relevant papers.

With a thorough search strategy using relevant keywords, we extensively searched the databases and gathered many papers, studies, and articles. In addition, we used the search tool ConnectedPapers to extend our research with current and similar documents. Finally, we evaluated the papers based on their abstract, introduction, methodology, and results, considering their quality.

The valuable papers underwent a thorough analysis to ensure a comprehensive, relevant, and high-quality selection of documents and articles. To conclude the search strategy, we identified the most popular computational tools for Bayesian MMM and reviewed related research.

2.1 Digital Marketing

The digital age has revolutionized the way businesses approach marketing. With more data than ever and constantly new algorithmic methods, traditional marketing methods that existed before the internet, such as print advertisements, radio, and TV commercials, are slowly being replaced by digital marketing practices. These practices have included social media marketing, email campaigns, and search engine optimization using the internet and electronic devices.

"Digital marketing is widely considered as an umbrella term, including online marketing, Internet marketing, and mobile marketing. It can be defined as marketing that utilizes digital technologies (hardware, software, communication technologies) for the deployment of marketing strategies." [Miklosik et al., 2019].

The shift to digital marketing has come with several challenges and risks, especially when using machine learning (ML) and artificial intelligence (AI) applications. Many marketing managers want to avoid the high entry cost, do not have the required know-how, and are confronted with market limitations. Still, according to a questionnaire by Miklosik et al., managers think that ML and AI are the future of digital marketing [Miklosik et al., 2019].

Additional challenges for businesses have been the influence of real-time events on digital platforms, the vast amount of data generated on all digital platforms, and the fast changes in consumer behavior. Nevertheless, the benefits of digital marketing have outweighed the risks and challenges. These methods can be cost-effective, more accessible to track, and they adjust themselves in real-time. The ability to produce and adjust digital marketing campaigns to the customers' individual preferences and behaviors and the possibility of directly interacting with consumers through digital platforms have made such campaigns increasingly attractive to marketers.

2.1.1 Digital Marketing in the Telecommunications Industry

This marketing change has also been apparent in the telecommunications industry. As customer behavior and preferences have evolved in the digital age, the industry has had to adapt accordingly. Research has shown that consumers expect relevant, helpful, and personalized marketing content that predicts and matches their needs and desires [BCG, 2018]. Telecom companies have used digital marketing strategies like SEO, PPC, social media, and content advertising to engage with customers. However, companies must find the optimal mix of different marketing approaches and gather data from these digital channels to achieve the desired results. They have also needed to collaborate with relevant agencies and marketing tech providers to gain valuable insights and make data-driven decisions. Data and analytic specialists must transform gathered insights into an actionable strategy. These specialists then need to connect customer interaction data to business outcomes like sales impact, which can be analyzed [BCG, 2018].

Despite increased growth possibilities, there have been accompanying challenges. Growing competition in the industry leads to an ongoing and more serious fight for the consumer's attention. Accompanied by an increasing price battle between the competitors, keeping customers and preventing them from churning has become more and more difficult. However, competing telecom companies are only one of the threats to telecom companies in the industry. WhatsApp, Apple's FaceTime, and Tencent's WeChat pose a significant threat to the industry. These companies decrease the traditional fixed and mobile communications services and pulling more and more customers away from traditionally operating companies. In addition to these challenges, many telecom companies have lagged behind, needing more skills and resources to modernize their infrastructure. Consumers expect telecom companies to stay ahead of technological trends, and finding out this is not the case might negatively affect the company's reputation. Data privacy and security issues have also been a concern as telecom companies generate a vast amount of sensitive customer data. Being transparent in handling this data and complying with the General Data Protection Regulation (GDPR) further enhances trust in their business [McKinsey Company, 2019].

2.2 Marketing Budget Allocation

"While marketing budget allocation has been studied for decades in traditional business, nowadays online business brings much more challenges due to the dynamic environment and complex decision-making process." [Zhao et al., 2019a]. Allocating the optimal amount of money

towards specific marketing campaigns has always been crucial for companies looking to increase their profitability. In the early days of marketing, budget allocation was mainly based on intuition, the availability of funds, and the past year's sales. The most significant portion of the budget was allocated to strategies that had historically provided the best returns [Ramaseshan, 1990]. New advertising strategies have emerged, such as Market Segmentation, where budgets are allocated based on audience target demographics [Johnson, 1971]. Traditional methods that supported decision-making have involved optimization models using mathematical programming to calculate the optimal allocation under various constraints or market segmentation. Another popular approach has been Multi-touch Attribution Models (MTA), which allows multiple ads to receive credit based on how much they contributed individually [Johnson, 1971].

With the advent of digital channels like social media and search engines, effective and optimized budget allocation strategies have become increasingly important. Attribution is a crucial issue for marketers. A simplified model in Figure 1 shows the relationship between key elements in the current online advertisement environment. Digital advertisers have used a variety of channels to reach consumers. Publishers, including websites, display these ads by selling ad spots. Real-time bidding (RTB) is how publishers and advertisers interact through an ad exchange that uses an auction model to function in real-time. Users engage with various publishers and channels, producing impressions. Conversions, or purchases, can result from a user's interaction with these ads and can be monitored using a funnel model. An attribution system tracks ad impressions, conversions, and user identification across media to determine how effective these ads are. For thorough tracking, this system frequently uses tracking pixels and browser cookies to create online profiles. Brands rely on the attribution system to evaluate the effectiveness of individual ads and the campaigns they run as a whole, which directly impacts the costs, revenues, and ad payments [Katsov, 2017].

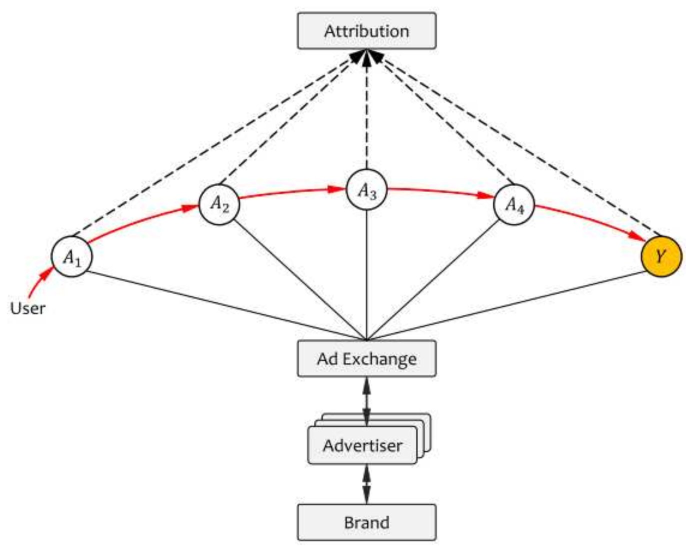


Figure 1: Simplified model of online advertising attribution.

Businesses that have allocated their budget ineffectively might miss out on opportunities and lose customers, which could ultimately lead to financial losses. Traditional and conservative decision-making in marketing has led to overspending on inefficient campaigns and underspending on successful ones. This has limited the adaptability of budget allocation [Becker, 1995]. While complying with GDPR helps gain consumers' trust, it has also posed challenges for optimal marketing budget allocation. As consumers have demanded more data protection, new restrictions on data tracking and potential data protection laws can limit what data can be tracked, making it challenging to develop an efficient budget allocation strategy. GDPR and various cookie rules have introduced new challenges in tracking customers and managing digital marketing campaigns effectively. Additionally, digital marketing providers such as Google and Meta have used walled gardens, which means that while you can get data about digital marketing campaigns, the data is aggregated and has a silo view for each operator. This has made measuring the relationship between investment and results difficult, especially when using approaches such as MTA, which requires precise user-level data to produce optimal results.

In this regard, marketing mix models (MMMs) have become attractive to marketers, as they can produce valuable insights and results from aggregated data instead of data with a high level of granularity [Harvard Business Review, 2023]. Our research focused on the challenge of using aggregated data, the high measurability of digital marketing campaigns, and how the collected data can be used to make more efficient data-driven decisions in the future. The main goal in marketing budget allocation is to balance maximizing customer acquisition and ensuring profitability. Research has shown that the use of real-time data, predictive analytics, and machine learning algorithms can provide the necessary information for optimal performance, faster decision-making, and automation of predictable activities, which will ultimately help to optimize marketing budget allocation [Miklosik et al., 2019, Zhao et al., 2019a].

Zhao et al. highlighted that by using their proposed machine learning model, daily marketing strategies in Taopiaopiao saw a sales improvement of over 6% while reducing marketing expenses by 40% [Zhao et al., 2019a]. Doordash has optimized its marketing budget allocation by building a marketing automation platform using a machine learning approach that optimally allocates its marketing budget on a campaign level while automatically publishing bids to its channel partners. By doing so, Doordash expected to reduce its marketing costs by 10 to 30 percent while reaching the same amount of customers. Furthermore, by digitally transforming their digital marketing team, employees can focus their time and energy on creative and strategic tasks, innovation, and experimentation instead of on operational functions that can be automated by machine learning [DoorDash, 2020].

Lyft has also been working on a marketing automation tool to improve its cost and volume efficiency. They have built an automated lifetime value forecaster and budget allocator, saving them time and achieving an overall higher ROI [Lyft, 2019]. We will explore different machine learning techniques focusing on MMMs that might help overcome this challenge.

2.3 Computational Approaches in Marketing Budget Allocation

The literature analyzed in the following chapter has shown various effective approaches and models for optimizing marketing budget allocation. The section will delve into the most commonly used methods and provide academic references to support each method. These papers aim to find the optimal model for their problem or discuss the challenges and opportunities.

2.3.1 Bayesian Methods in Media/Marketing Mix Modeling

Existing literature has provided valuable insights into how marketers can use Marketing Mix Models (MMMs) to measure the effectiveness of their advertising campaigns. These models have primarily relied on aggregated time-series data and regression analysis to assess the relationship between marketing activities and demand. Their ultimate goal has been to understand how advertising spending across different marketing channels impacts sales. The simplified model used in this approach can be summarized as follows:

$$\text{kpi} = \alpha + \text{trend} + \text{seasonality} + \text{media channels} + \text{other factors}$$

Where *kpi* is the total leads or sales per day or week, meaning the target KPI the model is trying to predict and further the metric that will be optimized during the optimization process. α is the model intercept that represents the baseline amount of leads or sales that the model would predict if no marketing spent or other factors were included in the model. *trend* represents the change of leads or sales over time and has a flexible and nonlinear form. *seasonality* captures possible seasonal trends and is a sinusoidal function with configurable parameters representing daily, weekly, or yearly seasonality. *media channels* is a matrix of the marketing spent per period and digital channel, whereas the media values must not contain negative values. *other factors* is a second matrix of factors that could influence our main KPI (leads or sales) without directly affecting them [Jin et al., 2017].

It is crucial to understand the impact of advertisements on consumers when considering marketing mix modeling (MMM), particularly when considering the carryover effect in the Bayesian framework [Jin et al., 2017]. Users may not convert immediately after seeing an ad on Facebook but may do so later when they search on Google. This carryover effect has often been left out by traditional models that only attributed the conversion to the ad that made it, leading to wrong decisions about marketing budgets. Jin et al. suggested that the shape effect leads to ads losing their effectiveness after spending a certain amount of money on them, so spending more on ads only sometimes brings in additional sales. The researchers used a Bayesian approach with algorithms to estimate a nonlinear model that includes prior knowledge, making decision-making more flexible. This model has been beneficial for estimating large data sets but may have produced biased results for smaller ones. Open-source packages like Lightweight, PYMC, and Robyn are available for using the model.

There are various challenges that traditional marketing mix modeling (MMM) faces, as mentioned in a study by Ng et al. (2021). One of the main challenges has been the need to update the models frequently due to the rapidly changing media and advertising landscape. In the past, marketing activities were limited to a few options, such as television advertising and promotional discounts. However, with the emergence of digital marketing, new activities such as social media marketing and search engine marketing have become popular. The authors of the study also emphasized the importance of balancing data granularity and the reliability of historical data. Increasing the granularity of the data may have made the historical data unreliable or noisy, making it difficult to predict long-term patterns. On the other hand, reducing the granularity may have led to missing essential trends. Finding the right balance between these two practises is crucial.

Moreover, traditional MMMs have been prone to correlated errors, endogeneity, multicollinearity, and self-selection bias. Ng et al. proposed a solution to these challenges using a Bayesian Time-Varying Coefficient (BTVC) model that combines Bayesian modeling and kernel regression. The BTVC model is based on the Bayesian approach, which can provide a range of possible outcomes with a probability assigned to each outcome instead of just a single prediction. This has helped marketers understand the risks associated with each marketing strategy. The model's performance has further been enhanced by integrating experimental data, another strength of the Bayesian approach. By incorporating the experiment results into their prior knowledge, they have made better and more informed predictions.

Kernel regression has been used to capture the relationship between variables that change over time, which helps understand the dynamically changing effectiveness of different marketing strategies. The BTVC model has predicted the outcome by considering several factors and breaking them down into hidden variables. The model has assigned each variable a weight based on how much it influences the outcome, offering more accurate predictions.

In their study, Chen et al. presented two models that aim to capture the various effects of advertising. The first model is a base model that only considers fixed effects. In contrast, the second model is a hierarchical model that includes fixed and random advertising effects to account for heterogeneity. A hierarchical model has taken multiple data hierarchy levels into account, which means it can factor in variations in different marketing campaigns or channels. Chen et al. also discussed the importance of using nonlinear functions with unknown variables to incorporate the carryover and shape effects. Using a Bayesian approach, the two models have learned the unknown variables to make better final predictions. Lastly, the model used sign constraints to ensure that the outcome of the model had the expected positive or negative sign. For example, by increasing the spending on a specific marketing campaign, the researchers expected increased sales. This has ensured an intuitive outcome of marketing activities, resulting in a realistic marketing mix model [Chen et al., 2021].

The insights obtained from the papers help optimize the marketing budget allocation. By comprehending the widely discussed carryover and shape effects, it is possible to make informed decisions about the budget allocation strategy and the use of our models. The Marketing

Mix Model (MMM) gives a thorough understanding of the effectiveness of campaigns over time while considering critical external factors. The Bayesian methods offer a structure to analyze the effectiveness of the Facebook and Google campaigns discussed in the papers. Furthermore, they help us understand uncertainties by providing a wide range of possible results, allowing us to comprehend the risks of each marketing strategy.

It is important to note that the models we use for advertising are designed for a specific marketing environment and may not be suitable for all channels. In our case, we will only work with digital marketing channels such as social media and search engine marketing. Since we have yet to gain prior knowledge of our marketing campaigns, it may affect our results. Moreover, traditional models rely on aggregated time-series data, which may include something other than individual campaign data required for efficient optimization. Additionally, these models need to account for the dynamic and interactive nature of digital marketing, which may lead to suboptimal results.

2.3.2 Machine Learning Models

Numerous machine learning models have been studied to assist in allocating marketing budgets. Using machine learning models is an effective approach for predicting multivariate time series. These models can identify patterns and relationships in past data to forecast future values for the outcome, such as total daily sales. Cai et al. used Reinforcement Learning (RL) to optimize marketing budget allocation. To do so, they used previously collected offline data to simulate a potential outcome. RL is a type of machine learning where an algorithm learns through trial and error by performing various actions. The algorithm is trained by receiving rewards or penalties for its actions. The aim is to find the actions that result in the most rewards over time. However, this approach has provided some challenges. Since the model has learned offline, it can only learn from historical data and not experiment with different actions as in an online setting. This might have resulted in the model needing to cover the optimal strategy. Additionally, there may have been a marketing strategy that was more popular at the time the data was collected or different consumer preferences, leading to skewed or biased results today.

Cai et al. used a game-theoretic approach to address these challenges while incorporating mechanisms to reduce possible biases from historical data. The model has learned from past marketing campaigns to find the optimal strategy and converge to the optimal budget allocation policy by using a mix of the best parts of multiple methods instead of remembering every single action. This has made it more flexible, practical, and efficient for real-world business cases, such as coming up with the optimal marketing strategy, compared to traditional RL techniques used to tackle this problem, demonstrating superior performance. By doing so, the model has handled large amounts of data and complex relationships [Cai et al., 2023].

Based on recent research, a multi-armed bandit system has addressed the challenges of marketing budget allocation. This model has framed the prediction problem as a contextual modified bandit, dynamically allocating budget across different marketing campaigns based on their performance and continually adjusting the budget according to the campaign's performance. By analyzing the relationship between the allocated budget and respective payout in labeled historical data, supervised learning was used to predict the outcome of the campaigns.

These scenarios and methodologies have aligned with current research, focusing on digital multichannel marketing campaigns and using historical data to understand consumer behavior better. By considering long-term effects, reinforcement learning has effectively captured consumer behavior, which can help tackle the attribution challenge. Here, attribution refers to understanding how much each marketing strategy contributes to the intended goal, which ultimately helps to understand the customer journey better. Han and Gabor extensively tested these methods on a large-scale marketing campaign [Han and Gabor, 2020].

Neural Networks (NN) were is another approach in machine learning. Here, the techniques imitate the functioning of the human brain. Over time, they have become incredibly effective in capturing complex relationships in data. NN models learn from the data they receive and, based on that, make predictions or classifications. Although NN models might initially make incorrect predictions, they understand and train with the data over time, improving their accuracy and allowing them to model complex relationships.

In a recent study, researchers used a semi-black-box model that combined the logit demand curve with NN to capture the relationship between sales and marketing costs without explicitly defining any relationships or patterns. Their approach was based on the rapidly changing business environment and the need for agile marketing strategies that utilize abundant data. Using the learned model, they formulated the optimization of budget allocation as an optimization problem to determine the best allocation strategy that maximizes sales while staying within budget constraints.

The researchers implemented their approach within the Alibaba Group and demonstrated its effectiveness, indicating that it could apply to other large-scale strategies [Zhao et al., 2019a]. Based on these insights, the methodology introduced by the researchers could be a possible approach to solving our budget allocation problem.

2.3.3 Optimization Models/Techniques

Researchers have proposed a model for solving the problem of allocating advertising budget, which is based on a budget-performance optimization approach [Park et al., 2022]. An optimization model is a mathematical technique used to find the best solution from numerous options available. This approach involves defining an objective to maximize or minimize a particular variable while imposing constraints to limit the possible options. The model attempts to identify the optimal budget allocation strategy that maximizes the advertising performance from various marketing channels while working within a given budget constraint. Considering the shape effect, the model has simulated expected conversions or returns generated by different advertising campaigns. The researchers found that their model suggests a more efficient budget allocation than traditional strategies.

Another approach in addressing this problem has been to merge optimization techniques from Operation Research (OR) with machine learning [Zhou et al., 2023]. The researchers developed a decision factor to bridge the gap between these two principles and simplify complex OR problems. They combined the strengths of machine learning, such as making predictions based on insights gathered from data, with the strengths of OR models, which is the optimization of decisions based on given criteria. The authors also introduced a customized loss function for direct heterogeneous causal learning, which can learn from different causes affecting marketing outcomes, resulting in unbiased estimation and accurate predictions.

Given these scenarios and results, it could be a valuable approach to consider incorporating an optimization model combined with machine learning techniques to solve the problem of marketing budget allocation discussed in this paper.

2.3.4 Attribution Modeling

The concept of attribution modeling has already been briefly discussed, including the carryover effects. Attribution modeling is a method of assigning credit to different marketing channels or touchpoints. It analyzes a customer's journey and determines how much each channel or touchpoint contributed toward achieving the desired outcome. Marketers can then evaluate which marketing strategy is most effective and adjust their marketing budget allocation accordingly [Zhao et al., 2019b].

The researchers suggested semiparametric additive models should be used instead of traditional linear models to achieve more accurate and flexible results in modeling the relationship between the attribution values. Their semiparametric model had a fixed functional form that adapts to the data it uses. Further, the additive part assumed that the relationship between the outcome and predictor variable is additive and not necessarily linear. The sum of the individual effects of each predictor was combined to form the combination of each predictor's effect. Using this model, the researchers captured the interactions between different marketing channels. This approach has ensured the results are consistently robust and scalable for large-scale multichannel marketing budget allocation optimization.

Kakalejˇc ık et al. introduced the use of Markov Chains to solve the problem of attribution modeling. In this approach, each previously mentioned channel or touchpoint has been considered as a state in the Markov chain [Kakalejčík et al., 2018]. Historical data has been used to calculate the probability of a customer moving from one state to another, such as seeing an ad on Facebook, using Google to search for the company, and clicking on the next ad. Using Markov chains, the modelers attributed more credit to a channel frequented regularly and less credit to less influential channels, thus eliminating the noise from those channels. This has helped the model understand the complex customer journey more effectively and flexibly, allowing for a holistic view of the relationships between the different marketing channels.

However, although these approaches seem very promising, we might only be able to use them if we have data on the user's conversion path. We need to find out what and how many channels the user has visited before converging. Nevertheless, we will incorporate the idea of Markov Chains in our approach.

2.4 Conclusion

Numerous studies have been conducted on both single and multiple marketing channels. There is, however, a lack of research on digital marketing campaigns, especially in the telecommunication sector. This industry poses specific challenges in understanding the effectiveness of digital marketing campaigns due to its well-established customer base and competition. Our research aims to address this gap and provide valuable solutions in this sector.

In addition, some research papers do not account for significant effects in digital marketing attribution, such as spillover effects. It is crucial to incorporate such effects into an efficient model that works in this regard. With the rapid development of digital marketing channels, it is essential to develop models that can adapt to new trends and platforms. Our research aims to create a model that considers this dynamic and can react to new developments. Finally, despite the extensive research on the relationship between marketing expenditure and sales, the actual effect of this newfound knowledge with real-world data is yet to be applied.

3 Methodology

The upcoming chapter aims to provide the reader with a clear understanding of the business case discussed in the paper. It further forms the basis of our research methodology and serves as a comprehensive guide for the subsequent analysis. Here, we will also introduce the relevant data processed and used for modeling. Moreover, we will discuss the three modeling techniques we chose to address the problem at hand, and how we evaluated the results. The structured flowchart of the process we have followed can be found in Figure 5.

Figure 2: The process flowchart of the methodology.

We followed specific guidelines to maintain scientific integrity while safeguarding the confidentiality of proprietary information at NOS. We anonymized digital channels by replacing their names with references (Channels 1-4) that don't reveal their format. This approach ensures that sensitive information remains undisclosed. Moreover, we transformed investment data to obscure the original investment information. Instead of presenting it in euros, we established one period's investment as the baseline and expressed investments for other periods relative to this baseline. These measures were taken to balance the confidentiality of business strategies with the scientific value of the thesis.

3.1 Business Understanding and Context

During the thesis, we collaborated with NOS, the largest telecommunications provider in Portugal. The company's digital marketing team allocates the marketing budget for Google and Meta campaigns to generate more leads and sales. This budget allocation is crucial for the company's growth and profitability. Hence, a balanced budget allocation is essential to maintain sales volume and ensure profitability. We aim to optimize marketing campaigns using data insights and historical performance data generated by the campaigns on their respective channels. We strive to maximize return on investment through this data-driven approach.

The digital marketing team manages multiple digital campaigns by adjusting budgets based on intuition and historical sales. Successful campaigns receive more budget, while less successful ones are deactivated. However, this manual approach, combined with the complexity of managing multiple campaigns and segments, leads to suboptimal results, especially when it comes to optimizing. The team needs to decide which channel is most worth investing in, requiring a differentiated view of the Google and Meta campaigns' effectiveness in generating leads and sales or conversions. However, this poses a challenge, especially for the Meta campaigns, because finding the actual value of the conversions for these digital marketing channels takes a lot of work and effort.

Previous research has suggested that machine learning and artificial intelligence can help find a more valuable solution to this problem. Our approach is to develop a machine learning model that automatically proposes the optimal budget allocation per digital marketing channel. This model would evaluate and optimize campaign performance using a decision-process model. The models should be able to analyze current campaign data and consider seasonality and trends to achieve a successful strategy in the long term.

Another important aspect is developing a specialized tool for the digital marketing team. This tool would support precise daily budget allocation while enabling the team to make well-founded decisions based on current and historical data. These decisions should result in higher leads and sales than the previous allocation approach.

3.2 Data Understanding and Preparation

For our marketing campaign analysis, we received two sets of data. Firstly, NOS provided us with marketing data from Google and Meta campaigns. This data covers a specific period and includes performance KPIs such as impressions, link clicks, and the budget amount. The data from Meta is daily data per campaign, while the Google data covers campaign details by day and hour.

Secondly, we received a telephony file that includes all the contacts from the call center associated with the digital marketing campaigns. The data came from a contact form that can be filled out after a potential lead clicks on the advert link. This data was used to determine the number of calls/leads and sales made each day. This data is essential as it provides a global view of the results rather than the partial view each provider reports on their platform.

We excluded contacts that did not come from the web as they did not originate from the digital channels that we are trying to analyze and optimize. We processed the raw data to ensure it was suitable for machine learning models. This involved cleaning, normalizing, transforming, and scaling the data. We then performed feature engineering to add variables and improve the model's accuracy. We prepared the data set by including performance KPIs such as Click-Through-Rate (CTR) and Cost per Click (CPC). We tested the impact of these new variables by running a regression with and without them, ultimately keeping them in the model, as they produced lower Root Mean Square Error (RMSE) scores. The RMSE is a commonly used metric to evaluate the accuracy of statistical and machine-learning models. Using it, we compared and evaluated our models on the same scale as the original data, allowing us to understand how well the model applies to real-world data.In doing so, we are sure that we created a reliable and relevant dataset that can be used train valuable prediction models. The data dictionary of the grouping variables can be seen in Table 1.

Table 1: Data dictionary of the grouping variables.

Variable	Description
total_sales	Total sales revenue for the specified period.
cost_channel_X	Marketing campaign spend for each channel.
impressions_channel_X	Number of impressions for each channel's campaigns.
clicks_channel_X	Number of clicks received on ads for each channel.
CPC_channel_X	Cost Per Click for each channel, indicating the cost of an individual click on an ad.
CTR_channel_X	Click-Through Rate for each channel, representing the percentage of impressions that resulted in a click.

3.3 Exploratory Data Analysis

The final dataset used for modeling consists of daily data, including performance metrics per marketing channel, as well as total leads and total sales per day. It included 355 rows, representing 355 consecutive days, covering nearly one year. The average daily marketing spend across all channels was 433€, and the average total daily sales was approximately 33.

In Figure 3, we can see the trend of the primary independent variables per day over time, which are the costs and the performance metrics (clicks and impressions) per channel. The expenses show a relatively constant development with a few irregularities for all channels. Channel 2 receives the most investment, with an upward trend towards the end of the year, and also has the most fluctuations and spikes, indicating it has the most volatile spending. This could be due to more intense marketing activities in this channel and seasonal budget increases. Channel 0 and 1 campaigns have fewer and less volatile costs, indicating a more constant allocation strategy. Campaigns from Channel 3 receive the lowest investment, which could be attributed to a lower focus and lower efficiency expectations. The spikes in the costs do not automatically lead to higher clicks and impressions, which might indicate an inefficiency in the budget allocation. Only Channel 3 shows a more abnormal trend regarding their clicks and impressions, with many spikes, volatility, and generally higher amounts of clicks on average, although they receive the lowest investment. This could be due to specific campaigns that lead to higher user engagement or inconsistent budget allocation. Channel 2 has the second-highest number of clicks but the lowest number of impressions, which is unusual, since it receives the most investment. This suggests that saturation may have set in, meaning there is no significant increase in impressions despite investing more money. Channels 0 and 1 show a relatively constant development in clicks and impressions.

The figure labeled Figure 4 shows the seasonal additive decomposition of total sales, which displays the trend, seasonality, and residuals of the time series data. This decomposition helps comprehend how these components contribute to the overall patterns in the sales data. The total sales show considerable fluctuations, and the discrepancies between the costs and performance

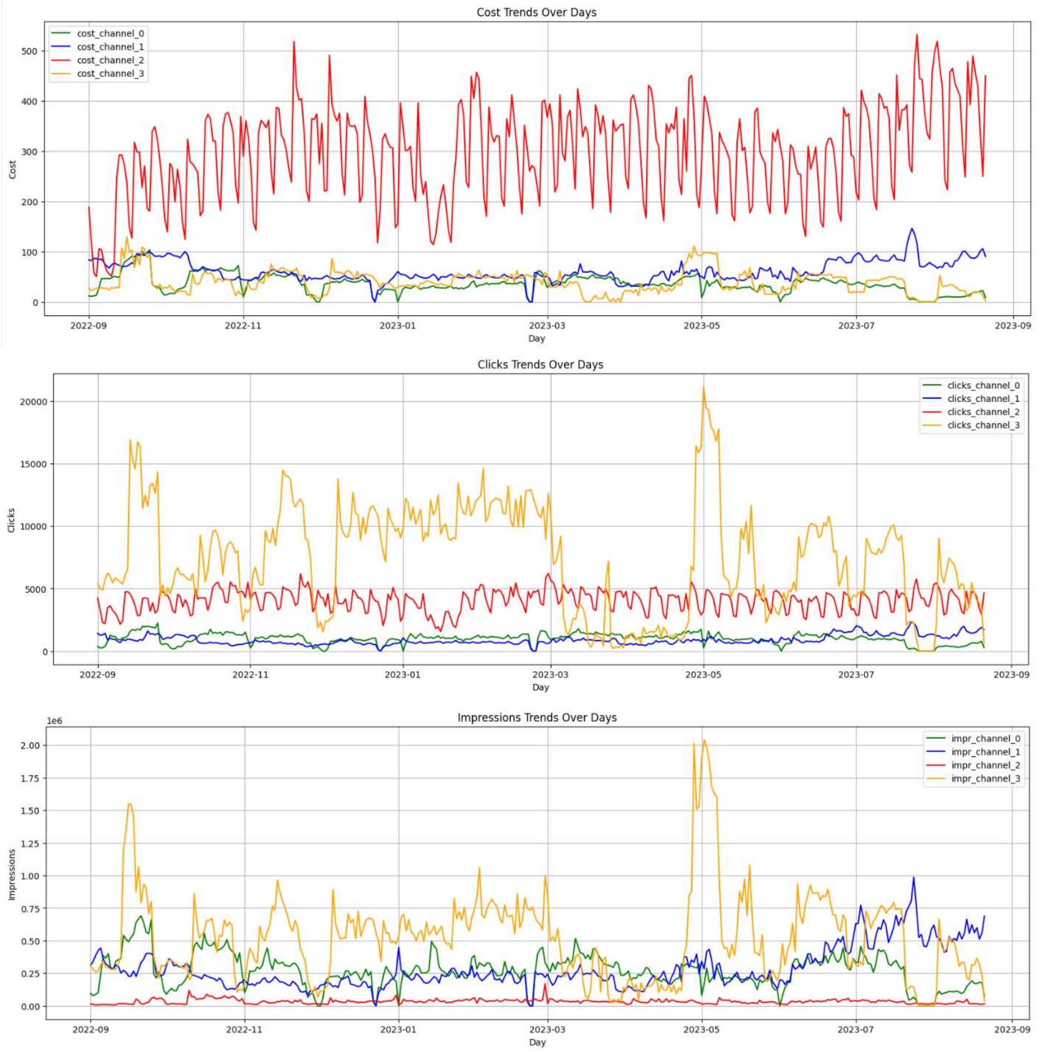


Figure 3: Trend of the main independent variables per day over time.

metrics shown in ”Figure 3” undermine the importance of valuable modeling to capture underlying patterns and adjust the marketing budget accordingly. We observe a relatively constant trend with no significant downward or upward spirals. A seasonality is visible in the data, and the residuals imply that there might be some discrepancies in the model.

We conducted a correlation analysis to gain further insights into the relationship between total sales and spending in each channel. Our findings showed that Channel 2 had the strongest positive correlation at 0.6331, indicating that an increase in spending in this channel leads to the highest growth in sales. Channel 1 followed with a correlation coefficient of 0.0428. Channel 0 and 3 had weak negative correlations of -0.0195 and -0.0091, respectively. While negative correlations are not typical, we have thoroughly checked and cleaned our data and are confident that data quality issues do not distort these results. Market saturation or ineffective targeting strategies may have contributed to these findings, or external factors that influenced sales were not accounted for in our analysis.

We shifted the sales data by one day and one week and computed correlations with spend.

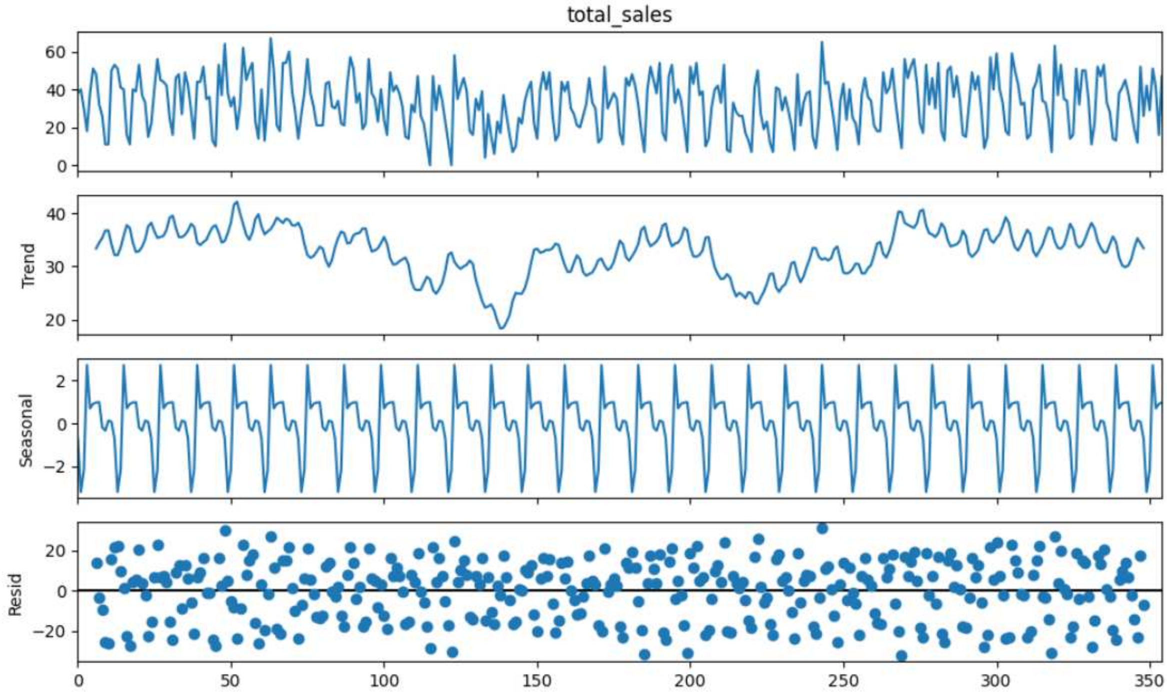


Figure 4: Seasonal decomposition into trend, seasonality, and residual components.

We found similar results after analyzing the lagged correlations between spend and sales. Channel 2 had the strongest positive correlations, with one-day and one-week lag. This undermines the importance of delayed time effects in marketing spend on sales and the need for a more rigorous adstock transformation. The low and negative correlations in our lag analysis show that evaluating and adjusting marketing activities is essential.

3.4 Modeling

We chose three different modeling approaches to solve the marketing budget allocation problem in more detail. All of the models can use the daily dataset that had been prepared previously. This diversity of modeling approaches enabled us to consider various aspects of the problem from different angles, such as detecting patterns like seasonality or trends and analyzing short-term fluctuations. By comparing and validating each model's results, we increased our predictions' accuracy and gained valuable insights for making any necessary adjustments. Each model's unique strengths and specializations contributed to developing a more precise strategy for optimizing our budget while minimizing risks and potential challenges arising from data or model limitations or limited assumptions. Ultimately, our goal was to gain a deep understanding of the complex relationship between marketing spend, customer behavior, and marketing success. By doing so, we made informed and precise decisions that maximize the companies' return on investment.

3.4.1 Baseline Model

We began by setting up a baseline machine learning model to analyze the relationship between marketing expenses and target variables. To ensure reproducibility, we established a random seed and set up logging before beginning modeling. We prepared the data set for predictive modeling and analysis for future time steps by including the features for the prediction day alongside the original data. This allowed us to make better predictions based on current and past data. We added numerical time features to extract valuable "Day" column information to capture time cycles, trends, or seasonal effects. We tried different combinations of time variables, including the day of the week, month, and year, to see which produced the best results. The approach that produced the lowest RMSE was the inclusion of all time variables. We then one-hot encoded the new variables to make them suitable for modeling.

Next, we created a temporal-dependent holdout data set to ensure that our model was robust and could generalize well. This set aside 20% of the data to evaluate how well the model performed on unseen data. By doing so, we developed a model that effectively predicted total sales and avoided overfitting, which occurs when a model is too closely fitted to the training data and doesn't generalize well to new data.

We defined some functions crucial in preparing and optimizing a prediction model to maximize its accuracy in predicting future outcomes. The first function is a feature extractor, which identifies the names of the numeric and categorical features in the dataset. This is essential in machine learning as it helps differentiate different data types. The second function is an Optuna objective function used for hyperparameter tuning. It helps find the model's best hyperparameters and precisely predicts future outcomes. A low RMSE indicates that the model is accurate.

We also defined the models we want to evaluate and the parameter ranges used in the Optuna hyperparameter search. These models are Linear Regression, Ridge, Lasso, Random Forest, Gradient Boosting, and XGBoost, each with strengths and limitations. The hyperparameters we defined are adjustable settings that influence the model's performance. During the search for the optimal hyperparameters, the grid search automatically tested different combinations of these parameters to find the combination that delivers the best model performance and the highest accuracy in predicting the future outcome.

We then defined a model pipeline for preprocessing the trained data. This pipeline included certain pre-processors as well as numerical and categorical transformers. Numerical transformations included normalizing numeric values by subtracting and dividing the mean by the standard deviation. By doing so, we can be sure that the values in the data are on a similar scale, increasing the model's convergence. The categorical features used a OneHotEncoder that transforms these values into a format that the machine learning model can use by creating a new binary column for every category per feature. Following these steps is crucial to ensure accurate results.

Next, we ran an Optuna optimizing study to find the optimal hyperparameter configuration that best predicted the future outcome and maximized the performance of our model. We based this on the data with one-day lagged features as they help us predict the outcome relatively in real-time. We ran 100 trials, and the study tested 100 configurations to find the best one. Afterward, we used the results of the five best trials based on the lowest RMSE. Based on the Optuna Hyperparameter study and cross-validation, we produced, trained, and evaluated the best model and accessed the best model's parameters and results.

Lastly, we added more features to our dataset using the TSFresh library. This library extracts various features from time-series data, including trend and seasonality information and statistical properties. These features helped to capture complex relationships and patterns that may not be identifiable through traditional methods. After adding these features, we trained and evaluated this model using the holdout dataset we created in the baseline modeling approach.

3.4.2 Marketing Mix Modeling (MMM)

We started with a non-informed machine learning model that needed to become familiar with the MMM ideas. Then, we integrated additional features that considered the carryover and saturation effect while keeping the ML approach. For this purpose, we used a specially developed open-source MMM framework based on Meta's MMM framework, Robyn [Kisilevich, 2023]. We aimed to enhance the model's functions and improve its performance by predicting uncertainties and injecting prior marketing knowledge through Bayesian modelling.

MMM helps capture the typical effects of marketing activities that a standard regression model may be unable to do. To do this, we incorporated three main ideas from MMM. Firstly, the coefficients of the marketing channels had to be favorable to include the positive effect of the marketing activity. Secondly, we applied an adstock transformation to capture the carryover effect of marketing activities. Lastly, we used the saturation transformation to account for the shape effect and diminishing returns.

To perform these transformations, we needed to know the initial adstock effect. The decay rate measures the effectiveness of the money spent on the marketing channel while declining over time. A higher decay rate (closer to 1) means that the advertisement's impact lasts longer. Conversely, a lower decay rate means the ad effect diminishes faster. For instance, a value of 0.3 means that 30% of the ad's impact from one period carries over to the next.

To obtain this value for each digital marketing channel, we ran an Optuna optimization study to find the optimal hyperparameters of the adstock and decay rate based on industry standards, to find the set of parameters that results in the lowest RMSE. We effectively search for the optimal parameters that could predict the total sales most accurately. These parameters were later refined using the Optuna hyperparameter optimization framework to then fit our data most accurately and built a valuable MMM.

3.4.3 Lightweight MMM

The LightweightMMM package, developed by Google engineers, is based on a paper called "Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects." We reviewed this paper in our literature review [Duque, 2023]. This package can be used to build a Bayesian MMM model that can be scaled to meet our specific needs. It helps us evaluate models, optimize marketing budget allocations per channel, and plot useful graphs for data visualization.

The model can be run at two levels. The most common approach is at the national level, where national data is aggregated to show leads or sales per period. The second level is the geo level or the subnational hierarchical approach, which uses KPIs such as sales per time period for each state in a country. This approach can provide more precise and accurate results as it has more data points. However, we used the standard approach since we only have national-level metrics.

As previously discussed, marketing campaigns have the most significant impact immediately after they occur. However, their effectiveness gradually reduces over time. To address this issue, the LightweightMMM model employs three different approaches. Firstly, it uses adstocks to apply an infinite lag on the KPI with a decreasing weight over time. Secondly, the model uses hill-adstock that applies a sigmoid-like function for diminishing returns to the adstock function's output. This function states that as the marketing spend in one channel increases, the sales generated after a certain point will not increase if other variables remain constant. Lastly, the model implements the carryover effect, which gives more weight to recent KPI values than more distant ones, as previously discussed. You can refer to the flowchart in Figure 5 below for implementing the LightweightMMM model. We will discuss the steps in more detail and explain how we implemented them.

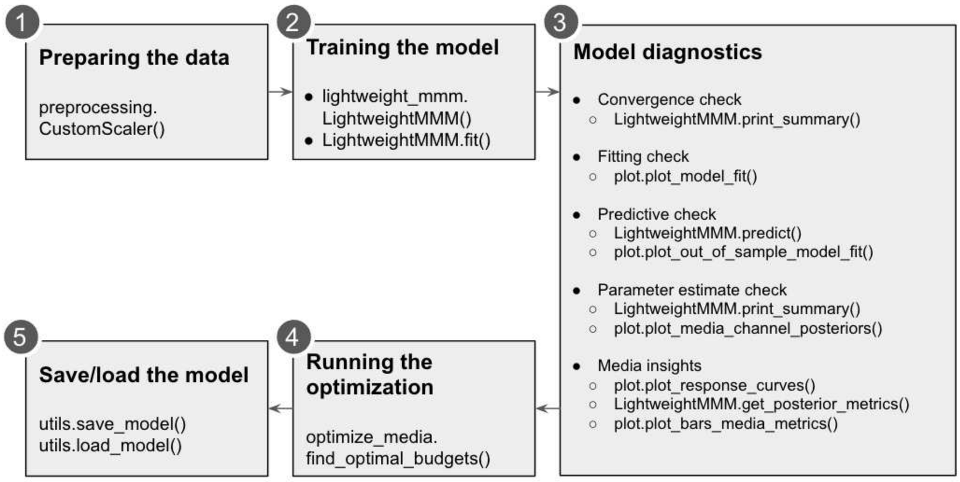


Figure 5: LightweightMMM implementation flowchart.

Our process began with a carefully curated, preprocessed data set optimized for LightweightMMM. The framework only uses the impressions, costs and total sales of our dataset. Scaling the data is crucial in modeling, as it improves performance and produces more accurate media effect estimates. We accomplished this by using predefined scaler objects within the LightweightMMM package to normalize our media data, costs, and target variables around a mean of 1. This approach facilitates comparing different data types and periods, ultimately leading to more excellent modeling performance and stability.

To account for the lagged advertising effect that decays slowly over time, the LightweightMMM model offers three different approaches, each with unique benefits. We tested all three and selected the one with the best out-of-sample fit. The methods involve applying an infinite lag that decreases in weight over time (adstock), a sigmoid-like function that diminishes returns (hill-adstock), and finally, a causal convolution that gives more weight to recent values (carryover). We then initialized these three approaches with their respective transformations and trained the model on the provided data to find the model approach with the best fit. To ensure that our model was performing as expected, we analyzed the R^2 on the training and holdout dataset and investigated several media plots.

4 Results

This section will thoroughly evaluate each modeling approach by assessing relevant performance metrics. We will also illuminate individual results' implications for our research and practical use case. To guarantee a valuable evaluation of the modeling approaches, we developed an evaluation framework that includes the individual variables' performance characteristics and influence. We paid particular attention to the impact of marketing spend on overall sales. We used a holdout dataset to evaluate the models, which enables a realistic assessment of model performance. In addition, specific evaluation metrics were used to quantify the accuracy and relevance of the various features of the prediction.

4.1 Baseline Model

After conducting the Optuna hyperparameter optimization process for various regression models, we determined the best hyperparameters for each model. The optimization history plot in Figure 6 illustrates the performance of the models over 100 trials, having the classifier and parameters that had the best RMSE per trial. We observed that the most significant performance improvements occurred at the beginning of the process, indicating the effectiveness of the study in identifying the optimal hyperparameter configuration as soon as possible. Furthermore, the ongoing stabilization of the optimization process suggests that the exploration was efficient.

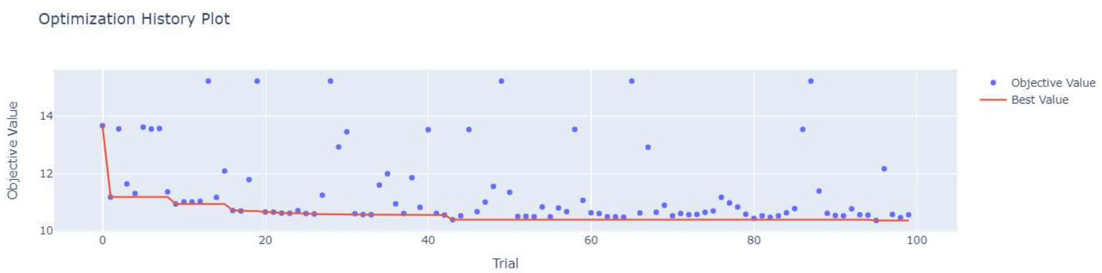


Figure 6: Optuna optimization history plot.

During the optimization process, Figure 7 illustrates the distribution of the RMSE for all trials per classifier. On average, the Random Forest model had the best values, indicating a more optimal fit to the data. All top 5 trials, ranked by their objective value (RMSE), use the Random Forest model. Of the 100 trials, 65 used this classifier, indicating that it performs the best on the given data. The highest value achieved in the Optuna study was approximately 10.3597.

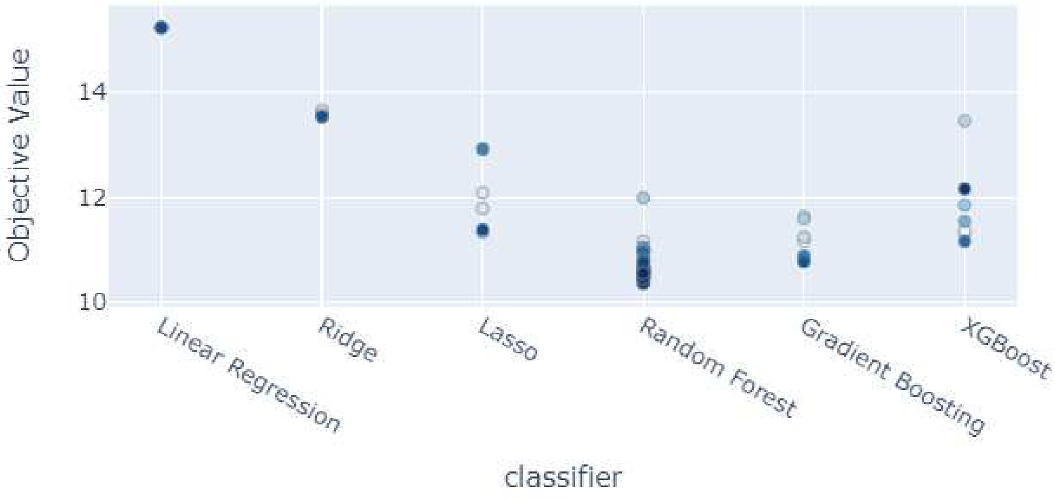


Figure 7: Distribution of RMSE per classifier.

After analyzing the holdout data, the performance of the chosen optimal model and hyper-parameters in predicting the sales against the actual total sales is depicted in Figure 8. The model achieved an RMSE of approximately 9.2395 on the holdout dataset. Although the model showed seasonality and trend patterns in sales, it failed to capture all the spikes accurately.

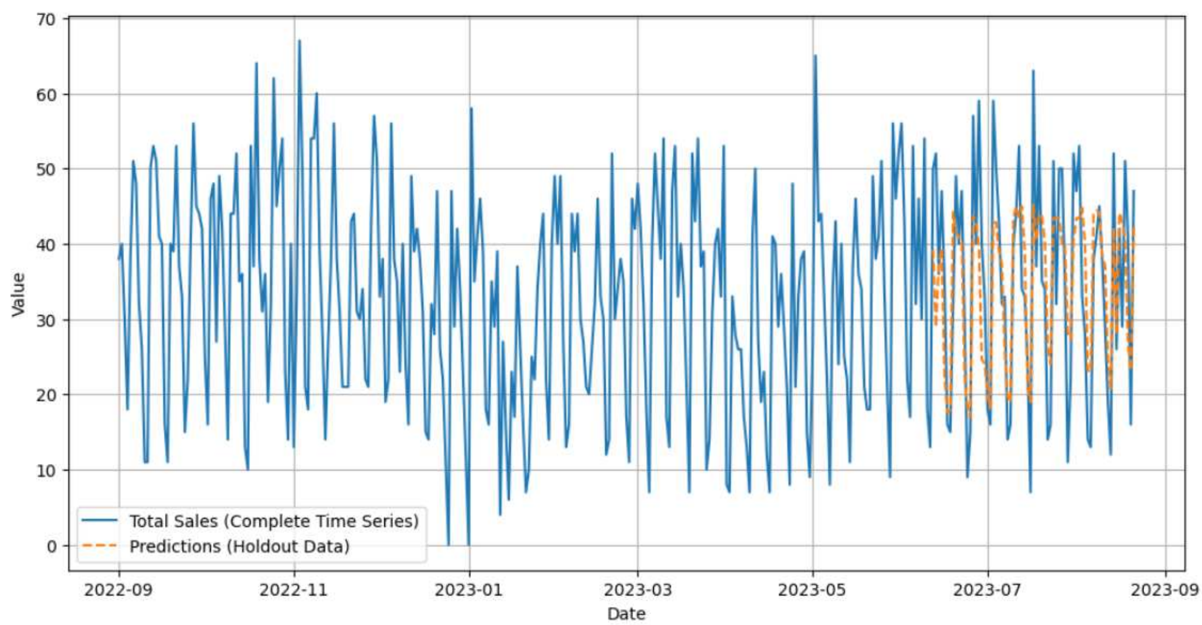
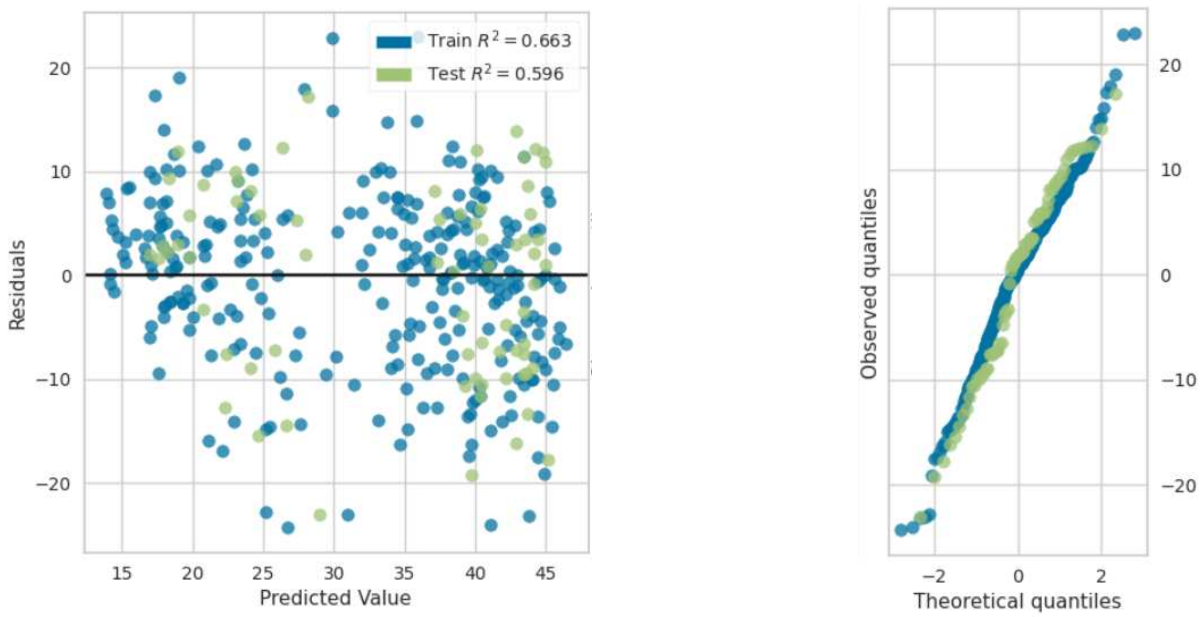


Figure 8: One-time step sales prediction.

The residual plot in Figure 9 helps us understand the variance and possible patterns in our prediction model. It visualizes the difference between the observed and predicted values. If the residuals are randomly scattered around 0 without any identifiable pattern, we can be sure that the model accurately captures the variance and is reliable for the whole dataset. The R^2 values in the plot further indicate how well the model fits the data, with a perfect fit having a value of 1. Our findings revealed that the R^2 for the training data was 0.663, while the R^2 for the testing data was 0.596. This suggests that the model captured a moderate amount of the variance in the data. However, the R^2 for the training data was only marginally better than the one for the testing data, indicating some overfitting.

The Q-Q plot reinforces this assumption, showing that the residuals were relatively normally distributed, with the points clustering closely along the line. This is another indication of the model's fitness.



(a) Residual plot (b) Q-Q plot

Figure 9: Residual plot and Q-Q plot for Random Forest Model.

Our analysis relies on utilizing a tool called SHAP value (SHapley Additive exPlanations), which helps determine the individual feature's contribution weight. This method makes the models more transparent and understandable, helping to clarify the decision-making process. By using SHAP values, we can understand how the variables interact with each other and impact the model's predictions. The SHAP values for each feature are presented in Figure 10. Since our research primarily focused on the influence of marketing costs on total sales, we concentrated on these variables. Our analysis shows that 'cost_channel_2' had a significantly higher impact on the outcome and the highest influence of all the different channel costs. Therefore, we focused on this variable in our ongoing research.

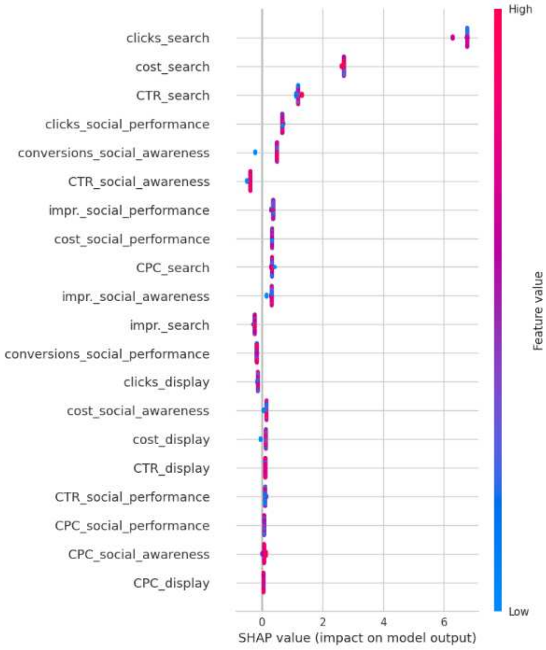


Figure 10: SHAP values for data variables.

Figure 11 displays a partial dependence plot that shows the impact of Channel 2 spending on predicted total sales. We can see that sales increase significantly if the spending reaches a threshold of approximately 270€. Beyond this point, investing more money into Channel 2 does not increase substantially total sales. This suggests that the channel has reached a saturation point.

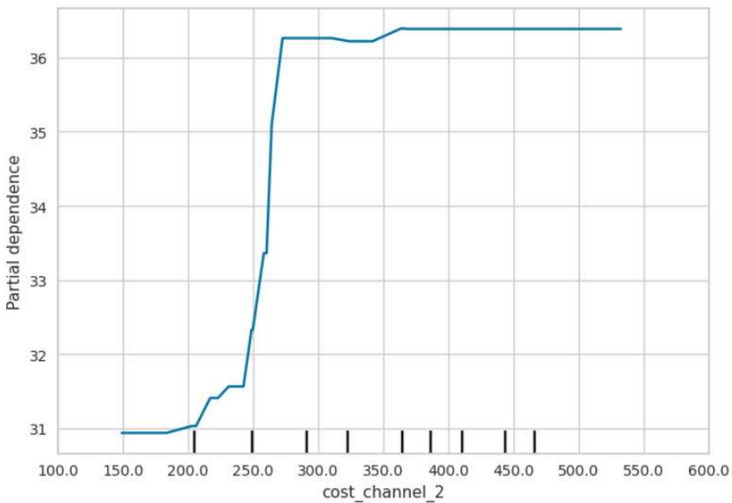


Figure 11: Partial dependence plot for Channel 2.

We added TSFresh features to our dataset and ran an Optuna optimization study to determine if these features improved our results. The best model was a Random Forest Model with an RMSE of 9.2395, the same as without the TSFresh variables. This is due to the integrated feature selection in the model architecture. The algorithm internally determines which features are most important for improving the accuracy of the prediction. Thus, we can assume that the original data already contains strong predictors of the target variable and consider the additional TSFresh features as less important or may even ignore them altogether. As a result, adding the additional variables did not impact the outcome.

4.2 Marketing Mix Modeling (MMM)

Before starting the modeling approach, we analyzed the trend and seasonality. Figure 12 shows that from September 2022 to January 2023, there is a slight downward trend, followed by a short stabilization and upward trend. Additionally, we observed a weekly seasonality pattern where sales are highest at the beginning and decrease over the week.

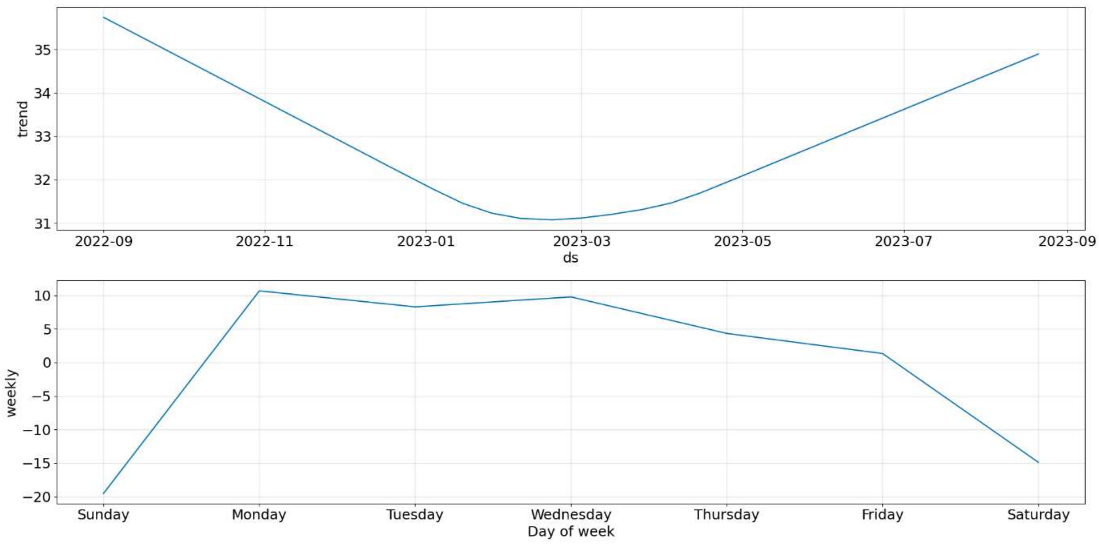


Figure 12: Trend and seasonality for MMM.

In Figure 13, we can see the coefficients of the marketing channels. The blue line represents Channel 0, the purple line Channel 1, the blue line Channel 2, and the red line Channel 3. As previously stated, we expected these coefficients to be favorable and reflect the positive impact of marketing activities. All marketing channels showed a positive effect, except Channel 3. This suggests that no matter how much money is invested here, it fails to produce positive results or sales. This may be due to inconsistencies in the data or modeling approach, which we have thoroughly checked, or it may produce adverse outcomes.

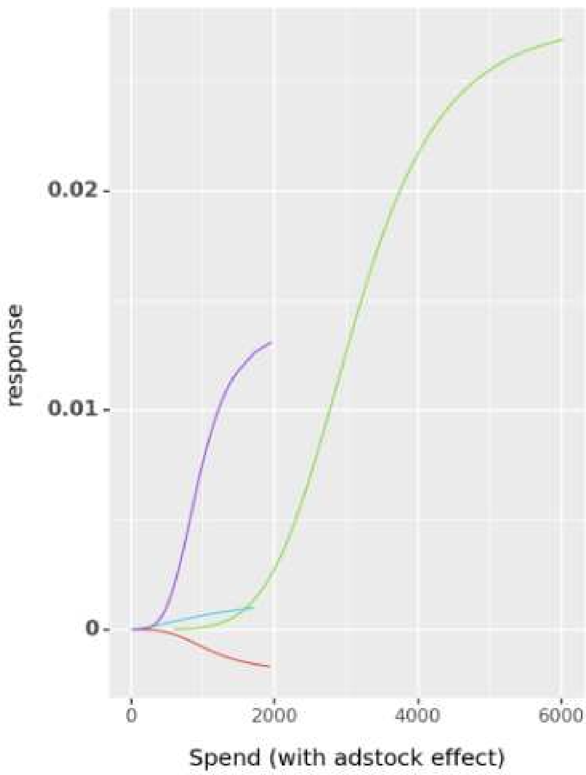


Figure 13: Response curves of all marketing channels.

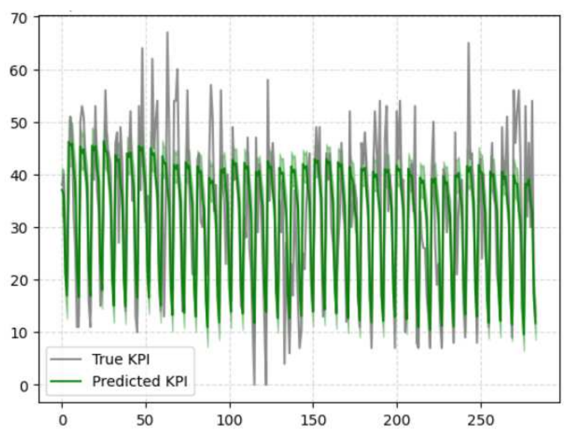
We obtained the optimal set of values for the adstock and decay rate by running the Optuna optimization study. These parameters for the adstock transformation where then refined with a second optimization study that applied the respective transformations with optimal values to calculate the carryover effect of various marketing activities. This helped us determine the adstock effects for different costs, which were: cost_channel_0 = 0.4107, cost_channel_1 = 0.2881, cost_channel_2 = 0.1373, and cost_channel_3 = 0.4521. These values indicate that Channel 3 has the longest-lasting advertising impact, while Channel 2 has the shortest duration despite having the highest coefficient, which signifies the most positive effect on the outcome. We then incorporated the saturation transformation to account for the shape effect and diminishing returns. The combination of these transformations allowed us to generate a nuanced model with an RMSE (root mean square error) of 9.114 and an R^2 value of 0.601.

While implementing the MMM approach, we faced a challenge in cross-validation using the same holdout dataset we used to compare all models. As the framework and architecture of the model with its intricate functions were predefined, we could not apply the necessary cross-validation. Hence, we cannot accurately compare the models with each other. However, the model produced valuable results with moderate performance metrics, providing us with more insights into the data.

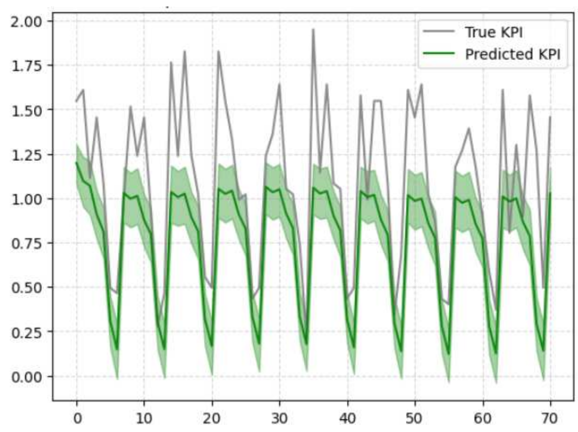
4.3 Lightweight MMM

We initialized three LightweightMMM methods with their respective transformations and trained each model using the given data. Based on our previous experience with MMM, we know a weekly seasonality is present, so we set the seasonality frequency to 52. We tried different hyperparameters, such as the number of warmups and samples and the number of chains, to improve the performance of the Markov Chain Monte Carlo method. Increasing the number of chains can help improve the model's reliability by checking for convergence across chains. We also experimented with different degrees of seasonality to check if capturing more nuanced seasonal patterns improves the model performance. After experimenting, we found that the carryover approach with 1000 warmups and samples, one chain, seasonality degree of one, and weekday seasonality performed the best.

The model combination we used had an R^2 value of 0.666 when we ran one-time step predictions on a holdout dataset that we also used for the baseline model. We have visualized the model's performance in Figure 14, where we can see the predicted and actual key performance indicator (KPI) representing sales. Plot 14a displays how the model fits on the training dataset, while Plot 14b visualizes the fit on unseen data. The y-axis on Plot 14b shows the scaled values of sales after applying a scaler that divides the mean of the target variable. This type of scaling can help normalize the data for modeling and get an accurate media effect estimate. The green area around the predicted KPI shows the confidence interval, representing the model's uncertainty. Although there were some discrepancies between the actual and predicted KPI values, we can see that the predicted KPI values generally followed the path of the actual KPI values. This indicates a moderate model performance. The deviation between the actual and predicted KPI values indicates where the model could be improved.



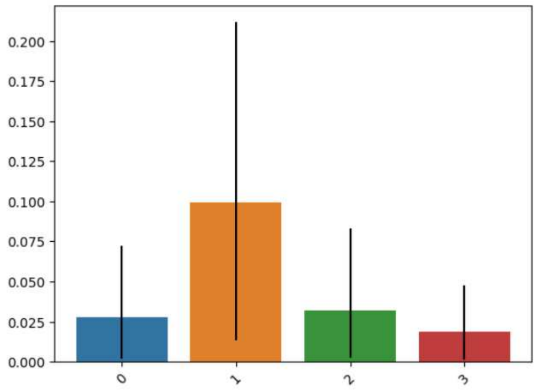
(a) Predicted and true KPI on training dataset



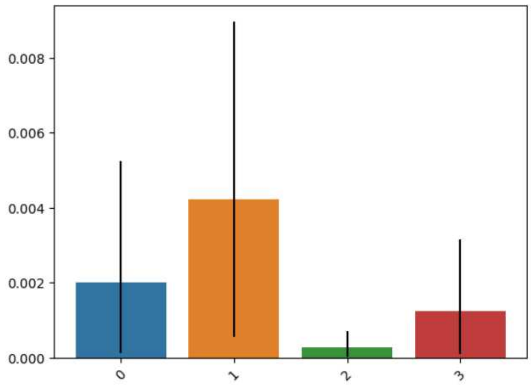
(b) Predicted and true KPI on holdout dataset

Figure 14: Predicted and true key performance indicator (KPI) on training and holdout dataset.

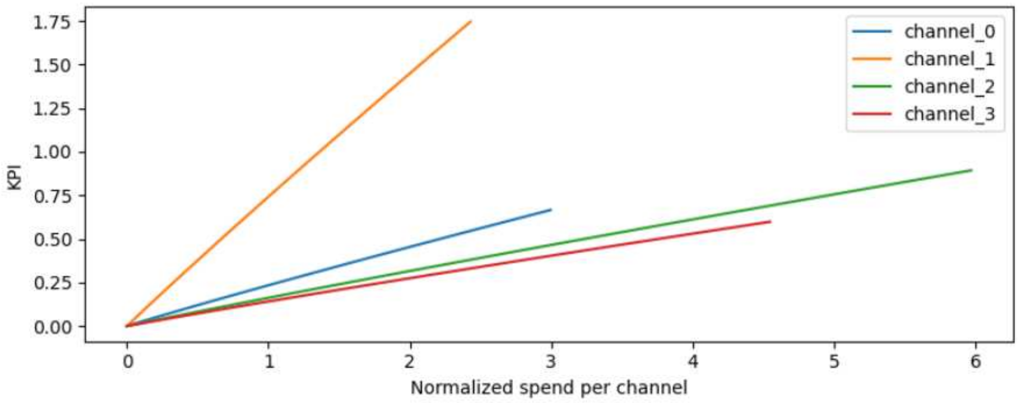
The LightweightMMM media findings are visualized through three plots in Figure 15. Plot 15a shows the estimated median contribution per channel, where Channel 1 had the highest impact on sales, while Channel 3 had the lowest. Channel 0 and 1 were in-between with similar results. However, Channel 1 had the longest error bar or confidence interval. This means there is higher uncertainty for this channel, requiring further analysis to understand their impact. The Plot 15a visualize the estimated median ROI. It reveals that Channel 1 had again the highest value with high uncertainty, and Channel 2 had the the lowest value. Generally, choosing a channel with a higher ROI makes more sense. However, selecting a channel with a lower ROI but a smaller confidence interval may be reasonable since there is less uncertainty. Finally, plot 15b represents the relationship between normalized spend per channel and the KPI's performance for each channel. The steeper the line, the more additional investment in the channel leads to a disproportionate increase in total sales. All channels showed a continuous increase, indicating that more spending consistently generates more sales growth. Channel 1 had the steepest increase, making it the most efficient channel. Based on these findings, Channel 1 outperformed the other channels, but the results should be interpreted cautiously due to the high uncertainty.



(a) Media contribution



(b) ROI analysis



(c) Normalized spend

Figure 15: LightweightMMM media findings.

5 Discussion

We conducted a study on three modeling approaches, including a baseline machine learning model (Random Forest) that utilizes the automatic creation of multivariate time-series data. We also used two Marketing Mix Models: one strategy implementing Meta's Robyn framework and another using Google's Lightweight-MMM framework. Our findings show that the LightweightMMM model delivered the best performance compared to the other two models. It achieved an R^2 value of 0.666 on the holdout dataset. Meanwhile, the Random Forest model obtained an R^2 value of 0.596 to predict on unseen data, and the other MMM achieved an R^2 value of 0.601. We must, however, interpret the latter result cautiously since we used a different holdout dataset to the other two models. We were unable to include it in the model architecture. This is because it had predefined functions and packages which where too complicable to alter in time. Despite this, we were able to analyze marketing spending more efficiently by implementing the adstock and saturation effect. This allowed us to understand how the spending impacted sales over time. The LightweightMMM model also showed greater efficiency in analyzing different marketing channels and their impact on total sales. Nonetheless, all three models produced only moderate results, indicating that there is still room for improvement.

Overall, we found that the marketing Channel 2 had the highest impact on total sales. However, this had already been apparent as it had received the most investment. Campaigns from Channel 0 and 1 had a longer-lasting after-effect, meaning there is a more extended possibility for the advertisement to take effect. The fluctuations, uncertainties, and differences in the developed models show that even though they produced valuable results, they cannot capture all the dynamics and interactions in digital marketing. Furthermore, the difference in the training and testing data's performance indicates possible overfitting, leading to reduced model generability.

The results are relevant to marketing budget allocation, especially in the telecom industry. Although the models are not fully finalized, they have produced valuable insights for the corresponding channels and satisfactory results for the time being, that can be used to optimize future marketing investments. It is also a good starting point for future work that would improve these models to develop an actual recommendation for the marketing budget allocation at NOS.

A MMM is the most suitable model for further development. While the baseline model provides a better understanding of data and potential connections, the MMM offers more detailed insights and better results, which are necessary for a long-term efficient budget allocation strategy. The MMM achieves this by incorporating the crucially important adstock and saturation effect, which we now know is evident in digital marketing activities. Although the LightweightMMM offers a fast and easy solution, its relatively low performance might prevent it from capturing the complex intricacies of the digital marketing environment at NOS. Nevertheless, the LightweightMMM and the first MMM should be further explored to determine the most valuable ongoing marketing strategy.

6 Limitations

We faced a few limitations while attempting to solve the budget allocation problem. A significant challenge in this process was imperfect data, as the data available from each digital provider is only aggregated at a high level, showing key performance indicators (KPIs) for a specific timeframe or per campaign instead of individual user-level interactions. This limited our ability to perform detailed analysis and customization. Furthermore, as we received data from different providers, such as Google and Meta, it could not be easily compared or combined due to different data formats and metrics. These metrics come with individual embedded assumptions, meaning that although both providers provide the same metric, it might have been calculated differently, leaving us unable to compare and interpret that feature correctly. Unfortunately, time constraints prevented us from implementing the budget-optimizing approach, which would have provided the digital marketing team with specific budget adjustments. Instead, we focused on training effective models that could be utilized to build an optimizer in the future. Additionally, our initial plan was to analyze campaigns more granularly, predicting outcomes per campaign type rather than per marketing channel. However, we could not assign contacts to individual campaigns using the target ground truth data, which resulted in us training a model for marketing channels instead, as it would still provide valuable information.

Additionally, we could not consider attribution modeling, as we cannot accurately credit each marketing channel or touchpoint influencing a user's conversion or action. Google and search campaigns benefit the most, as they are often the last point of contact before a conversion occurs. This created a challenge, as we could not accurately measure the true number of conversions per channel. It is essentially a black box that remains unknown.

We would have required additional historical data points spanning at least two to three years to gain even better insights and create a more comprehensive optimization model. Further data variables such as segment, performance, and campaign data could have also improved our results. Lastly, our model only predicted the next day rather than a specific period, as we initially intended. This short-term view does not account for long-term trends and seasonal fluctuations, crucial for planning an optimal budget allocation. Developing a long-term predictive model is a complex challenge that goes beyond the scope of this paper.

7 Conclusion and Recommendations

Our research lays the groundwork for incorporating machine learning into digital marketing campaigns to improve accuracy and performance. Using the provided data, we developed a successful model that predicts sales for the next day, revealing immense potential for further development and utilization of our models to optimize budget allocation at NOS.

Our main suggestion is to further develop and enhance the MMM approaches, as the methods we have developed still leave room for improvement. To improve the accuracy of the models, we recommend gathering more data that covers the last two to three years. In addition to this, expanding the short-term model to predict sales for a more extended period is critical in discovering the best budget allocation strategy. Furthermore, researchers should investigate individual SHAP values to understand the features better and potentially remove some to enhance the model's accuracy. Engineering and introducing additional features could also boost the performance. External variables such as holidays, economic indicators, or industry trends can also be considered here.

Moreover, the researchers could carry out a more comprehensive hyperparameter tuning. It's also essential to check for any data leakage, such as total leads, and determine if predicting the number of leads is better than sales. To better compare the different modeling strategies, the researchers should implement cross-validation that uses the same holdout dataset for all models to be able to compare the results accurately.

The researchers should investigate other MMM frameworks, such as PYMC, which can handle a more complex environment. To improve performance, the Digital Marketing team needs to know how much of the budget should be invested in which digital channel. Therefore, future work should focus on developing and deploying a marketing budget optimizer for their MMM approaches that can provide accurate results. This will help optimize the marketing strategy approach and ensure the proper budget is allocated to each channel. Optimizations can maximize sales by adjusting media inputs to keep the sum of media costs constant. Following these recommendations, researchers can develop a marketing budget allocation strategy for NOS that optimizes spend, ultimately increasing sales and ROI. hello

References

[BCG, 2018] BCG (2018). How digital marketing maturity helps you personalise marketing at scale. Think with Google. Accessed November 30, 2023.

[Becker, 1995] Becker, H. S. (1995). The power of inertia. *Qualitative Sociology*, 18:301–309.

[Cai et al., 2023] Cai, T., Jiang, J., Zhang, W., Zhou, S., Song, X., Yu, L., Gu, L., Zeng, X., Gu, J., and Zhang, G. (2023). Marketing budget allocation with offline constrained deep reinforcement learning. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 186–194.

[Chen et al., 2021] Chen, H., Zhang, M., Han, L., and Lim, A. (2021). Hierarchical marketing mix models with sign constraints. *Journal of Applied Statistics*, 48(13-15):2944–2960.

[DoorDash, 2020] DoorDash (2020). Optimizing doordash's marketing spend with machine learning. DoorDash Engineering Blog. Accessed December 3, 2023.

[Duque, 2023] Duque, P. (2023). Lightweight (bayesian) marketing mix modeling. GitHub repository. Accessed September 7, 2023.

[Han and Gabor, 2020] Han, B. and Gabor, J. (2020). Contextual bandits for advertising budget allocation. In *Proceedings of the 17th ADKDD Conference on Knowledge Discovery and Data Mining*.

[Harvard Business Review, 2023] Harvard Business Review (2023). A new gold standard for digital ad measurement? Harvard Business Review. Accessed December 3, 2023.

[Jin et al., 2017] Jin, Y., Wang, Y., Sun, Y., Chan, D., and Koehler, J. (2017). Bayesian methods for media mix modeling with carryover and shape effects. Technical report, Google Inc.

[Johnson, 1971] Johnson, R. M. (1971). Market segmentation: A strategic management tool. *Journal of Marketing Research*, 8(1):13–18.

[Kakalejčík et al., 2018] Kakalejčík, L., Bucko, J., Resende, P. A., and Ferencova, M. (2018). Multichannel marketing attribution using markov chains. *Journal of Applied Management and Investments*, 7(1):49–60.

[Katsov, 2017] Katsov, I. (2017). *Introduction to algorithmic marketing: Artificial intelligence for marketing operations*. Ilya Katsov.

[Kisilevich, 2023] Kisilevich, S. (2023). Practical approaches to optimizing budget in marketing mix modeling. GitHub repository. Accessed October 14, 2023.

[Lyft, 2019] Lyft (2019). Building lyft's marketing automation platform. Lyft Engineering Blog. Accessed December 3, 2023.

[McKinsey Company, 2019] McKinsey Company (2019). A new path for telco customer engagement. McKinsey Company. Accessed December 1, 2023.

[Miklosik et al., 2019] Miklosik, A., Kuchta, M., Evans, N., and Žák, (2019). Towards the adoption of machine learning-based analytical tools in digital marketing. *IEEE Access*, 7:85705–85718.

[Pandey et al., 2021] Pandey, S., Gupta, S., and Chhajed, S. (2021). Marketing mix modeling (mmm) - concepts and model interpretation. *International Journal of Engineering Research Technology (IJERT)*, 10.

[Park et al., 2022] Park, S. H., Lee, M., Kim, K., and Shin, D. (2022). A nonlinear optimization model of advertising budget allocation across multiple digital media channels. In *Proceedings of the 55th Annual Hawaii International Conference on System Sciences (HICSS 2022)*, pages 2166–2173. IEEE Computer Society.

[Ramaseshan, 1990] Ramaseshan, B. (1990). Research note: Marketing budgeting practices of retailers. *European Journal of Marketing*, 24(8):40–45.

[Zhao et al., 2019a] Zhao, K., Hua, J., Yan, L., Zhang, Q., Xu, H., and Yang, C. (2019a). A unified framework for marketing budget allocation. page 1820–1830.

[Zhao et al., 2019b] Zhao, K., Mahboobi, S. H., and Bagheri, S. R. (2019b). Revenue-based attribution modeling for online advertising. *International Journal of Market Research*, 61(2):195–209.

[Zhou et al., 2023] Zhou, H., Li, S., Jiang, G., Zheng, J., and Wang, D. (2023). Direct heterogeneous causal learning for resource allocation problems in marketing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 5446–5454.