UHASSELT

KNOWLEDGE IN ACTION

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Implementing Reinforcement Learning on Next Best Action: the data requirements and preparation

Khanh Ha Dao

Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :

Prof. dr. Koenraad VANHOOF

2023
2024

UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT

KNOWLEDGE IN ACTION

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Implementing Reinforcement Learning on Next Best Action: the data requirements and preparation

Khanh Ha Dao

Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :

Prof. dr. Koenraad VANHOOF

Implementing Reinforcement Learning on Next Best Action: the data requirements and preparation

Author: Khanh Ha Dao
Academic discipline: Business and Information Systems Engineering
Education institution: University of Hasselt
Promotor: Prof.dr.Koenraad Vanhoof
Assistant: Leen Jooken

2023 - 2024

Abstract

In the ever-evolving landscape of marketing, the quest for precision and personalization has led to the concept of "Next Best Action" (NBA), where the most suitable marketing action is recommended for individual customers in real-time. The decision of choosing the right marketing action at the right time is one of the most impactful aspects in improving the return on investment (ROI) of marketing campaigns. This master's thesis explores the integration of reinforcement learning (RL) techniques into marketing models, focusing specifically on the data requirements and preparation necessary for implementing RL in the context of NBA strategies. RL is a sub-branch of machine learning (ML) and a powerful technique in application to marketing models, which outperforms traditional marketing techniques by dynamically optimizing marketing processes through iterative learning and sequential decision making. RL algorithms however, require very specific input data, which is not always available within marketing companies. For this reason, the feasibility and limitations of the application of RL on NBA is tested, with the scope on marketing data requirements. The results have shown that marketing data can certainly be transformed into inputs for RL models if state transition probability matrices can be obtained after conducting analyses on the data. The first part of this thesis delves into the theoretical foundations of RL, its application in NBA and the scientific literature upon data requirements for NBA systems using RL. The second part of the research focuses on the technical aspect of data requirements and preparation for the implementation of RL in NBA. This includes the preparation of marketing data and results in an overview of different key components, which can be used as input to execute an RL experiment. Lastly, the feasibility of the deployment of RL algorithms on NBA models is discussed in terms of limitations and difficulties, regarding to the data preparation experiment.

Keywords— Reinforcement Learning, Next Best Action, Marketing, Data Requirements, Data preparation

Contents

1 Introduction

The modern marketing landscape is characterized by its ever-increasing complexity, driven by a dynamic interplay of customer preferences, evolving technologies, and data-driven decision-making [2, 24]. In this environment, the concept of "Next Best Action" (NBA) has emerged as a powerful strategy to enhance customer engagement and optimize marketing campaigns [28]. NBA entails real-time recommendation of the most relevant marketing action for individual customers, with the ultimate aim of creating more personalized and effective interactions [37]. More personalized and efficient interactions between companies and their customers implies significant improvement of marketeers' productivity and marketing campaigns' effectiveness which therefore generate a higher return on investment (ROI) for the company [37]. More and more companies are jumping on the AI train and are starting to implement reinforcement learning (RL) to improve their marketing processes. The use of AI-based techniques is interesting since they take advantage of real-life data and can therefore significantly improve efficency and effectiveness of marketing campaigns. There are however still numerous of challenges that are impeding the implementation of RL in a marketing context. One of these challenges can be related to the non-stationarity of the real-world, especially in the marketing world, where trends and seasonality are always present [32, 38]. Others can relate to data, specifically how RL models have to learn from data that are not in the form of sequences of states, actions and rewards [32]. This master's thesis specifically tackles this latter challenge by conducting research on how raw marketing data can be transformed into useful inputs (states, actions and rewards) for the implementation of RL models. In this manner, the feasibility of retrieving the necessary inputs from raw marketing data for executing NBA using RL models can be verified.

This master's thesis thus delves into the evolving intersection of marketing and artificial intelligence (AI), specifically focusing on the feasibility of employing RL techniques for NBA recommendations, with the scope of data requirements and preparation for the RL algorithms. By applying RL to the domain of NBA in marketing, we seek to address the fundamental questions: What are the requirements regarding marketing data, which is used as input in RL, in application to NBA? Are RL models suitable to implement in the context of NBA, starting from raw traditional marketing data present in a company?

2 Methodology

An exploratory literature review is conducted on the field of RL and NBA to first thoroughly understand these concepts, their theoretical foundations and the application of RL to NBA. This methodology is also used to conduct literature research upon data requirements for NBA using RL. Scientific databases used are *Science Direct*, *Springer Link*, *Research Gate* and *Towards Data Science*. Additionally, multiple online libraries are also consulted, including *John Wiley*

& Sons, *Public Library of Science* and *ACM Digital Library*. Search terms applied are "Reinforcement Learning", "Next Best Action", "Machine Learning", "Marketing" and "Data requirements". Furthermore, citation chaining is also being used when in-depth information about a specific subject or term is needed.

Literature findings upon data requirements for NBA using RL are utilized to support the composition of the action plan (Section 4.1), which acts as a road map to carry out the empirical experiment on data requirements and preparation for RL in supporting NBA. The empirical experiment makes use of real-world marketing data of a superstore in the United States[1]. According to the action plan, we first look at which input data we need, then we check whether our dataset contains the required input data to implement the RL-algorithm. If this is not the case, we check whether the data is transformable and if so, the dataset can be transformed into necessary input data by carrying out an RFM-analysis in R. It is important to mention that the scope of this experiment only includes data requirements and preparation for RL models executed in an offline setting, which means that there is no direct feedback from the real-life dynamic environment.

3 Exploratory literature review

In this section, findings resulted from exploratory literature review will be elaborated. The section is divided into four subsections, which includes *What is Reinforcement Learning?*, *What is Next Best Action?*, *Applying RL to NBA* and *RL input requirements in the context of NBA*. The first two subsections give a detailed description about the theoretical fundamentals of RL and NBA. The third subsection illustrates on the application of RL to NBA. The last subsection focuses on literature findings about input data that is required in the execution of NBA using RL.

3.1 What is Reinforcement Learning?

Reinforcement learning (RL) is a sub-branch of ML that trains a model to return an optimum solution for a problem by taking a sequence of decisions itself [3, 22]. RL refers to techniques in which an agent or a learner, learns how to make sequential decisions based on delayed reinforcement to maximize cumulative rewards and long-term outcomes, given a state of a specific environment [1, 22, 34].

RL learns from interaction with six key concepts: agent, environment, state, action, policy and reward [31, 37]. In a RL setting (Figure 1), the agent, which is the decision-making unit, is situated and operates in an environment, which is every external condition that the agent cannot modify or the training situation that the model must optimize [10, 30, 31]. The environment can demonstrate multiple states in which the agent takes actions, acts upon the environment to change its state and thereby receives a reward or penalty, which is determined by

[1] https://www.kaggle.com/datasets/vivek468/superstore-dataset-final

the achieved state and the agent's objective [26, 33]. In a marketing context, the agent's objective is commonly to maximize the customer lifetime value (CLV) or to minimize the cost of launching a marketing campaign [37]. In this manner, the agent can use this feedback to update his decision-making process to make better and improved decisions [28]. The state can be defined as the current position or condition that is returned by the environment, and the reward functions as a guide to help the agent move in the right direction and behave optimally [19]. The policy acts as a mapping between states and actions linked to each other and it determines how an agent will behave [17, 22]. To be able to structurally learn from the feedback received from the environment, it is important to have a policy, which can be used to determine actions that maximize the rewards, in a certain state [31]. This can simply be a look-up table where all the past actions taken are be recorded [10, 31]. After learning for a while, the agent will be able to select the action in a given state that yields the highest cumulative rewards, which is also the final outcome of the RL process, the optimal policy [33].

RL is furthermore a method which incorporates a trade-off between exploration and exploitation. The process of exploration and exploitation is being executed by the RL agent, who gradually learns the best strategies through interactions with the random environment, which can be seen as a black box, and through incorporation of the responses from these interactions to improve the overall performance [35]. Since the agent doesn't have any initial knowledge of the result of their action in a certain state, they will randomly select to execute an action at the beginning of the learning process. The agent thus "explores" the environment to gather knowledge on the results of his actions [33, 35]. When the agent is able to select the optimal action that maximizes his reward, he has successfully exploited the gathered knowledge in the exploration phase. Since exploration is costly in terms of time and resources, it is crucial that the agent can find a balance between exploiting what has been learned and continuing to explore the environment to keep obtaining new information for long-term improvements [35].

3.2 What is Next Best Action?

Next Best Action (NBA) is a marketing technique that helps businesses determine the most effective marketing actions, for every customer at different touch-points through their purchase journey, which will push the customers closer towards a desired conversion event [28]. A conversion event is often a purchase of a product, but it can also be a sign up for emails or a subscription renewal [25]. NBA utilizes a combination of AI and real-time interaction data to improve customer engagement by analyzing customers' unique needs and preferences [37, 38]. In other words, NBA consists of traditional targeting and personalization models including look-alike and collaborative filtering combined with RL, or other ML or AI-based techniques, to optimize multi-step marketing

[2]https://www.scribbr.com/ai-tools/reinforcement-learning/

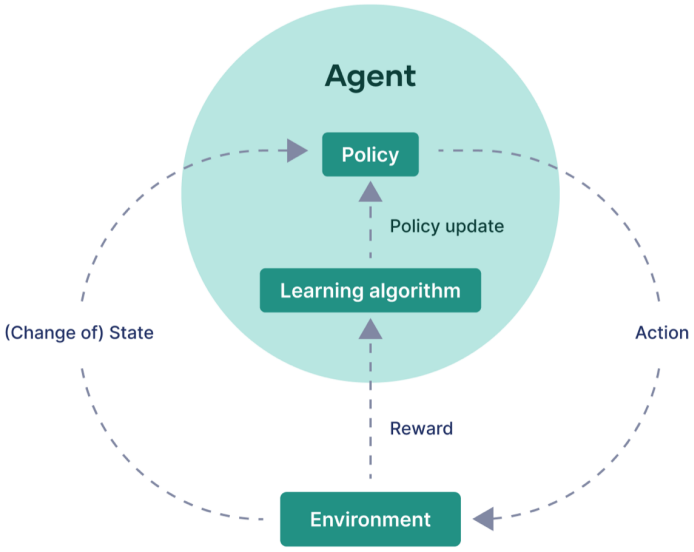

Figure 1: Key concepts of Reinforcement Learning [2]

action policies, also known as, NBA policies [11, 13].

The goal of NBA is to optimize marketing efforts and to improve the ROI of marketing campaigns. In other words, the executed marketing actions are personalized for each customer [28]. In this manner, businesses can execute more effective marketing campaigns which efficiently drive customers into conversion and hence minimize costs in the marketing department [28]. This technique is considered as the best practice in modern personalized marketing [38] since it is being utilized to make proposals based on customer's attributes and behaviors, purchase context and the company's strategic goals. NBA is considered to be a key element in Customer Relationship Management (CRM) as a Data-Driven Marketing approach [18].

The main idea of NBA is to assign to each customer three or more marketing actions that are considered the best actions for the customer. These can range from product offering and recommendation to retention and upselling pro-active actions [18, 37].

As already briefly mentioned above, the implementation of NBA relies mainly on the use of data-driven insights and the application of AI and ML to analyze and process these data [11]. According to [27], RL is one of the ML models that are the most suitable for NBA due to its multiple success in finding optimal winning strategies for turn-based games, which are comparable to the operation of NBA [28]. According to [9], NBA can also be modelled as a classification problem on which we can apply supervised and unsupervised learning.

3.3 Applying RL to NBA

RL models are mostly appropriate for finding the optimal winning strategies of turn-based games like chess, board games and Go [28]. It can therefore also be applied to finding the optimal Next Best Action to support and optimize marketing efforts [27]. Since there is a great variety of marketing actions, it is important for marketeers to send the right message, through the right channel and at the right time. It is therefore also crucial to rank customers or group them in different segments to be able to personalize marketing campaigns more effectively [23].

To explain RL applied to NBA, consider the following scenario: Your company wants to improve conversions and sales and say that the customer has bought a product from your business. If you want to keep the customer, you can recommend other products from your business that your customer may be interested in. If your customer pays a visit to your website and possibly makes a purchase again, you can rely on the marketing action that you have utilized to make even better decisions on which marketing actions to use in the future, for this specific customer, in this specific customer state. If the customer doesn't purchase your product or visit the website, then you know that the current marketing action being used is not suitable for this customer, or for the state in which the customer finds himself at that moment. When this happens, you can try to execute another marketing action.

When applying the Reinforcement Learning model, the *agent* can either be the customer or the marketeer in a marketing model [28]. We want to increase the customer's value by presenting them the right offer or recommendation, at the right time. On the other hand, the value of the marketeer is increased when the customer makes a purchase because of the offer received, hence the profit of their company increases. The *environment* is the digital advertising platform/channel in which the customer and the marketeer operate and interact with each other [28]. Examples of environments are email platforms, websites, social media platforms. The *state* of the environment can be seen as a certain situation in which the agent has to make a decision [31]. The *actions* are marketing actions that can be taken by the marketeers, so in this case it can be sending an email to the customer presenting the recommended products that the customer might be interested in [28]. Depending on the impact of the marketing action, the customer can receive a reward if their state has improved or a punishment if their state has worsen. In this case, if the customer shows interest in our offer or recommendation and visits the website, that means that the marketing action has a positive impact on the *state* of the customer/marketeer, which implies that we get a reward from the environment and the value of the customer/marketeer is increased. If not, we receive a negative feedback or a punishment from the environment, which means that we didn't successfully increase the agent's value. The RL model will use this feedback and learn from them to make better decisions on the next marketing action to be taken.

3.4 RL input requirements in the context of NBA

Findings from sources found on data requirements in the context of NBA using an RL model state that, when applying RL to NBA, marketing data have to contain the key components of an RL model. These components can be directly obtained from raw data, but can also be retrieved by transforming the data. The key components of RL models that can be obtained from the data are the *state space*, the *action space* and the *rewards* [14, 20, 28, 34]. These key components are required to establish the state transition probability matrices, which express the probabilities of a state transitioning to another, when a certain action is taken [29]. Furthermore, the data have to also contain a time series element since information about state transitions is essential to apply RL.

What we typically encounter when working with marketing data are demographic and purchasing information about the customers. Information about customer's buying behavior is essential for defining the *state space* since an improvement in the buying behavior, or customer state, implies a positive impact on the company's ROI. Often, the date in which customers made a purchase is also recorded. This type of information can be considered as a time series element and is indispensable since we want to be able to retrieve the impact of the marketing campaign on a certain customer over a period of time.

A typical objective of an NBA scenario is to optimize marketing campaigns. But there are also NBA scenarios with other objectives, including reduction of waste of marketing resources, improvement of customer engagement, customer experience and personalization [37].

According to [34], the first step to tackle when applying an RL algorithm is to define the *state* and *action space*. This process usually requires an in-depth analysis of the customers and their buying behavior. The *state space* represents a context that the environment presents to the agent to take actions for. An example of a state can be a loyalty level of the customers, but a state can also be something specific, like "not subscribed for newsletter" and "subscribed to newsletter". The *action space* represents all possible actions that the *agent* can execute, that can change the state of the environment [12, 26]. In an NBA context, the most common actions are offering a discount, receiving free shipping, referring a friend for a voucher. The *agent* component can be chosen freely by the analyst since the *agent* can be the customers or the marketeers in a marketing context. After obtaining the *state* and *action space*, state transition probability matrices can be established, which give information about the probability of a state transitioning to another after taking an action. This information is crucial to be able to retrieve the optimal policy, which consists of state-action pairs that yield the highest cumulative rewards [11, 33]. The *rewards* represent freely chosen scores that we give to a certain state transition after engaging with a certain action. When the agent experiences an improvement in their state, a positive score will be assigned, otherwise, the agent will receive a negative score, or a punishment [33].

Furthermore, there are also some hyperparameters that are typically used in RL models, which include the *number of episodes*, *discount factor*, *exploration*

rate and *learning rate*. These parameters can be set up by the analyst and are essential for the execution of the RL algorihtm. *Number of episodes* is the number of iterations that the RL model will generate. The *discount factor* is used to elate the *rewards* to the time domain since *rewards* can be achieved in the past, present and future [21]. The *exploration rate* represents the probability of exploring the environment by the *agent* [7]. The *learning rate* represents the rate in which the learning parameters are modified. A high *learning rate* allows fast changes, which is commonly used at the start of the experiment. As time progresses, the *learning rate* decreases [6].

4 Experiment

In this section, we want to analyze the feasibility of RL-implementation on NBA by addressing the problem of not be able to directly retrieve input data from raw traditional marketing data. We thus conduct an empirical experiment to test the feasibility of transforming real-world marketing data into useful inputs that can be fed to the RL algorithm, according to our findings in section 3.4. The NBA scenario chosen for this experiment has the objective to optimize marketing campaigns, in which the marketeer acts as the *agent*, who wants to know which marketing actions are the best to improve the customer's loyalty state. In section 4.1, we illustrate the action plan based on findings in section 3.4, which acts as a guide for the empirical experiment. In section 4.2, we conduct a exploratory data analysis to prepare the data for the experiment. Section 4.3 elaborates on the process of acquiring the necessary input data for the application of RL to NBA.

4.1 Action plan

The experiment starts by obtaining a suitable marketing dataset that can be used on an RL algorithm. The dataset used in this case, is a dataset from a superstore in the United States[1] that sells all sort of products, ranging from office supplies to technology and furniture. The dataset has 9994 observations, each representing an order from a certain customer. There are 23 variables. The description of each variable can be consulted in table 1.

Before getting started on the experiment, we first have to check whether the necessary inputs can directly be obtained from the dataset. Using our findings in section 3.4, we check for a time series element, the *state space*, *action space* and *rewards*. We want the loyalty levels to represent the *state space* but since they cannot be retrieved directly from the data, we can utilize information related to purchasing behaviour of the customers (recency, frequency and monetary). Loyalty levels can thus be defined by conducting an RFM-analysis. The *action space* is also not available from the dataset but we do have information about the amount of discount that the customers receive on certain product sub-categories. The different kinds of discount combined with the product sub-categories can therefore represent the marketing actions. The *rewards* can be

Row ID	Unique ID for each row
Order ID	Unique order ID for each customer
Order Date	Order date of the product
Ship Date	Shipping date of the product in yy-mm-dd
Ship Mode	Shipping mode specified by the customer (first class, second class, standard class, same day)
Customer ID	Unique ID to identify each customer
Customer Name	Name of the customer
Segment	The segment where the customer belongs (consumer, corporate, home office)
Country	Country of residence of the customer
City	City of residence of the customer
State	State of residence of the customer
Postal Code	Postal code of every customer
Region	Region where the customer belongs
Product ID	Unique ID of the product
Category	Category of the product ordered (office supplies, technology, furniture)
Sub-Category	Sub-category of the product ordered
Product Name	Name of the product
Sales	Sales of the product
Quantity	Quantity of the product
Discount	Discount provided
Profit	Profit/loss incurred

Table 1: Data description

determined based on the impact of the marketing actions on the customer's loyalty. The specific values of the rewards can be chosen freely, but accordingly to the degree of loyalty level's improvement. When defining the state transition probability matrices, we need to know which state the agent transitions to after applying a certain marketing action, so in this case to which loyalty level they transition to. Since state transitions are not available in the data, a time series element is essential, which is the date in which the customers make a purchase. Based on this element, we can define suitable periods in which the marketing campaigns are valid. The loyalty level of a customer before and after these periods are compared, and thereby the transition probabilities are calculated. A state transition probability matrix is composed of these transition probabilities, for each marketing action.

4.2 Exploratory data analysis

In this section, exploratory data analysis is conducted to check the quality of the data before getting started on the actual experiment. This process includes data cleaning by checking for missing values, duplicated rows and outliers of relevant variables.

In general, the dataset does not contain missing values or duplicated rows. Before checking for outliers, we first make a selection of the variables which are of interest for our experiment. Since we are only interested in the purchase behavior of the customers and the products that they purchase, the variables of interest are *Row ID*, *Order ID*, *Order Date*, *Customer ID*, *Customer Name*, *Product ID*, *Category*, *Sub-Category*, *Product Name*, *Sales*, *Quantity*, *Discount* and *Profit*. Boxplots are utilized to check for outliers. Notice that this procedure is only being carried out for numeric variables.

The first numeric variable we look at is *Sales*. Since the store sells goods of different values, it is not unusual that we will find some outliers since office supplies can be very cheap but technology goods and furniture can be quite expensive. We see that most of the orders have a *Sales* value below 5000. Furthermore, the *Sales* value also depends on the quantity and not only on the unit price. The second numeric variable is *Quantity*. We can see that the highest quantity is 14, which is not unusual. The third numeric variable is *Discount*. Here, we check whether all discount are below 1 since discounts cannot exceed 100%. The last numeric variable is *Profit*. This reasoning for *Sales* goes the same for *Profit*. The profit can be very high for orders with high value items, high quantity or with low or no discount. This variable also represents loss when its value is negative, a loss can incur due to multiple reasons, such as an order containing a low value item while having a small quantity and a high discount. In conclusion, there are some variables that contain outliers but they are not invalid and will not impact the result of our data preparation process for the RL experiment. The boxplots of these variables can be consulted in table 2.

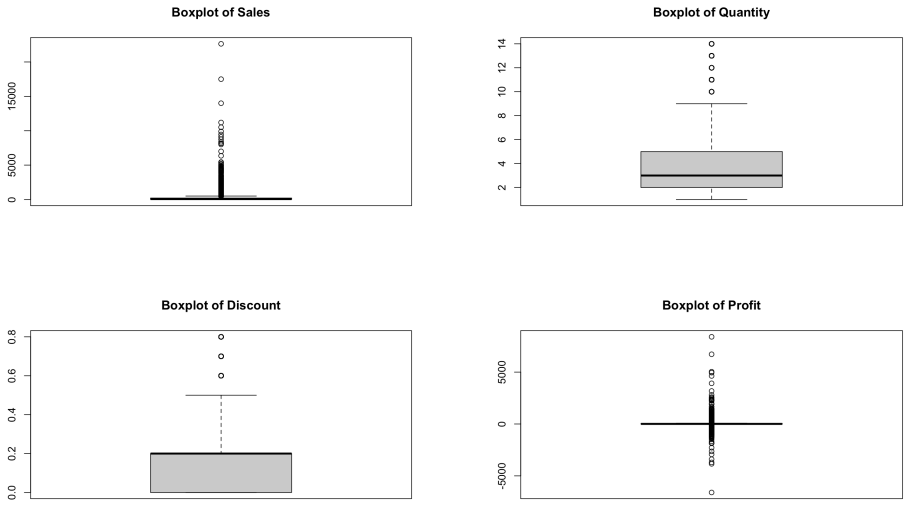

Table 2: Boxplots of Sales, Quantity, Discount and Profit

4.3 Data preparation analysis

This section elaborates on the process of data preparation, which consists mainly of the RFM-analysis, which is used to determine the *state space* (Section 4.3.1). The process of defining the *action space*, state transition probability matrices and *rewards* are also illustrated in section 4.3.2, 4.3.3 and 4.3.4 respectively.

4.3.1 Defining the state space

One of the key RL component to use as input for the RL algorithm is the *state space*. The *state space* cannot be directly retrieved from the dataset, but can be obtained by conducting an RFM-analysis. This technique will deliver loyalty levels as output, which are converted into customer segmentation, which can then be used to represent the different states of the customers.

The RFM-analysis is a powerful database marketing technique used for evaluating or segmenting customers based on their buying behavior. It is a method used to define customer segmentation, in other words, a method to cluster customers into different segments based on similarities and differences to identify valuable customers for the company [4, 5]. The analysis consists of a scoring method which is used to rank customers based on their past purchasing history [5]. The RFM-analysis is relevant in a wide range of applications that involve a large number of customers. The method is based on three dimensions, recency (R), frequency (F) and monetary (M) [4, 5, 16]. Recency expresses the number of days since the customer last made a purchase. A high recency value implies that the customer has not made a purchase recently. Frequency is the number

of times which the customers make a purchase in a certain period. So a higher frequency value corresponds with a customer who frequently buys something. Monetary is defined as the amount of money that the customer spends during a certain period [4]. The technical execution of the RFM-analysis in R is based on [8].

The required variables to execute this analysis, which are *recency*, *frequency*, *monetary*, are not directly present in the dataset, but can be derived from other variables that are given in the dataset. Customer segments determined by RFM-values can be used to represent the states since it clusters customers in different groups based on their purchasing behavior. Furthermore, these values are not static since the state of the customers can change over time, depending on the period in which they are calculated. These two elements make it quite reasonable to use the RFM-analysis to define the *state space*.

- Calculating Recency, Frequency and Monetary

In our dataset, the **recency** is calculated based on the variable *Order Date*. The last date of the period in which we calculate the RFM-scores is our analysis date. To retrieve the recency, the most recent date that each customer has made a purchase within the analysis period, is deducted from the analysis date. For the **frequency**, we also use the variable *Order Date*. We calculate for each customer the amount of different order dates in which they made a purchase. The sum of this amount makes up the frequency in which customers make a purchase. There are cases in which the customers place multiple orders/items within a certain date. In these cases, we have chosen to consider multiple orders within one date as one, so a frequency value of one. The **monetary** value can be considered as the amount of sales that the company obtained from each customer. So for each customer, we calculate the total amount of sales from the *Sales* variable. The reason we choose to use *Sales* over *Profit* is because sales represents the actual amount that the customer spends, without taking the costs of the company, such as discounts given to customers, into account.

- Defining RFM-segmentation

The next step is to assign the Recency-, Frequency- and Monetary-score to each observation. The three variables are divided into four categories, based on their median, first and third quartile. A score of 1 to 4 is assigned to each category of the variables. When the value of Recency is above the third quartile, all recency values in this category is assigned a Recency-score 1, since a higher recency means that it has been longer since the customers have made a purchase. Recency values higher than the median and lower than the third quartile are assigned a Recency-score of 2. Recency values higher then the first quartile and lower than the median are assigned a Recency-score of 3. A Recency-score of 4 is assigned to all recency values below the first quartile. This same process is executed for Monetary, but since higher monetary value means more money spent, the monetary values that are higher than the third quartile get assigned the highest Monetary-score. For Frequency, we notice that the median, first and

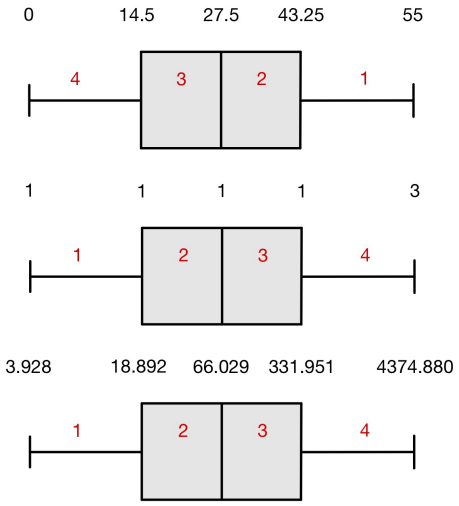

Figure 2: Boxplots of Recency, Frequency and Monetary

third quartile values of the frequency all equal to 1. For this case, it is decided to assign the maximum score (Frequency-score of 4) to all frequency values since we clearly see that a frequency value equals to 1 is oft-recurring. When looking at the calculated frequency values, frequency values that are higher than 1 do occur, but these are quite rare so that is why we have decided to consider frequency values of 1 to be "frequent". This score assignment procedure is illustrated in boxplots in figure 2, with the numbers in red representing the RFM-scores and the numbers in black representing the minimum, first quartile, median, third quartile and maximum, from left to right. Now that each value of all three variables are assigned a certain score between 1 and 4, a general RFM-score is calculated by taking a sum of the Recency-score multiplied by 100, Frequency-score multiplied by 10 and Monetary-score. This multiplication ensures that the Recency-score is the first digit, Frequency-score is the second digit and the Monetary-score is the third digit of this general RFM-score. All customers will now have a RFM-score between 111 and 444 and can now be divided in different customer segments.

- Defining loyalty states

The distribution of the customer segments is done based on a commonly accepted method used in practice, [8] and [36]. According to [8] and [36], customers can be divided into six segments or loyalty states, according to their purchase behavior. The first loyalty state is "Loyalists", these are customers that have completed a recent purchase, buy frequently and spend the most money. The second loyalty state is "Potential loyalists", they are customers who recently

Customer segment	RFM-scores
Loyalists	444, 443, 434, 344, 433, 343, 334, 333, 244
Potential loyalists	324, 243, 234, 233, 224, 442, 441, 431, 422, 421, 342, 432
New customers	414, 413, 412, 411, 314, 313, 424, 423, 332, 323
Promising	331, 341, 322, 321, 312, 242, 241, 231, 222, 212, 311
Need attention	232, 223, 214, 213, 144, 143, 134, 133, 142, 141, 114
Detractors	132, 131, 124, 123, 113, 221, 211, 122, 121, 112, 111

Table 3: Customer segmentation based on RFM-scores

spent a fair amount of money in the store more than once. "New customers" is the third loyalty state, which contains customers who recently completed a recent purchase but have a low purchasing frequency since they are "new" to the company. The fourth loyalty state is "Promising", which consists of customers who have recently completed a purchase but did not spend much. The fifth loyalty state is "Need attention", these are customers that completed big and frequent purchases a long time ago, efforts have to be made to keep them as customers. The last loyalty state is "Detractors", these are customers that purchased a long time ago, with a small amount of orders and a low amount of money. More information on the customer segmentation and their corresponding RFM-scores can be consulted in table 3.

4.3.2 Defining the action space

To define the *action space*, we first identify the marketing actions that we want to use in our *action space*. Since we need a variable that represents the marketing campaigns, which is not directly obtainable from the dataset, we create this by combining variables. In our case, a marketing campaign/action is defined by two variables, "Sub-category" and "Discount". We have chosen for "Sub-category" instead of "Category" because it is easier to define the periods in which the discount is valid since "Sub-category" gives us a specific item of a product.

In the dataset, there are 40 different combinations of "Sub-category" and "Discount", which means that there are potentially 40 different marketing actions that we could use. For every marketing action, a number of periods have to be defined because we want to obtain the state transitions of the customers when a marketing action is being used. In other words, we want to see whether the state of the customer has improved or worsen after applying a certain marketing action. Since marketing actions that are valid throughout the year are not useful to determine their impact on the state transitions, we have to make a selection of marketing actions that are suitable for defining the periods. Based on the table in figure 3, which presents the top ten combinations of "Sub-category"

Sub-Category <chr>	Discount <dbl>	Number of observations <int>
Binders	0.20	573
Paper	0.20	513
Phones	0.20	469
Binders	0.70	380
Storage	0.20	316
Accessories	0.20	304
Art	0.20	298
Chairs	0.20	250
Furnishings	0.20	248
Binders	0.80	233

Figure 3: Sub-category and Discount

and "Discount" with the most observations, we make a selection of suitable "Sub-category" and "Discount" combinations. The first criterion for choosing which marketing actions are suitable to work with is the ability to define periods in which the marketing actions are being used by the customers. Marketing actions that have too many observations are very challenging to define periods since there are no sufficient large gaps between the order dates. Therefore, a lower bound is defined for the number of periods. This implies that marketing actions must have at least four periods. After evaluating the marketing actions based on this first criterion, it is noticed that there are certain marketing actions in which too many periods can be defined (too many gaps between the order dates) due to the lack of observations. Therefore, the second criterion is formed, which is an upper bound for the number of periods. This means that the number of periods cannot exceed 10. After applying these two criteria, five marketing actions met the conditions, namely *Storage 20%*, *Accessories 20%*, *Art 20%*, *Chairs 20%* and *Furnishings 20%*. These five marketing actions form our *action space*.

4.3.3 State transition probability matrices

After obtaining the *state* and *action space*, the objective is to create a state transition probability matrix for each marketing action to acquire information about the impact of each marketing action on the different customer states.

We first iterate a base scenario, in which we will only use a sample of the dataset. This step is necessary since we want to obtain the transition of the states. The base scenario is used as a reference start point to compare the recency, frequency and monetary values as we progress through the different periods. The sample of the dataset contains data from order date 01/01/2014 up to 01/03/2014 (the first two months). In this sample, we will calculate the recency, frequency and monetary value for all observations. Afterwards, we

calculate the first quartile, the median and the third quartile of the recency, frequency and monetary, which will be used as comparison values to assign the RFM-scores on the relevant data samples. Specific explanation on this process can be consulted in section 4.2.1 under *Defining RFM-segmentation.*

The next step is to define the customer state transitions for each marketing action. The goal of this process is to compare the customer's loyalty state after each marketing action. In other words, we want to check whether the loyalty state of the customer is changed after using a certain marketing action. In this manner, we can obtain the impact of each marketing action on each loyalty state and retrieve information about which marketing action to use when a customer is in a certain loyalty state to maximize the customer's and the company's value. For example: For marketing action "Storage 20%", we calculate the RFM-score and determine the corresponding loyalty state two months before the start and two months before the end of a certain period. The span of two months is chosen to match with the base scenario data sample to make sure that the comparison would be well grounded since we use the base scenario as a reference point to execute the comparison. As a result, we would obtain the loyalty state of the customer before and after the customer engages with the marketing action. In this manner, we can check whether the marketing action has an impact on the improvement of the customer's loyalty state.

To compose the probability transition matrix for each marketing action, we look at the loyalty states before and after each period, for all defined periods. For each loyalty state before engaging with the marketing action, we calculate its transition probability to other loyalty states after engaging with the marketing action. The transition probability for a certain loyalty state before engaging with the marketing action equals the amount of customers transitioning from this loyalty state to another, divided by the total number of all state transitions that occur after engaging with the marketing action. An example for illustration: When applying marketing action *Storage 20%*, for state "Need attention", there are 48 transitions in total. 11 customers remain in the same state, so the transition probability would be 11 divided by 48, equals 23%. Three customers transition to "Promising", so transition probability is three divided by 48, yields 6%. Nine customers transition to "Potential Loyalists", so nine divided by 48 makes up 19% and 25 customers transition to "Loyalists", divided by 48 equals 52%.

4.3.4 Assigning rewards

The rewards can be defined based on the amount of improvement in customer's loyalty state. When the loyalty state of a customer remains the same under influence of a marketing action, a reward of zero is assigned. Except for the state "Loyalists" since there is no loyalty state that is higher than "Loyalists", a reward of plus one will be assigned. When the loyalty state increases by one level, a reward of plus one is assigned. This process is identically applied to when the loyalty state is increased by two or three levels. The same procedure is valid for when the loyalty state decreases but with a negative reward or a

punishment. The reward assignment can be visually consulted in the reward matrix in figure 9. Each cell in this matrix corresponds to each cell in the state transition probability matrices.

5 Results

After conducting the experiment, the following results are obtained. For the marketing action *Storage 20%*, four periods are defined with a total of 129 observations. For *Accessories 20%*, four periods and 143 observations. Marketing action *Art 20%* has four periods and a total of 94 observations. *Chairs 20%* has seven periods and 410 observations. Lastly, marketing action *Furnishings 20%* has eight periods with 419 observations. Further, it is noticed that only certain RFM-scores are present and that the loyalty state "New Customers" and "Detractors" never occur. This is due to the fact all customers have a Frequency-score of 4, a decision that we made in section 4.3.1, under *Defining RFM-segmentation*. When looking at table 3, we see that the second digit of the RFM-scores, the digit that represents the Frequency-score, of "New customers" and "Detractors" are less than 4, so this is why these two customer segments never occur in the results. These five state transition probability matrices can be consulted in figures 4, 5, 6, 7 and 8.

A probability transition matrix is composed of four rows and four columns, with the rows and columns representing the loyalty states before and after engaging with the marketing action, respectively. The loyalty states that occurred are "Need attention", "Promising", "Potential loyalists" and "Loyalists". For illustration, cell 11 (Figure 4) is the probability that customers with loyalty state "Need attention" would remain in the same state after using 20% discount to buy storage items, cell 12 is the probability that customers with loyalty state "Need attention" would transition to state "Promising", and so forth...

For marketing action *Storage 20%* in figure 4, it is noticed that the action works well for all loyalty states since the probability of transitioning to the state "Loyalists" is the highest for all states. For marketing action *Accessories 20%* in figure 5, all loyalty states have the highest transition probability to "Loyalists" except for "Promising". The state "Promising" would most likely transition to "Need attention", with a probability of 45%. When looking at *Art 20%* in figure 6, we see that it has the same effect on the loyalty states as *Storage 20%*. Marketing action *Chairs 20%* in figure 7 only has a positive effect on "Loyalists" since it remains in "Loyalists" with a probability of 51%. "Promising" and "Potential loyalists" are most likely to transition to "Need attention", so they experience a decrease in loyalty level with a probability of 44% and 28% respectively. "Need attention" has the highest chance of remaining in the same state, with 46%. In *Furnishings 20%* in figure 8, the only loyalty state that is most likely to have a positive transition is "Loyalists", with 62%, the rest of the loyalty states has the highest chance of experiencing a neutral effect. With 51%, 36% and 30% for "Need attention", "Promising" and "Potential loyalists" respectively. Overall, it it noticed that the customers that are "Loyalists" are

	Need attention	Promising	Potential loyalists	Loyalists
Need attention	0.23	0.06	0.19	0.52
Promising	0.13	0.07	0.13	0.67
Potential loyalists	0.21	0.11	0.05	0.63
Loyalists	0.29	0.11	0.11	0.49

Figure 4: Probability transition matrix: Storage 20%

	Need attention	Promising	Potential loyalists	Loyalists
Need attention	0.18	0.15	0.20	0.48
Promising	0.45	0.10	0.10	0.35
Potential loyalists	0.24	0.10	0.14	0.52
Loyalists	0.17	0.13	0.15	0.55

Figure 5: Probability transition matrix: Accessories 20%

most likely to remain as "Loyalists" in all marketing actions considered in this analysis. This can potentially mean that once the customers become "Loyalists", it is hard for them to drop down to lower loyalty states.

6 Discussion

In this section, we want to discuss whether the deployment of RL algorithms on NBA models are feasible based on the findings of our research.

So how can the results from the previous section be interpreted in the context of NBA when we apply RL? With NBA, we want to determine the next best marketing action to offer to a customer, implying the marketing action that gets the customer as close as possible to the desired end state. In the scenario of the experiment, this would be the highest level of loyalty. To be able to execute this process, we first have to look at the loyalty state in which the customer is in. When the state of the customer is determined, the RL algorithm will find through trial and error, with the help of the state transition probability matrices as input data, the best marketing action for this loyalty state, that yields the

	Need attention	Promising	Potential loyalists	Loyalists
Need attention	0.15	0.05	0.15	0.65
Promising	0.33	0.07	0.07	0.53
Potential loyalists	0.25	0.13	0.13	0.50
Loyalists	0.26	0.16	0.09	0.49

Figure 6: Probability transition matrix: Art 20%

	Need attention	Promising	Potential loyalists	Loyalists
Need attention	0.46	0.07	0.11	0.35
Promising	0.44	0.21	0.08	0.27
Potential loyalists	0.28	0.23	0.24	0.24
Loyalists	0.28	0.05	0.16	0.51

Figure 7: Probability transition matrix: Chairs 20%

	Need attention	Promising	Potential loyalists	Loyalists
Need attention	0.51	0.06	0.16	0.27
Promising	0.34	0.36	0.04	0.26
Potential loyalists	0.19	0.24	0.30	0.28
Loyalists	0.22	0.06	0.11	0.62

Figure 8: Probability transition matrix: Furnishings 20%

	Need attention	Promising	Potential loyalists	Loyalists
Need attention	0	1	2	3
Promising	-1	0	1	2
Potential loyalists	-2	-1	0	1
Loyalists	-3	-2	-1	1

Figure 9: Reward matrix

highest cumulative rewards. As mentioned in section 3.1, RL is useful to determine an optimum solution by taking a sequence of decision and its objective is to maximize cumulative rewards. So when implementing RL to determine NBA, the RL algorithm will not solely consider the immediate rewards, but also takes the cumulative rewards of the sequence of marketing actions into account.

However, maximizing cumulative rewards is not always ideal when the marketeer is focused on short-term in stead of long-term objectives. This happens when the company, for example wants to focus on strengthening their short-term competitive position [15]. This implies that determining the next best action that maximizes the immediate rewards are more valuable than cumulative rewards. In this case, implementation of RL for NBA would be less interesting.

According to the literature review in section 3, when companies are considering implementing ML models to improve their NBA-systems, RL would be one of the greatest fit since it is specialized in taking sequential decisions, which is appropriate when marketers want to evaluate a series of marketing campaigns. When working with raw marketing data, data analysis and transformation have to be conducted to be able to obtain the necessary inputs for the RL algorithms. Findings from this experiment suggest that it is indeed feasible to use RL to support NBA-systems. The data preparation analysis is however quite extensive and requires exhaustive procedures. This is a very important aspect that has to be kept in mind when making decisions about deploying RL, specifically when we look at the available resources to execute this data preparation process. However, cases where raw marketing data are not transformable also exist (ex. no time series element), which is an enormous barrier to the implementation of RL models. This research does confirm the feasibility of RL deployment in the context of NBA, but not for all cases of raw marketing data.

Moving forward, further research endeavors should focus on critically evaluating the methodology used in this master's thesis and exploring new methodologies to prepare and transform raw existing marketing data into suitable input for RL in an offline environment.

7 Conclusion

In conclusion, this master's thesis has delved into the critical aspects of data requirements and preparation process essential as input for the implementation of RL in a marketing context, specifically for NBA. In other words, it has addressed the problem of not be able to directly retrieve the right inputs from raw marketing data to implement an RL model. To support the comprehension of the relevant concept about RL, NBA and the application of RL in an NBA context, an extensive review of literature was carried out. Through an empirical analysis, it has been demonstrated that the implementation of RL algorithms in NBA heavily relies on the quality and relevance of the data fed into the learning process. Therefore, meticulous data preprocessing has to be conducted to support facilitating the deployment of RL algorithms for NBA applications, by making data suitable as input for an RL experiment. The data preparation

experiment consists largely of the RFM-analysis to obtain the customer states, marketing actions and rewards, the key components of an RL model. The results acquired are state transition probability matrices for each marketing action, which can be utilized as input for the RL experiment. Furthermore, this thesis also addressed the challenges inherent in the data acquiring and preparation process in terms of feasibility and limitations.

By conducting a research and advancing our understanding upon the data requirements and preparation for RL in NBA, this thesis has shed light on the feasibility of the use of RL in a marketing context, specifically for NBA systems to enhance marketing strategies in the pursuit of customer-centricity.

References

[1] Prerna Agarwal et al. *Goal-Oriented Next Best Activity Recommendation using Reinforcement Learning*. May 2022.

[2] Longbing Cao and Chengzhang Zhu. "Personalized next-best action recommendation with multi-party interaction learning for automated decision-making". English. In: *PLOS ONE* 17.1 (Jan. 2022). Publisher: Public Library of Science, e0263010. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0263010. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0263010.

[3] Mengmeng Chen. "Programmatic Marketing via Reinforcement Learning". In: *Journal of Business Management & Economics* 05 (Nov. 2017), pp. 01–04. DOI: 10.15520/jbme.2017.vol5.iss11.269.pp01-04.

[4] A. Joy Christy et al. "RFM ranking – An effective approach to customer segmentation". In: *Journal of King Saud University - Computer and Information Sciences* 33.10 (Dec. 2021), pp. 1251–1257. ISSN: 1319-1578. DOI: 10.1016/j.jksuci.2018.09.004. URL: https://www.sciencedirect.com/science/article/pii/S1319157818304178.

[5] Onur Dogan, Ejder Aycin, and Zeki Bulut. "Customer segmentation by using RFM model and clustering methods: A case study in retail industry". English. In: *International Journal of contemporary economics and administrative sciences* (2018). ISSN: 1925-4423.

[6] Eyal Even-Dar and Yishay Mansour. "Learning Rates for Q-Learning". English. In: *Computational Learning Theory*. Ed. by G. Goos et al. Vol. 2111. Series Title: Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 589–604. ISBN: 978-3-540-42343-0 978-3-540-44581-4. DOI: 10.1007/3-540-44581-1_39. URL: http://link.springer.com/10.1007/3-540-44581-1_39.

[7] *Exploitation and Exploration in Machine Learning - Javatpoint*. English. URL: https://www.javatpoint.com/exploitation-and-exploration-in-machine-learning#:~:text=In%20reinforcement%20learning%2C%20whenever%20agents,agent%20about%20the%20state%2C%20actions%2C (visited on 04/13/2024).

[8] Harshita Garg. *Customer Segmentation using RFM analysis in R*. English. May 2021. URL: https://medium.com/analytics-vidhya/customer-segmentation-using-rfm-analysis-in-r-cd8ba4e6891 (visited on 03/09/2024).

[9] Nikhil Goel. *Next Best Action using Machine Learning (XGBoost)*. Oct. 2023. URL: https://www.linkedin.com/pulse/next-best-action-using-machine-learning-xgboost-nikhil-goel-ktxzc/ (visited on 04/14/2024).

[10] Gabriel Gómez-Pérez et al. "Assigning discounts in a marketing campaign by using reinforcement learning and neural networks". In: *Expert Systems with Applications* 36.4 (May 2009), pp. 8022–8031. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2008.10.064.

[11] *How Deep Reinforcement Learning Can Take Content Marketing to the Next Level - Contrend*. English. Mar. 2022. URL: https://contrend.com/blog/how-deep-reinforcement-learning-can-take-content-marketing-to-the-next-level/,%20https://contrend.com/blog/how-deep-reinforcement-learning-can-take-content-marketing-to-the-next-level/ (visited on 02/01/2024).

[12] Lodewijk Kallenberg. *Lecture Notes Markov Decision Processes - version 2022*. Apr. 2022.

[13] Ilya Katsov. "Building Next Best Action Engines for B2C and B2B Use Cases". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. CIKM '22. New York, NY, USA: Association for Computing Machinery, Oct. 2022, pp. 5088–5089. ISBN: 978-1-4503-9236-5. DOI: 10.1145/3511808.3557511. URL: https://doi.org/10.1145/3511808.3557511.

[14] Ilya Katsov. *Next Best Action Model And Reinforcement Learning*. English. May 2019. URL: https://blog.griddynamics.com/building-a-next-best-action-model-using-reinforcement-learning/ (visited on 02/23/2024).

[15] Sev K Keil, David Reibstein, and Dick R Wittink. "The impact of business objectives and the time horizon of performance evaluation on pricing behavior". English. In: *ScienceDirect* 18 (June 2001), pp. 67–81. ISSN: 0167-8116. DOI: 10.1016/S0167-8116(01)00027-1. URL: https://www.sciencedirect.com/science/article/pii/S0167811601000271.

[16] Mahboubeh Khajvand et al. "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study". In: *Procedia Computer Science*. World Conference on Information Technology 3 (Jan. 2011), pp. 57–63. ISSN: 1877-0509. DOI: 10.1016/j.procs.2010.12.011.

[17] Yuxi Li. "Reinforcement Learning in Practice: Opportunities and Challenges". In: *ResearchGate* (Apr. 2022).

[18] João Luís Trindade Milheiro. "Next best action – a data-driven marketing approach". English. Accepted: 2020-02-18T19:37:24Z. MA thesis. Jan. 2020. URL: https://run.unl.pt/handle/10362/92945 (visited on 09/21/2023).

[19] Eduardo F. Morales and Hugo Jair Escalante. "Chapter 6 - A brief introduction to supervised, unsupervised, and reinforcement learning". In: *Biosignal Processing and Classification Using Computational Learning and Intelligence*. Ed. by Alejandro A. Torres-García et al. Academic Press, Jan. 2022, pp. 111–129. ISBN: 978-0-12-820125-1. DOI: 10.1016/B978-0-12-820125-1.00017-8.

[20] *Next best action recommendation - part 3: recommending actions using reinforcement learning*. English. URL: https://dataroots.io/blog/https-dataroots-io-research-contributions-next-best-action-recommendation-part-3-recommending-actions-using-reinforcement-learning (visited on 04/13/2024).

[21] Dr Barak Or. *Penalizing the Discount Factor in Reinforcement Learning*. en. Oct. 2023. URL: https://towardsdatascience.com/penalizing-the-discount-factor-in-reinforcement-learning-d672e3a38ffe (visited on 04/13/2024).

[22] Edwin Pednault, Naoki Abe, and Bianca Zadrozny. *Sequential cost-sensitive decision making with reinforcement learning*. English. Conference: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26, 2002, Edmonton, Alberta, Canada. DBLP, July 2002. DOI: 10.1145/775047.775086.

[23] Rui Jorge Eduardo Ramos. "Next Best Action Recommendation". In: (2022). URL: https://repositorio-aberto.up.pt/bitstream/10216/142911/2/572743.pdf.

[24] Albérico Travassos Rosário and Joana Carmo Dias. "How has data-driven marketing evolved: Challenges and opportunities with emerging technologies". In: *International Journal of Information Management Data Insights* 3.2 (Nov. 2023), p. 100203. ISSN: 2667-0968. DOI: 10.1016/j.jjimei.2023.100203.

[25] Nandini Seth. "Essays on next best action in digital marketing using reinforcement learning". English. In: *University* (2021). Accepted: 2023-02-03T10:54:12Z Publisher: Bangalore. URL: https://shodhganga.inflibnet.ac.in:8443/jspui/handle/10603/455996.

[26] Mohit Sewak. "Introduction to Reinforcement Learning". English. In: *Deep Reinforcement Learning: Frontiers of Artificial Intelligence*. Ed. by Mohit Sewak. Singapore: Springer, 2019, pp. 1–18. ISBN: 9789811382857. DOI: 10.1007/978-981-13-8285-7_1. URL: https://doi.org/10.1007/978-981-13-8285-7_1.

[27] Gurbaksh Sharma. *Next Best Action Recommendation Using Reinforcement Learning*. English. Feb. 2023. URL: https://medium.com/@gurumail10/next-best-action-recommendation-using-reinforcement-learning-8b070e858d36 (visited on 04/14/2024).

[28] Gurbaksh Sharma. *What is Next-best Action In Marketing? - Treasure Data Blog*. English. Mar. 2023. URL: https://blog.treasuredata.com/blog/2023/03/03/next-best-action-marketing/ (visited on 10/25/2023).

[29] Ayush Singh. *Introduction to Reinforcement Learning : Markov-Decision Process*. en. May 2022. URL: https://towardsdatascience.com/introduction-to-reinforcement-learning-markov-decision-process-44c533ebf8da (visited on 05/15/2024).

[30] Vinay Singh et al. "How to Maximize Clicks for Display Advertisement in Digital Marketing? A Reinforcement Learning Approach". English. In: *Information Systems Frontiers* 25.4 (Aug. 2023), pp. 1621–1638. ISSN: 1572-9419. DOI: 10.1007/s10796-022-10314-0. URL: https://doi.org/10.1007/s10796-022-10314-0.

[31] Richard S. Sutton and Andrew Barto. *Reinforcement learning: an introduction*. English. Nachdruck. Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2014. ISBN: 978-0-262-19398-6.

[32] Georgios Theocharous et al. *Reinforcement Learning for Strategic Recommendations*. Sept. 2020.

[33] Marlies Vanhulsel et al. "Simulation of sequential data: An enhanced reinforcement learning approach". In: *Expert Systems with Applications* 36.4 (May 2009), pp. 8032–8039. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2008.10.056.

[34] Víctor A. Vargas-Pérez et al. "Deep reinforcement learning in agent-based simulations for optimal media planning". In: *Information Fusion* 91 (Mar. 2023), pp. 644–664. ISSN: 1566-2535. DOI: 10.1016/j.inffus.2022.10.029.

[35] Haoran Wang, Thaleia Zariphopoulou, and Xunyu Zhou. *Exploration versus exploitation in reinforcement learning: a stochastic control approach*. Dec. 2018.

[36] *What Are RFM Scores and How To Calculate Them*. English. May 2022. URL: https://connectif.ai/en/blog/what-are-rfm-scores-and-how-to-calculate-them/ (visited on 03/09/2024).

[37] *What is Next Best Action? — Pega*. English. Jan. 2023. URL: https://www.pega.com/next-best-action (visited on 10/25/2023).

[38] Yuxi Zhang and Kexin Xie. "Incorporating intent propensities in personalized next best action recommendation". In: *Proceedings of the 13th ACM Conference on Recommender Systems*. RecSys '19. New York, NY, USA: Association for Computing Machinery, Sept. 2019, p. 530. ISBN: 978-1-4503-6243-6. DOI: 10.1145/3298689.3346962. URL: https://doi.org/10.1145/3298689.3346962.