پژوهش‌های نوین در تصمیم‌گیری

دوره ۹، شماره ۲، تابستان ۱۴۰۳، صص ۱۹۰-۱۶۴

نوع مقاله: پژوهشی

ارائه چارچوبی داده‌محور مبتنی بر شبیه‌سازی عامل‌بنیان برای پیش‌بینی ریزش مشتری در صنعت مخابرات

محمد جواد جعفری [۱]، محمد جعفر تارخ [*۲]، پریا سلیمانی [۳]

۱. دانشجوی رشته دکتری مدیریت فناوری اطلاعات، گروه تخصصی مدیریت فناوری اطلاعات، دانشکده مدیریت و اقتصاد، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران

۲. استاد گروه مهندسی صنایع، دانشکده مهندسی، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران

۳. دانشیار، گروه مهندسی صنایع، دانشکده مهندسی، واحد تهران جنوب، دانشگاه آزاد اسلامی، تهران، ایران

تاریخ دریافت: ۱۴۰۲/۱۱/۱۳ تاریخ پذیرش: ۱۴۰۳/۰۳/۱۹

چکیده

ریزش مشتریان یک چالش مهم برای صنعت ارتباطات از راه دور است که نیاز به استراتژی‌های موثر برای پیش‌بینی و پیشگیری دارد. در حالی که تحقیقات قبلی روش‌های مختلفی از جمله مدل‌سازی مبتنی بر عامل (ABM) را بررسی کرده‌اند، محدودیت‌ها همچنان وجود دارد. رویکردهای موجود به شدت بر ساختارهای نظری متکی هستند که منجر به مدل‌های ساده‌سازی شده و استفاده محدود از داده‌ها می‌شود. این مطالعه به یک شکاف تحقیقاتی مهم می پردازد: عدم وجود یک چارچوب جامع داده محور که داده های مشتری را با تصمیمات فردی و تعاملات اجتماعی برای پیش بینی ریزش در بازارهای مخابراتی یکپارچه می کند. برای پر کردن این شکاف، پس از مطالعه کسب و کار و مجموعه داده، یک مدل مفهومی ایجاد می شود. داده ها برای برآوردن معیارهای مدل از پیش پردازش می شوند. یادگیری ماشین (ML) برای برون یابی ویژگی های گمشده با استفاده از مدل رگرسیون استفاده می شود. مدل در یک ABM پیاده سازی شده است. طبقه بندی ML برای تعیین رفتار ریزش عوامل استفاده می شود. ABM حاصل با داده های واقعی و برون یابی غنی شده و در شبیه سازی استفاده می شود. مدل معتبر برای استفاده پیشرفته‌تر با یک بهینه‌ساز جفت می‌شود و کل فرآیند در یک چارچوب یکپارچه شکل می‌گیرد. نشان داده شده است که این چارچوب در سناریوهای مختلف رفتار ریزش را به درستی تعیین کرده است. همچنین توانایی های خود را در شرایطی نشان داده است که یک رقیب به طور تهاجمی برای افزایش سهم بازار خود حرکت می کند و مدل آن فرموله کردن پاسخ است. پاسخ پیشنهادی نه تنها منجر به بازیابی سهم از دست رفته ۳ درصدی بازار شد، بلکه ۸.۶ درصد دیگر از سهم بازار را بدون آسیب رساندن به سود شرکت به دست آورد. این تحقیق به پیشرفت درک و مدیریت ریزش مشتری در صنعت مخابرات کمک می کند.

کلیدواژه‌ها: مدل‌سازی عامل‌بنیان، یادگیری ماشین، پیش‌بینی انحراف، مدل‌های داده‌بنیان، بازاریابی

---

* نویسنده مسئول مقاله E-mail: mjtarokh@kntu.ac.ir

A data-driven Agent-based model and framework for Churn prediction in Telecommunication Industry

Abstract

Customer churn presents a significant challenge for the telecommunications industry, necessitating effective strategies for prediction and prevention. While prior research has explored diverse methodologies, including Agent-based Modeling (ABM), limitations persist. Existing approaches rely heavily on theoretical constructs, resulting in oversimplified models and constrained data utilization. This study addresses a critical research gap: the absence of a comprehensive, data-driven framework that integrates customer's data with individual decisions and social interactions to predict churn in telecommunication markets. To fill this gap, after a study of the business and the dataset, a conceptual model is established. Data is preprocessed to meet model criteria. Machine Learning (ML) is used for extrapolating missing features using a regression model. The model is implemented into an ABM. ML classification is used to determine agents' churn behavior. The resulting ABM enriched with real and extrapolated data and used in simulations. The validated model is paired with an optimizer for more advanced usage , and the whole process is shaped into a unified framework. This framework has been shown to have correctly determined churn behavior in different scenarios. It has also demonstrated its capabilities in a situation where a competitor aggressively moves to increase its market share, and the model is to formulate a response. The proposed response resulted not only in regaining the lost 3% market share but also gained an additional 8.6% in market share without damaging the company's bottom line. This research contributes to advancing the understanding and management of customer churn in the telecommunications industry.

Keywords: Agent-based-modeling, machine-learning, churn-prediction, data driven models, marketing

1- Introduction

Customer churn is a phenomenon where customers stop doing business with a company, which can have a significant impact on the company's bottom line. Customer churn analysis is a complex system with factors like price, network performance, and customers' threshold of tolerance for performance degradation influencing balancing market share with profitability [1]. The literature has demonstrated that customer churn study is financially beneficial for companies due to several reasons. For one, acquiring new customers costs five times more than retaining existing ones[2]. For the other, losing customers can result in opportunity costs from decreased sales. Therefore, customer retention has recently received significant attention in the research literature.

Agent-based modeling (ABM) can optimize complex systems, providing insights into customer satisfaction and system performance, allowing companies to simulate different scenarios and interventions to identify the most effective strategies for churn reduction. ABM is useful for churn analysis because consumers in a market interact, through processes such as imitation and conditioning, with individuals and groups of individuals or friends, family, etc. In addition to these interactions, the brands' advertising campaigns influence the consumers' choices. The consumers' purchasing or adoption decisions reciprocally influence in the brands' marketing policy. To study consumers' behaviors, interactions between the different actors (consumers and brands) of a given market should be analyzed, too.

However, traditional ABMs often rely on theoretical knowledge and idealized assumptions, introducing limitations in establishing a one-to-one correspondence between the model and reality, thereby limiting predictive capabilities [3–6]. In these models, agent behaviors are derived from a series of rules, and attribute values are initialized arbitrarily using measures such as percentages or probability distributions. These practices introduce challenges in establishing a one-to-one correspondence between the model and reality, limiting predictive capabilities due to the lack of empirical grounding. It also often struggles to represent the complex decision-making processes of individual agents, which can be influenced by various factors such as the environment, imperfect knowledge, and social factors [7].

To bridge this gap and enhance the effectiveness of ABM in churn prevention, this paper introduces a new data driven ABM in which agents no longer rely on theory for decision-making.

In this study first, a review of the existing literature on using data in agent-based modeling was conducted. A framework was then proposed to design a data-driven ABM for use in simulations for churn analysis, addressing the research gap. Finally, the proposed framework was tested using benchmark data from IBM to validate its applicability.

2- Literature Review

The paper's proposal encompasses various subjects, including agent-based modeling, machine learning and churn prediction techniques. For each topic, different approaches are compared, and their benefits and drawbacks for the proposed approach are highlighted.

2-1- Agent based modeling

Agent-based simulation, a third-generation model introduced in the early 2000s, is closely associated with disciplines such as complexity science, systems science, system dynamics, computer science, and management science. This form of simulation builds upon the foundations laid by previous methods like discrete event simulation and system dynamics, offering new capabilities for modeling and understanding complex systems[8].

One of the techniques used by researchers and industry practitioners to understand customer behavior in the marketplace is ABMS, which allows models to become more realistic by incorporating additional characteristics of real systems. This makes ABMS a suitable technique for modeling complex and dynamic environments, such as business environments and marketing strategies. Traditional agent-based models in research often rely on simple rules based on theories, hypotheses, and assumptions. However, this approach has limitations when it comes to modeling real-world systems, as it neglects data that is incompatible with the theory. This happens more often than not in real-world complex systems. In recent years, access to detailed data on almost every subject has been facilitated because of more sophisticated sensors and today's computers' ability to process big data. Also, with more and more companies understanding the importance of big data, high-quality curated datasets are available on many subjects. This has led researchers to use these datasets to create more realistic agent-based models and shift from theory-based approaches toward data-driven ones.

2-2- Data driven agent-based models

Many of such data-driven agent-based models were also reviewed in this research. The objective was to identify methodologies and techniques that can be used to develop a data-driven ABM for churn behavior analysis.

Kennedy et al. [9] developed AIMSS, a simulation assistance system that combines data-driven agent-based modeling with classical ABM. Real-world data is collected simultaneously with model development and compared to simulation output. If there is a deviation, the simulation model and parameters are adjusted to fit real-world data. Hassan et al. [10] have developed an approach incorporating data from surveys, panels, interviews, and official documents. The data sources are primarily qualitative and can affect both the design and simulation phases of the model. The main focus of the data in the model is the agent population and their expected behaviors. The impact of data in this approach is both on changing initial values and modifying the model structure.

Ge et al. [11] developed Virtual City, an agent-based simulation platform to study urban populations, using statistical and geographic data to drive different model components. The population data is used for agents, while geographic data is used to create the environment. This approach generates essential characteristics of the population, including the exact number of individuals, from high-level statistical data. The population generated in this technique has the same statistical attributes (such as age and gender distribution) as the high-level data. The approach only uses quantitative statistical data and lacks realistic agent behaviors, as acknowledged by the authors.

In a study by Sajad et al. [12], Agents' attributes are initialized using census data, but unlike other approaches, data is not used in the model’s design. The model is developed as a classical rule-based ABM, where the changes in agent properties over time are monitored and compared to the corresponding census data. Data is only utilized during simulation and for validation purposes.

The study by Hayashi et al.[13] aimed to propose a method for improving the accuracy of agent-based simulation results using quantitative data. The authors successfully showed how using data helps parameter determination in the modeling process of agent-based simulation. By changing the parameters of the simulation based on the information obtained from the data, the accuracy of the simulation was improved. In this study data was indirectly used for model structure definition. Table 1 presents a summary of the use of data in ABM in the reviewed articles.

Table 1. Literature of using data in ABM

Study	Data utilization	Limitation
[14]	adjust the model and validate the parameters	Agents' behaviours and attribute values are not derived from the data
[10]	modify agent behaviors and structure	Population characteristics and Agents' attribute values are not derived from the data
[15]	generate population characteristics and create the environment	Agents' behaviour is not derived from the data
[12]	initialize agent attributes and validate the model during simulation	Agents' behaviour is not derived from the data
[13]	improve parameter determination and enhance simulation accuracy	Population characteristics and Agents' behaviour are not derived from the data
[16]	generate rules for determining customer behavior	Population characteristics are not derived from the data

The table provides an overview of how each study incorporates data into their agent-based models. Data usage varies across the studies, including data-driven validation, modification of agent behaviors and structure, population generation, initialization and validation of agent attributes, and parameter determination for improved accuracy.

Literature review shows gap in studies that use data for both the simulation environment initialization and the generation of agents' behavior. By leveraging machine learning techniques and incorporating data-driven approaches, our framework can handle the complexity of churn behavior analysis.

2-3- Machine Learning for Churn prediction

Several investigations in customer research have indicated that customer satisfaction is a powerful indicator of customer retention [17]. Awareness of customer perceptions and the quality of the outcomes they desire in receiving services is among the important priorities of all organizations to create satisfaction for them, as evidenced by a study conducted by Zarei[18]. Numerous studies have been conducted to forecast customer churn in the telecommunications industry, specifically using the IBM dataset. These researchers use machine learning techniques for accurate churn prediction [19–22].

Furthermore, the increasing quality of data (more data in finer levels of granularity), both at individual and macro levels, has made machine learning an increasingly suitable tool for improving ABM in churn analysis.

Dehkordi et al.[23] explored how machine learning (ML) techniques address challenges in agent-based modeling and simulation (ABMS). Common challenges involve improving behavioral modeling accuracy, data pre-processing, and computational efficiency. As a result, in the proposed framework, ML techniques are used to fulfill these challenges.

3- Proposed Framework for churn analysis

As uncovered in the literature review, an approach blending data-driven Agent-Based Modeling (ABM) with machine learning holds promise for integrating the strengths of both methods while mitigating their shortcomings. The following sections outline the framework in a step-by-step manner.

3-1- Contextual modeling

Contextual model is a means of communication that describes the details of a model in a form such as text or diagrams [24]. Providing a contextual model enhances researchers' and stakeholders' understanding of the model as well as requirements for model generation. This understanding enables informed decisions regarding the selection, integration, and preprocessing of relevant data, thereby improving the overall quality and effectiveness of the model. A contextual model of our proposed framework is depicted in Figure 1. This model serves as a representation of the key components and their relationships within the framework, and also provides a high-level overview of how various elements interact and contribute to the analysis of churn behavior. The main components of an ABMS are consisted of Data preprocessing and collection[25], finding patterns and agent behaviors using ML techniques[26].

The contextual model consists of several interconnected components, including data collection, preprocessing, agent behavior modeling using machine learning, and an ABMS for churn analysis. Each component plays a vital role in the overall framework and contributes to the comprehensive understanding of churn behavior.

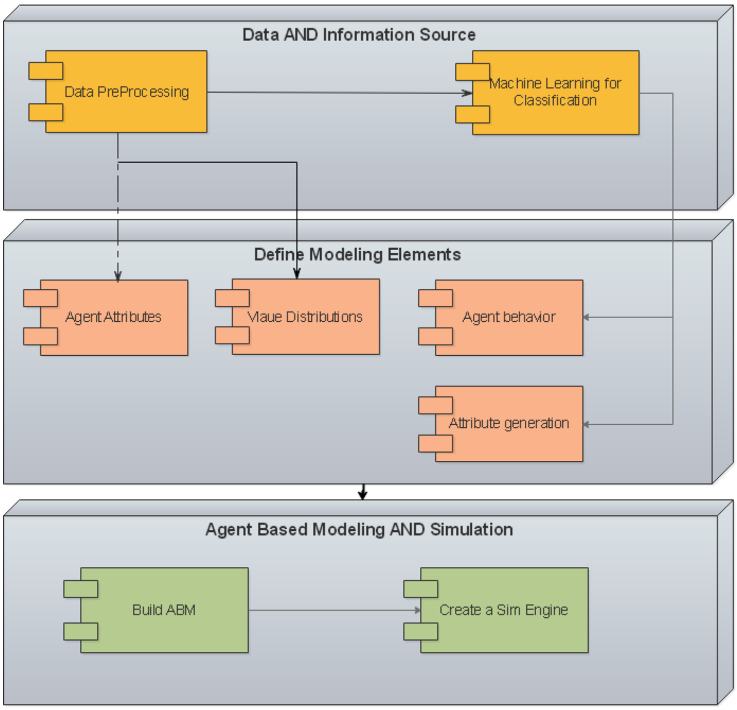

Figure 1. Contextual model of the proposed framework

3-2- Conceptual Modeling

The main objective of this model is to simulate the behavior of individuals in a telecommunications market with two mobile network operators (*MNO*). The simulation aims to predict the future market share, churn rate, and revenue of each *MNO* by employing different marketing strategies. In this simulation, the agents are represented as human agents (Human Agent). To demonstrate the causal relationships between variables and the feedback structure of the system, the best tool is the use of a causal loop diagram[27]. Figure 2 shows how an agent makes the decision to churn or stay.

The Human Agent exists within a closed network environment and possesses static and dynamic attributes. The static attributes include the agent's name, gender, marital status, number of dependencies, internet service usage, and bandwidth usage. These attributes are initialized at the beginning of the simulation using the IBM dataset. On the other hand, the dynamic attributes consist of contact information, age, tenure, churn status, and bandwidth usage.

The agents use *MNO* services such as internet and calls, and they are also influenced by word-of-mouth (WOM) from their contacts as well as advertisements. As mentioned in [28] WOM and advertisement are directly influence each which considered in this research's model implementation.

After a defined period of time, each agent makes a decision whether to churn or not based on parameters affecting the churn decision. The model captures the essential components and relationships within the agent, offering a comprehensive understanding of its decision-making process and behavior in the telecom market. This simulation provides insights into the factors influencing customer churn and helps evaluate the effectiveness of various marketing strategies.

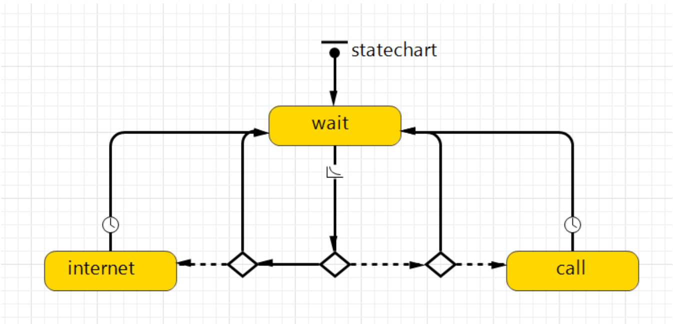

Figure 2. Customer agent behavior conceptual model

3-3- Dataset

The IBM Telco dataset[1] is a valuable resource used in this research. It comprises 21 attributes and 7,044 rows of data, focusing on customer behavior within a telecommunications company. The dataset provides insights into the

[1] TanKY (2020). Telco customer churn: IBM dataset . https://www.kaggle.com/yeanzc/telco-customer-churn-ibm-dataset

relationship between customer actions and churn. The dataset is widely recognized in the research community and has been extensively used by researchers for customer churn prediction. It includes both categorical features, representing object types, and numeric features, representing continuous variables. Using a common dataset makes it possible to compare and benchmark their results. The dataset can be accessed on Kaggle.

3-4- Attribute Model Creation Process

Assigning realistic attribute values will increase the model's correspondence with the real world. There are two ways for assigning values to an agent's attributes from a dataset: Translational Assignment and Algorithmic Assignment. Table 3 describes these methods.

Table 2. Value assignment in ABM model

Assignment Method	Description
Translational	Assigning attribute values through simple translation based on available data
Algorithmic	Determining attribute values using advanced algorithms. Involves cases where there are multiple records per agent for attribute identification. Requires training data for machine learning algorithms or mathematical operations. Ground-truth datasets are obtained by processing agent data for supervised classification/regression. Model training/fitting or creating simple algorithms. Testing for satisfactory fitness and generalizability of the model. Revision of attribute model creation if needed.

To enhance the analysis of churn behavior in the dataset, it was crucial to include the attribute of average monthly internet usage. However, this attribute was not initially available in the dataset. To address this, a linear regression model was employed as part of the algorithmic assignment approach. This model was used to predict the value of average monthly internet usage for each agent in the dataset. Linear regression was chosen as the method for estimating the average monthly internet usage due to its simplicity, interpretability, and ability to capture linear relationships between the independent variables and the dependent variable. The selected features, including gender, age, marital status, number of dependents, internet service type, and tenure in months, were believed to have a significant impact on internet usage patterns. By using linear regression, we aimed to quantitatively analyze the influence of these features and create a predictive model that could accurately estimate the average monthly internet usage for each customer agent. The interpretability of the regression coefficients also allowed us to assess the relative importance of each feature in determining internet consumption, providing valuable insights for understanding churn behavior and customer preferences.

3-5- Linear regression

In linear regression, the relationship between the independent variables (x) and the dependent variable (y) is represented by a linear equation 1:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon \quad (1)$$

Where: y is the dependent variable (average monthly internet usage) β_0 is the y-intercept (bias term) β_1, β_2, ..., β_p are the coefficients (weights) associated with each independent variable (gender, age, marital status, number of dependents, internet service type, tenure in months)

x_1, x_2, ..., x_p are the values of the independent variables ε represents the error term (the difference between the predicted and actual values)

The goal of linear regression is to estimate the coefficients (β_0, β_1, β_2, ..., β_p) that minimize the sum of squared residuals. This is typically done using a method called Ordinary Least Squares (OLS), which involves minimizing cost function Eq (2):

$$J(\beta) = \Sigma(y_i - \hat{y}_i)^2 \quad (2)$$

Where: y_i is the actual value of the dependent variable for the i^{th} observation $\hat{y}_i$ is the predicted value of the dependent variable for the i^{th} observation to estimate the coefficients, the partial derivative of the cost function with respect to each coefficient is calculated and set to zero, resulting in a system of equations. This system can be solved to obtain the optimal values for the coefficients, which minimize the cost function and provide the best fit line for the given data. The R-squared value (r^2) represents the proportion of the variance in the dependent variable that can be explained by the independent variables. It is calculated as Eq (3):

$$r^2 = 1 - ((\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2) \quad (3)$$

Where: $\Sigma(y_i - \hat{y}_i)^2$ is the sum of squared residuals (SSR) $\Sigma(y_i - \bar{y})^2$ is the total sum of squares (SST) $\bar{y}$ is the mean of the dependent variable The mean absolute error (MAE) is a measure of the average absolute difference between the predicted and actual values of the dependent variable and is calculated as Eq (4):

$$MAE = \sum |y_i - \hat{y}_i|/n \quad (4)$$

Where: n is the number of observations

During this process, each feature was assigned a weight to represent its influence on the average monthly usage, and through network training, a linear equation was derived to minimize the mean squared distance between the line and the data points. The model achieved an R-squared value of 0.503099, indicating that approximately 50.31% of the variability in the average monthly internet usage can be explained by the selected features. The mean absolute error was calculated as 11.237254, representing the average absolute difference between the predicted and actual values of the dependent variable. These evaluation metrics provide insights into the performance of the linear regression model in estimating the average monthly internet usage, indicating a moderate level of accuracy and explanatory power. The result of modeling monthly internet and call usage using linear regression method is presented in table 4.

On the other hand, for customer agent characteristics such as age, gender, marriage, tenure in months, number of dependencies, and bandwidth usage, the translational assignment method was employed. This involved a straightforward translation process to assign appropriate attribute values based on the available data. By combining both algorithmic and translational assignment methods, a comprehensive and accurate representation of the dataset was achieved, facilitating a more insightful analysis of churn behavior.

Table 3. Call and internet usage behavior model

	Call usage model	**Net usage model**
Feature	Weight	Weight
Age	0.2	-0.03
Number of Dependents	-0.11	0.03
CLTV	-0.001	0.02
Number of Referrals	-0.06	0.01

	Call usage model	Net usage model
Tenure in Months	0.111	0.015
Avg Monthly GB Download	0.14	0.1
Gender	0.01	0.02
Under 30	-0.6	0.96
Married	0.2	0.03
Bias	0.01	-0.2

3-6- Behavior Model Creation

The creation and testing of behavior models are the central aspects of this study approach, encompassing various sub-processes. These processes result in the development of behavior models that can be effectively employed to execute agent behaviors. To model the churn behavior from the dataset, we utilized a single-layer perceptron algorithm. This algorithm is a type of artificial neural network that consists of input nodes, a weighted sum calculation, an activation function, and an output node. For each feature, we assigned a weight that represents its influence on churn. The perceptron then calculates the weighted sum of these feature values and a bias term. Mathematically, the weighted sum (z) for a given input sample (x) can be calculated as Eq (5).

$$Z = w_1 x_1 + w_2 x_2 + \cdots + w_p x_p + b \quad (5)$$

Where w_1, w_2, ..., w_p are the weights associated with each feature, x_1, x_2, ..., x_p are the corresponding feature values, p is the number of features, and b is the bias term. The perceptron's output is obtained by passing the weighted sum through an activation function. In this case, we used a step function. The step function applies a threshold to the weighted sum, resulting in a binary output. Mathematically, the output ŷ can be calculated as:

$$\hat{y} = step(z) = \{1, if\ z \geq\ threshold |\ 0\ , if\ z\ <\ threshold\}$$

The threshold determines the decision boundary between churn and non-churn instances. The weights and bias terms are adjusted during the training process to minimize the error between the predicted churn status ŷ and the actual churn status y for the training samples. The perceptron learning algorithm updates the weights and bias term iteratively based on the difference between the predicted and actual churn status, aiming to minimize this difference. This process continues until the algorithm converges or reaches a predefined number of iterations. By training the single-layer perceptron using the dataset, we obtained the weights and bias term that define the relationship between the features and churn behavior. This enables us to predict the churn status for new customer instances based on their feature values.

The evaluation of the churn model using various metrics confirms its effectiveness in predicting churn behavior. With an accuracy of 0.92, the model accurately predicted churn in 92% of the instances. The macro average and weighted average precision, recall, and F1-score metrics also achieved a consistent score of 0.92, indicating a well-balanced performance across different classes and datasets. These evaluation results highlight the model's reliability and ability to generalize well. It is important to consider other factors such as business context, cost considerations, and domain expertise when interpreting these metrics. Overall, the churn model demonstrates strong predictive capabilities. At the end of our analysis, we derived the function of churn and determined the importance weight of each feature. This function provides valuable insights into how each attribute affects the churn behavior of individual customers. The importance weight of each feature is summarized in Table 5.

Table 4. Churn behavior function of each customer

Feature	Weight
Number of Referrals	-6.05
Tenure in Months	-3.2
Married	0.9
AVG minutes Call	1.8
Avg Monthly GB Download	3.2
Number of Dependents	2.0
Age	1.41
CLTV	-1.04
Under 30	4.02
Gender	0.97
Bias	-3.0

3-7- Simulation engine

Table 6 provides an overview of various parameters related to the telecom market. Parameters are categorized into groups such as Usage Patterns, Agent Characteristics, Cost and Pricing, Internet Quality and Capacity, Advertisement, and Satisfaction and Churn. Each parameter is described to explain its purpose and relevance in the simulation model.

The framework developed in this study is designed to facilitate a range of simulation experiments aimed at evaluating different marketing strategies and predicting their impact on market dynamics. Our main objective was to determine the most effective marketing strategy for *MNO 1* in response to *MNO 2*'s new marketing campaign. Table 7 summarizes the simulation experiments which can be conducted within this framework.

3-8- Model assumptions

Friedman's methodological credo, "good simulations rely on assumptions that are adequate for their purpose, rather than being descriptively accurate" is pragmatically relevant for agent-based modeling [29].

It is assumed that satisfaction due to QoS changes gradually for each customer agent during their tenure. It is also assumed that this change applies to the general appreciation of the current operator with the coefficient $\beta[0,1]$. In this study, it is assumed to be in the 0.5~0.7 interval. Other initial values of parameters are presented in Table 8.

In order to create a model that reflects reality in terms of sensitivity to bill prices, it is necessary to have an estimation of average talk time and internet usage per user. It is predictable that clients prefer to pay less, and this should show itself in the model if an *MNO* decides to dump prices aggressively to have a more significant market share.

To find the right tariff for talk and internet we have looked into Telecom market insights for 2021 which shows Call tariffs of 0.02~0.10$ for local and national calls and 0.5 ~4$ per gigabyte of internet consumption. Tariffs 0.05$/minute and 0.5$/GB were chosen as first estimates and applied to Monthly Cost based on average internet consumption. As the number of negative talk times were only a few (~300 out of 7024 records) and for simplicity's sake, these numbers were assumed to estimate talk time for all samples int the data. This also means that effects of tariff variations in different plans, VAS and other options are neglected in our study.

3-9- Model validation

Verification refers to the process that examines a model's performance against the intended designed study while validation evaluates to what extend the model explains the real-world system. Following the recommendations of Arifin and Madey[30], the model in this study was verified to implement the designed study illustrated in Fig1.

[30]Given the data collected from a real-world dataset, this study assumed that the agent's choices and attributes are likely to be representative of a community with 7000 customers. The churn behavior of this community is predicted based on extracted rules from the data using machine learning and agent-based modeling.

There are many types of validation for ABM, including requirement validation, data validation, face validation, process validation, theory validation, agent validation, and model output validation. This study captures four types of validations,

Including [31]: ***Data validation***: The data collected to represent agents in this study are based on the IBM telco churn dataset. ***Model validation***: This model is implemented based on real world data not theories. These models were validated by cross-validation technique. ***Model output validation(scenario):*** Broadly speaking, sensitivity analysis can considered as the exploration of a mathematical or numerical model. The model is typically regarded as a black box that processes inputs and calculates one or more quantities of interest (outputs). Thus, the exploration is not performed by a direct inspecting of the model. Instead, the properties of the model are obtained indirectly, by investigating how the output changes given variations in the inputs.

To test this, we can run the model with the same inputs for both operators, except for one parameter that we vary between them. This allows us to predict the outcome based on the parameter difference and then compare the simulation results with our prediction.

For instance, considering the advertising cost, If *MNO1* spends twice as much as *MNO2*, we would expect a positive evolution in market share for the bigger spender, *MNO1* assuming everything else remains equal.

Table 5. Model Parameters

Group	Parameters	Description	Agent
Usage Patterns	Probability Internet	Probability of using the internet when using the mobile	Customer
	Phone Usage Rate Daily	Average number of times the mobile is used per day	Customer
	Internet Duration	Duration of internet usage per session (hours)	Customer
	Call Duration	Duration of call usage per session (hours)	Customer
	Contact Size	Average number of contacts a person interacts with daily	Customer
Characteristics	Scale Factor	Number of people represented by each agent in the model	Main
	InitialOperator1Share	Number of people initially using Operator 1	MNO
	InitialOperator2Share	Number of people initially using Operator 2	MNO
	Total number of people	Total number of people in the agent properties	Main
Cost and Pricing	Cost Call Operator	Cost per minute of call	MNO
	Cost Internet Operator	Cost per gigabyte of internet usage	MNO
	Cost Call Capacity	Cost of increasing one unit of call capacity	MNO
	Cost Bandwidth Capacity	Cost of increasing one gigabyte of bandwidth capacity	MNO
Internet Quality and Capacity	Free Share High Quality	Unused share of Operator's internet capacity for high-quality experience	MNO
	Free Share Medium Quality	Unused share of Operator's internet capacity for medium-quality experience	MNO

Group	Parameters	Description	Agent
	Free Share Low Quality	Unused share of Operator's internet capacity for low-quality experience	MNO
	Total Capacity for Call	Total call capacity of the Operator	MNO
	Total Internet Bandwidth	Total internet bandwidth capacity of the Operator	MNO
Advertisement	Alpha	Coefficient for the impact of advertisements	MNO
	Advertisement Cover Level	Coverage level of individuals per advertisement	MNO
	Advertisement Cost	Cost per advertisement	MNO
	Advertisement Effectiveness	Effectiveness of advertisements in increasing satisfaction	Costumer
Satisfaction and Churn	Beta	Coefficient for satisfaction of QOS	Main
	Gamma	Coefficient for calibrating churn probability	Main

Table 6. Scenarios which model can experienced

Experiment	Description
Impact of Advertising	Examining the influence of advertising on churn behavior by varying advertisement coverage level and cost per advertisement.
Pricing Strategies	Exploring the impact of pricing strategies on churn behavior by manipulating the cost of call and internet services.
Operator Capacity Planning	Investigating the effect of operator capacity planning on churn behavior by adjusting call and internet capacity plans.
Comparative Analysis	Conducting a comparative analysis of different operator-specific factors, such as initial operator shares and service quality levels.
Sensitivity Analysis	Performing a sensitivity analysis to assess the impact of parameter variations on churn behavior and model outcomes.

Table 7. Model parameters' initial values

Probability	Value	Probability	Value
Probability of using internet	*0.6*	High quality internet experience	*0.5*
Phone usage daily rate	*60*	medium quality internet experience	*0.65*
internet usage	*0.1<x<2*	low quality internet experience	*0.85*
Call amount	*0.1<x<2*	Advertisement sequence MNO1	*1-30-1*
Contact size	*5*	Advertisement sequence MNO1	*1*
Higher price dissatisfaction threshold	*0.3*	Campaign level MNO1	*1-5-1*
Advertisement share MNO1	*0.8*	Campaign level MNO2	*5*
Advertisement share MNO2	*0.8*	call capacity increase plan MNO1	*10*
Call tariff MNO1	*1 – 10 - 1*	call capacity increase plan MNO2	*10*
Call tariff MNO2	*10*	call capacity increase plan MNO1	*20*
Internet tariff MNO1	*1 – 10 - 1*	call capacity increase plan MNO2	*20*
Internet tariff MNO2	*10*	High quality internet EXPERIENCE	*0.5*

By comparing the simulation results with this expectation, we can assess the accuracy of the model's prediction.

We repeat this validation process for each input parameter, systematically varying one parameter at a time while keeping the others constant. This allows us to assess the influence of each parameter on the model's output and confirm whether it matches our expectations and real-world outcomes. Once the model has been successfully validated and its behavior matches our predictions, we can run simulations on it. This means using various scenarios, changing multiple parameters simultaneously, and observing the corresponding outcomes.

Simulations provide insights into the system's dynamics and allow us to explore different "what-if" scenarios to understand the consequences of different strategies and interventions.

4- Scenario analysis

Scenario analysis in ABMS can help market strategists determine the method and timing of intervention [37][36][35][33]. Suppose the following scenario: *MNO2* has launched an aggressive campaign to increase its market share in an established market with two providers. *MNO1* needs some time to determine the rival’s campaign’s efficiency and formulate a response with maximal efficiency (taking back the lost market share) with most negligible impact on the bottom-line results (least cost including the loss of potential income). In order for *MNO1* to achieve its previous market share, a scenario was implemented in the designed Agent-Based Model (ABM Sim). The initial values were set, the model was run for a specific duration, and the results were thoroughly analyzed.

Parameters such as the probability of using the internet when using the mobile, average daily phone usage rate, and duration of internet and call usage per session were also considered.

The model was run with the given initial values for 10 months after which the main Key Performance Indicators: “market share”, “customer satisfaction”, “churn rate”, and “revenue” were compared with their initial value. At this point, *MNO1* can respond with four different strategies: **1-** a weaker response with the same mechanics, which is doomed to fail. **2-** a similar response with the same mechanics (shadowing the rival). **3-** a heavy handed response with the same mechanics and **4-** a response using one or more mechanics (including the one its rival has used).

Our initial hunch was for *MNO1* to shadow *MNO2*, bringing the market back to the same point as before *MNO2*’s intervention. However, our simulation showed that this strategy is inadequate and potentially disastrous: Between *MNO2*’s initial manipulation and the start of *MNO1*’s response, *MNO2* has increased its market share without sacrificing customer satisfaction. Although the situation seems similar to the initial conditions, the increased market share generates more positive WOM and contributes to an even grater increase in market share for *MNO2* before the market reaches a new balance point. Thus, it can be hypothesized that shadowing strategy only works if implemented as soon as possible, before the loss in market share becomes significant.

The other common scenario would be for the rivals to start a “price war”, each dumping the price for services with hopes that the rival would lose even more and exit the market before the loss cripples it, with hopes that in a market without competitors, one can fix higher prices and recover the losses endured during the price war. Although this strategy forces marginal/small competitors to exit, it is a lose-lose scenario for contenders in a two-player market with a sizable market share for each competitor. Our model’s simulation results aligned with the theory and showed a significant loss in income for both *MNOs* in the case of a price war.

So, another response for *MNO1* was devised, which included a few different mechanics and more aggressive campaigns. This strategy worked but adjusting the right weight for each action in order to recover the lost market share and increase (or at least minimize the decrease) in income can be an arduous job. We have shown that using optimization techniques (Any logic optimization toolbox), such a response can be formulated faster and with more precision.

4-1- Results

Simulation for shadowing and price-cutting are presented as in month 3 *MNO2* has taken the initiative to reduce prices in order to gain a more significant market share. In both scenarios *MNO1* starts intervening one month after *MNO2’s* manipulation (month 4) and a new balance situation has emerged after some time (9 weeks in shadowing scenario and 7 in price-cutting). Results are in accordance with our presumptions as in the shadowing scenario, *MNO1* continues to lose market share even after a price reduction. The new market share for *MNO2* acts as an engine to encourage *MNO1* subscribers to switch, reaching a balance at a 45/55 *MNO2* lead in market share. In a price-cutting scenario, *MNO1* performs better by losing less market share than *MNO2*, which leads to a 48/52 split. Also, achieving a new balance point takes less time than the shadowing scenario because having fewer churners for *MNO1* results in a lower churn rate for *MNO1* than in the previous scenario. It seems that these corrective actions are inadequate for regaining the market share. Figure 3 depicts the market status in the developed model after the shadowing scenario.

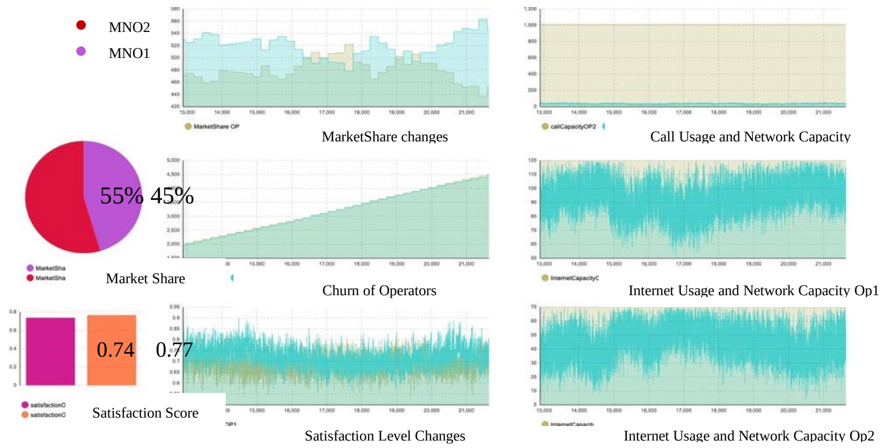

Figure 3. KPI monitoring of the market for shadowing

In our last simulation, we used **AnyLogic**'s optimization engine to formulate a better response with the objective of Maximizing *MNO1's* market share without sacrificing too much revenue. Table 9 shows optimizer's suggestions for *MNO1*.

Table 9. optimizer's suggestions for MNO1

Parameter	Best Value	MNO2 Value
Cost of call	6.75	8
Cost of internet	6	8
Sequence of advertising	6	10
Level of advertisement	4	2

After 8 weeks from applying the suggested strategy, a new balance emerges in market share (58.6/41.4 *MNO1* leads) which means *MNO1* has not only regained the lost market share but also has attracted more of its competitor's subscribers. This means more revenue for *MNO1*.

In conclusion, the implementation and analysis of scenarios in the designed ABM Sim model offered valuable insights and actionable strategies for *MNO1* to enhance its market position. Considering a wide range of parameters, the simulation provided a comprehensive understanding of the market and informed decision-making processes to drive *MNO1's* success in a competitive telecommunications industry.

5- Conclusion and future work

In this research, the gap in utilizing a data-driven approach for ABMS is successfully filled. Our framework overcomes previous agent-based modeling methods' limitations by combining translational and algorithmic approaches, offering a more practical and data-focused solution.

In contrast to approaches proposed in previous research [10,12,14,15,32], the framework outlined here incorporates data in both the Agents' initial attribute and behavior assignment. Compared to the study by Mgbemena and Bell[16] that analyzed churn decisions in the telecom industry using IBM dataset which is not capable of analyzing different market scenarios, our enhanced framework offers an improved tool for scenario analysis, enhancing its utility for decision-makers.

The validation process demonstrated the effectiveness of the framework's data-driven behaviors, as confirmed by cross-validation results, indicating the robustness and reliability of the proposed model. The established global model at the center of this study is validated by expert opinions through a survey that engaged 10 professionals from the industry.

To further demonstrate the practical utility of the framework, a testbed was set up as a hypothetical competitive market with two mobile network operators and a community of customers who interact socially within their network. These interactions influence customers' satisfaction level with their *MNO*, potentially impacting their decision to churn.

The novel framework presented in this research provides enhanced methodologies for conducting scenario analysis, specifically designed to address the need for populations with diverse and complex behavioral patterns, thus offering significant advantages for marketing department.

Validating the study's primary goal of forecasting customers' churn behavior in response to various market interventions like promotions, price reductions, service quality problems, etc. A series of "What-if" scenario analyses were used to test the framework. In this particular scenario, *MNO1* aimed to reclaim lost market share from *MNO2's* campaign. Valuable insights were gained from the framework, indicating that it would not be suitable to employ a shadowing strategy if a competitor launches a campaign in the market earlier, and other actors should consider incorporating dumping into their strategies. The optimizer suggested a course of action for MNO1 to reclaim its market share by executing intelligent and timely tactics such as shortening the ad sequence by 4 days and elevating the ad intensity to level 2, ultimately reducing revenue loss by surpassing the previous market share.

While this study has made a solid contribution in addressing churn evaluation in the telecom market, there are avenues for future research to enhance the framework's versatility and applicability.

1. Realistic Market Scenarios: Extend the framework to encompass realistic market scenarios by incorporating additional factors such as economic conditions, regulatory changes, and technological advancements.

2. Dynamic Parameterization: Investigate adjusting the model's parameters to mirror changing market conditions, enabling a more flexible and responsive framework.

3. Integration of External Data Sources: Investigating the integration of external data sources can enhance the accuracy and relevance of the model's predictions. This may involve incorporating data from social media or customer feedback.

4. Extended Industry Applications: Assess the framework's adaptability for application in other industries beyond telecom, considering its potential in diverse sectors with similar challenges.

By pursuing these avenues, future research can contribute to the ongoing refinement and expansion of our framework, ultimately advancing its effectiveness in capturing the intricacies of real-world market scenarios.

6- References

[1] Dutta A. Business Planning for Network Services: A Systems Thinking Approach. Information Systems Research 2001;12:260–83. https://doi.org/10.1287/isre.12.3.260.9713.

[2] Keramati A, Ardabili SM. Churn analysis for an Iranian mobile operator. Telecommunications Policy 2011;35:344–56.

[3] Bonabeau E. Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences 2002;99:7280–7.

[4] Bruch E, Atwell J. Agent-based models in empirical social research. Sociological Methods & Research 2015;44:186–221.

[5] Heppenstall A, Crooks A, Malleson N, Manley E, Ge J, Batty M. Future developments in geographical agent-based models: Challenges and opportunities. Geographical Analysis 2021;53:76–91.

[6] Kavak H, Padilla JJ, Lynch CJ, Diallo SY. Big data, agents, and machine learning: towards a data-driven agent-based modeling approach. Proceedings of the Annual Simulation Symposium, 2018, p. 1–12.

[7] DeAngelis DL, Diaz SG. Decision-making in agent-based modeling: A current review and future prospectus. Frontiers in Ecology and Evolution 2019;6:237.

[8] Nourozzadeh, F., Shahbazi, M., Karimi, T., Azar, A., & Farzam, S. (2023). Designing a resilience model for the distribution system of petroleum products using an agent-based simulation approach. Innovative Research in Decision Making, 8, 28–56. [in persian]

[9] Kennedy WG. Modelling human behaviour in agent-based models. Agent-based models of geographical systems, Springer; 2011, p. 167–79.

[10] Hassan S, Pavón J, Antunes L, Gilbert N. Injecting data into agent-based simulation. Simulating interacting agents and social phenomena, Springer; 2010, p. 177–91.

[11] Ge Y, Meng R, Cao Z, Qiu X, Huang K. Virtual city: An individual-based digital environment for human mobility and interactive behavior. Simulation 2014;90:917–35.

[12] Sajjad M, Singh K, Paik E, Ahn C-W. A data-driven approach for agent-based modeling: Simulating the dynamics of family formation. Journal of Artificial Societies and Social Simulation 2016;19:9.

[13] Hayashi S, Prasasti N, Kanamori K, Ohwada H. Improving Behavior Prediction Accuracy by Using Machine Learning for Agent-Based Simulation. In: Nguyen NT, Trawiński B, Fujita H, Hong T-P, editors. Intelligent Information and Database Systems, vol. 9621, Berlin, Heidelberg: Springer Berlin Heidelberg; 2016, p. 280–9. https://doi.org/10.1007/978-3-662-49381-6_27.

[14] Kennedy C, Theodoropoulos G, Sorge V, Ferrari E, Lee P, Skelcher C. AIMSS: An Architecture for Data Driven Simulations in the Social Sciences. In: Shi Y, Van Albada GD, Dongarra J, Sloot PMA, editors. Computational Science – ICCS 2007, vol. 4487, Berlin, Heidelberg: Springer Berlin Heidelberg; 2007, p. 1098–105. https://doi.org/10.1007/978-3-540-72584-8_144.

[15] Wang W, Wang X, Ge X, Deng L. Multi-objective optimization model for multi-project scheduling on critical chain. Advances in Engineering Software 2014;68:33–9.

[16] Mgbemena C, Bell D. Data-Driven customer behaviour model generation for agent based exploration. Proceedings of the 49th Annual Simulation Symposium, 2016, p. 1–7.

[17] Baumann C, Elliott G, Burton S. Modeling customer satisfaction and loyalty: survey data versus data mining. Journal of Services Marketing 2012;26:148–57.

[18] Zareei, A. (2021). Designing a structural model of customer churn in state-owned banks (Case study: Selected state-owned banks in Semnan city). Management Research in Iran, 21, 151–176. [in persian]

[19] Beeharry Y, Tsokizep Fokone R. Hybrid approach using machine learning algorithms for customers' churn prediction in the telecommunications industry. Concurrency and Computation: Practice and Experience 2022;34:e6627.

[20] Ebrah K, Elnasir S. Churn prediction using machine learning and recommendations plans for telecoms. Journal of Computer and Communications 2019;7:33–53.

[21] Lian-Ying Z, Amoh DM, Boateng LK, Okine AA. Combined appetency and upselling prediction scheme in telecommunication sector using support vector machines. International Journal of Modern Education and Computer Science 2019;10:1.

[22] Momin S, Bohra T, Raut P. Prediction of customer churn using machine learning. EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing, Springer; 2020, p. 203–12.

[23] Dehkordi MAE, Lechner JM, Ghorbani A, Nikolic I, Chappin EJL, Herder PM. Using machine learning for agent specifications in agent-based models and

simulations: A critical review and guidelines. Journal of Artificial Societies and Social Simulation 2023;26:9.

[24] McKenzie FD. Systems modeling: analysis and operations research. Modeling and Simulation Fundamental: Theoretical Underpinnings and Practical Domains 2010:147–80.

[25] Lamperti F, Roventini A, Sani A. Agent-based model calibration using machine learning surrogates. Journal of Economic Dynamics and Control 2018;90:366–89.

[26] Ratner AJ, De Sa CM, Wu S, Selsam D, Ré C. Data programming: Creating large training sets, quickly. Advances in Neural Information Processing Systems 2016;29.

[27] Eizadiar, M., Talooei Ashlaghi, A., & Mehri, Z. (2021). Application of system dynamics in assessing the performance sustainability of the large-scale supply chain in the automotive industry. Management Research in Iran, 25, 1–24.[in persian]

[28] Faghani, J., Zargar, S., Shirdel, Gh., & Pilevari, N. (2022). Presenting a model for simulating the factors affecting customer acquisition rate using system dynamics approach in the case study of Sepah Bank. Innovative Research in Decision Making, 7, 169–192.[in persian]

[29] Researcher. "A Pragmatic Reading of Friedman’s Methodological Essay." Google Scholar, Accessed November 21, 2023. [Online]. Available: https://scholar.google.com/scholar?hl=fa&as_sdt=0%2C5&q=A+Pragmatic+Reading+of+Friedman%27s+Methodological+Essay&btnG=

[30] Arifin SMN, Madey GR. Verification, Validation, and Replication Methods for Agent-Based Modeling and Simulation: Lessons Learned the Hard Way! In: Yilmaz L, editor. Concepts and Methodologies for Modeling and Simulation, Cham: Springer International Publishing; 2015, p. 217–42. https://doi.org/10.1007/978-3-319-15096-3_10.

[31] Macal CM, North MJ. Agent-based modeling and simulation: ABMS examples. 2008 Winter Simulation Conference, IEEE; 2008, p. 101–12.

[32] Hayashi S, Prasasti N, Kanamori K, Ohwada H. Improving behavior prediction accuracy by using machine learning for agent-based simulation. Asian Conference on Intelligent Information and Database Systems, Springer; 2016, p. 280–9.