ИСПОЛЬЗОВАНИЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЗНАЧИТЕЛЬНОГО СНИЖЕНИЯ ЗАТРАТ В СФЕРЕ ПРИВЛЕЧЕНИЯ ПОЛЬЗОВАТЕЛЕЙ С ПОМОЩЬЮ РЕКЛАМНЫХ КАМПАНИЙ*

УДК 007.51

АЛЕКСАНДР СИМОНЯН
магистрант кафедры управления, бизнеса и туризма
Института экономики и бизнеса
Российско-Армянского университета,
г.Ереван, Республика Армения
aleksandrsimonyan1996@gmail.com

Целью данного исследования является выявление возможностей снижения затрат при привлечении потребителей посредством машинного обучения. В данной научной работе были рассмотрены возможности использования моделей машинного обучения в сфере привлечения новых пользователей, проведен анализ агрегированных данных платформ, занимающихся привлечением новых пользователей, выявлены основные шаги по привлечению пользователей традиционными методами, показаны основные недостатки традиционных методов по привлечению пользователей с помощью рекламных кампаний. На основе проведенного исследования автором предлагается выделить основные элементы по использованию методов машинного обучения для достижения большей эффективности и значительных экономий затрат в сфере привлечения клиентов. В итоге была предложена новая совокупность процессов привлечения клиентов, показаны основные шаги по автоматизации данных процессов. Главными достоинствами предложенной модели по сравнению с традиционными методами привлечения клиентов с помощью рекламных кампаний является высокая скорость, эффективность, значительное снижение расходов, простота имплементации и полная автоматизация.

Ключевые слова: *большие данные, машинное обучение, решение, снижение затрат, рекламная кампания, эффективность.*

Данная исследовательская работа актуальна в силу резкого увеличения объемов инвестиций различных фирм в кампании по привлечению новых пользователей и потребителей. Часто такие инвестиции не приносят желаемого результата и становятся провальными. Различные технологии больших данных позволяют собирать и агрегировать терабайты информации относительно деятельностей потребителей и пользователей. Применение и модификация соответствующих алгоритмов машинного обучения способно

* Հոդվածը ներկայացվել է 20.03.2020թ., գրախոսվել՝ 23.03.2020թ., տպագրության ընդունվել՝ 10.07.2020թ.:

значительно оптимизировать структуры затрат относительно рекламных кампаний по привлечению клиентов и существенно повысить эффективность деятельность фирм.

Содержание и характеристика традиционного процесса привлечения новых пользователей посредством рекламных кампаний

Рассмотрим обыкновенный процесс привлечения новых пользователей для интернет-компаний. Обычно у крупных интернет-фирм всегда имеется свой сегмент рынка. Так, например, для самой крупной компании фоторедактирования Instagram сегментом является пользователи, которые постоянно делятся фотографиями с общественностью; для крупнейшей бизнес-платформы в мире, Linkedin, сегментом является люди, которые хотят расширить круг своих профессиональных знакомств. Безусловно в обоих случаях мы имеем дело с крайне успешными компаниями, у которых число активных пользователей превышает десятки миллионов. Естественно, эти фирмы разработали различные процессы по привлечению новых пользователей. Такие компании научились анализировать огромный рынок интернета и привлекать именно тех пользователей, которые, по их мнению, станут для них успешными (потенциальными) клиентами.

Очень часто крупные фирмы заключают между собой договоры по обмену определенной информацией и запуску рекламных кампании, которые приводят пользователей одной фирмы в другую. Например, в Facebook очень часто можно увидеть компании TikTok. Люди, нажавшие на рекламу, будучи пользователями Facebook, получают возможность установить приложение TikTok. Естественно, Facebook получает за такую услугу деньги от TikTok. Причем деньги поступают за каждую показанную рекламу (Impression) и, естественно, TikTok хочет минимизировать свои расходы и максимизировать количество пользователей, которых привлекут рекламные кампании. Для этого Facebook проводит анализ данных пользователей и показывает рекламу только тем пользователям, которые наиболее склонны к тому, чтобы стать пользователями TikTok. Часто Facebook делится информацией с TikTok для того, чтобы последний провел свой анализ и удостоверился в правдивости происходящего.

В действительности TikTok сотрудничает не только с Facebook, но и с десятками других компаний. Каждая из них проводит свои рекламные акции, в результате чего проходят десятки тысяч рекламных кампаний. Для того, чтобы крупные интернет - компании класса TikTok могли следить за всем происходящим, они обращаются за помощью к обслуживающим фирмам (third party platform)

Существуют множество различных обслуживающих фирм, но одной из самых известных является AppsFlyer. Эта компания представляет различные данные относительно аналитики разных фирм по привлечению пользователей. Информацию о партнерах получить довольно просто: необходимая информация заключается в разделе Partners Integrated Reports[536].

Теперь мы можем подробно рассмотреть процесс привлечения пользователей с помощью рекламных кампаний традиционными методами. Необходимо отметить, что, хотя отдельные этапы этого процесса могут варьироваться у разных фирм, его основные элементы являются общими для всех, так как зачастую фирмы обращаются за привлечением клиентов к одним и тем же компаниям, а аналитику получают от обслуживающих фирм.

[536] AppsFlyer's Integrated Partners: Partners Perspective, https://www.appsflyer.com/product/integrated-partners/.

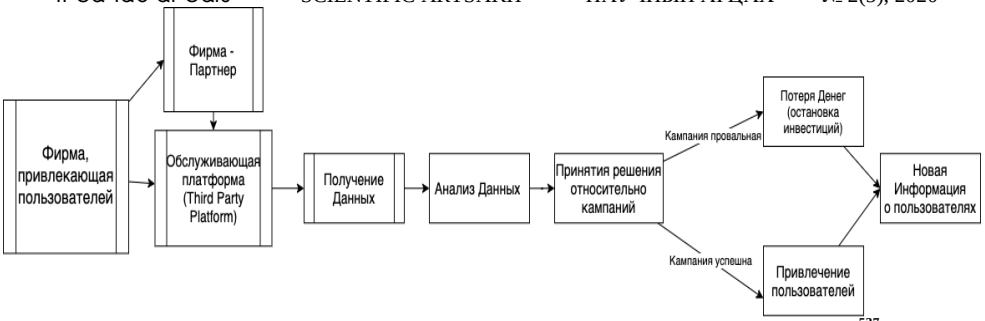

Рисунок 1. Процесс привлечения новых клиентов традиционными способами[537]

На рисунке 1 изображен стандартный процесс привлечения новых пользователей. На перовом этапе фирма заключает договор с организациями-партнерами, с помощью которых она будет привлекать новых пользователей, и обслуживающими фирмами, с помощью которых искомая организация будет получать необходимые данные. Так называемая “Third party platforms” предоставляет всю необходимую информацию о деятельности интернет-компаний, на основе чего фирма, привлекающая пользователей, решает, отставить ли рекламные кампании или удалить их. Естественно, чем дольше действуют рекламные кампании, тем больше денег фирма должна заплатить. Так как же фирма решает, успешна ли рекламная кампания в привлечении пользователей или нет?

Ответ находится в информации, полученной от обслуживающих фирм. После того, как рекламная кампания была запущена, а деньги инвестированы, фирма тщательно следит за результатами рекламных кампаний, а именно, сколько новых пользователей реклама привела, сколько пользователей увидели рекламу, нажали ли они на рекламу, сколько прибыли эти пользователи принесли и т.д. Другими словами, фирма следит за динамикой рекламных кампаний и на основе этого решает оставить кампанию или нет. Обычно процесс анализа, во время которого компания инвестирует в интернет-рекламу, длится две недели.

Изучение и предварительная обработка различных данных и показателей кампаний, полученных с помощью обслуживающих фирм

Мы разобрались с традиционным процессом анализа динамики привлечения новых пользователей и теперь задаемся вопросом, существуют ли методы машинного обучения, которые сделают данный процесс гораздо более эффективным и менее затратным? И если они есть, сможем ли мы их применить?

Для ответа на этот вопрос нам необходимо воспользоваться помощью самого популярного языка программирования в сфере науки данных Python и проанализировать реальную информацию о рекламных кампаниях, полученную с сайта обслуживающей платформы AppsFlyer.

country	platform	country_code	revenue	cost	install_day	media_source	campaign	impressions	clicks	installs	purchases	trials
ROW	android	vg	0.0000	0.0	2020-02-01	Organic	None	0	0	1	0	0
ROW	android	my	33.1755	0.0	2020-02-01	Organic	None	0	0	4366	1	75
ROW	ios	dz	0.0000	0.0	2020-02-01	Organic	None	0	0	180	0	0
ROW	ios	is	0.0000	0.0	2020-02-01	Organic	None	0	0	30	0	1

Рисунок 2. Первоначальная информация о результатах различных кампаний[538]

[537] Составлено автором на основе информации с обслуживающих платформ

На рисунке 2 показаны примеры различных метрик, полученных с обслуживающей платформы. Здесь представлена следующая информация: страна, в которой была показана реклама, платформа (ios/android), сколько была потрачено на рекламу, сколько денег было заработано, количество инсталляций, дата, название рекламы, количество показанных реклам, количество покупок, сколько раз пользователям удалось зайти в платную версию и т.д.

После этого начинается процесс анализа и переработки данных. Целью нашего анализа является выяснить, возможно ли спрогнозировать, что кампания станет успешной через определенной промежуток времени на основе только первых восьми дней ее запуска, а не четырнадцати, как было в традиционном процессе. Таким образом мы сэкономим затраты на 6 дней, и, если наша гипотеза подтвердится, мы сможем сэкономить примерно 42% наших расходов. Важно отметить, что эффективность мы будем рассчитывать, как соотношение дохода к расходам, а кампанию будем считать успешной в том случае, если доход превысит расход.

```
In [28]: op= dff.groupby(['campaign', 'retargeting','media_source', 'agency', 'country', 'platform'], as_index = False)
         .agg(({'install_day' :np.min}))

In [29]: op['minimum_date'] =  pd.to_datetime(op['install_day'])### minimum day

In [30]: op['minimum_date+8'] = op['minimum_date'] + timedelta(days = 8 )

In [31]: op = op.drop(['install_day'], axis =1)

In [32]: ### getting all the data we need
         mmm = pd.merge(dff, op, how = 'inner', on = ['campaign', 'retargeting','media_source', 'agency', 'country', 'platform']

In [33]: mmm['install_date'] = pd.to_datetime(mmm['install_day'])

In [34]: ##### filtering only first7 days performance
         temp = mmm[mmm["install_date"]<=mmm["minimum_date+8"]]

In [35]:
         campaign', 'retargeting', 'media_source', 'agency', 'country', 'platform'], as_index = False).
         ,'cost': np.sum, 'impressions':np.sum, 'clicks': np.sum,  'purchases':np.sum, 'installs': np.sum, 'cohortd7r': np.sum}

In [36]: final_training_data.head()
```

Рисунок 3. Манипуляция первоначальной информации с помощью Python[539]

На рисунке 3 изображен процесс переработки данных. На первом этапе мы фильтруем только ту информацию, которая показывает агрегацию результатов первых 8 дней действия кампаний. Мы повторяем этот процесс для первых дней, чтобы выяснить возможно ли на основе первых 8 дней предугадать успех дальнейших 100 дней. После этого начинается процесс анализа информации.

На рисунке 4 показано количество компаний, сгруппированных по различным странам мира. Мы видим, что США и остальной мир (ROW) занимают первые места и в США показано почти столько же компаний, как и в остальном мире. В остальных крупных странах показано почти одно и тоже количество компаний.

[538] Appsflyer SDK example on GitHub. https://github.com/AppsFlyerSDK/segment-appsflyer-ios

[539] Составлено автором на основе информации, полученной из AppsFlyer Open Source Data, https://github.com/AppsFlyerSDK/segment-appsflyer-ios

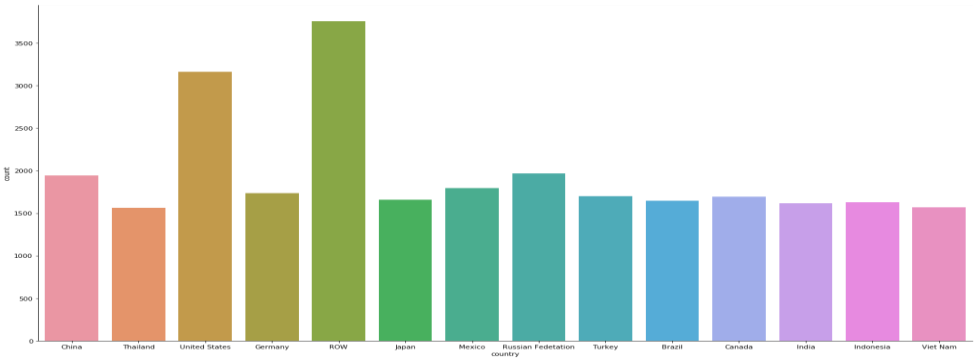

Рисунок 4. Количество показанных компаний, сгруппированных по различным странам мира[540]

На рисунке 5 мы видим, что, несмотря на что, есть корреляция между расходами и доходами, она не ярко выражена и вложенные деньги далеко не единственный фактор, влияющий на успех компании.

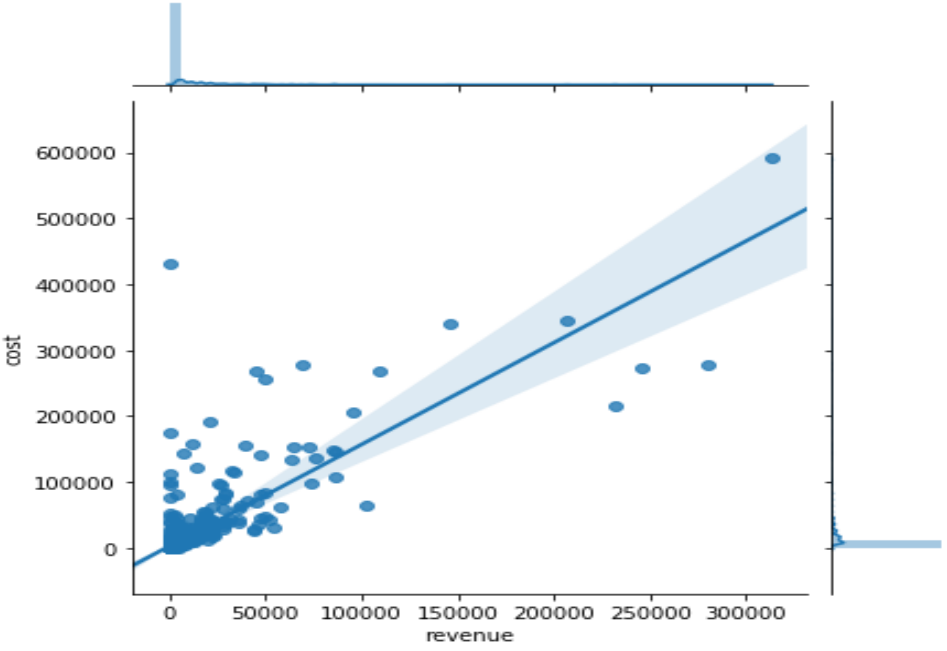

Рисунок 5. График зависимости между расходами и доходами рекламных кампаний [541]

Для достижения лучшего понимания того, с какой информацией мы имеем дело, мы определим новую метрику, ROAS, как соотношение между доходом компании и ее расходом, после чего проведем подробный анализ данной метрики.

[540] Составлено автором с помощью пакетов визуализации Питона
[541] Составлено автором с помощью пакетов визуализации Питона.

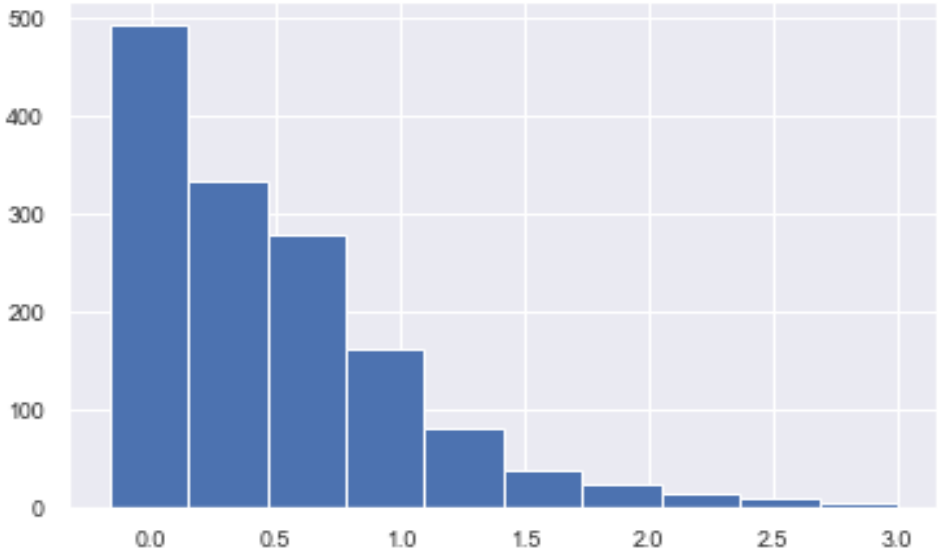

Рисунок 6. Гистограмма количества компаний по метрике ROAS[542]

На рисунке 6 показано количество компаний по ROAS. Мы видим, что чем выше ROAS, тем меньше число компаний, попавших в соответствующую группу. Тем не менее, более половины компаний смогли пройти ROAS в 50%. Безусловно, информация подается классификации и может быть оптимально прогнозируема.

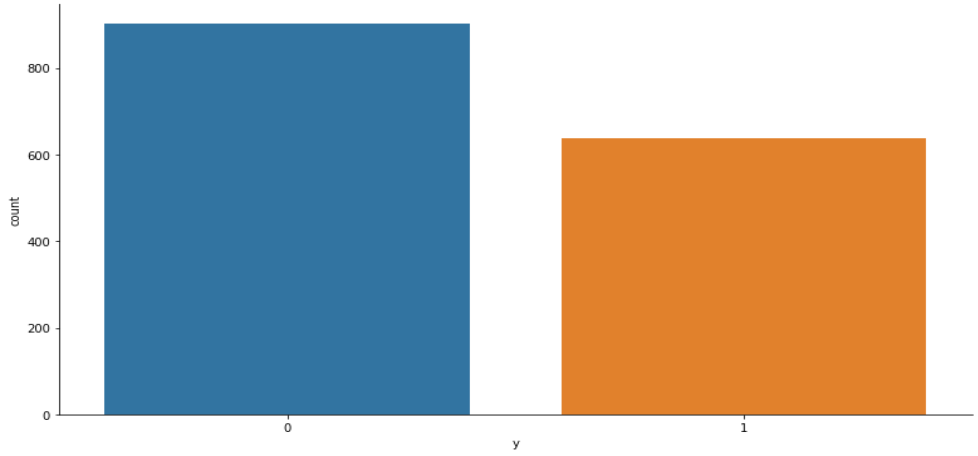

Рисунок 7. Соотношение успешных и провальных рекламных кампаний[543]

На рисунке 7 показано соотношение успешных и провальных рекламных кампаний. 0 показывает количество провальных. Видно, что провальные рекламы значительно превышают по количеству успешные.

[542] Составлено автором с помощью пакетов визуализации Питона.
[543] Составлено автором с помощью пакетов визуализации Питона.

Построение модели машинного обучения для оптимизации кампаний и анализ полученных результатов

Мы закончили анализировать информацию, которая у нас имеется, теперь мы с помощью алгоритма машинного обучения построим и получим результаты того, как модель будет прогнозировать течение кампании только на основе первых 8 дней. Мы используем мощную модель машинного обучения XGBoost, которая успешно справляется с проблемами классификации. XGBoost справляется с проблемами и overfitting и с underfitting, именно поэтому она считается успешной[544].

После построения модели, ее оттачивания и использования самых точных параметров, мы получили следующие результаты.

```
In [166]: print(classification_report(predict,y_test))
```

	precision	recall	f1-score	support
0	0.90	0.81	0.85	317
1	0.66	0.79	0.72	146
accuracy			0.81	463
macro avg	0.78	0.80	0.79	463
weighted avg	0.82	0.81	0.81	463

Рисунок 7. Classification Report, построенной на основе модели XGBoost [545]

Из рисунка 7, видно, что мы получили довольно хорошие данные и в среднем с точностью до 80% можем прогнозировать, будет ли кампания успешна или нет, только на основе первых 7-8 дней. Безусловно, этот подход поможет компаниям, работающим на привлечение пользователей, сэкономить около 40% дополнительных расходов, а также предоставит дополнительное время для нового маневра по привлечению пользователей.

Важно также осознавать, какие факторы влияют на успех кампании. С помощью оценки независимых переменных мы можем дать ответ на этот вопрос.

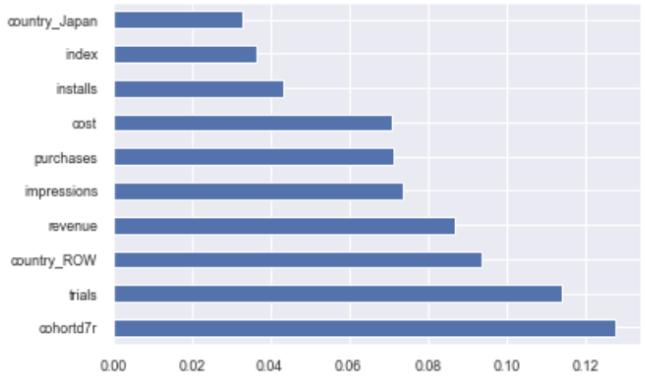

Рисунок 8. Относительная важность независимых переменных построенной модели XGBoost [546]

[544] A Gentle Introduction to XGBoost for Applied Machine Learning, https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/.

[545] Scikit Learn: Classification Report, https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html.

На рисунке 8 представлена сравнительная важность переменных. Самым важным фактором является прибыль, полученная в первые семь дней начала рекламы. Однако следует заметить, что разница между сравнительной важностью элементов не является высокой, и было бы неправильно выделять отдельные элементы в прогнозировании успеха; вместо этого нужно смотреть на совокупность факторов и на то, как эта совокупность взаимодействует.

Современные технологии позволяют повысить эффективность традиционных бизнес-процессов самых разных направлений и сфер. Привлечение пользователей на основе интернет- рекламы не является исключением. На сегодняшний день инновационные подходы меняют традиционное представление операций. Используя большие данные, компании могут прийти к оптимальному решению и использовать все возможности, которые открыты перед ними.

ՄԵՔԵՆԱՅԱԿԱՆ ՈՒՍՈՒՑՄԱՆ ՄՈԴԵԼՆԵՐԻ ՕԳՏԱԳՈՐԾՈՒՄԸ՝ ՈՐՊԵՍ ԳՈՎԱԶԴԱՅԻՆ ԱՐՇԱՎՆԵՐԻ ՄԻՋՈՑՈՎ ՕԳՏԱԳՈՐԾՈՂՆԵՐԻ ՆԵՐԳՐԱՎՄԱՆ ԾԱԽՍԵՐԸ ԷԱՊԵՍ ԿՐՃԱՏԵԼՈՒ ՄԻՋՈՑ

ԱԼԵՔՍԱՆԴՐ ՍԻՄՈՆՅԱՆ

Հայ-ռուսական համալսարանի բիզնեսի և տնտեսագիտության ինստիտուտի կառավարման, բիզնեսի և վբոսաշրջության ամբիոնի մագիստրանտ, ք.Երևան, Հայաստանի Հանրապետություն

Հոդվածի նպատակը մեքենայական ուսուցման միջոցով սպառողների ներգրավման ծախսերի նվազեցման հնարավորությունների բացահայտումն է: Աշխատանքում դիտարկվել են նոր օգտատերերի ներգրավման ոլորտում մեքենայական ուսուցման մոդելների կիրառման հնարավորությունները, կատարվել է նոր օգտատերերի ներգրավմամբ զբաղվող հարթակների ագրեգացված տվյալների վերլուծություն, բացահայտվել են ավանդական մեթոդներով օգտատերերի ներգրավման հիմնական քայլերը, մատնանշվել են գովազդային արշավների միջոցով օգտատերերի ներգրավման ավանդական մեթոդների հիմնական թերությունները: Կատարված հետազոտության հիման վրա հեղինակի կողմից առաջարկվում է ընդգծել մեքենայական ուսուցման մեթոդների օգտագործման հիմնական տարրերը՝ հասնելու առավել մեծ արդյունավետության և հաճախորդների ներգրավման ոլորտի ծախսերի շոշափելի կրճատման: Արդյունքում առաջարկվել է հաճախորդների ներգրավման գործընթացների նոր ամբողջություն, ցուցադրվել են տվյալ գործընթացների ավտոմատացման հիմնական քայլերը: Հաճախորդների ներգրավման առաջարկված մոդելի գլխավոր արժանիքներն են բարձր արագությունը, ծախսերի էական նվազեցումը, իմպլեմենտավորման պարզությունը և ամբողջական ավտոմատացումը՝ գովազդային արշավների միջոցով հաճախորդների ներգրավման ավանդական մեթոդների հետ համեմատությամբ:

Հիմնաբառեր՝ *մեծ տվյալներ, մեքենայական ուսուցում, որոշում, ծախսերի նվազում, գովազդային արշավ, արդյունավետություն:*

[546] Составлено автором.

USING MACHINE LEARNING MODELS TO SIGNIFICANTLY REDUCE COSTS IN ATTRACTING USERS THROUGH ADVERTISING CAMPAIGNS

ALEKSANDR SIMONYAN
Russian-Armenian University,
The Institute of Economics and Business,
Management, Business and Tourism Department, Master Student,
Yerevan, Republic of Armenia

The aim of the scientific work is to identify opportunities to reduce the cost of attracting consumers through machine learning. In this scientific work, the possibilities of using machine learning models in the field of attracting new users are examined, the aggregated data of platforms engaged in attracting new users are analyzed, the main steps to attract users by traditional methods are identified, and the main disadvantages of traditional methods for attracting users using advertising campaigns are shown. Based on the study, the author proposes to highlight the main elements on the use of machine learning methods to achieve greater efficiency and significant cost savings in attracting customers. As a result, a new set of processes for attracting customers is proposed, the basic steps for automating these processes are shown. Thus the main advantages of the proposed customer acquisition model are high speed, efficiency, significant cost reduction, ease of implementation and full automation compared to traditional methods of attracting customers through advertising campaigns.

Keywords*: Big Data, Machine Learning, decision, cost decrease, advertising campaigns, efficiency.*