

УДК: 004.85, 339.138

АНАЛИЗ ЭФФЕКТИВНОСТИ МАРКЕТИНГОВОЙ КАМПАНИИ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Зубарев К.М.

Старший преподаватель кафедры

*Московский государственный технический университет имени Н. Э.
Баумана,*

Москва, Россия

Безрученко Т.С.

студент

*Московский государственный технический университет имени Н. Э.
Баумана,*

Москва, Россия

Аннотация. Данная работа посвящена анализу эффективности маркетинговой компании с использованием методов машинного обучения. В статье рассматривается применение алгоритмов машинного обучения для оценки результатов маркетинговых кампаний, прогнозирования поведения потребителей и оптимизации маркетинговых стратегий. Исследование включает в себя анализ данных о рекламных активностях, потребительском спросе. Для анализа используются модели линейной регрессии и классификации. Для оценки качества построенных моделей используется расхождение Кульбака-Лейблера. Задача классификации решается на основе коэффициента ROMI. Авторы работы определили наиболее эффективные каналы продвижения продукта. Полученные результаты позволят компаниям улучшить эффективность своих маркетинговых усилий и повысить ROMI.

Ключевые слова: модель линейной регрессии, эффективность маркетинга, ROMI, модель классификации, машинное обучение.

***ANALYSIS OF THE EFFECTIVENESS OF A MARKETING CAMPAIGN
USING MACHINE LEARNING METHODS***

Zubarev K. M.

Senior Lecturer

Bauman Moscow state technical University,

Moscow, Russia

Bezruchenko T.S.

student

Bauman Moscow state technical University,

Moscow, Russia

Abstract. This work is devoted to analyzing the effectiveness of a marketing company using machine learning methods. The article discusses the use of machine learning algorithms to evaluate the results of marketing campaigns, predict consumer behavior and optimize marketing strategies. The research includes analysis of data on advertising activities and consumer demand. Linear regression and classification models are used for analysis. To assess the quality of the constructed models, the Kullback-Leibler divergence is used. The classification problem is solved based on the ROMI coefficient. The authors of the work identified the most effective channels for promoting the product. The results will allow companies to improve the effectiveness of their marketing efforts and increase ROMI.

Key words: linear regression model, marketing effectiveness, ROMI, classification model, machine learning.

Введение

В современном мире маркетинг играет ключевую роль в успехе бизнеса. С развитием технологий и увеличением объема данных, компании все чаще обращаются к методам машинного обучения для анализа и оптимизации своих маркетинговых кампаний [1,6,8]. Машинное обучение позволяет не только более точно прогнозировать поведение потребителей, но и выявлять скрытые закономерности и тенденции, которые могут быть упущены при традиционных методах анализа [2,3,7].

Цель данной статьи – рассмотреть основные методы машинного обучения, применяемые для анализа эффективности маркетинговых кампаний, а также продемонстрировать их практическое применение на реальных примерах. В статье будут рассмотрены такие методы, как классификация и линейная регрессия [11,12,13,16].

Особое внимание будет уделено тому, как эти методы помогают улучшить стратегию маркетинга, повысить возврат на инвестиции в маркетинг (ROMI).

Все эти аспекты в совокупности позволят компаниям принимать более обоснованные решения в области маркетинга, оптимизировать свои бюджеты и достичь более высоких показателей эффективности своих маркетинговых усилий [1,9,10,14].

Прогноз продаж по модели линейной регрессии

Рассматривается выборка данных маркетинговой компании, которая имеет вид:

$$D = (\bar{x}^i, y^i), \quad (1)$$

где D — обучающая выборка;

\bar{x}^i — переменные признаков (направления реклам);

y^i — переменная отклика (величина общей прибыли с реклам).

Модель линейной регрессии будет иметь следующий вид:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon, \quad (2)$$

где y — переменная отклика;

β_0, \dots, β_6 — коэффициенты регрессии, которые следует вычислить;

x_1, \dots, x_6 — переменные признаков;

ε — ненаблюдаемая случайная величина, распределенная по нормальному закону.

В данных о маркетинговой компании направления реклам разделены на 6: TV, Billboards, Google_Ads, Social_Media, Influencer_Marketing и Affiliate_Marketing. За данные о величине продаж отвечает вектор данных Product_Sold. Считаем, что величина Product_Sold линейно зависит от 6 направлений реклам [15].

	TV	Billboards	Google_Ads	Social_Media	Influencer_Marketing	Affiliate_Marketing	Product_Sold
0	281.42	538.80	123.94	349.30	242.77	910.10	7164.0
1	702.97	296.53	558.13	180.55	781.06	132.43	5055.0
2	313.14	295.94	642.96	505.71	438.91	464.23	6154.0
3	898.52	61.27	548.73	240.93	278.96	432.27	5480.0
4	766.52	550.72	651.91	666.33	396.33	841.93	9669.0
...
295	770.05	501.36	694.60	172.26	572.26	410.56	6851.0
296	512.38	250.83	373.78	366.95	987.14	509.03	6477.0
297	998.10	858.75	781.06	60.61	174.63	213.53	6949.0
298	322.35	681.22	640.29	343.65	534.22	648.71	7737.0
299	476.45	884.77	301.79	802.30	434.82	401.69	8187.0

300 rows x 7 columns

Рис. 1 — Иллюстрация данных маркетинговой компании [15]

После очистки данных от выбросов, датасет уменьшился на 9 позиций, что составляет 3% от изначального количества данных.

	TV	Billboards	Google_Ads	Social_Media	Influencer_Marketing	Affiliate_Marketing	Product_Sold
0	281.42	538.80	123.94	349.30	242.77	910.10	7164.0
1	702.97	296.53	558.13	180.55	781.06	132.43	5055.0
2	313.14	295.94	642.96	505.71	438.91	464.23	6154.0
3	898.52	61.27	548.73	240.93	278.96	432.27	5480.0
4	766.52	550.72	651.91	666.33	396.33	841.93	9669.0
...
286	708.03	846.72	425.30	555.71	208.48	144.39	6814.0
287	770.05	501.36	694.60	172.26	572.26	410.56	6851.0
288	512.38	250.83	373.78	366.95	987.14	509.03	6477.0
289	322.35	681.22	640.29	343.65	534.22	648.71	7737.0
290	476.45	884.77	301.79	802.30	434.82	401.69	8187.0

291 rows x 7 columns

Рис. 2 — Иллюстрация данных маркетинговой компании после очистки от выбросов [15]

Коэффициенты линейной регрессии подбираются при помощи минимизации суммы квадратов отклонений [6,7,8]. Поиск минимума реализуется методом сопряженных градиентов [2,3,4,5].

Решение задачи регрессии дало нам следующие результаты:

Коэффициенты регрессии:
Признак 1: 1.9981724243200742
Признак 2: 2.9995685933764813
Признак 3: 1.4952604361464594
Признак 4: 2.501356152359082
Признак 5: 1.1959724091212989
Признак 6: 3.9989495343549386
Свободный коэффициент: 2.967365658514609
Расхождение Кульбака-Лейблера: 0.8781798078016436

Рис. 3 — Иллюстрация значений ключевых показателей задачи регрессии. Авторская разработка.

Высокое влияние на прибыль оказывает признак 6 (Affiliate Marketing), в то время как Influencer Marketing имеет минимальное влияние. Расхождение Кульбака-Лейблера на уровне 0.878 указывает на значительные различия между распределениями данных [4,5,7]. Это значение может быть полезным для дальнейших исследований, но требует осторожности из-за возможных ошибок в данных.

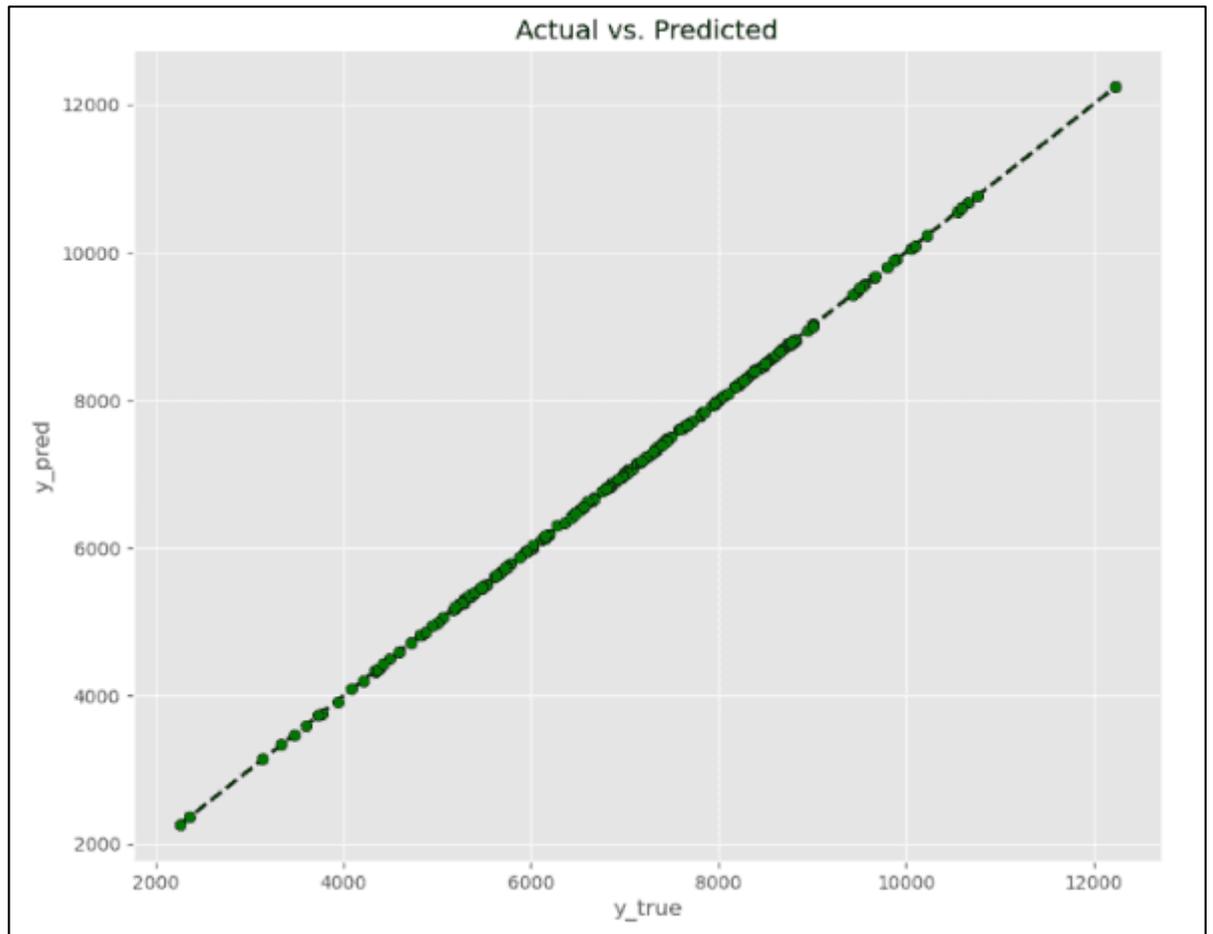


Рис. 4 — Иллюстрация графика зависимости между предсказанными значениями величины прибыли и фактическими значениями. Авторская разработка.

На графике, отражающем зависимость между фактическими и предсказанными значениями (рис. 4), наблюдается хорошая точность модели. Предсказания, сделанные моделью, тесно соответствуют реальным данным, что указывает на эффективность ее работы. Узкий диапазон разброса между предсказанными и фактическими значениями свидетельствует о стабильности и надежности модели. Это говорит о том, что модель хорошо улавливает закономерности в данных и способна делать точные прогнозы. Наличие минимальных расхождений между предсказаниями и реальными значениями подтверждает адекватность модели для данного набора данных. Такие результаты указывают на высокую степень соответствия модели реальным

данным и ее способность к точному прогнозированию. Это позволяет сделать вывод о том, что модель обладает хорошей предсказательной способностью и может быть использована для прогнозирования целевой переменной с высокой точностью.

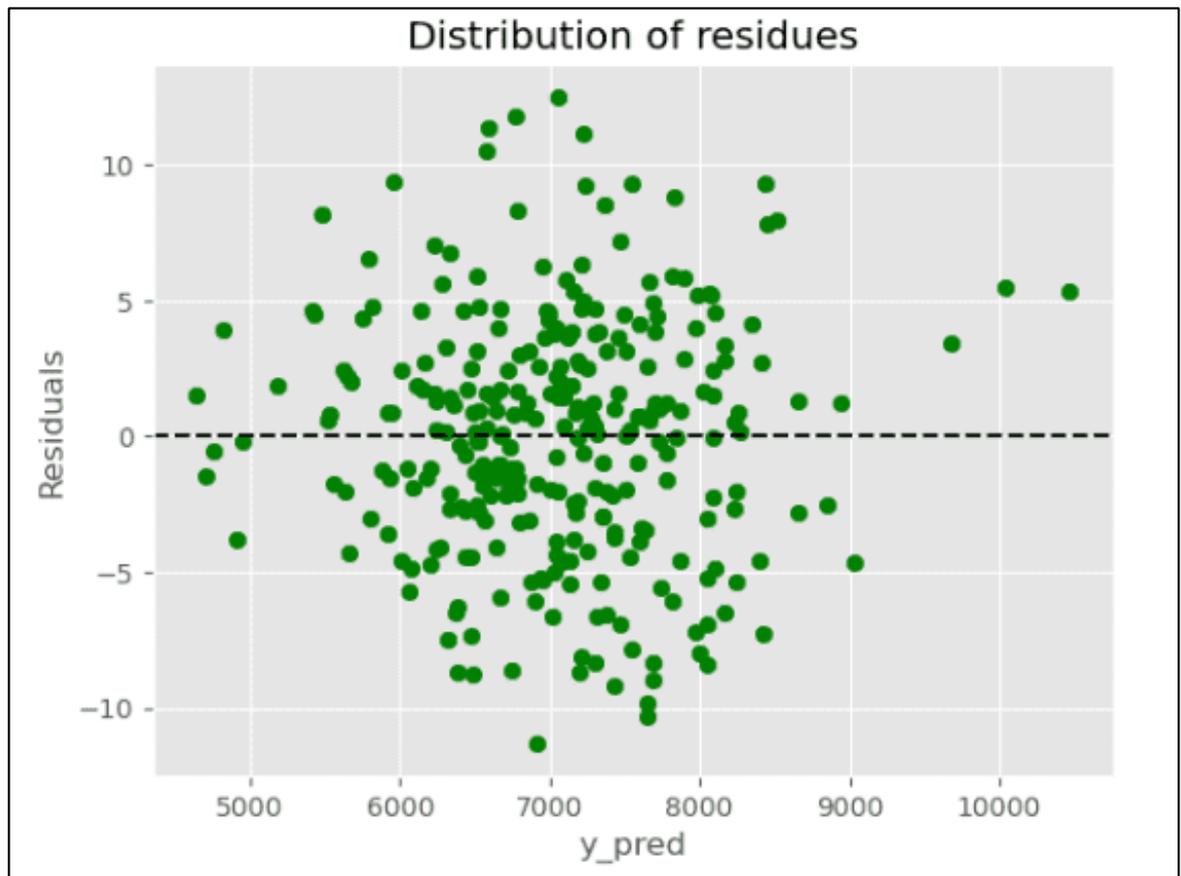


Рис. 5 — Иллюстрация графика остатков модели. Авторская разработка.

Из анализа графика остатков (рис.5) можно сделать вывод о том, что модель достаточно хорошо улавливает зависимости в данных, поскольку не наблюдается явных трендов или систематических отклонений. Диапазон разброса остатков остается умеренным, что свидетельствует о стабильности модели и отсутствии значительных выбросов. Общее распределение остатков модели кажется нормальным, с нулевым средним значением, что указывает на то, что модель хорошо справляется с предсказанием целевой переменной. Наличие нормально распределенных остатков свидетельствует о том, что

Дневник науки | www.dnevniknauki.ru | СМИ ЭЛ № ФС 77-68405 ISSN 2541-8327

модель хорошо подходит для данных и не содержит систематических ошибок. Эти признаки указывают на то, что модель является надежной и эффективной для прогнозирования целевой переменной в данном контексте.

Однако, несмотря на вышеперечисленные результаты, решение задачи регрессии как задачи классификации поможет упростить анализ и повысить точность прогнозирования. Важно провести дополнительные проверки и корректировки данных перед применением этого подхода. Адекватный выбор метода анализа данных с учетом их специфики поможет избежать искажений результатов и повысить достоверность выводов.

Применение модели классификации при анализе эффективности

Задача предсказания объёма продаж является классической задачей регрессии, но погрешность прогноза при прогнозировании может быть достаточно велика. На практике для анализа эффективности маркетинга используется показатель ROMI (Return on Marketing Investment) — это коэффициент возврата маркетинговых инвестиций. ROMI считают одним из самых показательных KPI в маркетинге. Еще ROMI называют универсальной метрикой, ведь на ее расчет не влияют различные погрешности — например, бухгалтерские расходы или сезонность спроса.

$$ROMI = \frac{R - C}{C}, \quad (53)$$

где R — величина выручки (Revenue);

C — величина затрат на маркетинговые стратегии (Costs).

В датасете, используемом ранее [15], каждому элементу при своем классе опираясь на значение коэффициента ROMI, следующим образом:

- Класс 0, низкая маркетинговая эффективность (величина убытков превалирует над прибылью): $-\infty < ROMI \leq 0.9$;

- Класс 1, нулевая маркетинговая эффективность (величина прибыли незначительна, а убытки приемлемы для компании): $0.9 < ROMI \leq 1.3$;
- Класс 2, высокая маркетинговая эффективность (величина прибыли преобладает над убытками): $1.3 < ROMI < +\infty$.

Значение коэффициента отображено на рисунке 6

	TV	Billboards	Google_Ads	Social_Media	Influencer_Marketing	Affiliate_Marketing	Product_Sold	Total Costs	ROMI
0	281.42	538.80	123.94	349.30	242.77	910.10	7164.0	2446.33	1.93
1	702.97	296.53	558.13	180.55	781.06	132.43	5055.0	2651.67	0.91
2	313.14	295.94	642.96	505.71	438.91	464.23	6154.0	2660.89	1.31
3	898.52	61.27	548.73	240.93	278.96	432.27	5480.0	2460.68	1.23
4	766.52	550.72	651.91	666.33	396.33	841.93	9669.0	3873.74	1.50
...
286	708.03	846.72	425.30	555.71	208.48	144.39	6814.0	2888.63	1.36
287	770.05	501.36	694.60	172.26	572.26	410.56	6851.0	3121.09	1.20
288	512.38	250.83	373.78	366.95	987.14	509.03	6477.0	3000.11	1.16
289	322.35	681.22	640.29	343.65	534.22	648.71	7737.0	3170.44	1.44
290	476.45	884.77	301.79	802.30	434.82	401.69	8187.0	3301.82	1.48

291 rows x 9 columns

Рис. 6 — Иллюстрация данных маркетинговой компании после переклассификации объектов по индексу ROMI. Авторская разработка.

Результаты решения задачи классификации представлены на рисунке 7.

Расхождение Кульбака-Лейблера: 0.0260692156560991
 Кол-во объектов класса 0 по ROMI: 6
 Кол-во объектов класса 1 по ROMI: 114
 Кол-во объектов класса 2 по ROMI: 171

Рис. 7 — Иллюстрация значений ключевых показателей задачи классификации. Авторская разработка.

Теперь величина РКЛ составляет 0.026, что говорит об улучшении предсказательной точности модели. Низкоэффективных стратегий 2%,

нулевая эффективность у 39% маркетинговых стратегий и высокоэффективных стратегий — 59%.

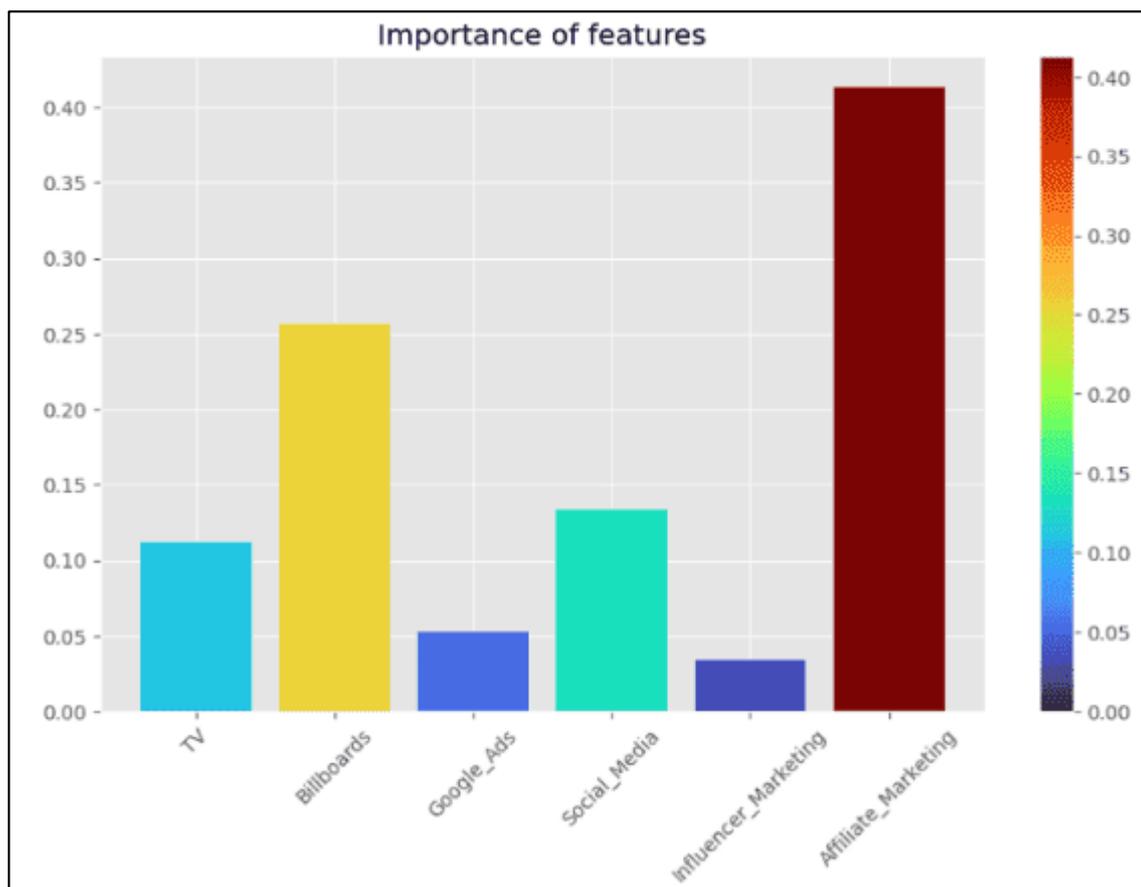


Рис. 8 — Иллюстрация графика важности признаков Авторская разработка.

На рис.12 представлен график, отражающий величину важности различных признаков. Самое большое влияние на величину продаж оказывает партнерская реклама. В задаче классификации, так же, как и в задаче регрессии менее всего влияет реклама у лидеров мнений. Говоря в общем, признаки с низкой важностью могут быть менее информативными или даже шумовыми. Поэтому исключение наименее влиятельного признака может улучшить производительность модели и упростить ее интерпретацию.

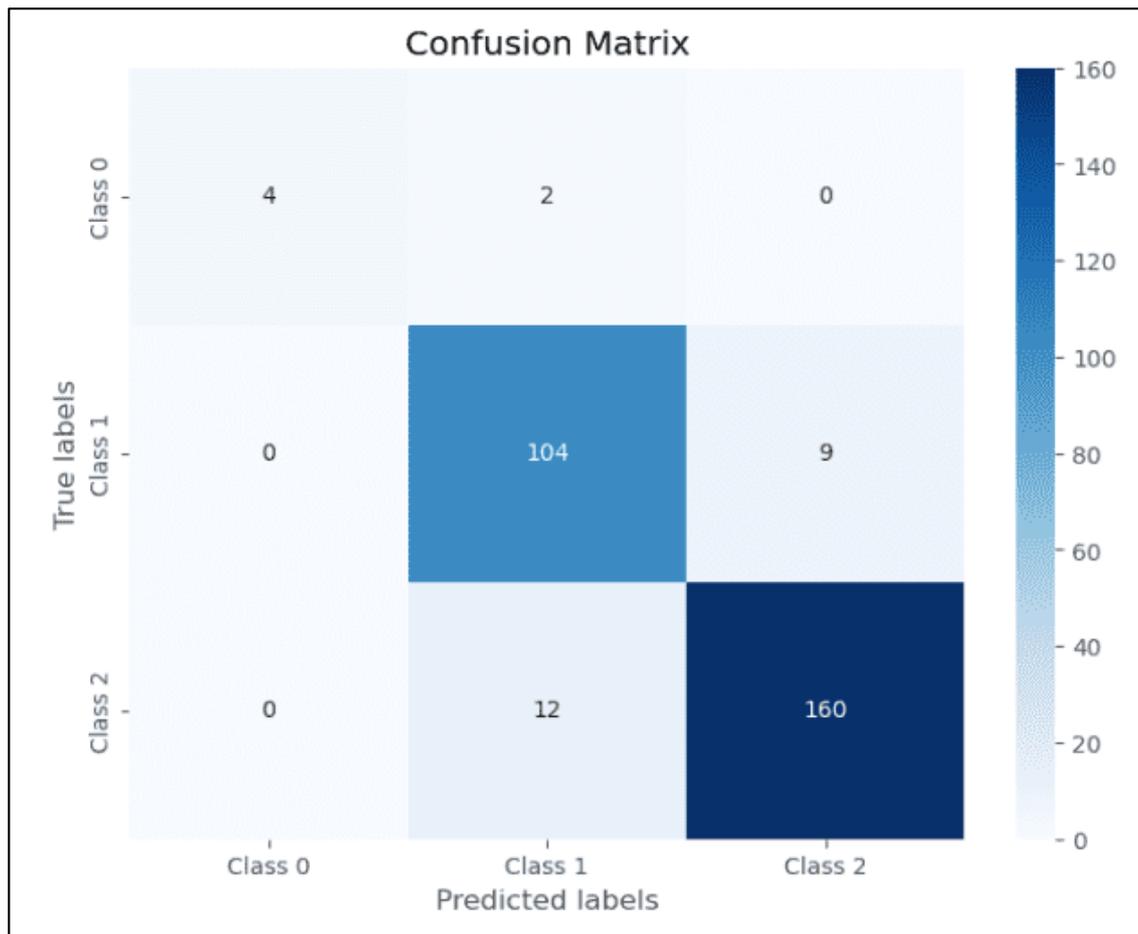


Рис. 9 — Иллюстрация матрицы ошибок. Авторская разработка.

График матрицы ошибок (рис. 9) отражает эффективность алгоритма классификации, то есть какие классы модель предсказала хорошо, а какие недостаточно точно. На диагоналях матрицы расположены кол-ва правильно предсказанных элементов.

Таким образом, 4 элемента из класса 0, 104 элемента из класса 1 и 160 элементов из класса 2 были предсказаны верно. Количество неправильно предсказанных положительных примеров (истинный класс 0, который отражает низкую маркетинговую эффективность, а модель определила этот класс, как класс 1, который отражает нулевую маркетинговую эффективность) составило 2 элемента, это менее 1% от общего количества стратегий. Также количество неправильно предсказанных положительных примеров (истинный класс 1, а модель определила этот класс, как класс 2, который отражает

высокую маркетинговую эффективность) составило 9 элементов, это 3% от общего кол-ва стратегий. И наконец, количество неправильно предсказанных отрицательных примеров (истинный класс 2, модель же определила этот класс, как класс 1) составило 12 элементов — 4% от общего количества стратегий.

Выводы

Можно отметить, что использование методов машинного обучения в анализе маркетинговых кампаний представляет собой мощный инструмент для современных компаний. Машинное обучение позволяет более точно определять целевую аудиторию, выявлять наиболее эффективные каналы продвижения, а также оптимизировать рекламные бюджеты.

Эффективность машинного обучения в маркетинге заключается в способности обрабатывать большие объемы данных и выявлять сложные взаимосвязи между различными маркетинговыми переменными. Это позволяет компаниям принимать обоснованные решения, основанные на фактах и статистике, а не только на интуиции.

Кроме того, использование методов машинного обучения в маркетинге способствует автоматизации процессов анализа данных, что позволяет сэкономить время и ресурсы компании.

Таким образом, интеграция методов машинного обучения в анализ эффективности маркетинговых кампаний открывает новые перспективы для компаний в достижении своих бизнес-целей, улучшении взаимодействия с клиентами и повышении конкурентоспособности на рынке.

Библиографический список

1. Байкова, И. А. Исследование эффективности продвижения образовательных услуг инструментами интернет-маркетинга / И. А. Байкова, В. В. Канафьева // Петербургский экономический журнал. – 2021. – № 1. – С. 60-77. – DOI 10.24411/2307-5368-2020-10058. – EDN XJQIDT.

2. Дж. Дэннис мл., Шнабель Р. Численные методы безусловной оптимизации и решения нелинейных уравнений, Москва, “Мир”, 1988.
3. Исаков С. Градиентный спуск: все, что нужно знать. — URL: <https://neurohive.io/ru/osnovy-data-science/gradient-descent/> (дата обращения: 13.05.2024).
4. Кульбак С., Лейблер Р. А. Об информации и достаточности // *Ан-налы математической статистики*. — 1951. — Т. 22. — №. 1. — С. 79-86.
5. Маккей, Дэвид Дж. С. Теория информации, логический вывод и алгоритмы обучения. — Первое изд. — Издательство Кембриджского университета. — 2003. — С. 34.
6. Математические и инженерные примеры законов распределений случайных величин в ЦОС Nomotex / Т. В. Облакова, К. М. Зубарев, А. А. Сальникова, Д. С. Шинаков // *Дневник науки*. — 2022. — № 12(72). — DOI 10.51691/2541-8327_2022_12_29. — EDN VLCPIK.
7. Николенко С., Кадурич А., Архангельская Е. Глубокое обучение. — СПб.: Питер, 2018. — 480 с.
8. Облакова, Т. В. Анализ распределения высоты морских волн. Сравнение оценок и применение критерия согласия Пирсона / Т. В. Облакова, К. М. Зубарев, Д. Ю. Яковлев // *Дневник науки*. — 2023. — № 12(84). — DOI 10.51691/2541-8327_2023_12_32. — EDN UUYBYS.
9. Плетнева, Н. А. Оценка эффективности распределения рекламного бюджета посредством мультитач-исследований / Н. А. Плетнева, А. В. Кашковский, К. В. Евдокимов // *Вестник Белгородского университета кооперации, экономики и права*. — 2018. — № 4(71). — С. 191-201. — EDN XTGFZZ.
10. Прудникова О.М., Корешкова Е.С. Применение линейной модели множественной регрессии в экономических расчетах / *Наука и современность*. — 2016 — Т. — №14 — с. 134 — 140.
11. Рубашкин Г. В. Прогнозирование объемов продаж промышленных предприятий на основе моделей множественной линейной регрессии / *Экономический анализ*. — 2006 — Т. — №8 — с. 51 — 57.
12. Стрижов В.В., Крымова Е.А. Методы выбора регрессионных моделей, *Сообщения по прикладной математике*. — М.: Вычислительный центр РАН, 2010 — 21 с.
13. Тартаковский А. Г. Об асимптотической оптимальности при последовательном обнаружении точки изменения // *Труды по теории информации*. — 2017. — Т.63. — №6. — С. 3433-3450.

14. Dimitrienko, Y. I. Modeling nonlinear dielectric properties of laminated composites / Y. I. Dimitrienko, E. A. Gubareva, K. M. Zubarev // Journal of Physics: Conference Series, Moscow, 20 ноября 2020 года. – Moscow, 2021. – P. 012057. – DOI 10.1088/1742-6596/1990/1/012057. – EDN AKRFNF.
15. Kaggle: [Электронный ресурс]. URL: <https://www.kaggle.com/datasets/singhnavjot2062001/product-advertising-data>. (Дата обращения: 08.04.2024).
16. Murphy, Kevin P. Machine learning: a probabilistic perspective. — 2012. — (Adaptive computation and machine learning series).

Оригинальность 90%