

УДК 004.852

ИСПОЛЬЗОВАНИЕ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ ДЛЯ АНАЛИЗА ДАННЫХ И ОЦЕНКИ УСПЕВАЕМОСТИ СТУДЕНТОВ (СОТРУДНИКОВ) В ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММАХ

Медетова К.М

Ташкентский государственный экономический университет, г. Ташкент, Узбекистан, medetovakm@gmail.com

Аннотация. Данная научная статья исследует применение метода логистической регрессии для оценки уровня знаний сотрудников на основе их учебных данных. Авторы представляют методику сбора и анализа данных, а также обсуждают результаты и интерпретацию модели логистической регрессии в контексте оценки обучения в организации.

Ключевые слова. Логистическая регрессия, анализ данных, успеваемость, образовательные программы, сотрудники, машинное обучение, прогнозирование успеха.

Современное корпоративное обучение становится все более цифровым и доступным. Компании инвестируют в обучение своих сотрудников с целью увеличения производительности, укрепления конкурентоспособности и поддержания развития персонала. Однако для эффективного управления обучением и мониторинга успехов сотрудников необходимо иметь систему анализа данных, способную выявлять паттерны и тенденции в учебных прогрессах [1].

В современном мире образование и обучение сотрудников стали ключевыми элементами успешного функционирования организаций. Однако эффективное оценивание и управление успеваемостью сотрудников в образовательных программах остаются сложными задачами. Необходимость в разработке методов, которые могли бы точно прогнозировать успех обучения и выявлять паттерны в учебных прогрессах сотрудников, стала критически важной для управленческих решений и обеспечения качественной подготовки персонала. Эта проблема также актуальна в контексте высших учебных заведений и образовательных учреждений.

Оценка знаний сотрудников играет ключевую роль в современной корпоративной среде. Организации инвестируют в обучение и развитие своих сотрудников с целью повышения квалификации и производительности. В данной статье исследуется использование метода логистической регрессии для оценки успешности обучения сотрудников на основе их учебных данных.

Логистическая регрессия – это мощный статистический метод, который доказал свою эффективность в анализе данных и оценке знаний сотрудников. Ее способность моделировать вероятность событий делает ее ценным инструментом в различных сферах, включая образование и корпоративное обучение.

В нашей статье мы обсудили, как логистическая регрессия может быть применена для бинарной и многоклассовой классификации, а также какие преимущества она предоставляет при анализе данных и оценке знаний сотрудников. Мы также рассмотрели примеры применения этого метода в реальных сценариях.

Для анализа данных и выявления паттернов в учебных прогрессах сотрудников с использованием

искусственного интеллекта (ИИ), можно использовать различные алгоритмы и методы машинного обучения [2].

1. Кластеризация:

К-средних (K-Means): Это метод машинного обучения без учителя, используемый для кластеризации данных. Он позволяет автоматически группировать данные на основе их сходства. Позволяет группировать сотрудников в кластеры на основе их учебных прогрессов, выявляя схожие паттерны в обучении.

2. Классификация:

– логистическая регрессия: Может быть использована для предсказания успеха или неуспеха сотрудников в зависимости от их учебных данных.

– случайные леса (Random Forests): Это мощный метод машинного обучения, который широко используется для решения задач классификации и регрессии. Позволяют создать модель классификации на основе учебных данных и оценить важность различных факторов.

3. Ассоциативные правила:

– алгоритм Apriori: Это алгоритм ассоциативного анализа данных, используемый для поиска частых наборов элементов в транзакционных данных. Может использоваться для выявления ассоциативных правил между разными курсами или элементами обучения, что может помочь в оптимизации учебных программ.

4. Анализ временных рядов:

– ARIMA (AutoRegressive Integrated Moving Average): Этот компонент отражает зависимость текущего значения временного ряда от его предыдущих значений. Применим для анализа изменений в учебных показателях с течением времени и прогнозирования будущих результатов.

5. Методы обработки текстовых данных:

– TF-IDF (Term Frequency-Inverse Document Frequency): Это статистическая мера, используемая в обработке текстов и информационном поиске для оценки важности слова в документе относительно корпуса текстов. Используется для анализа текстовых данных, таких как отзывы и комментарии сотрудников о курсах или тренингах.

– Word Embeddings (например, Word2Vec или GloVe): Она позволяет компьютерам работать с текстовой информацией, преобразуя слова в численные



формы, которые могут быть использованы в различных задачах обработки естественного языка и машинного обучения. Позволяют векторизировать текстовые данные и находить семантические связи между словами и фразами.

6. Нейронные сети:

– рекуррентные нейронные сети (RNN): Это класс нейронных сетей, спроектированный для работы с последовательными данными и данных с зависимостями во времени. Могут использоваться для анализа последовательных данных, таких как динамика учебных показателей сотрудников.

– сверточные нейронные сети (CNN): Это класс нейронных сетей, разработанных специально для обработки и анализа данных, организованных в виде сеток или матриц, таких как изображения или видео. Могут быть применены к анализу изображений и графиков, связанных с учебными данными.

7. Снижение размерности:

– метод главных компонент (PCA): Это статистический метод, используемый для уменьшения размерности данных, при этом сохраняя наибольшую часть их изменчивости. Позволяет уменьшить размерность данных и выделить наиболее важные признаки.

8. Обучение с подкреплением:

– Q-обучение (Q-Learning): это один из методов обучения с подкреплением в машинном обучении. Может использоваться для моделирования взаимодействия сотрудников с обучающими ресурсами и оптимизации их выбора [3].

Логистическая регрессия может быть использована для оценки знаний сотрудников, когда имеются данные о результатах их обучения и необходимо предсказать, успешно ли они освоили определенный курс или задание. Рассмотрим пример использования логистической регрессии для этой цели.

Предположим, есть набор данных, в котором каждый сотрудник оценивается по результатам сдачи онлайн-курса. Данные включают в себя два признака: количество часов обучения (например, время, проведенное на курсе) и оценку сотрудника (например, баллы за тест). Нашей задачей является определить, успешно ли сотрудник освоил курс (1 – успешно, 0 – неуспешно) на основе этих признаков.

Шаги для использования логистической регрессии:

Подготовка данных:

Загрузим набор данных и проведем его предварительную обработку. Включим в этот шаг удаление отсутствующих значений, обработку выбросов и масштабирование признаков, если это необходимо.

Выбор признаков:

Определим, какие признаки (факторы) будем использовать для предсказания целевой переменной (класса). Это может потребовать анализа данных и выбора наиболее релевантных признаков.

Разделение данных:

Разделим набор данных на обучающую и тестовую выборки. Обычно данные разделяют в соотношении 70-80 % обучающих данных и 20-30 % тестовых данных.

Создание модели:

Создадим модель логистической регрессии, используя библиотеки машинного обучения, такие как Scikit-Learn в Python.

Предположим, есть данные о преподавателях в высшем учебном заведении, и мы хотим оценить, насколько успешно они обучают студентов. Мы собрали данные о преподавательской активности, включая количество часов лекций, количество проведенных семинаров и студенческие оценки за их курсы [4].

Таблица 1 – Таблица данных

Преподаватель	Часы лекций	Часы семинаров	Средняя оценка студентов	Успешность
Преподаватель 1	200	50	4.8	1
Преподаватель 2	150	40	3.2	0
Преподаватель 3	180	60	4.7	1
Преподаватель 4	120	30	2.8	0
Преподаватель 5	220	70	4.5	1

В этой таблице каждая строка представляет данные об отдельном преподавателе, включая количество часов лекций, количество проведенных семинаров, среднюю оценку студентов и метку успешности (1 – успешный преподаватель, 0 – неуспешный преподаватель).

Шаги для использования логистической регрессии [5]:

1. Подготовка данных:

Загрузим и предварительно обработаем данные, включая масштабирование признаков, если это необходимо.

2. Разделение данных:

Разделим данные на обучающую и тестовую выборки.

3. Обучение модели:

Применим метод логистической регрессии для обучения модели на обучающих данных.

Код на Python

```
from sklearn.linear_model import
LogisticRegression

# Создание модели логистической регрессии
model = LogisticRegression()

# Обучение модели на обучающих данных
model.fit(X_train, y_train)
```

Оценка модели:

Оценим производительность модели на тестовой выборке, используя метрики точности, полноты, F1-меры и другие.

```
from sklearn.metrics import accuracy_score, classification_report

# Сделать прогнозы на тестовой выборке
y_pred = model.predict(X_test)

# Оценить точность модели
accuracy = accuracy_score(y_test, y_pred)
print(f"Точность модели: {accuracy}")

# Вывести отчет о классификации
report = classification_report(y_test,
y_pred)
print(report)
```

2. Интерпретация результатов:

Интерпретация результатов в контексте машинного обучения и анализа данных означает понимание и объяснение выводов, полученных из модели или алгоритма. Это важный этап в процессе анализа данных, поскольку позволяет делать выводы, принимать решения и выявлять важные закономерности. Проанализируем веса признаков модели, чтобы понять, какие факторы (часы лекций, часы семинаров, средние оценки студентов) влияют на успешность преподавателей.

Веса признаков

```
feature_weights = model.coef_[0]
print("Веса признаков:", feature_weights)
```

Таким образом, в этом конкретном примере мы использовали логистическую регрессию для оценки успешности преподавателей в высшем учебном заведении на основе их учебной деятельности и студенческих оценок. Модель может быть применена для прогнозирования успешности преподавателей и выявления факторов, влияющих на их успех в обучении студентов.

В данной статье было исследовано использование логистической регрессии для оценки успешности обучения сотрудников в организации на основе их учебных данных. Мы применили этот метод к конкретным данным о сотрудниках, включая количество часов обучения и оценки за задания, и провели анализ результатов.

Логистическая регрессия представляет собой мощный инструмент для оценки успешности обучения сотрудников и прогнозирования вероятности их успешного завершения курсов или заданий. Были продемонстрированы шаги для использования логистической регрессии, начиная с подготовки данных и заканчивая оценкой модели. Эти шаги включают в себя выбор признаков, разделение данных, создание и обучение модели, а также оценку производительности и интерпретацию результатов. Наша цель – продолжать развивать и совершенствовать методы оценки знаний, чтобы обеспечить более эффективное образование и профессиональное развитие сотрудников. Мы надеемся, что наши исследования будут вносить вклад в улучшение си-

стем контроля знаний и помогать сотрудникам достигать высших результатов.

Результаты анализа весов признаков позволяют выявить факторы, оказывающие наибольшее влияние на успешность обучения сотрудников. Это может быть полезной информацией для принятия решений в области образовательных программ и развития персонала. Важно помнить, что успешность модели зависит от качества данных и выбора подходящих признаков. Дополнительная настройка и оптимизация модели могут улучшить ее производительность.

Таким образом, логистическая регрессия представляет собой полезный инструмент для оценки и анализа данных об обучении сотрудников, что может способствовать более эффективному управлению образовательными программами и повышению производительности персонала в организации.

Литература

1. Zhang, A. Dive into deep learning. Release 0.7.1/ A. Zhang, Z.C. Lipton, M. Li et al. // 2019. – 900 p. URL: <https://d2l.ai/d2l-en.pdf>
2. Medetova K.M., Marisheva L.T. Применение цифровых технологий и средств для повышения эффективности управления кадрами. Scientific Journal Research and education, стр. 4-8.
3. Marisheva L.T., Medetova K.M. Classification of decision support systems in HR management. Международная научно-техническая конференции по теме: «Компьютер иллари ва мухандислик технологиялари». Том 2. стр. 279-281.
4. Шибайкин С.Д., Никулин В.В., Аббакумов А.А. Анализ применения методов машинного обучения компьютерных систем для повышения защищенности от мошеннических текстов // Вестник АГТУ. Серия: Управление, вычислительная техника и информатика. 2020. №1. URL: <https://cyberleninka.ru/article/n/analiz-primeneniya-metodov-mashinnogo-obucheniya-kompyuternyh-sistem-dlya-povysheniya-zaschischennosti-ot-moshennicheskikh-tekstov> (дата обращения: 18.01.2024).
5. Тарасов, И.В., Тарасова, А.И. (2019). Применение логистической регрессии в образовательной аналитике.

APPLICATION OF LOGISTIC REGRESSION FOR DATA ANALYSIS AND PERFORMANCE EVALUATION OF STUDENTS (EMPLOYEES) IN EDUCATIONAL PROGRAMS

K.M.Medetova

Tashkent state university of economy, Tashkent, Uzbekistan, medetovakm@gmail.com

Abstract. This scientific article explores the application of the logistic regression method to assess the level of knowledge of employees based on their training data. The authors present a methodology for data collection and analysis, as well as discuss the results and interpretation of the logistic regression model in the context of evaluating learning in an organization.

Keywords. Logistic regression, data analysis, academic performance, educational programs, employees, machine learning, predicting success.