

Соловьёв М.А

студент

Научный руководитель: Бабешко В.Н., к.т.н

Университет науки и технологий МИСИС

**МОДЕЛЬ ЗАВИСИМОСТИ ФИНАНСОВЫХ ПОКАЗАТЕЛЕЙ
МЕДИЦИНСКИХ ОРГАНИЗАЦИЙ ОТ КОММЕНТАРИЕВ
ПОЛЬЗОВАТЕЛЕЙ**

Аннотация: Онлайн-отзывы пациентов оказывают значимое влияние на финансовые показатели медицинских организаций. В исследовании на основе методов NLP и машинного обучения анализируется связь между тональностью, тематикой отзывов и экономической эффективностью клиник Москвы. Разработанная модель выявляет ключевые зависимости, что позволяет формулировать практические рекомендации для улучшения качества услуг и управления репутацией.

Ключевые слова: Машинное обучение, регрессионный анализ, NLP, математическая модель, тональность текста, отзывы пользователей

Solovyov M.A.

Student

Supervisor: Babeshko V.N., Ph.D. (Engineering)

National University of Science and Technology MISIS

**A MODEL OF THE DEPENDENCE OF FINANCIAL INDICATORS OF
MEDICAL ORGANIZATION ON USER REVIEWS**

Abstract: Online patient reviews significantly influence the financial performance of medical organizations. This study employs NLP and machine learning methods to analyze the connection between sentiment, review topics, and the economic efficiency of clinics in Moscow. The developed model identifies key dependencies, enabling practical recommendations for improving service quality and reputation management.

Keywords: Machine learning, regression analysis, NLP, mathematical model, text sentiment analysis, user reviews

Введение

Сегодня, отзывы пациентов — это не просто мнения, а реальный фактор, который может повлиять на успех медицинской клиники. Особенно это важно в сфере здравоохранения, где удовлетворённость пациентов напрямую сказывается не только на репутации, но и на доходах учреждения. В этой работе мы разбираемся, как именно отзывы в интернете связаны с финансовыми показателями клиник и что можно сделать, чтобы улучшить их экономическую эффективность.

Таким образом, улучшение управленческих решений может быть достигнуто за счет анализа зависимости финансовых показателей медицинских организаций от отзывов пользователей. В работе проведен такой анализ.

Метод и результаты исследования

Для анализа использовалось несколько источников данных: список медицинских организаций, лицензированные адреса, виды лицензированной медицинской деятельности были получены из открытых данных правительства Москвы, бухгалтерские отчеты для организаций были собраны из ресурса ГИР БФО, отзывы пользователей были получены из карт 2GIS. Для получения координат из адресов использовалась открытая база данных карт — OpenStreetMap.

Всего, в изначальном наборе данных содержалось порядка 20 тысяч медицинских лицензий. Из них порядка 10 тысяч имели доступную бухгалтерскую отчетность, остальные были бюджетными организациями, НКО, или по иным причинам бухгалтерской отчетности для них получить не удалось. Для организаций, для которых неизвестны финансовые показатели, невозможно выделить зависимости этих показателей от отзывов, так что эти организации из исследования были

исключены. Кроме того, были исключены организации, имеющие несколько лицензированных адресов, так как для таких организаций нет возможности связать отзывы на конкретном адресе с ее финансовыми показателями.

Для анализа, организации были кластеризованы [1] по видам лицензированной медицинской деятельности. С помощью метода локтя [2] было определено оптимальное количество кластеров — 4. После кластеризации получились осмысленные кластеры.

В первом кластере содержались организации, лицензированные под осуществление медицинских осмотров — в основном, организации, чьим основным видом деятельности является другая деятельность, но, по тем или иным причинам, нуждающиеся в лицензии на медицинские осмотры. Основной вид деятельности этих организаций не связан с медицинской деятельностью, так что для исследования они не подходят.

Во втором кластере содержались разнообразные салоны красоты, для определенных их видов деятельности требуется лицензия на осуществление медицинской деятельности. Эти организации так же не в полном смысле подходят под определение медицинских, и так же, в основном, зарабатывают не медицинской деятельностью, так что для исследования не подходят.

В третьем кластере содержались медицинские организации, оказывающие разнообразный спектр медицинских услуг, всего около 1000 штук. Эти организации подходят для исследования.

В четвертом кластере содержались медицинские организации, занимающиеся стоматологией. Медицинские организации, но, однако, осуществляющие достаточно ограниченный и однообразный спектр услуг, с отзывами, значительно связанными с стоматологической спецификой. Эти организации заслуживают отдельного исследования, так что они были исключены из исследования.

Для выбранных медицинских организаций, предоставляющих широкий спектр медицинских услуг были собраны отзывы на картах 2GIS. Организации связывались по юридическому названию и адресу. Выбирались организации, имеющие несколько отзывов. Таким образом удалось найти около 270 медицинских организаций, обладающих отзывами, которые удалось связать лицензией и бухгалтерскими показателями. Удалось собрать около 7000 отзывов для этих организаций.

В дальнейшем осуществлялся анализ настроений [3] и тем в отзывах. Для этого использовалась большая языковая модель GigaChat light [4]. За счет оптимизации показателей, влияющих на стабильность ответа, таких как температура, удалось добиться достаточно стабильных оценок отзывов. С помощью большой языковой модели для отзывов сначала определялись темы, из которых формировался закрытый список тем, затронутых в отзывах. А, дальше, с помощью модели для каждого отзыва определялись содержащиеся темы и общее настроение, а также, настроения по отдельным темам.

Всего удалось выделить несколько тем, которые были достаточно распространены между отзывами на разных организациях: услуги, отношение, цена, локация, чистота, очереди, персонал.

Для каждой организации из полученных настроений определялись средние значения, эти средние значения в дальнейшем использовались для построения моделей зависимости финансовых показателей от отзывов.

Дополнительно, для всех медицинских организаций был включен дополнительный параметр, который, как ожидалось, заметно влиял на финансовые показатели и отражал положение этих организаций — расстояние до центра Москвы. Этот параметр был достаточно ограничен, однако, так как сравнительно плохо отражал транспортную доступность, специфику богатства тех или иных районов, однако, тем не менее, от него ожидалось улучшения качества модели.

В качестве предсказываемого параметра была выбрана удельная прибыль — отношение прибыли к затратам. Это был единственный показатель, который являлся достаточно однообразным как для больших, так и для маленьких медицинских организаций, при этом, что важно, его повышение непосредственно связано с улучшением выручки, так как при его повышении ожидается либо рост прибыли, либо снижение затрат, либо промежуточное состояние, когда прибыль растет быстрее затрат.

На основании этих параметров были построены интеграционные регрессионные модели [5] — общая линейная модель с двумя параметрами — расстояние и общее отношение, линейная модель по отдельным настройкам и расстоянию, и квадратичная модель настройкам, с дополнительным линейным членом по расстоянию. Квадратичная модель использовалась для изучения ситуации, когда повышение того или иного тематического показателя требовал превосходящего роста затрат, что является не выгодным для медицинских организации. Это определяет некоторое предельное значение, выше которого повышение настроек для организации является нерациональным.

При этом, для моделей по отдельным темам была значимая мультиколлинеарность между отдельными темами, что говорит о том, что в отзывах часто позитив в разных темах коррелирует. Для учета этой мультиколлинеарности использовался L2 [6] метод регуляризации.

Итого, было получено, что расстояние во всех моделях является статистически значимым фактором на уровне $p\text{-value} = 0,05$, его повышение сопровождается снижением предсказываемой удельной прибыли, соответственно, быть ближе к центру является рациональным для медицинских организации и, например, для открытия филиалов разумно выбирать положения ближе к центру. Общее настроение значимый фактор в единственной модели, его повышение связывается

повышением удельной прибыли, соответственно, получена зависимость между настроением отзывов с финансовыми показателями.

По настройкам по отдельным темам удалось установить статистически значимую связь с удельной прибылью для тем персонал и услуги, остальные параметры получились статистически не значимыми. При этом, оба тематических параметра получили оптимальное значение в пределах интервала между -1 и 1, в пределах которого находились настройки отзывов. Это говорит о том, что действительно существует некоторый момент, когда повышать настроение отзывов становится не выгодным. При этом, для большинства организаций, этот показатель выше, чем достигнутый ими, что говорит о том, что для них является выгодным повышение этих показателей за счет принятия управленческих решений.

При этом, значимость моделей получилась не очень большая — от 0,09 до 0,13, что говорит о том, что существует большое количество иных факторов, оказывающих более значимое влияние на финансовые показатели медицинских организаций. Этими параметрами могут быть маркетинг, реальные компетенции персонала, оказываемые услуги, репутация или какие-то контрактные связи, которые влияют на финансовые показатели.

Из полученной модели можно определить примерное улучшение финансовых показателей, которое можно ожидать. Получить его можно если подставить текущие и оптимальные показатели в модель и сравнить. По итогам сравнения можно сделать вывод о том, что модель предсказывает рост удельной прибыли примерно на 12%, однако, с учетом небольшого коэффициента детерминации с уверенностью можно ожидать, что реальный рост финансовых показателей будет ниже предсказываемого и оценку ему можно дать в предположении, что только 11% (коэффициент детерминации для квадратичной по отдельным темам настройкам) изменчивости реально предсказывается моделью, соответственно,

ожидаемый рост удельной прибыли будет порядка 1,2%, что, несмотря на не слишком большое значение остается статистически значимым.

Заключение

Итого, была построена интеграционная модель зависимости финансовых показателей медицинских организации от отзывов пользователей. Она показывает статистически значимую зависимость, из этого можно сделать вывод, что качество работы персонала и качество оказываемых услуг оказывают статистически значимое влияние на финансовые показатели, так же, как и положение организации в пространстве. При этом, небольшой коэффициент детерминации говорит о том, что использовать модель как единственный источник для принятия решений неверно, и модель может быть использована лишь как средство для улучшения принятия решений и мотивирования стратегических управленческих решений в медицинских организациях.

Использованные источники:

1. Бородич С. А. Вводный курс эконометрики: Учебное пособие. – 2000.
2. Kuraria A., Jharbade N., Soni M. Centroid selection process using WCSS and elbow method for K-mean clustering algorithm in data mining //International Journal of Scientific Research in Science, Engineering and Technology. – 2018. – Т. 4. – №. 11. – С. 190-195.
3. Прокапович А. А., Егошина А. А. АНАЛИЗ МЕТОДОВ ОПРЕДЕЛЕНИЯ ТОНАЛЬНОСТИ ЕСТЕСТВЕННО-ЯЗЫКОВЫХ ТЕКСТОВ //ТЕХНОЛОГИИ И СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА. – С. 376.
4. Никита Дмитриевич Мальцев Структурные и филологические особенности текстовых генеративных нейронных сетей // Неофилология. 2024. №2 (38).

5. Вучков И., Бояджиева Л., Солаков Е. Прикладной линейный регрессионный анализ. – 1987.
6. Steiger J. H. Tests for comparing elements of a correlation matrix //Psychological bulletin. – 1980. – Т. 87. – №. 2. – С. 245.