Ординальная классификация тональности на форумах поддержки онкологических больных: контролируемый бенчмарк моделей машинного обучения и трансформеров

Источник: Frontiers in Digital Health

Оригинал: https://www.frontiersin.org/articles/10.3389/fdgth.2026.1836536

ALBERT | BioBERT | NLP | классификация | ментальное здоровье | мониторинг | онкология | трансформеры | цифровая фенотипизация

Введение

Онлайн-форумы поддержки онкологических пациентов содержат естественные описания страха, неопределённости, совладания и нагрузки на лиц, осуществляющих уход. Такие тексты могут способствовать цифровому фенотипированию как один компонент лонгитюдного мониторинга под наблюдением человека, однако операционная связь между психосоциальным дистрессом и метками сентимента требует явной оценки.

Методы

Мы провели бенчмаркинг репрезентативных классических, рекуррентных и трансформерных моделей для четырёхклассовой ординальной классификации сентимента с использованием набора данных Mental Health Insights—Vulnerable Cancer Patients (N = 10 392). Модели оценивались в рамках единого валидационно-ориентированного разделения 60/20/20 с использованием взвешенного F1, макро AUC one-vs-rest, производительности по классам и парных сравнений.

Результаты

Трансформерные модели продемонстрировали наилучшую общую производительность. ALBERT показал наивысшие значения взвешенного F1 и макро AUC (0,7667 и 0,931 соответственно), в то время как BioBERT был близко сопоставим (взвешенный F1 = 0,7613; макро AUC = 0,917) и продемонстрировал несколько более высокую полноту для класса «очень негативный» (0,8019 против 0,7736). Анализ ошибок показал, что ошибки трансформерных моделей концентрировались вокруг ординальных границ решений, в то время как остаточные ошибки положительного класса оставались операционно значимыми для поддерживающих рабочих процессов.

Обсуждение

Эти специфичные для разделения данные поддерживают использование тонкой настройки трансформерных моделей в качестве компонента поддержки принятия решений для мониторинга, ориентированного на уязвимость, при этом подчёркивая важность калибровки, прозрачного анализа ошибок и человеческого надзора, а не автономной клинической оценки.
