

ВЫЯВЛЕНИЕ СИГНАЛОВ НЕБЛАГОНАДЁЖНОСТИ КОМПАНИЙ В НОВОСТЯХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Маньков В.А. (ИТМО)
Научный руководитель – Булушова Л.С.,
Университет ИТМО

Введение

Традиционно анализ рисков основывается на финансовой отчётности, однако данный подход имеет ряд ограничений [1]: во-первых, не все компании являются публичными и обязаны раскрывать подробную отчётность; во-вторых, публикация отчётности осуществляется с периодичностью раз в квартал или год, что затрудняет оперативное выявление проблем; в-третьих, финансовые показатели не всегда отражают юридические и репутационные риски.

В зарубежных исследованиях активно развиваются методы автоматизированного анализа текстов новостей и отчётных документов с применением машинного обучения и обработки естественного языка [2, 3]. Новостные источники могут содержать ранние сигналы о финансовых трудностях, судебных разбирательствах, расследованиях регуляторов и иных событиях, способных повлиять на устойчивость компании. Поэтому становится актуальной задача разработки модели, способной классифицировать новости по признаку наличия в них сигналов неблагонадёжности компаний.

Основная часть

В работе предлагается подход к множественной классификации новостных сообщений об американских компаниях на три класса: содержащие сигнал неблагонадёжности (финансовый или юридический) и не содержащие такого сигнала. Под сигналами понимаются упоминания о снижении финансовых показателей, угрозе банкротства, дефолтах, судебных исках, расследованиях со стороны регуляторов и иных событиях, потенциально свидетельствующих о росте рисков. Предлагаемое решение включает следующие этапы: формирование корпуса новостей по американским компаниям из открытых источников; разработка критериев разметки данных и создание обучающей выборки; предобработка текстов; обучение и сравнение нескольких алгоритмов классификации.

В качестве методов планируется использование как классических алгоритмов (логистическая регрессия, градиентный бустинг), так и трансформеров [3]. Особое внимание уделяется выявлению юридических сигналов, которые часто не отражаются в финансовой отчётности, но могут существенно влиять на устойчивость компании.

Выводы

Ожидается, что модель продемонстрирует способность выявлять релевантные сигналы неблагонадёжности в новостных текстах с приемлемым уровнем точности и полноты. Практическая значимость работы заключается в возможности использования результатов для: мониторинга рисков инвестиционных портфелей; поддержки принятия решений аналитиками и риск-менеджерами.

Литература

1. Pei J., Vadlamannati S, Modeling and Huang L. Detecting Company Risks from News// University of Michigan, Ann Arbor, MI, USA. 2024. 10 p.
2. Loughran T., McDonald B. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks // The Journal of Finance. 2011. Vol. 66. No. 1. P. 35–65.
3. Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. 2019. P. 4171–4186.