

Обзор литературы по теме «Применение машинного обучения в управлении кредитными рисками»

Научно-исследовательский семинар “Методология проектной работы”

Столярова Анастасия, группа БЭАД223

Мотивация и актуальность работы

- С одной стороны, в настоящее время сфер, где можно применять методы машинного обучения, становится все больше.
- С другой стороны, эффективное управление кредитными рисками в настоящий момент является областью активных исследований, особенно в связи с турбулентностью, вызванной факторами среды.
- Мотивацией работы стало изучение конвергенции этих двух сфер и рассмотрение того, как именно методы машинного обучения позволяют улучшить существующие методы оценки кредитных рисков.

Основные акценты исследования

- Для начала был проведен предварительный анализ литературы, цель которого - в целом проследить исторический процесс внедрения методов МО при анализе кредитных рисков.
- Предварительный анализ показал, первые публикации по теме датируются 1995-2000 годами, то есть в целом тема не является содержательно новой. Более того, за последнее десятилетие темпы прироста количества статей по данной теме значительно выросли в силу бурного развития и улучшения методов машинного обучения.
- Активно публикуются масштабные исследовательские работы по данной теме.
- В связи с этим, акцент в данном исследовании был сделан в первую очередь на изучении тенденций последних 4-х лет.

Методология работы

1. Критерии отбора статей:

- Фокус на профильных международных научных журналах.
- Высокий уровень цитирования и актуальность работ.
- Включены статьи, опубликованные за последние три года.

2. Систематический обзор исследований:

- Изучение ранее опубликованных обзоров литературы.
- Определение ключевых вопросов и ограничений в применении методов машинного обучения.
- Анализ свежих статей для выявления новых подходов и решений.

3. Анализ применения на реальных данных:

- Понимание практической работы методов машинного обучения.
- Оценка результатов их использования в реальных условиях.

4. Цели исследования:

- Обзор современных подходов и тенденций в оценке кредитного риска.
- Выявление потенциал для дальнейшего развития методов машинного обучения.
- Определение проблем и ограничений при использовании этих методов.

Методология выбора источников

Для того, чтобы обеспечить актуальность работы, использовался следующий алгоритм отбора статей для итогового шорт-листа статей, вошедших в исследование:

1. дата публикации статьи - не ранее 2019 года
2. для обеспечения релевантности исследуемых источников:
 - а) использовались статьи с большим количеством цитирований
 - б) если статья была опубликована в 2023 году, то приоритетно учитывался рейтинг журнала

Вопросы, исследованные в ходе работы

Цель работы заключалась в ответ на следующие ключевые вопросы:

1. Какие ограничения были выявлены в ходе изучения применения методов машинного обучения при анализе кредитных рисков за последнее десятилетие?
2. Какие существуют способы решить проблемы с ограничениями применения методов машинного обучения в анализе кредитных рисков?
3. Как методы машинного обучения позволяют справляться с ситуациями высокой неопределенности и асимметричности информации, улучшают ли они результаты кредитного скоринга в сравнении с использованием обычных статистических моделей и человеческого опыта?

Этапы работы

1. Первый этап:

Изучение основных тенденций в применении методов машинного обучения в анализе кредитных рисков на основе обзоров литературы, опубликованных до целевых годов исследования (то есть до 2019 года)

2. Второй этап:

На основании источников из пункта 1 были выделены основные проблемы в применении методов машинного обучения и дальнейшее исследование строилось с целью изучения предлагаемых решений выявленных проблем

3. Третий этап:

В силу того, что нелинейность и непредсказуемость моделей машинного обучения позволяет выявлять неуловимые статистическими моделями паттерны и закономерности, отдельное внимание в исследовании уделялось применению методов машинного обучения для анализа кредитных рисков в нетривиальных кейсах с ярко выраженной асимметрией информации и высокой неопределенностью

Результаты первого этапа работы

В результате систематического анализа литературы были сформулированы следующие выводы:

1. Сложность интерпретации результатов, препятствующая повсеместному внедрению методов машинного обучения для анализа кредитных рисков в финансовых организациях из-за законодательных ограничений
2. Отсутствие агрегированной модели, применимой к любому набору данных без потери устойчивости

Результаты второго этапа работы

1. Существуют алгоритмы машинного обучения, позволяющие сделать модель более устойчивой к факторам среды (в частности, к экономико-политическим факторам)
2. С проблемой плохой агрегируемости моделей машинного обучения справляется нейросеточная модель, использующая нейросеть, основанная на известном методе Нистрёма. Более того, эта имеет преимущества в практическом применении
3. Проблеме сложности интерпретации моделей машинного обучения при анализе рисков посвящено много исследований: популярное решение - использовать методы машинного обучения для улучшения результатов работы статистических методов. Исследования демонстрируют высокую эффективность

Результаты третьего этапа работы

Были рассмотрены нетривиальные кейсы, в которых использование машинного обучения позволило улучшить результаты работы:

- социальное кредитование
- peer-to-peer кредитование
- кредитование в условиях высокого уровня невозвратов, вызванных серым сектором экономики на примере сельских районов Перу
- кредитование МСП
- кредитные риски в условиях неопределенности, вызванной эпидемией covid-19
- методы предотвращения дефолтов по кредитным картам

В исследованиях отмечалось, что используемые подходы могут быть применимы в индустрии для преодоления ограничения асимметричности информации.

Результаты исследования

- Использование методов машинного обучения значительно улучшает эффективность и точность задачи кредитного скоринга
- Наблюдается тенденция использования методов, комбинирующих статистические методы, подкрепленные методами машинного обучения для уточнения результатов и выявления скрытых нелинейных зависимостей
- Преимущество данного подхода заключается в том, что он использует в качестве базового метода легко интерпретируемый статистический метод, который широко применяется в индустрии, однако за счет дополнительного обучения модели он способен гарантировать результат, сопоставимый по точности с методами, использующими чисто методы машинного обучения
- Дальнейшие актуальные исследования в области применения методов машинного обучения для анализа кредитных рисков будут направлены на улучшения работы статистических методов при помощи методов машинного обучения и получении моделей машинного обучения с высокой интерпретируемостью результатов.

Список литературы

- [1] Guina Sotomayor Alzamora, Miguel Romilio Aceituno-Rojo и Henry Ivan Condori-Alejo. “An Assertive Machine Learning Model for Rural Micro Credit Assessment in Peru”. B: *Procedia Computer Science* 202 (2022), с. 301—306.
- [2] Siddharth Bhatore, Lalit Mohan и Y Raghu Reddy. “Machine learning techniques for credit risk evaluation: a systematic literature review”. B: *Journal of Banking and Financial Technology* 4 (2020), с. 111—138.
- [3] Alessandro Bitetto и др. “Machine learning and credit risk: Empirical evidence from small- and mid-sized businesses”. B: *Socio-Economic Planning Sciences* (2023), с. 101746.
- [4] Elena Dumitrescu и др. “Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects”. B: *European Journal of Operational Research* 297.3 (2022), с. 1178—1192.
- [5] Marta Ramos Gonza'lez, Antonio Partal Ure'na и Pilar Go'mez Fern'andez-Aguado. “Forecasting for regulatory credit loss derived from the COVID-19 pandemic: A machine learning approach”. B: *Research in International Business and Finance* 64 (2023), с. 101907.
- [6] Tie Li, Gang Kou и Yi Peng. “A new representation learning approach for credit data analysis”. B: *Information Sciences* 627 (2023), с. 115—131.
- [7] David Maloney, Sung-Chul Hong и Barin N Nag. “Two class Bayes point machines in repayment prediction of low credit borrowers”. B: *Heliyon* 8.11 (2022).
- [8] Saba Moradi и Farimah Mokhatab Rafiei. “A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks”. B: *Financial Innovation* 5.1 (2019), с. 1—27.
- [9] Fabio Sigrist и Nicola Leuenberger. “Machine learning for corporate default risk: Multi- period prediction, frailty correlation, loan portfolios, and tail probabilities”. B: *European Journal of Operational Research* 305.3 (2023), с. 1390—1406.
- [10] Tanmay Srinath и HS Gururaja. “Explainable machine learning in identifying credit card defaulters”. B: *Global Transitions Proceedings* 3.1 (2022), с. 119—126.

Список литературы

[11] Yuan Wang и др. “National student loans default risk prediction: A heterogeneous ensemble learning approach and the SHAP method”. В: *Computers and Education: Artificial Intelligence* 5 (2023), с. 100166.

[12] Xu Zhu и др. “Explainable prediction of loan default based on machine learning models”. В: *Data Science and Management* (2023).