

Андрій Лавренюк, Володимир Месюра (Вінниця)

## ДОСЛІДЖЕННЯ МОДЕЛЕЙ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ДЛЯ ПОБУДОВИ СИСТЕМ КРЕДИТНОГО СКОРИНГУ

Кредитним скорингом називається методологія оцінювання кредитоспроможності потенційних позичальників, класифікації потенційних клієнтів банку по рівню ризику, набір моделей прийняття рішень щодо вирішення питання надання споживчого кредиту. Найчастіше постановкою задачі для скорингу в поняттях ризику контрагента є прогнозування показників, які є ранніми індикаторами по відношенню до дефолту, але призводять до нього з високою ймовірністю.

Швидкий розвиток машинного навчання і заснованого на ньому інтелектуального аналізу даних (ІАД) надав нові методологічні можливості для побудови вдосконалених моделей кредитного скорингу, які можуть базуватись щонайменше на шести його класичних задачах: класифікації, кластеризації, прогнозуванні, оцінюванні, асоціації та візуалізації, – і використовуються на етапах кореляційного аналізу і аналізу прогностичної сили вхідних змінних, при постановці задачі моделювання, побудові моделі та оцінюванні якості прогнозів. Дуже важливими для побудови моделей кредитного скорингу є і методи попередньої обробки даних ІАД. Отже, значний інтерес викликає дослідження ефективності застосування як окремих методів ІАД, так і їх комбінації на різних етапах кредитного скорингу.

**Постановка задачі.** Необхідно здійснити порівняльний аналіз методів ІАД і визначити основні переваги і недоліки методів ІАД при їх використанні на різних етапах побудови систем кредитного скорингу.

**Розв'язання задачі.** Першою задачею при порівнянні методів ІАД є підготовка наборів даних, які були взяті з архіву центру машинного навчання та інтелектуальних систем [3]. Для зменшення розмірності даних з мінімальними втратами інформації та відбору найбільш значущих параметрів з базового набору даних були отримані два зменшені набори: за допомогою методу головних компонент і випадкового лісу. Для порівняння були відібрані моделі: методу опорних векторів, логістичної регресії, бустингу і випадкового лісу. Отримані результати також передбачається порівняти з результатами, отриманими раніше при побудові систем кредитного скорингу на основі генетичних алгоритмів і нечітких нейронних мереж [1,2].

Важливою задачею при порівнянні різних моделей і методів машинного навчання є вибір параметрів кожної моделі, які не можуть бути оптимізовані на навчальному наборі і мають бути апіорними. Для вибору параметрів моделі вирішено використовувати крос-валідацію. Для оцінювання точності визначення як гарних, так і поганих позичальників використовується графік ROC (Characteristic Operator Receiver).

При попередньому тестуванні моделей найкращі результати показали логістична регресія і випадковий ліс. По тестуванню на наборах з кількох різних баз даних перевагу отримав випадковий ліс, хоча він потребує більше часу ніж логістична регресія..

**Висновки.** Отримані результати є попередніми і вимагають подальшого вдосконалення методології порівняння моделей. Це стосується як створення «рівних» умов при тестуванні різних моделей, так і вибору атрибутів набору даних і визначення їх значущості.

По результатам попередніх досліджень найпростішим і найшвидшим засобом аналізу даних став метод головних компонент. Найпростішою і найбільш стабільною моделлю виявилась логістична регресія.

### Література

1. Лавренюк А.О., Месюра В.І. Методи інтелектуального аналізу даних у кредитному скорингу [Електронний ресурс] // XLV НТК ВНТУ.- Вінниця, 2016. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2016/paper/view/1123/694> (дата звернення 08.09.2018).
2. Лавренюк А.О., Месюра В.І. Система скорингової оцінки кредитоспроможності позичальника на основі нечіткої нейронної мережі [Електронний ресурс] // XLVI НТК ВНТУ.- Вінниця, 2017. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2016/paper/view/1123/694> (дата звернення 08.09.2018).
3. Machine Learning Repository UCI [Електронний ресурс]. – Режим доступу: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing> (режим звернення 08.09.2018). – Назва з екрану.