

**АНАЛІЗ ТА ІНТЕРПРЕТАЦІЯ ДАНИХ З ВІДКРИТИХ ДЖЕРЕЛ ІНФОРМАЦІЇ**

Інтернет торгівля є основою сучасного ринку і пропонує широкий спектр товарів (автомобілі, нерухомість, побутова техніка і т.д.). Як наслідок, збільшується також кількість злочинів у сфері інтернет-торгівлі.

**Метою роботи** є дослідження методів інтелектуального аналізу даних (ІАД) з метою отримання моделі, яка буде застосована для аналізу відкритих даних про продавців на торгових інтернет-площадках. Це дозволить попередити кіберзлочини шляхом передбачення ненадійних продавців, створювати рейтинги довіри на інтернет-ресурсі з використанням широкого спектру критеріїв.

**Результати дослідження.** Задача прогнозування надійності продавця на основі аналізу відгуків, інформації про магазин та продавця є типовою задачею класифікації. Подібні задачі вирішуються за допомогою систем підтримки прийняття рішень, які використовуються в банківській сфері, зокрема, у кредитуванні, дозволяють не тільки попередити можливі ризики, але і дають можливість коректного вибору кредитного рейтингу.

При вирішенні задач класифікації однією із основних вимог є велика кількість правильних результатів у межах тестової вибірки. Згідно CRISP-DM [1], першим етапом при застосуванні методів ІАД є підготовка даних, що полягає у створенні і дослідженні початкового набору даних, та визначення його повноти. Наступним етапом є попередня обробка даних, яка використовується для підготовки даних до побудови і навчання аналітичної моделі за допомогою методів машинного навчання. Третім етапом є створення аналітичної моделі, що буде здійснювати аналіз нових даних. Результати цього етапу залежать від вибраних методів навчання та аналізу даних. Останнім етапом є перевірка моделі шляхом її апробації на тестових даних.

З метою отримання вибірки даних про інтернет-магазин, створено скраппер – клієнт, що збирає дані для навчання моделі: дату реєстрації, наявність легальних документів, кількість товарів та їх категорії, а також перелік відгуків щодо магазину, які включають рейтинг (оцінку) результату виконання певного замовлення, інформацію про підтвердження замовлення, а також відповідність замовлення за параметрами ціни, опису, наявності, та вчасності виконання замовлення.

Класичними методами вирішення схожих задач є нейромережі, логістична регресія, метод опорних векторів, дерево вибору чи випадковий ліс подібних дерев, а також метод k-найближчих сусідів. Проаналізувавши світовий досвід рішення задач бінарної класифікації багатовимірних даних в умовах малого об'єму навчальної вибірки [2], а також шляхом експериментальних досліджень, було встановлено, що найбільш доцільно застосувати глибинну нейронну мережу. В контексті вирішуваної задачі пропонується відокремити навчання на відгуках про магазин і загальних характеристиках магазину, тому перший шар перцептронів потрібно розділити.

Враховуючи особливості задачі та форму даних навчальної вибірки, запропоновано архітектуру нейромережі, що складається з 4 шарів перцептронів: 11 - у першому шарі, 6 - у другому, 4 - у третьому і один перцептрон на виході. Перші 4 входи, що відповідають загальним даним про інтернет - магазин, з'єднані виключно з 2-ма перцепторами другого рівня, інші 7 входів з'єднані виключно з іншими 4-ма перцептронами. Для корекції вагових коефіцієнтів нейромережі та запобігання перенавчання моделі застосовано алгоритм мутацій [3,4].

Для перевірки запропонованої моделі розроблено її програмну реалізацію. Навчальну вибірку сформовано на основі інформації про 40 продавців, 30 з яких були надійні і 10 не надійні. В ході навчання було досягнуто 87% точності при пошуку підозрілих продавців. Навчання відбулося за 600 тис. ітерацій, що було на 800 тис. ітерацій швидше у порівнянні з повністю з'єднаною нейронною мережею. Відносно невелика точність моделі пояснюється малим обсягом навчальної вибірки та ресурсів для навчання.

**Висновки.** Дослідження та експериментальне моделювання показало, що запропонована модель є придатною для бінарної класифікації даних і може бути застосована для аналізу даних із відкритих джерел інформації. Показано, що за рахунок поділу вхідного шару нейронів досягається аналогічна точність, але більш ніж вдвічі швидше, що дозволяє скоротити час навчання нейронних мереж на подібних форматах даних.

**Література**

1. Rdigier Wirth. CRISP-DM: Towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 2000, pp. 29–39.
2. Дударов С. П., Диев А. Н. Нейросетевое моделирование на основе перцептронных комплексов при малых объемах обучающих выборок. Математические методы в технике и технологиях – ММТТ-26: сб. трудов XXVI Междунар. науч. конф. Секции 6, 7, 8, 9. Нижний Новгород, 2013. с. 114–116.
3. W.M. Jenkins. Neural network weight training by mutation. Computers & Structures. Volume 84. Issues 31–32. 2006. pp. 2107–2112. ISSN 0045-7949
4. V. Tkachuk, M. Kozlenko, M. Kuz, I. Lazarovych, and M. Dutchak, "Function optimization based on higher-order quantum genetic algorithm," Electronic Modeling, vol.41,no.3,pp.43–58,2019,doi: 10.15407/emodel.41.03.043