РЕФЕРАТ

Дисертацію виконано на 78 аркушах, вона містить 3 додатки та перелік посилань на використані джерела з 27 найменувань. У роботі наведено 10 рисунків та 8 таблиць.

Актуальність теми. У сучасному світі проблема порушення групової анонімності даних, як одного з шляхів розкриття навмисно прихованої інформації, оволодіння якою може нашкодити її власникам, набуває великого значення. Стрімке зростання обсягів даних, неконтрольоване поширення даних за рахунок розподіленої архітектури зберігання даних, а також відновлення даних за тими, що знаходяться у загальному, відкритому доступі, потребують більш відповідального ставлення до проблем порушення анонімності даних. Задача забезпечення індивідуальної та групової анонімності сприяє можливостям створення наборів даних для навчальних цілей без остороги порушення особистих прав респондента.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась згідно з планом науково-дослідних робіт кафедри прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Мета і задачі дослідження. Метою дисертаційної роботи є визначення можливості розкриття групової анонімності даних методами машинного навчання, пояснення побудованих структур та визначення закономірностей прийняття рішень опрацьованими методами.

Для досягнення вказаної мети було розв'язано такі задачі:

– визначення стану розробки питань обраної наукової проблеми у вітчизняній та іноземній літературі;

– опрацювання статистичних даних перепису населення для визначення ключових атрибутів потрібних для порушення групової анонімності;

– побудова та аналіз класифікаційних моделей;

– проведення аналізу адекватності побудованих моделей;

– визначення зв’язків між прийнятими рішеннями різними методами машинного навчання.

Об’єктом дослідження є методи розкриття групової анонімності, а також методи машинного навчання, які застосовуються для класифікації даних (дерева рішень, зокрема ансамбль «випадковий ліс»; штучні нейронні мережі).

Предметом дослідження є розробка математичної моделі та програмного забезпечення для оцінки можливості розкриття групової анонімності методами машинного навчання, трактування структур класифікаційних моделей, побудованих різними методами машинного навчання, їх взаємозв’язок.

Методи дослідження. Для розв’язання поставленої задачі використовувалися такі методи: методи машинного навчання, які застосовуються для класифікації даних, дерева рішень, а саме ансамбль «випадковий ліс», та штучні нейронні мережі, а саме багатошаровий перцептрон.

Наукова новизна одержаних результатів складається з таких положень:

– перевірена можливість порушення групової анонімності даних за допомогою нейронних мереж та методу «випадкового лісу»;

– перевірена адекватність побудованих моделей, визначені оптимальні характеристики методів для вирішення поставленої задачі;

– порівняна структура побудованих моделей та визначені відповідності між класифікаційними рішеннями прийнятими ансамблем «випадкового лісу» та лінійною трансформацією простору багатошаровим перцептроном.

Практичне значення одержаних результатів. Отримані результати можуть бути застосовані для розкриття групової анонімності даних. Змістовне пояснення структур побудованих моделей дозволяє провести паралелі між рішеннями прийнятими ансамблем «випадкового лісу» та нейронною мережею, та таким чином пояснити структуру «чорного ящика» - нейронної мережі, за допомогою структури «білого ящика» - «випадкового лісу».

Апробація результатів дисертації. Основні положення й результати роботи представлено на VI міжнародній науково-практичній інтернет-конференції «Тенденції та вектор розвитку науки в сучасному світі» (2018).

Публікації. Результати дисертації викладено в 2 наукових працях, у тому числі:

– у 2 статтях у наукових журналах, включених до Переліку наукових фахових видань України з технічних наук.

Ключові слова: групова анонімність, «випадковий ліс», нейронна мережа.