УДК 004.85
DOI https://doi.org/10.24144/2616-7700.2020.1(36).112-122

М. М. Шаркаді[1], М. В. Роботишин[2], М. М. Маляр[3]

[1] ДВНЗ «Ужгородський національний університет», Ужгород,
доцент кафедри кібернетики і прикладної математики,
кандидат економічних наук
marianna.sharkadi@uzhnu.edu.ua
ORCID: https://orcid.org/0000-0002-1850-996X

[2] ДВНЗ «Ужгородський національний університет», Ужгород,
магістр прикладної математики
mykolaroboteshyn@gmail.com
ORCID: https://orcid.org/0000-0001-6567-6974

[3] ДВНЗ «Ужгородський національний університет», Ужгород,
професор кафедри кібернетики і прикладної математики,
доктор технічних наук
mykola.malyar@uzhnu.edu.ua
ORCID: https://orcid.org/0000-0002-2544-1959

МОДЕЛІ І МЕТОДИ МАШИННОГО НАВЧАННЯ ДЛЯ ЗАВДАНЬ ПЕРЕДБАЧЕННЯ

В процесі еволюції людства змінюється характер діяльності людини і необхідний «інструментарій» для вирішення нових задач. Останнім часом все більшу увагу заслуговують проблеми пов'язані з прийняттям рішень. Особливо актуальними є проблеми підтримки рішень у процесі управління соціально-економічними системами. Приймаючи рішення, як правило, стикаються з проблемами пошуку інформації, невпевненістю, невизначеністю, а в деяких випадках і з конфліктністю у процесі вироблення рішення. При цьому припускається, що реалізація будь-якого з варіантів рішень передбачає настання певних наслідків, аналіз та оцінка яких повністю характеризує обраний варіант. Для оцінювання можливих наслідків традиційно використовуються складні аналітичні розрахунки, знання фахівців-експертів, засоби сучасних інформаційних технологій.

Проведений аналіз існуючої практики управління соціальними та економічними системи дає можливість запропонувати нові напрями її оптимізації, котра, в свою чергу, передбачає орієнтацію на запрограмовані показники розвитку як внутрішніх системних характеристик, так і параметрів зовнішнього середовища з урахуванням прогнозних значень ключових параметрів об'єкта управління. Саме орієнтація на прогнозні показники розвитку дозволяє розробляти та втілювати в життя дієві стратегії управління процесами в соціальних та економічних системах. Важливість володіння інструментарієм та методиками розробки прогнозів для економіста і управлінця в сучасних умовах є беззаперечною.

Мета даної роботи, на основі аналізу літературних джерел зробити висновки щодо особливостей, перспектив використання та можливостей розвитку інтелектуального аналізу даних у сучасних умовах розвитку комп'ютерних технологій.

У роботі розглянуто основні методи машинного навчання і проаналізовано особливості та результати їх застосування до вирішення проблем завдань передбачення. Для вирішення проблеми, що існує потрібно визначити, які напрямки розвитку технологій потрібно удосконалювати та досліджувати науковцям.

Машинне навчання являється підрозділом доволі широкої області науки, яка вивчає штучний інтелект. Алгоритми, які відносяться до даного напрямку, використовуються для вирішення завдань, для яких часто складно або неможливо придумати явний алгоритм розв'язку.

Ключові слова: передбачення, інтелектуальний аналіз даних, алгоритми машинного навчання.

1. Вступ. Управління сучасним бізнесом немислимо без передбачення, прогнозування та аналізу даних. З кожним роком бізнес компанії збирають нові дані і використовують їх для розв'язання задач, щоб збільшувати власний прибуток. Тенденція обробки великих обсягів інформації та їх аналіз неможлива без використання методів інтелектуального аналізу даних та їх подальшого використання у сфері життєдіяльності людини. Вирішення проблем обумовлених великими обсягами даних (*Big Data*), які пов'язані з їх складністю, мережевою природою, динамікою і різноманітністю інформації привели до стрімкого зростання кількості і потужності моделей, методів і засобів інтелектуального аналізу даних.

Інтелектуальний аналіз даних (*Data Mining*) спрямований на виявлення прихованих закономірностей у даних та передбачає безпосереднє виявлення знань [1]. Тобто, на його основі можна отримати моделі, що дозволяють краще розуміти дані і передбачати їх поведінку.

Передбачення це різновиди технологій одержання інформації про майбутнє. Технологічне передбачення це обґрунтоване бачення стану можливої поведінки в майбутньому різного роду явищ, процесів і об'єктів, які невідомі в даний час, але які піддаються виявленню [2]. Розрізняються наукове і ненаукове передбачення. Наукове передбачення базується на наукових методах дослідження, а ненаукове – ґрунтується на передчуттях, інтуїції, міфології, релігії і т. п. Основною формою наукового передбачення є прогнозування, як процес вироблення прогнозів, тобто імовірне судження про стан певного явища в майбутньому, яке базується на статистичній інформації, переважно у кількісному виді. Технологічне передбачення доцільно використовувати у процесі прийняття управлінських рішень складними соціо-економічними системами в основі яких лежить людський фактор.

2. Постановка задач. Загальна постановка задачі передбачення може бути сформульована наступним чином.

Відома деяка сукупність об'єктів (ситуацій) і множина відповідей (реакцій, відгуків), а також множина у вигляді пар «об'єкт – відповідь», яка називається навчальною вибіркою. Існує деяка залежність між відповідями і об'єктами, але вона невідома. Потрібно, на основі цих даних відновити залежність, тобто побудувати модель (алгоритм), яка здатна для довільного об'єкта надати досить точну відповідь. Для вимірювання точності відповідей вводиться визначеним чином функціонал якості.

3. Мета роботи. Описати підходи, які використовуються для вирішення проблеми передбачення у різних сферах суспільного життя. Представити моделі і методи машинного навчання для різних класів завдань.

4. Огляд підходів. Ефективне застосування моделей і методів передбачення не можливе без широкого використання сучасних новітніх інформаційних технологій, яскравим представником яких є інтелектуальний аналіз даних.

Інтелектуальний аналіз даних (ІАД) це інформаційна технологія, яка дозволяє видобувати корисні знання за допомогою обробки інформації та виявлення в ній закономірностей та тенденцій, які, як правило, використовуються для підтримки прийняття управлінських рішень. Технологія ІАД базується на моделях і методах виявлення знань у великих наборах даних.

Основою ІАД є виявлення різних закономірностей у «сирих», необроблених

даних. На сьогоднішній день можна виділити такі найбільш поширені завдання, які найчастіше розв'язуються методами ІАД [1].

Класифікація – це розподіл об'єктів до одного із заздалегідь відомих наперед класів. Вона дає можливість встановити функціональну залежність між вхідними даними і дискретними вихідними змінними, що відповідають певним класам. Класифікація розділяє об'єкти за визначеною заздалегідь ознакою і вважається самим популярним завданням у всьому машинному навчанні.

Кластеризація – це групування об'єктів за подібністю, на основі певних, наперед невідомих, суттєвих властивостях (ознаках). Подібні об'єкти повинні бути віднесені до одного кластеру і відрізнятись від об'єктів з іншого кластеру. При кластеризації класи об'єктів наперед невідомі на відміну від класифікації. Точність кластеризації визначається схожістю об'єктів у середині кластера і відмінностями між кластерами.

Регресія – це встановлення деякої форми функціональної залежності вхідних змінних від вихідних. Вона безпосередньо пов'язана з прогнозуванням. Регресія показує або передбачає взаємозв'язок між процесом та тим, що цей процес може спонукати.

Асоціація – виявлення закономірностей між пов'язаними подіями. До прикладу закономірності слугує правило, що із події X випливає подія Y. Якщо події рознесені в часі, тоді асоціативні правила задають послідовні шаблони, тобто, це асоціації, які вказують на закономірності між пов'язаними в часі подіями.

ІАД безпосередньо зв'язаний з машинним навчанням (*Machine Learning*), наукою про мислення (*Cognitive Science*), а великі обсяги даних (*Big Data*), в свою чергу є підрозділом науки про аналіз даних (*Data Science*). Основу ІАД складають методи машинного навчання. Крім того, у підходах і до роботи з великими даними використовується машинне навчання, для того, щоб комп'ютер сам шукав результати опрацьованих даних. Натомість за допомогою Machine learning алгоритмів, комп'ютер сам аналізує і видає результат з обробленої інформації.

Машинне навчання (МН) – це підрозділ штучного інтелекту, який розглядає побудову алгоритмів, які можуть навчатися на наявних даних [3,4,5]. Навчання – це річ, знайома будь-якій людині, оскільки люди навчаються щодня і показують у цьому процесі прекрасні результати. Спостерігаючи закономірності в зміні середовища навколо, вони конструюють певну модель зміни цього середовища і приймають ті чи інші рішення. Середовище певним чином реагує на прийняті рішення і люди знову корегують модель світу.

Машинне навчання дозволяє знаходити закономірності в існуючих даних, щоб потім передбачати потрібну інформацію для нових об'єктів. Приблизно так і працює машинне навчання, ідея якого дуже проста: знайти закономірність і поширити її на нові дані. Машинне навчання – це спрощена версія процесу навчання, яке відбувається з людиною. Як правило, в машинному навчанні наявний певний набір прикладів, спостережень, реакцій до цих спостережень. Задача полягає у тому, щоб сконструювати такі моделі, які будуть максимально ефективно описувати наявні дані і робити достовірні прогнози. Машинне навчання є індуктивним навчанням або «навчання за прецендентами» на основі пар «об'єкт – відповідь», оскільки в основному вчимо машину вчитися на прикла-

дах, спостерігати велику кількість прикладів із реального життя, будувати на них моделі, перевіряти та застосовувати їх на подальших прикладах.

Математична модель та методи машинного навчання. Завдання МН виглядає так: уявімо собі, що в нас є певний набір об'єктів-прикладів і певний набір міток, тобто, реакцій, відповідей. Між прикладами (спостереженнями) і відповідями (реакціями) є певна прихована залежність. Задача МН – знайти цю приховану залежність для прогнозування відповідей на основі нових даних. Математичне формулювання та модель такої задачі виглядає наступним чином:

Нехай X – деяка множина, елементи якої називаються об'єктами або прикладами, ситуаціями, входами (samples); а Y – множина, елементи якої називаються відповідями або відгуками, мітками, виходами (responses). Існує деяка залежність (детермінована і імовірнісна), що дозволяє за елементами $x \in X$ передбачити $y \in Y$. Зокрема, якщо залежність детермінована, то існує функція $\varphi^* : X \to Y$. Залежність відома тільки на об'єктах навчальної вибірки $\{(x^{(i)}, y^{(i)}) : x^{(i)} \in X, y^{(i)} \in Y (i = 1, \ldots, N)\}$, N – кількість об'єктів у навчальній виборці. Упорядкована пара "об'єкт – відповідь"$\{(x^{(i)}, y^{(i)}) : x^{(i)} \in X \times Y\}$ називається прецедентом. Потрібно встановити залежність між входом і виходом на основі даних навчальної вибірки.

Модель задачі машинного навчання за прецедентами.

Задано: множина об'єктів X і множина відповідей Y .

Відомо: навчальна вибірка $\{x^{(i)} \in X, (i = 1, \ldots, N)\}$ і відповідно відповіді на цій вибірці $\{y^{(i)} = y(x^{(i)}) : y \in Y\, (i = 1, \ldots, N)\}$.

Знайти : алгоритм $a : X \to Y$, тобто алгоритм побудови вирішальної функції $\varphi \in \Phi$, яка наближує, найбільш точно, $y \in Y$ не тільки на навчальній виборці, а і на всій множині X.

Для різних типів задач множина об'єктів X і відповідей Y може задаватись по-різному. Наприклад, для задачі класифікації: $Y = \{-1; +1\}$ – класифікація на два класи; $Y = \{1, \ldots, M\}$ – класифікація на M класи, які не перетинаються; $Y = \{0; 1\}^M$ – класифікація на M класи, які перетинаються. Для задачі регресії : $Y = R$ або $Y = R^m$; для задачі ранжування Y – скінченна впорядкована множина. Множина об'єктів X, як правило, задається не самими об'єктами, а їх описами. Найбільш поширеним є ознаковий опис. Ознака (*feature*) f об'єкта $x \in X$– це результат вимірювання деякої характеристики об'єкта x. При такому підході об'єкт $x \in X$ представляється як вектор $x = (x_1, x_2, \ldots, x_k)$, k – кількість ознак, а $x_j = f_j(x)$ $(j = 1, 2, \ldots, k)$. Формально ознака це відображення $f : X \to D_j$, де D_j – множина допустимих значень ознаки.

У МН виділяються чотири види навчання [3-5]:

1) *Контрольоване.* Навчання з учителем (*supervised learning*) – найбільш поширений випадок. Кожен прецедент являє собою пару «об'єкт – відповідь». Потрібно знайти функціональну залежність відповідей від описів об'єктів і побудувати алгоритм, який бере на вході опис об'єкта і видає на виході відповідь. Функціонал якості, зазвичай, визначається як середня помилка відповідей, виданих алгоритмом, за всіма об'єктами вибірки. Як правило, вирішуються завдання класифікації та регресії.

2) *Неконтрольоване.* Навчання без вчителя (*unsupervised learning*) – тут відповіді не задаються, а потрібно шукати залежності між об'єктами. Для

кожного претендента задається тільки ситуація і потрібно згрупувати об'єкти в кластери, використовуючи дані про парну схожість об'єктів. Функціонали якості можуть визначатися по-різному, наприклад, як відношення середніх міжкластерних і внутрішньокластерних відстаней.

3) *Навчання з підкріпленням (reinforcement learning).* Навчання як окремий випадок контрольованого навчання, але вчителем є певне «середовище», в якому є певний агент (машина), що контролюється комп'ютером. Агент може вчиняти певні дії, які приводять до позитивних чи негативних відкликів «середовища» і тим самим надає агенту дані, які дозволяють йому реагувати на них і вчитися. Таким чином, агент і середовище утворюють систему за зворотнім зв'язком. Задача – максимізувати позитивні і мінімізувати негативні відклики.

4) *Напівавтоматичне навчання (semi-supervised learning)* – евристичний спосіб, в якому нерозмічені дані (ті, які не мають міток) також використовуються в тренуванні разом з розміченими даними.

Процес та алгоритми машинного навчання. У процесі вирішення завдань МН можна виділити наступні етапи [6].

– Розуміння задачі і даних;

– Попередня обробка даних і винахід (вибір, підбір) ознак;

– Побудова моделі;

– Зведення навчання до оптимізації;

– Вирішення проблеми оптимізації перенавчання;

– Оцінка якості;

– Впровадження і експлуатація.

Кожний із етапів є сам по собі складним процесом, який потребує відповідних знань і компетенції.

До типових задач машинного навчання належать: класифікація (розділяє об'єкти за визначеною заздалегідь ознакою), кластеризація (розділяє об'єкти за невідомою ознакою), регресія (показує або передбачає взаємозв'язок між процесом та його чинниками), прогнозування (знаходить значення часового ряду у майбутньому за попередніми даними), зменшення розмірності ознак (збирає конкретні ознаки у абстракції більш високого рівня), виявлення аномалій (виявлення відхилень від стандартних значень), пошук правил (шукає закономірності в потоці даних).

Результати МН у великій мірі залежать від рівня підготовки даних. Попередня обробка даних це очищення (виділення із загального масиву даних корисну інформацію), консолідація (об'єднання даних із різних джерел інформації) і підготовка їх у зручний для аналізу формат. Дані за різновидністю діляться на наступні категорії: статистичні, структуровані, напівструктуровані (слабо структуровані), неструктуровані.

Для представлення даних для МН найчастіше використовується метод ознакового описання об'єкту. Якість роботи систем МН досить сильно залежить від множини ознак, які обираються для описання вхідних даних і яким чином вони будуть описуватись. Виокремлення ознак означає перетворення початкових «сирих» даних у придатне представлення на вхід. Описання ознак, як правило, проводиться у певних шкалах таких як: бінарна, номінальна, порядкова, кількісна. Відповідно, якщо $D_j = \{0, 1\}$, тоді f бінарна ознака, D_j – скінченна множина, тоді f – номінальна ознака, D_j – скінченна упорядкована множина, тоді f – порядкова ознака, $D_j = R$, тоді f – кількісна ознака. Підбір правильних ознак займає більше часу ніж все машинне навчання. Не правильність підбору ознак може приводити до продукування неправильних результатів.

Під моделюванням розуміється використання алгоритму МН для пошуку інформації в існуючих даних для конкретної прикладної задачі. Найбільш поширеними і вживаними у практичному використанні вважаються наступні алгоритми. Алгоритми побудовані на основі регресійного аналізу. Це лінійна, багатомірна, логістична та нелінійна моделі регресії, які, як правило, використовуються для бінарної класифікації [7,8].

Лінійну модель регресії задають у вигляді рівняння прямої, яке найбільш точно показує взаємозв'язок між вхідними і вихідними змінними. Для складання такого рівняння потрібно знайти відповідні коефіцієнти для вхідних змінних. Для ситуації, коли простір ознак – лінійний, ознаки задані певним звичайним числом, простір відповідей (реакцій) – не лінійний, заданий набором класів (наприклад, проміжком), доцільно використовувати логістичну модель регресії. У нелінійній і логістичній моделях регресії вихідні дані перетворюються за допомогою нелінійних або логістичних функцій. Математичний апарат, який використовується у даних моделях – метод Нютона-Рафсона. Багатомірна модель регресії побудована на основі матричного представлення та його сингулярного розкладу.

Лінійний дискримінатний аналіз базується на статистичних властивостях даних розрахованих для кожного класу. Для кожної вхідної змінної це включає: середнє значення для кожного класу та дисперсію розраховану за всіма класами. Цей алгоритм доцільно застосовувати якщо класів є більше як два, а дані мають нормальний закон розподілу.

Дерева рішень належать до самих популярних і потужних інструментів, що дозволяють ефективно вирішувати задачі класифікації. В основі роботи дерев рішень лежить процес рекурсивної розбивки вхідної множини спостережень або об'єктів на підмножини, асоційовані із класами [9,10]. Дерево рішень будується на основі навчальної вибірки з використанням поняття інформаційної ентропії. Алгоритм «Дерева рішень» можна представити у вигляді двійкового дерева, де кожний вузол є вхідна змінна і точка розщеплення для даної змінної, при умові, що змінна є число. Існують різні критерії розщеплення. Найбільш відомі – міра ентропії й індекс Gini в основі яких лежить нормований приріст інформації. Листові вузли це вихідна змінна, яка використовуються для передбачення, яке проводиться шляхом проходу по дереву до листового вузла і виводу значення класу в цьому вузлі.

Алгоритм «Наївний Байесовський класифікатор» [9]. Суть даного алгоритму базується на припущенні, що кожна вхідна змінна є незалежною у теоремі

Байеса. Модель використовує два типи ймовірностей – імовірність кожного класу і умовну імовірність для кожного класу для всіх значень вхідної змінної, які розраховуються на тренувальних даних. Після чого дану модель можна використовувати для нових даних згідно теореми Байеса.

Алгоритм «$\boldsymbol{K}$ – найближчих сусідів» [10]. Даний алгоритм базується на оцінці подібності об'єктів. Оскільки кожний об'єкт може характеризуватись різнорідними ознаками, то основна проблема даного алгоритму у виборі метрики відстані між об'єктами. Проблему можна вирішити шляхом відбору невеликої кількості інформативних ознак, для кожної з яких будується своя функція близькості і проводиться їх згортка.

Метод опорних векторів [11,12]. У фокусі даного алгоритму лежить ідея використання поняття гіперплощини, тобто лінії, яка розділяє простір вхідних змінних. По суті, це означає провести дві прямі між категоріями так, щоб між ними утворився найбільший зазор. Найкращою гіперплощиною вважається лінія з найбільшою відстанню між гіперплощиною і найближчими точками даних. Ці точки називаються опорними векторами і відіграють головну роль при побудові гіперплощини і класифікатора. Для визначення коефіцієнтів гіперплощини, які максимізують відстань, використовуються спеціальні методи оптимізації. Сьогодні даний алгоритм є найефективним класичним класифікатором і самим популярним для спам-фільтрів.

На сьогоднішній день широке застосування для класифікації отримали нейронні мережі [10,11,13], які добре себе проявили у роботі зі складними даними, типу картинок, відео та незрозумілих бігдат. Навчання нейронних мереж з погляду математики це багато параметрична задача нелінійного програмування.

Останнім часом дедалі більшого поширення набуває клас методів глибинного навчання [12]. Суть даного підходу полягає у навчанні ознак, що дозволяє автоматично відокремлювати особливості із вхідних даних і застерігає модель від «перенавчання». Глибинне навчання характеризується, як клас алгоритмів машинного навчання, який використовує багатошарову систему нелінійних фільтрів для вилучення ознак з перетвореннями, де кожен наступний шар отримує на вході вихідні дані попереднього шару.

Для підвищення точності моделі доцільно застосовувати ансамблеві алгоритми [14-16]. Ансамбль або композиція алгоритмів – це об'єднання декількох алгоритмів у один. Кожний алгоритм може вивчити свої закономірності в даних, а якщо алгоритмів багато, то відповідно є багато закономірностей і це краще. Суть ідеї у тому, щоб навчити алгоритми, а потім усереднити отримані від них відповіді [17-19]. Головні підходи до побудови ансамблевих моделей: стекінг (*stacking*); беггінг (*bagging/bootstrap aggregation*); бустинг (*boosting*).

Стекінг – спершу навчають кілька алгоритмів, потім результати їх роботи показують останньому алгоритму. Саме він і приймає остаточне рішення. Стекінг – хороший, але найменш точний ансамбль серед інших методів.

Беггінг – це тип навчання, коли багато разів навчаємо ансамбль на випадкових вибірках даних. Тобто, тренувальні дані розбиваються на множину вибірок, для кожної із яких створюється модель. В кінцевому підсумку усереднюється відповідь за кожною із моделей. Це виглядає як голосування за найбільш популярну відповідь, де багато моделей працюють паралельно.

Бустинг – включає послідовне навчання алгоритмів. Тобто, спочатку навча-

ємо перший алгоритм і відзначаємо місця, де він помилився. Потім навчаємо другий, особливу увагу приділяючи місцям на яких помилявся перший. І так далі до необхідного результату, тобто, добавляються моделі до тих пір поки тренувальні дані не будуть ідеально передбачатись або не буде перевищена максимальна кількість моделей.

Ключовим моментом у досягненні бажаного результату в процесі навчання будь-якої системи машинного навчання є вибір правильного функціоналу якості. Функціонал якості – це певна функція, яка видає нам рівень помилки, яку робить система. В навчанні моделі головна вимога – мінімізувати помилку, яку видає система. Система спочатку видає результати, які дуже відрізняються від того, що очікується. Відповідно по кожному із прикладів ми можемо вирахувати помилку і скорегувати систему таким чином, щоб помилка була меншою. В залежності від того, яку функцію втрат функціоналу якості вибрано, буде залежати яким чином буде навчатися система. Функціонал якості відрізняється у залежності від задачі. Наприклад, для задачі регресії досить доцільне використання так званої функції квадрат помилки, тобто квадрат різниці між очікуваним результатом і результатом, який видала система. Для задачі мультикласової класифікації можна обрати абсолютну помилку, тобто функцією помилки може бути – віднесла система до правильного класу чи не віднесла.

Системи машинного навчання це представлення даних і функцій оцінки цих даних на основі уявлення і узагальнення. Якщо система будує дуже узагальнені або дуже часткові моделі і не генералізує закономірність, то виникає проблема її перенавчання. Явище при якому результат моделі на наборі даних для тестування сильно відрізняється від актуальних даних називається перенавчанням. Суть процесу перенавчання полягає у поділі даних на дані для тренування і дані для тестування моделі. Для навчання використовується лише набір даних для тренування. Тестовий набір даних використовується для перевірки роботи алгоритму на раніше невідомих даних. Процес перенавчання системи залежить від якості тренувальної вибірки.

На сьогоднішній день дані стали таким самим важливим ресурсом у всіх галузях виробництва як трудові ресурси чи виробничі активи. За рахунок їх використання компанії можуть отримувати відчутні конкурентні переваги. Наприклад, у промисловості моделі та методи МН можуть бути використані для вирішення наступних завдань: прогнозування ринкової ситуації, маркетинг і оптимізація продажів, ухвалення управлінських рішень, ефективна логістика, моніторинг стану основних фондів і т.д.

Методи машинного навчання використовуються у системах машинного зору, для аналізу людської мови і текстів, ідентифікації об'єктів на зображеннях, веб-пошуку і фільтрації контенту тощо [20]. Сфери їх застосування медична діагностика, біоінформатика, передбачення відтоку клієнтів, категоризація документів, фінансовий нагляд, технічна діагностика, кредитний скорінг і т.д.

5. Висновки та перспективи подальших досліджень. Машинне навчання — це процес застосування алгоритмів для автоматичного знаходження закономірностей у даних і використання їх для прийняття великої кількості однотипних рішень. У результаті роботи алгоритмів певний відсоток помилок є допустимими. Методів машинного навчання існує дуже багато. Тому, набагато

важливіше розуміти, коли використання тих чи інших методів буде найбільш доцільним. З наукової точки зору машинне навчання – це процес моделювання, настройки параметрів, підготовки даних і оптимізації компонент. Ціль машинного навчання, як дослідницького процесу, це пошук оптимальних відповідей та прогнозів.

Перспективи подальших досліджень полягають у розв'язанні практичних прикладних задач використовуючи методи машинного навчання.

Роботу виконано в рамках держбюджетної науково-дослідної теми Ужгородського національного університету «Розробка математичних моделей і методів для оброблення інформації та інтелектуального аналізу даних» (номер державної реєстрації 0115U004630).

Список використаної літератури

1. Гладун А.Я., Рогушена Ю.В. Data mining: Пошук знань в даних. – К.: ТОВ «ВД «АДЕФ - Україна», 2016. – 452 с.
2. Згуровский М.З., Панкратова Н.Д. Технологическое предвидение. — Киев: ІВЦ «Вид-во «Політехніка»», 2005. — 156 с.
3. Machine learning by Stanford university: веб-сайт. URL: https://www.coursera.org/learn/machine-learning/home/welcome (дата останнього звернення 01.06.2019).
4. Машинне навчання. Типи навчання: веб-сайт. URL: https://courses.prometheus.org.ua/courses/IRF/ML101/2016_T3/about (дата останнього звернення 01.06.2019).
5. Курс "Машинне навчання" від Prometheus: веб-сайт. URL: https://courses.prometheus.org.ua/assets/courseware/cdf163c83c64f8357ddbcdac82f7d624/c4x/IRF/ML101/asset/Тиждень_1_конспект (дата останнього звернення 08.04.2019).
6. Voroncov K. V. Algoritmy klasterizacii i mnogomernogo shkalirovanija : kurs lekcij [Algorithms for clustering and multidimensional scaling : course of lectures], Moskovskij gosudarstvennyj universitet, Moscow, Russia. – 2007.
7. Bishop C.M. Pattern Recognition and Machine Learning / C.M. Bishop. — NY: Springer. — 2006.
8. Ghamisi P. Advanced Spectral Classifiers for Hyperspectral Images: A review / P. Ghamisi, J. Plaza, Y. Chen et al // IEEE Geoscience and Remote Sensing Magazine. — Vol. 5, N 1. — 2017. — P. 8–32.
9. Воронцов К.В. Лекции по логическим алгоритмам классификации: веб-сайт. URL: http://www.ccas.ru/voron/download/LogicAlgs.pdf (дата останнього звернення 01.10.2019).
10. Hastie T. The Elements of Statistical Learning / T. Hastie, R. Tibshirani, J. Friedman. – Springer-Verlag, 2008.
11. Novikov A. The synthesis of information protection systems with optimal properties /Novikov, A. Rodionov // Complexity and Security. — Vol. 37. — 2008. — 307 p.
12. LeCun Y. Deep learning / Y. LeCun, B. Yoshua, H. Geoffrey // Nature. — Vol. 521, N 7553. — 2015. — P. 436–444.
13. Maulik U. Remote Sensing Image Classification: A survey of support-vectormachine-based advanced techniques / U. Maulik, D. Chakraborty // IEEE Geoscience and Remote Sensing Magazine. — Vol. 5, N 1. — 2017. — P. 33–52.
14. Huang F.J. Large-scale learning with SVM and convolutional nets for generic object categorization / F.J. Huang, Y. LeCun // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — 2006. — P. 284–291.
15. Pirotti F. Benchmark of machine learning methods for classification of a Sentinel-2 image / F. Pirotti, F. Sunar, M. Piragnolo // International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences. — Vol. 41. — 2016. — P. 335–340.
16. Breiman L. Random forests / L. Breiman // Machine learning. — Vol. 45, N 1. — 2001. — P. 5–32.
17. Chen T. XGBoost: A Scalable Tree Boosting System/ Chen T., Guestrin C: веб-сайт. URL: https://arxiv.org/abs/1603.02754 (дата останнього звернення 15.11.2019)
18. Lior Rokach Ensemble Methods for Classifiers - Boston: Springer. - 2005.

19. Ensemble Methods in Machine Learning: веб-сайт. URL: https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f (дата останнього звернення 15.10.2019).
20. Zgurovsky *M.Z.* System Analysis: Theory and Applications / M.Z. Zgurovsky, N.D. Pankratova // Springer. — 2007. — 448 p.

Sharkadi, M. M., Robotyshyn, M. V., Malyar, M. M. Modeling of risk level of the socio-economic systems functioning.

In the process of human evolution, the nature of human activity changes and it is necessary "tools" are available for solving new problems. Recently, problems related to decision-making have become increasingly important. Particularly relevant are the problems of supporting decisions in the process of managing socio-economic systems. Decision-makers are usually faced with issues of information retrieval, uncertainty, and in some cases, conflict in the decision-making process. At the same time, it is assumed that the implementation of any of the variants of decisions implies the occurrence of certain consequences, the analysis and evaluation of which fully characterizes the chosen variant. Traditionally, complex analytical calculations, expert knowledge, modern information technology tools are used to evaluate the possible consequences.

The analysis of the existing practice of managing social and economic systems makes it possible to propose new directions of its optimization, which, in turn, provides an orientation to the programmed indicators of development of both internal system characteristics and parameters of the external environment, taking into account the forecast values of key parameters of the management object. It is the orientation to the projected development indicators that allows you to develop and implement effective strategies for managing processes in social and economic systems. The importance of owning the tools and techniques of forecasting for the economist and manager in today's context is undeniable.

The purpose of this work, based on the analysis of literature sources, is to draw conclusions about the features, perspectives of use and opportunities for the development of data mining in the current environment of computer technology.

The basic methods of machine learning are considered in the paper and the peculiarities and results of their application to solving problems of prediction problems are analyzed. In order to solve the problem, there is a need to identify what areas of technology development scientists need to improve and research.

Machine learning is a unit of a fairly broad field of science that studies artificial intelligence. Related algorithms are used to solve problems that often make it difficult or impossible to come up with an explicit algorithm for solving them.

Keywords: prediction, data mining, machine learning algorithms.

References

1. Gladun, A.Y., & Rogushena, Y. V. (2016). *Data mining: Search for knowledge in data*. Kiev: ADEF - Ukraine LLC. [in Ukrainian]
2. Zgurovsky, M.Z., & Pankratova, N.D. (2005). *Technological foresight*. Kiev: Publishing House «Polytechnic Publishing House». [in Russian]
3. Machine learning by Stanford university: Website.(2019, June 1). Retrieved from https://www.coursera.org/learn/machine-learning/home/welcome.
4. Machine learning. Types of training: Website. (2019, June 1). Retrieved from https://courses.prometheus.org/en/courses/IRF/ML101/2016_T3/about. [in Ukrainian]
5. Machine Training course from Prometheus: website. (2019, April 4). Retrieved from : https://courses.prometheus.org/assets/courseware/cdf163c83c64f8357ddbcdac82f7d624/c4x/IRF/ML101/asset/Week_1_summary. [in Ukrainian]
6. Voroncov, K. V. (2007). *Algorithms of clustering and many-dimensional scaling: a course of lessons [Algorithms for clustering and multidimensional scaling: a course of lectures]*. Moscow State University, Moscow, Russia. [in Russian]
7. Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. NY: Springer.
8. Ghamisi, P., Plaza, J. , Chen, Y., Li, J. & Plaza, A. J. (2017). Advanced Spectral Classifiers

for Hyperspectral Images: A review. *IEEE Geoscience and Remote Sensing Magazine*, 5(1), 8–32. 10.1109/MGRS.2016.2616418
9. Vorontsov, K.V. Lectures on logical classification algorithms: website. (2019, october 1) Retrieved from http://www.ccas.ru/voron/download/LogicAlgs.pdf .
10. Hastie, T. Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning.* Springer-Verlag.
11. Novikov, & Rodionov, A. (2008). *The synthesis of information protection systems with optimal properties.* Complexity and Security, 37.
12. LeCun, Y., Yoshua, B., & Geoffrey, H. (2015). Deep learning. *Nature*, 521(7553), 436–444.
13. Maulik, U. & Chakraborty, D. (2017). Remote Sensing Image Classification: A survey of support-vectormachine-based advanced techniques. *IEEE Geoscience and Remote Sensing Magazine*, 5(1), 33–52.
14. Huang F.J. Large-scale learning with SVM and convolutional nets for a generic object categorization / F.J. Huang, Y. LeCun // IEEE Computer Society Conference on Computer Vision and Pattern Recognition. - 2006. P. 284–291.
15. Pirotti, F., Sunar, F., & Piragnolo, M. (2016).Benchmark of machine learning methods for the classification of a Sentinel-2 image. *International Archives of Photogrammetry, Remote Sensing & Spatial Information Sciences*,41, 335–340.
16. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
17. Chen, T., & Guestrin, C. (2019, November 15). XGBoost: A Scalable Tree Boosting System. Retrieved from https://arxiv.org/abs/1603.02754
18. Rokach, L. (2005). *Ensemble Methods for Classifiers.* Boston: Springer.
19. Ensemble Methods in Machine Learning: Website. (2019, November 15). Retrieved from https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f.
20. Zgurovsky, M.Z., & Pankratova, N.D. (2007). *System Analysis: Theory and Applications.* Springer.

Одержано 17.04.2020