

Актуальність теми

Об'єктом дослідження є процес побудови алгоритму універсальної прогностичної моделі.

Предметом дослідження є методи побудови прогностичних моделей та алгоритми класифікації даних.

Мета роботи: створення нового алгоритму побудови прогностичної моделі, що буде демонструвати точність передбачення не меншу, ніж аналогічні моделі для схожого роду вхідних даних, та мати просту реалізацію.

Методи дослідження. В роботі використовуються методи збору даних, методи класифікації текстових даних та статистичні методи.

Наукова новизна роботи полягає в наступному:

1. Запропоновано підхід, результатом якого є універсальна прогностична модель, що дає змогу абстрагуватися від конкретних реалізацій і використовувати її для тих самих даних з аналогічними показниками точності та кращими показниками швидкодії.
2. Наведено процес перетворення будь-якої прогностичної моделі чи деякої композиції моделей для перетворення в універсальну модель.
3. Підтверджено значно більші показники швидкодії моделі, розробленої за допомогою даного підходу.

На даному етапі роботи отриманих даних досить для того, щоб почати використовувати даний підхід для роботи з реальними даними та заміною існуючих алгоритмів. Практична цінність роботи полягає в наступному: Апробація роботи роботи полягає в наступному: Структура та обсяг роботи. Магістерська дисертація складається з вступу, п'яти розділів, висновків та додатків. У вступі надано загальну характеристику роботи, виконано оцінку поточного стану проблеми, обґрунтовано актуальність напрямку досліджень. У першому розділі розглянуто теоретичні відомості, існуючі алгоритми класифікації текстових даних, наведено математичні основи, що використовуються для побудови моделей. Розглянуті загальні підходи до автоматизованої класифікації текстових даних та поширені алгоритми, що застосовуються в даній галузі. Основну увагу приділено всьому процесу обробки даних: від їх початкового збору до безпосереднього застосування прогностичної моделі. У другому розділі розглянуто платформу, яка надає можливості для спрощення операцій рефакторингу; проведено огляд типів індексів, які дозволяють пришвидшити операцію пошуку; запропоновано метод рефакторингу структури бази даних для підвищення швидкодії обробки запитів вибірки; описано архітектуру компонентів та технології, обрані для реалізації проекту; наведено вхідні дані та результати роботи методу, надано відповідні ілюстративні матеріали; зазначено перелік та детальний опис кроків методу. У третьому розділі запропоновано засоби реалізації для кожного з етапів методу; наведено огляд архітектурних підходів до організації програмного забезпечення; обґрунтовано вибір мікросервісної архітектури; запропоновано структуру та особливості реалізації кожного з мікросервісів, наведено відповідні графічні матеріали, що ілюструють взаємодію елементів системи. У четвертому розділі наведено результати роботи алгоритму, підтверджено на практиці гіпотезу про те, що застосування розробленого алгоритму надає вигоду у швидкодії; отримано підтвердження того, що використання однорідних інструкцій дозволяє зменшити витрати ресурсів процесора; здійснено порівняння точності та швидкості роботи з існуючими алгоритмами; зроблено висновок щодо можливості застосування даного підходу для використання з різними алгоритмами та вхідними даними для вирішення задачі класифікації; запропоновано шляхи покращення та вектори розвитку для подальшої роботи. У п'ятому розділі подано аналіз програмного продукту, його оцінку та перспективи для виходу на ринок. Наведені слабкі та сильні сторони проекту, порівняння з аналогами та конкурентноспроможність. Проведено оцінку розміру необхідних інвестицій, обсягу ресурсів, що потрібно залучити та показників прибутку за умови подальшої комерціалізації проекту. У висновках проаналізовано отримані результати роботи. У додатках наведено фрагменти програмної реалізації запропонованого способу та копії графічних матеріалів. Робота виконана на 80 аркушах, містить 2 додатки та посилання на список використаних літературних джерел з 30 найменувань. У роботі наведено 14 рисунків та 4 таблиці. Ключові слова: класифікація, прогностичні моделі, апроксимація моделі, датасет, машинне навчання.