

*Алексєєнко В.В.*

*асистент кафедри комп'ютерних наук та інформаційних технологій,*

*Гуменюк С.П.*

*асистент кафедри комп'ютерних наук та інформаційних технологій*

*Житомирський державний університет імені Івана Франка*

## **РОЗРОБКА МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ НА ОСНОВІ ВЕЛИКИХ ДАНИХ**

У сучасному світі обсяг даних, що генеруються щодня, зростає в геометричній прогресії. Цей процес охоплює всі сфери життя, включаючи бізнес, медицину, освіту та розваги. Великі дані (Big Data) стали ключовим ресурсом для прийняття рішень, розробки стратегій та покращення різних процесів. Математичне моделювання та аналіз великих даних дозволяють виявляти закономірності, прогнозувати результати та оптимізувати діяльність у різних галузях.

Великі дані характеризуються великим обсягом, високою швидкістю створення та великою різноманітністю. Традиційні методи обробки даних часто не можуть впоратися з такими обсягами інформації, тому необхідні нові підходи та інструменти. Великі дані можна розділити на структуровані, напівструктуровані та неструктуровані. До структурованих даних належать, наприклад, таблиці баз даних, тоді як неструктуровані дані включають текстові документи, зображення та відео.

Розробка моделей для прогнозування на основі великих даних включає кілька ключових етапів:

1. Збір даних. Включає процес отримання даних з різних джерел, таких як сенсори, соціальні мережі, бази даних, веб-сайти тощо.
2. Підготовка даних. Очищення даних, заповнення пропусків, обробка аномалій та перетворення даних до форми, зручної для аналізу.
3. Аналіз даних. Використання статистичних методів та алгоритмів машинного навчання для виявлення закономірностей та побудови моделей.
4. Моделювання. Створення прогнозних моделей за допомогою методів регресії, класифікації, кластеризації тощо.
5. Оцінка моделі. Перевірка точності та надійності моделі на тестових даних, оцінка її ефективності.

6. Впровадження та моніторинг. Використання моделі в реальних умовах та постійний моніторинг її продуктивності.

Розглянемо різноманітні методи та алгоритми, які використовуються для прогнозування на основі великих даних:

- Лінійна регресія (простий та ефективний метод, який використовується для передбачення значення залежної змінної на основі значень однієї або кількох незалежних змінних);
- Логістична регресія (метод класифікації, який прогнозує ймовірність належності об'єкта до певного класу);
- Дерева рішень (алгоритм, який використовує ієрархічну структуру для прийняття рішень, базуючись на значеннях вхідних змінних);
- Метод опорних векторів (використовується для класифікації та регресії, добре працює з великими обсягами даних);
- Нейронні мережі (потужний інструмент для моделювання складних нелінійних взаємозв'язків, особливо популярний у глибокому навчанні);
- Методи кластеризації (використовуються для групування об'єктів на основі їхніх характеристик, наприклад, алгоритм k-середніх).

Розробка моделей для прогнозування результатів на основі великих даних є важливим та перспективним напрямком у комп'ютерних науках. Використання сучасних методів машинного навчання та аналізу даних дозволяє отримувати точні та надійні прогнози, що сприяє прийняттю обґрунтованих рішень у різних сферах. Незважаючи на виклики, розвиток технологій та нових підходів відкриває широкі перспективи для подальшого вдосконалення моделей та їх застосування.

### **Список використаних джерел та літератури**

1. Han J., Pei J., Kamber M. Data Mining: Concepts and Techniques. Elsevier, 2012. 740 p.
2. Murphy K. P. Machine Learning: A Probabilistic Perspective. MIT Press, 2012. 1098 p.
3. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006. 758 p.

4. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2009. 764 p.
5. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016. 800 p.