

DOI: <https://doi.org/10.36910/6775-2524-0560-2021-42-26>

УДК: 004.89

**Мороз Борис Іванович**, д.т.н., професор

<https://orcid.org/0000-0002-5625-0864>

**Кабак Леонід Віталійович**, к.т.н., доцент

<http://orcid.org/0000-0001-6267-1772>

**Ширін Артем Леонідович**, к.т.н., доцент

<https://orcid.org/0000-0003-0026-2767>

**Овчаренко Світлана Олександрівна**, студентка.

Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна.

## ВИКОРИСТАННЯ DATA MINING В ІНФОРМАЦІЙНИХ БІБЛІОТЕЧНИХ СИСТЕМАХ

**Мороз Б. І., Кабак Л. В., Ширін А. Л., Овчаренко С. А. Використання Data Mining в інформаційних бібліотечних системах.** В даний час для аналізу даних використовуються алгоритми та інструменти для аналізу даних, які називаються Data Mining. Data Mining успішно використовується в різних галузях промисловості. У статті розглядається можливість впровадження цієї технології в бібліотечні системи, оскільки аналіз даних можна використовувати для отримання різних прогнозів. Сучасні технології надходять і до бібліотек. Для запобігання дефіциту книг, та запобігання переповнення бібліотеки застарілими книгами ми можемо використовувати інформаційну систему з функціями аналізу даних з інтеграцією інструментів аналізу даних у системи управління ризиками, які можуть допомогти отримати корисну інформацію з бази даних з метою прийняття рішення про списання книг або переміщення в сховище.

**Ключові слова:** База даних Oracle, аналіз даних, аналіз даних, виявлення аномалій, бібліотека, кластерний аналіз, кореляційний аналіз, система управління ризиками.

**Мороз Б. И., Кабак Л.В., Ширин А. Л., Овчаренко С. А. Использование Data Mining в информационных библиотечных системах.** В настоящее время для анализа данных используются алгоритмы и инструменты для анализа данных, которые называются Data Mining. Data Mining успешно используется в различных отраслях промышленности. В статье рассматривается возможность внедрения этой технологии в библиотечные системы, поскольку анализ данных можно использовать для получения различных прогнозов. Современные технологии используются и в библиотеках. Для предотвращения дефицита книг и предотвращения переполнения библиотеки устаревшими книгами мы можем использовать информационную систему с функциями анализа данных с интеграцией инструментов анализа данных в системы управления рисками, которые могут помочь получить полезную информацию из базы данных с целью принятия решения о списании книг или перемещение в хранилище.

**Ключевые слова:** База данных Oracle, анализ данных, анализ данных, выявления аномалий, библиотека, кластерный анализ, корреляционный анализ, система управления рисками.

**Moroz B., Kabak L., Shyrin, A., Ovcharenko S. Using of the Data Mining technics in Library Systems.** Algorithms and tools for data analysis called Data Mining are currently used for data analysis. Data Mining is successfully used in various industries. The paper considers the possibility of introducing this technology in library systems because data mining can be used to obtain various forecasts. Modern technologies comes to libraries too. To prevent book shortages, and to prevent the library from overflowing with obsolete books, we can use an information system with data analysis functions with the integration of data analysis tools into risk management systems, which can help to obtain other useful information from the data for the purpose of making a decision to write off books or move to repository.

**Keywords:** Oracle database, data mining, data analysis, anomaly detection, library, cluster analysis, correlation analysis, risk management system.

### Постановка наукової проблеми.

Відтепер інформаційних технологій стає все більше. Тому більше людей відходять від старого бачення ведення паперової документації, і переходять на використання електронних інформаційних систем. Це дає змогу заощаджувати час та робити деякі справи не виходячи з дому. Майже кожна людина у світі полюбає читати книги, проте щоразу купувати книги в домашню бібліотеку досить дорого та не дуже екологічно, тому на допомогу приходять громадські бібліотеки. Розглянемо старе бачення бібліотеки, нам було потрібно прийти отримати читацький квиток, переглянути паперову картотеку щоб дізнатися, чи є потрібна нам книга у наявності, для замовлення взяття книги додому бібліотекар записував нас у книгу замовлень і віддавав нам книгу - це все займало купу часу. З появою мережі Інтернет передавати та отримувати інформацію стало швидше і простіше, тому щоб зробити роботу бібліотеки швидшою і простішою постає потреба у створенні додатку, що дозволив би частину зі справ, що роблять читачі і бібліотекарі виконувати онлайн, а саме: зручно відслідковувати статус замовлень, наявність книжок та статус користувачів, переглядати список, змінювати асортимент бібліотеки та фільтрувати користувачів. У кожного веб-додатку є якась база даних для збереження, обробки і накопичення інформації, аналізуючи раніше зібрану інформацію можна отримувати корисні висновки і прогнози та використовувати їх для розвитку місця введення інформаційної системи або передбачення негативних явищ і їх запобігання. Для цього і застосовується інтелектуальний аналіз

даних. Нині для аналізу даних використовуються алгоритми та інструменти для аналізу даних які називаються Data Mining. Data Mining успішно використовується у різноманітних галузях. В роботі розглядається можливість запровадження цієї технології у бібліотечних системах адже інтелектуальний аналіз даних можна використовувати для отримання різного роду прогнозів. Наприклад дану технологію можна інтегрувати до системи аналізу кошків споживачів, щоб оптимально використовувати ресурси книжок які зберігаються у бібліотеці, та використовуючи технологію побудови регресії та прогнозування виявити книжки та підручники, які більш затребувані та прогнозувати їх закупівлю на майбутнє. І це не єдиний можливий варіант використання технології інтелектуального аналізу даних у бібліотеці. Багато організацій та підприємств вже використовують цю технологію і успішно використовують можливості інтелектуального аналізу даних для досягнення необхідних результатів. Метою роботи є розробка та створення: моделі інформаційної системи бібліотеки з використанням Data Mining, та додаток «Library», що буде здійснювати занесення та перегляд інформації, проводити аналіз внесених даних за допомогою сучасної технології Oracle Data Mining, дозволяти генерувати звіти, створювати текстові файли, з використанням мови Java та технологій Thymeleaf, jsp, jstl, css.

#### Аналіз досліджень.

Нині в інформаційних системах все частіше використовують технології інтелектуального аналізу даних з метою виявлення моделей та тенденцій. Прикладами таких продуктів є: Oracle Data Mining, Microsoft SQL Server Integration Services, RapidMiner Studio, STATISTICA і багато інших [1], [2]. Технологія Data Mining вперше була запропонована Piatetsky-Shapiro G в 1989 році і використовується для визначення безлічі алгоритмів аналізу даних.

Інтелектуальний аналіз даних є важливим кроком у отриманні знань і процесів з баз даних. За допомогою технологій інтелектуального аналізу даних можливо створити корисні зразки або моделі з даних [3]. Інтелектуальний аналіз даних відноситься до загального процесу виявлення корисних знань з даних. Data Mining відноситься до виявлення нових зразків із безлічі даних у базах даних, орієнтуючись на алгоритми для отримання корисних знань [3].

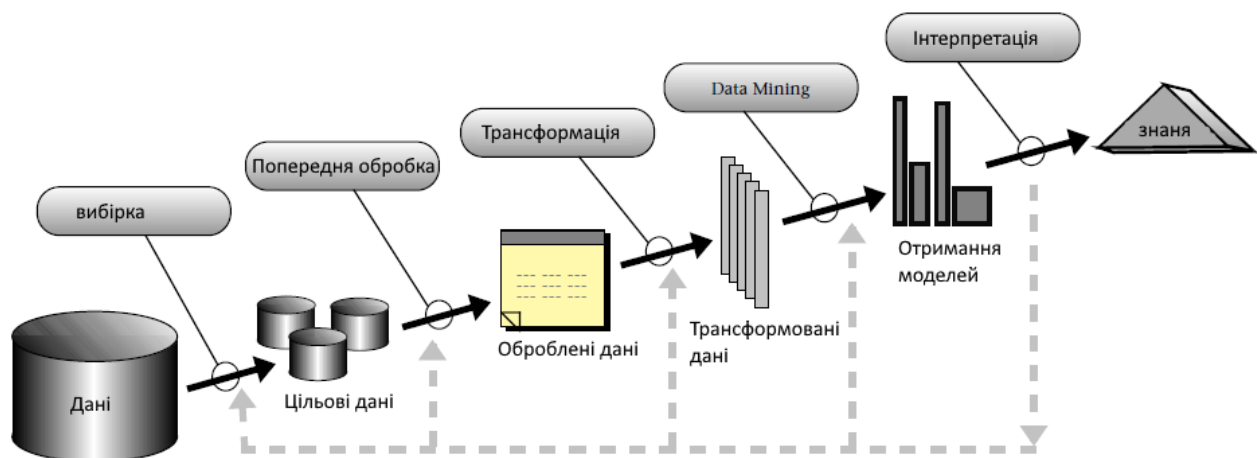


Рис. 1 Процес отримання знань [3]

Процес отримання знань складається з наступних етапів. С початку проводиться вибірка даних. На цьому етапі відбираються данні з бази даних, що в подальшому будуть проаналізовані. Далі проводиться попередня обробка. На цьому етапі проводиться видалення шуму та непослідовних даних; поєднання декількох джерел даних. Наступний етап це трансформація даних. На цьому етапі проводиться перетворення даних у відповідні форми для здійснення обміну даними. На далі використовується алгоритм Data Mining. На цьому етапі поводится вибір алгоритму для дослідження даних, який відповідає потребам предметної області. В процесі роботи алгоритмів Data Mining будуть побудовані моделі даних відповідно до обраних алгоритмів. Останній етап інтелектуального аналізу даних це Інтерпретація та Оцінка. На цьому етапі проводиться Інтерпретація отриманих моделей даних у знання шляхом видалення надмірних або неактуальних примірників. Після цього проводиться представлення корисних моделей даних на терміни, що зрозумілі для людини. У роботах [8,9] було розглянуто можливість використання обчислювальних підходів Data Mining для аналізу освітніх даних та для вивчення освітніх питань. У роботі [8] були розглянуті найбільш актуальні дослідження, проведені в галузі навчання.

Дослідженням технологій використання технологій Data Mining займались також вітчизняні вчені. У навчальних посібниках [10,11] були розглянуті і досліджені актуальні проблеми Data Mining. Виклад матеріалу у цих посібниках було зосереджено на задачах класифікації, кластеризації та побудови асоціативних правил.

**Мета статті.** На підставі дослідженого та проаналізованого матеріалу наукових досліджень було запропоновано використання технології Data Mining в предметній області «Бібліотека». Та подальшої її інтеграції до систем аналізу кошику споживача та розглянути можливість застосування даної технології в інформаційній системі бібліотеки. Основна мета статті полягає в розгляді та запропоновані можливі варіанти використання алгоритмів Data Mining в предметній області «Бібліотека». Для реалізації цієї задачі в статті викладено побудову інформаційної системи «Бібліотеки» та побудована інформаційна модель системи аналізу даних за допомогою технології Data Mining.

**Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.**

Основна ідея технології Data Mining полягає у виявленні в великих наборах даних прихованих закономірностей, залежностей і взаємозв'язків, корисних при прийнятті рішень на різних рівнях управління. У результаті роботи алгоритмів Data Mining будуть отримані закономірності, які повинні представлятися у вигляді інтелектуальних, або дескриптивних моделей різного типу, що дозволяють проводити класифікацію ситуацій або об'єктів, прогнозувати їх поведінку, виявляти групи подібних об'єктів і т.п. Істотно, що моделі будуються автоматично на основі аналізу наявних даних про об'єкти, спостереженнях і ситуаціях за допомогою спеціальних алгоритмів.

Методи Data Mining застосовуються в різних областях для вирішення таких практичних завдань, як виявлення можливої некредитоспроможності клієнтів, управління відтоків клієнтів, боротьба з шахрайством, наприклад, виявлення підозрілих операцій з кредитними картками, аналіз споживчого кошика, підвищення ефективності маркетингових заходів, персоналізоване просування продуктів і послуг, виявлення причин збоїв обладнання і так далі.

В сучасних СУБД вже вбудовані технології Data Mining. Вони складаються з процедур, що реалізують різні алгоритми побудови моделей регресії, класифікації кластеризації. Починаючи з версії Oracle 12c підтримується наступні алгоритми, які відображені на Рис. 2.

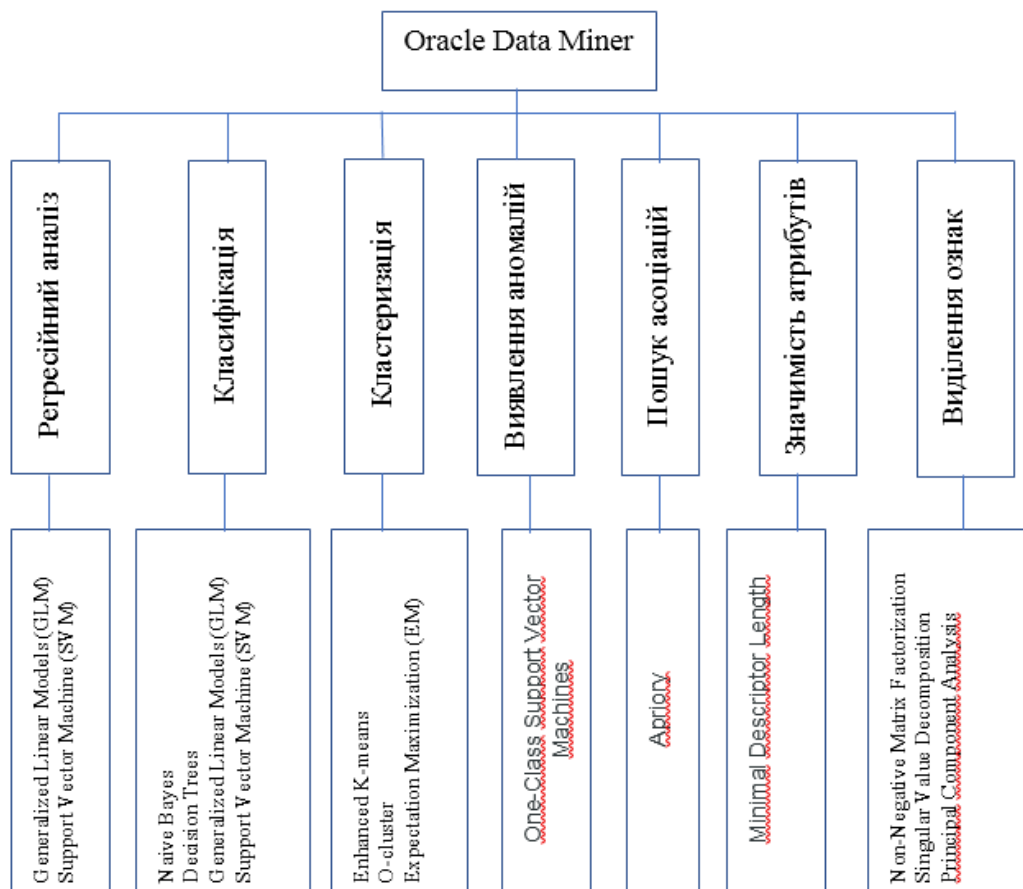


Рис. 2 Алгоритми Oracle Data Mining

Ці алгоритми можна використовувати в різних предметних галузях. Наприклад, правила асоціації можна застосувати подібним чином, як їх застосовують для аналізу ринкової корзини в супермаркетах для бібліотеки, щоб дізнатись які з книжок користуються більшим попитом. Аналізуючи раніше замовлені книги, можна побудувати асоціацію і дізнатись, які книги будуть замовлені в майбутньому і з якою вірогідністю, цю інформацію можна використовувати для замовлення більшої кількості екземплярів. Класифікацію можна використовувати для віднесення нової книги до одного з встановлених жанрів. Регресію можна використати для аналізу в часі замовлених і не повернутих книг. Кластеризація може бути використана для групування книг замовлених однією людиною.

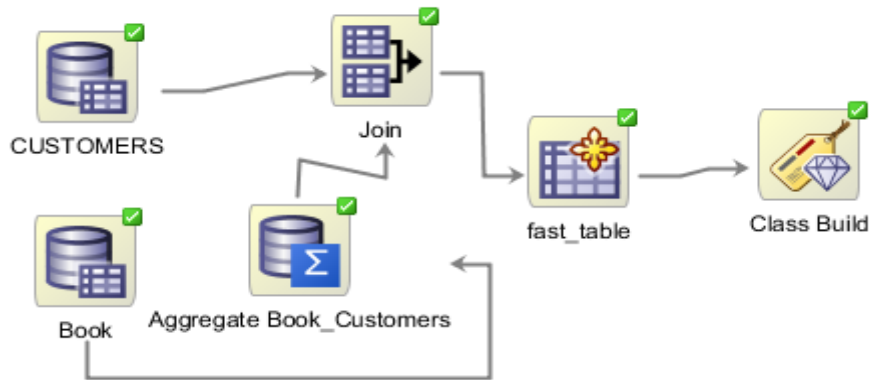


Рис. 3 Кластеризація здійснення замовлень книг

Для визначення порушень у роботі інформаційної системи бібліотеки запропоновано використовувати технологію «Виявлення аномалій». Цільовими користувачами запропонованої системи будуть працівники бібліотеки та читачі. Розроблювана система може зменшити обсяг паперової роботи, а також допомогти виявити порушення, та покращити рівень роботи і обслуговування, використовуючи сучасні технології інтелектуального аналізу даних.

Використовуючи графічний інтерфейс Oracle Data Miner було створено модель інтелектуального аналізу даних операцій замовлення книг, зображену на рис. 4.

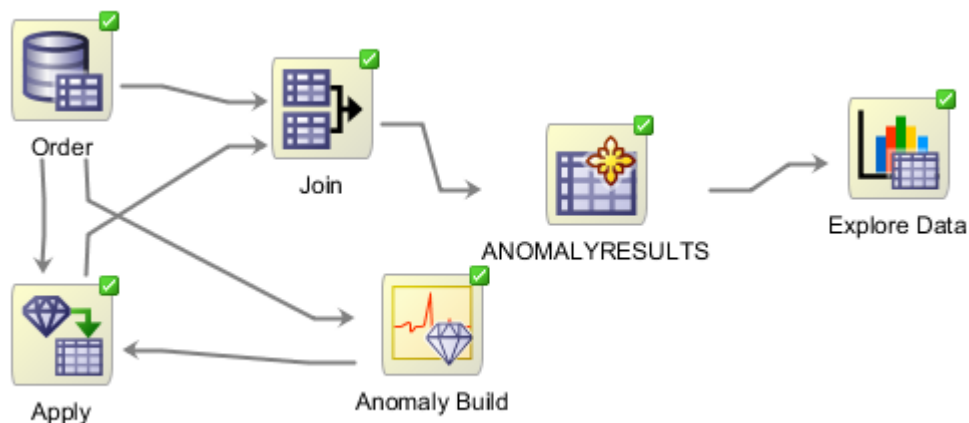


Рис. 4 Пошук аномалей при замовленні книжок

Джерелом даних є таблиця про користування книжками LIBRARY. Цільовий атрибут аналізу orderId та amount - ідентифікатор та обсяг замовлення. Після того як аналіз аномалій буде виконано, дані аналізу будуть накладатися на дані в таблиці Orders і в кінцевому результаті отримується таблиця, де наведена інформація про операції, які операції є аномальними, процентна ймовірність того чи іншого висновку.

Виявити аномальні записи не так і складно, найскладнішою і важливішою задачею під час аналізу аномалій є встановлення чиннику аномалій. Виконавши аналіз даних в результатуючій таблиці було встановлено що аномалію викликають атрибути, які стосуються взятих і не повернених книг. Для отримання таблиці, яку можна використовувати для подальшого прийняття рішень проводиться необхідна фільтрація даних. Так можна обрати лише аномальні записи, після чого зменшити їх кількість за процентним показником, який вказує наскільки аномалія є вірогідною в записі, далі можна обрати необхідну інформацію пов'язану з аномальними записами, наприклад контактні дані того хто не повернув багато книг таблицю з організованими таким чином даними можна використовувати надалі в обробці для виявлення ризиків та прийняття необхідних заходів. Також можна проводити аналіз про закупівлю бібліотекою книжок які не користуються попитом.

Методологія моделювання інформаційних систем IDEF0 може використовуватись для любого типу предметних областей. В роботі модель IDEF0 використовується для моделювання діяльності бібліотеки. На рисунку 5 наведена декомпована модель автоматизованої інформаційно-пошукової системи бібліотеки в нотації IDEF0. Як ми бачимо з рисунку 5 основну діяльність бібліотеки можливо поділити на 6 блоків:

- Оформлення книжок, які надходять до бібліотеки;
- Ведення обліку книжок і попиту;
- Ведення карток читачів;
- Видача книжок читачам;
- Приймання книжок у читачів;
- Проведення списання книжок та переміщення книжок у книго сховище.

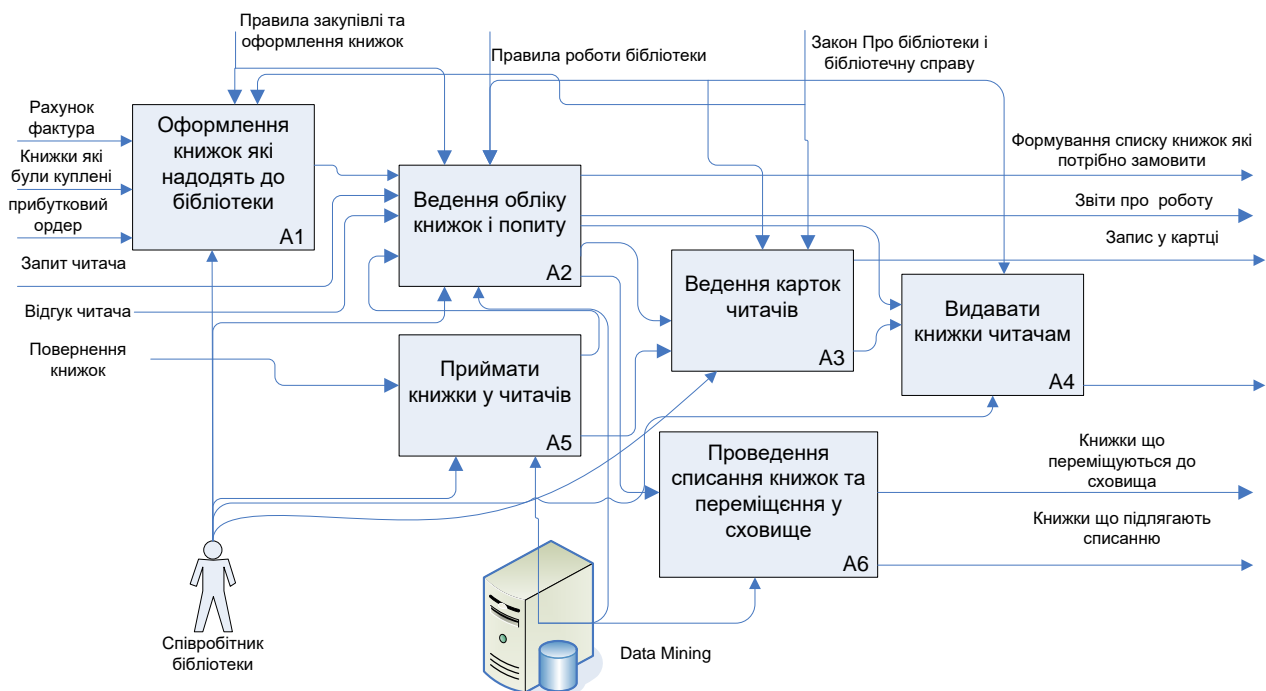


Рис. 5 Декомпована модель діяльності бібліотеки

Бібліотечний фонд формується на підставі запитів кафедр (книжки що необхідні для навчання), запитів студентів, запитів співробітників ВУЗа. На підставі цих запитів формується бібліотечний фонд. Запропонована автоматизована інформаційна система проводить аналіз існуючих у бібліотеці книжок, на основі попиту на книжки за допомогою технології Data Mining, та надає поради щодо списання книг, переміщення книг у сховище, та закупівлі книжок.

Бібліотека складається з декількох відділів. Кожен відділ має назву. В кожному відділі є зберігаються книжки з певними рубриками. Кожна книга має назву, авторів, рік видання, та кількість сторінок. Одна і та сама книжка може відноситись до декількох рубрик одного відділу. Книга може бути замовленою якщо вона є в наявності. З бібліотекою взаємодіють користувачі: адміністратори, бібліотекари та читачі. Бібліотекар може закрити доступ до книг читачам, що не повернули книги. Адміністратор може оновити перелік книг та змінити права бібліотекара. Для демонстрації

використання техніки Anomaly Detection було створено невелику базу даних з декількох таблиць, що зв'язані між собою, коли в центрі є головна таблиця з якою зв'язана решта, за допомогою механізму зовнішніх ключів.

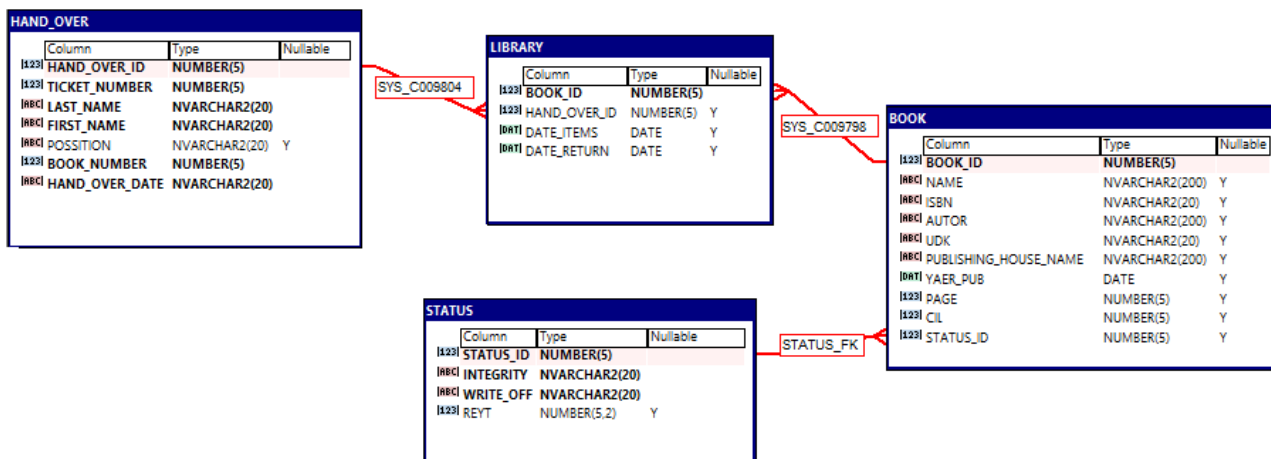


Рис. 6 – ER діаграма

Для роботи Data Mining було створено користувача Dataminer зі схемою DATAMINER. Дані на сервері було організовано наступним чином: одна схема LIBRARY слугує для накопичення даних бібліотеки зберігає описані на ER моделі вище таблиці і їх дані, друга схема DATAMINER слугує для маніпуляцій під час інтелектуального аналізу даних і збереження таблиць з результатами прогнозування. Модель використовує цільові дані, які зберігаються в одній схемі та зберігає отримані результати (прогнози), а також таблиці з проміжними даними в схемі DATAMINER. Дана архітектура забезпечує стабільну і безпечну роботу, без втрати цінних даних. В таблиці статус у стовпці Readability знаходиться такий показник, як читаність книжок який підраховується щорічно.

$$R=N/U \quad (1)$$

N це кількість разів певна книга була видана за рік;

U – кількість читачів зареєстрованих у бібліотеці.

З часом цінність книжок зменшується є книжки що були закуплені та не використовувались не жодного разу. Тому для підрахунку цінності книжок використовуємо формулу.

$$Цк= N/k \quad (2)$$

Де k це кількість книжок, яка зберігається у бібліотечному фонді. Зараз у бібліотеках знаходиться велика кількість книжок, які вже застаріли та не використовуються вже певний час. Тому для визначення книжок, які підлягають списанню та переміщенню у книгосховище пропонується використовувати лінійну регресію рис. 7. На осі Y відображається цінність книжок яка підраховується за формулою (2), на осі X відображається кількість книг у бібліотеці. Книги, які знаходяться нижче лінії підлягають списанню, чи переміщенню у книгосховище.

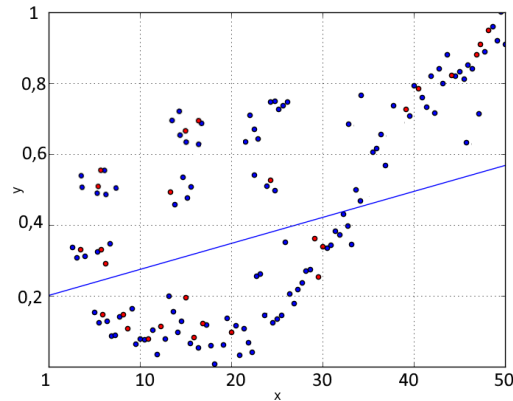


Рис. 7 Ленійна регресія

### Висновки та перспективи подальшого дослідження.

В ході виконання роботи було розроблено та створено: інформаційну модель. В перше запропоновано використання сучасних технологій Data Mining в бібліотечній діяльності для пошуку аномалій та визначення книжок для списання та переміщення в книгосховище.

Для створення бази даних, що буде зберігати всю необхідну інформацію про користувачів, книги та замовлення було обрано СКБД Oracle, її структура, особливості та можливості реалізації технології інтелектуального аналізу даних дають змогу гнучко працювати з SQL реляційною базою даних; було проаналізовано і досліджено останні тенденції розвитку інформаційних технологій, популярність використання засобів інтелектуального аналізу організаціями; тенденції щодо запровадження використання даної технології для вдосконалення роботи системи контролю ризиків.

Набула подальший розвиток можливість використання технології Data Mining для аналізу операцій замовлення та видачі книжок з метою оптимального навантаження системи. Результати даної роботи можуть бути використані для подальших досліджень і розробок, а також для запровадження використання технології Data Mining у бібліотечній діяльності.

### Список бібліографічного опису

1. Brendan Tierney Predictive Analytics Using Oracle Data Miner: Develop & Use Data Mining Models in Oracle Data Miner, SQL & PL/SQL: - Oracle Press, 2014. – 429 - ISBN: 978-0-07-182175-9
2. Data Mining software [Electronic resource] – Access mode: <https://www.predictiveanalyticstoday.com/top-data-mining-software/>
3. Fayyad U., Piatetsky-Shapiro G., Smyth P. From Data Mining to Knowledge Discovery in Databases. //AI Magazine, 17(3), 1996, P. 37-54.
4. Ashish D. A Systematic Review on Educational Data Mining./ Ashish D. , Maizatul A. Ismail1 Tutut H. //IEEE Access ( Volume: 5 ) P. 15991 – 16005. Doi: 10.1109/ACCESS.2017.2654247.
5. Oracle Advanced Analytics Customer Success Stories [Electronic resource] – Access mode: <https://www.oracle.com/technetwork/database/options/advanced-analytics/odm/odm-customers-086483.html>
6. Free data mining software [Electronic resource] - Access mode: <https://www.predictiveanalyticstoday.com/top-free-data-mining-software/#>
7. What is DM? [Electronic resource] – Access mode: <https://www.itweek.ru/idea/article/detail.php?ID=51482>
8. Romero C. Educational Data Mining: a Survey from 1995 to 2005. / Romero, C., Ventura, S. // Expert Systems with Applications. – 2007. – No 1. – Vol 33. – P. 135–146.
9. Baker R. The State of Educational Data Mining in 2009: A Review and Future visions. / Baker R., Yacef K. // Journal of Educational Data Mining. – 2009. – Vol 1. – No 1.— Pp. 3–17.
10. Шумейко А. А., Сотник С. Л. Интеллектуальный анализ данных (Введение в Data Mining) : учеб. пособ. // Днепропетровск: Белая Е. А., 2012. – 212 с.
11. Марченко О.О., Россада Т.В. Актуальні проблеми Data Mining: Навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. // Київ. -2017. -150 с.

### References

1. Brendan Tierney Predictive Analytics Using Oracle Data Miner: Develop & Use Data Mining Models in Oracle Data Miner, SQL & PL/SQL: - Oracle Press, 2014. – 429 - ISBN: 978-0-07-182175-9
2. Data Mining software [Electronic resource] – Access mode: <https://www.predictiveanalyticstoday.com/top-data-mining-software/>
3. Fayyad U., Piatetsky-Shapiro G., Smyth P. From Data Mining to Knowledge Discovery in Databases. //AI Magazine,

- 17(3), 1996, P. 37-54.
4. Ashish D. A Systematic Review on Educational Data Mining./ Ashish D. , Maizatul A. Ismail1 Tutut H. //IEEE Access ( Volume: 5 ) P. 15991 – 16005.Doi: 10.1109/ACCESS.2017.2654247.
  5. Oracle Advanced Analytics Customer Success Stories [Electronic resource] – Access mode: <https://www.oracle.com/technetwork/database/options/advanced-analytics/odm/odm-customers-086483.html>
  6. Free data mining software [Electronic resource] - Access mode: <https://www.predictiveanalyticstoday.com/top-free-data-mining-software/#>
  7. What is DM? [Electronic resource] – Access mode: <https://www.itweek.ru/idea/article/detail.php?ID=51482>
  8. Romero C. Educational Data Mining: a Survey from 1995 to 2005. / Romero, C., Ventura, S. // Expert Systems with Applications. – 2007. – No 1. – Vol 33. – P. 135–146.
  9. Baker R. The State of Educational Data Mining in 2009: A Review and Future visions. / Baker R., Yacef K. // Journal of Educational Data Mining. – 2009. – Vol 1. – No 1.— Pp. 3–17.
  10. Shumejko A. A., Sotnik S. L. Intellektual'nyj analiz danny'kh (Vvedenie v Data Mining) : ucheb. posob. // Dnepropetrovsk: Belaya E. A., 2012. – 212 s.
  11. Marchenko O.O., Rossada T.V. Aktual'ni` problemi Data Mining: Navchal`nij posi`bnik dlya studentiv`v fakul`tetu komp'yuternikh nauk ta ki`bernetiki. // Kiyiv. -2017. -150 s.