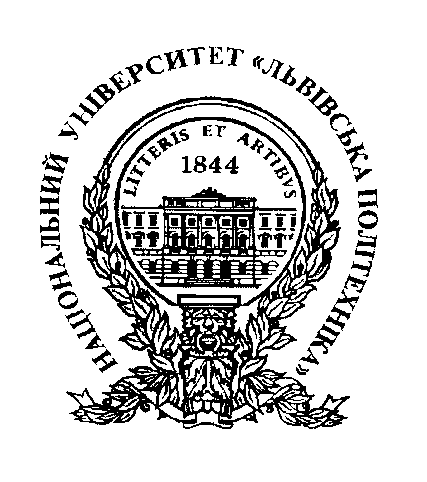
**Міністерство Освіти І НАУКИ України**

**Національний університет "Львівська політехніка"**

Інститут **КНІ**

Кафедра **ПЗ**



***ПОЯСНЮВАЛЬНА ЗАПИСКА***

**На тему:** *“ Дослідження роботи алгоритму apriori для напівструктурованих даних”*

**З дисципліни:** *“ Методи та засоби досліджень в ІПЗ ”*

**Виконав:**

ст. гр. ПЗІП-12

Гаврилюк А.М.

**Перевірив:**

професор каф. ПЗ

Федасюк Д.В.

**Керівник:**

доцент каф. ПЗ

Левус Є.В.

« \_\_\_\_ » \_\_\_\_\_\_\_\_ 2017 р.

∑= \_\_\_\_ .

Львів – 2017

**Зміст**

[1. АНОТАЦІЯ 3](#_Toc501391519)

[2. ПРОБЛЕМАТИКА ДОСЛІДЖЕННЯ 4](#_Toc501391520)

[2.1. Прикладна проблема, на вирішення якої спрямовано проект: 4](#_Toc501391521)

[2.2. Об’єкт дослідження і розробки: 5](#_Toc501391522)

[2.3. Предмет дослідження і розробки: 5](#_Toc501391523)

[3. СТАН ДОСЛІДЖЕНЬ ПРОБЛЕМИ І ТЕМАТИКИ 5](#_Toc501391524)

[4. МЕТА, ОСНОВНІ ЗАВДАННЯ ТА ЇХ АКТУАЛЬНІСТЬ 6](#_Toc501391525)

[4.1. Ідеї та робочі гіпотези проекту: 6](#_Toc501391526)

[4.2. Мета і завдання, на вирішення яких спрямовано проект: 7](#_Toc501391527)

[5. ПІДХІД, МЕТОДИ, ЗАСОБИ ТА ОСОБЛИВОСТІ ДОСЛІДЖЕНЬ І РОЗРОБОК ЗА ПРОЕКТОМ 7](#_Toc501391528)

[6. ОЧІКУВАНІ РЕЗУЛЬТАТИ ВИКОНАННЯ ПРОЕКТУ ТА ЇХ НАУКОВА НОВИЗНА 8](#_Toc501391529)

[6.1. Докладно представити зміст очікуваних результатів, навести попередні описи методик та продуктів-аналогів: 8](#_Toc501391530)

[6.2. Заплановано представлення доповідей і наукових публікацій 9](#_Toc501391531)

[7. ПРАКТИЧНА ЦІННІСТЬ ПРОЕКТУ 9](#_Toc501391532)

[8. ЛСМ 9](#_Toc501391533)

[8. Календарний графік виконання робіт і фінансування, подані у вигляді діаграми Ганта. 13](#_Toc501391534)

[Висновки 14](#_Toc501391535)

[Список використаної літератури 15](#_Toc501391536)

**Структура пояснювальної записки для наукового проекту студентів**

**Назва проекту:** Дослідження роботи алгоритму apriori для напівструктурованих даних

**Пропоновані строки виконання проекту:** 03.09.2017 – 25.05.2019

**Обсяг фінансування:** 68 135 грн.

1. **АНОТАЦІЯ**

ІТ є однією з найбільш динамічніших сфер, активно розробляються нові технології, створюються сотні програмних продуктів та накопичаються приблизно декілька сотень петабайт інформації у день. Ця інформація надходить з різних джерел, та містить багато напівструктурованих даних.

Прикладами напівструктурованих даних є XML та JSON. Структуровані дані зберігаються у базах даних. Вони легко опрацьовуються,отже з них легко отримати результат.

Big Data відноситься до надзвичайно великих масивів даних, які важко аналізувати традиційними інструментами. Великі дані можуть включати в себе як структуровані, так і неструктуровані дані. Багато інструментів, призначених для аналізу великих даних, можуть обробляти напівструктуровані дані.

Також для досягнення конкурентоспроможності у бізнесі реалізуються багато алгоритмів, що призначені опрацьовувати такі дані, та шукати у них залежності, асоціативні правила. Найбільш відомим та широко застосовуваним зараз є алгоритм Apriori. Отже, у магістерському дослідженні буде реалізовано алгоритм і використано для оцінки його роботи на структурованих та напівструктурованих великих даних.

Реалізована під час магістерського дослідження система у подальшому може бути використана як основа для створення рекомендаційної системи, оскільки використовує стек технологій Hadoop екосистеми, що дозволяє опрацювувати величезні масиви даних. А враховуючи темпи накопичення даних у найближчі роки Big Data стане найпопулярнішим напрямком ІТ.

Багато спостерігачів даної галузі кажуть, що Hadoop став де-факто галузевим стандартом для управління великими даними. Це проект з відкритим кодом управляється Apache Software Foundation.Основною метою проекту є покращення рівня успішності реалізації проектів з розробки програмного забезпечення з використанням аналізу специфікації вимог та інших документів, за допомогою засобів нейронних мереж.

1. **ПРОБЛЕМАТИКА ДОСЛІДЖЕННЯ**
   1. **Прикладна проблема, на вирішення якої спрямовано проект:**

Основною прикладною проблемою на вирішення якої спрямовується проект є деталізація та формалізація моделі управління ризиками розроблення ПЗ та створення системи яка б використовувала дану модель. Також потрібно створити механізм прогнозування провалу проекту на основі даних про проект.

Дана проблема розв’язується шляхом аналізу існуючих моделей та використання засобів нейронної мережі, що дозволить отримувати найбільш точні результати враховуючи процес навчання мережі, на основі бази даних інформації про проекти які вже завершились.

* 1. **Об’єкт дослідження і розробки:**

Об`єктом дослідження і розробки є процес аналізу управління ризиками.

* 1. **Предмет дослідження і розробки:**

Предметом дослідження і розробки є модель управління ризиків.

1. **СТАН ДОСЛІДЖЕНЬ ПРОБЛЕМИ І ТЕМАТИКИ**

В ході аналізу предметної області було розглянуто ряд статей та публікацій. Було проаналізовано міжнародні стандарти з управління ризиками[1, 2] та проведено ознайомлення з основними моделями управління ризиків.

Існують різні моделі управління ризиками [3-7], серед яких найбільшого використання набула модель Інституту програмної інженерії Карнеги-Меллон (англ. Software Engineering Institute, SEI), що містить як вимоги стандартів [1, 2], так і відомі кращі практичні рекомендації щодо їх запобігання чи знешкодження. Хоча модель SEI й наведена у вигляді текстових рекомендацій та плану виконання дій, однак у ній відсутній формалізований метод управління ризиками, що призводить, зазвичай, до специфічного її використання та вільного трактування отриманих результатів.

Також, проведено аналіз публікацій із застосування нейронних мереж[], це дозволило краще зорієнтуватись в необхідності використання цього засобу при розробці даної системи.

1. **МЕТА, ОСНОВНІ ЗАВДАННЯ ТА ЇХ АКТУАЛЬНІСТЬ**
   1. **Ідеї та робочі гіпотези проекту:**

Основною ідеєю проекту є створення формальної математичної моделі, та розробити систему яка буде базуватись на цій моделі.

Перша гіпотеза полягає у створенні і впроваджені нейронної мережі яка буде прогнозувати ймовірність провалу проекту за допомогою аналізу інформації, щодо проекту який використовує дану систему.

Друга гіпотеза: якщо застосувати систему для реального проекту, і вона зарекомендує себе з позитивної сторони, то необхідно застосовувати систему для інших проектів.

* 1. **Мета і завдання, на вирішення яких спрямовано проект:**

Основною метою проекту є покращення рівня успішності реалізації проектів з розробки програмного забезпечення з використанням аналізу специфікації вимог та інших документів.

Виходячи з мети, необхідно розв’язати такі завдання:

1.      Ознайомлення із літературними джерелами в даній предметній області.

2.      Ознайомлення із існуючими моделями управління ризиків, які базується на текстових рекомендаціях та планах виконання дій.

3.      Вибір моделі управління ризиками.

4.      Поділ джерел появи ризиків та потенційно ризикових подій по групам.

5.      Визначення рівнів пріоритезації окремих груп джерел появи ризиків та потенційно ризикових подій.

6.      Створення формальної математичної моделі.

7.      Визначення адекватності створеної формальної математичної моделі, на основі експериментальних проектів.

8.  Створення програмної системи, яка на основі створеної формальної математичної моделі, буде визначати ймовірність провалу проекту.

9.  Створення нейронної мережі, яка на основі експериментальних даних навчиться визначати чи проект буде успішним, чи провалиться(на основі пошуку проектів які мають аналогічні показники прояву джерел появи ризиків, та потенційно ризикових подій.

1. **ПІДХІД, МЕТОДИ, ЗАСОБИ ТА ОСОБЛИВОСТІ ДОСЛІДЖЕНЬ І РОЗРОБОК ЗА ПРОЕКТОМ**

Розробка та тестування програмного продукту буде здійснюватись на мові С#, у середовищі Visual Studio 2017 Enterprice. Також, планується використовувати базу даних MS SQL Server, і нейронну мережу для прогнозування провалу проекту. Ці технології були обрані тому що, вони надають необхідний набір можливостей для досягнення поставленої мети у розробці даного наукового проекту.

Особливості структури та складових проведення досліджень:

1. Аналіз публікацій та статей за тематикою дослідження.
2. Аналіз переваг і недоліків існуючих моделей управління ризиками.
3. Створення формальної математичної моделі управління ризиками.
4. Дослідження логіки роботи нейронної мережі для прогнозування провалу проекту.
5. Розробка системи контролю ризиками.
6. Створення і впровадження нейронної мережі для даної системи.
7. Тестування розробленої системи та аналіз результатів.

В загальному структуру проведення дослідження можна розділити на 3 етапи:

1. Аналітична частина, яка включає аналіз публікацій, створення та перевірка моделі.
2. Створення системи контролю ризиками.
3. Створення нейронної мережі для даної системи.
4. **ОЧІКУВАНІ РЕЗУЛЬТАТИ ВИКОНАННЯ ПРОЕКТУ ТА ЇХ НАУКОВА НОВИЗНА**
   1. **Докладно представити зміст очікуваних результатів, навести попередні описи методик та продуктів-аналогів:**

У проекті передбачається :

* Створення моделі управління ризиками.
* Перевірка моделі на адекватність.
* Збирання та наповнення бази даних з даними про проекти які були завершені, щоб обійти всі можливі варіанти провалу проекту.
* Створення системи.
* Створення нейронної мережі.
* Тестування даної системи.
* Перевірка системи для реальних проектів.

На даний час існує багато моделей управління ризиків, але вони базуються лише на описі, та не містять математичного апарату.

Кінцевим результатом проекту повинен бути програмний продукт, який дозволить управляти ризиками на всіх етапах розробки програмного забезпечення, та зможе прогнозувати ймовірність провалу проекту за певною вхідною інформацією про проект на основі нейронної мережі.

До основних вимог до моделі можна віднести: адекватність, відповідність моделі міжнародним стандартам, перевірка моделі для різноманітних проектів.

Вимоги до системи: мінімальний час на відповідь системи на певні дії користувача, повідомлення користувачу статусу виконання довготривалих операцій системи.

* 1. **Заплановано представлення доповідей і наукових публікацій**

В результаті виконання даного наукового проекту заплановано виступ на конференціях, та публікації. При можливості публікація буде видана англійською мовою, та попаде в науково метричні бази даних.

1. **ПРАКТИЧНА ЦІННІСТЬ ПРОЕКТУ**

В результаті програмної реалізації розроблених методів планується створення системи, яка зможе прогнозувати провал проекту за допомогою засобів нейронних мереж, на основі створеної формальної математичної моделі та даних про проект який буде використовувати цю систему.

Основними користувачами системи будуть проектні менеджери, або інші учасники команди з розробки програмного забезпечення на яких будуть поставлені обов’язки з контролю управління ризиками проектів.

Очікувані наукові результати направлені на:

* подальше покращення створеної моделі;
* застосування системи на нових не стандартних проектах;
* додавання нового функціоналу до системи, який покращить реалізацію програмних продуктів.

2. **Логіко-структурна матриця.**

*Таблиця 1. Логіко-структурна матриця*

|  |  |  |  |
| --- | --- | --- | --- |
| Ширші цілі. | Показники прогресу. | Як оцінювати показники. |  |
| Що являє собою більш широку ціль, якій сприятиме проект? | Які основні показники, пов’язані з більш ширшою метою? | Які є джерела інформації за цими показниками? |  |
| Сприяти покращенню рівня успішності реалізації проектів з розробки програмного забезпечення, на основі управління ризиків, з використанням аналізу специфікації вимог та інших документів. | Рівень успішності реалізації проектів з розробки програмного забезпечення. Даний показник  можна розділити на три взаємозалежні параметри успішності:  1.      Проект закінчений відповідно до поставленого бюджету;  2.      Проект закінчений відповідно до поставлених часових обмежень;  3.      Проект закінчений з виконанням всіх поставлених функціональних можливостей. | На основі експериментальних даних про проекти які вже були реалізовані, однак потерпіли невдачі, необхідно на основі експертних оцінок визначити чи даний продукт покращив би параметри успішності(бюджет, час, функціональність) |  |
| Конкретні цілі. | Показники прогресу. | Як оцінювати показники. | Припущення та ризики. |
| Які є особливі цілі, які досягне проект? | Які є кількісні та  якісні індикатори, що показують чи та якою мірою конкретні цілі проекту  буде досягнуто? | Які джерела інформації існують і які можуть бути  зібрані? Які методи  є необхідні для отримання цієї інформації? | Які чинники та  умови, які не перебувають під прямим контролем проекту, необхідні для досягнення цих цілей? Які небезпеки необхідно взяти до уваги? |

*Продовження таблиці 1.*

|  |  |  |  |
| --- | --- | --- | --- |
| 1.      Визначення можливих джерел появи ризиків.  2.      Ідентифікація потенційних ризикових подій.  3.      Створення формальної математичної моделі.  4.      Визначення ймовірності провалу експериментальних проектів.  5.      Оцінка створеної системи на основі випробувань. | Джерела появи ризиків та потенційно ризикові події повинні відповідати стандарту, на основі експертної оцінки.    Відсоток правильних визначень ймовірності провалу проектів, для експериментальних даних.(більше 75%, для більше 10 експериментальних проектів) | Стандарти: ISO-31000 Risk Management.    База даних із даними про експериментальні проекти.    Звіт про випробування системи.    Порівняння даних які надаватиме система з статистичними даними про проекти які взяті як експериментальні зразки. | Поява конкурентів.  Додавання подібного функціоналу до вже існуючих систем для управління проектами.  Не можливість отримати необхідний відсоток правильних визначень ймовірностей провалу.    Якість програмної системи може бути незадовільною. |
| Підсумки і результати | Показники прогресу. | Як оцінювати показники. | Припущення та ризики. |
| Список конкретних результатів - підсумки / результати, які призведуть до вирішення конкретних цілей | Які є показники  для вимірювання  чи і в якій мірі проект досягає запланованих резуьтатів та ефекту? | Які є джерела інформації за  цими показниками? | Які зовнішні чинники і умови повинні бути реалізовані для отримання очікуваних результатів та результатів за графіком? |

*Продовження таблиці 1.*

|  |  |  |  |
| --- | --- | --- | --- |
| 1.      Створена формальна математична модель.  2.      Програмна система з управління ризиками створена.  3.      Проведено випробування створеної системи на основі формальної математичної моделі на тестових, експериментальних проектах. | Формальна математична модель є адекватною для різноманітних експериментальних проектів.    Програмна система реалізована та протестована на експериментальних проектах(10+).    Впровадження системи в розробку першого реального проекту. | Експертна оцінка адекватності моделі.    Звіт про виконання тестування системи.    Технічне завдання реального проекту, в якому міститься інформація про використання даної системи. | Експертна оцінка має велику похибку, тому результати оцінки можуть бути хибними.  Тестовий набір експериментальних проектів може бути не створения, або ж не включати в себе всю необхідну множину даних для опрацювання всіх можливих варіантів роботи системи.  Не можливість застосування системи для реального проекту. |
| Дії | Ресурси. | Витрати. | Припущення, ризики та передумови. |
| Які основні заходи будуть здійснюватись (згруповані у етапи роботи) та в якій послідовності вони будуть проводитись для того, щоб отримати очікувані результати? | Які ресурси  необхідні для здійснення цих видів діяльності? | Якою буде загальна вартість реалізації проекту? | Які попередні умови потрібні до початку проекту? Які умови  поза  прямого контролю проекту повинні бути присутніми  для здійснення запланованих заходів? |

*Продовження таблиці 1.*

|  |  |  |  |
| --- | --- | --- | --- |
| 1.      Ознайомлення із літературними джерелами в даній предметній області.  2.      Ознайомлення із існуючими моделями управління ризиків, які базується на текстових рекомендаціях та планах виконання дій.  3.      Вибір моделі управління ризиками.  4.      Поділ джерел появи ризиків та потенційно ризикових подій по групам.  5.      Визначення рівнів пріоритезації окремих груп джерел появи ризиків та потенційно ризикових подій.  6.      Створення формальної математичної моделі.  7.      Визначення адекватності створеної формальної математичної моделі, на основі експериментальних проектів.  8.  Створення програмної системи, яка на основі створеної формальної математичної моделі, буде визначати ймовірність провалу проекту.  9.  Створення нейронної мережі, яка на основі експериментальних даних навчиться визначати чи проект буде успішним, чи провалиться(на основі пошуку проектів які мають аналогічні показники прояву джерел появи ризиків, та потенційно ризикових подій)  10.      Пошук проектів, які вже реалізовані з невтішними наслідками, та витягнення з них експериментальних даних.  11.      Створення бази даних, яка містить експериментальні дані про проекти.  12.  Реалізація хоча б одного реального проекту з використанням створеної системи. | Персональний комп’ютер з всім додатковим обладнанням.(1 шт.)    Принтер.(1 шт.)    Папір А-4(500 шт.)    Ручка.(2 шт.)  240 робочих днів для програміста та його керівника(2 год на день.)  Ліцензія на необхідне програмне забезпечення(Visual Studio 2017 Enterprice, Microsoft Ofice)    Диск DWD RW(1 шт.) | 30 тис. грн. + 2 тис. грн. + 100 грн. + 15 грн. +36 тис. грн. + 2 тис. грн. + 20 грн. =  70 135 грн. | Можливі додаткові витрати на техніку чи залучення додаткового персоналу.    Проект може не бути завершений за 240 робочих днів, тому можливі перевитрати бюджету та часу.    Технічна база програміста може не зовсім відповідати потребам даного проекту, тому необхідний додатковий час для його навчання потрібним технологіям. |

1. **Календарний графік виконання робіт і фінансування, подані у вигляді діаграми Ганта.**

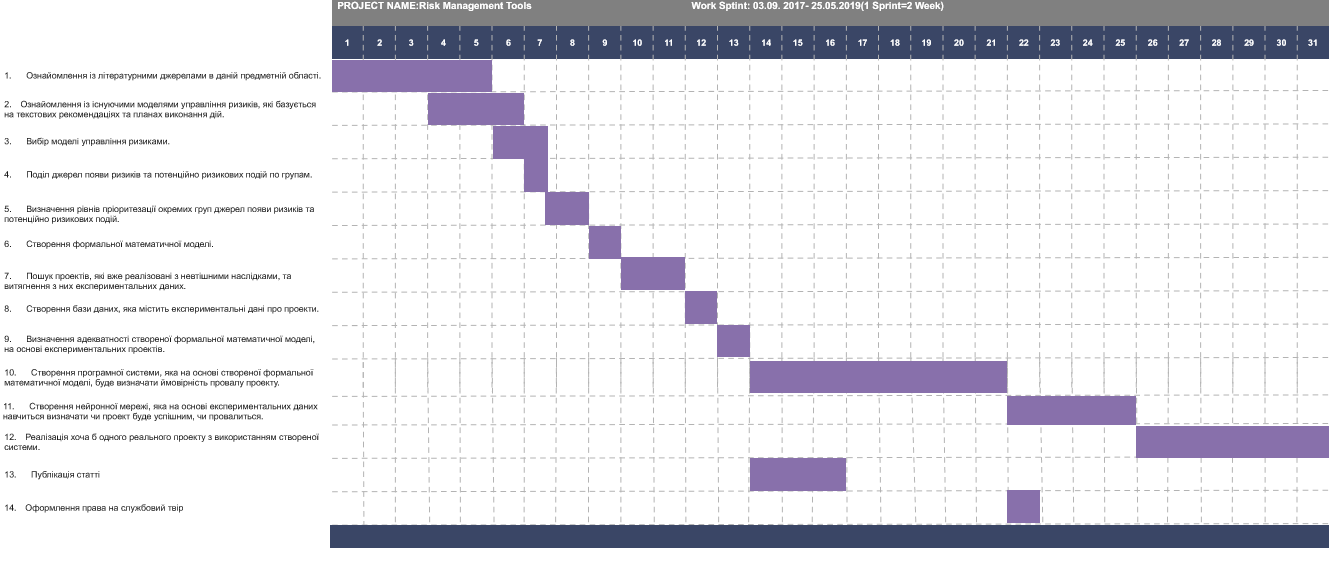


Рис.1 Графік виконання робіт у вигляді діаграми Гантта

**Висновки**

На основі проведеного аналізу було вирішено створення нової моделі для управління ризиками. На основі створеної мною моделі потрібно визначити проекти які будуть успішними, а які будуть мати певні проблеми з часом, бюджетом чи виконанням функціоналу. Щоб виконати поставлене завдання я вирішив використовувати нейронну мережу, на вхід якій будуть: характеристики проекту на поточному етапі розробки, а на виході на основі аналогічних проектів буде прогнозуватись чи проект може провалитись та з якою ймовірністю.

**Список використаної літератури**

1. ISO / IEC 12207:2008. Systems and software engineering – Software life cycle processes [Електронний ресурс] – Режим доступу: https://www.iso.org/standard/43447.html
2. ISO / IEC 31000:2015. Risk Management [Електронний ресурс] – Режим доступу: <https://www.iso.org/iso-31000-risk-management.html>
3. Abdullah Al Murad Chowdhury and Shamsul Arefeen, “Software Risk Management: Importance and Practices” [Електронний ресурс] – Режим доступу: http://ijcit.org/ijcit\_papers/vol2no1/IJCIT-110740.pdf
4. Md. Forhad Rabbi, Khan Olid Bin Mannan, “A Review of Software Risk Management for Selection of best Tools and Techniques”, [Електронний ресурс] – Режим доступу: http://ieeexplore.ieee.org/abstract/document/4617465/?reload=true
5. Linda Westfall, “SOFTWARE RISK MANAGEMENT” [Електронний ресурс] – Режим доступу: https://search.proquest.com/openview/af85ceff70c3654c8a6e9e79a9fa73d8/1?pq-origsite=gscholar&cbl=39817
6. Kathy Schwalbe, “IT Project Management” [Електронний ресурс] – Режим доступу: https://books.google.com.ua/books?hl=uk&lr=&id=mPeoBAAAQBAJ&oi=fnd&pg=PR7&dq=Kathy+Schwalbe,+%E2%80%9CIT+Project+Management%E2%80%9D&ots=FMkt-u--3d&sig=64H4dXj5Tdo3-x\_XUsg6y0F-DUw&redir\_esc=y#v=onepage&q=Kathy%20Schwalbe%2C%20%E2%80%9CIT%20Project%20Management%E2%80%9D&f=false
7. Bloch M. Delivering large-scale IT projects on time, on budget, and on value / M. Bloch, S. Blumberg, J. Laartz – October, 2012 // [Електронний ресурс] – Режим доступу: http://www.mckinsey.com / insights/business technology/delivering large-scale it projects on time on budget and on value
8. CHAOS Manifesto: Think big, act small, 2013 // [Електронний ресурс] – Режим доступу: <http://www.versionone.com/assets/img/files/CHAOSManifesto2013.pdf>
9. Risk Management Guide for DoD Acquisition [Електронний ресурс] = Руководство по управлению рисками заказов для Департамента обороны / U.S. Department of Defense, Defense Acquisition University. – Fifth Edition. – Електронні, текстові дані – Virginia (USA): DAU Press, 2002. – Режим доступу: <http://www.dau.mil/pubs/pubs-main.asp>.
10. [Green, Brian Patrick](https://search.proquest.com/indexinglinkhandler/sng/au/Green,+Brian+Patrick/$N), [Choi, Jae Hwa](https://search.proquest.com/indexinglinkhandler/sng/au/Choi,+Jae+Hwa/$N). Assessing the risk of management through neural network technology [Електронний ресурс] – Режим доступу: https://search.proquest.com/openview/87e9dd4190efa4245a81c49837900922/1?pq-origsite=gscholar&cbl=31718