Міністерство освіти і науки України

Сумський державний університет

Навчально-науковий інститут бізнесу, економіки та менеджменту

Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

на тему «DATA MINING ДЛЯ ДОСЛІДЖЕННЯ СТАНУ ЗАХИЩЕНОСТІ ІНФОРМАЦІЙНОГО ПРОСТОРУ СУМСЬКОГО ДЕРЖАВНОГО УНІВЕРСИТЕТУ»

Виконав студент IV курсу, групи ЕК.М-01а
(номер курсу) (шифр групи)

Спеціальності 051 «Економіка»

(«Економічна кібернетика»)

_____Довга Ю.О._____
(ініціали, прізвище студента)

Керівник професор, д.е.н. Кузьменко О.В.
(посада, науковий ступінь, ініціали, прізвище)

Суми – 2021 рік

РЕФЕРАТ

дипломної роботи на тему:

«Data Mining для дослідження стану захищеності інформаційного простору Сумського державного університету»

студентки

Довгої Юлії Олександрівни

Актуальність теми дослідження. Актуальність теми визначається щоденним розвитком IT-технологій, а з ними і кібератак. На фоні пандемії Covid-19 кіберзлочинність опинилася на стадії підйому і є значущим ризиком сьогодення.

Мета даної роботи полягає у побудові нейромережевої моделі дослідження стану та прогнозування можливості успішних кібератак на інформаційний простір, а саме інформаційну систему Сумського державного університету.

Об'єктом дослідження є обсяги успішних кібератак на інформаційний простір СумДУ, які приведуть до повної або часткової втрати інформації співробітників чи тимчасового виведення з ладу технічного забезпечення.

Предметом дослідження є методи і моделі аналізу та прогнозування часових рядів.

Методи дослідження. Для дослідження поставлених завдань були використані методи аналізу часових рядів та нейромережевого прогнозування.

Основний науковий результат роботи. У роботі проведено аналіз стану захищеності інформаційного простору СумДУ; розглянуто напрямки кібератак, способи та засоби їх реалізації; визначено дії для захисту від них; проаналізовано найпопулярніші способи атак; проаналізовано та порівняно методи Data Mining для прогнозування; описано моделі та сформовано вимоги до них; систематизовано та охарактеризовано вхідні дані; використано метод Ірвіна для перевірки даних на однорідність; розглянуто та порівняно найпопулярніше програмне забезпечення для побудови моделі; побудовано п'ять моделей

нейронних мереж для прогнозування; визначена адекватність побудованих моделей; побудований прогноз успішних кібератак на інформаційний простір СумДУ.

Рекомендації щодо використання результатів дослідження. Прогноз може бути використаний відділом технічного захисту інформації для прийняти правильних рішень, що покращать ситуації з небажаними втручаннями.

Інформаційною базою дипломної роботи є дані, зібрані у Центрі телекомунікаційних технологій та комп'ютерного забезпечення (ЦТТКЗ) Сумського державного університету.

Апробація результатів дослідження. Основні положення кваліфікаційної магістерської роботи використані у матеріалах статті Cyberattack Features in Smart Manufacturing and Industry 4.0, яка розглядалася на конференції 5th International Conference on Design, Simulation, Manufacturing: The Innovation Exchange (DSMIE-2022).

Ключові слова: Data Mining, кібератака, прогнозування, моделювання, нейронні мережі.

Основний зміст дипломної роботи викладено на 40 сторінках, у тому числі список використаних джерел з 54 найменування, який розміщено на 6 сторінках. Робота містить 3 таблиць, 19 рисунків, а також 1 додаток.

Рік виконання дипломної роботи – 2021 рік.

Рік захисту роботи – 2021 рік.

Міністерство освіти і науки України

Сумський державний університет

Навчально-науковий інститут бізнесу, економіки та менеджменту

Кафедра економічної кібернетики

ЗАТВЕРДЖУЮ
Завідувач кафедри
д.е.н., професор
_____ О.В. Кузьменко
“__” _____ 2021 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ МАГІСТРА
спеціальність 051 «Економіка (Економічна кібернетика)
студентки 5 курсу, групи ЕК.м-01а

_____Довгій Юлії Олександрівні_____
(прізвище, ім’я, по батькові студента)

1. Тема роботи Data Mining для дослідження стану захищеності інформаційного простору Сумського державного університету
затверджена наказом по університету від «__»_____ 2021 року № _____
2. Термін подання студентом закінченої роботи «13» грудня 2021 року
3. Мета кваліфікаційної роботи - побудова нейромережевої моделі дослідження стану та прогнозування можливості успішних кібератак на інформаційний простір, а саме інформаційну систему Сумського державного університету
4. Об’єкт дослідження - обсяги успішних кібератак на інформаційний простір СумДУ, які приведуть до повної або часткової втрати інформації співробітників чи тимчасового виведення з ладу технічного забезпечення
5. Предмет дослідження – методи і моделі аналізу та прогнозування часових рядів
6. Кваліфікаційна робота виконується на матеріалах Центру телекомунікаційних технологій та комп’ютерного забезпечення (ЦТТКЗ) Сумського державного університету_____
7. Орієнтовний план кваліфікаційної роботи, терміни подання розділів керівникові та зміст завдань для виконання поставленої мети

Розділ 1._____
АНАЛІЗ СТАНУ ЗАХИЩЕНОСТІ ІНФОРМАЦІЙНОГО ПРОСТОРУ СУМСЬКОГО ДЕРЖАВНОГО УНІВЕРСИТЕТУ
_____
(назва – термін подання)

У розділі 1. Охарактеризувати підрозділ дослідження, а саме Центру телекомунікаційних технологій та комп’ютерного забезпечення (ЦТТКЗ)

Сумського державного університету; розглянути поняття, види та засоби захисту від кібератак; поставити задачі дослідження _____
(зміст конкретних завдань до розділу, які має виконати студент)

Розділ 2._____
РОЗРОБКА МОДЕЛІ ПРОГНОЗУВАННЯ МОЖЛИВИХ ОБСЯГІВ УСПІШНИХ КІБЕРАТАК НА ІНФОРМАЦІЙНИЙ ПРОСТІР СУМДУ_____
(назва – термін подання)

У розділі 2. Розглянути і порівняти методи та моделі прогнозування Data Mining; сформувати вимоги до моделі; описати вхідні дані._____
(зміст конкретних завдань до розділу, які повинен виконати студент)

Розділ 3._____
МОДЕЛЮВАННЯ МОЖЛИВИХ ОБСЯГІВ УСПІШНИХ КІБЕРАТАК НА ІНФОРМАЦІЙНИЙ ПРОСТІР СУМДУ
_____
(назва – термін подання)

У розділі 3. Розглянути і порівняти програмне забезпечення для прогнозування; побудувати моделі нейронних мереж для прогнозування кібератак._____
(зміст конкретних завдань до розділу, які повинен виконати студент)

8. Консультації з роботи:

Розділ	Прізвище, ініціали та посада Консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1			
2			
3			

9. Дата видачі завдання: « » _____ 2021 року

Керівник кваліфікаційної роботи _____ _____
(підпис) (ініціали, прізвище)

Завдання до виконання одержав _____ _____
(підпис) (ініціали, прізвище)

ЗМІСТ

ВСТУП

Концентрація нових кібератак на ІТ-системи підприємств, організацій та приватних осіб зросла з розповсюдженням інформаційних технологій. Зауважимо, що у сучасному світі традиційні заходи і методи з кібербезпеки уже в повному обсязі не запобігають чи стримують кіберзагрози через високу швидкість та регулярність їх проявів. Система безпеки інформації на підприємстві є комплексною, включає постійний експрес-аудит загроз на ІТ-сферу підприємства, прослідковує його стратегію розвитку та інформаційну політику. Вона визначає суб'єкти загроз, їх ціль, наміри атак, вразливі до проникнення місця системи інформаційної безпеки. Для виконання таких задач необхідні нові рішення, які задовольняють не тільки реалії сьогодення, а й здатні працювати на перспективу у розвитку, беручи за основу тенденції у галузі безпеки інформації в цілому.

Тематика дослідження кібератак, їх передбачення та прогнозування є досить проблематичною, так як складно передбачити інтенсивність атак і наразі недостатньо сучасних суттєвих методів їх прогнозу. Література з удосконалення графів небажаних втручать для аудиту кібербезпеки, обробки циклів, маніпулювання помилками, автоматичного вибору методів захисту є працями таких вчених як: Савченко В. А., Юдін О. К., Добринін І. С., Копитін Ю. В., Шуклін Г. В., Корченко О. Г., Грищук Р. В., Барабаш О. В., Даник Ю. Г., Смірнов О. А. Розробкою концептуальних моделей інформаційних систем впливу займалися Лужацький В. А. та Дудатьєв А. В. Комплексні захисні системи на основі нейронних мереж будували Тюлюпа С. В., Хлапонін Ю. І., Пархоменко І. І., Козловський В. В., Міщенко А. В. Прогнозуванням, аналітикою, ідентифікацією, розробкою плану протидій та ліквідації еквівалентних кібератак займалися такі зарубіжні науковці як: L. Bilge, У. Han, M. Dell'Amico, Sauerwein, C. Sillaber, M.M. Huber, A. Mussmann, R. Breu, S.K. Lim, A.O. Muis, W. Lu, C.H. Ong, A.P. Moore,

R.J. Ellison, R.C. Linger, A. Sapienza, A. Bessi, S. Damodaran, P. Shakarian, K. Lerman, E. Ferrara, N. Kheir, N. Cuppens-Boulahia, F. Cuppens, H. Debar, A.P. Moore, R.J. Ellison, R.C. Linger та інші. Комп'ютерне програмування з основою на нейронних мережах, що для прогнозу застосовує загальнодоступні на веб-ресурсах зовнішні сигнали, описали у своїх працях Palash Goyal, Ashok Deb, Nazgol Tavabi, Nathan Bartley, Andres Abeliuk, Emilio Ferrara and Kristina Lerman.

Актуальність теми визначається щоденним розвитком кібератак та їх підйомом на фоні пандемії Covid-19. Кіберзлочинність є значущим ризиком сьогодення.

Мета кваліфікаційної магістерської роботи полягає у побудові нейромережевої моделі дослідження стану та прогнозування можливості успішних кібератак на інформаційний простір, а саме інформаційну систему Сумського державного університету.

Об'єктом дослідження є обсяги успішних кібератак на інформаційний простір СумДУ, які приведуть до повної або часткової втрати інформації співробітників чи тимчасового виведення з ладу технічного забезпечення.

Предметом дослідження є методи і моделі аналізу та прогнозування часових рядів.

Для реалізації поставлених цілей сформовано детальні задачі проведення дослідження:

— охарактеризувати підрозділ досліджуваної інформаційної системи;
— розглянути поняття, види та засоби кібератак і способи їх реалізації;
— здійснити аналіз методів і моделей прогнозу кібератак;
— згенерувати вимоги обов'язкові для моделі;
— проаналізувати програмне забезпечення для реалізації моделі;
— побудувати та перевірити на адекватність модель.

Для дослідження поставлених завдань були використані методи аналізу часових рядів та нейромережевого прогнозування.

Інформаційною базою кваліфікаційної магістерської роботи є дані, які були отримані у Центрі телекомунікаційних технологій та комп'ютерного забезпечення (ЦТТКЗ) Сумського державного університету.

Основні положення кваліфікаційної магістерської роботи використані у матеріалах статті Cyberattack Features in Smart Manufacturing and Industry 4.0, яка розглядалася на конференції 5th International Conference on Design, Simulation, Manufacturing: The Innovation Exchange (DSMIE-2022).

1. АНАЛІЗ СТАНУ ЗАХИЩЕНОСТІ ІНФОРМАЦІЙНОГО ПРОСТОРУ СУМСЬКОГО ДЕРЖАВНОГО УНІВЕРСИТЕТУ

1.1 Характеристика підрозділу

Центр телекомунікаційних технологій та комп'ютерного забезпечення (ЦТТКЗ) є структурним підрозділом Сумського державного університету та проводить свою діяльність відповідно до чинної загальнодержавної та внутрішньоуніверситетської нормативної бази.

Мета та завдання центру:

— проектувати, будувати, вводити в експлуатацію, модернізувати, технічно обслуговувати та підтримувати, розвивати єдину інформаційну телекомунікаційну систему;

— технічно обслуговувати, установлювати, адмініструвати та системно супроводжувати сервери мережі;

— створювати та впроваджувати комплексну систему захисту інформації телекомунікаційної мережі СумДУ;

— надавати консультації, проводити навчальні семінари, вебінари в сфері захисту інформації в телекомунікаційних мережах;

— обслуговувати, ремонтувати, налагоджувати периферійне обладнання та устаткування (ксерокси, сканери, принтери і т.п.), комп'ютерної техніки.

До структури ЦТТКЗ входять такі відділи, лабораторія та групи:

1) Відділ технічного захисту інформації, напрямом діяльності якого є здійснення безпосереднього контролю за створенням і рухом нормативно-правової документації стосовно інформаційної безпеки в єдиній телекомунікаційній ІТ-системі університету;

2) Лабораторія мережевих технологій підтримки навчальної та наукової діяльності, напрямом діяльності якої є забезпечення роботи

структурованої кабельної системи технологій сучасності, а також технічне обслуговування, ремонт та модернізація мережі та мережевого обладнання;

3) Відділ технічного забезпечення ІТ-діяльності, напрямами якого є проведення обслуговування, ремонту, налагодження, модернізація та демонтаж комп'ютерної техніки відповідного програмного забезпечення;

4) Відділ системного забезпечення та адміністрування, що проектує комп'ютерні класи, мультимедійні аудиторії, лінії телефонного зв'язку, системи відеоспостереження, системи контролю доступу;

5) Інші групи, які займаються адмініструванням і технічною підтримкою систем телекомунікацій, телефонних зв'язків, охоронної безпеки та єдиної системи часу.

Робота кожного з відділів, зазвичай, базується створеними власними технологічними, програмними й технічними рішеннями, що реалізовують роботу ІТ-системи загалом та дозволяють досягти кінцевої мети забезпечення якісної роботи єдиної інформаційної системи.

Значущим ризиком сьогодення є кібератаки. Кіберзлочинність активізувалася за останні роки і ніяких прогнозів до її спаду немає. Пандемія Covid-19 тільки погіршила ситуацію.

Важливо розуміти можливі проблеми, популярні види атак і звідки саме їх очікувати. Саме цим і займається відділ технічного захисту інформації СумДУ. Кількість атак зростає кожного дня. Та завдяки висококваліфікованим співробітникам та комплексному своєчасному підходу до захисту інформаційного простору, успішність кібератак є незначною та не завдає великої шкоди.

У випадку коли кількість кібератак вдасться спрогнозувати, то відділ технічного захисту інформації зможе прийняти правильне рішення для поліпшення ситуації з небажаними втручаннями.

1.2 Поняття, види та засоби захисту від кібератак

Кібератака – шкідлива свідома спроба кіберзловмисника чи програми проникнути в інформаційну систему іншої людини або організації. Під атакою на інформаційну систему розуміють дії (процеси) або послідовність зв'язаних між собою дій порушника, які приводять до реалізації загроз інформаційним ресурсам, шляхом використання уразливостей цієї інформаційної системи [4]. Як правило, порушуючи роботу мережі жертви, хакер прагне отримати зиск.

Важливо розуміти можливі проблеми, популярні види атак і звідки саме їх очікувати. На сьогодні найпопулярнішими серед атак є фішинг, шкідливі програми, психологічні кібератаки, DDoS атаки, шкідливі програми з вимогою викупу. Статистика компанії Symantec показує, що: на 1 з 36 мобільних пристроях встановлені небезпечні програми; 1 із 13 пошукових запитів призводять до шкідливих програм; 48% від усіх шкідливих email-вкладень відносяться до офісних файлів.

Кількість інцидентів в першому кварталі 2021 року у порівнянні з аналогічним періодом 2020 року збільшилася на 17%, а відносно останнього кварталу 2020 року приріст становить 1,2% [4]. Статистика наведена на рисунку 1. На організації було спрямовано 88% атак і 12% на приватних осіб. Кіберзлочинність розповсюдилася найрізноманітнішими сферами. Частіше зловмисники атакували держустанови, промислові компанії і організації науки та освіти. Основним мотивом атак залишається отримання даних. Головними цілями зловмисників являються персональні і облікові дані, а у випадку організацій додається ще й комерційна таємниця.

Кількість інцидентів у першому кварталі цього року становить 607 атак: січень – 191; лютий – 211; березень – 205. Основним методом атак на організації є використання шкідливого програмного забезпечення – 58% випадків. Соціальна інженерія – 52%, хакінг – 26%, експлуатація веб-вразливості – 15%, підбір облікових даних – 5%, інші – 2%. Держустанови, промисловість, наука і освіта, телекомунікації найбільше потерпіли від методі

використання шкідливого ПЗ; фінансові та медичні установи від соціальної інженерії; ІТ-компанії від хакінгу.

Рисунок 1 – Кількість інцидентів кіберзлочинності в 2020 і 2021 роках

Найбільш популярним шкідливим ПЗ звісно стали програми-вимагачі, а переважаючим способом доставки залишається електронна пошта. Зловмисники використовували її в шести з десяти атак на організації. Приватних осіб як і раніше атакують використовуючи банківські трояни та шпигунські програми, що надають віддалений доступ до пристрою.

Кількість виявлених інцидентів говорить про те, наскільки сильно пандемія Covid-19 вплинула на організації. Кількість кібератак значно збільшилася у порівнянні з 2019 роком. Перехід на дистанційну роботу, який спричинила пандемія, відкрив нові можливості для кіберзлочинності [4]. Умови для цього були сприятливі. Більшість компаній не були готові до пов’язаних з переходом на дистанційну роботу вимогам безпеки і захисту даних.

Аналіз даних Positive Technologies виявив, що наукові та освітні організації є в трійці найпопулярніших установ, які підвернені атакам кіберзлочинців. Ціллю таких атак є отримання персональних даних облікових записів заради продажу їх на «чорному ринку», а саме «дарквебі». Близько 80%

інформації, яка там розповсюджується, це саме дані облікових записів та банківських карт.

Більшість облікових даних продається за ціною до 10$. Викрадені аккаунти від соціальних мереж і інших інтернет-сервісів розповсюджуються партіями від декількох тисяч до мільйонів записів. Ціни за такі обсяги варіюються від десятків доларів до десятків тисяч. Облікові записи для доступу в особисті кабінети онлайн-банку перепродаються поодиноку. Середньою ціною доступу є 22$.

Нещодавно хакери атакували розробника Kaseya і розповсюдили програми-вимагачі на клієнтів даної компанії. Преса назвала кібернапад «витонченим» і запропонувала користувачам не використовувати тимчасово функцію віддаленого керування системою для клієнтів VSA server. В результаті удару постраждало більше тисячі компаній. У числі яких мережа продуктових магазинів Швеції Coop, яка вимушена була закрити ледве не всі вісімсот точок. В магазинах припинили роботу касові системи. Маніпуляції здійснювалися також освітніми закладами, державними установами, кредитними спілками, туристичними та розважальними організаціями. Вимоги, одна з яких $70 млн, було опубліковано в блозі dark web, що зазвичай використовують хакери угрупування REvil.

Бразильська компанія JBS [9] є найбільший виробник м’яса свинини та яловичини у світі. Атака на неї була здійснена програмою-вимагачем, яка зашифрувала інформаційну базу компанії. Хакери запропонували ключ до даних у обмін на викуп. Компанія заздалегідь потурбувалася про резервні копії своїх баз даних і своїми силами відновила роботу систем. Але експерти та консультанти компанії застерігалися того, що хакери нанесуть інший удар. Переговори з нападниками продовжилися. У результаті чого була виплачена винагорода у біткоїнах вартістю $11 млн. У атакі звинуватили угрупування REvil, яке також має інше ім’я – Sodinokibi.

Кібератака хакерів DarkSide [11] зупинила доставку палива найбільшого турбопроводу у США на кілька днів. Небажане втручання призвело до зупинки

турбопроводу, що стало причиною нестачі налива, підвищеного попиту його на південному сході США і, як наслідку, зростанню ціни на бензин. Згідно даних автомобільної асоціації Америки, середньою ціною стала понад $3, а це найвищий показник із жовтня 2014-ого року. Керівництво Colonial Pipeline заплатило хакерам майже $5 млн у криптовалюті. У Colonial Pipeline не бажали розкривати подробиці щодо того, як саме атакували їх систему. Аналітика серверів експертами з безпеки інформаційного простору вказала на нюанси у захищеності, що є можливостями для зловмисники щодо втручання до системи чи перехоплення баз даних. Наприклад, велика кількість об'єктів відеоспостереження приєднана до ІТ-інфраструктури підприємста чи програмне забезпечення відкритого дистанційного керування та обміну файлами, що дає доступ хакерам до внутрішньої мережі компанії, якщо їм відомі логіни.

Європейська компанія Toyota Boshoku Corporation [20] втратила $37 млн через зловмисників і компроментацію шляхом ділового листування. Хакери розіслали співробітникам повідомлення, що містили інструкції щодо оплати замовлень. Фальсифіковані документи були складені настільки ідентично, що працівники дізналися про збитки лише після перерахунку коштів.

Кібератака відбулася і на платформу для обміну криптовалютами Poly Network. Хакери викрали $604 млн. Poly Network вказала адреси рахунків, на які, ймовірно, пішли кошти користувачів. Зокрема, було виведено 2 858 токенів Ethereum на суму близько $267 млн, 6 610 монет binance на суму понад $252 млн і приблизно $85 млн в токенах USDC. Компанія Slowmist, що займається безпекою на основі блокчейнів ідентифікувала електронну пошту, IP-адресу і пристрої зловмисника.

Наймасштабнішою кібератакою в Україні наразі є NotPetya 2017 року. Її наслідків здобули комп'ютерні мережі Кабміну, Пенсійного фонду, Держказначейства, Укрзалізниці, Мінфіна та інших державних закладів і компаній. Шкідливий програмний код розповсюджувався через програмне забезпечення МЕДОК.

Не стане новиною, що організації недостатньо підготовлені до кібератак. Звичні індустріальні системи управління сфери освіти й інші типові комплекси були створені насамперед, щоб зберігати інформацію в своїх межах та стабільно й надійно контролювати її. Тому, у подібних системах майже відсутні передбачені засоби для забезпечення захисту, які б мали можливість не надати стороннім здобути доступ до них.

Більшість кібератак не піддається розголосу через ризик втрати репутації. А отже, підрахувати точні збитки чи кількість інцидентів є неможливим.

Сліди кібератак можливо виявити, якщо проводити наступні рекомендації: 1) контролювати цілісність баз даних, програмного забезпечення та ресурсів в цілому, що мають потребу у забезпеченні безпеки; 2) аналізувати роботу трафіку в мережі, задіяних процесів і користувачів; 3) контролювати фізичні форми нападу на IT-систему, включаючи віддалені бази зберігання даних; 4) проводити аудит діяльності адміністратора щодо опрацювання первинних випадків кібератак.

Головні вразливі місця і недоліки системи захищеності досліджуваного інформаційного простору СумДУ можна розділити на чотири класи:

− вразливість web-додатків (так звані SQL-ін’єкції);

− низька безпека мережі (відкриті протоколи передачі даних, програмне забезпечення віддаленої доступності і надання доступу інтернет користувачам);

− дефекти конфігурації серед серверів (робота зі застарілим програмним забезпеченням, зберігання інформації у відкритому доступу);

− недоліки створення облікових записів і паролів (застосування словарних мало надійних паролів).

Більшість сучасних організацій встановили підвищений рівень захищеності мережі. Та для будь-якої структури найбільш уразливою ланкою є персонал і Сумський державний університет не є виключенням. У процесі аналізу вразливості щодо інформаційного захисту виявлено: 50% співробітників переходили за посиланням, зазначеним у фішинговому листі;

25% – вводили власні дані авторизації в недійсну форму; ще 25% – запускали на своєму ПК шкідливе вкладення. Та оцінюючи обізнаність про ризики серед персоналу лише 8% переходило за фішинговим посиланням, 2% запускали вкладені файли та менше 1% вводили свої облікові дані за фейковими формами аутентифікації.

Рівень обізнаності у питання інформаційної безпеки серед співробітників Сумського державного університету є вищим відносно деяких інших галузь. Та для кіберзлочинників досить, щоб хтось із користувачів виконав неправильну дію, і кібератака наддасть доступ відразу до внутрішньокорпоративної мережі вцілому.

Будь-яку кібератаку можливо виявити аналізом вручну або за допомогою вбудованих в операційну систему засобів. Це дозволяє значно знизити витрати на розробку комплексу аналізу кібератак. Але недоліками такого підходу є довгий час на його виконання. Та головне, що такий метод аналізу не завжди надає можливість для своєчасного запобігання небажаним втручанням. А отже, правильним рішенням є залучення додаткових засобів прогнозування можливих кібератак.

Ефективність системи виявлення кібератак значно підвищується при застосовуванні інноваційних методів аналізу отриманої інформації (статистичного підходу; експертних систем; нейронних мереж) [45]. Зазначені методи мають як переваги так і недоліки, тому системи зазвичай використовують комплексний підхід. Звісно, для ефективності цих методів доцільно використовувати їх у сукупності та компетентними IT-службами безпеки.

1.3 Постановка задачі дослідження

Для побудови економіко-математичної моделі прогнозування можливості успішних кібератак на досліджуваний інформаційний простір, а саме

інформаційну систему Сумського державного університету, виділено наступні задачі:

— провести аналіз стану захищеності інформаційного простору СумДУ;

— проаналізувати та порівняти методології прогнозування;

— описати моделі прогнозування;

— сформувати вимоги до моделі;

— провести аналіз вхідних даних програмним пакетом STATISTICA;

— перевірити вхідні дані на однорідність методом Ірвіна;

— проаналізувати та порівняти програмне забезпечення для побудови прогнозу;

— побудувати нейронні мережі для прогнозування;

— перевірити побудовані моделі на адекватність;

— провести прогнозування успішних кібератак на інформаційний простір Сумського державного університету, які приведуть до повної або часткової втрати інформації співробітників чи тимчасового виведення з ладу технічного забезпечення.

2. РОЗРОБКА МОДЕЛІ ПРОГНОЗУВАННЯ МОЖЛИВИХ ОБСЯГІВ УСПІШНИХ КІБЕРАТАК НА ІНФОРМАЦІЙНИЙ ПРОСТІР СУМДУ

2.1 Методологія та моделі прогнозування

Методи Data Mining - це методи моделювання та прогнозування даних, за допомогою яких проводиться аналітична робота щодо виявлення алгоритмів у великих масивах інформації. Ці методи дозволяють людям, які не мають спеціальної математичної підготовки, використовувати інструментарій Data Mining та приймати на їх основі ефективні управлінські рішення.

До методів та алгоритмів Data Mining відносяться:

— штучні нейронні мережі;

— дерева рішень, символьні правила;

— методи найближчого сусіда та k-найближчого сусіда;

— метод опорних векторів;

— байєсовські мережі;

— лінійна регресія;

— кореляційно-регресійний аналіз;

— ієрархічні методи кластерного аналізу;

— неієрархічні методи кластерного аналізу, у тому числі алгоритми k-середніх та k-медіани;

— еволюційне програмування та генетичні алгоритми;

— метод обмеженого перебору;

— еволюційне програмування та генетичні алгоритми;

— різноманітні методи візуалізації даних та безліч інших методів.

Кожен із перелічених методів застосовується для вибірки конкретних даних, побудови прогнозуючої моделі чи очищення БД від помилкових відомостей.

Класифікація методів Data Mining на категорії:

— Методи, пов'язані з безпосереднім використанням (збереженням) даних. Дані під час обробки деталізуються при побудові прогностичної моделі або під час аналізу винятків. Однак такі методи є малоефективними при роботі з великими масивами даних.

— Дистиляція шаблонів – формування та застосування закономірностей, що мають упорядкований вигляд, тобто вилучення інформації з початкових даних з її перетворенням на певну систематизовану конструкцію. Задіяння цих методів забезпечує ефективне застосування отриманих під час вільного пошуку результатів і перетворення цих відомостей на зрозумілі для користувачів закономірності.

Різноманітність методів Data Mining характеризується певними вагомими якостями, які можуть стати вирішальними у виборі методу для проведення аналізу даних. Порівняння деяких з них наведено у таблиці 2.1. Порівняння методів Data Mining проведено за такими якостями: точність, масштабованість, інтерпретованість, перевірюваність, трудомісткість, гнучкість, швидкість та популярність .

Таблиця 2.1 – Порівняльна таблиця методів Data Mining

Алгоритм	Лінійна регресія	Нейронні мережі	Методи візуалізації	Дерева рішень	К-найближчого сусіда
точність	+–	+	+	–	–
масштабованість	+	+	–	+	–
інтерпретованість	+–	+	+	+	+–
перевірюваність	+	+–	+	+–	+–
трудомісткість	+–	+	+	+	+–
гнучкість	+–	+	–	+	–
швидкість	+	+	–	–	+
популярність	–	–	+	+	–

Кожен із методів має свої сильні та слабкі сторони. Але жоден метод неспроможний забезпечити вирішення всього спектра завдань Data Mining. У дослідженні використано метод нейронних мереж.

У світі безліч речей, які є цікавими для вивчення і дослідження, а саме виявлення їх розвитку і змін за певний проміжок часу. Це може бути курс валюти, зміна тиску чи температури, ріст акцій чи цін та інше. Усі вони непостійні і з часом змінюються. А зміна однієї характеристики в часі називається часовим рядом.

Часовий ряд коротко можна записати у наступному вигляді:

$$y_t,\ t = 1,2,\ \dots\ n, \qquad (2.1)$$

де t –проміжки часу, що обов'язково є рівновіддаленими (хвилина, година, доба, місяць, рік та ін.).

Зміни для певного показника часовий ряд досить вірно відображає за умови, коли дані для нього мають достатню кількість спостережень, відібрані за однакові проміжки часу, представлені в одних одиницях виміру.

У методі часового ряду однією із цілей є визначення характеру і напряму змін даних у майбутньому, це дозволяє отримати нові прогнозовані значення. О.В. Козьменко і О.В. Кузьменко пропонують визначення прогнозу як «науково обгрунтованого судження стосовно можливих станів об'єкта в майбутньому, альтернативні шляхи і терміни їх здійснення».

Різких змін значень у часовому ряді не повинно бути – такі значення негативно впливають на прогноз, а отже, їх необхідно знайти і ліквідувати. У такому випадку доцільно використати метод Ірвіна. Він здатен, проаналізувавши часовий ряд, відображати аномальні, тобто нетипові, значення.

В основі цього метода лежить порівняння сусідніх значень та розрахунок допоміжного обов'язкового значення λ, що дорівнює:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\hat{\sigma_y}};\ t = 2,3,\dots,n; \qquad (2.2)$$

де $\sigma^{\wedge}$ − середньоквадратичне відхилення вибірки

$$\sigma_y^{\wedge} = \sqrt{\frac{\sum_{t=1}^{n}(y_t - y^{-})^2}{n}}, y^{\wedge} = \frac{\sum_{t=1}^{n} y_t}{n}. \qquad (2.3)$$

Розраховані значення λ_2, λ порівнюють із табличним $\lambda\alpha$. Якщо вони не перевищують табличні, то відповідні показники y_t, можна вважати нормальними. Табличні значення для значущості α = 0,05 (помилка 5 %) наведено в таблиці 2.2.

Таблиця 2.2 – Значення табличного $\lambda\alpha$

n	*2*	*3*	*10*	*20*	*30*	*50*	*100*
λ_α	*2,8*	*2,3*	*1,6*	*1,3*	*1,2*	*1,1*	*1,0*

Модифікований метод Ірвіна відрізняється тим, що $\sigma^{\wedge}$ обчислюється лише за трьома показниками спостереження, а не вибіркою вцілому. Особливістю є можливість маніпулювати не всією вибіркою, а лише у місцях де є підозра у аномальності значень.

Середнє відхилення між сусідніми показниками обчислюється наступним чином:

$$\bar{y}_t = \frac{y_{t-1} + y_{t+1}}{2} \qquad (2.4)$$

$$\widehat{\sigma_y} = \sqrt{\frac{\sum_{t=1}^{n}(y_{t-1} - \bar{y})^2 + (y_{t+1} - \bar{y})^2}{2}} \qquad (2.5)$$

$$F_{\text{розрах}} = \begin{Bmatrix} \hat{\sigma}_2^2 / \hat{\sigma}_1^2, \text{якщо } \hat{\sigma}_2^2 > \hat{\sigma}_1^2 \\ \hat{\sigma}_1^2 / \hat{\sigma}_2^2, \text{якщо } \hat{\sigma}_1^2 > \hat{\sigma}_2^2 \end{Bmatrix} \qquad (2.6)$$

Значення, які виявилися аномальними замінюють на середнє між сусідніми, коли причиною аномальності є помилки першого роду [40].

Усе частіше чути про системи штучного інтелекту, які базуються на використанні апарата штучних нейронних мереж. За допомогою них вирішується велика кількість проблем, таких як побудова моделі об’єктів при великій кількості шуму, недостатній кількості інформації, кластеризації [54].

Ідея нейронних мереж з’явилася і грунтується на спробі змоделювати поведінку об’єкта, що відчуває вплив зовнішнього середовища на собі і навчається на отриманому власному досвіді. Сьогодні штучні нейронні мережі знайшли своє місце і у прогнозуванні.

Існує багато структур нейронних мереж, які відрізняються кількістю і розміщенням нейронів, зв’язків. Найпопулярнішою є багатошаровий персептрон (MLP). Він представляє повнозв’язну модель, що не має зворотніх зв’язків. Кількість шарів моделі і нейронів залежить від задачі, яку потрібно виконати. Схема структурування у персептроні випадкових шарів і нейронів зображена на рисунку 2.2.

В основі нейромережі є вимушений нейрон, що копіює наближенні якості біологічного. На вхід штучного нейрону надходить велика кількість сигналів, які також є виходами, але для інших нейронів. Потужність мережі визначається саме кількістю нейронів [44].

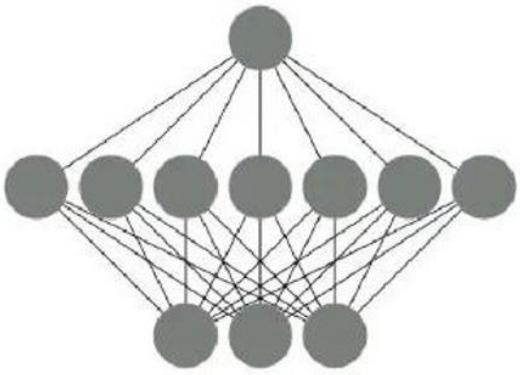

Рисунок 2.1 – Схема структурування у персептроні MLP

Скупчення нейронів поєднують у шари, на основі яких створюються одношарові або багатошарові нейронні мережі. Багатошарові нейромережі дозволяють реалізувати більше можливостей, але вони можуть привести до

збільшення потужності лише коли активаційна функція, що поєднує шари не є лінійною [55].

Оцінка ризику успішних кібератак на інформаційний простір Сумського державного університету, які приведуть до повної або часткової втрати інформації співробітників чи тимчасового виведення з ладу технічного забезпечення пропонується до реалізації шляхом створення нейронної мережі. Економіко-математичну модель пропонується представити у вигляді мережі радіальних базисних функціях та багатошарового персептрону. Вона набуває наступного вигляду:

$$f(x) = F\left(\sum_{k_N} w_{k_N n_N N} \dots \sum_{k_2} w_{k_2 n_2 2}\, F\left(\sum_{k_1} w_{k_1 n_1 1} x_{k_1 n_1 1} - \theta_{n_1 1}\right) - \theta_{n_2 2} \dots - \theta_{n_N N}\right) \qquad (2.7)$$

де $F\left(\sum_{k_1} w_{k_1 n_1 1} x_{k_1 n_1 1} - \theta_{n_1 1}\right)$ – перший шар;

$\sum_{k_2} w_{k_2 n_2 2}\, F\left(\sum_{k_1} w_{k_1 n_1 1} x_{k_1 n_1 1} - \theta_{n_1 1}\right) - \theta_{n_2 2}$ – другий шар;

$F\left(\sum_{k_N} w_{k_N n_N N} \dots \sum_{k_2} w_{k_2 n_2 2}\, F\left(\sum_{k_1} w_{k_1 n_1 1} x_{k_1 n_1 1} - \theta_{n_1 1}\right) - \theta_{n_2 2} \dots - \theta_{n_N N}\right)$ –*N*-ий шар;

k – індекс входу;

n – індекс нейрона;

$x_{k_1 n_1 1}$ – вхідний сигнал k нейрону n у першому шарі;

$w_{k_N n_N N}$ – ваговий коефіцієнт вхідного сигналу k нейрону n у *N*-у шарі;

$\theta_{n_N N}$ – пороговий рівень нейрону n у *N*-у шарі.

Економіко-математична модель нейромережі радіальних базисних функцій набуває вигляду:

$$f(x) = \sum_{k=1}^{N} v_k \varphi(\|r - r_k\|) \qquad (2.8)$$

де - v_k – ваговий коефіцієнт вхідного сигналу k;

r_k – центри радіальних базисних функцій.

Для побудови нейронних мереж можна застосовувати алгоритм Бройдена-Флетчера-Гольдфарба-Шанно (BFGS). Даний алгоритм підходить для побудови мереж типу багатошарового персептрону. BFGS відноситься до

квазіньютонівських методів. Основна ідея методу це пошук екстремумів локального типу для функцій нелінійного типу без обмежень. Метод виконує процедури ітеративної числової оптимізації.

Для реалізація алгоритму BFGS необхідно виконати наступні кроки:

1) Розрахувати коефіцієнти малих величин та значення зворотнього гессіана (початкові значення наближення). Розмірність матриці визначається довжиною вектора градієнта; T – матриці розміру lxl, де l – довжина вектор градієнта g;

2) Розрахувати параметри градієнта G;

3) Розрахувати вагові коефіцієнти (їх кореляції) $\Delta W=G\cdot m$, $W_{y+1}=W_y-\Delta W$, де m величина швидкості навчання;

4) Розрахувати нове значення градієнту $G = G(W)$ та його зміну в порівнянні з попереднім етапом ітерації $\Delta G = G - G_p$;

5) Розрахувати гессіана та зміну параметрів проводити за наступними формулами:

$$T_{y+1} = T_y - \frac{T_y \cdot s \cdot s^t \cdot T_y}{s^T \cdot T_y \cdot s} + \frac{r \cdot r^t}{s^t \cdot s}, \quad (2.9)$$

$$r = \Delta G_y = G_y - G_{y-1}$$

$$s = \Delta W_y = W_y - W_{y-1}$$

$$\Delta W = W \cdot G$$

$$W = W - \Delta W$$

6) Визначити значення похибки, якщо значення вище допустимого, алгоритм повторяється, починаючи з 4-ого етапу, інакше, алгоритм зупиняється.

Представимо проектовану форму нейронної мережі в узагальненій схемі економіко-математичного моделювання (рис. 2.2). Вхідними даними є часовий ряд і індикатор часу. На виході маємо прогнозне значення рівня кібератак. Некерованими змінними виступила кількість прихованих шарів, а керованими – ваговий коефіцієнт k-ого вхідного сигналу n-ого нейрону у шарі N; пороговий рівень n-ого нейрону у шарі N.

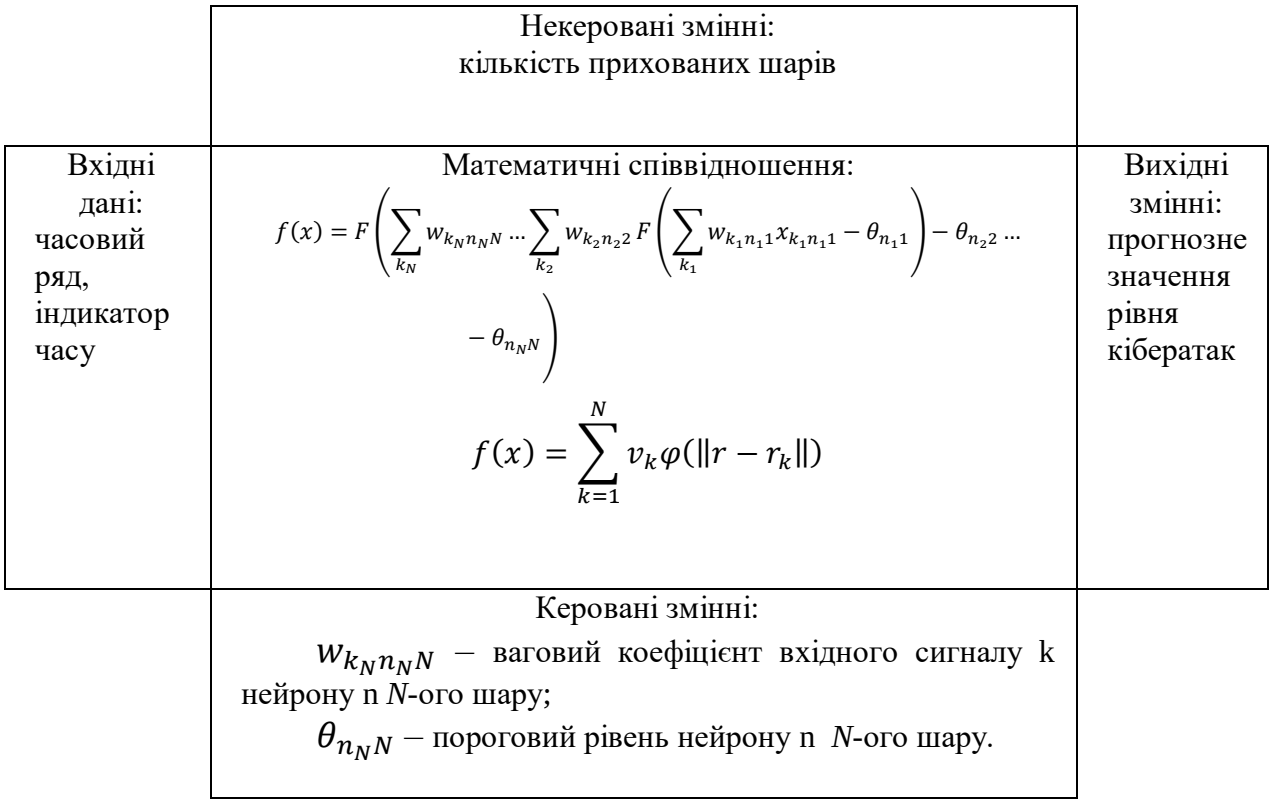

Рисунок 2.2 – Узагальнена схема економіко-математичного моделювання

Перша та друга модель нейронних мереж архітектури RBF 1-14-1, що має загальних шарів – 1, прихованих шарів – 14, у загальному математичному вигляді представлена наступним чином:

$$ab_1^{(2)} = f(d_{11}^{(1)} c_1 + a_1^{(1)}) \quad (2.10)$$

$$ab_2^{(2)} = f(d_{21}^{(1)} c_1 + a_2^{(1)})$$

$$ab_3^{(2)} = f(d_{31}^{(1)} c_1 + a_3^{(1)})$$

$$ab_4^{(2)} = f(d_{41}^{(1)} c_1 + a_4^{(1)})$$

$$ab_5^{(2)} = f(d_{51}^{(1)} c_1 + a_5^{(1)})$$

$$ab_6^{(2)} = f(d_{61}^{(1)} c_1 + a_6^{(1)})$$

$$ab_7^{(2)} = f(d_{71}^{(1)} c_1 + a_7^{(1)})$$

$$ab_8^{(2)} = f(d_{81}^{(1)} c_1 + a_8^{(1)})$$

$$ab_9^{(2)} = f(d_{91}^{(1)} c_1 + a_9^{(1)})$$

$$ab_{10}^{(2)} = f(d_{101}^{(1)}c_1 + a_{10}^{(1)})$$
$$ab_{11}^{(2)} = f(d_{111}^{(1)}c_1 + a_{11}^{(1)})$$
$$ab_{12}^{(2)} = f(d_{121}^{(1)}c_1 + a_{12}^{(1)})$$
$$ab_{13}^{(2)} = f(d_{131}^{(1)}c_1 + a_{13}^{(1)})$$
$$ab_{14}^{(2)} = f(d_{141}^{(1)}c_1 + a_{14}^{(1)})$$

$$S = k^{(3)} = f(d_1^{(2)}ab_1^{(2)} + d_2^{(2)}ab_2^{(2)} + d_3^{(2)}ab_3^{(2)} + d_4^{(2)}ab_4^{(2)} + d_5^{(2)}ab_5^{(2)} + d_6^{(2)}ab_6^{(2)} + d_7^{(2)}ab_7^{(2)} + d_8^{(2)}ab_8^{(2)} + d_9^{(2)}ab_9^{(2)} + d_{10}^{(2)}ab_{10}^{(2)} + d_{11}^{(2)}ab_{11}^{(2)} + d_{12}^{(2)}ab_{12}^{(2)} + d_{13}^{(2)}ab_{13}^{(2)} + d_{14}^{(2)}ab_{14}^{(2)} + a^{(2)})$$

де f – функція активації прихованих нейронів;

$ab_n^{(2)}$ – вихід прихованого нейрону у розрізі 2-ого шару;

$ab_n^{(3)}$ – вихід прихованого нейрону у розрізі 3-ого шару.

Функцією активації у даному випадку є лінійна:

$$QUT = net$$

де *QUT* – виходи прихованих нейронів мережі в розрізі шару;

net – обсяг вхідних сигналів.

Третя модель архітектури MLP 1-2-1 (загальна кількість шарів – 1, кількість прихованих шарів – 2) у загальному математичному вигляді представлена наступним чином:

$$ab_1^{(2)} = f(d_{11}^{(1)}c_1 + a_1^{(1)}) \quad (2.11)$$
$$ab_2^{(2)} = f(d_{21}^{(1)}c_1 + a_2^{(1)})$$
$$S = k^{(3)} = f(d_1^{(2)}ab_1^{(2)} + d_2^{(2)}ab_2^{(2)} + a^{(2)})$$

Четверта модель нейронних мереж архітектури RBF 1-16-1, що має загальних шарів – 1, прихованих шарів – 16, у загальному математичному вигляді представлена наступним чином:

$$ab_1^{(2)} = f(d_{11}^{(1)}c_1 + a_1^{(1)}) \quad (2.12)$$
$$ab_2^{(2)} = f(d_{21}^{(1)}c_1 + a_2^{(1)})$$

$$ab_3^{(2)} = f(d_{31}^{(1)}c_1 + a_3^{(1)})$$
$$ab_4^{(2)} = f(d_{41}^{(1)}c_1 + a_4^{(1)})$$
$$ab_5^{(2)} = f(d_{51}^{(1)}c_1 + a_5^{(1)})$$
$$ab_6^{(2)} = f(d_{61}^{(1)}c_1 + a_6^{(1)})$$
$$ab_7^{(2)} = f(d_{71}^{(1)}c_1 + a_7^{(1)})$$
$$ab_8^{(2)} = f(d_{81}^{(1)}c_1 + a_8^{(1)})$$
$$ab_9^{(2)} = f(d_{91}^{(1)}c_1 + a_9^{(1)})$$
$$ab_{10}^{(2)} = f(d_{101}^{(1)}c_1 + a_{10}^{(1)})$$
$$ab_{11}^{(2)} = f(d_{111}^{(1)}c_1 + a_{11}^{(1)})$$
$$ab_{12}^{(2)} = f(d_{121}^{(1)}c_1 + a_{12}^{(1)})$$
$$ab_{13}^{(2)} = f(d_{131}^{(1)}c_1 + a_{13}^{(1)})$$
$$ab_{14}^{(2)} = f(d_{141}^{(1)}c_1 + a_{14}^{(1)})$$
$$ab_{15}^{(2)} = f(d_{151}^{(1)}c_1 + a_{15}^{(1)})$$
$$ab_{16}^{(2)} = f(d_{161}^{(1)}c_1 + a_{16}^{(1)})$$

$$S = k^{(3)} = f(d_1^{(2)}ab_1^{(2)} + d_2^{(2)}ab_2^{(2)} + d_3^{(2)}ab_3^{(2)} + d_4^{(2)}ab_4^{(2)} + d_5^{(2)}ab_5^{(2)} + d_6^{(2)}ab_6^{(2)} + d_7^{(2)}ab_7^{(2)} + d_8^{(2)}ab_8^{(2)} + d_9^{(2)}ab_9^{(2)} + d_{10}^{(2)}ab_{10}^{(2)} + d_{11}^{(2)}ab_{11}^{(2)} + d_{12}^{(2)}ab_{12}^{(2)} + d_{13}^{(2)}ab_{13}^{(2)} + d_{14}^{(2)}ab_{14}^{(2)} + d_{15}^{(2)}ab_{15}^{(2)} + d_{16}^{(2)}ab_{16}^{(2)} + a^{(2)})$$

П’ята модель нейронних мереж архітектури RBF 1-16-1, що має загальних шарів – 1, прихованих шарів – 16, у загальному математичному вигляді представлена наступним чином:

$$ab_1^{(2)} = f(d_{11}^{(1)}c_1 + a_1^{(1)}) \quad (2.13)$$
$$ab_2^{(2)} = f(d_{21}^{(1)}c_1 + a_2^{(1)})$$
$$ab_3^{(2)} = f(d_{31}^{(1)}c_1 + a_3^{(1)})$$
$$ab_4^{(2)} = f(d_{41}^{(1)}c_1 + a_4^{(1)})$$
$$ab_5^{(2)} = f(d_{51}^{(1)}c_1 + a_5^{(1)})$$

$$ab_6^{(2)} = f(d_{61}^{(1)}c_1 + a_6^{(1)})$$
$$ab_7^{(2)} = f(d_{71}^{(1)}c_1 + a_7^{(1)})$$
$$ab_8^{(2)} = f(d_{81}^{(1)}c_1 + a_8^{(1)})$$
$$ab_9^{(2)} = f(d_{91}^{(1)}c_1 + a_9^{(1)})$$
$$ab_{10}^{(2)} = f(d_{101}^{(1)}c_1 + a_{10}^{(1)})$$
$$ab_{11}^{(2)} = f(d_{111}^{(1)}c_1 + a_{11}^{(1)})$$
$$ab_{12}^{(2)} = f(d_{121}^{(1)}c_1 + a_{12}^{(1)})$$
$$ab_{13}^{(2)} = f(d_{131}^{(1)}c_1 + a_{13}^{(1)})$$
$$ab_{14}^{(2)} = f(d_{141}^{(1)}c_1 + a_{14}^{(1)})$$
$$ab_{15}^{(2)} = f(d_{151}^{(1)}c_1 + a_{15}^{(1)})$$
$$ab_{16}^{(2)} = f(d_{161}^{(1)}c_1 + a_{16}^{(1)})$$

$$S = k^{(3)} = f(d_1^{(2)}ab_1^{(2)} + d_2^{(2)}ab_2^{(2)} + d_3^{(2)}ab_3^{(2)} + d_4^{(2)}ab_4^{(2)} + d_5^{(2)}ab_5^{(2)} + d_6^{(2)}ab_6^{(2)} + d_7^{(2)}ab_7^{(2)} + d_8^{(2)}ab_8^{(2)} + d_9^{(2)}ab_9^{(2)} + d_{10}^{(2)}ab_{10}^{(2)} + d_{11}^{(2)}ab_{11}^{(2)} + d_{12}^{(2)}ab_{12}^{(2)} + d_{13}^{(2)}ab_{13}^{(2)} + d_{14}^{(2)}ab_{14}^{(2)} + d_{15}^{(2)}ab_{15}^{(2)} + d_{16}^{(2)}ab_{16}^{(2)} + a^{(2)})$$

2.2 Формування вимог до моделі

Для побудови моделі прогнозу можливості успішних кібератак на інформаційний простір СумДУ, які приведуть до повної або часткової втрати інформації співробітників чи тимчасового виведення з ладу технічного забезпечення, слід визначитися із вимогами до неї. Насамперед модель повинна бути адекватною, точною та простою.

Адекватність – це здатність моделі відповідати вимогам ситуації та очікуванням. До початку побудови моделі оцінити адекватність важко, але є можливість орієнтуватися на раніше отриманий досвід у побудові моделей схожої тематики [12].

Ступінь вірності прогнозування відображає точність. Вона оцінюється за первинними даним вибірки. Суттєвими факторами, що дають неправильний прогноз є невірно підібрана модель і малий обсяг вхідних даних.

Простота виражається обсягом затрачених ресурсів на реалізацію моделі прогнозування. Мається на увазі час і оперативна пам’ять. Чим менше їх затрачено, тим ліпшою є модель, тобто економічнішою.

До моделі прогнозування можливості успішних кібератак на інформаційний простір СумДУ сформовані такі вимоги:

− створена модель мають бути практично застосоваі;

− модель повинна швидко виконувати поставлені задачі та оперативно формувати результати;

− модель має відповідати показнику адекватності;

− модель має бути стійкою ;

− інформаційна база для моделі має бути достовірною та повною;

Таким чином, модель прогнозування успішних кібератак має відповідати вищенаведеним вимогам для того, щоб прогноз був достовірним та корисним при обробці співробітником із кібербезпеки.

2.3 Опис вхідних даних

Вхідними даними для побудови моделі є вибірка успішних кібератак зібрана у Центрі телекомунікаційних технологій та комп’ютерного забезпечення Сумського державного університету (рис. 2.3).

Дата	Кількість	Дата	Кількість	Дата	Кількість
01.01.2016	3	01.01.2018	5	01.01.2020	9
01.02.2016	3	01.02.2018	2	01.02.2020	9
01.03.2016	4	01.03.2018	2	01.03.2020	9
01.04.2016	1	01.04.2018	2	01.04.2020	4
01.05.2016	0	01.05.2018	2	01.05.2020	9
01.06.2016	4	01.06.2018	2	01.06.2020	4
01.07.2016	5	01.07.2018	2	01.07.2020	5
01.08.2016	2	01.08.2018	4	01.08.2020	9
01.09.2016	5	01.09.2018	2	01.09.2020	8
01.10.2016	4	01.10.2018	3	01.10.2020	5
01.11.2016	2	01.11.2018	4	01.11.2020	9
01.12.2016	6	01.12.2018	4	01.12.2020	5
01.01.2017	2	01.01.2019	3	01.01.2021	5
01.02.2017	5	01.02.2019	4	01.02.2021	4
01.03.2017	1	01.03.2019	4	01.03.2021	6
01.04.2017	2	01.04.2019	6	01.04.2021	5
01.05.2017	3	01.05.2019	3	01.05.2021	4
01.06.2017	2	01.06.2019	1	01.06.2021	4
01.07.2017	4	01.07.2019	4	01.07.2021	3
01.08.2017	3	01.08.2019	1	01.08.2021	5
01.09.2017	3	01.09.2019	2	01.09.2021	5
01.10.2017	2	01.10.2019	7	01.10.2021	6
01.11.2017	6	01.11.2019	6		
01.12.2017	2	01.12.2019	7		

Рисунок 2.3 – Вхідні дані

Для кращого сприйняття даних та візуальній оцінці скористаємося функціоналом STATISTICA та побудуємо графік вхідних даних (рис. 2.4).

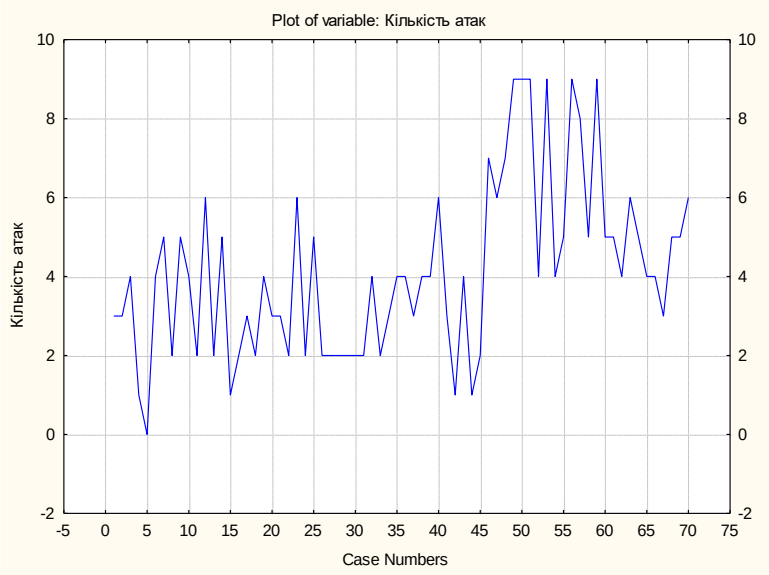

Рисунок 2.4 – Графік вхідних даних

Середнє значення даної вибірки – 4,1. Медіана – 4. Модальне значення невизначене. Коефіцієнт варіації дорівнює 53,9. Мінімальним значення рівне нулю, максимальне – 9. Дані характеристики отримані програмним пакетом STATISTICA, основні статистики і табиці (рис. 2.5).

Variable	Descriptive Statistics (Spreadsheet1.sta)									
	Mean	Median	Mode	Frequency of Mode	Minimum	Maximum	Std.Dev.	Coef.Var.	Skewness	Kurtosis
Кількість атак	4,114286	4,000000	Multiple	15	0,00	9,000000	2,216819	53,88101	0,697928	0,036378

Рисунок 2.5 – Аналіз вхідних даних

Вхідні дані було перевірено модифікованим методом Ірвіна, результати перевірки наведені на рис 2.6. За результатами дослідження λ розраховане є меншим від табличного значення (табл. 2.2), тому досліджуваний ряд є однорідним, а значить використовувати даний часовий ряд для наступного аналізу можливо.

Модифікований метод Ірвіна					
Year	yi	yt	(yt-1-yt)^2	(yt+1-yt)^2	λ
01.01.2016	3	-	-	-	-
01.02.2016	3	3,5	0,250	0,25	0
01.03.2016	4	2	1,000	1	0,10706
01.04.2016	1	2	4,000	4	0,32117
01.05.2016	0	2,5	2,250	2,25	0,10706
01.06.2016	4	2,5	6,250	6,25	0,42823
01.07.2016	5	3	1,000	1	0,10706
01.08.2016	2	5	0,000	0	0,32117
01.09.2016	5	3	1,000	1	0,32117
01.10.2016	4	3,5	2,250	2,25	0,10706
01.11.2016	2	5	1,000	1	0,21412
01.12.2016	6	2	0,000	0	0,42823
01.01.2017	2	5,5	0,250	0,25	0,42823
01.02.2017	5	1,5	0,250	0,25	0,32117
01.03.2017	1	3,5	2,250	2,25	0,42823
01.04.2017	2	2	1,000	1	0,10706
01.05.2017	3	2	0,000	0	0,10706
01.06.2017	2	3,5	0,250	0,25	0,10706
01.07.2017	4	2,5	0,250	0,25	0,21412
01.08.2017	3	3,5	0,250	0,25	0,10706
01.09.2017	3	2,5	0,250	0,25	0
01.10.2017	2	4,5	2,250	2,25	0,10706
01.11.2017	6	2	0,000	0	0,42823
01.12.2017	2	5,5	0,250	0,25	0,42823
01.01.2018	5	2	0,000	0	0,32117

Рисунок 2.6 – Фрагмент перевірка на однорідність

3. МОДЕЛЮВАННЯ МОЖЛИВИХ ОБСЯГІВ УСПІШНИХ КІБЕРАТАК НА ІНФОРМАЦІЙНИЙ ПРОСТІР СУМДУ

3.1 Програмне забезпечення для прогнозу

Інформаційні технології використовуються у найрізноманітніших галузях, у тому числі і у сфері економіко-математичного моделювання. Отже, засобів для реалізації прогнозування багато і найкращими для дослідження часових рядів є STATISTICA, MATLAB, пакет SAS (Statistical Analysis System), EViews [7].

STATISTICA – це комплексний аналітичний інструмент, призначений для побудови точних прогнозів у будь-яких областях, використовуючи різні методи прогнозування. Вона включає в себе модуль для аналізу часових рядів, який дозволяє побудувати прогноз без використання допоміжних факторів, тобто прогнозування поведінки ряду базується на основі його власної історії. Використовуються найбільш ефективні та популярні методи для аналізу часових рядів: експоненційне згладжування, модель авторегресії і ковзного середнього, сезонна декомпозиція, спектральний аналіз Фур'є.

STATISTICA також має нейромережу, що місить в собі збірку потужних вбудованих інтелектуальних можливостей, які дозволяють вирішити реальні задачі, навіть користувачу, що ще не користувався нейронними мережами. В той самий час, досвідчений користувач зможе повністю керувати майже всіма аспектами нейромережевих структур і навчання.

При дослідженні часових рядів часто найефективнішими виявляються графічні і описові методи аналізу. Модуль також містить повний набір засобів для проведення будь-яких видозмін часового ряду, таких як взяття різниць різних порядків (вивчення мінливості ряду), згладжування ряду (виявлення тенденцій в поведінці ряду), виділення тренду (виділення детермінованої

систематичної складової ряду), обрахунок автокореляційних і кроскореляційних функцій, а також побудова їх графіків (корелограм) .

MATLAB – це пакет прикладних програм для вирішення задач технічних розрахунків, в тому числі і з часовими рядами.

Інструменти MATLAB дозволяють отримувати доступ, візуалізувати і аналізувати історичні і поточні дані часових рядів, для того, щоб виявити і проаналізувати залежність. За допомогою пакету MATLAB можливо:

− отримувати доступ до даних із різних джерел (файли, електронні таблиці, бази даних);

− зберігати дані в об’єктах часових рядів, для того, щоб полегшити керування даними, обробки пропущених даних;

− виконувати технічний аналіз із різними фільтрами, стохастиками та індексами;

− створювати користувацькі процедури аналізу, візуалізації і анімації для демонстрації процесу аналізу;

− налаштування середовища аналізу із розширеною функціональністю для обробки сигналів, статистики чи економетрики [16].

MATLAB дає можливість оцінити спектри часових рядів, які описують варіації часових рядів, використовуючи циклічні компоненти на різних частотах. Також є можливість проаналізувати авторегресію (AR), авторегресію з ковзним середнім (ARIMA) .

Пакет SAS (Statistical Analysis System) – професійний статистичний пакет від компанії SAS Institute Inc. Основний додаток SAS – Business Intelligence. Дана система є наштовуваною та призначена для фінансово менеджменту, керування ризиками, маркетингу, прогнозування. Всі рішення базуються на загальній технологічній платформі – SAS Enterprise Intelligence Platform, яка забезпечує базові необхідні всім додаткам функціональні можливості:

− ETL/ELT – процес добутку даних із різних джерел із наступною обробкою та очищенням;

− зберігання даних у спеціалізованому аналітичному банку даних;

− поглиблена аналітика – середовище для проведення поглибленого аналізу даних (data mining), описового і прогнозного моделювання, прогнозування часових рядів, оптимізації.

SAS включає в себе точні методи та потужні інструменти статичного моделювання для невеликих наборів даних і задач із великими даними, а також сучасні методи аналізу даних, що містять відсутні значення. Пакет SAS дозволяє:

− будувати дерева класифікації та регресії;

− відокремлювати дані на навчальні, контрольні і тестові ролі;

− використовувати сучасні методи підбору моделей, такі як еластична сітка і група LASSO;

Інформація подається сотнями вбудованих графіків і діаграм. Через це результати аналізу легко зрозуміти. Так як метадані зберігаються в централізованому сховищі, то є можливість включати моделі SAS/STAT в інші рішення SAS.

EViews – це статичний пакет для Windows, що використовується в основному для орієнтованого на часові ряди економічного аналізу. Засіб економетрики, статистики та прогнозування, що поєднує потужні аналітичні інструменти в гнучкому і зручному інтерфейсі. В порівнянні з конкурентами EViews немає модульної системи, але є доступним вікно робочого файлу, де є можливість зберігати ряд об’єктів. В додатку EViews для побудови моделі ARMA використовується розширений тест Дікі-Фулера, перевірка стаціонарності виконується автоматично після взяття різниць першого чи другого порядку. EViews включає в себе технологію електронних таблиць і реляційних баз даних із традиційними задачами, що використовуються в статистичному програмному засобі і використовує графічний інтерфейс Windows [7].

Для точного уявлення переваг і недоліків було створено порівняльну таблицю 3.1.

Таблиця 3.1 – Переваги і недоліки програм для побудови моделі

ПЗ	Переваги	Недоліки
STATISTICA	- Можливість паралельноїроботи в різних модулях; - велика кількістьдовідкової літератури; - зрозумілий інтерфейс; - швидкодія; - легкий імпорт/експорт даних в електронні ітекстові процесори.	- Складно опанувати для не фахівця в області математичної статистики; - висока ціна.
MATLAB	- Зручний інтерфейс; - простота в роботі.	- Дорога ліцензія; - заплутана інтеграція із JAVA додатками.
SAS	-Швидке оброблення дуже великих обсягів даних; - можливість перетворювати формули у програмний код; - створення користувацьких модулів.	- Дорога ліцензія ; - складність опанування.
EViews	- Швидкодія; - можливість роботи із декількома файламиодночасно; - великий вибір сучасних методів для обробки даних.	- Відсутність україномовної або русифікованої версії; - невелика кількість україномовної літератури.

Розглядаючи вищенаведені переваги і недоліки та враховуючи минулу практику роботи з наведеними програмними засобами для прогнозування можливості кібератак було обрано пакет STATISTICA.

3.2 Побудова моделей прогнозування успішних кібератак

Реалізуємо прогнозування кількості успішних кібератак за допомогою програмного продукту STATISTICA, а саме пакету Аналіз, в якому використаємо функції Нейронні мережі та Регресія. Для початку введемо вхідні дані, які були зібрані у Центрі телекомунікаційних технологій та комп'ютерного забезпечення Сумського державного університету. Далі знайдемо на головному меню необхідний пункт Аналізу (автоматизовані нейронні мережі) та налаштуємо правильні параметри для проведення дослідження (рис. 3.1-.3.4). Цільовий показник – кількість кібератак, вхідні змінні – індикатор часу.

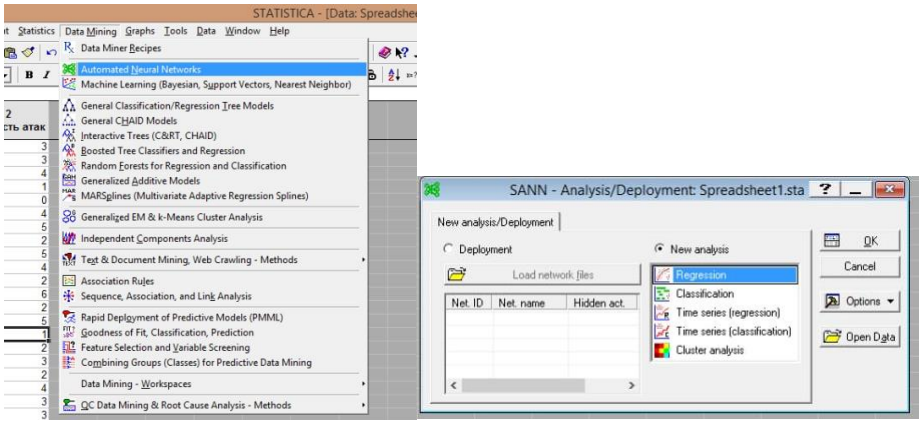

Рисунок 3.1-3.2 – Алгоритм використання пакету Аналіз STATISTICA

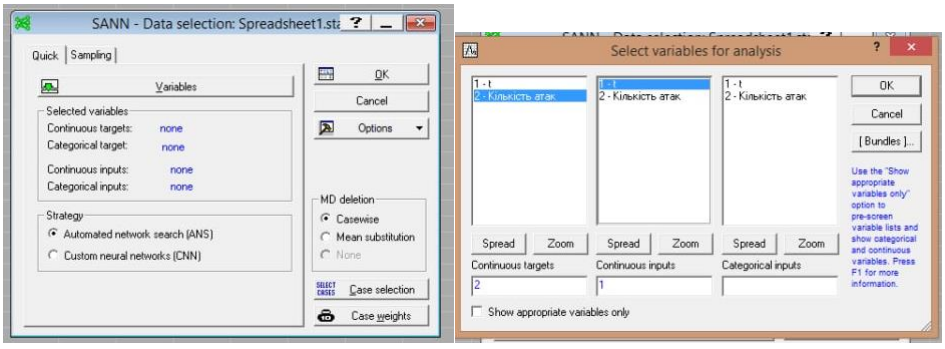

Рисунок 3.3-3.4 – Алгоритм використання пакету Аналіз STATISTICA

У результаті проведення аналізу побудовано п'ять нейронних мереж (рис. 3.5). При моделюванні використано моделі RBF (на основі радіальних базисних функцій) та MLP (на основі багатошарового персептрона).

Summary of active networks (Spreadsheet1.sta)									
Index	Net. name	Training perf.	Test perf.	Training error	Test error	Training algorithm	Error function	Hidden activation	Output activation
1	RBF 1-14-1	0,726444	0,676019	0,016179	0,009750	RBFT	SOS	Gaussian	Identity
2	RBF 1-14-1	0,777042	0,735469	0,013228	0,008493	RBFT	SOS	Gaussian	Identity
3	MLP 1-2-1	0,679831	0,616785	0,017957	0,010290	BFGS 77	SOS	Logistic	Sine
4	RBF 1-16-1	0,738458	0,624694	0,015251	0,010747	RBFT	SOS	Gaussian	Identity
5	RBF 1-16-1	0,736223	0,614843	0,015289	0,011060	RBFT	SOS	Gaussian	Identity

Рисунок 3.5 – Результати нейронного моделювання

Training perf та Test perf (продуктивність навчання та тест продуктивності відповідно) це показники за якими можемо оцінити рівень адекватності моделі. У моделей типу RBF діапазон варіації даних показників від 0,726 до 0,777. У моделі типу MLP – 0,679. Такі значення характеризують високий рівень адекватності моделей. Достовірність побудованих моделей нейронних мереж підтверджується також показниками помилки тестової, контрольної та навчальної вибірок, що приймають близькі до нуля значення.

Основою для прогнозування є створені нейронні мережі: модель архітектури RBF 1-14-1, загальна кількість шарів – 1, кількість прихованих шарів – 14; модель архітектури RBF 1-14-1, загальна кількість шарів – 1, кількість прихованих шарів – 14; модель архітектури MLP 1-2-1, загальна кількість шарів – 1, кількість прихованих шарів – 2; модель архітектури RBF 1-16-1, загальна кількість шарів – 1, кількість прихованих шарів – 16; модель архітектури RBF 1-16-1, загальна кількість шарів – 1, кількість прихованих шарів – 16. За показниками найкраще спрацювала друга модель. Результати відображені на рисунку 3.6-3.7.

1	t --> hidden neuron 1	0,898551	t --> hidden neuron 1	0,434783	t --> hidden neuron 1	10,83344	t --> hidden neuron 1	0,492754	t --> hidden neuron 1	0,420290
2	t --> hidden neuron 2	0,579710	t --> hidden neuron 2	0,826087	t --> hidden neuron 2	13,06999	t --> hidden neuron 2	0,695652	t --> hidden neuron 2	0,086957
3	t --> hidden neuron 3	0,304348	t --> hidden neuron 3	0,101449	input bias --> hidden neuron 1	-7,58074	t --> hidden neuron 3	0,623188	t --> hidden neuron 3	0,724638
4	t --> hidden neuron 4	0,478261	t --> hidden neuron 4	0,797101	input bias --> hidden neuron 2	-8,77110	t --> hidden neuron 4	0,159420	t --> hidden neuron 4	0,434783
5	t --> hidden neuron 5	0,753623	t --> hidden neuron 5	0,710145	hidden neuron 1 --> Кількість атак	-4,13074	t --> hidden neuron 5	0,275362	t --> hidden neuron 5	0,057971
6	t --> hidden neuron 6	0,101449	t --> hidden neuron 6	0,855072	hidden neuron 2 --> Кількість атак	4,15201	t --> hidden neuron 6	0,710145	t --> hidden neuron 6	0,971014
7	t --> hidden neuron 7	0,275362	t --> hidden neuron 7	0,188406	hidden bias --> Кількість атак	0,34164	t --> hidden neuron 7	0,000000	t --> hidden neuron 7	0,217391
8	t --> hidden neuron 8	0,173913	t --> hidden neuron 8	0,275362			t --> hidden neuron 8	0,898551	t --> hidden neuron 8	0,623188
9	t --> hidden neuron 9	0,000000	t --> hidden neuron 9	0,391304			t --> hidden neuron 9	0,565217	t --> hidden neuron 9	0,202899
10	t --> hidden neuron 10	0,724638	t --> hidden neuron 10	0,971014			t --> hidden neuron 10	0,449275	t --> hidden neuron 10	0,260870
11	t --> hidden neuron 11	0,565217	t --> hidden neuron 11	0,753623			t --> hidden neuron 11	0,101449	t --> hidden neuron 11	0,391304
12	t --> hidden neuron 12	0,884058	t --> hidden neuron 12	0,782609			t --> hidden neuron 12	0,608696	t --> hidden neuron 12	0,956522
13	t --> hidden neuron 13	0,695652	t --> hidden neuron 13	1,000000			t --> hidden neuron 13	0,579710	t --> hidden neuron 13	0,405797
14	t --> hidden neuron 14	0,840580	t --> hidden neuron 14	0,318841			t --> hidden neuron 14	0,304348	t --> hidden neuron 14	0,942029
15	input bias --> hidden neuron 1	0,014493	input bias --> hidden neuron 1	0,043478			t --> hidden neuron 15	0,289855	t --> hidden neuron 15	0,231884
16	input bias --> hidden neuron 2	0,014493	input bias --> hidden neuron 2	0,028986			t --> hidden neuron 16	0,840580	t --> hidden neuron 16	0,000000
17	input bias --> hidden neuron 3	0,028986	input bias --> hidden neuron 3	0,086957			input bias --> hidden neuron 1	0,043478	input bias --> hidden neuron 1	0,014493
18	input bias --> hidden neuron 4	0,086957	input bias --> hidden neuron 4	0,014493			input bias --> hidden neuron 2	0,014493	input bias --> hidden neuron 2	0,028986
19	input bias --> hidden neuron 5	0,028986	input bias --> hidden neuron 5	0,043478			input bias --> hidden neuron 3	0,014493	input bias --> hidden neuron 3	0,101449
20	input bias --> hidden neuron 6	0,072464	input bias --> hidden neuron 6	0,028986			input bias --> hidden neuron 4	0,057971	input bias --> hidden neuron 4	0,014493
21	input bias --> hidden neuron 7	0,028986	input bias --> hidden neuron 7	0,086957			input bias --> hidden neuron 5	0,014493	input bias --> hidden neuron 5	0,028986
22	input bias --> hidden neuron 8	0,072464	input bias --> hidden neuron 8	0,043478			input bias --> hidden neuron 6	0,014493	input bias --> hidden neuron 6	0,014493
23	input bias --> hidden neuron 9	0,101449	input bias --> hidden neuron 9	0,043478			input bias --> hidden neuron 7	0,101449	input bias --> hidden neuron 7	0,014493
24	input bias --> hidden neuron 10	0,028986	input bias --> hidden neuron 10	0,028986			input bias --> hidden neuron 8	0,057971	input bias --> hidden neuron 8	0,101449
25	input bias --> hidden neuron 11	0,014493	input bias --> hidden neuron 11	0,028986			input bias --> hidden neuron 9	0,014493	input bias --> hidden neuron 9	0,014493
26	input bias --> hidden neuron 12	0,014493	input bias --> hidden neuron 12	0,014493			input bias --> hidden neuron 10	0,043478	input bias --> hidden neuron 10	0,028986

Рисунок 3.6 – Фрагмент архітектури нейронних мереж з прихованими шарами

27	input bias --> hidden neuron 13	0,028986	input bias --> hidden neuron 13	0,028986			input bias --> hidden neuron 11	0,057971	input bias --> hidden neuron 11	0,014493
28	input bias --> hidden neuron 14	0,043478	input bias --> hidden neuron 14	0,043478			input bias --> hidden neuron 12	0,014493	input bias --> hidden neuron 12	0,014493
29	hidden neuron 1 --> Кількість атак	0,019551	hidden neuron 1 --> Кількість атак	0,009611			input bias --> hidden neuron 13	0,014493	input bias --> hidden neuron 13	0,014493
30	hidden neuron 2 --> Кількість атак	-0,004040	hidden neuron 2 --> Кількість атак	0,004546			input bias --> hidden neuron 14	0,014493	input bias --> hidden neuron 14	0,014493
31	hidden neuron 3 --> Кількість атак	0,008020	hidden neuron 3 --> Кількість атак	-0,024699			input bias --> hidden neuron 15	0,014493	input bias --> hidden neuron 15	0,014493
32	hidden neuron 4 --> Кількість атак	0,027490	hidden neuron 4 --> Кількість атак	0,023247			input bias --> hidden neuron 16	0,057971	input bias --> hidden neuron 16	0,057971
33	hidden neuron 5 --> Кількість атак	0,034104	hidden neuron 5 --> Кількість атак	0,066145			hidden neuron 1 --> Кількість атак	0,013202	hidden neuron 1 --> Кількість атак	-0,013979
34	hidden neuron 6 --> Кількість атак	0,039401	hidden neuron 6 --> Кількість атак	0,023657			hidden neuron 2 --> Кількість атак	0,015324	hidden neuron 2 --> Кількість атак	0,010711
35	hidden neuron 7 --> Кількість атак	0,003543	hidden neuron 7 --> Кількість атак	0,024796			hidden neuron 3 --> Кількість атак	-0,013355	hidden neuron 3 --> Кількість атак	0,136267
36	hidden neuron 8 --> Кількість атак	0,004860	hidden neuron 8 --> Кількість атак	-0,032488			hidden neuron 4 --> Кількість атак	0,001993	hidden neuron 4 --> Кількість атак	0,002688
37	hidden neuron 9 --> Кількість атак	-0,036247	hidden neuron 9 --> Кількість атак	-0,031151			hidden neuron 5 --> Кількість атак	0,003810	hidden neuron 5 --> Кількість атак	-0,021659
38	hidden neuron 10 --> Кількість атак	-0,025603	hidden neuron 10 --> Кількість атак	-0,002170			hidden neuron 6 --> Кількість атак	0,013186	hidden neuron 6 --> Кількість атак	0,017475
39	hidden neuron 11 --> Кількість атак	0,011711	hidden neuron 11 --> Кількість атак	-0,001070			hidden neuron 7 --> Кількість атак	0,000443	hidden neuron 7 --> Кількість атак	-0,013553
40	hidden neuron 12 --> Кількість атак	-0,018920	hidden neuron 12 --> Кількість атак	-0,011872			hidden neuron 8 --> Кількість атак	-0,023959	hidden neuron 8 --> Кількість атак	-0,064221
41	hidden neuron 13 --> Кількість атак	0,069188	hidden neuron 13 --> Кількість атак	0,023370			hidden neuron 9 --> Кількість атак	0,013463	hidden neuron 9 --> Кількість атак	0,000264
42	hidden neuron 14 --> Кількість атак	0,064236	hidden neuron 14 --> Кількість атак	0,036048			hidden neuron 10 --> Кількість атак	-0,025884	hidden neuron 10 --> Кількість атак	-0,006166
43	hidden bias --> Кількість атак	0,231946	hidden bias --> Кількість атак	0,354818			hidden neuron 11 --> Кількість атак	-0,023294	hidden neuron 11 --> Кількість атак	-0,011639
44							hidden neuron 12 --> Кількість атак	0,006778	hidden neuron 12 --> Кількість атак	-0,022866
45							hidden neuron 13 --> Кількість атак	-0,013195	hidden neuron 13 --> Кількість атак	0,009886
46							hidden neuron 14 --> Кількість атак	0,004265	hidden neuron 14 --> Кількість атак	0,010720
47							hidden neuron 15 --> Кількість атак	-0,010488	hidden neuron 15 --> Кількість атак	0,007822
48							hidden neuron 16 --> Кількість атак	0,069962	hidden neuron 16 --> Кількість атак	-0,009448
49							hidden bias --> Кількість атак	0,414278	hidden bias --> Кількість атак	0,426331

Рисунок 3.6 – Фрагмент архітектури нейронних мереж з прихованими шарами

Запишемо моделі на основі реальних даних. Перша модель:

$$ab_1^{(2)} = f(0{,}8985c_1 + 0{,}0144) \quad (3.1)$$

$$ab_2^{(2)} = f(0{,}5797c_1 + 0{,}0144)$$

$$ab_3^{(2)} = f(0{,}3043c_1 + 0{,}0289)$$

$$ab_4^{(2)} = f(0{,}4782c_1 + 0{,}0869)$$

$$ab_5^{(2)} = f(0{,}7536c_1 + 0{,}0289)$$

$$ab_6^{(2)} = f(0{,}1014c_1 + 0{,}0724)$$

$$ab_7^{(2)} = f(0{,}2753c_1 + 0{,}0289)$$

$$ab_8^{(2)} = f(0{,}1739c_1 + 0{,}0724)$$

$$ab_9^{(2)} = f(0{,}0c_1 + 0{,}1014)$$

$$ab_{10}^{(2)} = f(0{,}7246c_1 + 0{,}0289)$$

$$ab_{11}^{(2)} = f(0{,}5652c_1 + 0{,}0144)$$

$$ab_{12}^{(2)} = f(0{,}8841c_1 + 0{,}0144)$$

$$ab_{13}^{(2)} = f(0{,}6956c_1 + 0{,}0289)$$

$$ab_{14}^{(2)} = f(0{,}8405c_1 + 0{,}0434)$$

$$S = k^{(3)} = f(0{,}0195b_1^{(2)} - 0{,}0040ab_2^{(2)} + 0{,}0080b_3^{(2)} + 0{,}0274b_4^{(2)} + 0{,}0341b_5^{(2)} + 0{,}0394b_6^{(2)} + 0{,}0035b_7^{(2)} + 0{,}0048b_8^{(2)} - 0{,}0362ab_9^{(2)} - 0{,}0256b_{10}^{(2)} + 0{,}0117b_{11}^{(2)} - 0{,}0189ab_{12}^{(2)} + 0{,}0691b_{13}^{(2)} + 0{,}0642b_{14}^{(2)} + 0{,}2319)$$

Друга модель:

$$ab_1^{(2)} = f(0{,}4347c_1 + 0{,}0434) \quad (3.2)$$

$$ab_2^{(2)} = f(0{,}8260c_1 + 0{,}0289)$$

$$ab_3^{(2)} = f(0{,}1014c_1 + 0{,}0869)$$

$$ab_4^{(2)} = f(0{,}7971c_1 + 0{,}0144)$$

$$ab_5^{(2)} = f(0{,}7101c_1 + 0{,}0434)$$

$$ab_6^{(2)} = f(0{,}8551c_1 + 0{,}0289)$$

$$ab_7^{(2)} = f(0{,}1884c_1 + 0{,}0869)$$

$$ab_8^{(2)} = f(0{,}2753c_1 + 0{,}0434)$$

$$ab_9^{(2)} = f(0{,}3913c_1 + 0{,}0434)$$

$$ab_{10}^{(2)} = f(0{,}9711c_1 + 0{,}0289)$$

$$ab_{11}^{(2)} = f(0{,}7536c_1 + 0{,}0289)$$

$$ab_{12}^{(2)} = f(0{,}7826c_1 + 0{,}0144)$$

$$ab_{13}^{(2)} = f(1{,}0c_1 + 0{,}0289)$$

$$ab_{14}^{(2)} = f(0{,}3188c_1 + 0{,}0434)$$

$$S = k^{(3)} = f(0{,}0096ab_1^{(2)} + 0{,}0045ab_2^{(2)} - 0{,}0246ab_3^{(2)} + 0{,}0232ab_4^{(2)} + 0{,}0661ab_5^{(2)} + 0{,}0236ab_6^{(2)} + 0{,}0247ab_7^{(2)} - 0{,}0324ab_8^{(2)} - 0{,}0311ab_9^{(2)} - 0{,}0021ab_{10}^{(2)} - 0{,}0011ab_{11}^{(2)} - 0{,}0118ab_{12}^{(2)} + 0{,}0233ab_{13}^{(2)} + 0{,}0361ab_{14}^{(2)} + 0{,}3548)$$

Третя модель на основі реальних даних:

$$ab_1^{(2)} = f(10{,}8334c_1 - 7{,}5807) \quad (3.3)$$

$$ab_2^{(2)} = f(13{,}0699c_1 - 8{,}7711)$$

$$S = k^{(3)} = f(-4{,}1307ab_1^{(2)} - 4{,}1520ab_2^{(2)} + 0{,}3416)$$

Вигляд четвертої моделі на основі реальних даних:

$$ab_1^{(2)} = f(0{,}4927c_1 + 0{,}0434) \quad (3.4)$$

$$ab_2^{(2)} = f(0{,}6956c_1 + 0{,}0144)$$

$$ab_3^{(2)} = f(0{,}6231c_1 + 0{,}0144)$$

$$ab_4^{(2)} = f(0{,}1594c_1 + 0{,}0579)$$

$$ab_5^{(2)} = f(0{,}2753c_1 + 0{,}0144)$$

$$ab_6^{(2)} = f(0{,}7101c_1 + 0{,}0144)$$

$$ab_7^{(2)} = f(0{,}0c_1 + 0{,}1014)$$

$$ab_8^{(2)} = f(0{,}8985c_1 + 0{,}0579)$$

$$ab_9^{(2)} = f(0{,}5652c_1 + 0{,}0144)$$

$$ab_{10}^{(2)} = f(0{,}4492c_1 + 0{,}0434)$$

$$ab_{11}^{(2)} = f(0{,}1014c_1 + 0{,}0579)$$

$$ab_{12}^{(2)} = f(0{,}6086c_1 + 0{,}0144)$$

$$ab_{13}^{(2)} = f(0{,}5797c_1 + 0{,}0144)$$

$$ab_{14}^{(2)} = f(0{,}3043c_1 + 0{,}0144)$$

$$ab_{15}^{(2)} = f(0{,}2898c_1 + 0{,}0144)$$

$$ab_{16}^{(2)} = f(0{,}8405c_1 + 0{,}0579)$$

$$S = k^{(3)} = f(0{,}0132b_1^{(2)} + 0{,}0153b_2^{(2)} - 0{,}0133b_3^{(2)} + 0{,}0019b_4^{(2)} + 0{,}0038b_5^{(2)} + 0{,}0131b_6^{(2)} + 0{,}0004b_7^{(2)} - 0{,}0239b_8^{(2)} + 0{,}0134b_9^{(2)} - 0{,}0258b_{10}^{(2)} - 0{,}0232b_{11}^{(2)} + 0{,}0067b_{12}^{(2)} - 0{,}0131b_{13}^{(2)} + 0{,}0042b_{14}^{(2)} - 0{,}0104b_{15}^{(2)} + 0{,}0699b_{16}^{(2)} + 0{,}4142)$$

Вигляд п’ятої моделі на основі реальних даних:

$$ab_1^{(2)} = f(0{,}4202c_1 + 0{,}0144) \quad (3.5)$$

$$ab_2^{(2)} = f(0{,}0869c_1 + 0{,}0289)$$

$$ab_3^{(2)} = f(0{,}7246c_1 + 0{,}1014)$$

$$ab_4^{(2)} = f(0{,}4347c_1 + 0{,}0144)$$

$$ab_5^{(2)} = f(0{,}0579c_1 + 0{,}0289)$$

$$ab_6^{(2)} = f(0{,}9710c_1 + 0{,}0144)$$

$$ab_7^{(2)} = f(0{,}2173c_1 + 0{,}0144)$$

$$ab_8^{(2)} = f(0{,}6231c_1 + 0{,}1014)$$

$$ab_9^{(2)} = f(0{,}2028c_1 + 0{,}0144)$$

$$ab_{10}^{(2)} = f(0{,}2606c_1 + 0{,}0289)$$

$$ab_{11}^{(2)} = f(0{,}3913c_1 + 0{,}0144)$$

$$ab_{12}^{(2)} = f(0{,}9565c_1 + 0{,}0144)$$

$$ab_{13}^{(2)} = f(0{,}4057c_1 + 0{,}0144)$$

$$ab_{14}^{(2)} = f(0{,}9420c_1 + 0{,}0144)$$

$$ab_{15}^{(2)} = f(0{,}2318c_1 + 0{,}0144)$$

$$ab_{16}^{(2)} = f(0{,}0c_1 + 0{,}0579)$$

$$S = k^{(3)} = f(-0{,}0139b_1^{(2)} + 0{,}0107b_2^{(2)} + 0{,}1362b_3^{(2)} + 0{,}0026b_4^{(2)} - 0{,}0216b_5^{(2)} + 0{,}0174b_6^{(2)} - 0{,}0135b_7^{(2)} - 0{,}0642b_8^{(2)} + 0{,}0002b_9^{(2)} - 0{,}0061b_{10}^{(2)} - 0{,}0116b_{11}^{(2)} - 0{,}0228b_{12}^{(2)} + 0{,}0098b_{13}^{(2)} + 0{,}0107b_{14}^{(2)} + 0{,}0078b_{15}^{(2)} - 0{,}0094b_{16}^{(2)} + 0{,}4263)$$

Наступним етапом є прогнозування майбутніх показників об’єкта дослідження. Для наглядності побудуємо діаграму розсіювання (рис. 3.7). Вона показує розташування знайдених значень відносно фактичних. Достатньо щільне положення одних значень відносно інших свідчить, що достовірність моделі є високою.

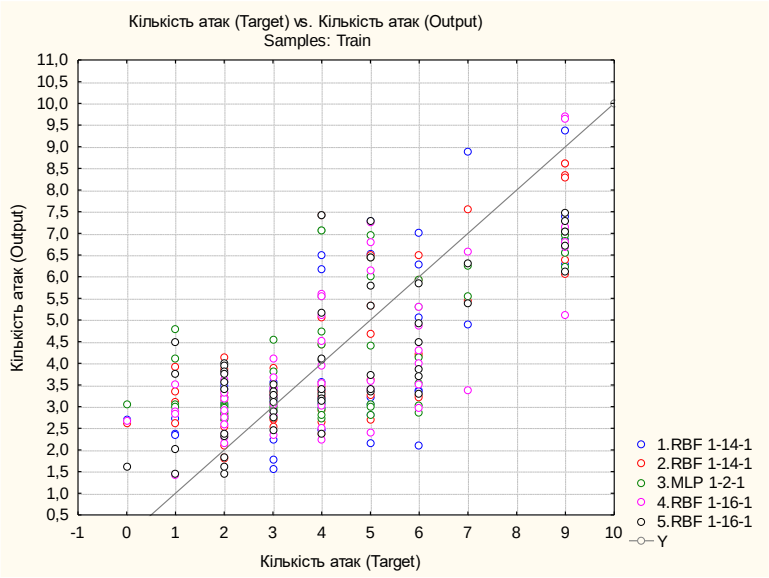

Рисунок 3.7 – Графік розсіювання фактичних та прогнозованих значень

Проаналізуємо характеристики якості моделей (рис.3.8) та її чутливість до зміни вхідних даних (рис. 3.9). Невеликий розмах варіацій серед мінімальних та максимальних рівнів навчальної і тестової вибірок у межах від 1,4 до 9,6 свідчить, що спроектовані моделі є якісними. Показник чутливості в рамках 2,3-2,8 визначає модель досить чутливою до змін масштабів вхідних даних.

	Predictions statistics (Spreadsheet1.sta) Target: Кількість атак				
Statistics	1.RBF 1-14-1	2.RBF 1-14-1	3.MLP 1-2-1	4.RBF 1-16-1	5.RBF 1-16-1
Minimum prediction (Train)	1,55089	1,81123	2,70965	1,41856	1,44521
Maximum prediction (Train)	9,37638	8,61287	7,09750	9,69348	7,47231
Minimum prediction (Test)	2,07268	2,22480	2,76502	1,96271	1,87281
Maximum prediction (Test)	6,87523	8,07790	6,65790	7,07075	6,77189
Minimum prediction (Validation)					
Maximum prediction (Validation)					
Minimum residual (Train)	-3,90610	-2,92902	-3,14524	-3,89236	-2,88183
Maximum residual (Train)	2,69555	3,42679	3,80149	2,67949	3,48924
Minimum residual (Test)	-2,89198	-2,09210	-1,96540	-2,56851	-2,12719
Maximum residual (Test)	1,42478	2,55538	3,17982	1,80520	2,92725
Minimum residual (Validation)					
Maximum residual (Validation)					
Minimum standard residual (Train)	-2,41271	-2,00084	-1,84406	-2,47629	-1,83113
Maximum standard residual (Train)	1,66498	2,34087	2,22882	1,70467	2,21709
Minimum standard residual (Test)	-2,30108	-1,78354	-1,52223	-1,94665	-1,58918
Maximum standard residual (Test)	1,13366	2,17851	2,46282	1,36815	2,18689
Minimum standard residual (Validation)					
Maximum standard residual (Validation)					

Рисунок 3.8 – Статистика передбачуваних значень

	Sensitivity analysis (Spreadsheet1.sta) Samples: Train
Networks	t
1.RBF 1-14-1	2,396024
2.RBF 1-14-1	2,813456
3.MLP 1-2-1	2,467673
4.RBF 1-16-1	2,360820
5.RBF 1-16-1	2,541353
Average	2,515865

Рисунок 3.9 – Показник чутливості прогнозованих моделей нейронних мереж

Отже, побудувавши прогноз на основі п’яти моделей нейронних мереж, отримали наступні результати, які відображені на рисунку 3.10 та за допомогою графіка (рис. 3.11) . Прогнозовані значення були розраховані на 18 періодів, тобто на 1,5 року функціонування організації. Тенденцій до різких перепадів чи піків графік не відображає. Стосовно результатів даного дослідження можна стверджувати, що прогнозування методом нейронних мереж є ефективним, наочним та зручним інструментом. Даний тип дослідження доцільно використовувати при великих масивах даних. Перевагою відносно інших традиційних алгоритмів є здатність виявляти залежності між вхідними і вихідними даними та навчатися на їх основі при прогнозуванні.

	Custom predictions spreadsheet (Spreadsheet1.sta)				
Cases	1.Кількі	2.Кількі	3.Кількі	4.Кількі	5.Кількі
1	2,089297	5,660767	4,040203	3,575807	3,963258
2	2,087961	4,912781	3,943138	3,632476	3,891226
3	2,087616	4,121362	3,857197	3,672778	3,871119
4	2,087537	3,582147	3,781355	3,698503	3,858575
5	2,087521	3,319963	3,714615	3,713467	3,850368
6	2,087518	3,225430	3,656031	3,721469	3,845111
7	2,087517	3,199684	3,604718	3,725426	3,841818
8	2,087517	3,194332	3,559860	3,727242	3,839800
9	2,087517	3,193478	3,520712	3,728017	3,838591
10	2,087517	3,193373	3,486599	3,728325	3,837882
11	2,087517	3,193363	3,456913	3,728440	3,837475
12	2,087517	3,193362	3,431110	3,728480	3,837247
13	2,087517	3,193362	3,408708	3,728493	3,837121
14	2,087517	3,193362	3,389276	3,728497	3,837054
15	2,087517	3,193362	3,372436	3,728498	3,837018
16	2,087517	3,193362	3,357853	3,728498	3,837000
17	2,087517	3,193362	3,345235	3,728498	3,836990
18	2,087517	3,193362	3,334324	3,728498	3,836986

Рисунок 3.10 – Прогнозовані значення успішних кібератак моделями нейронних мереж

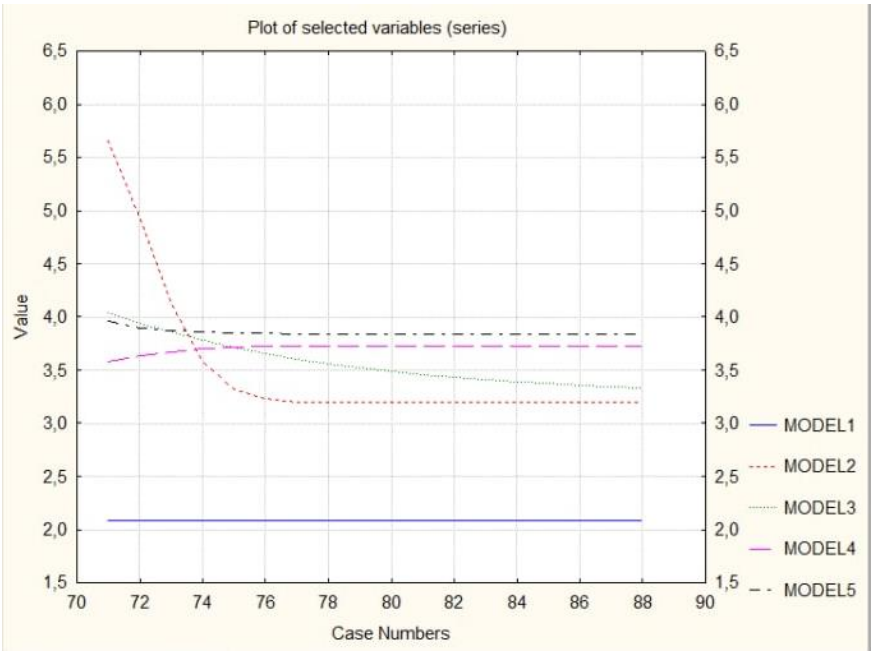

Рисунок 3.11 – Графік прогнозу обсягу успішних кібератак

У підсумку, моделі дають схожі результати, які близькі до реальних даних. Це у свою чергу свідчить про адекватність моделей та коректність їх використання при прогнозуванні можливості успішних кібератак на інформаційний простір Сумського державного університету, які приведуть до повної або часткової втрати інформації співробітників чи тимчасового виведення з ладу технічного забезпечення.

ВИСНОВКИ

У результаті проведеної роботи було побудовано модель прогнозування успішних кібератак на інформаційний простір СумДУ, які приведуть до повної або часткової втрати інформації співробітників чи тимчасового виведення з ладу технічного забезпечення. Дане дослідження основувалося на одному із методів Data Mining, а саме побудові моделей нейронних мереж. Для побудови моделі використали пакет STATISTICA враховуючи його простоту та можливості функціоналу.

Під час виконання дипломної роботи були виконані наступні завдання:

— проведено аналіз стану захищеності інформаційного простору СумДУ;

— розглянуто напрямки кібератак, способи та засоби їх реалізації; визначено дії для захисту від них; проаналізовано найпопулярніші способи атак;

— проаналізовано та порівняно методи Data Mining для прогнозування;

— описано моделі та сформовано вимоги до них;

— систематизовано та охарактеризовано вхідні дані;

— використано метод Ірвіна для перевірки даних на однорідність;

— розглянуто та порівняно найпопулярніше програмне забезпечення для побудови моделі;

— побудовано п’ять моделей нейронних мереж для прогнозування;

— визначена адекватність побудованих моделей;

— побудований прогноз успішних кібератак на інформаційний простір СумДУ.

У підсумку, моделі показали схожі результати, які близькі до реальних даних. Це свідчить про їх адекватність та коректність. Прогноз може бути використаний для прийняти правильних рішень, що покращать ситуації з небажаними втручаннями.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. 2021: Data theft as a result of cyber attacks. URL: https://tadviser.com/index.php/Company:Toyota_Auto_Body (дата звернення 08.11.2021).

2. A. Korchenko, K. Warwas, A. Kłos-Witkowska, «The Tupel Model of Basic Components' Set Formation for Cyberattacks», Proceedings of the 2015 IEEE 8th International Conference on «Intelligent Data Ac-quisition and Advanced Computing Systems: Technology and Applications» (IDAACS'2015), Warsaw, Poland, Septem-ber 24-26,: Vol. 1, pp. 478-483, 2015.

3. B. Akhemetov, A. Korchenko, S. Akhmetova, N. Zhumangalieva, «Improved method for the for-mation of linguistic standards for of intrusion detection systems», Journal of Theoretical and Applied Information Technology, vol.87. №.2, p. 221-232, 2016

4. Cybersecurity in 2021: Cost Projections and Top Trends URL: https://techexpert.ua/ru/cybersecurity-in-2021/ (дата звернення 12.11.2021).

5. Статистичний аналіз даних за допомогою пакету STATISTICA URL: http://matphys.rpd.univ.kiev.ua/downloads/courses/mmatstat/StatAn.doc (дата звернення 10.11.2021).

6. Data Mining / Дюк В., Самойленко А. - СПб. ; М. ; Харьков : Питер, 2001. – 366 с.

7. EViews Econometric Modeling Software URL: https://www.ihs.com/products/eviews-econometric-modeling-analysissoftware.html (дата звернення 10.11.2021).

8. Firmino P., de Mattos Neto P. & Ferreira T. Correcting and combining time series forecasters // Neural Networks. — 2014. — 50. — P. 1–11

9. Hackers behind Colonial Pipeline attack reportedly received $90 million in bitcoin before shutting down. URL: https://www.cnbc.com/2021/05/18/colonial-

pipeline-hackers-darkside-received-90-million-in-bitcoin.html Accessed 20 Sept 2021 (дата звернення 08.11.2021).

10. Information technology – Security techniques – Guidelines for cybersecurity . URL: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=4641 (дата звернення 10.11.2021).

11. JBS says it paid $11 million ransom after cyberattack. URL: https://edition.cnn.com/2021/06/09/business/jbs-cyberattack-11-million/index.html Accessed 20 Sept 2021 (дата звернення 08.11.2021).

12. Lawrence R. Using Neural Networks to Forecast Stock Market Prices. New York; London: Oxford University Press, 1997. 326 p.

13. M. Karpinski, A. Korchenko, P. Vikulov, R. Kochan, «The Etalon Models of Linguistic Variables for Sniffing-Attack Detection», Proceedings of the 2017 IEEE 9th International Conference on «Intelligent Data Ac-quisition and Advanced Computing Systems: Technology and Applications» (IDAACS'2017), Romania, Bucharest, Sep-tember 21-23, Vol. 1, p. 258-264, 2017

14. Robert J. Van Eyden The Application of Neural Networks in the Forecasting of Share Prices. New York: Finance and Technology Publishing, 1996. 326 p

15. Rombovsky M. Y. Automation of Research Area Recognition Under Weak-Contrast Borders on a Picture of Transparent Objects / M. Y. Rombovsky, R. V. Radchenko. // Theory and Practice of Forencic Science and Criminalistics. – 2019. – №19(1). – pp. 556–567. doi.org/10.32353/KHRIFE.1.2019.45

16. Schmidhuber, J. Deep Learning in Neural Networks: An Overview // Neural Networks. – 2015. – Vol. 61. – P. 85–117.

17. Sen. J. A Robust Mechanism for Defendeing Distributed Denial OF Service Attacks on Web Servers / J. Sen // International Journal of Network Security & Its Applications (IJNSA). – 2011, March. – Vol. 3, N 2. – P. 162-179.

18. Setiono N. & Liu R. Improving backpropagation learning with feature selection. Applied Intelligence / N. Setiono & R. Liu // The International Journal of Artificial Intelligence, Neural Networks, and Complex ProblemSolving Technologies. — 1996. — № 62. — P. 129–139.

19. The Root Cause Of Poly Network Being Hacked. URL: https://slowmist.medium.com/the-rootcause-of-poly-network-being-hacked-ec2ee1b0c68f (дата звернення 08.11.2021).

20. Toyota Parts Supplier Hit By $37 Million Email Scam. URL: https://www.forbes.com/sites/leemathews/2019/09/06/toyota-parts-supplier-hit-by-37-million-email-scam/?sh=4cd0aeef5856 (дата звернення 08.11.2021).

21. Yoon Y., G. Swales Applying Artificial Neural Networks to Investment Analysis. London: Taylor & Francis, 2011. 80 c

22. Алгоритми навчання нейронних мереж. URL: https://uadoc.zavantag.com/text/24469/index-1.html (дата звернення 12.11.2021).

23. Бахрушин В. Є. Методи аналізу даних: навчальний посібник / В. Є. Бахрушин. – Запоріжжя: КПУ, 2011. – 268 с.

24. Бестужев-Лада И. В. Рабочая книга по прогнозированию / И. В. Бестужев-Лада. – М.: Мысль, 1982. – С. 10–21.

25. Бідюк П. І., Романенко В. Д., Тимощук О. Л. Аналіз часових рядів: навч. посіб. / ННК «Інститут прикладного системного аналізу» Національний технічний університет України «Київський політехнічний інститут», 2010. 317 с.

26. Бідюк П.І. Аналіз часових рядів: навчальний посібник. К: Політехніка, 2010. 317 с.

27. Бурячок В. Л. Політика інформаційної безпеки : підручник / В. Л. Бурячок, Р. В. Грищук, В. О. Хорошко ; під заг. ред. проф. В. О. Хорошка. – К.: ПВП «Задруга», 2014. – 222 с.

28. Винничук Р. О. Особливості розвитку ІТ-ринку в Україні: стан та тенденції / Р. О. Винничук, Т. В. Склярук // Вісник Національного університету "Львівська політехніка". Логістика. – 2015. – № 833. – С. 3-8

29. Вишнев С.М. Основы комплексного прогнозирования / С.М. Вишнев. - М.: Наука, 1977. - 287 с.

30. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учеб.пособ.– 2-е изд., испр. и доп. – М.: Форум, 2008.– 464 с.

31. Грешилов А. А. Математические методы построения прогнозов. Москва: Радио и связь, 1997. 112 с.

32. Данильченко О.М., Данильченко А.О. Інтелектуальний аналіз даних: Навч. посібник. – Житомир: ЖДТУ, 2009. – 405 с.

33. Жуков Ю.В. Основы веб-хакинга: нападение и защита / Ю.В. Жуков – СПб.: Питер, 2011. – 176 с.

34. Зенкін А.И. Про математичні методи прогнозування М: 2001. - 90с.

35. Інформаційні технології. Словник термінів. Частина 34. Штучний інтелект. Нейронні мережі : ДСТУ ISO/IEC 2382-34-2003. – К. : Держспоживстандарт, 2005. – 20 с.

36. Каллан, Р. Основные концепции нейронных сетей / Р. Каллан. – М.: Издательский дом «Вильямс», 2001. – 287 с.

37. Корченко О.Г. Системи захисту інформації: Монографія / Корченко О.Г. - К.: НАУ, 2004. – 264 с.

38. Ларіонцева, Е.А. Проблеми і засоби захисту інформації. / Е.А. Ларіонцева // Наука і освіта. - 2011. - № 4.- С. 83-87.

39. Мозолевська М.О., Ставицький О.В. Використання нейронних мереж для прогнозування/ – Київ: КПІ, 2017. – 4 с.

40. Носко В. П. Эконометрика. Введение в регрессионный анализ временных рядов. Москва: Литкон, 2002. 273 с

41. Овчаров В.В. Програмні засоби моделювання штучних нейронних мереж прогнозування / Науковий вісник ТДАТУ. – Випуск 2, 2007. – 27с

42. Олійник А. О. Інтелектуальний аналіз даних: навчальний посібник/ А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя: ЗНТУ, 2012. – 271 с.

43. Осовський С. Нейронні мережі для обробки інформації.М.:фінанси та статистика, 2002. –344с.

44. Персептрон URL: https://znaimo.com.ua/ (дата звернення 12.11.2021).

45. Разработка моделей и методов для создания системы информационной безопасности корпоративной сети предприятия с учетом различных критериев. URL: http://masters.donntu.org/2009/fvti/khimka/diss/index.htm (дата звернення 10.11.2021).

46. Руденко О. Г. Штучні нейронні мережі/ О. Г. Руденко, Є. В. Бодянський.– Харків : Компанія СМІТ, 2006. – 404 с.

47. Рутковская, Д. Нейронные сети и генетические алгоритмы/ Д. Рутковская, М. Пипиньский, Л. Рутковский. – М.: Горячая линияТелеком, 2006. – 452 с.

48. Ситник В. Ф., Краснюк М. Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. посібник. — К.: КНЕУ, 2007. — 376 с.

49. Ставицький А. В. Навчально-методичний комплекс з курсів «Прогнозування» та «Фінансове прогнозування». Київ: Центр учб. літ., 2006. 107 с.

50. Субботін, С. О. Нейронні мережі : навчальний посібник / С. О. Субботін, А. О. Олійник ; під заг. ред. проф. С. О. Субботіна. – Запоріжжя : ЗНТУ, 2014. – 132 с.

51. Хайкин, С. Нейронные сети: полный курс / С. Хайкин ; пер. с англ. – М. : Вильямс, 2006. – 1104 с.

52. Черняк О.І. Інтелектуальний аналіз даних: підручник. – К: Знання, 2014. –599с.

53. Щука В. Г. Дослідження методів прогнозування та обґрунтування вибору кращого/ В. Г. Щука, Д. І. Мандрик // Вісник Хмельницького національного університету. – 2015. – № 1. – С. 102-104.

54. Ярушкина Н.Г. Интеллектуальный анализ временных рядов : учебное пособие / Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева. – Ульяновск: УлГТУ, 2010. – 320 с.

ДОДАТКИ

Додаток А

SUMMARY

Dovga Y. O. Data Mining to study the security of the Sumy State University information space: economic risks' modeling – Masters-level Qualification Thesis. Sumy State University, Sumy, 2021.

The analysis of the state of protection of the information space of Sumy State University is performed in the work; types of cyberattacks are considered and actions for protection against them are defined; analyzed and compared Data Mining methods; considered and compared the most popular software for model building; built and tested for the adequacy of the neural network model for forecasting. The main purpose of the study is to build a model for predicting the possibility of successful cyber attacks on the studied information space.

Keywords: Data Mining, cyber attack, forecasting, modeling, neural networks.

АНОТАЦІЯ

Довга Ю. О. Data Mining для дослідження стану захищеності інформаційного простору Сумського державного університету – кваліфікаційна магістерська робота. Сумський державний університет, Суми, 2021р.

У роботі виконано аналіз стану захищеності інформаційного простору Сумського державного університету; розглянуто види кібератак та визначено дії для захисту від них; проаналізовано та порівняно методи Data Mining; розглянуто та порівняно найпопулярніше програмне забезпечення для побудови моделі; побудовано та перевірено на адекватність моделі нейронних мереж для прогнозування. Основна мета дослідження полягає у побудові моделі прогнозування можливості успішних кібератак на досліджуваний

інформаційний простір.

Ключові слова: Data Mining, кібератака, прогнозування, моделювання, нейронні мережі.