

<https://doi.org/10.31891/2219-9365-2022-71-3-6>

УДК 681.3.06

ХРИСТИНА ЛІП'ЯНИНА-ГОНЧАРЕНКО

Західноукраїнський національний університет

<https://orcid.org/0000-0002-2441-6292>

e-mail: xrustya.com@gmail.com

МИРОСЛАВ КОМАР

Західноукраїнський національний університет

<https://orcid.org/0000-0001-6541-0359>

e-mail: mko@wunu.edu.ua

АНАТОЛІЙ САЧЕНКО

Західноукраїнський національний університет

<https://orcid.org/0000-0002-0907-3682>

e-mail: as@wunu.edu.ua

ТАРАС ЛЕНДЮК

Західноукраїнський національний університет

<https://orcid.org/0000-0001-9484-8333>

e-mail: tl@wunu.edu.ua

ОЦІНКА ІНВЕСТИЦІЙНИХ РИЗИКІВ ВІРТУАЛЬНОЇ ІТ-КОМПАНІЇ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Розроблено модуль для оцінки інвестиційних ризиків віртуальної ІТ-компанії на основі машинного навчання, що дозволяє зменшити часові витрати на оцінку ризиків інвестора віртуальної ІТ-компанії. Проведено детальне обґрунтування кожного вибраного параметру-ризика, що впливає на успішність інвестиційного проекту віртуальної ІТ-компанії. Алгоритм оцінки інвестиційних ризиків віртуальної ІТ-компанії на основі машинного навчання сформований на основі бального оцінювання 10 експертами з 20 реалізованих проектів за 23 параметрами-ризиками. Проведено оцінку прогнозування з моделювання оцінки інвестиційних ризиків віртуальної ІТ-компанії використовуючи машинне навчання на основі восьми методів: Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression. Та розроблено модуль для підтримки прийняття рішень на основі трьох методів з найкращим показником прогнозу, а саме: Support Vector Classifier, Random Forest Classifier, K-Neighbors Classifier.

Ключові слова: інвестиційні ризики, віртуальне підприємство, Support Vector Classifier, Random Forest Classifier, K-Neighbors Classifier, машинне навчання.

KHRYSTYNA LIPIANINA-HONCHARENKO, MYROSLAV KOMAR,
ANATOLIY SACHENKO, TARAS LENDIUK

West Ukrainian National University

ASSESSING THE INVESTMENT RISK OF VIRTUAL IT COMPANY BASED ON MACHINE LEARNING

A module for assessing the investment risks of a virtual IT company has been developed. It enables to reduce the time spent on assessing the investor's risks of a virtual IT company. A detailed justification of each selected risk parameter that influences on the success of the investment project of the virtual IT Company has done. A developed algorithm for assessing the investment risk of the virtual IT company is based on machine learning and using the expert scoring method (10 experts from 20 implemented projects were involved) by 23 risk parameters. Forecasting of investment risk assessment modeling of the virtual IT company using machine learning is based on eight methods: Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression. In addition, a module was developed to support decision-making based on three methods with the best forecast, namely: Support Vector Classifier, Random Forest Classifier, K-Neighbors Classifier.

Keywords: investment risk, virtual enterprise, Support Vector Classifier, Random Forest Classifier, K-Neighbors Classifier, machine learning.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

У сучасних умовах господарювання виникає потреба в продукті (послугах), що вимагає інноваційного підходу до його виробництва. Все це створює потребу в нових формах ведення бізнесу, однією з яких є віртуальні підприємства, що набувають все більшого розвитку. Віртуальне підприємство створюється шляхом відбору людських, фінансових, матеріальних, організаційно-технологічних та інших ресурсів з різних підприємств та їх інтеграції з використанням комп'ютерних мереж [45]. Збільшенню прибутку віртуального підприємства сприяє ефективне управління ризиками віртуального підприємства.

Під інвестиційним проектом розуміємо сукупність певних елементів і зв'язків між ними, що

забезпечує досягнення інвестиційних цілей протягом встановленого часу та обмежених ресурсів та сприяє подальшому розвитку економічних процесів. Оскільки інвестиційний проект націлений на довготривалий період реалізації початкових вкладень чим передбачає забезпечення процесів розширеного відтворення, то повинен компенсувати інвестору не тільки суму вкладень, а й забезпечити отримання вигоди від фінансування ним подальшого розвитку об'єкту. Разом з тим, зовнішнє оточення характеризується не лише сукупністю факторів, але й певними характеристиками.

Причиною появи ризику є невизначеність та відсутність повної вичерпної інформації, яка зумовлює неможливість адекватного прогнозу економічного явища. Таким чином, реалізуючи інвестиційний проект, будь-який суб'єкт стикається з ризиком, що являє собою суб'єктивно-об'єктивну категорію, яка пов'язана з прийняттям рішення в ситуації невизначеності і конфліктності. Вона відображає міру досягнення очікуваного результату і відхилення від мети з урахуванням дії контрольованих і неконтрольованих факторів. І віртуальні підприємства не є винятком.

У зв'язку з цим можна вважати, що розробка модуля оцінки інвестиційних ризиків віртуальної IT-компанії на основі машинного навчання є одним із перспективних напрямків у управлінні ризиками при формуванні інвестиційного портфелю віртуальної компанії.

Аналіз відомих рішень і постановка задачі дослідження

У роботах [1-6, 8, 9] запропоновані концептуальні моделі віртуальних підприємств та представлена інноваційна інфраструктура інформаційної системи та її мережеве рішення.

У роботі [7] вибір партнерів формується як тип нечіткої гібридної багатокритеріальної групової проблеми прийняття рішень зі ступенями нечіткої істинності порівнянь альтернатив, представлених як нечіткі числа (TrFN).

У роботі [10] описано алгоритм групового прийняття рішень на основі INPR, який можна застосувати для усунення неповних та непослідовних INPR.

У статті [11] описані результати дослідження кібернетичного управління підприємством на основі моделі та технології цифрового побратиму та концепції системи інтелектуального управління підприємством.

У статтях [12, 16-18] наведено рекомендації щодо зменшення інноваційних та фінансових ризиків віртуального підприємства.

У роботах [13-15, 20, 21] обговорено процеси управління ризиками, встановлених твердо в контексті завдання управління проектами в цілому та орієнтованого на підвищення ефективності.

У [22, 21, 24-27] досліджено стратегію управління ризиками в IT-проектах на прикладах і встановлено, що підприємства практикуючі в IT-індустрії, як правило, є суб'єктами ризику, які хочуть пройти навчання з управління ризиками.

У документі [28] представлено методологію та комп'ютерну модель, розроблену Міжнародним енергетичним агентством для кількісного оцінювання наслідків невизначеності політики зміни клімату на інвестиції в енергію за допомогою підходу ROA.

У документі [19] представлена інноваційна методологія оцінки інвестицій в ESS. Робота [29] є порівняльним дослідженням між сучасними та старими інструментами інвестування. У роботі [30] окреслюються особистості, які впливають на інвестиційні структури інвестора.

У [32] представлено систематизований огляд однокласних класифікаторів підтримки. У роботі [33] використано методи векторних машин підтримки (SVM) та нейронних мереж, де модель SVM отримала найвищі показники ефективності серед класифікаторів для кожного набору даних.

Дослідження [34] дає розуміння використання алгоритмів стохастичного градієнтного спуску для застосувань великих даних, наприклад, для прискорення SVM або регульованої регресії у великих масштабах, або для підвищення ефективності онлайн-навчання чи прогнозування в реальному часі (управління).

У роботі [35] виконано порівняння методів ансамблю (Random Forest, AdaBoost та Kernel Factory) з моделями одиночних класифікаторів (Нейронні мережі, Логістична регресія, Підтримка векторних машин та К-Найближчий сусід).

У роботі [36] представлені основні поняття стратегії багатоступеневої класифікації та обговорюються два методи проектування дерев рішень. У роботі [37] описано сценарії побудови класифікатора дерева рішень.

У роботі [38] використано статистичні методи для оцінки непараметричної щільності.

У роботі [39] пропонується нове рішення для k-найближчої класифікації сусідів на основі Spark, а в [40] описано вибір параметру K в алгоритмі k-найближчого сусіда.

У роботі [41] проаналізовано алгоритм прискорення AdaBoost, а також його пізніші версії Gentle та Real AdaBoost. Робота [42] пропонує новий алгоритм прискорення, який створює меншу помилку узагальнення порівняно із згаданими алгоритмами ціною дещо більшої помилки навчання.

У [43] описуються причини популярності логістичної моделі та її ключові особливості.

Переважає більшість розглянутих вище робіт чітко окреслює окрему прикладну область, а саме: управління ризиками, управління віртуальним підприємством, управління інвестиційними ризиками на підприємстві, управління ризиками ІТ-підприємства, використання методів машинного навчання в соціально-економічній сфері. Деякі дослідження [12,16, 19, 21, 22, 24] частково об'єднують кілька областей.

Робота [31], де застосовано методи машинного навчання для сприяння процесу штучного відбору державних облігацій, є найбільш близьким аналогом. Тут проаналізовані існуючі інвестиційні фонди, які мають вже оцінку доходів та витрат, що спрощує прийняття рішень.

У розвиток роботи [31], автори даної статті пропонують дослідити інвестиційні ризики віртуальної ІТ-компанії на основі машинного навчання. Для цього, спершу потрібно виконати аналіз методів машинного навчання (рис. 1).

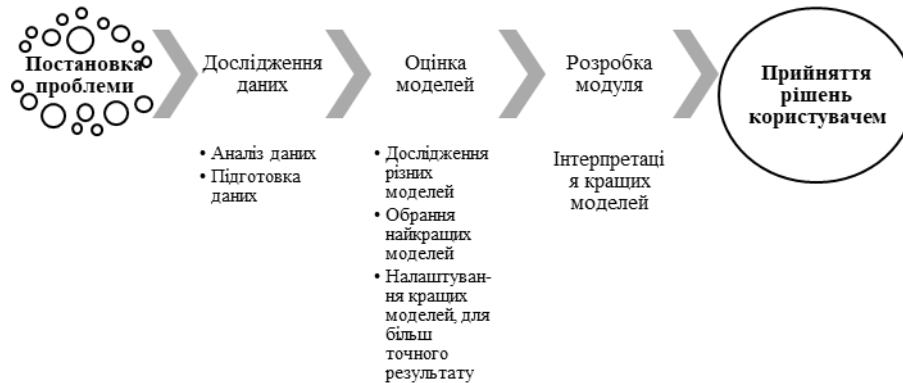


Fig. 1. Покроковий аналіз оцінки кращих моделей на основі машинного навчання

Розроблений модуль дозволяє на основі методів машинного навчання та експертної оцінки, виводити найбільш оптимальний результат оцінки інвестиційних ризиків віртуальної ІТ-компанії, на основі, якого користувач може приймати управлінські рішення стосовно вкладання коштів в нову віртуальну компанію.

Аналіз даних

Головною проблемою в оцінці ризику є аналіз та прогноз можливих збитків. Збитки оцінюються за обсягом та ймовірністю їх появи. Це стосується збитків матеріальних, фінансових, інформаційних, трудових, часових та спеціальних видів збитків.

Джерелами можливих збитків [12, 14, 18, 23-26, 29] є політичні фактори, економічна та соціальна нестабільність, темпи зростання науково-технічного прогресу, старіння товарів та послуг, недосконалість методів управління, некомпетентність кадрів, непорядність клієнтів, недосконалість кредитно-грошова та правова політика.

У кожній країні є своє законодавство відносно ризиків, узагальнити їх дуже складно, тому автори розглядають ризики відповідно до чинного законодавства України [44]: отримання дозволів і ліцензій; попит та ринок збуту; операційні ризики (постачання та вхідні матеріали; недооцінка витрат; недоліки технології); нормативно-правові ризики; фінансові та макроекономічні ризики (наявність фінансування; фінансова здатність приватного партнера; відсоткова ставка; інфляція; курс гривні до іноземної валюти); соціальні та політичні ризики, заперечення або опір з боку основних заінтересованих сторін; форс-мажорні обставини та інші непередбачені події з високим рівнем.

Якщо розглядати ризики, які стосуються відносно віртуальних підприємств то до загальної класифікації варто віднести ще:

1. Ризик упущеної фінансової вигоди. Це ризики настання непрямого (побічного) фінансового збитку (неодержаний прибуток) у результаті нездійснення якого-небудь заходу (наприклад, недосягнення планового обсягу продажів) або ж, якщо розглядати глобальний варіант – припинення господарської діяльності підприємства.

2. Ресурсний ризик. Полягає в підвищенні цін в наслідок ринкових коливань або погіршення якості. Крім того, даний ризик може виникнути у випадку зростання вартості робочої сили.

3. Ризик впровадження. Цей вид ризику пов'язаний з тим, що в ході реалізації проекту чи здійсненні стратегії підприємства не буде досягнуто кінцевих запланованих результатів: не вдається підібрати персонал необхідної кваліфікації, який би знав специфіку ринку та який би мав досвід роботи; невиконання (неякісне, несвоєчасне) виконання зобов'язань компанією, яка відповідає за надання супутніх послуг; помилки у проектуванні (не точність в розрахунках).

4. Бюрократичні та адміністративні ризики. Такі ризики виникають в результаті прийняття компетентним органом юридично значимих рішень нормативного характеру, які прямим або опосередкованим способом негативно впливають на діяльність підприємства.

5. Фінансові ризики. До даної категорії ризиків відносяться ризики, які можуть спричинити за собою можливість неповернення залученого капіталу в плановані терміни і при плановій вартості інвестицій. До них відносяться: виникнення невизначених обставин, які можуть призвести до збільшення вартості проекту; збільшення вартості позикового капіталу; зміна платоспроможності клієнтів, що може привести до неможливості вчасно і в повному обсязі розраховуватися за своїми зобов'язаннями; коливання цін на вироблену продукцію.

6. Репутаційні. Наявність офісу для проведення зустрічей з клієнтами, брендovanість, впізнаваність.

7. Робота персоналу. Відтік інформації в тому числі в наслідок використання персоналом не стабільного програмного забезпечення.

8. Стрімкий розвиток та зміни ринку технологій. Втрата актуальності продукту.

9. Акумуляування та зберігання інформації.

Враховуючи різноманітність та специфіку ризиків, які виникають у процесі реалізації інвестиційного проекту, важливою та обов'язковою складовою має бути управління проектними ризиками.

При достатньо високому ступені ризику в альтернативних стратегіях часто приймають те рішення, яке є менш ефективним, але володіє шансом на своєчасну та успішну його реалізацію.

Аналізуючи ризик використовують такі моделі: апіорні та емпіричні. Перші будуються за допомогою певних теоретичних припущень і на основі цього дають рекомендації стосовно результатів певних рішень. Інші будуються на основі узагальнення минулих спостережень (статистичної інформації).

Ризик може вимірюватися кількісно як в абсолютному, так і у відносному вираженні. У доповнення до методів вимірювання ризику за рівнем відхилень можливих результатів діяльності від середньо-очікуваних ризиків може визначатися величиною можливих збитків. Для якісної оцінки ризиків використовують 10-бальну та 5-ти бальну шкалу (табл. 1).

Таблиця 1.

Характеристика компонентів ризику					
Ступінь впливу – $Q(R)$			Ймовірність виникнення – $I(R)$		
5-ти бальна			10-ти бальна		
5	Дуже висока	80-100% інвестицій	Завжди	Кожен день	10
			Практично неминуче	Частіше, ніж 1 раз на місяць	9
4	Висока	60-79% інвестицій	Дуже часто	Частіше, ніж 1 раз в рік, і рідше, ніж 1 раз на місяць	8
			Часто	Частіше, ніж 1 раз на 2 роки, й рідше, ніж 1 раз в рік	7
3	Середня	40-59% інвестицій	Можливо	Частіше, ніж 1 раз в 3 роки, і рідше, ніж 1 раз в рік	6
			Випадковість, частіше викликана людським фактором	1 раз в 3 роки, і частіше, ніж 1 раз в 5 років	5
2	Низька	20-39% інвестицій	Час від часу	1 раз в 5 років, і частіше, ніж 1 раз в 7 років	4
			Дуже рідко	1 раз в 7 років, і рідше, ніж 1 раз в 5 років	3
1	Дуже низька	0-19% інвестицій	Практично не можливо	Рідше, ніж 1 раз в 7 років	2
			Не можливо	Рідше, ніж 1 раз в 10 років	1

Підготовка даних

Для даного дослідження проведено експертну оцінку 10 експертами 20 проектів створення віртуальних ІТ-компаній в бальній системі. Оцінка проведена на основі параметрів розглянутих вище в таблиці характеристики компонентів ризику (див. таблицю 1) з оцінками рівнів ризику

$$O(R) = Q(R) \times I(R).$$

Отримані результати сформовані в таблиці 2.

Таблиця 2.

Характеристика даних для оцінки інвестиційних ризиків віртуального ІТ-підприємства			
№	Параметр	Позначення	Тип даних
	Код компанії	ID p	object
	Код експерта	ID exp	object
	Попит та ринок збуту	DM	int64
	Постачання та вхідні матеріали	SIM	int64
	Недооцінка витрат	UC	int64
	Недоліки технології	ST	int64
	Нормативно-правові ризики	RegR	int64
	Наявність фінансування	AF	int64
	Фінансова здатність приватного партнера	FCPP	int64
	Відсоткова ставка	IR	int64

№	Параметр	Позначення	Тип даних
	Інфляція	inflation	int64
	Курс гривні до іноземної валюти	ERHFC	int64
	Форс-мажорні обставини	FM	int64
	Ризик упущеної фінансової вигоди.	RLFG	int64
	Ризик впровадження.	RSup	int64
	Не вдалося підібрати персонал необхідної кваліфікації	PS	int64
	Невиконання зобов'язань компанією	Failure	int64
	Помилки у проектуванні.	DE	int64
	Зміна платоспроможності клієнтів	CSC	int64
	Коливання цін на вироблену продукцію.	FPMP	int64
	Репутація	Reputational	int64
	Робота персоналу	IS	int64
	Втрата актуальності продукту –	LR	int64
	Акумуляування та зберігання інформації –	ISe	int64
	Ведення бухгалтерський обліку	accounting	int64

Оцінка даних

Оцінка даних буде проводитись методами машинного навчання, тому мову програмування для аналізу обрано Python, адже вона найкраще працює з аналізом даних на основі машинного навчання. Для аналізу використано наступні бібліотеки: pandas, numpy, train_test_split, SVC, GridSearchCV, SGDClassifier, RandomForestClassifier, DecisionTreeClassifier, GaussianNB, KNeighborsClassifier, AdaBoostClassifier, LogisticRegression, KFold та cross_val_score.

Перед дослідженням даних слід розділити їх на дві групи: навчальний набір 80%, тестовий набір 20%. Набір для навчання слід використовувати для побудови моделей машинного навчання. Тестовий набір слід використовувати для того, щоб побачити, наскільки добре працює побудована модель на незнайомих даних. Для визначення найкращого проекту з мінімальними ризиками додаємо параметр fit. Звичайно, алгоритми машинного навчання оперують цифровими значеннями, тому призначаємо відповідні дискретні значення 0 або 1.

Розглядаючи дані в статистичному описі (Табл. 3) та візуальному (Рис. 2) видно, що всі показники є щільними та не мають пропусків, що означає, що дані не потребують попередньої обробки та очистки відповідно.

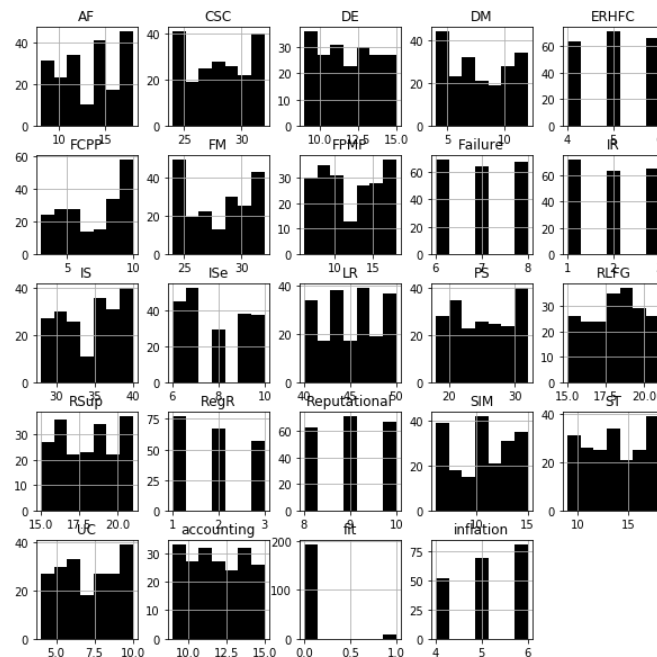


Рис. 2. Візуалізація та статистичний опис даних

У таблиці 3 представлено кількість значень, середнє значення, мінімальне та максимальне значення в наборі даних. Рядок std показує стандартне відхилення, яке вимірює, наскільки розсіяні значення – для всіх параметрів показник не значний. Для розуміння того, наскільки кожен атрибут співвідноситься з оцінкою ризиків, обчислюють стандартний коефіцієнт кореляції. З таблиці видно, що всі значення близькі до 0, що говорить про низьку залежність параметрів від ознаки fit. А ще, більшість параметрів мають негативний вплив на атрибут, що логічно, адже параметри є – ризиками, які в своїй природі мають негативний вплив.

Таблиця 2.

Статистичний опис даних						
Параметр	Кореляційне відношення до значення fit	count	mean	std	min	max
fit	1,00	201	1	0	0	1
IS	-0,23	201	34	4	28	40
DM	-0,19	201	8	3	4	12
SIM	-0,18	201	11	3	6	15
ST	-0,17	201	14	3	9	18
RegR	-0,15	201	2	1	1	3
FPMP	-0,14	201	12	4	6	18
RLFG	-0,13	201	18	1	2	30
accounting	-0,12	201	12	2	9	15
PS	-0,12	201	25	4	18	32
ISe	-0,11	201	8	1	6	10
IR	0,10	201	2	1	1	3
Failure	-0,09	201	7	1	6	8
AF	-0,08	201	13	3	8	18
Reputational	-0,07	201	9	1	8	10
LR	-0,06	201	45	3	40	50
ERHFC	0,06	201	24	1	5	20
CSC	-0,05	201	28	2	24	32
inflation	-0,04	201	5	1	4	6
RSup	0,03	201	5	1	3	6
UC	0,02	201	7	2	4	10
DE	-0,02	201	12	2	9	15
FM	-0,02	201	8	1	4	12
FCPP	-0,02	201	7	2	3	10

У зв'язку з тим, що в наборі даних немає відсутніх значень, ознак, які не дають не корисної інформації для завдання. То можна зробити висновок, що набір даних досить чистий. Таким чином, не потрібно робити будь-які складні процеси з підготовки даних.

Перехресна оцінка моделей

Оцінка моделі машинного навчання може бути досить складною. Зазвичай модель оцінюється на основі показника похибок. Цей спосіб не дуже надійний, оскільки точність, отримана для одного тестового набору, може сильно відрізнятися від точності, отриманої для іншого тестового набору.

Запорукою справедливого порівняння алгоритмів машинного навчання є забезпечення того, що кожен алгоритм оцінюється однаково на одних і тих же даних. Перехресне підтвердження в К-кратному стані (CV), що забезпечує вирішення цієї проблеми, розділивши дані на частини та забезпечити, щоб кожна частина була використана як тестовий набір у певний момент.

Перехресну перевірку проведемо на восьми класичних методах класифікації: Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression.

```
# prepare configuration for cross validation test harness
seed = 7
# prepare models
models = []
models.append(('SupportVectorClassifier', SVC()))
models.append(('StochasticGradientDecentC', SGDCClassifier()))
models.append(('RandomForestClassifier', RandomForestClassifier()))
models.append(('DecisionTreeClassifier', DecisionTreeClassifier()))
models.append(('GaussianNB', GaussianNB()))
models.append(('KNeighborsClassifier', KNeighborsClassifier()))
models.append(('AdaBoostClassifier', AdaBoostClassifier()))
models.append(('LogisticRegression', LogisticRegression()))
# evaluate each model in turn
results = []
names = []
scoring = 'accuracy'
for name, model in models:
    kfold = model_selection.KFold(n_splits=10, random_state=seed)
    cv_results = model_selection.cross_val_score(model, X_train, y_train, cv=kfold, scoring=scoring)
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
```

Із результатів перехресної оцінки (Table 4) видно, що найкращими є методи: Support Vector Classifier з прогнозованою оцінкою 0,98 та середнім квадратичним відхиленням близьким до 0,1; Random

Forest Classifier з прогнозованою оцінкою 0,98 та середнім квадратичним відхиленням близьким до 0,2; KNeighbors Classifier з прогнозованою оцінкою 0,98 та середнім квадратичним відхиленням близьким до 0,3.

Таблиця 4.

Результати перехресної оцінки

№	Метод	Оцінка прогнозу	Середнє квадратичне відхилення
1	SupportVectorClassifier	0.982	0.014
2	StochasticGradientDecentC	0.963	0.024
3	RandomForestClassifier	0.984	0.021
4	DecisionTreeClassifier	0.951	0.042
5	GaussianNB	0.962	0.023
6	KNeighborsClassifier	0.981	0.024
7	AdaBoostClassifier	0.953	0.044
8	LogisticRegression	0.984	0.027

Для кращого оцінювання цими методами варто провести їх більш детальне налаштування та інтерпретацію, що дасть можливість зробити більш точні висновки.

Дослідження кращих моделей.

Support Vector Classifier.

Головна мета Support Vector Classifier (SVC), як класифікатора – знайти рівняння, що розділяє гіперплощину [46] $w_1x_1 + w_2x_2 + \dots + w_nx_n + w_0 = 0$ в просторі R^n , яка б розділила два класи оптимальним чином. Загальний вигляд перетворення F об'єкта x в мітку класу Y : $F(x) = \text{sign}(w^Tx - b)$. Та позначимо $w = (w_1, w_2, \dots, w_n)$, $b = -w_0$. Після настройки ваг алгоритму w і b , всі об'єкти, які потрапляють по одну сторону від побудованої гіперплощини, будуть визначені як перший клас, а об'єкти, що потрапляють по іншу сторону – другий клас. Іншими словами, алгоритм максимізує зазор (Margin) між гіперплощиною і об'єктами класів, які розташовані найближче до неї. Такі об'єкти і називають опорними векторами.

Для оцінки якості виходу класифікатора використовується метрика експлуатаційної характеристики приймача (ROC), що дозволяє визначити значення класу відносно методу. Це графік помилкової позитивної швидкості навчання (вісь x) проти справжньої позитивної швидкості (вісь y) для ряду різних значень порогових значень між 0,0 та 1,0. Площа під кривою (AUC) може використовуватися як підсумок навички моделі. Функція AUC приймає як справжні результати (0,1) з тестового набору, так і прогнозовану ймовірності. AUC є в межах між 0 та 1. Чим вище показник AUC, тим якісніше діє класифікатор, при цьому значення 0,5 демонструє непридатність обраного методу класифікації.

З рисунку 3 видно, що AUC=0,91 та ROC крива набуває позитивного значення 1 при 0,1 значенні хибно позитивному, що говорить про дуже хорошу побудову алгоритму SVC на даних про інвестиційні ризики для створення віртуальної IT-компанії.

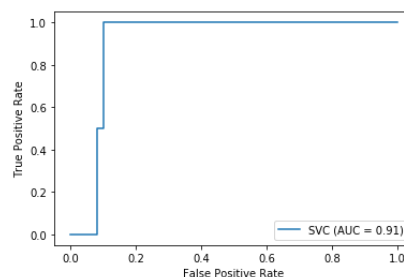


Рис. 3. Графік кривої ROC для класифікатора SVC на даних про інвестиційні ризики для створення віртуальної IT-компанії

Тому можемо проводити класифікацію на основі методу SVC при тестовій вибірці в 80 % від загального набору з параметром регуляризації рівним 1 та поліноміальним ядром з 4 ступенем та рівними вагами для всіх параметрів.

Наступним кроком є виведення текстового звіту із основними показниками класифікації та confusion matrix оцінки точності класифікації, побудованих на основі тестової вибірки.

Основними показниками класифікації є: precision, recall, f1-score, support. Точність показує, наскільки точно модель передбачає отриманий клас, відповідно до кількості раз, коли вона вірно визначила його. Повнота також називають чутливістю, вона показує, як часто вибираються релевантні класи. Підтримка для класу показує кількість контрольних екземплярів, які брали участь в обчисленні оцінок. Оцінка F1 – це середнє гармонічне з точності і повноти; цей показник є більш інформативною, ніж проста точність, тому що враховує внесок кожного класу в загальний результат.

Кожен рядок confusion matrix (Table 3.) представляє екземпляри в передбачуваному класі, а кожен стовпець представляє екземпляри фактичного класу (або навпаки).

Таблиця 4.

Confusion matrix			
		Actual class	
		P = positive	N = Negative
Predicted class	P = positive	TP = True Positive	FP = False Positive
	N = Negative	FN = False Negative	TN = True Negative

В кінці виводимо найкращий результат за допомогою бібліотеки GridSearchCV, яка дозволяє зробити пошук за вказаними значеннями параметрів для тестової вибірки.

Random Forest.

Random Forest [47, 49] – алгоритм машинного навчання, який застосовується для задач класифікації, регресії і кластеризації. Основна ідея полягає в використанні великого ансамблю вирішальних дерев, кожне з яких саме по собі дає дуже невисоку якість класифікації, але за рахунок їх великої кількості результат виходить хорошим.

Random Forest, складається з N дерев та b_n гілок по вибірці x : по заданому критерію відбираються кращі ознаки, далі проводиться розбиття в дереві по всій вибірці; дерево будується, поки в кожному аркуші буде не більше як n_{min} об'єктів або поки не буде досягнуто певної висоти дерева; при кожному розбитті спочатку вибирається m випадкових ознак з n вихідних, і оптимальний розподіл вибірки робиться пошук тільки серед них. Підсумковий класифікатор:

$$a(x) = \frac{1}{N} \sum_{i=1}^N b_{i(x)}$$

для завдання класифікації ми вибираємо рішення за більшістю, а в задачі регресії за середнім.

При побудові алгоритму класифікації на основі методу Random Forest найоптимальніше використовувати 80% даних для навчальної вибірки з рандомним визначенням з всієї вибірки. Кількість дерев для аналізу в лісі рівне = 320. При візуалізації дерева виводяться показники: feature names – зі списку ознак; node_ids – ідентифікаційний номер на кожному вузлі; criterion: gini – показує оцінку якості класифікації, він досягає свого мінімуму (нуля), коли всі випадки у вузлі потрапляють в одну цільову категорію; samples; value; class_names – назви кожного із цільових класів у порядку зростання.

В завершенні методу отримаємо найкращий результат для тестової вибірки на основі методу Random Forest.

k Nearest Neighbor.

k Nearest Neighbor [48] – це один з найпростіших алгоритмів класифікації, також іноді використовується в задачах регресії. Для класифікації кожного з об'єктів тестової вибірки необхідно послідовно виконати наступні операції: обчислити відстань до кожного з об'єктів навчальної вибірки; відібрати k об'єктів навчальної вибірки, відстань до яких мінімальна; віднести кожну ознаку до окремого класу, який найчастіше трапляється серед k найближчих сусідів.

Для визначення найоптимальнішого k для методу k Nearest Neighbor, потрібно провести шведке тестування (Fig. 4), яке покаже на якому значенні k тестова оцінка стабілізується. Тестування проводиться в межах від 1 до 9 значень k . Для оцінки інвестиційного ризиків для створення віртуальної IT-компанії $k=2$.

Наступним кроком є побудова алгоритму Nearest Neighbor використовуючи 80% даних всієї вибірки. Ваги прописуємо за методом 'distance': у цьому випадку ближчі сусіди матимуть більший вплив, ніж сусіди, які знаходяться далі. Алгоритм для обчислення найближчих сусідів: 'auto' що визначить найбільш відповідний алгоритм на основі значень fit .

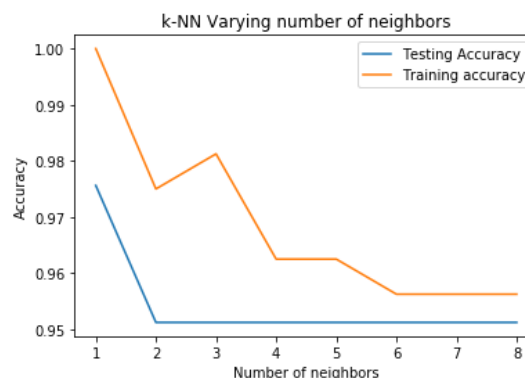


Рис. 4. Значення k для KNN

Візуалізація та обговорення результатів

Діаграма послідовностей (рис. 6) показує обмін повідомленнями (тобто виклики методів) між різними об'єктами в специфічній, обмеженої за часом, ситуації. Діаграма послідовностей спеціально виділяє порядок і часи відсилання повідомлень об'єктам. У діаграмах послідовностей об'єкти представляються вертикальними переривистими лініями з ім'ям об'єкта нагорі. Таким чином, повідомлення, посилаємо від одного об'єкта до іншого, відображаються стрілками з зазначенням операції і параметрів.

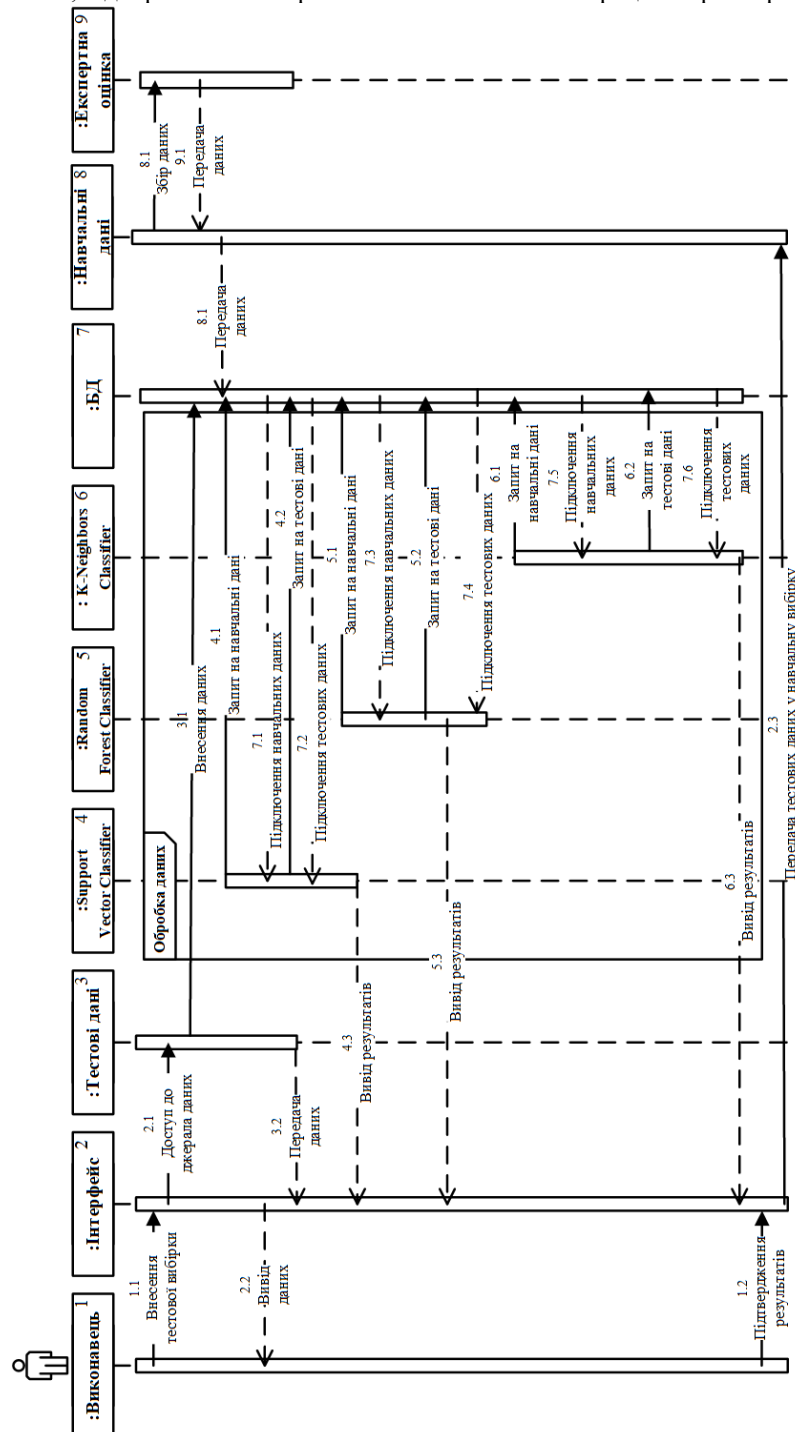


Рис. 5. Діаграма послідовності UML MLInvestRisks

На першому етапі виконавець (блок 1) вносить (1.1) тестові дані в інтерфейс MLInvestRisks (блок 2), це дані проєктів для створення віртуальної ІТ-компанії, що його цікавлять з відповідною оцінкою ризиків. Дані також виводяться назад (3.2) в інтерфейс (блок 2), щоб виконавець (блок 1) міг їх бачити.

Дані переносяться (2.1) у базу (блок 3) з тестовою вибіркою та вносяться (3.1) у загальну базу даних (блок 7). Далі дані поступають (8.2) у загальну базу даних (блок 7) з навчальних даних (блок 8). Навчальні дані формуються (8.1, 9.1) на основі експертної оцінки (блок 9).

На наступному етапі проводиться обробка даних за трьома методами, що показали найкраще переєхресне оцінювання – Support Vector Classifier (блок 4), Random Forest Classifier (блок 5), KNeighborsClassifier (блок 6). Кожен блок (блок 4-6) подають запит на навчальну вибірку (4.1, 5.1, 6.1) до загальної бази даних (блок 7) та отримується підключення до неї (7.1, 7.3, 7.5) (див. рис 6). Після чого проводиться моделювання на основі навчальної вибірки.

Далі проводиться моделювання на основі тестової вибірки, для чого спершу проводиться запит в загальну базу даних (блок 7) на тестові дані (4.2, 5.2, 6.2) для кожного методу окремо та відповідно отримується підключення до даних (7.2, 7.4, 7.6).

Отримані результати для тестової вибірки виводяться (4.3, 5.3, 6.3) у інтерфейс (блок 2). Після чого виконавець (блок 1) переглядає отримані результати та на основі внесених даних (2.2) підтверджує результат (1.2) з відповідним рішенням чи система правильно вибрала найкращий результат оцінки інвестиційних ризиків для віртуальної ІТ-компанії чи ні. Ці дані автоматично передаються (2.3) у навчальні дані (блок 8), для подальшого роботи системи) (див. рис 6).

Після внесення нової тестової вибірки в MLInvestRisks отримуємо наступні результати для методу Support Vector Classifier (рис. 7).

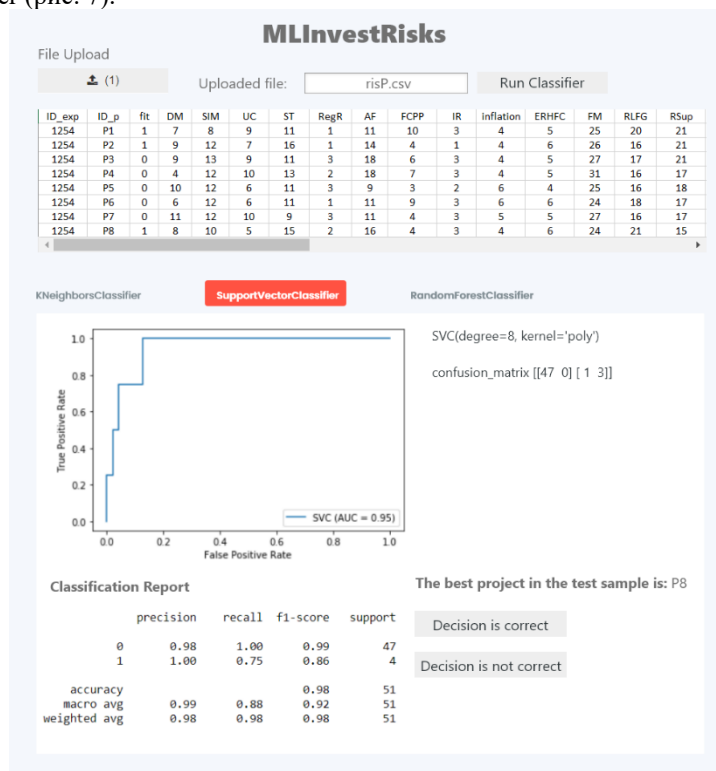


Рис. 6. Вікно результатів Support Vector Classifier в MLInvestRisks

У confusion matrix Support Vector Classifier визначено 47 проєктів, в яких ризики найоптимальніші, також алгоритм передбачив, що три – проєкти найбільш вигідні. З основних показників класифікації оцінки f1, при визначенні проєктів, що не підходять для інвестування алгоритм визначає з точністю 0,99 та проєкти в які варто інвестувати з точністю 0,86. Загальна оцінка точності моделі рівна 0,98, що є достатньо добрим результатом. Найбільш привабливим інвестиційним проєктом для створення віртуальної ІТ-компанії, система на основі Support Vector Classifier, обрала проєкт з кодом – P8.

З результатів (рис. 8) Random Forest Classifier середня абсолютна похибка моделювання рівна 0,05,. Оцінка моделювання = 0,98 – що є хорошим результатом. З дерева видно, що результати вірні при IR <= 0.7, ST <= -1.2, RLFG <= 0.8, а це 3% від всіх даних. Найкращим проєктом відносно рівня ризиків система на основі Random Forest Classifier обрала також проєкт з кодом – P8, як і за методом Support Vector Classifier.

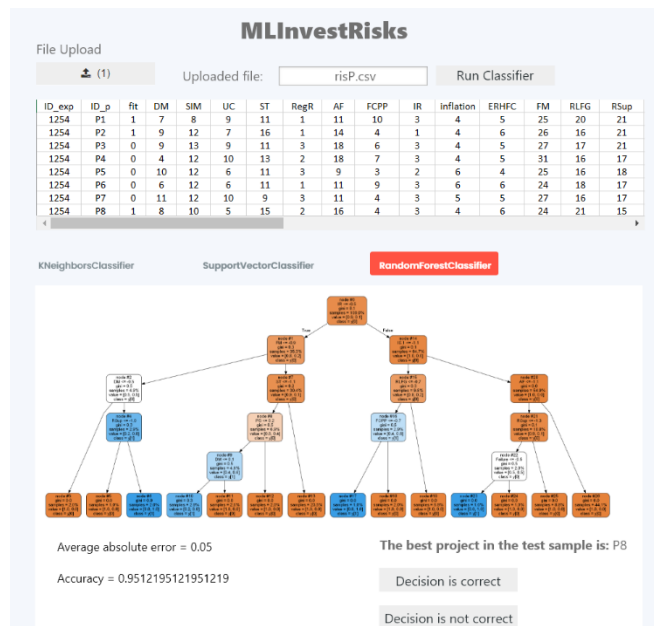


Рис. 7. Вікно результатів Random Forest Classifier в MLInvestRisks

У confusion matrix KNeighborsClassifier (рис.9) визначено 39 проектів створення віртуальної ІТ-компанії, в яких ризики найоптимальніші, також алгоритм передбачив, що один – проект найбільш вигідний. З основних показників класифікації видно, що по оцінці f1, при визначенні проекти, що не підходять для інвестування алгоритм визначає з точністю 0,99 та проекти в які варто інвестувати з точністю 0,67. Загальна оцінка точності моделі рівна 0,98, що є дуже хорошим результатом. Система на основі KNeighborsClassifier, обрала проект з кодом – P1.

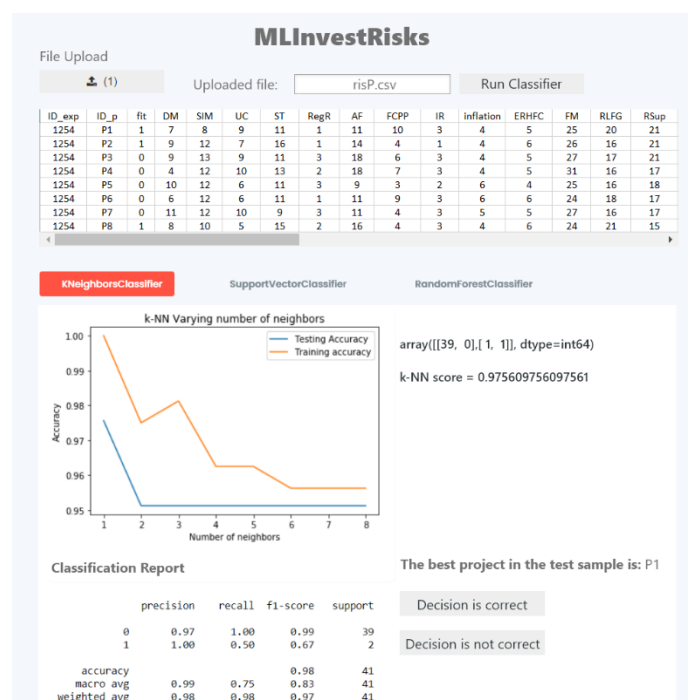


Рис. 8. Вікно результатів KNeighborsClassifier в MLInvestRisks

Найбільш привабливим інвестиційним проектом для створення віртуальної ІТ-компанії, система на основі Support Vector Classifier та Random Forest Classifier обрала проект з кодом – P8. Якщо розглянути даний проект з точки зору оцінки експертів то він потрапляє у категорію ризиків з середнім впливом на проект та з випадковістю, що частіше викликана людським фактором та проявляється 1 раз в 3 роки. На основі KNeighborsClassifier, системи обрала проект з кодом – P1. Він потрапляє у категорію ризиків з середнім впливом на проект та з можливою випадковістю ризиків та частотою: частіше, ніж 1 раз в 3 роки, і

рідше, ніж 1 раз в рік. Автори вважають що обидва проекти створення віртуальної ІТ-компанії – можна інвестувати, проте Р8 менш ризиковий. Остаточне рішення приймається користувачем-інвестором.

Висновки та перспективи подальшого розвитку у даному напрямі

У статті проведено перехресне оцінювання інвестиційних ризиків для створення віртуальної ІТ-компанії на основі восьми методів класифікації з застосуванням машинного навчання, а саме: Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression. З результатів оцінювання отримано, що кращими методами є, Support Vector Classifier, Random Forest Classifier, K-Neighbors Classifier з оцінками точності прогнозування на рівні 0,98.

Проведено більш детальне уточнення параметрів найкращих методів (Support Vector Classifier, Random Forest Classifier, K-Neighbors Classifier), що дало можливість частково підвищити точність та зменшити похибки для цих методів в середньому на 1-2%. Також проведена інтерпретація результатів для даних тестової вибірки.

Розроблено програмне середовище, яке дозволяє на основі Support Vector Classifier, Random Forest Classifier, K-Neighbors Classifier формувати рішення для потенційних інвесторів. Результати проведених експериментів у розробленій інформаційній системі підтвердили ефективність використання методів, що обрані в перехресній оцінці. Також дана система дає можливість поновлювати навчальну вибірку для подальшої роботи системи. Найбільш привабливим інвестиційним проектом для створення віртуальної ІТ-компанії, система на основі Support Vector Classifier та Random Forest Classifier обрала проект з кодом – Р8 та на основі KNeighborsClassifier – Р1.

Література

1. Esposito, E. Investigating virtual enterprise models: literature review and empirical findings / E. Esposito, P. Evangelista // International Journal of Production Economics. – 2014. Vol. – 148. – P. 145–157. doi:10.1016/j.ijpe.2013.10.003.
2. Martinez, M. Virtual enterprise – organisation, evolution and control / M. Martinez, P. Fouletier, K. Park, J. Favrel // International Journal of Production Economics. – 2001. – Vol. 74. – No. 1-3. – P. 225–238. doi:10.1016/s0925-5273(01)00129-3.
3. Park, K. H. Virtual enterprise – Information system and networking solution / K. H. Park, J. Favrel // Computers & Industrial Engineering. – 1999. – Vol. 37. – No. 1-2. – P. 441–444. doi:10.1016/s0360-8352(99)00113-8.
4. Gou, H. A framework for virtual enterprise operation management / H. Gou, B. Huang, W. Liu, X. Li // Computers in Industry. – 2003. – Vol. 50. – No. 3. – P. 333–352. doi:10.1016/s0166-3615(03)00021-6.
5. Sadigh, B. L. An ontology-based multi-agent virtual enterprise system (OMAVE): part 1: domain modelling and rule management / B. L. Sadigh, H. O. Unver, S. Nikghadam, E. Dogdu, A. M. Ozbayoglu, S. E. Kilic // International Journal of Computer Integrated Manufacturing. – 2016. – Vol. 30. – No. 2-3. – P. 320–343. doi:10.1080/0951192x.2016.1145811.
6. Sadigh, B. L. An ontology-based multi-agent virtual enterprise system (OMAVE): part 2: partner selection / B. L. Sadigh, S. Nikghadam, A. M. Ozbayoglu, H. O. Unver, E. Dogdu, S. E. Kilic // International Journal of Computer Integrated Manufacturing. – 2017. – Vol. 30. – No. 10. – P. 1072–1092. doi:10.1080/0951192x.2017.1285424.
7. Dong, J.-Y. Virtual enterprise partner selection integrating LINMAP and TOPSIS / J.-Y. Dong, S.-P. Wan // Journal of the Operational Research Society. – 2016. – Vol. 67. – No. 10. – P. 1288–1308. doi:10.1057/jors.2016.22.
8. Yuan, M. Q. SI system housing virtual enterprise partners selection based on vector angle cosine / M. Q. Yuan, Z. F. Li, L. Li // J. Civ. Eng. Manag. – 2018. – Vol. 35. – P. 117–122.
9. Ferreira, L. Virtual Enterprise integration management based on a Meta-enterprise – a PMBoK approach / L. Ferreira, N. Lopes, P. S. Ávila, H. Castro, M. L. R. Varela, G. D. Putnik, M. M. Cruz-Cunha // Procedia Computer Science. – 2017. – Vol. 121. – P. 1112–1118. doi: 10.1016/j.procs.2017.12.120.
10. Meng, F. Interval neutrosophic preference relations and their application in virtual enterprise partner selection / F. Meng, N. Wang, Y. Xu // Journal of Ambient Intelligence and Humanized Computing. – 2019. – Vol. 10. – P. 5007–5036. doi:10.1007/s12652-019-01178-5.
11. Tsvetkov, V. Y. Digital Enterprise Management in Cyberspace / V. Y. Tsvetkov, S. V. Shaytura, N. L. Sulstaeva // In: Proceedings of the 2nd International Scientific and Practical Conference “Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth” (MTDE’2020), Atlantis Press, 2020, pp. 361–365. doi: 10.2991/aebmr.k.200502.059.

12. Apatova, N. Mechanisms and Resources of Virtual Enterprise Risk Management / N. Apatova // In: Proceedings of the International Scientific Conference "Far East Con" (ISCFEC 2020), Atlantis Press, 2020, pp. 834-844. doi: 10.2991/aebmr.k.200312.116.
13. Chapman, C. Project risk management: processes, techniques and insights / C. Chapman, S. Ward // John Wiley, 1996. ISBN 0-470-85355-7.
14. Lytvyn, V. The risk management modelling in multi project environment / V. Lytvyn, V. Vysotska, O. Veres, I. Rishnyak, H. Rishnyak // In: Proceedings of the 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), 2017, pp. 32-35. doi: 10.1109/stc-csit.2017.8098730.
15. Marodin, G. A. Model of risk interactions hindering lean production implementation / G. A. Marodin, T. A. Saurin, G. L. Tortorella, D. C. Fettermann // Gestao & Producao. – 2018. – Vol. 25. – No. 4. – P. 696-712.
16. Wang, N. Research on Virtual Enterprise Risk Control Based on Optimization / N. Wang // Advanced Materials Research. 2010. – No. 129-131. – P. 1267–1272. doi: 10.4028/www.scientific.net/amr.129-131.1267.
17. Lu, F. Virtual Enterprise risk management under asymmetric information / F. Lu, H. Bi, M. Huang, X. Wang // In: Proceedings of the 2013 10th International Conference on Service Systems and Service Management, 2013, pp. 202-207. doi: 10.1109/icsssm.2013.6602655.
18. Wan, J. Research on risk factors of entrepreneurship in Internet industry with the grounded theory / J. Wan, Q. Jiang, L. Xie // In: Proceedings of the Wuhan International Conference on e-Business, Association for Information Systems, 2017, pp. 99-107.
19. Locatelli, G. Investment and risk appraisal in energy storage systems: A real options approach / G. Locatelli, D. C. Invernizzi, M. Mancini // Energy. – 2016. – Vol. 104. – P. 114–131. doi:10.1016/j.energy.2016.03.098
20. Ahmad, Z. Building information modeling as a risk transformer: An evolutionary insight into the project uncertainty / Z. Ahmad, M. J. Thaheem, A. Maqsoom // Automation in Construction. – 2018. – Vol. 92. – P. 103–119. doi: 10.1016/j.autcon.2018.03.032.
21. Javani, B. Risk management in IT projects – a case of the South African public sector / B. Javani, P. M. D. Rwelamila // International Journal of Managing Projects in Business. – 2016. – Vol. 9. – No. 2. – P. 389–413. doi: 10.1108/ijmpb-07-2015-0055.
22. Cabral, J. S. Project Risk Management Strategies for IT Project Managers / J. S. Cabral // PhD thesis, Walden University, 2017.
23. ChePa, N. A review on risk mitigation of IT governance / N. ChePa, B. A. Jnr, R. N. H. Nor, M. A. A. Murad // Information Technology Journal. – 2015. – Vol. 14. – No. 1. – P. 1-9. doi: 10.3923/itj.2015.1.9.
24. Kumsuprom, S. ICT risk management in organizations: Case studies in Thai business / S. Kumsuprom, B. Corbitt, S. Pittayachawan // In: Proceedings of the Australasian Conference on Information Systems ACIS 2008, Christchurch, New Zealand, 2008, paper 98. doi: 10.13140/2.1.2489.1689.
25. Hwang, B.-G. Sustainable risk management in the construction industry: lessons learned from the IT industry / B.-G. Hwang, M. Chen // Technological and Economic Development of Economy. – 2015. – Vol. 21. – No. 2. – P. 216–231. doi:10.3846/20294913.2014.979455.
26. Pa, N. C. A model of mitigating risk for IT organisations / N. C. Pa, B. Anthony // In: Proceedings of the IEEE 2015 4th International Conference on Software Engineering and Computer Systems (ICSECS), Kuantan, 2015, pp. 49-54. doi: 10.1109/ICSECS.2015.7333082.
27. Anthony Jnr, B. The Development and Initial Results of a Component Model for Risk Mitigation in IT Governance / B. Anthony Jnr, N. Che Pa, R. Haizan Nor, Y. Yah Josoh // Journal of Science, Technology and Innovation Policy. – 2017. – Vol. 2. – No. 2. – P. 1-13. Retrieved from <http://jostip.org/index.php/jostip/article/view/39/16>
28. Yang, M. Modeling investment risks and uncertainties with real options approach / M. Yang, W. Blyth // International Energy Agency Working Paper Series. Paris. – 2007. – Paper Number LTO/2007/WP01.
29. Swami, R. Managing Investment Risks: Modern vs Traditional Knowledge and Practices / R. Swami, T. K. Jain // 2019. Electronic copy available at: <https://ssrn.com/abstract=3310062>. doi: 10.2139/ssrn.3310062
30. Chen, T.-H. Investigating the Predictive Power of Investor Personality in Forecasting Investment Performance using machine learning models / T.-H. Chen, R.-J. Ho, Y.-W. Liu // Computers in Human Behavior. – 2018. – Vol. 101. – P. 409-416. doi: 10.1016/j.chb.2018.09.027.
31. Tao, T. Classification of Mutual Fund Investment Types with Advanced Machine Learning Models / T. Tao, K. Yan, S. Yang // In: Proceedings of the 2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD), 2019, pp. 84-89.
32. Alam, S. One-class support vector classifiers: A survey / S. Alam, S. K. Sonbhadra, S. Agarwal, P. Nagabhushan // Knowledge-Based Systems. – 2020. – Vol. 196. – Paper 105754. doi: 10.1016/j.knosys.2020.105754.

33. Costa, N. L. Using Support Vector Machines and neural networks to classify Merlot wines from South America / N. L. Costa, L. A. G. Llobodanin, I. A. Castro, R. Barbosa // *Information Processing in Agriculture*. – 2018. – Vol. 6. – No. 2. – P. 265-278. doi: 10.1016/j.inpa.2018.10.003.
34. He, W. To regularize or not: Revisiting SGD with simple algorithms and experimental studies / W. He, Y. Liu // *Expert Systems with Applications*. – 2018. – Vol. 112. – P. 1–14. doi: 10.1016/j.eswa.2018.06.026.
35. Ballings, M. Evaluating multiple classifiers for stock price direction prediction / M. Ballings, D. Van den Poel, N. Hespeels, R. Gryp // *Expert Systems with Applications*. – 2015. – Vol. 42. – No. 20. – P. 7046–7056. doi: 10.1016/j.eswa.2015.05.013.
36. Swain, P. H. The decision tree classifier: Design and potential / P. H. Swain, H. Hauska // *IEEE Transactions on Geoscience Electronics*. – 1977. – Vol. 15. – No. 3. – P. 142–147. doi: 10.1109/tge.1977.6498972.
37. Du, W. Building decision tree classifier on private data / W. Du, Z. Zhan // In: *Proceedings of the IEEE International Conference on Privacy, Security and Data Mining CRPIT'14*, 2002, vol. 14, pp. 1–8.
38. John, G. H. Estimating continuous distributions in Bayesian classifiers / G. H. John, P. Langley // *arXiv preprint arXiv:1302.4964*, 2013.
39. Maillo, J. kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data / J. Maillo, S. Ramirez, I. Triguero, F. Herrera // *Knowledge-Based Systems*. – 2017. – Vol. 117. – P. 3–15. doi: 10.1016/j.knosys.2016.06.012.
40. Hassanat, A. B. Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach / A. B. Hassanat, M. A. Abbadi, G. A. Altarawneh, A. A. Alhasanat // *arXiv preprint arXiv:1409.0919*, 2014.
41. Schapire, R. E. Explaining AdaBoost / R. E. Schapire // In: Schölkopf B., Luo Z., Vovk V. (eds) *Empirical Inference*. Springer, Berlin, Heidelberg, 2013, pp. 37–52. doi: 10.1007/978-3-642-41136-6_5
42. Vezhnevets, A. Modest AdaBoost-teaching AdaBoost to generalize better / A. Vezhnevets, V. Vezhnevets // *Graphicon*. – 2005. – Vol. 12. – No. 5. – P. 987-997.
43. Kleinbaum, D. G. Introduction to Logistic Regression / D. G. Kleinbaum, M. Klein // In: *Logistic Regression. Statistics for Biology and Health*. Springer, New York, NY, 2010, pp. 1–39. doi: 10.1007/978-1-4419-1742-3_1.
44. Про затвердження Методики ідентифікації ризиків державно-приватного партнерства, їх оцінки та визначення форми управління ними. Постанова Кабінету Міністрів України. – 2011. – № 232.
45. Davidow, W. H. The virtual corporation: Structuring and revitalizing the corporation for the 21st century / W. H. Davidow // 1st edition, Harpercollins. – 1992. ISBN-13: 978-0887305931.
46. Blumer, A. Learnability and the Vapnik-Chervonenkis dimension / A. Blumer, A. Ehrenfeucht, D. Haussler, M. K. Warmuth // *Journal of the ACM*. – 1989. – Vol. 36. – No. 4. – P. 929–965. doi: 10.1145/76359.76371.
47. Breiman, L. Random Forests / L. Breiman // *Machine Learning*. – 2001. – Vol. 45. – No. 1. – P. 5–32. doi: 10.1023/a:1010933404324
48. Dudani, S. A. The Distance-Weighted k-Nearest-Neighbor Rule / S. A. Dudani // *IEEE Transactions on Systems, Man, and Cybernetics*. – 1976. – Vol. 6. – No. 4. – P. 325–327. doi: 10.1109/tsmc.1976.5408784.
49. Lipyana, H. Decision Tree Based Targeting Model of Customer Interaction with Business Page / H. Lipyana, A. Sachenko, T. Lendyuk, S. Nadvynychny, S. Grodskyi // In: *Proceedings of the third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020)*, CEUR Workshop Proceedings, 2020, vol. 2608, pp. 1001-1012. Electronic copy available at: <http://ceur-ws.org/Vol-2608/paper75.pdf>.

References

1. E. Esposito, P. Evangelista. Investigating virtual enterprise models: literature review and empirical findings. *International Journal of Production Economics*, 148, 145–157, 2014. doi:10.1016/j.ijpe.2013.10.003.
2. M. Martinez, P. Fouletier, K. Park, J. Favrel. Virtual enterprise – organisation, evolution and control. *International Journal of Production Economics*, 74 (1-3), 225–238, 2001. doi:10.1016/s0925-5273(01)00129-3.
3. K. H. Park, J. Favrel. Virtual enterprise – Information system and networking solution. *Computers & Industrial Engineering*, 37 (1-2), 441–444, 1999. doi:10.1016/s0360-8352(99)00113-8.
4. H. Gou, B. Huang, W. Liu, X. Li. A framework for virtual enterprise operation management. *Computers in Industry*, 50(3), 333–352, 2003. doi:10.1016/s0166-3615(03)00021-6.
5. B. L. Sadigh, H. O. Unver, S. Nikghadam, E. Dogdu, A. M. Ozbayoglu, S. E. Kilic. An ontology-based multi-agent virtual enterprise system (OMAVE): part 1: domain modelling and rule management. *International Journal of Computer Integrated Manufacturing*, 30 (2-3), 320–343, 2016. doi:10.1080/0951192x.2016.1145811.
6. B. L. Sadigh, S. Nikghadam, A. M. Ozbayoglu, H. O. Unver, E. Dogdu, S. E. Kilic. An ontology-based multi-agent virtual enterprise system (OMAVE): part 2: partner selection. *International Journal of Computer Integrated Manufacturing*, 30 (10), 1072–1092, 2017. doi:10.1080/0951192x.2017.1285424.
7. J.-Y. Dong, S.-P. Wan. Virtual enterprise partner selection integrating LINMAP and TOPSIS. *Journal of the Operational Research Society*, 67 (10), 1288–1308, 2016. doi:10.1057/jors.2016.22.

8. M. Q. Yuan, Z. F. Li, L. Li. SI system housing virtual enterprise partners selection based on vector angle cosine. *J. Civ. Eng. Manag.*, 35, 117-122, 2018.
9. L. Ferreira, N. Lopes, P. S. Ávila, H. Castro, M. L. R. Varela, G. D. Putnik, M. M. Cruz-Cunha. Virtual Enterprise integration management based on a Meta-enterprise – a PMBoK approach. *Procedia Computer Science*, (121), 1112–1118 (2017). doi: 10.1016/j.procs.2017.12.120.
10. F. Meng, N. Wang, Y. Xu. Interval neutrosophic preference relations and their application in virtual enterprise partner selection. *Journal of Ambient Intelligence and Humanized Computing*, 10, 5007–5036, 2019. doi:10.1007/s12652-019-01178-5.
11. V. Y. Tsvetkov, S. V. Shaytura, N. L. Sultaeva. Digital Enterprise Management in Cyberspace. In: *Proceedings of the 2nd International Scientific and Practical Conference “Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth” (MTDE 2020)*, Atlantis Press, 2020, pp. 361-365. doi: 10.2991/aebmr.k.200502.059.
12. N. Apatova. Mechanisms and Resources of Virtual Enterprise Risk Management. In: *Proceedings of the International Scientific Conference “Far East Con” (ISCFC 2020)*, Atlantis Press, 2020, pp. 834-844. doi: 10.2991/aebmr.k.200312.116.
13. C. Chapman, S. Ward. *Project risk management: processes, techniques and insights*. John Wiley, 1996. ISBN 0-470-85355-7.
14. V. Lytvyn, V. Vysotska, O. Veres, I. Rishnyak, H. Rishnyak. The risk management modelling in multi project environment. In: *Proceedings of the 2017 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, 2017, pp. 32-35. doi: 10.1109/stc-csit.2017.8098730.
15. G. A. Marodin, T. A. Saurin, G. L. Tortorella, D. C. Fettermann. Model of risk interactions hindering lean production implementation. *Gestao & Producao*, 25(4), 696-712, 2018.
16. N. Wang. Research on Virtual Enterprise Risk Control Based on Optimization. *Advanced Materials Research*, 129-131, 1267–1272, 2010. doi: 10.4028/www.scientific.net/amr.129-131.1267.
17. F. Lu, H. Bi, M. Huang, X. Wang. Virtual Enterprise risk management under asymmetric information. In: *Proceedings of the 2013 10th International Conference on Service Systems and Service Management*, 2013, pp. 202-207. doi: 10.1109/icsssm.2013.6602655.
18. J. Wan, Q. Jiang, L. Xie. Research on risk factors of entrepreneurship in Internet industry with the grounded theory. In: *Proceedings of the Wuhan International Conference on e-Business*, Association for Information Systems, 2017, pp. 99-107.
19. G. Locatelli, D. C. Invernizzi, M. Mancini. Investment and risk appraisal in energy storage systems: A real options approach. *Energy*, 104, 114–131, 2016. doi:10.1016/j.energy.2016.03.098.
20. Z. Ahmad, M. J. Thaheem, A. Maqsoom. Building information modeling as a risk transformer: An evolutionary insight into the project uncertainty. *Automation in Construction*, 92, 103–119, 2018. doi: 10.1016/j.autcon.2018.03.032.
21. B. Javani, P. M. D. Rwelamila. Risk management in IT projects – a case of the South African public sector. *International Journal of Managing Projects in Business*, 9(2), 389–413, 2016. doi: 10.1108/ijmpb-07-2015-0055.
22. J. S. Cabral. *Project Risk Management Strategies for IT Project Managers*. PhD thesis, Walden University, 2017.
23. N. ChePa, B. A. Jnr, R. N. H. Nor, M. A. A. Murad. A review on risk mitigation of IT governance. *Information Technology Journal*, 14(1), 1-9, 2015. doi: 10.3923/itj.2015.1.9.
24. S. Kumsuprom, B. Corbitt, S. Pittayachawan. ICT risk management in organizations: Case studies in Thai business. In: *Proceedings of the Australasian Conference on Information Systems ACIS 2008*, Christchurch, New Zealand, 2008, paper 98. doi: 10.13140/2.1.2489.1689.
25. B.-G. Hwang, M. Chen. Sustainable risk management in the construction industry: lessons learned from the IT industry. *Technological and Economic Development of Economy*, 21(2), 216–231, 2015. doi:10.3846/20294913.2014.979455.
26. N. C. Pa, B. Anthony. A model of mitigating risk for IT organisations. In: *Proceedings of the IEEE 2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*, Kuantan, 2015, pp. 49-54. doi: 10.1109/ICSECS.2015.7333082.
27. B. Anthony Jnr, N. Che Pa, R. Haizan Nor, Y. Yah Josoh. The Development and Initial Results of a Component Model for Risk Mitigation in IT Governance. *Journal of Science, Technology and Innovation Policy*, 2(2), 1-13, 2017. Retrieved from <http://jostip.org/index.php/jostip/article/view/39/16>.
28. M. Yang, W. Blyth. Modeling investment risks and uncertainties with real options approach. *International Energy Agency Working Paper Series*. Paris, 2007, Paper Number LTO/2007/WP01.
29. R. Swami, T. K. Jain. Managing Investment Risks: Modern vs Traditional Knowledge and Practices. Electronic copy available at: <https://ssrn.com/abstract=3310062>. doi: 10.2139/ssrn.3310062.
30. T.-H. Chen, R.-J. Ho, Y.-W. Liu. Investigating the Predictive Power of Investor Personality in Forecasting Investment Performance using machine learning models. *Computers in Human Behavior*, 101, 409-416, 2018. doi: 10.1016/j.chb.2018.09.027.
31. T. Tao, K. Yan, S. Yang. Classification of Mutual Fund Investment Types with Advanced Machine Learning Models. In: *Proceedings of the 2019 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD)*, 2019, 84-89.
32. S. Alam, S. K. Sonbhadra, S. Agarwal, P. Nagabhushan. One-class support vector classifiers: A survey. *Knowledge-Based Systems*, 196, 105754, 2020. doi: 10.1016/j.knsys.2020.105754.
33. N. L. Costa, L. A. G. Llobodanin, I. A. Castro, R. Barbosa. Using Support Vector Machines and neural networks to classify Merlot wines from South America. *Information Processing in Agriculture*, 6 (2), 265-278, 2018. doi: 10.1016/j.inpa.2018.10.003.
34. W. He, Y. Liu. To regularize or not: Revisiting SGD with simple algorithms and experimental studies. *Expert Systems with Applications*, 112, 1–14, 2018. doi: 10.1016/j.eswa.2018.06.026.
35. M. Ballings, D. Van den Poel, N. Hespeels, R. Gryp. Evaluating multiple classifiers for stock price direction prediction. *Expert Systems with Applications*, 42(20), 7046–7056, 2015. doi: 10.1016/j.eswa.2015.05.013.
36. P. H. Swain, H. Hauska. The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics*, 15(3), 142–147, 1977. doi: 10.1109/tge.1977.6498972.
37. W. Du, Z. Zhan. Building decision tree classifier on private data. In: *Proceedings of the IEEE International Conference on Privacy, Security and Data Mining CRPIT'14*, 2002, vol. 14, pp. 1–8 (2002).
38. G. H. John, P. Langley. Estimating continuous distributions in Bayesian classifiers. *arXiv preprint arXiv:1302.4964*, 2013.
39. J. Maillio, S. Ramirez, I. Triguero, F. Herrera. kNN-IS: An Iterative Spark-based design of the k-Nearest Neighbors classifier for big data. *Knowledge-Based Systems*, 117, 3–15, 2017. doi: 10.1016/j.knsys.2016.06.012.
40. A. B. Hassanat, M. A. Abbadi, G. A. Altarawneh, A. A. Alhasanat. Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach, *arXiv preprint arXiv:1409.0919*, 2014.
41. R. E. Schapire. Explaining AdaBoost. In: *Schölkopf B., Luo Z., Vovk V. (eds) Empirical Inference*. Springer, Berlin, Heidelberg, 2013, pp. 37–52. doi: 10.1007/978-3-642-41136-6_5
42. A. Vezhnevets, V. Vezhnevets. Modest AdaBoost-teaching AdaBoost to generalize better. *Graphicon*, 12 (5), 987-997, 2005.

-
43. D. G. Kleinbaum, M. Klein. Introduction to Logistic Regression. In: Logistic Regression. Statistics for Biology and Health. Springer, New York, NY, 2010, pp. 1–39. doi: 10.1007/978-1-4419-1742-3_1.
44. About approval of the methodology for identifying risks of public-private partnership, their assessment and determination of the form of their management. Resolution of the Cabinet of Ministers of Ukraine, #232, 2011. (in Ukrainian)
45. W. H. Davidow. The virtual corporation: Structuring and revitalizing the corporation for the 21st century. 1st edition, Harpercollins, 1992. ISBN-13: 978-0887305931.
46. A. Blumer, A. Ehrenfeucht, D. Haussler, M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. Journal of the ACM, 36(4), 929–965, 1989. doi: 10.1145/76359.76371.
47. L. Breiman. Random Forests. Machine Learning, 45 (1), 5–32, 2001. doi: 10.1023/a:1010933404324
48. S. A. Dudani. The Distance-Weighted k-Nearest-Neighbor Rule. IEEE Transactions on Systems, Man, and Cybernetics, 6(4), 325–327, 1976. doi: 10.1109/tsmc.1976.5408784.
49. H. Lipyana, A. Sachenko, T. Lendyuk, S. Nadvynychny, S. Grodskyi. Decision Tree Based Targeting Model of Customer Interaction with Business Page. In: Proceedings of the third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), CEUR Workshop Proceedings, 2020, vol. 2608, pp. 1001-1012. Electronic copy available at: <http://ceur-ws.org/Vol-2608/paper75.pdf>.