Національний технічний університет України
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

ММСА

Кафедра математичних
методів системного аналізу

Інтелектуальний аналіз даних

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	***Перший (бакалаврський)***
Галузь знань	*12 Інформаційні технології*
Спеціальність	*122 «Комп'ютерні науки»*
Освітня програма	*«Системи і методи штучного інтелекту»*
Статус дисципліни	*Вибіркова*
Форма навчання	*очна(денна)*
Рік підготовки, семестр	*3 курс, осінній семестр*
Обсяг дисципліни	*36 год.лекції, 18 год.практичні заняття*
Семестровий контроль/ контрольні заходи	*залік, письмовий / контрольні роботи, письмові*
Розклад занять	*Понеділок - перша пара, четвер – четверта пара, https://schedule.kpi.ua/.*
Мова викладання	*Українська*
Інформація про керівника курсу / викладачів	Лектор: *д.т.н., професор, доцент каф.ММСА Недашківська Надія Іванівна, n.nedashkivska@gmail.com* Практичні заняття: *д.т.н., професор, доцент каф.ММСА Недашківська Надія Іванівна, n.nedashkivska@gmail.com*
Розміщення курсу	Платформа дистанційного навчання «Сікорський», Googleclassroom, код курсу **37lzkci**

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Метою кредитного модуля є формування у студентів здатностей:

- *аналізу і обробки даних у різних форматах з метою підтримки прийняття рішень,*
- *пошуку шаблонів у великих і надвеликих базах даних,*
- *побудови прогнозів з використанням сучасних методів і алгоритмів інтелектуального аналізу даних, машинного навчання та інформаційних технологій;*
- *застосування сучасних методів інтелектуального аналізу даних та машинного навчання;*
- *використання програмного забезпечення для інтелектуального аналізу даних та машинного навчання в практичній роботі.*

Після засвоєння кредитного модуля мають продемонструвати такі результати навчання:

компетентності:

здатність застосовувати знання в практичних ситуаціях, здатність абстрактно мислити, застосовувати методи аналізу і синтезу, здатність знати та розуміти предметну область і професійну діяльність, здатність до пошуку, оброблення та аналізу інформації з різних джерел, здатність до адаптації та дії в новій ситуації, здатність забезпечувати та оцінювати якість виконуваних робіт,

здатність використовувати системний аналіз в якості сучасної міждисциплінарної методології, заснованої на прикладах математичних методів та сучасних інформаційних технологіях, і орієнтована на вирішення задач аналізу і синтезу технічних, економічних, соціальних, екологічних та інших складних систем,

здатність будувати математично коректні моделі статичних та динамічних процесів і систем із зосередженими та розподіленими параметрами із врахуванням невизначеності зовнішніх та внутрішніх факторів,

здатність до комп'ютерної реалізації математичних моделей реальних систем і процесів; проектувати, застосовувати і супроводжувати програмні засоби моделювання, прийняття рішень, обробки інформації, інтелектуального аналізу даних,

здатність розробляти експериментальні та спостережувальні дослідження і аналізувати дані, отримані в них,

застосовувати методи і засоби роботи з даними і знаннями, методи математичного моделювання, технології системного і статичного аналізу,

проектувати, реалізовувати, тестувати, впроваджувати, супроводжувати, експлуатувати програмні засоби роботи з даними і знаннями в комп'ютерних системах і мережах,

розуміти і застосовувати на практиці методи статичного моделювання і прогнозування, оцінювати вихідні дані

знання:
сучасних методів і алгоритмів інтелектуального аналізу даних та машинного навчання, методики застосування цих методів, алгоритмів для побудови прогнозів на основі статистичних даних, системного вирішення практичних задач аналізу і пошуку шаблонів у великих і надвеликих базах даних, методики розв'язання практичних задач класифікації на основі дерев рішень, опорних векторів, багатошарового персептрона, ансамблів моделей, задач кластеризації на основі ієрархічних, ітераційних, графових та щільнісних алгоритмів з використанням інформаційних технологій, знання бібліотеки scikit-learn python

уміння:
застосовувати сучасні методи і алгоритми інтелектуального аналізу даних та машинного навчання з метою підтримки прийняття рішень, будувати прогнози на основі статистичних даних, розв'язувати практичні задачі класифікації та кластеризації, системно вирішувати практичні задачі аналізу і пошуку шаблонів у великих і надвеликих базах даних, використовувати програмне забезпечення scikit-learn python для машинного навчання в практичній роботі

досвід:
теоретичний та практичний досвід аналізу і обробки даних у різних форматах з метою підтримки прийняття рішень, побудови прогнозів, використання програмного забезпечення scikit-learn python для інтелектуального аналізу даних та машинного навчання в практичній роботі

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

При вивченні дисципліни використовуються знання дисциплін «Теорія ймовірностей та математична статистика», «Математичний аналіз», «Дискретна математика (розділ «Теорія графів»)», «Об'єктно-орієнтовне програмування», «Методи оптимізації».

Знання, набуті при вивченні цієї дисципліни, використовуються при опануванні дисциплін «Інтелектуальний аналіз великих сховищ даних», «Інтелектуальні системи прийняття рішень», «Інтелектуальні системи підтримки прийняття рішень», в дипломному

проектуванні, у практичній самостійній роботі випускника в галузі інтелектуального аналізу даних під час аналізу великих і надвеликих баз даних, при побудові прогнозів на основі статистичних даних, при розробці корпоративних інформаційно-аналітичних систем в державних і приватних управлінських структурах.

3. Зміст навчальної дисципліни

Розділ 1. Основні поняття інтелектуального аналізу даних (ІАД) та машинного навчання (МН)

Тема 1. Загальні відомості про ІАД. Досвід в задачах МН: задачі навчання з вчителем – supervised learning і без вчителя – unsupervised learning. Класифікація, регресія, машинний переклад, пошук асоціативних правил, структурний вивід, синтез і вибірка, оцінка функції ймовірності або функції щільності ймовірності, кластеризація, сегментація. Огляд методів ІАД та машинного навчання.

Тема 2. Оцінювання якості алгоритмів МН. Перехресна перевірка (cross-validation, CV). Поняття перенавчання (overfitting) моделі МН та регуляризації моделі. Компроміс між систематичною помилкою і дисперсією моделі. Порівняння алгоритмів. Вибір гіперпараметрів моделі методами решітчастого Grid Search CV та рандомізованого пошуку Random Search CV.

Тема 3. Байесівський підхід та підхід максимальної правдоподібності до оцінювання якості моделі.

Розділ 2. Методи та алгоритми класифікації

Тема 1. Дерева рішень Decision Trees. Алгоритм розбиття, його властивості. Критерії вибору змінної розбиття: ентропійний, Джині. Алгоритми ID3, C4.5 вибору змінної розбиття, їх властивості. Міри ефективності дерев рішень. Проблема зупинки побудови дерева.

Тема 2. Дерева рішень в scikit-learn python. Реалізація алгоритму розбиття CART для класифікації та регресії. Регуляризація дерев рішень в scikit-learn python. Приклад практичної задачі багатокласової класифікації ірисів Фішера. Переваги і недоліки дерев рішень. Алгоритм покриття. 1-R алгоритм класифікації.

Тема 3. Оцінювання ефективності алгоритмів класифікації. K-Fold CV та його модифікації. Матриця неточностей (confusion matrix), метрики accuracy, precision, recall, specificity, F1-score для вибору моделі. Криві ROC-curve, PR-curve. Вибір моделі на прикладі задачі класифікації ірисів Фішера. Проблема незбалансованих класів.

Тема 4. Байесівський підхід до класифікації. Оптимальний байесівський класифікатор. Оцінювання апріорних імовірностей та функцій правдоподібності за вибіркою. Наївний байесівський класифікатор. Реалізація алгоритму Naïve Bayes в scikit-learn python. Приклад практичної задачі класифікації алгоритмом Naïve Bayes в scikit-learn python.

Тема 5. Задача розділу суміші. Алгоритм Expectation-Maximization та його модифікації. Реалізація алгоритму GaussianMixture в scikit-learn python. Приклад практичної задачі класифікації алгоритмом GaussianMixture.

Тема 6. Побудова математичних функцій класифікації. Метод опорних векторів (support vector machine, SVM): лінійний та нелінійний випадки. Реалізація методу в scikit-learn python. Приклад розв'язання задачі класифікації алгоритмами SVC, NuSVC та LinearSVC scikit-learn python. Приклад розв'язання задачі регресії алгоритмами SVR, NuSVR та LinearSVR scikit-learn python. Обгрунтування методу опорних векторів. Вибір параметру розмиття смуги. Функція ядра, властивості, типи ядер. Переваги і недоліки методу опорних векторів.

Тема 7. Основи теорії штучних нейронних мереж. Класичний і сучасні персептрони (multiple layer perceptron, MLP). Функції активації: сигмоїдна, гіперболічний тангенс, ReLU, LeackyReLU та їх модифікації. Реалізація побудови і навчання багатошарового персептрону в scikit-learn python.

Приклад розв'язання задачі класифікації набору даних MNIST алгоритмом MLPClassifier scikit-learn python. Приклад розв'язання задачі регресії – прогнозування попиту на велосипеди на основі сезонного циклічного часового ряду fetch_openml – алгоритмом MLPRegressor scikit-learn python. Метод стохастичного градієнтного спуску. Пакетний, стохастичний (SGD) та міні-пакетний (mini-batch) варіанти реалізації градієнтного спуску (gradient descent, GD). Проблема вибору гіперпараметру швидкості навчання learning rate.

Розділ 3. Методи та алгоритми кластеризації

Тема 1. Функції відстані. Ієрархічна кластеризація: агломеративний алгоритм найближчого сусіда, дівізимний алгоритм. Алгоритм AgglomerativeClustering scikit-learn python. Методи розрахунку відстані між кластерами. Приклад кластеризації наборів даних різної форми алгоритмом AgglomerativeClustering scikit-learn python. Поняття дендрограми та її побудова засобами scipy.cluster.hierarchy.

Тема 2. Алгоритми k-середніх, нечітких k-середніх та g-середніх. Реалізація в scikit-learn python: алгоритми KMeans, MiniBatchKMeans. Приклади кластеризації наборів даних різної форми, використовуючи KMeans scikit-learn python, вибір значень гіперпараметрів. Емпірична оцінка впливу ініціалізації в методі k-середніх. Порівняння алгоритмів KMeans та MiniBatchKMeans на наборах даних make_blobs.

Тема 3. Аналіз результатів кластеризації. Метрики якості homogeneity, completeness, v-measure, adjusted rand score, adjusted mutual info score, коефіцієнт силуету. Вибір кількості кластерів за допомогою аналізу силуетів у кластеризації KMeans. Порівняння результатів, оцінювання часу виконання та якості результатів на прикладі практичної задачі кластеризації рукописних цифр з набору sklearn.datasets.load_digits.

Тема 4. Методи кластеризації на основі теорії графів. Алгоритми Прима, Крускала і Борувки побудови мінімального покриваючого дерева. Алгоритм Форел та його модифікації.

Тема 5. Кластеризація на основі мережі Кохонена. Конкурентне навчання. Метод самоорганізуючих карт Кохонена. Інтерпретація карт Кохонена.

Тема 6. Щільнісні алгоритми кластеризації Mean Shift, DBSCAN та OPTICS. Реалізація в scikit-learn python та вибір гіперпараметрів. Аналіз результатів кластеризації наборів даних різної форми.

Алгоритми Affinity propagation, SpectralClustering та Birch.

Розділ 4. Ансамблі моделей ІАД

Тема 1. Види ансамблів. Беггінг. Композиції дерев рішень та випадковий ліс Random Forest. Реалізація в scikit-learn python: алгоритми BaggingClassifier, BaggingRegressor, RandomForestClassifier, RandomForestRegressor, ExtraTreesClassifier, ExtraTreesRegressor. Методи випадкових ділянок (random patches method) і випадкових підпросторів (random subspaces method). Переваги і недоліки випадкового лісу. Класифікатор з жорстким і м'яким голосуванням. Реалізація в scikit-learn python: VotingClassifier, VotingRegressor.

Тема 2. Бустинг. Методи AdaBoost та градієнтний бустинг. Реалізація в scikit-learn python: алгоритми AdaBoostClassifier, AdaBoostRegressor, GradientBoostingClassifier, GradientBoostingRegressor. Зміст параметрів. Переваги і недоліки бустингу. Стекінг. Реалізація в scikit-learn python: StackingClassifier, StackingRegressor.

Напрямки розвитку та перспективи подальших досліджень в області ІАД та машинного навчання. Невирішені проблеми.

4. Навчальні матеріали та ресурси

Базова

1. *Н.І. Недашківська. Конспект і слайди лекцій з кредитного модуля «Інтелектуальний аналіз даних», 2023. https://classroom.google.com/c/NjIwNzk2OTE0NjIx?cjc=37lzkci*
2. *Н.І. Недашківська. Інтелектуальний аналіз даних : Практикум [Електронний ресурс] : навч. посіб. для студ. спеціальності 124 «Системний аналіз», освітніх програм «Системний аналіз і управління», «Системний аналіз фінансового ринку»/ Н. І. Недашківська; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 6 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 105 с. https://ela.kpi.ua/handle/123456789/53763*

Знайти зазначені в п. 1 і 2 матеріали можна в Електронному Кампусі та на Платформі дистанційного навчання «Сікорський», Googleclassroom, код курсу **37lzkci**.

Додаткова література

3. *Scikit-Learn Documentation. https://scikit-learn.org/*
4. *Jake VanderPlas. Python Data Science Handbook. Essential Tools for Working with Data. O'Reilly Media Inc., 2017. 576 p. (за запитом викладачу)*
5. *Sebastian Raschka, Vahid Mirjalili. Python Machine Learning. Third Edition. Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2. Packt Publishing, 2019. https://github.com/rasbt/python-machine-learning-book-3rd-edition*
6. *Ian Goodfellow, Yoshua Bengio, Aaron Courville. Deep Learning. The MIT Press Cambridge, Massachusetts London, England, 2017. https://www.deeplearningbook.org/*
7. *Wes McKinney. Python for Data Analysis. O'Reilly Media Inc., 2013. 482 p. (за запитом викладачу)*
8. *Henrik Brink Joseph W. Richards Mark Fetherolf. Real-World Machine Learning. Manning Publications. 2016. 336 p. (за запитом викладачу)*
9. *Andreas C. Mueller and Sarah Guido. An Introduction to Machine Learning with Python. O'Reilly Media Inc., 2017. 392 p. (за запитом викладачу)*
10. *Davy Cielen, Arno Meysman, Mohamed Ali. Introducing Data Science: Big Data, Machine Learning, and more, using Python tools. Manning Publications, 2016. 320 p. (за запитом викладачу)*

Навчальний контент

5. Методика опанування навчальної дисципліни(освітнього компонента)

Структура кредитного модуля

Назви розділів і тем	Кількість годин				
	Всього	у тому числі			
		Лекції	Практичні (семінарські)	Лабораторні (комп'ютерний практикум)	СРС
1	2	3	4	5	6
Розділ 1. Основні поняття інтелектуального аналізу даних (ІАД) та машинного навчання (МН)					
Тема 1. Загальні відомості про ІАД. Досвід в задачах МН: задачі навчання з вчителем – supervised learning і без вчителя – unsupervised learning. Класифікація, регресія, машинний переклад, пошук асоціативних правил, структурний вивід, синтез і вибірка, оцінка функції ймовірності або функції	4	2			2

1	2	3	4	5	6
щільності ймовірності, кластеризація, сегментація. Огляд методів ІАД та машинного навчання.					
Тема 2. Оцінювання якості алгоритмів МН. Перехресна перевірка (cross-validation, CV). Поняття перенавчання (overfitting) моделі МН та регуляризації моделі. Компроміс між систематичною помилкою і дисперсією моделі. Криві навчання і перевірки. U-крива. Порівняння алгоритмів. Вибір гіперпараметрів моделі методами решітчастого Grid Search CV та рандомізованого пошуку Random Search CV.	6	3	2		1
Тема 3. Байесівський підхід та підхід максимальної правдоподібності до оцінювання якості моделі.	2	1			1
Разом за розділом 1	**12**	**6**	**2**		**4**
Розділ 2. Методи та алгоритми класифікації					
Тема 1. Дерева рішень Decision Trees. Алгоритм розбиття, його властивості. Критерії вибору змінної розбиття: ентропійний, Джині. Алгоритми ID3, C4.5 вибору змінної розбиття, їх властивості. Міри ефективності дерев рішень. Проблема зупинки побудови дерева.	6	2	1		2
Тема 2. Дерева рішень в scikit-learn python. Реалізація алгоритму розбиття CART для класифікації та регресії. Регуляризація дерев рішень в scikit-learn python. Приклад практичної задачі багатокласової класифікації ірисів Фішера. Переваги і недоліки дерев рішень. Алгоритм покриття. 1-R алгоритм класифікації.	6	2	1		3
Тема 3. Оцінювання ефективності алгоритмів класифікації. K-Fold CV та його модифікації. Матриця неточностей (confusion matrix), метрики accuracy, precision, recall, specificity, F1-score для вибору моделі. Криві ROC-curve, PR-curve. Вибір моделі на прикладі задачі класифікації ірисів Фішера. Проблема незбалансованих класів.	5	2	2		2
Тема 4. Байесівський підхід до класифікації. Оптимальний байесівський класифікатор. Оцінювання апріорних імовірностей та функцій правдоподібності за вибіркою. Наївний байесівський класифікатор. Реалізація алгоритму Naïve Bayes в scikit-learn python. Приклад практичної задачі класифікації алгоритмом Naïve Bayes в scikit-learn python.	9	2	1		6
Тема 5. Задача розділу суміші. Алгоритм Expectation-Maximization та його модифікації. Реалізація алгоритму GaussianMixture в scikit-learn python. Приклад практичної задачі класифікації алгоритмом GaussianMixture.	6	2	1		3
Тема 6. Побудова математичних функцій класифікації. Метод опорних векторів (support vector machine, SVM): лінійний та нелінійний випадки. Реалізація методу в scikit-learn python. Приклад розв'язання задачі класифікації алгоритмами SVC, NuSVC та LinearSVC scikit-learn python. Приклад розв'язання задачі регресії алгоритмами SVR, NuSVR та LinearSVR scikit-learn	7	2	1		3

1	2	3	4	5	6
python. Обгрунтування методу опорних векторів. Вибір параметру розмиття смуги. Функція ядра, властивості, типи ядер. Переваги і недоліки методу опорних векторів.					
Тема 7. Основи теорії штучних нейронних мереж. Класичний і сучасні персептрони (multiple layer perceptron, MLP). Функції активації: сигмоїдна, гіперболічний тангенс, ReLU, LeackyReLU та їх модифікації. Реалізація побудови і навчання багатошарового персептрону в scikit-learn python. Приклад розв'язання задачі класифікації набору даних MNIST алгоритмом MLPClassifier scikit-learn python. Приклад розв'язання задачі регресії – прогнозування попиту на велосипеди на основі сезонного циклічного часового ряду fetch_openml – алгоритмом MLPRegressor scikit-learn python. Метод стохастичного градієнтного спуску. Пакетний, стохастичний (SGD) та міні-пакетний (mini-batch) варіанти реалізації градієнтного спуску (gradient descent, GD). Проблема вибору гіперпараметру швидкості навчання learning rate.	4	2	1		4
Перша частина модульної контрольної роботи	4				4
Разом за розділом 2	**39**	**14**	**8**		**19**
Розділ 3. Методи та алгоритми кластеризації					
Тема 1. Функції відстані. Ієрархічна кластеризація: агломеративний алгоритм найближчого сусіда, дівізимний алгоритм. Алгоритм AgglomerativeClustering scikit-learn python. Методи розрахунку відстані між кластерами. Приклад кластеризації наборів даних різної форми алгоритмом AgglomerativeClustering scikit-learn python. Поняття дендрограми та її побудова засобами scipy.cluster.hierarchy.	5	2	1		2
Тема 2. Алгоритми k-середніх, нечітких k-середніх та g-середніх. Реалізація в scikit-learn python: алгоритми KMeans, MiniBatchKMeans. Приклади кластеризації наборів даних різної форми, використовуючи KMeans scikit-learn python, вибір значень гіперпараметрів. Емпірична оцінка впливу ініціалізації в методі k-середніх. Порівняння алгоритмів KMeans та MiniBatchKMeans на наборах даних make_blobs.	6	2	1		3
Тема 3. Аналіз результатів кластеризації. Метрики якості homogeneity, completeness, v-measure, adjusted rand score, adjusted mutual info score, коефіцієнт силуету. Вибір кількості кластерів за допомогою аналізу силуетів у кластеризації KMeans. Порівняння результатів, оцінювання часу виконання та якості результатів на прикладі практичної задачі кластеризації рукописних цифр з набору sklearn.datasets.load_digits.	5	2	1		3
Тема 4. Методи кластеризації на основі теорії графів. Алгоритми Прима, Крускала і Борувки побудови мінімального покриваючого дерева. Алгоритм Форел та його модифікації.	2	2			2

1	2	3	4	5	6
Тема 5. Кластеризація на основі мережі Кохонена. Конкурентне навчання. Метод самоорганізуючих карт Кохонена. Інтерпретація карт Кохонена.	3.5	2			2
Тема 6. Щільнісні алгоритми кластеризації Mean Shift, DBSCAN та OPTICS. Реалізація в scikit-learn python та вибір гіперпараметрів. Аналіз результатів кластеризації наборів даних різної форми. Алгоритми Affinity propagation, SpectralClustering та Birch.	2.5	2	1		2
Разом за розділом 3	**31**	**12**	**4**		**16**
Розділ 4. Ансамблі моделей ІАД					
Тема 1. Види ансамблів. Беггінг. Композиції дерев рішень та випадковий ліс Random Forest. Реалізація в scikit-learn python: алгоритми BaggingClassifier, BaggingRegressor, RandomForestClassifier, RandomForestRegressor, ExtraTreesClassifier, ExtraTreesRegressor. Методи випадкових ділянок (random patches method) і випадкових підпросторів (random subspaces method). Переваги і недоліки випадкового лісу. Класифікатор з жорстким і м'яким голосуванням. Реалізація в scikit-learn python: VotingClassifier, VotingRegressor.	6	2	2		2
Тема 2. Бустинг. Методи AdaBoost та градієнтний бустинг. Реалізація в scikit-learn python: алгоритми AdaBoostClassifier, AdaBoostRegressor, GradientBoostingClassifier, GradientBoostingRegressor. Зміст параметрів. Переваги і недоліки бустингу. Стекінг. Реалізація в scikit-learn python: StackingClassifier, StackingRegressor. Практична задача: прогнозування і оцінювання важливості ознак (features) за допомогою RandomForestClassifier. Порівняння різних класифікаторів на основі бегінгу, бустингу і стекінгу, їх метрик якості, кривих ROC і DET для трьох різних навчальних наборів даних: make_moons, make_circles, make_classification. Напрямки розвитку та перспективи подальших досліджень в області ІАД та машинного навчання. Невирішені проблеми.	8	2	2		4
Друга частина модульної контрольної роботи	4				4
Разом за розділом 4	**15**	**6**	**4**		**7**
Залік	4				16
Всього годин	**108**	**36**		**18**	**54**

Лекційні заняття

№ з/п	Назва теми лекції та перелік основних питань (перелік дидактичних засобів, посилання на літературу та завдання на СРС)
1	*Тема 1.1. Загальні відомості про ІАД. Досвід в задачах МН: задачі навчання з вчителем – supervised learning і без вчителя – unsupervised learning. Класифікація, регресія, машинний переклад, пошук асоціативних правил, структурний вивід, синтез і вибірка, оцінка функції ймовірності або функції щільності ймовірності, кластеризація, сегментація. Огляд методів ІАД та машинного навчання. [1, 2, 4, 6]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Розв'язання практичних задач в сфері інтернет-технологій. Методи і технології Web Mining [4, 8]*
2	*Тема 1.2. Оцінювання якості алгоритмів МН. Перехресна перевірка (cross-validation, CV). Поняття перенавчання (overfitting) моделі МН та регуляризації моделі. Компроміс між систематичною помилкою і дисперсією моделі. Криві навчання і перевірки. U-крива. Порівняння алгоритмів. Вибір гіперпараметрів моделі методами решітчастого Grid Search CV та рандомізованого пошуку Random Search CV. [1 – 11]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Етапи процесу виявлення знань. Управління знаннями (knowledge management). Метод, орієнтований на продукти. Метод, орієнтований на процеси. [8, 9]*
3	*Тема 1.3. Теорема Байеса. Максимальна апостеріорна гіпотеза. Метод максимальної правдоподібності. [1, 2, 6]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Задача лінійної регресії. Міра якості в задачі лінійної регресії. Функція помилки в задачі регресії та її обгрунтування. Регуляризація зі зниженням ваги та її обгрунтування за допомогою байесівського підходу. Гребнева регресія. [1 – 4] Лассо регресія. Еластична регресія [4, 5].*
4	*Тема 2.1. Дерева рішень Decision Trees. Алгоритм розбиття, його властивості. Критерії вибору змінної розбиття: ентропійний, Джині. Алгоритми ID3, C4.5 вибору змінної розбиття, їх властивості. Міри ефективності дерев рішень. Проблема зупинки побудови дерева. [1 – 5]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Вибір параметрів методу Decision Trees. Представлення результатів [5, 11]*
5	*Тема 2.2. Дерева рішень в scikit-learn python [5]. Реалізація алгоритму розбиття CART для класифікації та регресії. Регуляризація дерев рішень в scikit-learn python. Приклад практичної задачі багатокласової класифікації ірисів Фішера. Переваги і недоліки дерев рішень [1, 4]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Алгоритм покриття. 1-R алгоритм класифікації. [1, 2]*
6	*Тема 2.3. Оцінювання ефективності алгоритмів класифікації. K-Fold CV та його модифікації. Матриця неточностей (confusion matrix), метрики accuracy, precision, recall, specificity, F1-score для вибору моделі. [1 – 5] Криві ROC-curve, PR-curve. Вибір моделі на прикладі задачі класифікації ірисів Фішера. Проблема незбалансованих класів. [1, 2]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Крива DET-curve [5]. Приклади розв'язання практичних задач [5, 11].*
7	*Тема 2.4. Байесівський підхід до класифікації. Оптимальний байесівський класифікатор. Оцінювання апріорних імовірностей та функцій правдоподібності за*

	вибіркою. [1, 2] Наївний байєсівський класифікатор. Реалізація алгоритму Naïve Bayes в scikit-learn python [5]. Приклад практичної задачі класифікації алгоритмом Naïve Bayes в scikit-learn python. Наочність навчальних занять забезпечується використанням слайдів.
8	*Тема 2.5. Задача розділу суміші. Алгоритм Expectation-Maximization та його модифікації. [1 – 3] Реалізація алгоритму GaussianMixture в scikit-learn python [5]. Приклад практичної задачі класифікації алгоритмом GaussianMixture.* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Реалізація в scikit-learn python.*
9	*Тема 2.6. Побудова математичних функцій класифікації. Метод опорних векторів (support vector machine, SVM): лінійний та нелінійний випадки. [1, 2] Реалізація методу в scikit-learn python [3, 5]. Приклад розв'язання задачі класифікації алгоритмами SVC, NuSVC та LinearSVC scikit-learn python. Приклад розв'язання задачі регресії алгоритмами SVR, NuSVR та LinearSVR scikit-learn python [5]. Обгрунтування методу опорних векторів. [1, 2] Вибір параметру розмиття смуги. Функція ядра, властивості [4]. Переваги і недоліки методу опорних векторів.* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Вибір типу ядра в SVM [4, 5]. Настроювання параметрів методу SVM [1, 3, 4, 8]*
10	*Тема 2.7. Основи теорії штучних нейронних мереж. Класичний і сучасні персептрони (multiple layer perceptron, MLP). Функції активації: сигмоїдна, гіперболічний тангенс, ReLU, LeackyReLU та їх модифікації. [1, 2, 6, 7] Реалізація побудови і навчання багатошарового персептрону в scikit-learn python [5]. Приклад розв'язання задачі класифікації набору даних MNIST алгоритмом MLPClassifier scikit-learn python. Приклад розв'язання задачі регресії – прогнозування попиту на велосипеди на основі сезонного циклічного часового ряду fetch_openml – алгоритмом MLPRegressor scikit-learn python. Метод стохастичного градієнтного спуску. Пакетний, стохастичний (SGD) та міні-пакетний (mini-batch) варіанти реалізації градієнтного спуску (gradient descent, GD). Проблема вибору гіперпараметру швидкості навчання learning rate [1,2]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Адаптивні алгоритми градієнтного спуску* [6, 7]. *Функція активації Swish.*
11	*Тема 3.1. Функції відстані. Ієрархічна кластеризація: агломеративний алгоритм найближчого сусіда, дівізимний алгоритм. [1 – 4] Алгоритм AgglomerativeClustering scikit-learn python. Методи розрахунку відстані між кластерами. Приклад кластеризації наборів даних різної форми алгоритмом AgglomerativeClustering scikit-learn python [5]. Поняття дендрограми.* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *побудова дендрограми засобами scipy.cluster.hierarchy [5].*
12	*Тема 3.2. Алгоритми k-середніх, нечітких k-середніх та g-середніх. [1 – 3, 8 – 11] Реалізація в scikit-learn python: алгоритми KMeans, MiniBatchKMeans [5]. Приклади кластеризації наборів даних різної форми, використовуючи KMeans scikit-learn python, вибір значень гіперпараметрів. Емпірична оцінка впливу ініціалізації в методі k-середніх. Порівняння алгоритмів KMeans та MiniBatchKMeans на наборах даних make_blobs.* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Порівняння алгоритмів KMeans та MiniBatchKMeans на різних наборах даних.*
13	*Тема 3.3. Аналіз результатів кластеризації. Метрики якості homogeneity, completeness, v-measure, adjusted rand score, adjusted mutual info score, коефіцієнт силуету. [1, 2, 5] Вибір кількості кластерів за допомогою аналізу силуетів у кластеризації KMeans. Порівняння результатів, оцінювання часу виконання та якості результатів на прикладі практичної задачі кластеризації рукописних цифр з*

	набору sklearn.datasets.load_digits. Наочність навчальних занять забезпечується використанням слайдів.
14	*Тема 3.4. Методи кластеризації на основі теорії графів. Алгоритми Прима, Крускала і Борувки побудови мінімального покриваючого дерева. [1]* *Алгоритм Форел та його модифікації. [1]* *Кластеризація на основі мережі Кохонена. Конкурентне навчання. Метод самоорганізуючих карт Кохонена. Інтерпретація карт Кохонена. [1]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Розв'язання практичної задачі в ППП Deductor.*
15	*Тема 3.5. Щільнісні алгоритми кластеризації Mean Shift, DBSCAN та OPTICS. [1, 2, 4] Реалізація в scikit-learn python та вибір гіперпараметрів [5]. Аналіз результатів кластеризації наборів даних різної форми.* *Алгоритми Affinity propagation, SpectralClustering та Birch.* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Реалізація алгоритмів в scikit-learn python.*
16	*Тема 4.1. Види ансамблів. Беггінг. Композиції дерев рішень та випадковий ліс Random Forest. [1 – 4] Реалізація в scikit-learn python: алгоритми BaggingClassifier, BaggingRegressor, RandomForestClassifier, RandomForestRegressor, ExtraTreesClassifier, ExtraTreesRegressor. [5] Методи випадкових ділянок (random patches method) і випадкових підпросторів (random subspaces method). Переваги і недоліки випадкового лісу.* *Класифікатор з жорстким і м'яким голосуванням. Реалізація в scikit-learn python: VotingClassifier, VotingRegressor. [5]* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Настройка параметрів. Інтерпретація результатів [4, 5, 8 – 11]*
17	*Тема 4.2. Бустинг. Методи AdaBoost та градієнтний бустинг. [1 – 4] Реалізація в scikit-learn python: алгоритми AdaBoostClassifier, AdaBoostRegressor, GradientBoostingClassifier, GradientBoostingRegressor. [5] Зміст параметрів. Переваги і недоліки бустингу.* Наочність навчальних занять забезпечується використанням слайдів. **Завдання на СРС:** *Настройка параметрів. Інтерпретація результатів [4, 5, 8 – 11]*
18	*Тема 4.3. Стекінг. [1 – 4] Реалізація в scikit-learn python: StackingClassifier, StackingRegressor. [5]* *Практична задача: прогнозування і оцінювання важливості ознак (features) за допомогою RandomForestClassifier. Порівняння різних класифікаторів на основі бегінгу, бустингу і стекінгу, їх метрик якості, кривих ROC і DET для трьох різних навчальних наборів даних: make_moons, make_circles, make_classification.* *Методи розрахунку коефіцієнтів відносних важливостей (ваг) моделей ансамблю [1]* *Напрямки розвитку та перспективи подальших досліджень в області ІАД та машинного навчання. Невирішені проблеми.* Наочність навчальних занять забезпечується використанням слайдів.

Практичні заняття

Метою практичних занять є закріплення теоретичних положень навчальної дисципліни, отримання практичних навичок використання методів інтелектуального аналізу даних і машинного навчання з метою підтримки прийняття рішень. В результаті виконання робіт студенти повинні вміти будувати прогнози на основі статистичних даних, використовуючи вказані методи, розв'язувати практичні задачі класифікації і кластеризації з використанням

сучасного програмного забезпечення Scikit-Learn Python, системно вирішувати практичні задачі аналізу і пошуку шаблонів у великих і надвеликих базах даних, використовувати сучасне програмне забезпечення Scikit-Learn Python для інтелектуального аналізу даних та машинного навчання в практичній роботі.

№ з/п	Назва роботи	Кількість ауд. годин
1	Отримання навичок роботи в середовищі Python	2
2	Побудова та оцінювання якості моделей дерев рішень та опорних векторів для класифікації та регресії засобами бібліотеки Scikit-Learn Python	4
3	Класифікація та регресія на основі багатошарового персептрона в Scikit-Learn Python	2
4	Побудова та оцінювання якості моделей кластеризації в Scikit-Learn Python	6
5	Побудова та оцінювання ансамблів моделей класифікації та регресії засобами Scikit-Learn Python	4

Для виконання практичних робіт використовується open-source програмне забезпечення Python (https://www.python.org/), Scikit-Learn 1.2.1 – open source, commercially usable – BSD license (https://scikit-learn.org/), TensorFlow v.2.11.0 – Apache-2.0 license (https://www.tensorflow.org), Keras – Apache-2.0 license (https://keras.io)

6. Самостійна робота студента

Самостійна робота студента включає підготовку до практичних робіт, підготовку до модульної контрольної роботи, в тому числі опрацювання тем, які не увійшли до лекцій та розв'язок задач.

№ з/п	Назва теми, що виноситься на самостійне опрацювання	К-ть годин СРС
1	Процес ІАД. Підготовлення даних	3
2	Практичне застосування ІАД	3
3	Методи побудови дерев рішень	3
4	Байесівські методи класифікації	3
5	Метод опорних векторів	3
6	Нечітко-нейронні системи для розв'язання задач класифікації	3
7	Ієрархічна кластеризація	3
8	Статистичні алгоритми k-середніх, EM та їх модифікації	3
9	Метод самоорганізуючих карт Кохонена	3
10	Ансамблі моделей ІАД.	3
11	Методи розрахунку коефіцієнтів відносних важливостей (ваг) моделей в ансамблі	3

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

***Пропущені контрольні заходи оцінювання**. Кожен студент має право відпрацювати пропущені з поважної причини (лікарняний, мобільність тощо) заняття за рахунок самостійної роботи. Детальніше за посиланням: https://kpi.ua/files/n3277.pdf.*

***Процедура оскарження результатів контрольних заходів оцінювання**. Студент може підняти будь-яке питання, яке стосується процедури контрольних заходів та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами. Студенти мають право аргументовано оскаржити результати контрольних заходів, пояснивши з яким критерієм не погоджуються відповідно до оціночного.*

***Календарний контроль** проводиться з метою підвищення якості навчання студентів та моніторингу виконання студентом вимог силабусу.*

***Академічна доброчесність.** Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: https://kpi.ua/code.*

***Норми етичної поведінки.** Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: https://kpi.ua/code.*

***Інклюзивне навчання.** Засвоєння знань та умінь в ході вивчення дисципліни «Сталий інноваційний розвиток» може бути доступним для більшості осіб з особливими освітніми потребами, окрім здобувачів з серйозними вадами зору, які не дозволяють виконувати завдання за допомогою персональних комп'ютерів, ноутбуків та/або інших технічних засобів.*

***Навчання іноземною мовою**. У ході виконання завдань студентам може бути рекомендовано звернутися до англомовних джерел.*

8. Види контролю та рейтингова система оцінювання результатів навчання (РСО)

Поточний контроль: модульна контрольна робота.

Модульна контрольна робота складається з трьох частин: КР – 1, КР – 2 і КР – 3. Кожна КР містить теоретичні питання і/або задачу. Оцінки за теоретичні питання визначаються за шкалою:

- «відмінно», повна відповідь (не менше 95% потрібної інформації) – 4,75 – 5 балів;
- «добре», достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 3,75 – 4,7 бали;
- «задовільно», неповна відповідь (не менше 60% потрібної інформації) та незначні помилки – 3 – 3,7 бали;
- «незадовільно» (не відповідає вимогам на «задовільно») – 0 – 2,9 балів.

Оцінки за задачі визначаються за шкалою:

- завдання виконано правильно – 4,75 – 5 балів;
- завдання виконано з незначними помилками або неточностями – 3,75 – 4,7 бали;
- завдання виконано не повністю або з суттєвими помилками – 3 – 3,7 бали;
- завдання не виконано – 0 балів.

Максимальна оцінка за КР – 1 складає 10 балів, КР – 2 складає 5 балів і КР – 3 – 15 балів.

Максимальна кількість балів за три частини модульної КР складає 30 балів.

***Календарний контроль:** проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.*

***Семестровий контроль:** залік.*

***Умови допуску до семестрового контролю:** виконання та захист першої, другої та четвертої практичних робіт, семестровий рейтинг не менше 40 балів.*

Рейтингова система оцінювання результатів навчання

Рейтинг студента з кредитного модуля складається з балів, які він отримує за:

1) виконання та захист 5 практичних робіт;
2) виконання модульної контрольної роботи.

1. Практичні роботи. Упродовж семестру студент має виконати 5 практичних робіт (ПР).

Рейтингова оцінка кожної ПР складається з 2 частин, які оцінюються окремо.

а. Якість підготовки до роботи, її виконання та оформлення звіту.

- за умови правильно оформленого звіту з точним виконанням завдання ПР – 6,5 -7 балів;
- за наявності несуттєвих неточностей в оформленні або процедурі виконання ПР – 5-6 балів;
- за наявності порушень в оформленні, неповного або неточного виконання – 3-4 бали.

б. Якість захисту матеріалу. В цій частині оцінюється ступінь володіння теоретичним і практичним матеріалом за темою роботи.

- відмінне володіння матеріалом – 6,5 – 7 балів;
- добре володіння матеріалом – 5 – 6 балів;
- задовільне володіння матеріалом – 4 бали.

	ПР – 1	ПР – 2	ПР – 3	ПР – 4	ПР – 5
Строк здачі та захисту роботи	28.09	12.10	26.10	16.11	07.11

За несвоєчасну подачу звіту з ПР та несвоєчасний захист ПР нараховуються штрафні бали: мінус 0.3 бали за кожний тиждень запізнення.

Максимальна кількість балів за всі ПР дорівнює: 5*14 =70 балів.

2. Модульна контрольна робота. Модульна контрольна робота складається з двох частин – КР№1 і КР№2. Кожна КР містить два теоретичні питання і задачу. Оцінки за теоретичні питання визначаються за шкалою:

- «відмінно», повна відповідь (не менше 95% потрібної інформації) – 5 балів;
- «добре», достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 4 бали;
- «задовільно», неповна відповідь (не менше 60% потрібної інформації) та незначні помилки – 3 бали;
- «незадовільно», незадовільна відповідь (не відповідає вимогам на «задовільно») – 0 балів.

Оцінки за задачі визначаються за шкалою:

- завдання виконано правильно – 5 балів;
- завдання виконано з незначними помилками або неточностями – 4 бали;
- завдання виконано не повністю або з суттєвими помилками – 3 бали;
- завдання не виконано – 0 балів.

Максимальна оцінка за КР – 1 складає 10 балів, КР – 2 складає 5 балів і КР – 3 – 15 балів.

Максимальна кількість балів за три частини модульної КР складає 30 балів.

	КР – 1	КР – 2	КР – 3
Дата проведення роботи	09.10	06.11	04.12

За результатами навчальної роботи за перші 7 тижнів станом на 21.10 «ідеальний студент» має набрати 43 бали, які включають дві практичні роботи та першу частину МКР. **На першому календарному контролі на 8-му тижні студент отримує «зараховано», якщо його поточний рейтинг не менше 22 балів.**

За результатами 14 тижнів навчання станом на 09.12 «ідеальний студент» має набрати 85 балів, що включає п'ять зданих і захищених практичних роботи та першу частину МКР. **На другому календарному контролі на 15-му тижні студент отримує «зараховано», якщо його поточний рейтинг не менше 43 балів.**

Максимальна сума балів за роботу в семестрі складає 100. Необхідною умовою допуску до заліку є отримання рейтингу 40 балів і вище. Для отримання заліку з кредитного модуля «автоматом» потрібно мати рейтинг не менше 60 балів.

Студенти, які наприкінці семестру мають рейтинг менше 60 балів і виконали Умови допуску до семестрового контролю (див. вище), а також ті, хто хоче підвищити оцінку, виконують залікову контрольну роботу. При цьому до балів за практичні роботи додаються бали за залікову контрольну роботу, і ця рейтингова оцінка є остаточною. Завдання залікової контрольної роботи складається з двох теоретичних питань різних розділів робочої програми і двох практичних завдань.

Кожне теоретичне і практичне питання залікової контрольної роботи оцінюється у 7,5 балів відповідно до системи оцінювання:

- «відмінно», повна відповідь (не менше 95% потрібної інформації) – 7 – 7,5 балів;
- «добре», достатньо повна відповідь (не менше 75% потрібної інформації або незначні неточності) – 6-7 балів;
- «задовільно», неповна відповідь (не менше 60% потрібної інформації та деякі помилки) – 4,5- 5 балів;
- «незадовільно», незадовільна відповідь – 0 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Кількість балів	*Оцінка*
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Менше 40	Не допущено

9. Додаткова інформація з дисципліни (освітнього компонента)

Перелік питань, які виносяться на семестровий контроль, наведено в додатку А.

Сертифікати проходження дистанційних чи онлайн курсів за тематикою дисципліни можуть бути зараховані з додатковими 5 – 10 балами до загального рейтингу студента.

Робочу програму навчальної дисципліни (силабус):

Складено професор, д.т.н. Недашківська Надія Іванівна

Ухвалено кафедрою ММСА НН ІПСА (протокол № 13 від 05.06.2024)

Погоджено Методичною комісією НН ІПСА (протокол № 10 від 24.06.2024)

Додаток А

Перелік питань з дисципліни

«Інтелектуальний аналіз даних»

третій курс, ММСА ІПСА НТУУ «КПІ ім. Ігоря Сікорського»
викладач д.т.н., доц. Недашківська Н.І.

1. Задачі машинного навчання (МН). *
2. Задачі навчання з учителем і без учителя. *
3. Перенавчання (overfitting) моделі МН. Ємність моделі. *
4. Компроміс між систематичною помилкою і дисперсією.
5. Крива перевірки.
6. Теорема Байеса і максимальна апостеріорна гіпотеза.
7. Лінійна регресія. *
8. Постановка задачі та ідея методу опорних векторів для лінійно роздільного випадку. Функціонал помилок. Означення опорного вектору. *
9. Теоретичне обгрунтування методу опорних векторів.
10. Метод опорних векторів для лінійно нероздільного випадку. Ідея методу. Типи векторів-порушників роздільчої смуги.
11. Метод опорних векторів для лінійно нероздільного випадку. Обгрунтування методу.
12. Нелінійне узагальнення методу опорних векторів. Означення ядра і приклади. Обгрунтування методу.
13. Функції класифікації за методом опорних векторів для лінійно роздільного і нероздільного випадків. Переваги і обмеження методу. Як визначається параметр C ? *
14. Поняття дерева рішень. Структура дерева рішень. Листові і проміжні вершини дерева. Навести приклад. *
15. Алгоритм розбиття побудови дерев рішень. Властивості алгоритму розбиття. *
16. Означення ентропії. Ентропійний критерій вибору змінної розбиття. *
17. Проблема перенавчання (зверхчутливості) дерев рішень. Алгоритм C4.5 вибору змінної розбиття. Модифікований C4.5 для випадку неперервних атрибутів.
18. Дерева рішень для класифікації в Scikit-Learn Python. Атрибути вузла дерева. Міра забрудненості Джині. *
19. Алгоритм CART для класифікації. Алгоритм розбиття побудови дерев рішень класифікації *
20. Переваги і обмеження алгоритму розбиття. Регуляризація дерев рішень. Міри ефективності і проблема зупинки побудови дерева рішень *
21. Дерева рішень для регресії в Scikit-Learn Python, атрибути вузла. Алгоритм розбиття CART для регресії *
22. Алгоритм покриття побудови дерев рішень. 1R алгоритм *

23. Основи байесівської класифікації. Максимум апостеріорної імовірності. Штраф за помилку. Середній ризик. Оцінювання апріорних імовірностей та функцій правдоподібності
24. Наївний байесівський класифікатор (алгоритм Naive Bayes) *
25. Оцінка ефективності класифікації: перехресна перевірка, K-fold та його модифікації, решітчатий і рандомізований пошук *
26. Оцінка ефективності класифікації: confusion matrix, accuracy, precision, recall, PR-curve, F1 score, ROC крива *
27. Постановка задачі розділу суміші. Базовий алгоритм EM: ідея та етапи *
28. Обгрунтування базового алгоритму EM (виведення формул)
29. Алгоритм EM з фіксованою кількістю компонент. Недоліки алгоритму
30. Узагальнений алгоритм EM. Стохастичний алгоритм EM
31. Алгоритм EM з послідовним додаванням компонент.
32. Постановка задачі класифікації на два класи, поняття відступу, функціоналу помилок та функції втрат. Лінійний класифікатор.
33. Методи градієнтного спуску і стохастичного градієнтного спуску. *
34. Алгоритм стохастичного градієнтного спуску: етапи, оцінка функціоналу якості.
35. Ініціалізація ваг, порядок пред'явлення об'єктів в алгоритмі стохастичного градієнта. Переваги і недоліки, проблема перенавчання цього алгоритму.
36. Загальна постановка задачі кластеризації. Функції відстані в задачах кластеризації. *
37. Агломеративний ієрархічний алгоритм найближчого сусіда. *
38. Розділяючий ієрархічний алгоритм DIANA.
39. Алгоритм k-середніх. Переваги і недоліки алгоритму k-середніх. *
40. Алгоритм нечітких k-середніх. Переваги і недоліки.
41. Алгоритм G-середніх.
42. Алгоритм знаходження зв'язних компонент. Переваги і недоліки.
43. Базові принципи побудови мінімального покриваючого дерева (МПД) «Жадібний» алгоритм побудови МПД. *
44. Алгоритми Крускала і Прима побудови мінімального покриваючого дерева. *
45. Алгоритм Борувки побудови мінімального покриваючого дерева.
46. Щільнісний алгоритм кластеризації MeanShift.*
47. Щільнісний алгоритм кластеризації DBSCAN.*
48. Ідея базового алгоритму FOREL.
49. Етапи базового алгоритму FOREL. Переваги і недоліки.
50. Ідея та етапи FOREL-3.
51. Алгоритм FOREL-4.
52. Вибір кількості кластерів.
53. Метрики якості кластеризації: Homogeneity, Completeness, V-measure, Silhouette Coefficient. *
54. Метрики якості кластеризації: Adjusted Rand Index, Adjusted Mutual Information.

55. Метрики якості кластеризації: Calinski-Harabasz Index, Davies-Bouldin index.
56. Постановка задачі кластеризації на основі мережі Кохонена. Опис мережі Кохонена. *
57. Алгоритм стохастичного градієнта для розрахунку центрів кластерів в мережі Кохонена.*
58. Поняття карти Кохонена. Метрики на карті. Навести ілюстрацію карти.*
59. Алгоритм навчання карти Кохонена.
60. Візуалізація карт Кохонена, типи карт. Переваги і недоліки методу карт Кохонена.
61. Означення нейронної мережі. Модель МакКаллока-Піттса. *
62. Функції активації в нейронних мережах: порогові, сигмоїдальна, гіперболічний тангенс. Їх властивості.
63. Побудова ансамблю моделей методом бустингу. Алгоритм AdaBoostClassifier. *
64. Побудова ансамблю моделей методом бегінгу.Алгоритм BaggingClassifier. *
65. Алгоритм GradientBoostingClassifier.
66. Побудова ансамблю моделей методом стекінгу. Алгоритм StackingClassifier. *
67. Побудова ансамблю моделей методом голосування (Voting). Алгоритм VotingClassifier. *