

Код та назва дисципліни українською мовою/ Назва дисципліни англійською мовою	1-E8-112-5 Практична статистика для Machine Learning/ Practical statistics for Machine Learning
Рекомендується для галузі знань	А – Освіта, Е – Природничі науки, математика та статистика, F – Інформаційні технології, G – Інженерія, виробництво та будівництво, 09 Біологія, 10 Природничі науки, 11 Математика та статистика, 12 Інформаційні технології, 13 Механічна інженерія, 14 Електрична інженерія, 16 Хімічна інженерія та біоінженерія, 17 Електроніка, автоматизація та електронні комунікації, 18 Виробництво та технології, 014.04 Середня освіта (Математика), 014.08 Середня освіта (Фізика та астрономія)
Кафедра	Статистики й теорії ймовірностей
П.І.П. НПП	Бондаренко Яна Сергіївна
Рівень ВО	Перший (бакалаврський)
КУРС, семестр (в якому буде викладатись)	2-4 курси, парний
Мова викладання	Українська
Пререквізити (передумови вивчення дисципліни)	Знання з дисциплін «Математичний аналіз» або «Вища математика», «Інформатика та програмування» або «Програмування» або «Алгоритмізація та програмування»
Чому це цікаво/треба вивчати	У сучасному світі цифрових технологій студенти дедалі частіше шукають перспективні та актуальні професії, які відповідають викликам часу. Одним із найзатребуваніших напрямів є машинне навчання (Machine Learning) – сфера, що лежить в основі штучного інтелекту, аналітики та автоматизованих рішень. Фахівці з Machine Learning вже багато років посідають провідні позиції у рейтингах кадрових агентств і залишаються одними з найперспективніших професіоналів на світовому ринку праці. Постійний дефіцит кваліфікованих спеціалістів створює широкі можливості для студентів, які прагнуть побудувати кар'єру у сфері аналітики, прогнозування та оптимізації бізнес-процесів. Ця професія гарантує високу конкурентну заробітну плату, відкриває двері до міжнародного ринку праці та дозволяє безпосередньо впливати на розвиток інноваційних рішень у світі даних. Машинне навчання – це не просто тренд, а фундамент сучасного цифрового бізнесу, фінансів та технологій.
Перелік тем з дисципліни	Розділ 1. Основи статистичного навчання Тема 1. Визначення та роль статистичних методів у машинному навчанні Тема 2. Типи навчання: контрольоване, неконтрольоване Тема 3. Робота з навчальними та тестовими вибірками Тема 4. Підготовка даних: нормалізація, масштабування, заповнення пропущених значень Тема 5. Основні математичні та ймовірнісні моделі у статистичному навчанні Розділ 2. Лінійні та логістичні моделі Тема 1. Лінійна регресія Тема 2. Логістична регресія Тема 3. Поліноміальна регресія

	Тема 4. Регуляризація моделей: Ridge, Lasso, ElasticNet Тема 5. Визначення якості прогнозів та оцінка точності Розділ 3. Ансамблеві методи статистичного навчання Тема 1. Побудова ансамблів моделей Тема 2. Беггінг (Bagging): випадкові ліси, усереднення моделей Тема 3. Бустинг (Boosting): XGBoost, LightGBM, AdaBoost Тема 4. Стекінг та блендінг моделей Тема 5. Оптимізація ансамблевих методів для підвищення точності Розділ 4. Метрики оцінки якості моделей Тема 1. Оцінка якості регресійних моделей, метрики MSE, RMSE, MAE Тема 2. Оцінка якості класифікаційних моделей, метрики Accuracy, Precision, Recall, F1-score, ROC-AUC, PR-AUC Тема 3. Дисбаланс класів. Методи балансування класів Oversampling, Undersampling, Class Weights, Augmentation Тема 4. Перехресна валідація. Основні види перехресної валідації Hold-out validation, K-fold Cross-Validation, Stratified K-fold, Leave-One-Out (LOO) Cross-Validation. Перевірка узагальнення моделі Тема 5. Типи помилок у статистичних моделях. Методи аналізу помилок. Інтерпретованість статистичних моделей Розділ 5. Кластеризація та аналіз груп даних Тема 1. Поняття кластеризації та її роль у аналізі даних. Відмінність між класифікацією та кластеризацією. Тема 2. K-means алгоритм. Вибір оптимальної кількості кластерів. Недоліки K-means. Тема 3. Ієрархічна кластеризація та її можливості. Дендрограма: графічне представлення кластерів. Тема 4. DBSCAN алгоритм кластеризація на основі щільності. Порівняння DBSCAN з K-means: коли варто використовувати кожен. Тема 5. Візуалізація кластерів та аналіз груп
Як можна користуватися набутими знаннями і уміннями	Здатність розв'язувати практичні задачі в галузі бізнесу.
Очікувані результати навчання	Після опанування дисципліни ви зможете самостійно проводити статистичний аналіз, застосовувати машинне навчання для розв'язання реальних задач та використовувати методи оцінювання та інтерпретації моделей.
Інформаційне забезпечення	Робоча програма навчальної дисципліни, інтернет-ресурси
Види навчальних занять	Лекції, практичні заняття
Вид семестрового контролю	Диференційований залік
Максимальна кількість здобувачів на семестр	без обмежень

Декан факультету _____ Олександр ХАМІНІЧ