


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Механіко-математичний факультет
Кафедра математичної статистики і диференціальних рівнянь

Затверджено

На засіданні кафедри математичної
статистики
і диференціальних рівнянь
механіко-математичного
факультету
Львівського національного
університету
імені Івана Франка
(протокол №1 від 28.08.2020 р.)

В.о. завідувача кафедри:

 _____ Олег БУГРИЙ

Силабус з навчальної дисципліни
“ Тематичне моделювання ”,
що викладається в межах ОПП “Статистичний аналіз даних”
першого (бакалаврського) рівня вищої освіти для здобувачів із
спеціальності 112 – Статистика

Львів 2020

Назва дисципліни	Тематичне моделювання
Адреса викладання дисципліни	м. Львів, вул. Університетська 1
Факультет та кафедра, за якою закріплена дисципліна	Механіко-математичного факультет Кафедра математичної статистики і диференціальних рівнянь
Галузь знань, шифр та назва спеціальності	11 – Математика і статистика 112 – Статистика
Викладачі дисципліни	Бокало Микола Михайлович, доктор фізико-математичних наук, професор кафедри математичної статистики і диференціальних рівнянь
Контактна інформація викладачів	mykola.bokalo@lnu.edu.ua
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю).
Інформація про дисципліну	Курс розроблено для ознайомлення студентів з основними тематичними моделями для аналізу текстів, текстових архівів документів, для аналізу зміни тем у наборах документів.
Коротка анотація дисципліни	Дисципліна “Тематичне моделювання” є вибірковою дисципліною зі спеціальності 112 -- Статистика для освітньої програми “Статистичний аналіз даних”, яка викладається в 7-ому семестрі в обсязі 5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Мета та цілі дисципліни	<i>Мета:</i> ознайомити з основними способами побудови моделей колекцій текстових документів, до яких тем належить кожен з документів. <i>Цілі:</i> викласти основні положення теорії статистичних моделей для знаходження прихованих тем, що зустрічаються в колекції документів.
Література для вивчення дисципліни	<ol style="list-style-type: none"> 1) Daud Ali; Li Juanzi; Zhou Lizhu; Muhammad Faqir (2010). Knowledge discovery through directed probabilistic topic models: a survey. In Proceedings of Frontiers of Computer Science in China. (вид. www.researchgate.net). 2) Christos Papadimitriou; Prabhakar Raghavan; Hisao Tamaki; Santosh Vempala (1998). Latent Semantic Indexing: A probabilistic analysis (вид. Proceedings of ACM PODS). Архів оригіналу за 9 травня 2013. 3) David M. Blei; Andrew Y. Ng; Michael I. Jordan (2003). Latent

	<p>Dirichlet Allocation (вид. Journal of Machine Learning Research). Архів оригіналу за 1 травня 2012.</p> <p>4) David Blei (2012). Introduction to Probabilistic Topic Models (вид. Communications of the ACM). с. 77–84. Архів оригіналу за 15 лютого 2017.</p> <p>5) David Blei; J.D. Lafferty (2009). Topic models (web). Архів оригіналу за 31 травня 2013.</p> <p>6) David Blei; J.D. Lafferty (2007). Introduction to Probabilistic Topic Models (вид. Annals of Applied Statistics). с. 17–35. doi:10.1214/07-AOAS114. Архів оригіналу за 15 лютого 2017.</p> <p>7) David Mimno (2012). Computational Historiography: Data Mining in a Century of Classics Journals (журнал) (вид. Journal on Computing and Cultural Heritag). doi:10.1145/2160165.2160168.</p> <p>8) Matthew L. Jockers (2010). Who's your DH Blog Mate: Match-Making the Day of DH Bloggers with Topic Modeling.</p>
Обсяг курсу	Загальний обсяг -- 150 годин. Аудиторних занять -- 64 год., з них 32 год. лекційних та 32 годин практичних годин. Самостійної роботи -- 86 год. Кредитів – 5.
Очікувані результати навчання	У результаті вивчення даного курсу студент буде: знати: методи та техніки тематичного моделювання для виявлення та аналізу тем у текстах. вміти: використовувати методи та техніки тематичного моделювання для виявлення та аналізу тем у текстах.
Ключові слова	Байєсівські методи, ймовірнісне моделювання, нейронна мережа
Формат курсу	Очний
Теми	Див. “Схема курсу”
Підсумковий контроль, форма	Залік в кінці 7-го семестру
Пререквізити	Для вивчення даного курсу студенти повинні мати базові знання з: - теорії ймовірностей - математичної статистики
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Лекції, практичні заняття, презентації та самостійна робота
Необхідне об-	Комп’ютер із необхідним програмним забезпеченням, доступ до Internet

ладнання	мережі.
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються так: контрольна робота №1 – 16 балів; контрольна робота №2 – 18 балів; контрольна робота №3 – 21 бал; колоквіум №1 – 8 балів; колоквіум №2 – 8 балів; виконання домашніх завдань – 24 бали; премія за активну роботу на заняттях – 5 балів. Всього – 100 балів.</p> <p>Академічна доброчесність: Очікується, що роботи студентів будуть оригінальними дослідженнями чи міркуваннями. Списування та втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в написанні завдань є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та практичні заняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися визначених термінів для виконання всіх видів робіт, передбачених курсом.</p> <p>Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Політика виставлення балів. Враховуються бали, набрані при поточному контролі та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання і т. ін.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</p> <p>Оцінювання колоквіуму та контрольних робіт відбувається шляхом оцінки письмових відповідей студента на поставлені запитання. Відсотки нарахування балів оцінювання відповіді на кожне запитання нараховуються за наступним співвідношенням:</p> <p>75-100% – тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно;</p> <p>50-75% – відтворюється значна частина розглянутої теми, проте присутні неточності та/або невідповідності;</p> <p>25-50% – виявлено множинні неточності та невідповідності, пояснення відсутні чи частково помилкові;</p> <p>0-25% – тему майже не розкрито, кількість викладеного матеріалу не відповідає загальним нормам обраного виду роботи.</p> <p>Критерії оцінювання результатів неформальної освіти: Нарахування балів відбувається за публікацію студентом тез доповідей на конференціях, наукових статей, за участь студента у діяльності наукових</p>

	<p>гуртків, семінарів, круглих столів, конкурсів, участь у заходах неформальної освіти, за отримання сертифікатів про проходження навчання на різних освітніх платформах (Coursera, Prometheus тощо), курсах провідних ІТ компаній за тематикою навчальної дисципліни. Кількість балів визначається відсотком покриття результатів відповідної активності до вимог результатів навчання з навчальної дисципліни.</p>
Питання для контролю	<p>Основи текстового аналізу</p> <ul style="list-style-type: none"> • Вступ до текстового аналізу • Основні поняття: термінологія та методи • Інструменти для обробки текстових даних <p>Байєсівські методи тематичного моделювання</p> <ul style="list-style-type: none"> • Введення в ймовірнісне моделювання • Модель LDA (Latent Dirichlet Allocation) • Реалізація LDA на прикладі <p>Оцінка та вибір тематичної моделі</p> <ul style="list-style-type: none"> • Метрики якості для тематичного моделювання • Порівняння різних тематичних моделей • Оптимізація параметрів моделі <p>Глибоке навчання для тематичного моделювання</p> <ul style="list-style-type: none"> • Застосування нейронних мереж для аналізу текстів • Word embeddings та їхнє використання • Порівняння глибоких методів з класичними тематичними моделями <p>Прикладні аспекти тематичного моделювання</p> <ul style="list-style-type: none"> • Застосування тематичного моделювання в індустрії та науці • Виявлення аномалій та важливих подій за допомогою тематичного аналізу • Етичні питання тематичного моделювання <p>Тематичне моделювання в реальних задачах</p> <ul style="list-style-type: none"> • Аналіз настрою на основі тематичного моделювання • Визначення асоціацій та зв'язків між темами • Проектна робота: Застосування тематичного моделювання до власних даних <p>Перспективи розвитку тематичного моделювання</p> <ul style="list-style-type: none"> • Тенденції у сфері тематичного моделювання • Перспективи вдосконалення методів та інструментів
Опитування	<p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p>

Схема курсу

Тиж- день	Лекції			Практичні заняття		СР К-ть год Л-ра
	Назва теми	К- сть год		Назва теми	К- сть год	
7-ий семестр						
1	Вступ до текстового аналізу	2		Інструменти для обробки текстових даних.	2	3 [1-8]
2	Основні поняття: термінологія та методи	2		Інструменти для обробки текстових даних.	2	3 [1-8]
3	Введення в ймовірнісне моделювання	2		Контрольна робота №1	2	3 [1-8]
4	Модель LDA (Latent Dirichlet Allocation)	2		Реалізація LDA на прикладі	2	3 [1-8]
5	Метрики якості для тематичного моделювання	2		Оптимізація параметрів моделі	2	3 [1-8]
6	Порівняння різних тематичних моделей	2		Оптимізація параметрів моделі	2	3 [1-8]
7	Застосування нейронних мереж для аналізу текстів	2		Word embeddings та їхнє використання	2	3 [1-8]
8	Порівняння глибоких методів з класичними тематичними моделями	2		Контрольна робота №2	2	3 [1-8]
9	Застосування тематичного моделювання в індустрії та науці	2		Виявлення аномалій та важливих подій за допомогою тематичного аналізу	2	3 [1-8]
10	Колоквіум №1	2		Виявлення аномалій та важливих подій за допомогою тематичного аналізу	2	3 [1-8]
11	Етичні питання тематичного	2		Проектна робота: Застосування	2	3

	моделювання. Аналіз настрою на основі тематичного моделювання			тематичного моделювання до власних даних		[1-8]
12	Визначення асоціацій та зв'язків між темами	2		Проектна робота: Застосування тематичного моделювання до власних даних	2	3 [1-8]
13	Колоквіум №2	2		Проектна робота: Застосування тематичного моделювання до власних даних	2	3 [1-8]
14	Тенденції у сфері тематичного моделювання	2		Проектна робота: Застосування тематичного моделювання до власних даних	2	3 [1-8]
15	Тенденції у сфері тематичного моделювання	2		Контрольна робота №3	2	3 [1-8]
16	Перспективи вдосконалення методів та інструментів	2		Підсумкове обговорення та рефлексія	2	3 [1-8]
Всього:	150	32			32	86