

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДВНЗ “ПРИКАРПАТСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАСИЛЯ СТЕФАНИКА”**

Факультет математики та інформатики

Кафедра математичного і функціонального аналізу

**СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
ТЕХНОЛОГІЇ BIG DATA**

Освітня програма “Прикладна та теоретична статистика”

Спеціальність 112 “Статистика”

Галузь знань 11 “Математика та статистика”

Затверджено на засіданні кафедри
Протокол № 1 від “27” серпня 2020 р.

ЗМІСТ

1. Загальна інформація
2. Анотація до курсу
3. Мета та цілі курсу
4. Результати навчання та компетентності
5. Організація навчання курсу
6. Система оцінювання курсу
7. Політика курсу
8. Рекомендована література

1. Загальна інформація					
Назва дисципліни	Технології Big Data				
Рівень вищої освіти	Другий (магістерський)				
Викладач (-і)	Дмитришин Роман Іванович				
Контактний телефон викладача					
Е-mail викладача	roman.dmytryshyn@pnu.edu.ua				
Формат дисципліни					
Обсяг дисципліни	6 кредитів ЄКТС				
Посилання на сайт дистанційного навчання	https://d-learn.pnu.edu.ua/				
Консультації	Очні консультації: згідно розкладу консультацій				
2. Анотація до курсу					
Курс “Технології Big Data” присвячений базовим поняттям великих даних і питанням аналізу великих даних та пов’язаних з ними технічні, концептуальні та етичні проблеми. Курс містить практичні виправи для ознайомлення студентів з форматом великих даних і практикою обробки та аналізу великих, складних структур даних.					
3. Мета та цілі курсу					
Мета та цілі курсу полягають у формуванні в студентів нових теоретичних знань і практичних навичок із обробки та аналізу великих і складних структур даних як інструменту для розв’язування прикладних проблем конкретних предметних областей реального світу.					
4. Результати навчання та компетентності					
Результати навчання:					
ПРН-9. Уміти використовувати в практичній діяльності та розробляти спеціалізоване статистичне програмне забезпечення.					
ПРН-12. Уміти знаходити науково-технічну інформацію із застосуванням раціональних способів пошуку, включаючи засоби електронних інформаційних мереж.					
ПРН-13. Уміти інтегрувати знання з різних галузей для розв’язання теоретичних та практичних задач і проблем.					
Компетентності:					
ЗК-6. Здатність до пошуку, обробки, аналізу та дослідження інформації з різних джерел.					
ФК-7. Здатність аналізувати статистичні алгоритми, оцінювати їхню обґрунтованість та ефективність.					
ФК-8. Здатність використовувати спеціалізовані мови програмування та програмне забезпечення для розв’язання задач дослідницького або практичного характеру.					
5. Організація навчання курсу					
Обсяг курсу					
Вид заняття			Загальна кількість годин		
лекції			30		
семінарські заняття / практичні / лабораторні			30		
самостійна робота			120		
Ознаки курсу					
Семестр	Спеціальність	Курс (рік навчання)		Нормативний / вибірковий	
3	112 Статистика	II		Вибірковий	
Тематика курсу					
Тема, план	Форма заняття	Література	Завдання, год	Вага оцінки	Термін виконання
Змістовий модуль 1. Вступ до Big Data					

Тема 1. Основні положення концепції Big Data. Що такі великі дані і чому ви майте? Що таке аналітика великих даних? Як отримати великі дані від традиційних джерел?	Лекція, практичне заняття	[1-10]	Опрацювання лекційного матеріалу, 2; розв'язування задач, 2.	0.2	За розкладом
Тема 2. BIG DATA, екосистема APACHE HADOOP. Основні концепції Hadoop®. Встановлення та налаштування: локально і в хмарі. HDFS: архітектура, реплікація, читання і запис даних. Команди HDFS.	Лекція, практичне заняття	[1-10]	Опрацювання лекційного матеріалу, 4; розв'язування задач, 2.	0.3	За розкладом
Тема 3. MapReduce в Hadoop. Робочий процес MapReduce. Hadoop MapReduce і HDFS. Виконання MapReduce. Архітектура і робота першої версії MapReduce. Hadoop API (типи, класи). WordCount (Configure Job, Mapper, Reducer). Reducer як Combiner. Типи даних в Hadoop. InputSplit, InputFormat, OutputFormat. Shuffle і Sort в Hadoop. Запуск і налагодження завдань. Hadoop Streaming. Streaming в MapReduce.	Лекція, практичне заняття	[1-10]	Опрацювання лекційного матеріалу, 6; розв'язування задач, 4.	0.5	За розкладом
Змістовий модуль 2. Основи аналізу Big Data					
Тема 4. Вступ в Pig і Hive. Що таке Pig, для чого використовується? Pig і MapReduce. Основні можливості. Компоненти. Режими виконання. Запуск Pig. Pig Latin. Операції DUMP і STORE.	Лекція, практичне заняття	[1-10]	Опрацювання лекційного матеріалу, 6; розв'язування задач, 6.	0.4	За розкладом

Великий обсяг даних. Команда LOAD. Типи даних для схеми. Pig Latin (засоби діагностики, угруповання, Inner і Outer bag, FOREACH, функція TOKENIZE, оператор FLATTEN, WordCount, Inner і Outer Join). Hive (архітектура, інтерфейс, концепція, створення таблиці, завантаження даних, виконання запиту, Inner і Outer Join, WordCount).					
Тема 5. Apache Mahout. Що таке Apache Mahout? Реалізовані алгоритми. Класифікація (Naive Bayes, k-Means). Рекомендації (коллаборативна фільтрація, Item-based, алгоритм Slop One, Apache.teste, Item-based з Hadoop, Mahout зі Spark, co-occurrence recommenders).	Лекція, практичне заняття	[1-10]	Опрацювання лекційного матеріалу, 6; розв'язування задач, 6.	0.3	За розкладом
Тема 6. NoSQL, HBase, Cassandra. Масштабування вгору. Масштабування RDBMS (master / slave, sharding). Що таке NoSQL? Dynamo і BigTable. Теорія CAP. Модель консистентності. Eventual Consistency. Типи NoSQL. Key / Value.	Лекція, практичне заняття	[1-10]	Опрацювання лекційного матеріалу, 6; розв'язування задач, 6.	0.3	За розкладом
6. Система оцінювання курсу					
Загальна система оцінювання курсу	100 балів: 40 балів за змістовний модуль 1 (20 балів – модульна контрольна робота, 20 балів – опитування);				

	<p>60 балів за змістовний модуль 2 (30 балів – модульна контрольна робота, 30 балів – опитування).</p> <p>Критерії оцінювання знань, умінь і навичок студентів:</p> <p>90 – 100 (відмінно) – студент демонструє повні і глибокі знання навчального матеріалу, достовірний рівень розвитку умінь та навичок, правильне й обґрунтоване формулювання практичних висновків, наводить повний обґрунтований розв’язок прикладів та задач, аналізує причинно-наслідкові зв’язки; вільно володіє науковими термінами;</p> <p>70 – 89 (добре) – студент демонструє повні знання навчального матеріалу, але допускає незначні пропуски фактичного матеріалу, вміє застосувати його до розв’язання конкретних прикладів та задач, у деяких випадках нечітко формулює загалом правильні відповіді, допускає окремі несуттєві помилки та неточності розв’язках;</p> <p>50 – 69 (задовільно) – студент володіє більшою частиною фактичного матеріалу, але викладає його не досить послідовно і логічно, допускає істотні пропуски у відповіді, не завжди вміє правильно застосувати набуті знання до розв’язання конкретних прикладів та задач, нечітко, а інколи й невірно формулює основні твердження та причинно-наслідкові зв’язки;</p> <p>0 – 49 (незадовільно) – студент не володіє достатнім рівнем необхідних знань, умінь, навичок, науковими термінами.</p>
Вимоги до письмової роботи	Відповідно до робочого навчального плану, студент виконує дві контрольні роботи. Головна мета їх – перевірка самостійної роботи студентів у процесі навчання, виявлення ступеня засвоєння ними теоретичних положень курсу. При розв’язанні задач студент має детально вказувати, яким саме був хід його роздумів, якими формулами і твердженнями він користувався.
Семінарські заняття	Практичні заняття проводиться з метою формування у студентів умінь і навичок з курсу, вирішення сформульованих завдань, їх перевірка та оцінювання. За метою і структурою практичні заняття є ланцюжком, який пов’язує теоретичне навчання і навчальну практику з курсу, а також передбачає попередній контроль знань студентів. Оцінки за практичні заняття враховуються при виставленні підсумкової оцінки з курсу.
Умови допуску до підсумкового контролю	Оцінка за опитування і модульні контрольні роботи має бути не менша, ніж 25 балів .
7. Політика курсу	
<p>Самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей). Посилання на джерела інформації у разі використання ідей, розробок, тверджень, відомостей.</p> <p>Надання достовірної інформації про результати власної навчальної (наукової, творчої) діяльності, використанні методики досліджень і джерела інформації.</p> <p>Засвоєння пропущеної теми лекції з поважної причини перевіряється під час складання підсумкового контролю. Пропуск лекції з неповажної причини відпрацьовується студентом відповідно вимог кафедри, що встановлені на засіданні кафедри (співбесіда, реферат тощо).</p>	

Пропущені практичні заняття, незалежно від причини пропуску, студент відпрацьовує згідно з графіком консультацій. Поточні незадовільні оцінки, отримані студентом під час засвоєння відповідної теми на практичному занятті, перескладаються викладачеві, який веде заняття до складання підсумкового контролю з обов'язковою відміткою у журналі обліку роботи академічних груп.

8. Рекомендована література

1. Вайгенд А. BIG DATA. Вся технология в одной книге. Москва, Эксмо, 2018.
2. Волкова С. Просто BIG DATA. СПб., Страта, 2019.
3. Ын А., Су К. Теоретический минимум по Big Data. Всё что нужно знать о больших данных. СПб., Питер, 2019.
4. Радченко И.А., Николаев И.Н. Технологии и инфраструктура Big Data. СПб., Университет ИТМО, 2018.
5. Akerkar R. Models of Computation for Big Data Cham. Springer International Publishing, 2018.
6. Ghavami P. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. CreateSpace Independent Publishing Platform, 2016.
7. Feeney K. et al. Engineering Agile Big-Data Systems. River Publishers, 2018.
8. Raheem N. Big Data: A Tutorial-Based Approach. Taylor & Francis Group LLC. CRC Press, 2019.
9. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. Springer, 2020.
10. Wiktorski T. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. Springer, 2019.

Викладач _____