

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
“ДНІПРОВСЬКА ПОЛІТЕХНІКА”**



**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
Кафедра інформаційних технологій та комп'ютерної інженерії**

Кожевников А.В.

**Програма та методичні рекомендації
щодо виконання курсового проекту
з дисципліни**

**“Методи та інформаційні технології обробки великих даних
(Big Data)”**

**для студентів четвертого курсу навчання бакалаврата
спеціальності 126 “Інформаційні системи та технології”**

**Дніпро
НТУ “ДП”
2020**

УДК 378.14:371.14

Програма та методичні рекомендації щодо виконання курсового проекту з дисципліни “Методи та інформаційні технології обробки великих даних (Big Data)” для студентів четвертого курсу навчання бакалаврата спеціальності 126 “Інформаційні системи та технології” / Кожевников А.В. – Д.: НТУ «ДП», 2020. – 15 с.

Наведено програму курсового проектування з дисципліни “Методи та інформаційні технології обробки великих даних (Big Data)”, а також методичні рекомендації щодо її організації та проведення. Розкриті компоненти, що сприяють формуванню професійних компетентностей, а також підкріпленню комплексу компетентностних характеристик студентів (знань, умінь, комунікацій, автономності та відповідальності) у процесі виконання курсового проекту.

Програма та методичні вказівки містять загальні положення та рекомендації щодо складання звіту про виконання курсового проекту.

Вказівки призначені для студентів спеціальності 126 “Інформаційні системи та технології”, галузі 12 “Інформаційні технології”.

Погоджено рішенням методичної комісії спеціальності 126 “Інформаційні системи та технології” (протокол № 7 від 27.08.2020).

ЗМІСТ

ВСТУП	4
1 МЕТА ТА ЗАДАЧІ КУРСОВОГО ПРОЕКТУ	4
2 ОРГАНІЗАЦІЯ ПРОВЕДЕННЯ КУРСОВОГО ПРОЕКТУВАННЯ ТА ТЕХНІКА БЕЗПЕКИ	6
3 ПОРЯДОК ВИКОНАННЯ ПРОЕКТУ ТА ВИМОГИ ДО ОФОРМЛЕННЯ ЗВІТУ	7
4 ВАРІАНТИ ЗАВДАНЬ	9
5 КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ	10
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	13

ВСТУП

Характерною рисою сучасного світа є бурхливий розвиток інформаційних технологій, їх проникнення в усі сфери людського життя та стрімке зростання об'ємів даних з якими вони оперують. Тому опанування практикою використання методів та інформаційних технологій аналізу великих даних (Big Data) та розробки програмних інструментів для їх обробки є актуальними.

В процесі виконання курсового проекту приділяється увага як процесу аналізу великих даних, так і проектування та кодування програми з їх обробки.

Тривалість курсового проекту визначається затвердженим навчальним планом за спеціальністю 126 “Інформаційні системи та технології”. Конкретні терміни проведення курсового проекту встановлюються графіком навчального процесу, який затверджується щорічно керівництвом НТУ “Дніпровська політехніка”.

1 МЕТА ТА ЗАДАЧІ КУРСОВОГО ПРОЕКТУ

Метою курсового проекту є практичне закріплення теоретичних знань та навичок, придбаних в ході вивчення дисципліни “Методи та інформаційні технології обробки великих даних (Big Data)”.

Згідно освітньо-професійної програми (ОПП) з підготовки магістрів спеціальності 126 “Інформаційні системи та технології”, інтегральна компетентність полягає в здатності “розв’язувати складні спеціалізовані задачі та практичні проблеми промисловості або навчання у галузі інформаційних технологій, що передбачає проведення досліджень з ефективного використання інформаційних технологій для потреб вітчизняної науки та виробництва, а також використання новітніх інформаційних технологій для розробки нових інформаційних систем для потреб суспільства”. Методи та інформаційні технології обробки великих даних надають певні інструменти, які використовуються для розв’язання широкого кола задач як для потреб науки, так і виробництва. Безпосередньо курсовий проект готує студентів для розробки певної частини кваліфікаційної випускової роботи бакалавра.

До спеціальних, фахових компетентностей ОПП, що пов’язані з

цим курсовим проектом, можна віднести:

КС 4. Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші).

КС 6. Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків.

КС 13. Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень.

Основними результатами навчання з курсового проекту можна вважати:

ПР4. Проводити системний аналіз об'єктів проектування та обґрунтовувати вибір структури, алгоритмів та способів передачі інформації в інформаційних системах та технологіях.

ПР6. Демонструвати знання сучасного рівня технологій інформаційних систем, практичні навички програмування та використання прикладних і спеціалізованих комп'ютерних систем та середовищ з метою їх запровадження у професійній діяльності.

ПР9. Здійснювати системний аналіз архітектури підприємства та його ІТ-інфраструктури, проводити розроблення та вдосконалення її елементної бази і структури.

ПР12. Використовувати сучасні методи розробки програмного забезпечення, яке є складовою інформаційних систем та технологій.

ПР14. Розробляти комп'ютерні системи розпізнавання та обробки зображень, зокрема аерокосмічних зображень, обробки просторових даних, а також програмне забезпечення різноманітної спрямованості.

Таким чином, основними задачами курсового проектування з методів та інформаційних технологій обробки великих даних в рамках спеціальності 126 “Інформаційні системи та технології”, враховуючи ОПП, виступають:

- закріплення теоретичних знань щодо методів аналізу масивів (таблиць) великих даних шляхом використання інструментів електронних таблиць EXCEL та аналітичного середовища Statistica;

- отримання навичок розробки програмних інструментів зберігання та обробки великих даних з використанням засобів мови Python.

2 ОРГАНІЗАЦІЯ ПРОВЕДЕННЯ КУРСОВОГО ПРОЕКТУВАННЯ ТА ТЕХНІКА БЕЗПЕКИ

Згідно навчальних планів, затверджених керівництвом НТУ “Дніпровська політехніка”, з 2020 року курсове проектування виконується студентом як на базі обладнання комп’ютерних класів випускової кафедри, так і на власному обладнанні за бажанням студента. Зокрема, використовуються комп’ютери на базі процесорів Intel Core із шостого покоління або більш новітні та потужніші з тактовою частотою процесорів 2,3 ГГц та вище і з об’ємами оперативної пам’яті 4 Гб або більше.

Для виконання поставлених задач курсового проектування студент використовує програмне забезпечення:

- операційна система (ОС) Windows 7 (або новіша);
- Microsoft EXCEL (або новіший);
- Statistica 13.3 (або новіша);
- Anaconda на базі Python 2.7 (або новішого);
- Oracle VM VirtualBox 6.1.16 (або новіша);
- HDP Sandbox 2.5.0 (або новіша);
- програма-браузер, наприклад, Mozilla FireFox.

Під час дії карантинних заходів студенти виконують завдання курсового проекту у віддаленому режимі за місцем самоізоляції. При цьому керівник курсового проекту періодично спілкується зі студентами за допомогою різноманітних програм, наприклад таких, як Microsoft Teams, або електронна пошта.

До кінця терміну виконання курсового проекту, студент зобов’язаний скласти звіт, який містить всі етапи і результати аналізу набору даних згідно варіанту завдання та створення фрагментів програм обробки великих даних, а також захистити його викладачеві. Викладач, згідно критеріїв оцінювання, зараховує студенту відповідні бали.

На початку курсу студенти обов’язково проходять інструктаж з техніки безпеки та охорони праці.

Студенти зобов’язані:

- 1) пройти індивідуальний інструктаж з техніки безпеки на робочому місці в комп’ютерному класі кафедри;
- 2) дотримуватися інструкцій з охорони праці для працівників своєї

професії;

3) при зміні робочого місця пройти додатковий інструктаж;

4) виконувати всі рекомендації з охорони праці, які надає викладач;

5) під час дії карантинних заходів, студент повинен дотримуватися всіх правил поведінки та безпеки, встановлених або рекомендованих Міністерством охорони здоров'я України, а також внутрішніми розпорядженнями та наказами керівництва НТУ “Дніпровська політехніка”.

Оскільки основна робота студента зосереджена за робочим місцем, що обладнано комп'ютером, який під'єднаний до електричної мережі, то перш за все потрібно виконувати всі норми та правила з техніки безпеки і охорони праці, які прописані в [1].

При багаторазовому порушенні студентом правил техніки безпеки та охорони праці, питання про подальше використання техніки комп'ютерного класу розглядається керівництвом кафедри та факультету з прийняттям відповідного рішення.

3 ПОРЯДОК ВИКОНАННЯ ПРОЕКТУ ТА ВИМОГИ ДО ОФОРМЛЕННЯ ЗВІТУ

Проект складається з теоретичної та практичної частин.

При виконанні теоретичної частини студент формує опис методу або алгоритму обробки великих даних, який вказується викладачем.

При виконанні практичної частини студент:

1. 1.1. Завантажує з репозиторію набір даних для обробки згідно варіанту завдання.
- 1.2. З'ясовує природу та предметну область використання даних. Імпортує дані в середовище EXCEL, проводить їх кореляційний аналіз.
- 1.3. Проводить стандартизацію даних. Для факторної ознаки, яка найбільш тісно корелює з результативною, будує парну лінійну регресію. Проводить перевірку значущості рівняння регресії та її параметрів.
- 1.4. Імпортує стандартизовані дані середовище Statistica. Проводить кореляційний та регресійний аналіз аналогічно п. 1.2., 1.3. Порівнює отримані результати.
- 1.5. Проводить множинний лінійний регресійний аналіз даних

методом покрокового включення факторів.

1.6. Проводить кластерний аналіз даних методом k-середніх та ієрархічним агломеративним методами.

2. 2.1. Встановлює та налаштовує на власному комп'ютері середовище Anaconda на базі Python.

2.2. Розробляє та реалізує мовою Python програму, яка дозволяє проводити імпорт, стандартизацію даних набору даних, що опрацьовується, їх парний лінійний регресійний аналіз та кластерний аналіз методом k-середніх аналогічно п. 1.4., 1.6. Порівнює отримані результати.

3. 3.1. Встановлює на власному комп'ютері віртуальні машини Oracle VM VirtualBox та HDP Sandbox та налаштовує діалог з сервісом Hadoop.

3.2. Розробляє та реалізує мовою Python програму, яка дозволяє проводити розміщення у файловій системі Hadoop набору даних, що опрацьовується та його зворотне завантаження,

Звіт оформлюється у вигляді окремого документу державною мовою, згідно загальних правил оформлення звітної документації, складається від третьої особи, містить певний варіант завдання та включає наступні елементи:

Титульний лист

Зміст

Вступ

Розділ 1. Теоретична частина

Розділ 2. Практична частина

Висновки

Список використаних джерел

Звіт починається з титульного листа, який представлений у додатку.

Всі елементи оформлення тексту звіта (висота та розмітка, подання рисунків, формул та таблиць) відповідають вимогам оформлення, які надані в [2, 3].

У вступі студент формулює мету курсового проекту, а також визначає задачі, які необхідно вирішити в процесі його виконання. Також наводиться перелік програмного забезпечення, яке використовується при проектуванні.

В першому розділі наводиться опис обраного методу або алгоритму

обробки великих даних. Опис повинен містити відповідний математичний апарат та ілюстративний матеріал: графічні схеми алгоритмів, діаграми, графіки, таблиці тощо. Доречним є наведення чисельних прикладів, які пояснюють метод (алгоритм).

В другому розділі опису реалізації проекту студент наводить:

стилий опис набору даних, що опрацьовувався та результати його аналізу за п. 1.2 – 1.6;

покроковий опис встановлення програмного забезпечення за п. 2.2., 3.2 у вигляді інструкцій користувача;

програмні коди, які були створені при реалізації п. 2.2., 3.2 проекту, в кодах обов'язково повинні бути коментарі;

порівняння результатів аналізу даних, отриманих за допомогою різних програмних засобів.

При опису розробки у другому розділі рекомендується користуватися матеріалами лекційного курсу “Проектування інформаційних систем” [4]. Доречним використання матеріалів [5, 6]. Необхідне для виконання проекту програмне забезпечення може бути завантажено з джерел [8-10].

У висновках студент надає стислий опис основних результатів роботи.

Обов'язковим є надання списку використаних джерел, на які студент робив посилання у тексті звіту.

4 ВАРІАНТИ ЗАВДАНЬ

Студент опрацьовує варіант завдання, наданий у таблиці 4.1, який збігається з його порядковим номером у журналі обліку академічної групи (за потреби варіант завдання може бути виданий для виконання двома студентами при великій кількості студентів в академічній групі). Набори даних для обробки завантажуються з репозиторію [7].

Таблиця 4.1

Варіанти завдань на курсове проектування

Номер варіанту	Набір даних для обробки
1	Wine+Quality
2	Auto+MPG
3	Computer+Hardware
4	Yacht+Hydrodynamics

5	Real+estate+valuation+data+set
6	Automobile
7	Forest+Fires
8	Concrete+Slump+Test
9	Concrete+Compressive+Strength
10	Gas+Turbine+CO+and+NOx+Emission+Data+Set
11	Solar+Flare
12	Algerian+Forest+Fires+Dataset++
13	Iranian+Churn+Dataset
14	Energy+efficiency
15	Bike+Sharing+Dataset
16	Airfoil+Self-Noise
17	Combined+Cycle+Power+Plant
18	Seoul+Bike+Sharing+Demand
19	Condition+Based+Maintenance+of+Naval+Propulsion+Plants
20	Real+estate+valuation+data+set
21	Residential+Building+Data+Set
22	Liver+Disorders

5 КРИТЕРІЇ ОЦІНЮВАННЯ РЕЗУЛЬТАТІВ

Оцінювання результатів курсового проектування студентів здійснюється за допомогою прозорих процедур, що ґрунтуються на об'єктивних критеріях відповідно до Положення університету «Про оцінювання результатів навчання здобувачів вищої освіти».

Досягнутий рівень компетентностей відносно очікуваних, що ідентифікований під час контрольних заходів, відображає реальний результат навчання студента.

Оцінювання навчальних досягнень студентів НТУ «Дніпровська політехніка» здійснюється за рейтинговою (100-бальною) та інституційною шкалами (табл. 5.1). Остання необхідна (за офіційною відсутністю національної шкали) для конвертації (переведення) оцінок мобільних студентів.

Таблиця 5.1

Шкали оцінювання навчальних досягнень
студентів НТУ “Дніпровська політехніка”

<i>Рейтингова</i>	<i>Інституційна</i>
90...100	відмінно / Excellent
74...89	добре / Good
60...73	задовільно / Satisfactory
0...59	незадовільно / Fail

Курсовий проект зараховується, якщо студент отримав підсумкову оцінку не менше 60-ти балів. Нижча оцінка вважається академічною заборгованістю, що підлягає ліквідації відповідно до Положення про організацію освітнього процесу НТУ “Дніпровська політехніка”.

Студент на контрольних заходах має надати звіт про виконане курсового проекту, орієнтований виключно на демонстрацію дисциплінарних результатів навчання.

Основні критерії оцінювання результату захисту звіту з курсового проекту представлені в таблиці 5.2.

Таблиця 5.2

Критерії оцінювання результатів виконання курсового проекту

<i>№ з/п</i>	<i>Результат виконання</i>	<i>Рейтингова кількість балів</i>	<i>Інституційна оцінка</i>
1.	- звіт містить текст з детальним описом розробки в розділах 1 та 2, а також висновки; - всі програмні фрагменти виконують операції за завданням.	90...100	відмінно / Excellent
2.	- звіт містить незначні відхилення від правил оформлення; - студент при захисті дав невпевнені відповіді на запитання викладача або помилявся; - всі програмні фрагменти виконують операції за завданням.	74...89	добре / Good

3.	<ul style="list-style-type: none"> - звіт містить значні відхилення від правил оформлення; - студент при захисті погано володіє описом створеної розробки; - не всі програмні фрагменти виконують операції за завданням. - звіт поданий до захисту несвоєчасно. 	60...73	задовільно / Satisfactory
4.	<ul style="list-style-type: none"> - програмні фрагменти не працюють; - звіт про виконання відсутній; - студент не з'являвся на залік. 	0...59	незадовільно / Fail

Кінцева кількість балів з визначених в таблиці 5.2 діапазонів встановлюється з врахуванням якості захисту виконаного звіту та вірних відповідей на поставлені викладачем запитання, відповідності сформованим компетентностям навчання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Вимоги щодо безпеки та захисту здоров'я працівників під час роботи з екранними пристроями. Затв. Наказом Міністерства соціальної політики України 14.02.2018. Офіційний вісник України від 18.05.2018 – 2018 р., No 38, стор. 121, стаття 1352, код акта 90123/2018 (URL: <https://zakon.rada.gov.ua/laws/show/z0508-18>).
2. Методичні рекомендації до виконання та оформлення кваліфікаційних робіт бакалаврів / Гаркуша І.М., Гнатушенко В.В., Коротенко Г.М. – Д.: НТУ «ДП», 2020. –27 с.
3. Програма та методичні рекомендації щодо виконання проектно-технологічної практики для студентів третього курсу навчання спеціальності 126 “Інформаційні системи та технології” / Гаркуша І.М. – Д.: НТУ «ДП», 2020. – 18 с.
4. Гаркуша І.М. Конспект лекцій з дисципліни “Проектування інформаційних систем”. [Електрон. ресурс]. Режим доступа: <https://do.nmu.org.ua/enrol/index.php?id=1167> (дата звернення: 20.08.2020).
5. Силен Д. Основы Data Science и Big Data. Python и наука о данных [Текст] /Д. Силен, А. Мейсман, М. Али // СПб.: Питер, 2017, 336 с.
6. Форман Дж. Много цифр: Анализ больших данных при помощи Excel [Текст] / Дж. Форман // М.: Альпина Паблишер, 2016, 464 с.
7. UCI Machine Learning Repository [Электрон. ресурс]. Режим доступа: <http://archive.ics.uci.edu/ml/index.php/> (дата звернення: 20.08.2020).
8. Download VirtualBox [Электрон. ресурс]. Режим доступа: <https://www.virtualbox.org/wiki/Downloads> (дата звернення: 20.08.2020).
9. Hortonworks Data Platform (HDP) on Hortonworks Sandbox [Электрон. ресурс]. Режим доступа: <https://www.cloudera.com/downloads/hortonworks-sandbox.html#install> (дата звернення: 20.08.2020).
10. Statistica Data Miner пробная версия Statistica [Электрон. ресурс]. Режим доступа: <http://statsoft.ru/products/trial/trial.php>.

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
“ДНІПРОВСЬКА ПОЛІТЕХНІКА”**



**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
Кафедра інформаційних технологій та комп'ютерної інженерії**

КУРСОВИЙ ПРОЕКТ

з дисципліни

“Методи та інформаційні технології обробки великих даних (Big Data)”

Тема: “ Аналіз масивів великих даних за допомогою інструментів електронних таблиць EXCEL, аналітичного середовища STATISTICA та засобів мови Python.”

Виконав:

ст. гр. 126-17-1
Шевченко Т.Г.

(прізвище, ініціали, підпис)

Перевірив:

доц. каф. ІТКІ
Кожевников А.В.

(прізвище, ініціали, підпис)

Дата захисту: _____

Рейтингова кількість балів: _____

Інституційна оцінка: _____

Дніпро
НТУ “ДП”
2020

Навчальне видання

Кожевников Антон Вячеславович

ПРОГРАМА ТА МЕТОДИЧНІ РЕКОМЕНДАЦІЇ
щодо виконання курсового проекту
з дисципліни
“Методи та інформаційні технології обробки великих даних
(Big Data)”
для студентів четвертого курсу навчання бакалаврата
спеціальності 126 “Інформаційні системи та технології”

Електронний ресурс

Видано
у Національному технічному університеті
«Дніпровська політехніка».
Свідоцтво про внесення до Державного реєстру ДК № 1842 від 11.06.2004.
49005, м. Дніпро, просп. Дмитра Яворницького, 19.