

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ

Кафедра теорії та технології програмування

«ЗАТВЕРДЖУЮ»

Заступник декана
з навчальної роботи

_____ Кашпур О.Ф.

«___» _____ 2017 року

**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
АКТУАЛЬНІ ПРОБЛЕМИ «DATA MINING»**

для студентів

галузь знань	12 «Інформаційні технології» <i>(шифр і назва)</i>
спеціальність	122 «Комп'ютерні науки» <i>(шифр і назва спеціальності)</i>
освітній рівень	магістр <i>(молодший бакалавр, бакалавр, магістр)</i>
освітня програма	«Штучний інтелект» <i>(назва освітньої програми)</i>
вид дисципліни	за вибором ВНЗ

Форма навчання

денна

Навчальний рік

2017/2018

Семестр

2

Кількість кредитів ECTS

5

Мова викладання, навчання
та оцінювання

англійська

Форма заключного контролю

екзамен

Викладач: **к.ф.-м.н., асистент Россада Т.В.** (лекції, лабораторні заняття)

Пролонговано: на 20__/20__ н.р. _____ (_____) «__» __ 20__ р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____ (_____) «__» __ 20__ р.
(підпис, ПІБ, дата)

КИЇВ – 2017

Робоча програма навчальної дисципліни «Data Mining Actual Problems» для студентів спеціальності 122 «Комп'ютерні науки» освітнього рівня магістр освітньої програми «Штучний інтелект»

«31» серпня 2017 року – 12 с.

Розробник: Россада Т.В., к.ф.-м.н., асист.

Робоча програма дисципліни «Data Mining Actual Problems» затверджена на засіданні кафедри теорії та технології програмування

Протокол №___ від «___» _____ 2017 року

В.О. Завідувача кафедри _____

Панченко Т.В.

Схвалено науково-методичною комісією факультету

Протокол №___ від «___» _____ 2017 року

Голова НМК _____

Хусаїнов Д.Я.

Робоча програма дисципліни «Data Mining Actual Problems» затверджена на засіданні Вченої ради факультету кібернетики

Протокол Протокол №___ від «___» _____ 2017 року

Вступ

Навчальна дисципліна «Актуальні проблеми «Data Mining» є складовою циклу професійної підготовки фахівців спеціальності 122 «Комп'ютерні науки» освітнього рівня магістр освітньої програми «Штучний інтелект».

Дисципліна «Актуальні проблеми «Data Mining» є дисципліною за вибором ВНЗ, що викладається на 1-му курсі магістратури у 2-му семестрі в обсязі 120 годин (5 кредитів ECTS), в тому числі 24 години лекцій, 12 годин лабораторних занять 2 години консультацій та 80 годин самостійної роботи. Викладання ведеться англійською мовою. Завершується дисципліна **екзаменом**.

Метою і завданням навчальної дисципліни є знайомство з теоретичними та прикладними аспектами технології Data Mining, методами інтелектуального аналізу даних, можливостями їх застосування.

Структура курсу. В результаті вивчення навчальної дисципліни студент повинен **знати** теоретичні та прикладні аспекти методів Data Mining та **вміти** їх застосовувати для розв'язання широкого спектру задач.

Для освоєння курсу необхідні базові знання з *програмування, теорії баз даних, вищої математики*.

Контроль знань і розподіл балів, які отримують студенти

Контроль знань студентів здійснюється за модульно-рейтинговою системою. Результати навчальної діяльності студентів оцінюються за 100-бальною шкалою.

Робота в семестрі поділяється на 2 змістових модуля. При виставленні балів за змістовий модуль враховується: робота на лабораторних заняттях – 10 балів, оцінка за модульну контрольну роботу – 15 балів. Завершується дисципліна іспитом.

Оцінювання за формами контролю:

	ЗМ1		ЗМ2	
	Min. – 18	Max. – 30	Min. – 18	Max. – 30
<i>Лабораторні роботи</i>	10	20	10	20
<i>Модульні контрольні роботи</i>	8	10	8	10

У випадку відсутності студента з поважних причин відпрацювання та перездачі МКР здійснюються у відповідності до «Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу» від 1 жовтня 2010 року.

При простому розрахунку отримаємо:

	ЗМ1	ЗМ2	Екзамен	Підсумкова оцінка
Мінімум	18	18	24	60
Максимум	30	30	40	100

При цьому, кількість балів:

- 1-34 відповідає оцінці «незадовільно» з обов'язковим повторним вивченням дисципліни;
- 35-59 відповідає оцінці «незадовільно» з можливістю повторного складання;
- 60-64 відповідає оцінці «задовільно» («достатньо»);
- 65-74 відповідає оцінці «задовільно»;
- 75 - 84 відповідає оцінці «добре»;
- 85 - 89 відповідає оцінці «добре» («дуже добре»);
- 90 - 100 відповідає оцінці «відмінно».

Шкала відповідності

За 100 – бальною шкалою	За національною шкалою
90 – 100	Відмінно
85 – 89	Добре
75 – 84	
65 – 74	Задовільно
60 – 64	
1 – 59	Не задовільно

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. Data Mining Tasks and Methods — 36 год.

Тема 1. Data Mining Tasks and Methods — 20 год.

Тема 2. Data Mining Standards — 16 год.

Змістовий модуль 2. Classification and Clusterization — 80 год.

Тема 3. Clasterization — 30 год.

Тема 4. Associative Rules Searching — 24 год.

Тема 5. Clasterization — 26 год.

СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
ТЕМАТИЧНИЙ ПЛАН ЛЕКЦІЙ ТА ЛАБОРАТОРНИХ ЗАНЯТЬ

№ п/п	Назва лекції	Кількість годин		
		лекції	лаборат.	сам.
Змістовий модуль 1. Data Mining Tasks and Methods				
1	Тема 1. Data Mining Tasks and Methods	8	2	10
2	Тема 2. Data Mining Standards	4	2	10
	<i>Модульна контрольна робота 1</i>			
Змістовий модуль 2. Classification and Clusterization				
5	Тема 3. Clasterization	6	4	20
6	Тема 4. Associative Rules Searching	2	2	20
7	Тема 5. Clasterization	4	2	20
	<i>Модульна контрольна робота 2</i>			
	ВСЬОГО	24	12	80

Загальний обсяг 120 год., в тому числі:

лекції – 24 год.

лабораторні заняття – 12 год.

самостійна робота – 80 год.

консультації — 2 год.

Змістовий модуль 1. Data Mining Tasks and Methods

Тема 1. Data Mining Tasks and Methods

Лекція 1. Introduction to the discipline. Concept and tasks of Data Mining – 2 год.

Лекція 2. Data set and their attributes. Data measurement – 2 год.

Лекція 3. Data processing. Map Reduce – 2 год.

Лабораторне заняття 1. Laboratory work No 1. – 2 год.

Лекція 4. Data Mining Methods and Tools – 2 год.

Самостійна робота. Laboratory work No 1. — 10 год.

Тема 2. Data Mining Standards

Лекція 1. Visualization Methods. - 2 год.

Лабораторне заняття 1. Laboratory work No 1.– 2 год.

Лекція 2. Data Mining Standards: CWM, CRISP, PMML. – 2 год.

Самостійна робота. Laboratory work No 1.— 10 год.

Контрольні запитання

1. The concept of Data Mining.
2. Concepts of the data. Scribes and their attributes.
3. Measurement of data.
4. Features processingData.
5. Map Reduce
6. VisualAnalizations. Methodology visualization
7. DataMining Standards: CWM
8. DataMining Standards: CRISP
9. DataMining Standards: PMML

Типове завдання модульної контрольної роботи No 1

1. Concepts of data. Data set and their attributes.
2. Data Mining Standards: CRISP
3. Describe map and reduce functions to find the largest number of sequence numbers.

Рекомендована література: [1-32]

Змістовий модуль 2. Classification and Clusterization

Тема 3. Clasterization– 30 год.

Лекція 1. Classification Task. Evalueting of classification — 2 год.

Лабораторне заняття 1. Laboratory work No 2. — 2 год.

Лекція 2. Classification Methods. Algorithm for constructing elementary rules (1-rule), Naive Bayes algorithm. — 2 год.

Лекція 3. Neural networks using in classification tasks. Decision Tree. Nearest Neighbor Algorithm — 2 год.

Лабораторне заняття 2. Laboratory work No 2. — 2 год.

Самостійна робота. Laboratory work No 2. — 20 год.

Тема 4. Associative Rules Searching - 24 год.

Лекція 1. Associative rules fsearching. The Apriori algorithms. — 2 год.

Лабораторне заняття 1. Laboratory work No 2 — 2 год.

Самостійна робота. Laboratory work No 2 — 20 год.

Тема 5. Clusterization — 26 год.

Лекція 1. Clusterization. Hierarchical clustering algorithms — 2 год.

Лекція 2. Algorithm k-Means. Method of the nearest neighbor — 2 год.

Лабораторне заняття 1. Laboratory work No 2 — 2 год.

Самостійна робота. Laboratory work No 2 — 20 год.

Рекомендована література: [1-32]

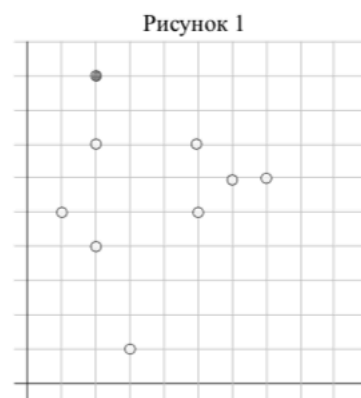
Контрольні запитання

1. The task of classification
2. Accuracy of classification. Estimation of the level of errors
3. Setting the task of finding associative rules, its varieties. Representation of results
4. Apriori algorithms
5. Setting the task of clusterization
6. Hierarchical clustering algorithms
7. Algorithm K-Means (Hard-c-means)
8. Method of the nearest neighbor
9. Trees decision-making
10. Algorithm of the nearest neighbor
11. Adaptive methods of clusterization.

Типове завдання модульної контрольної роботи № 2

1. Setting the task of classification. Algorithms for constructing elementary rules and Naive Bayes: theoretical foundations, basic stages, applications.
2. Application of neural networks for clustering tasks.
3. For data from tab. 1 build a decision tree ($S = 1$) using the theoretically-informative criterion for selecting the attribute and find the class of the last object.
4. Using the Apriori algorithm ($\text{minsup} = 3$), specify the associative rules for the data set from Table 2. Identify validation and support for each rule.
5. Break the data presented in Figure 1 into clusters using the k-means algorithm and using the nearest neighbor method (taking a threshold of 2).

Таблиця 1						Таблиця 2	
	q1	q2	q3	q4	S	1	
1	1	0	1	0	1	2	0, 1, 2, 3
2	1	1	0	1	0	3	1, 2, 3
3	0	0	0	1	0	4	1, 4, 5, 6, 7
4	0	0	1	0	1	5	1, 2, 3, 4, 6, 9
5	0	1	1	1	0	6	1, 2, 3, 4, 5
6	0	2	1	0	1	7	0, 5, 6
7	0	1	1	0	1	8	0, 8, 9
8	0	2	1	0	1	9	0, 1, 4, 7
9	0	1	0	0	1	10	5, 6, 7, 8, 9
10	1	0	0	1	0		4, 5, 7
11	0	1	0	0	?		



Питання на іспит

1. The concept of Data Mining.
2. Concepts of the data. Scribes and their attributes.
3. Measurement of data.
4. Features processingData.
5. Map Reduce
6. VisualAnalizations. Methodology visualization
7. DataMining Standards: CWM
8. DataMining Standards: CRISP
9. DataMining Standards: PMML
10. The task of classification
11. Accuracy of classification. Estimation of the level of errors
12. Setting the task of finding associative rules, its varieties. Representation of results
13. Apriori algorithms
14. Setting the task of clusterization
15. Hierarchical clustering algorithms
16. Algorithm K-Means (Hard-c-means)
17. Method of the nearest neighbor
18. Trees decision-making
19. Algorithm of the nearest neighbor
20. Adaptive methods of clusterization.

Рекомендована література

1. Марченко О. О., Россада Т.В. Актуальні проблеми Data Mining: навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с.
2. Leskovec J. Mining of Massive Datasets / Jure Leskovec Anand Rajaraman, Jeffrey David Ullman // Stanford Univ. — 2010.
3. Bradley, P., Fayyad, U., Reina, C. Scaling Clustering Algorithms to Large Databases, Proc. 4th Int'l Conf. Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, Calif., 1998.
4. Zhang, T., Ramakrishnan, R., Livny, M. Birch: An Efficient Data Clustering Method for Large Databases, Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1996.

5. Paul S. Bradley, Usama M. Fayyad, Cory A. Reina Scaling EM (Expectation-Maximization) Clustering to Large Databases, Microsoft Research, 1999.
6. Z. Huang. Clustering large data sets with mixed numeric and categorical values. In The First Pacific-Asia Conference on Knowledge Discovery and Data Mining, 1997.
7. Milenova, B., Campos, M. Clustering large databases with numeric and nominal values using orthogonal projections, Oracle Data Mining Technologies, 2002.
8. Z. Huang. A fast clustering algorithm to cluster very large categorical data sets in Data Mining. Research Issues on Data Mining and KDD, 1997.
9. Wang, K., Xu, C., Liu, B. Clustering transactions using large items. In Proc. CIKM'99, Kansas, Missouri, 1999.
10. Guha S., Rastogi R., Shim K. CURE: An Efficient Clustering Algorithm for Large Databases, Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1998.
11. Ganti V., Gerhke J., Ramakrishnan R. CACTUS – Clustering Categorical Data Using Summaries. In Proc KDD'99, 1999.
12. J. Bilmes. A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Tech. Report ICSI-TR-97-021, 1997.
13. Добыча данных в сверхбольших базах данных / В. Ганти, Й. Герке, Р. Рамакришнан // Открытые системы, No9-10, 1999.
14. Барсегян и др. Методы и модели анализа данных: OLAP и Data Mining. – СПб., 2004
15. Berry, Michael J. A. "Data mining techniques: for marketing, sales, and customer relationship management" / Michael J.A. Berry, Gordon Linoff. – 2nd ed.
16. Larose, Daniel T. "Discovering knowledge in data: an introduction to data mining" / Daniel T. Larose
17. J. Ross Quinlan. C4.5: Programs for Machine learning. Morgan Kaufmann Publishers 1993.
18. S.Murthy. Automatic construction of decision trees from data: A Multi-disciplinary survey.1997.
19. W. Buntine. A theory of classification rules. 1992.
20. Machine Learning, Neural and Statistical Classification. Editors D. Mitchie et.al. 1994.
21. К. Шеннон. Работы по теории информации и кибернетике. М. Иностранная литература, 1963
22. С.А. Айвазян, В.С Мхитарян Прикладная статистика и основы эконометрики, М. Юнити, 1998
23. Dirk Emma Baestaens, Willem Max Van Den Bergh, Douglas Wood, "Neural Network Solution for Trading in Financial Markets", Pitman publishing
24. R. M. Hristev, "Artificial Neural Networks"

25. R. Agrawal, T. Imielinski, A. Swami. 1993. Mining Associations between Sets of Items in Massive Databases. In Proc. of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data, 207-216.
26. R. Agrawal, R. Srikant. "Fast Discovery of Association Rules", In Proc. of the 20th International Conference on VLDB, Santiago, Chile, September 1994.
27. R. Srikant, R. Agrawal. "Mining Generalized Association Rules", In Proc. of the 21th International Conference on VLDB, Zurich, Switzerland, 1995.
28. R. Srikant, R. Agrawal. "Mining quantitative association rules in large relational tables". In Proceedings of the ACM SIGMOD Conference on Management of Data, Montreal, Canada, June 1996.
29. Savasere, E. Omiecinski, and S. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases", In Proc. 21st Int'l Conf. Very Large Data Bases, Morgan Kaufmann, San Francisco, 1995.
30. J.S. Park, M.-S. Chen, and S.Y. Philip, "An Effective HashBased Algorithm for Mining Association Rules", In Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1995.
31. S. Brin et al., "Dynamic Itemset Counting and Implication Rules for Market Basket Data", In Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1997.
32. J. Hipp, U. Guntzer, and G. Nakaeizadeh. Algorithms for Association Rule Mining – A General Survey and Comparison. In Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000.

Електронні джерела

1. Курс лекцій Николая Анохина (mail.ru DataGroup)
<https://www.youtube.com/playlist?list=PLrCZzMib1e9pyyrqknouMZbIPf4I3CwUP>
2. Data is the New Oil By Michael Palmer http://ana.blogs.com/maestros/2006/11/data_is_the_new.html
3. Анализ данных как область знания
<http://postnauka.ru/video/34960>
4. Материалы на тему анализа данных http://www.basegroup.ru/library/methodology/data_mining/
5. Наивный Байесовский классификатор в 25 строк кода
<http://habrahabr.ru/post/120194/>
6. Фильтрация смс спама с помощью наивного байесовского классификатора
<http://habrahabr.ru/post/184574/>
7. Лекции курса «Машинное обучение» от yandex
<https://yadi.sk/d/V9p7E6uAFjHcD>
8. Воронцов К. В. Лекции по алгоритмам кластеризации и многомерного шкалирования
<http://www.ccas.ru/voron/download/Clustering.pdf>

9. Котов А., Красильников Н. Кластеризация данных. 2006

<http://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf>

10. Информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных

www.machinelearning.ru/

11. Н.Ю. Золотых Как обучаются машины? научно-популярная лекция [http://](http://www.uic.unn.ru/~zny/ml/Pop/ml_pop.pdf)

www.uic.unn.ru/~zny/ml/Pop/ml_pop.pdf

12. Главы из книги на тему машинного обучения и презентации уроков Сергея Николенко

<http://logic.pdmi.ras.ru/~sergey/teaching/ml/>

Додаток до робочої програми
Завдання для самостійної роботи з елементами дистанційного навчання з
дисципліни «Актуальні проблеми Data Mining»
на період з 24 січня до 28 лютого 2018 р.
для студентів
1 курсу магістратури спеціальності 122 «Комп'ютерні науки»
освітньої програми «Штучний інтелект»

Викладач: к.ф.-м.н., асист. Россада Т.В. (e-mail: trossada@knu.ua)

Види та форми контрольних заходів з перевірки самостійної роботи студентів

Контроль за виконанням самостійної роботи студентами здійснюється у двох формах: у січні-лютому за допомогою електронних засобів (електронною поштою), у березні – шляхом проведення письмової контрольної роботи. Виконання самостійної роботи є допуском до написання контрольної роботи у березні 2018 р.

Впродовж січня-лютого (24 січня – 20 лютого 2018 р.) студенти мають вивчити запропоновані питання визначених тем на базовому рівні. Для підтвердження виконання завдання студенти мають надіслати відповіді на 3 теоретичних питання та виконану лабораторну роботу із звітом не пізніше **15 лютого 2018 р.** Завдання першого етапу, які мають бути виконані та надіслані на електронну пошту викладача (trossada@knu.ua), подано у **додатку 1.**

На контрольну роботу за підсумками самостійної роботи виносяться всі зазначені нижче теоретичні питання.

Теоретичні питання для самостійного опрацювання

1. The concept of Data Mining.
2. Concepts of the data. Scribes and their attributes.
3. Measurement of data.
4. Features processingData.
5. Map Reduce
6. VisualAnalizations. Methodology visualization
7. DataMining Standards: CWM
8. DataMining Standards: CRISP
9. DataMining Standards: PMML

Критерії оцінювання

Викладач оцінює виконані завдання в категоріях «зараховано» або «не зараховано». Щоб отримати оцінку «зараховано» потрібно набрати 4 і більше балів за відповіді на теоретичні питання та виконати лабораторну роботу. Відповіді на теоретичні питання оцінюються наступним чином: відповіді немає — 0 балів; відповідь є, але не повна — 1 бал; відповідь повна — 2 бали.

Якщо студент отримує оцінку «не зараховано», у нього є час до **20 лютого** переробити завдання та надіслати їх викладачу повторно.

Контрольна робота оцінюється максимум в **10 балів**. Вона включає в себе 5 тестових питань з проблематики, винесеної на самостійну роботу, та одне практичне завдання. Правильна відповідь на кожне тестове завдання оцінюється в 1 бал. За розв'язання задачі студент може отримати від 1 до 5 балів.

Контрольна робота проводиться на першому лекційному занятті з курсу у березні 2018 р. Її тривалість – 1 академічна година.

Рекомендована література

1. Марченко О. О., Россада Т.В. Актуальні проблеми Data Mining: навчальний посібник для студентів факультету комп'ютерних наук та кібернетики. — Київ. — 2017. — 150 с.
2. Leskovec J. Mining of Massive Datasets / Jure Leskovec Anand Rajaraman, Jeffrey David Ullman // Stanford Univ. – 2010.
3. Bradley, P., Fayyad, U., Reina, C. Scaling Clustering Algorithms to Large Databases, Proc. 4th Int'l Conf. Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, Calif., 1998.
4. Zhang, T., Ramakrishnan, R., Livny, M. Birch: An Efficient Data Clustering Method for Large Databases, Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1996.
5. Paul S. Bradley, Usama M. Fayyad, Cory A. Reina Scaling EM (Expectation-Maximization) Clustering to Large Databases, Microsoft Research, 1999.
6. Z. Huang. Clustering large data sets with mixed numeric and categorical values. In The First Pacific-Asia Conference on Knowledge Discovery and Data Mining, 1997.
7. Milenova, B., Campos, M. Clustering large databases with numeric and nominal values using orthogonal projections, Oracle Data Mining Technologies, 2002.
8. Z. Huang. A fast clustering algorithm to cluster very large categorical data sets in Data Mining. Research Issues on on Data Mining and KDD, 1997.
9. Wang, K., Xu, C., Liu, B. Clustering transactions using large items. In Proc. CIKM'99, Kansas, Missouri, 1999.
10. Guha S., Rastogi R., Shim K. CURE: An Efficient Clustering Algorithm for Large Databases, Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1998.
11. Ganti V., Gerhke J., Ramakrishnan R. CACTUS – Clustering Categorical Data Using Summaries. In Proc KDD'99, 1999.
12. J. Bilmes. A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Tech. Report ICSI-TR-97-021, 1997.
13. Добыча данных в сверхбольших базах данных / В. Ганти, Й. Герке, Р. Рамакришнан // Открытые системы, No9-10, 1999.
14. Барсегян и др. Методы и модели анализа данных: OLAP и Data Mining. – СПб., 2004

15. Berry, Michael J. A. "Data mining techniques: for marketing, sales, and customer relationship management" / Michael J.A. Berry, Gordon Linoff. – 2nd ed.
16. Larose, Daniel T. "Discovering knowledge in data: an introduction to data mining" / Daniel T. Larose
17. J. Ross Quinlan. C4.5: Programs for Machine learning. Morgan Kaufmann Publishers 1993.
18. S.Murthy. Automatic construction of decision trees from data: A Multi-disciplinary survey.1997.
19. W. Buntine. A theory of classification rules. 1992.
20. Machine Learning, Neural and Statistical Classification. Editors D. Mitchie et.al. 1994.
21. К. Шеннон. Работы по теории информации и кибернетике. М. Иностранная литература, 1963
22. С.А. Айвазян, В.С Мхитарян Прикладная статистика и основы эконометрики, М. Юнити, 1998
23. Dirk Emma Baestaens, Willem Max Van Den Bergh, Douglas Wood, "Neural Network Solution for Trading in Financial Markets", Pitman publishing
24. R. M. Hristev, "Artificial Neural Networks"
25. R. Agrawal, T. Imielinski, A. Swami. 1993. Mining Associations between Sets of Items in Massive Databases. In Proc. of the 1993 ACM-SIGMOD Int'l Conf. on Management of Data, 207-216.
26. R. Agrawal, R. Srikant. "Fast Discovery of Association Rules", In Proc. of the 20th International Conference on VLDB, Santiago, Chile, September 1994.
27. R. Srikant, R. Agrawal. "Mining Generalized Association Rules", In Proc. of the 21th International Conference on VLDB, Zurich, Switzerland, 1995.
28. R. Srikant, R. Agrawal. "Mining quantitative association rules in large relational tables". In Proceedings of the ACM SIGMOD Conference on Management of Data, Montreal, Canada, June 1996.
29. Savasere, E. Omiecinski, and S. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases", In Proc. 21st Int'l Conf. Very Large Data Bases, Morgan Kaufmann, San Francisco, 1995.
30. J.S. Park, M.-S. Chen, and S.Y. Philip, "An Effective HashBased Algorithm for Mining Association Rules", In Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1995.
31. S. Brin et al., "Dynamic Itemset Counting and Implication Rules for Market Basket Data", In Proc. ACM SIGMOD Int'l Conf. Management of Data, ACM Press, New York, 1997.
32. J. Hipp, U. Guntzer, and G. Nakaeizadeh. Algorithms for Association Rule Mining – A General Survey and Comparison. In Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000.

Рекомендовані електронні джерела

1. Курс лекцій Николая Анохина (mail.ru DataGroup)

[https://www.youtube.com/playlist?](https://www.youtube.com/playlist?list=PLrCZzMib1e9pyyrqknouMZbIPf4I3CwUP)

[list=PLrCZzMib1e9pyyrqknouMZbIPf4I3CwUP](https://www.youtube.com/playlist?list=PLrCZzMib1e9pyyrqknouMZbIPf4I3CwUP)

2. Data is the New Oil By Michael Palmer http://ana.blogs.com/maestros/2006/11/data_is_the_new.html

3. Анализ данных как область знания

<http://postnauka.ru/video/34960>

4. Материалы на тему анализа данных http://www.basegroup.ru/library/methodology/data_mining/

5. Наивный Байесовский классификатор в 25 строк кода

<http://habrahabr.ru/post/120194/>

6. Фильтрация смс спама с помощью наивного байесовского классификатора

<http://habrahabr.ru/post/184574/>

7. Лекции курса «Машинное обучение» от yandex

<https://yadi.sk/d/V9p7E6uAFjHcD>

8. Воронцов К. В. Лекции по алгоритмам кластеризации и многомерного шкалирования

<http://www.ccas.ru/voron/download/Clustering.pdf>

9. Котов А., Красильников Н. Кластеризация данных. 2006

<http://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf>

10. Информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных

www.machinelearning.ru/

11. Н.Ю. Золотых Как обучаются машины? научно-популярная лекция http://www.uic.unn.ru/~zny/ml/Pop/ml_pop.pdf

12. Главы из книги на тему машинного обучения и презентации уроков Сергея Николенко

<http://logic.pdmi.ras.ru/~sergey/teaching/ml/>

Додаток 1. Завдання для самостійної роботи Теоретичні питання

1. Give detailed definitions of concepts: data, database, scale, data classification, data clustering, Data Mining.
2. Map Reduce Technology
3. Standard CRISP-DM

Лабораторна робота No1

1. Download abalone data set
Abalone Data Set: <https://archive.ics.uci.edu/ml/datasets/abalone>
2. Find and process papers that cite this data set
3. Prepare abalone data set for Data Mining
Pyle D. Data Preparation for Data Mining: <https://pdfs.semanticscholar.org/470a/828d5e3962f2917a0092cc6ba46ccfe41a2a.pdf>
3. Use a several classifiers in scikit-learn on abalone datasets
Classifier comparison: http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html
4. Determine the quality of a multiclass classifier using precision-score
5. Prepare a report