DATA MINING
course syllabus

Educational and professional program	**"Artificial Intelligence Technologies"**
Specialty	126 – Information systems and technologies
Level of education	the first (bachelor's)
Code of the course in the educational program	ОКП.23

Competencies	КІ	КЗ 3	КС 1	КС 6	КС 11
Program results	ПР 4	ПР 9			

Lectures and practical classes according to the schedule http://rasp.kart.edu.ua.
The information resources of the course are placed on the Google Classroom platform.

Lecturer of the course – PhD, Sen. Lec. Oleksandr Ivaniuk.

The course is designed for one semester, consists of 30 hours of lectures, 45 hours of laboratory work. The course ends with an exam.

The course consists of six thematic blocks:
1. Introduction to machine learning and data mining (2 hours of lectures).
2. Linear regression method (6 hours of lectures).
3. Linear and Bayesian methods of classification and regression (6 hours of lectures).
4. Metric and logical methods of classification and regression (6 hours of lectures).
5. Unsupervised learning methods (6 hours of lectures).
6. Basics of recommender systems (4 hours of lectures).

Thematic and calendar plan of the course

Week	Topic of the lecture	Topic of laboratory work
1	Lec. 1. Introduction to machine learning and data mining	LW 1. Basics of linear algebra in Python
2	Lec. 2. Regression. Linear regression method	LW 2. Construction and application of the linear regression method
2	–	
3	Lec. 3. Polynomial regression method	
4	Lec. 4. Overfitting. Regularization	LW 3. Construction and application of the polynomial regression method
4	–	
5	Lec. 5. Classification. Logistic regression method	LW 4. Construction and application of the logistic regression method
6	Lec. 6. Classification quality metrics. Validation techniques	
6	–	LW 5. Construction and application of the Naive Bayesian classification method
7	Lec. 7 Naive Bayesian classification method	
8	Lec. 8. Metric classification and regression. k nearest neighbors method	LW 6. Construction and application of the k nearest neighbors method
8	Lec. 9. Logical classification and regression. Decision trees method	
9	–	

Week	Topic of the lecture	Topic of laboratory work
10	Lec. 10. Ensemble methods. Random forest method	LW 7. Application of the decision trees method
10	–	
11	Lec. 11. Clustering. Hierarchical clustering method. k-means method	LW 8. Construction and application of the hierarchical clustering method and the k-means method
12	Lec. 12. Dimension reduction. Principal component analysis	
12	–	LW 9. Application of the dimension reduction methods
13	Lec. 13. Anomalies detection. Local outlier factor method. Isolation forest method	
14	Lec. 14. Basics of recommender systems	LW 10. Application of the anomalies detection methods
14	–	
15	Lec. 15. Basics of recommender systems	

Recommended reading

Main

1. Basyuk T.M., Lytvyn V.V., Zakharia L.M., Kunanets N.E. Mashynne navchannya: navch. posib. Lviv: "Novyy Svit-2000", 2019. 315 p.
2. Hastie T., Friedman J., Tibshirani R. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2017. URL: https://web.stanford.edu/~hastie/ElemStatLearn.
3. Intelektual′nyy analiz danykh : praktykum / M. T. Fisun, I. O. Kravets, P. P. Kazmirchuk, S. G. Nikolenko. Lviv: "Novyy Svit-2000", 2019. 162 p.
4. Ng A. Machine Learning Yearning (draft). Deeplearning.ai, 2018.
5. Bishop C. Pattern recognition and machine learning. Springer, 2006. URL: https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/.
6. Mitchell T. Machine learning. Singapore: McGraw-Hill, 1997. URL: http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html.

Supporting

1. MacKay D. Information theory, inference, and learning algorithms. Cambridge: Cambridge University Press, 2019. URL: http://www.inference.org.uk/mackay/itila.
2. Burkov A. The hundred-page machine learning book.
3. Daume H. A Course in Machine Learning (v0.9), 2017. URL: http://ciml.info.
4. Wasserman L. All of Statistics. New York, NY: Springer, 2013. URL: http://www.stat.cmu.edu/~larry/all-of-statistics/index.html.
5. Coelho L.P., Richert W. Building Machine Learning Systems with Python. 2016. 302 c.
6. Ng A. Machine Learning course notes. Coursera, 2011.
7. Domingos, P. The Master Algorithm: How machine learning is reshaping how we live. 2015.

Useful Internet resources

1. Synopsis of course lectures «Machine Learning» (prof. Andrew Ng, Coursera), author – Alex Holehouse. URL: http://www.holehouse.org/mlclass/.
2. Specialization «Machine Learning Specialization» (Coursera), author – prof. Andrew Ns. URL: https://www.coursera.org/specializations/machine-learning-introduction.
3. Course «Machine Learning» (Udacity) developer – Georgia Institute of Technology. URL: https://www.udacity.com/course/machine-learning--ud262.
4. Course CS229 – «Machine Learning» (Stanford), author – prof. Andrew Ng. URL: https://see.stanford.edu/Course/CS229.

5. Course «Machine Learning» (YouTube), author – mathematicalmonk. URL: https://bit.ly/2YmuvYN.

6. Course «Machine Learning with Python» (YouTube), author – mathematicalmonk. URL: https://bit.ly/3h2Srae.

7. Course «Scikit-learn Machine Learning with Python and SKlearn» (YouTube), author – sentdex. URL: https://bit.ly/38I5qts.

8. List of AI and Machine Learning Resources from Around the Web (Medium), author – Robbie Allen. URL: https://bit.ly/3kQKzK8.

The procedure for assessment learning outcomes

Assessment of the student's knowledge of the course is carried out according to the following components:

- performance and protection of laboratory work;
- passing modular control;
- taking an exam (mandatory when receiving an overall modular grade of F, at the student's choice when receiving an overall modular grade of D or B).

Laboratory work can be protected provided that the task is completed independently and correctly. Protection of laboratory work is possible during laboratory classes and during consultation hours, but no more than one laboratory work per day.

If a student misses more than two lectures during one module, the current grade for the corresponding module is reduced by three points, for each missed lecture of more than two. Penalty points can be worked off by completing additional individual or group tasks, showing activity during lectures and laboratory classes, participating in student conferences (on the subject of the course) or data analysis competitions.

Module	Type of control	Type of work		Maximum scores	Sum	
1	Current control	Performance and protection of laboratory work	LW 1	12	60	100
			LW 2	12		
			LW 3	12		
			LW 4	12		
			LW 5	12		
	Modular control			40	40	
2	Current control	Performance and protection of laboratory work	LW 6	12	60	100
			LW 7	12		
			LW 8	12		
			LW 9	12		
			LW 10	12		
	Modular control			40	40	