Syllabus
«Data mining with Weka»

Викладач: *Guiko Gennadyi Petrovych*
Dr. Sc. (in Phys.-Math.), Professor,
Department of Computer Engineering,
Petro Mohyla Black Sea National University

The scope of the discipline: 3 ECTS credits (10 hours of lectures, 20 hours of practice, 60 hours of self-study, exam).

Purpose:
Data mining is an effective support tool that allows the researcher to gain additional knowledge of the subject area in which he is working and must make informed and informed decisions. Weka is a proven 20-year-old Java-based open source software that provides tools for data mining, including a variety of machine learning and visualization algorithms.

Originality of the academic discipline:
Author's course

Expected learning outcomes
As a result of studying the discipline, the student

must know:
- methods of automatic search for regularities in data;
- methods of working with cloud services for intercomponent interaction;
- basic data exchange protocols between system elements;

must be able to:
- use the existing component base to implement the task;
- apply machine learning and visualization algorithms through the graphical user interface (GUI) of Weka to your real data;
- design and implement the software part of the system, using existing libraries or creating your own;
- identify, verify and apply patterns to predict new data or knowledge;
- to make balanced and informed decisions during the collection and accumulation of various data.

The content of the discipline

Topic 1. **Intelligent data analysis in Weka software**
Attribute categories. ARFF file format. Datasets. Data loading and visualization in Weka.

Topic 2. **Weka software product interface**
Weka software product interface. Software package manager. Guide and work with it.

Topic 3. **The essence of classification processes**
Data preparation. Building a classification model. Testing the classification model.

Topic 4. **The essence of regression**
Data preparation. Building and testing a regression model.

Topic 5. **The essence of clustering**
Examples of clustering. Building a clustering model and testing it.

Topic 6. **The essence of associative rules and their search**
Contents of parameters of association search algorithms and their testing. Ranking attributes in datasets.

Prerequisites
Disciplines "Higher mathematics", "Physics", "Sensors and transducers", "Optimization methods", "Distributed systems with cloud architecture and data

Deadline policy
Works that are submitted in violation of deadlines without good reason are evaluated at a lower grade.

storage".

Consequences

The knowledge gained during the study of the discipline can be used for the preparation of a qualification work (dissertation) and professional activity.

Semester control: exam

Evaluation:

Evaluation of student's performance during semester: 60 points
Exam: 40 points

Types of work

Practical works – 45 points
Individual project task – 15 points

University Attendance Policy

A student enrolled in a course is expected to attend all scheduled classes. The teacher of each course informs the students about the rules regarding the absence of students. Students should be aware of this policy. The final decision on whether or not to excuse a student's absence is made by the teacher.

Attendance or participation is also expected in online courses. Participation in online courses can take different forms; it is the teacher who determines what form of attendance or participation is expected.

The teacher reserves the right to give a failing grade to a student for excessive absences.

Technical support

Practices on the discipline are carried out in computer classes using Weka 3-9-6, GNU General Public License version 3.0, URL: https://sourceforge.net/projects/weka

Academic Integrity Policy

Provides independent performance of practices. Write-off (including using mobile devices) is prohibited. The work is not credited in case of detection of plagiarism or writeoff.

Evaluation criteria of laboratory works / practices / individual works / reports / projects

Maximum number of points – a PhD student with high quality independently performed the entire scope of work, answers all questions related to the work performed, and makes additional calculations, for example, using the methods of fuzzy data processing offered to him by the teacher. The teacher has no complaints about the software implementation and performance requirements.

70 %–99 % of the maximum number of points – a PhD student with sufficient quality independently completed all tasks, but in the process he made some mistakes, which, after pointing to them by the teacher, corrected themselves. He answers some questions with a slight error. The additional calculations offered by the teacher make with some complexity. Not all work requirements are met.

40 %–69 % of the maximum number of points – a PhD student of average quality independently completed all tasks, but did not meet all the requirements for implementation. He answers the question with a slight error. The additional calculations offered by the teacher, for example, using fuzzy methods of decision making makes with insignificant errors. Not all requirements for the design of the work are met.

1 %–39 % of the maximum number of points – a PhD student performed all the work independently, but the quality of implementation is insufficient (errors in calculations, not all work requirements are met). The answers to the questions about the work are not entirely clear. There are errors in the answers.

0 points – a PhD student did not perform the entire amount of work, or performed with gross errors. He has problems with calculations by certain methods, does not know the theoretical material, the software implementation does not meet the requirements.