Syllabus
Course Program

MACHINE LEARNING

Specialty
123 – Computer engineering

Institute
Institute of Computer Science and Information Technology

Educational program
Modern programming, mobile devices, and computer games

Department
Computer Engineering and Programming (326)

Level of education
Bachelor's level
Master's level
Ph.D.

Course type
Professional, Elective

Semester
2

Language of instruction
English, Ukrainian

Lecturers and course developers

Gavrylenko Svitlana

Svitlana.Gavrylenko@khpi.edu.ua
Doctor of Technical Sciences, Professor, Professor at the Computer Engineering and Programming Department NTU "KhPI".
Has over 250 publications. Teaches courses: "Machine Learning," " Compiler Design Theory," "Formal Languages, Grammars, and Automata."
For more information about the professor, please visit the department's website.
https://web.kpi.kharkov.ua/cep/2021/09/03/gavrylenko-svitlana-yuriyivna/

General information

Summary

Machine learning is a branch of artificial intelligence and computer science which focuses on the use of data and algorithms capable of learning and solving tasks such as classification, recognition, prediction, clustering, visualization, automation, chatbot development, and more..

Course objectives and goals

The purpose of studying this discipline is to introduce students to the theoretical foundations and algorithms of machine learning, their potential practical implementations, their application in solving real-world tasks, and the development of the corresponding competence in this field.

Format of classes

The course includes lectures, laboratory and practical work, computational assignments, independent study, and consultations. The final assessment is conducted through an exam.

Competencies

CK4: Ability to construct and analyze models of computer systems and networks.
CK10: Ability to identify, classify, and describe the operation of software and technical tools, computer systems, networks, and their components.
CK11: Ability to select effective methods for solving complex problems in computer engineering, critically evaluate obtained results, and justify the decisions made.

Learning outcomes

RN2: Ability to find necessary data, analyze and evaluate them.
RN3: Ability to construct and investigate models of computer systems and networks, evaluate their adequacy, and determine the limits of applicability.

Student workload

The total volume of the course is 150 hours (5 ECTS credits): lectures - 32 hours, practical classes - 16 hours, laboratory classes - 16 hours, self-study - 86 hours.

Course prerequisites

To successfully complete the course, it is necessary to have knowledge of probability theory, linear algebra, and the programming language Python.

Features of the course, teaching and learning methods, and technologies

Lectures are conducted using multimedia technologies, including interactive cloud-based environments like Google Colaboratory and the Jupyter Notebook service for programming in Python.

Program of the course

Topics of the lectures

Topic 1. Introduction
Key concepts. Subject of machine learning. Types and tools of machine learning. Google Colab environment for Python.
Topic 2. Data visualization
Data visualization libraries: Matplotlib, Pandas, Seaborn, and others. Techniques of data visualization.
Topic 3. Data preprocessing
Steps of data preprocessing: Data cleaning, Data transformation, Data reduction, Data integration. Data balancing: Undersampling, Oversampling, SMOTEENN, SVMSMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE.
Topic 4. Model evaluation
Key performance for classification tasks: Accuracy, Precision, Recall, F1score, ROC curve.
Topic 5. Supervised learning. Data classification.
Classification methods: k-Nearest Neighbors, Naive Bayes, Support Vector Machines, Logistic Regression, and others. Python libraries for data classification.
Topic 6. Decision trees
Decision tree techniques: ID3, C4.5, CART. Overfitting of decision trees.
Topic 7. Ensemble methods for data classification problem
Process of generating samples: Pasting, Bootstrapping, Random Subspaces, Random Patches. Meta-algorithm Bagging. Random Forest algorithm. Meta-algorithms: Boosting, Stacking. Python libraries for working with ensembles.
Topic 8. Data prediction.
Methods of data prediction. Predictive analytics: Linear regression, Multiple regression, Decision trees for forecasting, SVM algorithm. Quality metrics for regression tasks. Python libraries for data forecasting.
Topic 9. Unsupervised learning
Data clustering, dimensionality reduction. Methods: K-means, DBSCAN, PCA. Python libraries for data clustering.
Topic 10. Reinforcement learning
Reinforcement learning. Value function. Methods of reinforcement learning. Application of reinforcement learning for building game agents, autonomous vehicles, and robotics. Python libraries for reinforcement learning.
Topic 11. Data Anomalies.
Data Anomalies. Types of anomalies. Methods for detecting anomalies in data. Isolation Forest algorithm.

Topics of the workshops

Topic 1. Google Colab Environment for Python.
Main libraries. Concept of DataFrame. Data visualization techniques.

Topic 2. Methods of data preprocessing.
Methods: Data cleaning, Data transformation, Data reduction, Data integration. Data balancing methods: Undersampling, Oversampling, SMOTEENN, SVMSMOTE, BorderlineSMOTE, ADASYN, SMOTE, KMeansSMOTE.
Topic 3. Quality metrics for classification tasks.
Calculation of quality metrics such as Accuracy, Precision, Recall, F1-score, and ROC curve using examples.
Topic 4. Methods of data classification.
Building models using examples of toy data. Evaluation of model quality.
Topic 5. Ensemble methods for classification.
Building ensembles using examples of toy data. Evaluation of model quality.
Topic 6. Cluster analysis.
Development of clustering models using examples of toy data.
Topic 7. Reinforcement learning.
Methods of reinforcement learning. Dynamic programming methods. Monte Carlo methods.
Topic 8. Methods of anomaly detection in data.
Algorithms: One-class SVM, Isolation Forest.

Topics of the laboratory classes

Topic 1. Working with DataFrames. Creating a DataFrame. Basic operations. Data grouping.
Topic 2. Python libraries for data visualization. Data Visualization in Python. Line Charts. Bar Graphs. Histograms. Scatter Plots. Heat Maps. Visual data analysis. Saving visualization results.
Topic 3. Investigating the impact of data preprocessing methods on the performance of a model.
Topic 4. Building a classification model. Investigating the impact of tuning model parameters on the quality of data classification. Visualizing the results of the investigation.
Topic 5. Ensemble methods for classification. Bagging. Selection of base classifier. Investigating the impact of tuning ensemble parameters on the quality of data classification. Visualizing the results of the investigation.
Topic 6. Ensemble methods for classification. Boosting. Selection of base classifier. Investigating the impact of tuning ensemble parameters on the quality of data classification. Visualizing the results of the investigation.
Topic 7. Building a predictive model. Quality metrics for predictive models. Investigating the impact of tuning model parameters on the quality of predictions. Visualizing the results of the investigation.
Topic 8. Building a data clustering model using different methods. Performing a comparative analysis. Visualizing the results of the investigation.

Self-study

The course includes completing an individual computational assignment on one of the following topics:
- Development of an anomaly detection model in data.
- Development of a decision-making system using reinforcement learning methods.
The result is documented in a written report in the form of a presentation of the developed software.
Students are also recommended additional materials (videos, articles, textbooks) for self-study and analysis, to prepare for lectures, practical sessions, and laboratory sessions.

Course materials and recommended reading

Primary literature:
1. Gavrylenko S.Y., Chelak V.V., Hornostal O.A., Zozulia V.D. Machine Learning. Laboratory Workshop / S.Y. Gavrylenko, V.V. Chelak, O.A. Hornostal, V.D. Zozulia. - Kharkiv: NTU "KhPI", 2022. - 86 p.
2. Machine Learning: Starter Course: electronic tutorial / Shtovba S.D., Kozachko O.M. - Vinnytsia: VNTU, 2020. - 81 p.
3. Matsuga, O.M. Educational manual for the course "Information Technologies of Image Recognition" / O.M. Matsuga, Yu.M. Arkhangelska, N.M. Yereshchenko. - Dnipro: RVV DNU, 2016. - 60 p.
4. Oliynyk A.O. Intelligent Data Analysis: Educational Manual / A.O. Oliynyk, O.O. Oliynyk, S.O. Subbotin. - Zaporizhzhia: ZNTU, 2011. - 278 p.

Additional literature:
5. Halterman R.L. Fundamentals of Python Programming. - Southern Adventist University. 2019. - 658 p.
Steele B., Chandler J., Reddy S. Algorithms for Data Science. - Springer. 2016. - 430 p.
6. Ethem Alpaydin. Introduction to Machine Learning. - 2009. - 584 p.

Assessment and grading

Criteria for assessment of student performance, and the final score structure

The final grade consists of the assessment results in the form of an exam (40%) and ongoing assessment (60%). Exam: written assignment (2 theory questions + problem-solving) and an oral presentation. Ongoing assessment: 2 online tests (15%), laboratory works (35%), and a computational assignment (10%).

Grading scale

Total points	National	ECTS
90–100	Excellent	A
82–89	Good	B
75–81	Good	C
64–74	Satisfactory	D
60–63	Satisfactory	E
35–59	Unsatisfactory (requires additional learning)	FX
1–34	Unsatisfactory (requires repetition of the course)	F

Norms of academic integrity and course policy

The student must adhere to the Code of Ethics of Academic Relations and Integrity of NTU "KhPI": to demonstrate discipline, good manners, kindness, honesty, and responsibility. Conflict situations should be openly discussed in academic groups with a lecturer, and if it is impossible to resolve the conflict, they should be brought to the attention of the Institute's management.
Regulatory and legal documents related to the implementation of the principles of academic integrity at NTU "KhPI" are available on the website: http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/

Approval

Approved by	Date, signature	Head of the department Oleksandr ZAKOVOROTNYI
	Date, signature	Guarantor of the educational program Svitlana GAVRYLENKO