SYLLABUS
Spring semester 2023-2024 academic year
Educational program "7M06301 – Information Security Systems"

ID and name of the discipline	Independent work of the student (IWS)	Number of credits			General number of credits	Independent work of the student under the guidance of a teacher (IWST)
		Lectures (L)	Pract. classes (PC)	Lab. classes (LC)		
102503 Machine learning to detect online threats	4	1.7	0	3.3	5	9

ACADEMIC INFORMATION ABOUT DISCIPLINE

Training format	Cycle, component	Lecture types	Types of practical exercises	Shape and platform final control
Offline	BC, CC	Problem-oriented	Exploring Natural Language Processing Concepts Using Machine Learning Models	Oral offline
Lecturer - (s)	Karyukin Vladislav Igorevich			
e-mail:	vladislav.karyukin@gmail.com vladislav.karyukin@kaznu.kz			
Phone :	+77019405992			
Assistant - (s)	Karyukin Vladislav Igorevich			
e-mail:	vladislav.karyukin@gmail.com vladislav.karyukin@kaznu.kz			
Phone:	+77019405992			

ACADEMIC PRESENTATION OF THE DISCIPLINE

Purpose of the discipline	Expected learning outcomes (LO) *	Indicators of achievement of RO (ID)
Gain skills in detecting Internet threats using machine learning models and neural networks, highlighting Internet threat anomalies, phishing, and SQL injections	1. (cognitive) Theoretical concepts of Internet threat detection	1.1 Understands basic and advanced concepts of Internet threats
		1.2 Analyzes the features of Internet threat detection methods
		1.3 Applies application development methods that use machine learning protection against Internet threats
	2. (functional) Working with libraries for creating machine learning models	2.1 Uses knowledge of installing libraries to work with machine learning models
		2.2 Uses library data to work with machine learning models
		2.3 Builds skills in working with machine learning libraries when creating applications
	3. (functional) Develop programs that detect Internet threats	3.1 Develops methods for generating Internet threat datasets
		3.2 Creates a fully functional application that detects Internet threats
		3.3 Develops methods for analyzing the accuracy of Internet threat detection
	4. (system) Create data protection modules	4.1 Creates data security configuration
		4.2 Analyzes application vulnerabilities
		4.3 Applies machine learning techniques to ensure application security
	5. (system) Create web-application that uses machine learning techniques to detect Internet threats	5.1 Creates a web application that uses machine learning models
		5.2 Configures machine learning modules in the web application
		5.3 Evaluates the accuracy of Internet threat detection
Prerequisites	Information security audit, Semantic analysis methods for ensuring information security	

Post-requisites	Web Application Security
Learning Resources	**Literature :** **Main :** – Natural Language Processing with Python and spaCy: A Practical Introduction, Yuli Vasiliev, 2021. – Machine Learning and Deep Learning in Natural Language Processing, Anitha S. Pillai, Roberto Tedesco, 2023. – Natural Language Processing: A Machine Learning Perspective Yue Zhang, Zhiyang Teng, 2021. – Natural Language Processing Projects: Build Next-Generation NLP Applications Using AI Techniques, Akshay Kulkarni, Adarsha Shivananda , Anoosh Kulkarni, 2021. – Security and Privacy for Big Data, Cloud Computing and Applications. Wei Ren , Lizhe Wang , Kim-Kwang Raymond Choo , Fatos Xhafa , 2019. – Big Data Security. Shibakali Gupta, Indradip Banerjee and Siddhartha Bhattacharyya, 2019. – Machine Learning and Security. Clarence Chio, David Freeman, 2018. **Additional :** – Learning Scientific Programming with Python, Christian Hill, 2021 – Deep Learning for Natural Language Processing: Creating Neural Networks with Python. Palash Goyal, Sumit Pandey, Karan Jain, 2018 **Professional scientific databases :** – Laboratory room 514 – Laboratory room 323 **Internet resources:** – Python Exercises, Practice, Solution – https://www.w3resource.com/python-exercises/ – Natural Language Toolkit – https://www.nltk.org/ – Tensorflow – https://www.tensorflow.org/?hl=en – Machine learning mastery – https://machinelearningmastery.com/start-here/ **Software provision :** Python IDE, Anaconda Navigator Python, NLTK, Microsoft Office Word, WinRAR, Power Point, Adobe Reader, Paint.
Academic discipline policy	The academic policy of the discipline is determined by the Academic Policy and the Academic Integrity Policy of Al-Farabi KazNU. Documents are available on the main page of the Univer IS. **Integration of science and education.** The research work of students, undergraduates, and doctoral students deepens the educational process. It is organized directly in departments, laboratories, scientific and design departments of the university, and in student scientific and technical associations. Independent work of students at all levels of education is aimed at developing research skills and competencies based on acquiring new knowledge using modern research and information technologies. A teacher at a research university integrates the results of scientific activity into the topics of lectures and seminar (practical) classes, laboratory classes, and into the tasks of the IWST and IWS, which are reflected in the syllabus and are responsible for the relevance of the topics of training sessions and tasks. **Attendance.** The deadline for each task is indicated in the calendar (schedule) for the implementation of the discipline content. Failure to meet deadlines will result in loss of points. **Academic integrity.** Practical/laboratory classes and SRL develop the student's independence, critical thinking, and creativity. Plagiarism, forgery, use of cheat sheets, and cheating at all stages of assignments are unacceptable. In addition to the main policies, the observance of academic integrity during theoretical training and exams is regulated by the "Rules for conducting final control", "Instructions for conducting final control of the autumn/spring semester of the current academic year", "Regulations on checking students' text documents for the presence of borrowings". Documents are available on the main page of the Univer IS. **Basic principles of inclusive education.** The educational environment of the university is conceived as a safe place where there is always support and equal treatment on the part of the teacher towards all students and students towards each other, regardless of gender, race/ethnicity, religious beliefs, socio-economic status, physical health of the student, etc. All people need the support and friendship of peers and fellow students. For all students, making progress is more about what they can do than what they can't do. Variety enhances all aspects of life. All students, especially those with disabilities, can receive advice by phone/e- mail vladislav.karyukin@gmail.com / +77019405992 or via video call in MS Teams https://teams.microsoft.com/l/team/19%3AcGGkY7DL7w1krzPs1cYvF4qbh4myCMg1y9gBtOWWCv81%40thread.tacv2/conversations?groupId=412d8a35-3027-4fb9-bd2e-b7b37d3219a6&tenantId=b0ab71a5-75b1-4d65-81f7-f479b4978d7b

INFORMATION ABOUT TEACHING, LEARNING AND ASSESSMENT

Point -rating letter system for assessing educational achievements

Grade	Digital equivalent points	Points, % content	Traditional assessment
A	4.0 _	95-100	Great
A-	3.67	90-94	
B+	3.33	85-89	Fine
B	3.0	80-84	
B-	2.67	75-79	
C+	2.33	70-74	
C	2.0	65-69	Satisfactory
C-	1.67	60-64	
D+	1.33	55-59	
D	1.0	50-54	
FX	0.5	25-49	Unsatisfactory
F	0	0-24	

Assessment methods

Criteria-based assessment is the process of correlating actually achieved learning outcomes with expected learning outcomes based on clearly developed criteria. Based on formative and summative assessment.

Formative assessment is a type of assessment that is carried out during everyday learning activities. Is a current indicator of academic performance. Provides operational communication between the student and the teacher. Allows you to determine the student's capabilities, identify difficulties, help in achieving the best results, and promptly correct the educational process for the teacher. The completion of assignments, activity in the classroom during lectures, seminars, practical classes (discussions, quizzes, debates, round tables, laboratory work, etc.) are assessed. Acquired knowledge and competencies are assessed.

Summative assessment – a type of assessment that is carried out upon completion of the study of a section in accordance with the discipline program. Conducted 3-4 times per semester when performing IWS. This is an assessment of mastery of expected learning outcomes in relation to descriptors. Allows you to determine and record the level of mastery of a discipline over a certain period. Learning outcomes are assessed.

Formative and summative assessment	Points % content
Activity in lectures	0
Work in practical classes	25
Independent work	25
Project and creative activities	10
Final control (exam)	40
TOTAL	100

Calendar (schedule) for implementing the content of the discipline. Teaching and learning methods.

A week	Topic name	Number of hours	Max. point
	MODULE 1 Basic concepts of network threats		
1	**L 1.** Introduction to Network Threat Analysis	1	
	LC 1. Application of methods for detecting network threats	2	5
2	**L 2.** Network threat detection technologies	1	
	LC 2. Creating a database to store network threat logs	2	5
	IWST 1. Consultations on the implementation of SRO 1 on the topic " Implementation of a project for analyzing and processing network threats"		
3	**L 3.** Perform network threat data processing operations	1	
	LC 3. Development of a program for processing network threat data	2	7
	IWST 2. Passing IWS 1		20
4	**L 4.** Performing an operation to retrieve data from a network threat dataset	1	
	LZ 4. Creating a network threat data sampling program	2	7
	IWST 3. Conducting a colloquium on topics for 1-3 weeks		5
5	**L 5.** Performing the vectorization operation of text data of network threats	1	
	LC 5. Creating a program for vectorizing text data using tf-idf, Word2Vec methods	2	7
	IWST 4. Consultation on the implementation of SRO 2 on the topic " Classification of network threats using machine learning methods"		
	MODULE 2 Machine learning models for detecting online threats		
6	**L 6.** Preparing network threat data for classification by machine learning models	1	
	LC 6. Creation of a program for processing datasets of network threats such as DDoS, Man in the middle, SQL injection, Phishing, Malware	2	7
7	**L 7.** Classification of network threats machine learning models	1	
	LC 7. Creating a program for classifying network threat models of Naive Bayes, Logistic Regression, Decision Tree, Random Forest, etc.	2	12
	IWST 5. Passing IWS 2		25
Frontier control 1			**100**
8	**L 8.** Classification of network threats with neural networks	1	
	LC 8. Creation of a program for classifying network threats using the Deep neural network, Convolutional neural network, and Long short-term memory neural network models	2	5
	IWST 6. Consultation on the implementation of SRO 3 on the topic " Development of a program for classifying network threats using BERT"		
9	**L 9.** Classification of network threats using ensemble models	1	

	LC 9. Creation of a program for classifying network threats using ensemble models	2	5
10	**L 10.** Data analysis and processing using ChatGPT queries	1	
	LC 10. Creating a data processing program with API ChatGPT	2	5
	IWST 7. Passing IWS 3		25
MODULE 3 Development of an application for detecting network threats			
11	**L 11.** Determine the basic requirements of the web application	1	
	LC 11. Installing and configuring libraries for developing a web application	2	5
	IWST 8. Consultation on implementation of SROP 4 on the topic "Creating an application using machine learning and neural network models"		
12	**L12.** Preparing machine learning models for a web application	1	
	LC 12. Integration of machine learning models into the developed web application	2	5
13	**L 13.** Configuring the web application database	1	
	LC 13. Creating a web application database	2	5
	IWST 9. Passing IWS 4		25
14	**L 14.** Visualization of web application network threat detection methods	1	
	LC 14. Create web pages that display online threat detection	2	10
15	**L 15.** Complete design and testing of the web application	1	
	LC 15. Creating the web application	2	10
Frontier control 2			**100**
Final control (exam)			**100**
TOTAL for discipline			**100**

SUMMATIVE ASSESSMENT RUBRICTOR

CRITERIA FOR ASSESSING LEARNING RESULTS

IWS 1. Implementation of the big data analysis and processing project (20 % of 100% BC1)

Criterion	"Great" 21-25%	"Fine" 11-20%	"Satisfactorily" 6-10%	"Unsatisfactory" 0-5%
Knowledge and understanding of basic concepts of big data analysis and processing	Understanding the degree of relevance and reliability of the data found. Knowledge and understanding of all basic elements and operations of big data analysis and processing	Understanding the degree of relevance and reliability of the data found. Knowledge of more than part of big data analysis and processing operations	Limited understanding of the relevance, relevance, and validity of big data analytics and processing elements and operations	Superficial understanding/lack of understanding of the degree of relevance, relevance, and reliability of the data found. Lack of knowledge of the elements and operations of big data analysis and processing
Coding skills analysis and processing of big data	Clear and clear presentation of program code, absence of syntax errors in the code	There are small logical errors in the program code	A large number of logical and syntax errors in the program code, which make it practically unworkable	No code or just a few lines of code
Writing a report	The writing demonstrates clarity, conciseness, and accuracy.	The writing demonstrates clarity, conciseness, and correctness. Mostly, there are no errors.	There are some key errors in the writing, and the clarity needs improvement.	The writing is unclear, and it is difficult to follow the content. Lots of errors in the text

IWS 2. Analysis of big data protection methods (25% of 100% BC1)

Criterion	"Great" 21-2 5%	"Fine" 11-20%	"Satisfactorily" 6-10%	"Unsatisfactory" 0-5 %
Working with big data protection methods	Understand the degree of compliance, relevance, and reliability of the data in the application. Knowledge and understanding of all major big data security techniques	Understanding the degree of relevance and reliability of the data found. Knowledge of most big data protection methods	Limited understanding of the degree of compliance, relevance, and reliability of creation operations big data protection methods _ _	Superficial understanding/lack of understanding of the degree of compliance, relevance, and reliability of working with databases. Lack of knowledge of the operations of creating big data protection methods
Coding skills	Clear and clear presentation of program code, absence of syntax errors in the code	There are small logical errors in the program code	A large number of logical and syntax errors in the program code, which make it practically unworkable	No code or just a few lines of code
Writing a report	The writing demonstrates clarity, conciseness, and accuracy.	The writing demonstrates clarity, conciseness, and correctness. Mostly, there are no errors.	There are some key errors in the writing, and the clarity needs improvement.	The writing is unclear, and it is difficult to follow the content. Lots of errors in the text

IWS 3. Development of a program for classifying Internet threats using BERT (25% of 100% BC2)

Criterion	"Great" 21 - 25 %	"Fine" 1 1 -2 0 %	"Satisfactorily" 6 - 10 %	"Unsatisfactory" 0-5 %
Working with machine learning models to classify online threats using BERT	Understand the degree of compliance, relevance, and reliability of the data in the application. Knowledge and understanding of all basic Internet threat classification operations using BERT	Understanding the degree of relevance and reliability of the data found. Knowledge of more than half of Internet threat classification operations using BERT	Limited understanding of the relevance, relevance, and confidence of BERT's online threat classification operations	Superficial understanding/lack of understanding of the degree of compliance, relevance, and reliability of working with databases. Lack of knowledge of Internet threat classification operations using BERT
Coding skills	Clear and clear presentation of program code, absence of syntax errors in the code	There are small logical errors in the program code	A large number of logical and syntax errors in the program code, which make it practically unworkable	No code or just a few lines of code
Writing a report	The writing demonstrates clarity, conciseness, and accuracy.	The writing demonstrates clarity, conciseness, and correctness. Mostly, there are no errors.	There are some key errors in the writing, and the clarity needs improvement.	The writing is unclear, and it is difficult to follow the content. Lots of errors in the text

IWS 4. Creating an application that uses big data protection methods (25% of 100% BC2)

Criterion	"Great" 21-25%	"Fine" 11-20%	"Satisfactorily" 6-10%	"Unsatisfactory" 0-5%
Knowledge and understanding of creating an application that uses big data security techniques	Understanding compliance, relevance, and trustworthiness when building an application that uses big data security techniques	Understanding compliance, relevance, and trustworthiness when building an application that uses big data security techniques	Limited understanding of building an application that uses big data security techniques	Superficial understanding/lack of understanding of basic operations for creating an application that uses big data protection methods
Coding skills	Clear and clear presentation of program code, absence of syntax errors in the code	There are small logical errors in the program code	A large number of logical and syntax errors in the program code, which make it practically unworkable	No code or just a few lines of code
Writing a report	The writing demonstrates clarity, conciseness, and accuracy.	The writing demonstrates clarity, conciseness, and correctness. Mostly, there are no errors.	There are some key errors in the writing, and the clarity needs improvement.	The writing is unclear, and it is difficult to follow the content. Lots of errors in the text

Dean _____ **Turar O.N.**

Head of the department _____ **Mussiraliyeva Sh.Zh.**

Lecturer _____ **Karyukin V.I.**