KBTU KAZAKH-BRITISH TECHNICAL UNIVERSITY

JSC «Kazakh-British Technical University»
School of IT and Engineering

APPROVED BY
Dean of SITE
Azamat Imanbayev
«05» 01 2026

SYLLABUS

Discipline: CSCI3234 Machine learning
Term: Spring 2026
Instructors full name: Adilet Yerkin

Personal Information about the Instructor	Time and place of classes		Contact information	
	Lessons	Office Hours	Tel.	e-mail
Adilet Yerkin, MS in Engineering, Senior Lecturer	According to the schedule	According to the schedule	MS Teams team code - **y4ntwlz**	a.yerkin@kbtu.kz

MS TEAMS LINK - (MS Teams team code - **y4ntwlz**)
https://teams.microsoft.com/l/team/19%3ADbAku6j09FY3Saw6CTLD4NIBuUq9OWIV0hHqY2iVUCo1%40thread.tacv2/conversations?groupId=728e4fd1-2448-44a7-868c-59a93fd1276b&tenantId=57081b5e-e66a-4993-8eaf-15b0b309293f

COURSE DURATION: 3 hours a week, 15 weeks, 45 class hours

GENERAL COURSE AIMS:

The course aims to equip students with fundamental knowledge and practical skills in machine learning for building, evaluating, and deploying predictive models. It focuses on the principles of learning from data, model generalization, and performance evaluation, while also addressing modern challenges related to scalability, reliability, and production deployment. The course prepares students to apply machine learning methods effectively in both research and real-world systems.

COURSE DESCRIPTION

This course provides a systematic introduction to machine learning, covering the complete lifecycle of machine learning solutions—from data preprocessing and feature engineering to model training, validation, and deployment. Students study supervised and unsupervised learning methods, including regression, classification, clustering, and ensemble techniques, as well as similarity measures, dimensionality reduction, and outlier analysis.

Special attention is given to model evaluation, validation strategies, handling imbalanced data, and preventing overfitting. The course also introduces machine learning system design, data engineering fundamentals, model interpretability, automated machine learning, and deployment paradigms. Practical work in Python complements theoretical lectures, enabling students to implement algorithms, analyze results, and understand the differences between experimental and production-level machine learning systems.

The course integrates theoretical lectures, independent practice works, a midterm exam, and a final course project, encouraging hands-on problem-solving and research-oriented thinking.

COURSE OBJECTIVES

1. The objectives of the course are high-quality training of students in the field of machine learning and the acquisition of the following competencies by students:
2. Understanding the fundamental principles of machine learning, including learning paradigms, model assumptions, and generalization.
3. Acquiring practical skills in data preprocessing, feature engineering, dimensionality reduction, and data exploration for machine learning tasks.
4. Developing the ability to design, train, and evaluate supervised learning models for regression and classification problems.
5. Gaining experience with unsupervised learning methods and clustering techniques, including appropriate evaluation metrics.
6. Learning to address common machine learning challenges such as overfitting, imbalanced data, and model selection.
7. Understanding ensemble learning methods and modern gradient boosting frameworks, including their advantages and limitations.
8. Developing awareness of machine learning system design, including data pipelines, validation strategies, and deployment considerations.
9. Strengthening skills in model interpretability, automated machine learning tools, and critical assessment of model behavior.
10. Applying theoretical knowledge in practice through hands-on assignments and evaluation-driven analysis of machine learning solutions.

COURSE OUTCOMES

By the end of the course, students will be able to:

1. Preprocess and analyze datasets for machine learning tasks, including cleaning, transformation, and feature selection. Apply similarity and distance measures to explore and structure data.
2. Design and implement supervised learning models using established machine learning algorithms.
3. Train, tune, and evaluate ensemble learning models using appropriate performance metrics.
4. Apply unsupervised learning techniques and assess clustering quality.
5. Implement reliable validation strategies and mitigate overfitting and data imbalance issues.
6. Interpret machine learning model outputs and assess model behavior using interpretability methods.
7. Understand the design and operation of machine learning systems in production environments.
8. Analyze potential failure points of deployed machine learning systems and propose improvements.
9. Integrate theoretical and practical knowledge to solve applied machine learning problems.

COURSE PREREQUISITES:

- Introduction to Machine Learning
- Basic Python programming
- Basic mathematics (algebra, simple calculus)
- Understanding of descriptive statistics (mean, median, variance)

LITERATURE

1. Foundations of Machine Learning: Adaptive Computation and Machine Learning / M. Mohri, A. Rostamizadeh, A. Talwalkar. - 2-nd ed. - Great Britain: The MIT Press, 2018. - 486p (KBTU Library)
2. The Hundred-Page Machine Learning Book / A. Burkov. - Gread Britain, 2019. - 141p (KBTU Library)
3. The Elements of Statistical Learning: Data Mining, Inference, and Prediction / T. Hastie, R. Tibshirani, J. Friedman. - USA: Springer, 2017, ISBN 978-0-387-84857-0. (KBTU Library)
4. Data Mining and Machine Learning Applications / R. Raja, K. K. Nagwanshi, S. Kumar [et al.]. - USA: John Wiley & Sons, Ltd, 2022, ISBN 9781119792529 (KBTU Library)
5. Data Science for Business:What You Need to Know about Data Mining and Data-Analytic Thinking, F. Provost, T. Fawcett, G. Grolemund. - 1-th ed. - USA: Published by O'Reilly Media, 2013. ISBN 10:1449361323. (KBTU Library)
6. Concept Data Analysis: Theory and Applications / C. Carpineto, G. Romano. - USA: John Wiley & Sons, Ltd, 2005. - 210 p, ISBN 9780470011294. (KBTU Library)
7. Machine Learning and Data Science: Fundamentals and Applications / P. Agrawal, C. Gupta, A. Sharma [et al.]. - USA: John Wiley & Sons, Ltd, 2022. - 394 p (KBTU Library)
8. An Introduction to Statistical Learning with Applications in Python" by James, Witten, Hastie, Tibshirani, Taylor (2023)
9. Python Machine Learning, Sebastian Raschka and Vahid Mirjalili
10. Murphy K.P. Machine learning: A probabilistic perspective. MIT Press, second edition (2012)

COURSE ASSESSMENT CRITERIA

Assessment occurs continuously throughout the course. The evaluation will be based on the levels of (maximums in %):

Type of activity	Final scores
Attendance /participation	**6**
Midterm/Endterm	**30**
SIS	**24**
Final exam - Project defense	**40**
Total	**100**

TASKS for students' independent study (SIS)

Week	Practice works	Cost (in points)
3	**SIS 1**	6
6	**SIS 2**	6
12	**SIS 3**	12

	Total		24

COURSE CALENDAR

Week	Class work			SIS (students independent study)
	Topic	Lectures (hours)	Practice (hours)	
1	**Lecture #1.** Intro to Machine learning. Python for DS. Data manipulation, cleaning, data types. Data and dataset. Data Preprocessing. Aggregation. Sampling. Dimensionality reduction. Feature subset selection. Feature creation. Discretization and binarization	2	1	
2	**Lecture #2.** Data exploration. Variable Transformation. Measures of Similarity and Dissimilarity. Distances. Jaccard Coefficient. Cosine Similarity. Correlation. Mutual Information. Feature selection and dimensionality reduction. PCA. Outlier analysis. Missing Data.	2	1	
3	**Lecture #3.** Types of learning methods. Supervised learning. Unsupervised learning. Linear regression, train/test/validation split.	2	1	**SIS 1**
4	**Lecture #4.** Logistic regression. Support vector machines (SVM). kNN: NEAREST NEIGHBOR CLASSIFIER. Decision Tree Classifier. Model estimation. Classification metrics: accuracy, precision, recall, AUC, confusion matrix, log loss. Regression metrics: MSE, MAE, R2-score.	2	1	
5	**Lecture #5.** Imbalanced data classification. Model Overfitting. Model Selection.	2	1	
6	**Lecture #6.** Ensemble learning. Bagging and boosting. Stochastic gradient boosting. LightGBM, XGBoost, CatBoost. Hyper-parameters.	2	1	**SIS 2**
7	**Lecture #7.** Unsupervised learning, K-means clustering, hierarchical, evaluation metrics	2	1	
8	**MIDTERM**	2		
9	**Lecture #8.** Essentials of machine learning system design. Machine Learning Systems in Production. Machine learning in research vs. in production	2	1	
10	**Lecture #9.** Designing ML Systems in Production. Build or buy, open source-based or proprietary tech. Loss functions and metrics. Gathering datasets. Properties of a healthy data pipeline	2	1	
11	**Lecture #10.** Data Engineering Fundamentals	2	1	

12	**Lecture #11.** InterpretML. AutoML	2	1	**SIS 3**
13	**Lecture #12.** Validation schemas. Ensuring reliable evaluation. Standard validation schemas Non-trivial validation schema	2	1	
14	**Lecture #13.** Model Deployment. Machine Learning Deployment Myths. Batch Prediction vs. Online Prediction. Why Machine Learning Systems Fail in Production	2	1	
15	**ENDTERM**	2		
16-17	**Final Exam**			**COURSE PROJECT DEFENCE**

No	Assessment criteria	1	2	3	4	5	6	7	8 / 1 at	9	10	11	12	13	14	15 /2 at	16-17	%
1.	Midterm								15									15%
3.	Endterm															15		15%
4.	SIS			6			6						12					24%
6.	Attendance and activity on lessons								3							3		6%
7.	Final exam: project defense																40	40%
	Total								30							30		**100%**

Class sessions – will be a mixture of information, discussion and practical application of skills.

In-class assessment – will prepare students for their mid-term and final assessment and identify the competence level they have achieved on a related subject matter, the aim being to diagnose potential discrepancies in students' understanding and performance in order to make specific adjustments to the course content and procedures and/or to assign additional assignments to certain individuals or the whole group.

Home assignments – will consolidate the concepts and materials taken during in-class activities, help students to expand the content through diverse background resources and/or practise certain skill areas; they will also develop the students' ability to work individually in exploring and examining related issues.

SIS (Student Independent Study) – project to be done by students on the independent basis. Students are supposed to use knowledge and skills acquired in class to do the project. Assistance and advice will be provided by teachers during office hours.

TSIS (Teacher Supervised Student Independent Study) – student self-made project.

End-term test – a diagnostic test used to identify the students' progress, their strengths and weaknesses, intended to force student to prepare for Final Exam.

Final examination – an attainment test designed to identify how successful the students have been achieving objectives.

Grading policy:

Intermediate attestations (on 7/8th and 15th week) join topics of all lectures, practice, SIS and materials for reading discussed to the time of attestation. Maximum number of points within attendance, activity, SIS and practice for each attestation is 30 points.

Final exam joins and generalizes all course materials, is conducted in the form with project defense. The maximum number of points is 40. At the end of the semester you receive overall total grade (summarized index of your work during semester) according to conventional KBTU grade scale.

ATTENTION!

1) If student missed without plausible reason more than **30% of lessons student receives «F (Fail)»** grade;
2) If for two attestations student receives 29.4 or less points, this student is not accepted to final exam and for all course he (she) receives **«F (Fail)» grade;**
3) If student receives on final exam 9.4 or less points, then independently on how many points he (she) received for two attestations, in whole he (she) receives **«F (Fail)» grade;**
In the case of missing or being late for final exam without plausible reason, independently on how many points he (she) received for two attestations, in whole he (she) receives **«F (Fail)» grade.**
4) It is forbidden to change the topic of the course project and change the composition of the team after **2 weeks of training**. If a student does not join a team during the first two weeks of study, he (she) develops a **course project individually.**
5) If a student missed more **than 30%** of the lectures due to health problems and has medical documents in the form of KBTU, but did not complete the course project, the student is not allowed to **defend the course project**, and it is recommended to take an **academic leave.**
6) In case of non-compliance of the course project with the **given assignment**, the student is not allowed to **defend the course** project and automatically receives an **«F (Fail)» grade**.
7) In case of detection **of plagiarism** in the course project, the student is automatically not allowed to defend the course project and receives **«F (Fail)» grade**.
8) At the exam, the student must prepare a **printed course project**, a **presentation** and a **software implementation** of the project. If any documents are missing, the student does not automatically give the right to defend the course project and receives an **«F (Fail)» grade**.
9) The delivery of the **electronic version** of **the course project** and **presentations in MS TEAMS** should be no later than **15 weeks**, if students do not upload the course project **on time**, they will not automatically finish the exam and receives an **«F (Fail)» grade**.
10) If the course project does not contain the mandatory parts of the assignment, it is not automatically considered in the exam. The exam is given a grade of **«F (Fail)» grade**.
11) The presence of all team members to defend the course project is mandatory.
12) Project protection in online format is not provided, unless it is required by a special order of the responsible authorities.
13) The teacher provides bonus assignments and incentives for students for excellent results in the form of additional points and certificates.
14) The teacher provides guest lectures from industry representatives to increase students' interest in learning.
15) The teacher has provided games and lotteries to increase students' activity in studying the discipline.
16) During the course, some materials are presented in video form.

17) The student is required to join the **MS Teams command** to receive important notifications.
18) If students are late for a lesson or miss an online mark for a lecture, a mark about the student's absence from the lesson is automatically entered into the **WSP portal.**
19) In case of failure to comply with the rules of conduct in class, the student may be removed from the discipline.

REQUIREMENTS FOR THE SUBMISSION OF ASSIGNMENTS OF THE SIS PARTS OF THE COURSE PROJECT!

1) If the documents are not prepared according to the KBTU standards, the task is not automatically considered for written work. **The grade is set to 0.**
2) After the deadline for accepting works that are not uploaded to the MS Teams task on time, the grade **is automatically set to 0.** The task completion time is one week, so a technical failure is not a valid reason for not uploading work to the MS Teams task.
3) Missed tasks must be uploaded to the next task for checking and forming parts of the course project, but the team's points will be lost.
4) The structure of the presentation for the final exam is strictly regulated. If the presentation does not meet the template requirements, the course project will not be considered.

Academic Policy:

1. Cheating, duplication, falsification of data, plagiarism are not permitted under any circumstances!
2. Students must participate fully in every class. While attendance is crucial, merely being in class does not constitute "participation". Participation means reading the assigned materials, coming to class prepared to ask questions and engage in discussion.
3. Students are expected to take an active role in learning (the instructor will provide the information and guidelines to do this).
4. Students must come to class on time.
5. Students are to take responsibility for making up any work missed.
6. Make up tests in case of absence will not normally be allowed.
7. Mobile phones must always be switched off in class.
8. Students should always show tolerance, consideration and mutual support towards other students.

Students are encouraged to:

9. consult the teacher on any issues related to the course;
10. make up within a week's time for the works undone for a valid reason without any grade deductions;
11. make any proposals on improvement of the academic process;
12. track down their continuous rating throughout the semester.

Grade		Achievement percentage	Assessment criterion
«Excellent»	A	95-100%	This grade is given when the student: demonstrated a complete understanding of the course material; did not make any errors or inaccuracies; completed control and laboratory work in a timely and correct manner, and submitted reports on them;

	A -	90-94%	demonstrated original thinking; submitted control quizzes on time and without any errors; completed homework assignments; engaged in research work; independently used additional scientific literature in studying the discipline; was able to independently systematize the course material.
«Good»	B+	85-89%	This grade is given when the student: Has mastered the course material at no less than 75%; Did not make gross errors in responses; Timely completed control and laboratory work and submitted them without fundamental remarks; Correctly completed and timely submitted control tests and homework assignments without fundamental remarks; Utilized additional literature as indicated by the instructor; Engaged in research work, made non-fundamental errors, and fundamental errors corrected by the student themselves; Managed to systematize the course material with the help of the instructor.
	B	80-84%	
	B-	75-79%	
	C+	70-74%	
«Satisfactory»	C	65-69%	This grade is given when the student: Has mastered the course material no less than 50%; Required assistance from the instructor when completing control and laboratory work, homework assignments; Made inaccuracies and non-fundamental errors when submitting control tests; Did not demonstrate activity in research work, relied solely on the educational literature indicated by the instructor; Experienced more difficulty in systematizing the material.
	C-	60-64%	
	D+	55-59%	
	D	50-54%	

MS in Engineering,
Senior Lecturer **Adilet Yerkin**

Minutes #9 of School of Information Technology and Engineering meeting on January 5, 2026