1. GENERAL INFORMATION	
Faculty	Information Technology
Major code and title	5B070400 Computer Systems and Software Engineering
Educational program code and title	-
Year, semester	2, 4
Subject category	Compulsory
Number of credits (ECTS)	4
Prerequisites	Application Development in Python, Mathematical analysis
Postrequisites	Machine Learning - 2
Lecturer	Nazgul K. Rakhimzhanova, Senior-lecturer, Master of technical science Room 409, +7-727-330-85-66 (ext. 2039), E-mail: n.rakhimzhanova@edu.iitu.kz Office-hours: Friday 12:30 – 14:00, 409 room
Instructors	lecturer, Vladlen Chsheglov, v.chsheglov@edu.iitu.kz senior-lecturer, Ersain Chinibayev, ycrsain@gmail.com

2. GOALS, OBJECTIVES AND LEARNING OUTCOMES OF THE COURSE

The course goal is the focus of this course is on learning the fundamentals of machine learning to have complete the understanding, so students can choose track of study on third and fourth courses.

The objectives of the course are

- learn the methodology of data science and open source tools for data science
- learn the fundamentals of math statistics necessary for ML
- learn how to perform explanatory data analysis (EDA)
- learn how to create and test hypothesis on data
- learn how to apply simple predictive models

Learning outcomes of the course

Students successfully completing the course will be able to:

1. Perform explanatory data analysis
2. Perform sampling and estimation
3. Create and test hypothesis
4. Visualize data in different ways
5. Implement basic ML algorithms from supervised learning methods
6. Build and apply predictive model on practical tasks

3. COURSE DESCRIPTION

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data. This course will provide you a foundational understanding of machine learning and statistic basics. This course doesn't cover such topics as logistic regression, multilayer perceptron, convolutional neural networks, natural language processing, etc. As this course is aimed to introduce students with Data Science area, so they will be able to make a conscious choice about their elective courses.

4. COURSE POLICY

Students are forbidden to:

- Submit any tasks after the deadline. Late submissions are graded with penalty.
- Cheat, plagiarized papers shall not be graded.
- Be late for classes, being tardy three times amounts to one absence.
- Retake any tests, unless there is a valid reason for missing them.

Students should always

- Be appropriately dressed (formal/semi- formal styles are acceptable).
- Show consideration for and mutual support of teachers and other students.
- Let the teacher know of any problems arising in connection with their studies.

Collaboration and Academic Integrity

Academic integrity is fundamentally about ethical behavior. Appropriate collaboration and research of previous work is an important part of the learning process. However, not all collaboration or use of existing work is ethical.

Different classes have different rules about collaboration. These are the standards students will be held to for the class.

Students may only get help on graded assignments from designated people.

Students are always welcome to get help on an assignment from your professors, professional school mentors, tutors, or any other faculty member. They may help you at the computer, on paper, or any way they believe will be effective.

To avoid suspicion of plagiarism, students must specify their sources together with all turned-in materials. List classmates they discussed their homework with and webpages from which students got inspiration or copied (short) code snippets. All students are expected to understand and be able to explain their turned-in materials. Plagiarism and cheating, as in copying the work of others, paying others to do work, etc., is obviously prohibited, is grounds for failing the course.

5. LITERATURE

Basic literature:

1. Python Data Science Handbook: Essential Tools for Working with Data / Jacob T. Vanderplas / O'Reilly Media Inc., 2016
2. Practical Statistics for Data Scientists: 50 Essential Concepts / Andrew Bruce and Peter C. Bruce, 2017
3. Data Science from Scratch / Joel Grus / O'Reilly Media Inc., 2017

Supplementary literature:

1. Harrington P., Machine Learning in Action., 2012
2. Hal Daumé III, (2013). A Course in Machine Learning, 2013
3. Sebastian Raschka, (2015). Python Machine Learning, 2015
4. https://www.kaggle.com/
5. https://www.datacamp.com/
6. https://powerbi.microsoft.com/ru-ru/

6. Course schedule

Week/ date	Course topics	References	Lectures (h/w)	Practical sessions (h/w)	Lab. sessions (h/w)	TSIS (h/w)	SIS (h/w)
#1 27/01/20 – 31/01/20	Intro: Data Science Area and open source tools for Data Science	B: [3], S: [1], [5]	1	0	2	1	4
#2 03/02/19 – 06/02/20	NumPy package for data science	B: [1], [2], S: [5]	1	0	2	1	4

#3 10/02/20 – 14/02/20	Pandas package for data science	B: [1], [2], S: [5]	1	0	2	1	4
#4 17/02/20 – 21/02/20	Visualization with matplotlib	B: [1], [2], S: [3], [5]	1	0	2	1	4
#5 24/02/19 – 28/02/20	Statistics: Distribution – Normal	B: [3], S: [1], [2]	1	0	2	1	4
#6 02/03/20 – 06/03/20	Exploratory Data Analysis (EDA)	B:[3] S:[1], [4]	1	0	2	1	4
#7 09/03/20 – 13/03/20	Summary for 6 weeks QA session	B: [1]	1	0	2	1	4
#8 16/03/20 – 20/03/20	Statistics: Distribution – Lognormal, Exponential	B: [3], S: [1], [2]	1	0	2	1	4
#9 23/03/20 – 27/03/20	Sampling and Estimation	B: [3], S: [1], [2]	1	0	2	1	4
#10 30/03/20 – 03/04/20	Visualization II	B: [1], [2], S: [3]	1	0	2	1	4
#11 06/04/20 – 10/04/20	Correlation and Covariance	B: [3], S: [1], [2]	1	0	2	1	4
#12 13/04/20– 17/04/20	Hypothesis testing	B: [3], S: [1], [2]	1	0	2	1	4
#13 20/04/20 – 24/04/20	Linear Regression	B:[1], [3]	1	0	2	1	4
#14 27/04/20 – 01/05/20	Summary for 6 weeks QA session	B:[1], [3]	1	0	2	1	4
#15 04/05/20 – 08/05/20	Course summary	B:[1], [3]	1	0	2	1	4
	Total hours:	**120**	**15**	**0**	**30**	**15**	**60**

7. List of topics/ assignments for laboratory classes

№	Topic Title	Number of hours	References	Form of reporting	Deadline
1	2	3	4	5	6
1	Laboratory 00: Starting with Jupyter Notebook MCQ00	2	B: [1], [2] S: [5]	.ipynb file, on DL	Week 1
2	Laboratory 01: Practicing NumPy package MCQ01	2	B: [1], [2] S: [5]	.ipynb file, on DL	Week 2
3	Laboratory 02: Practice with Pandas package	2	B: [1], [2] S: [5]	.ipynb file, on DL	Week 3

	MCQ02				
4	Laboratory 03: Creating histograms, boxplots, density, pie charts. MCQ03	2	B: [1], [2], [3]	.ipynb file, on DL	Week 4
5	Laboratory 04: Working with normal distributed data MCQ04	2	B: [3], S:[1], [2]	.ipynb file, on DL	Week 5
6	Laboratory 05: Performing EDA over chosen dataset MCQ05	2	B:[3] S: [1], [4]	.ipynb file, on DL	Week 6
7	**Mid-Term Assessment**	2	B: [1], [2]	.ipynb file, on DL	Week 7
8	Mid Term Review	2	-	-	Week 8
9	Laboratory 06: Sampling and Estimation MCQ06	2	B: [3] S: [1]	.ipynb file, on DL	Week 9
10	Laboratory 07: Creating correlation plots, color maps, parallel coordinates, violinplots, MCQ07	2	B: [1], [2] S: [3], [5]	.ipynb file, on DL	Week 10
11	Laboratory 08: Checking data on correlation and covariance MCQ08	2	B: [2], [3] S: [2]	.ipynb file, on DL	Week 11
12	Laboratory 09: Testing different hypothesis on chosen dataset MCQ09	2	B: [3], S: [1], [2]	.ipynb file, on DL	Week 12
13	Laboratory 10: Building linear regression model using sklearn package MCQ10	2	B: [1], [3] S: [1]	.ipynb file, on DL	Week 13
14	**End Term Assessment**	2	B: [1], [3] S: [1]	.ipynb file, on DL	Week 14
15	End Term Review Feedback on course	2	-	.ipynb file, on DL	Week 15
	Total hours:	**30**			

8. List of topics/assignments for Student Independent Study

Proper organization of students independent study is the key to the formation of skills in mastering, learning, assimilation and systematization of acquired knowledge, ensuring a high level of academic performance in the learning process

№	Topic/Assignment title	Number of hours	References	Form of reporting	Deadline
1	2	3	4	5	6
1	SIS01: Choosing a dataset, installing the IDE and environment	4	S: [4], [5]	Demonstration on laptop	Week 1
2	SIS02: Installing Numpy package Practicing solving problems	4	https://www.machinelearningplus.com/python/101-numpy-exercises-python/	Submission on DL	week 2
3	SIS03: Installing Pandas package Practicing solving problems	4	https://www.machinelearningplus.com/python/101-pandas-exercises-python/	Submission on DL	week 3
4	SIS04: Visualizing using Excel package	4	S: [4], [5]	Submission on DL	week 4
5	SIS05: Working with normal distributed data on Excel	4	S: [4], [5]	Submission on DL	week 5
6	SIS06: Preparation to Mid Term assignment	4	B: [1], [2], [3]	Discussion	week 6
7	SIS07: Reading and practicing	4	B: [1], [2], [3]	Discussion	week 7
8	SIS08: EDA on Excel	4	S: [4], [5]	Submission on DL	week 8
9	SIS09: Power BI dashboards	4	S: [6]	Submission on DL	week 9
10	SIS10: Plotting with Bokeah library	4	B:[1] S: [4]	Submission on DL	week 10
11	SIS11: Checking correlation using Excel	4	S: [4]	Submission on DL	week 11
12	SIS12: Hypothesis testing on Excel	4	S: [4]	Submission on DL	week 12
13	SIS13: Linear Regression model using Excel	4	S: [4], [5]	Submission on DL	week 13
14	SIS14: Preparation to End Term assignment	4	B: [1], [2], [3]	Discussion	week 14

15	SIS15: Preparation to Final Exam	4	B: [1], [2], [3]	Discussion	week 15
	Total hours:	**60**			

9. System for evaluating student performance in a discipline:

Period	Assignments	Score	Total
1st attestation	**Laboratory works:**	**40**	**100**
	Laboratory01,	8	
	Laboratory02,	8	
	Laboratory03,	8	
	Laboratory04,	8	
	Laboratory05	8	
	MCQs:	**20**	
	MCQ01,	4	
	MCQ02,	4	
	MCQ03,	4	
	MCQ04,	4	
	MCQ05,	4	
	Mid-term	**30**	
	SIS assignments	**10**	
	SIS02,		
	SIS03,		
	SIS04,		
	SIS05		
2nd attestation	**Laboratory works:**	**35**	**100**
	Laboratory06,	7	
	Laboratory07,	7	
	Laboratory08,	7	
	Laboratory09,	7	
	Laboratory10	7	
	MCQs:	**20**	
	MCQ06,	4	
	MCQ07,	4	
	MCQ08,	4	
	MCQ09,	4	
	MCQ10,	4	
	End-of-term	**30**	
	SIS assignments	**15**	
	SIS08	2.5	
	SIS09	2.5	
	SIS10	2.5	
	SIS11	2.5	
	SIS12	2.5	
	SIS13	2.5	
Exam			**100**
Total	**0,3*1stAtt+0,3*2ndAtt+0,4*Final**		

*If the number of absences exceeds 20%, student will be automatically scheduled for a Retake (summer semester)

10. Assessment criteria:

Letter Grade	Numerical equivalent	Points (%)	Grading scheme
A	4,0	95-100	«Excellent» - the laboratory work is complete (90% - 100%), all tasks are performed on high level, the notebook includes necessary comments and testing cells
A-	3,67	90-94	
B+	3,33	85-89	«Good» - the laboratory work is complete on 80 – 88%, not all recommended functions or methods are used, notebook has satisfactory amount of comments
B	3,0	80-84	
B-	2,67	75-79	«Good» - the laboratory work is complete on 70-80%, a few important tasks are skipped, notebook has a few comments
C+	2,33	70-74	
C	2,0	65-69	«Satisfactory» - laboratory work is complete on 50 – 60%, the notebook has no comments (or a few but unrelevant)
C-	1,67	60-64	
D+	1,33	55-59	
D	1,0	50-54	
FX	0	25-49	FX (25-49) «Fail» with re-exam – the work completion is less than 50%, the tasks are not solved, student has no idea where to start
F	0	0-24	F (0-24) «Fail» - only a few the easiest tasks are complete, no explanation comments, a student cannot explain the done work and cannot answer on teachers' questions