

International Information Technology University JSC

Faculty of Information Technology

Information Systems Department

Approved

Vice-Rector for Academic and Educational
Affairs of IITU JSC, PhD

_____ Umarov T.F.

« ___ » _____ 2019

SYLLABUS (ACADEMIC PROGRAM)

Course (code, title): « Big data processing »

Major: 5B050700- Management, 5B050900- Finance

Educational program (code, title): 6B04102 “E-business», 6B04103 “Innovation management», 6B04104 “Financial Engineering», 6B04105 «Financial technologies»

Year: 3; **Semester:** 6; **Number of credits:** 5 ECTS

Lectures: 15 hours

Practical classes: 30 hours

T/SIS: 105 hours

Total: 150 hours

Final assessment form: Examination

Almaty 2019

Academic program of the course (code, title) « Big data processing » has been developed on the basis of Standard Academic Program.

Academic program has been reviewed at the meeting of Information Systems department.

Minutes №. 1 dated 28.08.2019

Head of the Department _____ V.V.Serbin

Author _____ Abdullayeva A.S.

The academic program has been approved at the meeting of IITU SMC

Minutes №. 1 dated 29.08.2019

Head of IITU AMC, Vice-Rector
for Academic and Educational Affairs of IITU JSC, PhD _____ Umarov T.F.

Head of the Department
for Academic Affairs _____ Mustafina A.K.

Syllabus
for the course
« Big data processing »
Academic Year 2019-2020

1. General information	
Faculty	Information Technology
Major code and title	6B04102 «E-business», 6B04103 «Innovation management», 6B04104 «Financial Engineering», 6B04105 «Financial technologies»
Year, semester	3.6
Subject category	Major
Number of Credits	3
Language of Delivery:	English
Prerequisites:	Data Mining (DM 1)
Postrequisites	None
Lecturer	Seydakhmetova Kalamkas Seydakhmetkyzy
Instructors	Seydakhmetova Kalamkas Seydakhmetkyzy
2. Goals, objectives and learning outcomes of the course	
Course goal(s) Mastering students of professional and personal competences who will give the chance to use the modern information technologies in different areas of data management and advanced data analysis.	
Course objectives <ul style="list-style-type: none"> • to give an idea of use of data analysis packages; • to give an idea of ways of data searching, data mining and deduce outcomes; • to inform of the main concepts of data management, data sorting; • to explain the principles of data management; • to learn to carry independent creative search project; 	
Learning Outcomes: By the end of this course the students will be able to: <ul style="list-style-type: none"> • to define the main tendencies in the field of data management; • to get to and retrieve information wherever it is stored; • to combine different types of data with the help of data integration tools; • to practice with electronic spreadsheets, to execute consolidation of data, to visualize results of work with databases; 	
3. Course description	
This course introduces study about data management and transformation. Course demonstrates some of the power of R language and RStudio tool. Students will have some common tasks that come up when dealing with data. These tasks range from assembling different data sets into more convenient forms and ways to apply functions to different parts of the data sets.	
4. Course policy	
<ol style="list-style-type: none"> 1. Respect and helpfulness to fellow students. 2. No disruption of learning: <ul style="list-style-type: none"> – If you must leave early, let the instructor know and sit near an exit. – Same if you expect an urgent phone call. – Minimize flickering on your computer screen to avoid distraction of your fellow students' peripheral vision. 3. No electronic media in the class room (texting, email, facebook, instagram, browsing, games, working on extraneous tasks, ...) except if related to the class. 4. Professional expression in all forms of communication (oral, email, written), in particular no slang. In particular, there is a ban on the use of "like" as a filler word. 	

5. Honesty and openness in all proceedings, including homeworks and quizzes.

Students are forbidden to:

- submit any tasks after the deadline. The mark for late submissions is decreased by 80 points each week (ex. Student will get 20 for 1 week late submission etc.);
- cheat. Plagiarized papers shall not be graded and receive a "0";
- be late for classes;
- retake any tests, unless there is a valid reason for missing them;
- use mobile phones in class;

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.

Instructor may change course outline at any time during the course. Students may come to see Tutor only at Office Hours time or by appointment.

5. Literature

Basic literature:

1. Notes on R: A Programming Environment for Data Analysis and Graphics, W. N. Venables, D. M. Smith ,2018.
2. Using R and RStudio for Data Management, Nicholas J. Horton, 2015

Supplementary literature:

1. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, Hadley Wickham, Garrett Golemund, 2016.

6. Course schedule

Week No	Course Topic	Reference Materials	Lectures (1 h/w)	Practice sessions (1 h/w)	TSIS (2 h/w)	SIS (4 h/w)
1	Lecture 1. What is exploratory data management?	[3]	1	2	1	6
2	Lecture 2. Introduction to Data Management with R.	[1]	1	2	1	6
3	Lecture 3. R Basics	[9]	1	2	1	6
4	Lecture 4. Data input and output	[9]	1	2	1	6
5	Lecture 5. R Data Structures	[7]	1	2	1	6
6	Lecture 6. Data Management with R	[7]	1	2	1	6
7	Lecture 7. Web Data Extraction and Analysis. Web scraping.	[5]	1	2	1	6

8	Lecture 8. R Statistics	[8]	1	2	1	6
9	Lecture 9. Advanced Statistics	[11]	1	2	1	6
10	Lecture 10. Bootstrapping in R	[4]	1	2	1	6
11	Lecture 11. Graphs. Advanced Graphs	[10], [11]	1	2	1	6
12	Lecture 12. Databases using R	[5]	1	2	1	6
13	Lecture 13. Managing Big Data	[6]	1	2	1	6
14	Lecture 14. R integration with online reporting tools.	[2]	1	2	1	6
15	Lecture 15. Advanced data analytics and management for solving organization problems.	[3]	1	2	1	6
Totalhours		150	15	30	15	90

7.List of topics/ assignments for practical classes

№	Topic Title	Number of hours	References	Form of reporting	Deadline
1	2	3	4	5	6
1	What is exploratory data management?	2	[3]	Discussion	2
2	Introduction to Data Management with R.	2	[1]	Discussion	3
3	R Basics	2	[9]	Discussion	4
4	Data input and output	2	[9]	Discussion	5
5	R Data Structures	2	[7]	Discussion	6
6	Data Management with R	2	[7]	Discussion	7
7	Web Data Extraction and Analysis. Web scraping.	2	[5]	Discussion	8
8	R Statistics	2	[8]	Discussion	9
9	Advanced Statistics	2	[11]	Discussion	10
10	Bootstrapping in R	2	[4]	Discussion	11
11	Graphs. Advanced Graphs	2	[10], [11]	Discussion	12
12	Databases using R	2	[5]	Discussion	13
13	Managing Big Data	2	[6]	Discussion	14
14	R integration with online reporting tools.	2	[2]	Discussion	15
15	Advanced data analytics and management for solving organization problems.	2	[3]	Discussion	15

8. List of assignments for Student Independent Study

№	Assignments (topics) for Independent study	Hours	Recommended literature and other sources (links)	Form of submission	Deadline
1	2	3	4	5	6
1	Exploratory data management	6	[1]	Presentation	2
2	Install RStudio and packages, basic commands of R.	6	[1]	Presentation	3
3	Practice writing R scripts and inspect data sets.	6	[3]	Presentation	4
4	Explore many variables	6	[4]	Presentation	5
5	Data visualization	6	[10]	Presentation	6
6	Data manipulation and management	6	[10]	Presentation	7
7	Web Data Scraping	6	[9]	Presentation	8
8	Sorting different types of data	6	[7]	Presentation	9
9	Price predictions	6	[3]	Presentation	10
10	Forecasting with R	6	[4]	Presentation	11
11	Graphs. Advanced Graphs	6	[5]	Presentation	12
12	R integration with MongoDB	6	[10], [8]	Presentation	13
13	R integration with MongoDB	6	[3]	Presentation	14
14	R integration with PHP	6	[4]	Presentation	15
15	R integration with PHP	6	[10]	Presentation	15

9. Student performance evaluation system for the course

Period	Assignments	Number of points	Total
1 st attestation	laboratory works:	40	100
	1 LW,	6	
	2 LW,	6	
	3LW,	6	
	4LW,	6	
	5LW,	6	
	Practice:	30	
	1 Exercise,	5	
	2 Exercise,	5	
	3 Exercise,	5	
4 Exercise,	5		
5 Exercise,	5		
	Mid term	30	
2 nd attestation	laboratory works:	40	100
	1 LW,	6	
	2 LW,	6	
	3LW,	6	
	4 LW,	6	
	5 LW,	6	
	Practice:	30	

	1 Exercise, 2 Exercise, 3 Exercise, 4 Exercise, 5 Exercise, End of term	5 5 5 5 5 30	
Final exam	Exam		100
Total	0,3*1stAtt+0,3*2ndAtt+0,4*Final		100

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

Achievement level as per course curriculum shall be assessed according to the evaluation chart adopted by the academic credit system:

Letter Grade	Numerical equivalent	Points (%)	Traditional system assessment	General description of grading criteria
A	4,0	95-100	Excellent	The student has knowledge of the subject in the full scope of the curriculum, understands the discipline deeply enough; shows a high level of knowledge that exceeds the volume provided by the syllabus, gives an exhaustive answer
A-	3,67	90-94		The student has knowledge of the subject in the full scope of the curriculum, understands the discipline deeply enough; gives an exhaustive answer
B+	3,33	85-89	Good	The student shows a complete, well-founded knowledge of the subject, but the answers did not always highlight the main idea, rational methods of calculation were not always used; the answers were mostly brief and sometimes unclear.
B	3,0	80-84		
B-	2,67	75-79		
C+	2,33	70-74	Satisfactory	The student demonstrates sufficient knowledge of the subject, but without proper depth and justification, the answers are unclear and without proper logical sequence.
C	2,0	65-69		
C-	1,67	60-64		
D+	1,33	55-59		
D	1,0	50-54	Unsatisfactory	The student demonstrates insufficient knowledge of the subject, positive answers were not given to individual questions.
FX	0,5	25-49		
F	0	0-24		

