SHAKHMARDAN YESSENOV FOUNDATION

ALMA MATER MANAGEMENT UNIVERSITY

YESSENOV DATA LAB

10th June – 2nd August 2019
Almaty

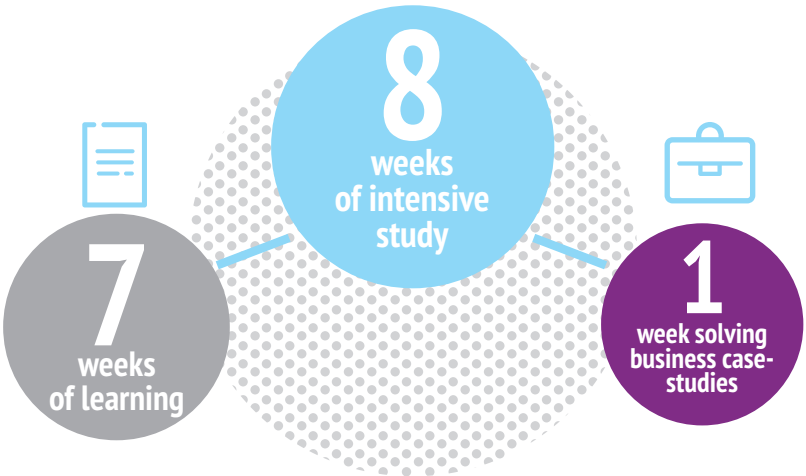
In partnership with: Яндекс BIGDATA TEAM

Contents

YESSENOV DATA LAB

What is the Yessenov Data Lab?

The Yessenov Data Lab summer school is an 8-week intensive course that allows you to familiarize yourself with the data analytics profession for 8 weeks, solve real business problems and continue to improve the knowledge you gain independently.



Course Dates: 10th June – 2nd August

Course location
Almaty Management University

COURSE GRADUATES WILL ACQUIRE THE SKILLS OF:

1. Python programming for data analysis
2. Preprocessing data in preparation for further analysis
3. Data visualization and finding correlations
4. Making a forecast based on historical data
5. Understanding different learning algorithms
6. Choosing the right learning model
7. Fundamental understanding of the work of neural networks

ALMA UNIVERSITY

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

Programme of Study

Week 1. Python Programming

Kuanish Abeshev
Dean of the School
Of Engineering and
Management, AlmaU

Timur Bakibayev
Professor, AlmaU

Day 1

09:00 – 10:00	Registration of participants
10:00 – 11:30	What is Data Mining, Big Data, examples
12:00 – 13:15	Introduction to Python: variables, lists
14:30 – 16:00	Introduction to Python: conditions, cycles
16:15 – 18:00	Lab: The Basics of Python

Day 2

10:00 – 11:30	Work with lines of code
12:00 – 13:15	Acquaintance with the library NumPy
14:30 – 16:00	Lab: NumPy tasks
16:15 – 18:00	Lab: NumPy tasks

Day 3

10:00 – 11:30	Data structures: lists, sets, dictionaries
12:00 – 13:15	Python Algorithms: Sorting, Search
14:30 – 16:00	Lab: data structures and NumPy
16:15 – 18:00	Lab: Implementation of Algorithms

Day 4

10:00 – 11:30	Introduction to the Pandas Library, MatPlotLib
12:00 – 13:15	Grouping of data. Filters, sorting.
14:30 – 16:00	Lab: work with Pandas.
16:15 – 18:00	Lab: work with MatPlotLib

Day 5

10:00 – 11:30	Fastening material by Pandas and MatplotLib
12:00 – 13:15	Acquaintance with the regression task, SciKitLearn
14:30 – 16:00	Lab: Linear Regression
16:15 – 18:00	Lab: Linear Regression

11:30 – 11:45	Coffee break
13:15 – 14:30	Lunch
16:00 – 16:15	Coffee break

30 ALMA UNIVERSITY

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

Programme of Study

Week 2. Linear models of classification and regression

Dimitri Rusanov
Data Scientist
elama.ru

Day 1

10:00 – 11:30	Optimization task, gradient descent method
12:00 – 13:15	Laboratory work
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

Day 2

10:00 – 11:30	Quality metrics
12:00 – 13:15	Laboratory work
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

Day 3

10:00 – 11:30	Cross-validation
12:00 – 13:15	Laboratory work
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

Day 4

10:00 – 11:30	Linear classification and regression models
12:00 – 13:15	Laboratory work
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

Day 5

10:00 – 11:30	Review of study, generalization
12:00 – 13:15	Laboratory work
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

11:30 – 11:45	Coffee break
13:15 – 14:30	Lunch
16:00 – 16:15	Coffee break

30 ALMA UNIVERSITY

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

Programme of Study

Week 3. Trees and Boosting

BIGDATA TEAM

Day 1

10:00 – 11:30 SVM, optimization tasks
12:00 – 13:15 Practice, master class
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 2

10:00 – 11:30 Trees
12:00 – 13:15 Practice, master class
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 3

10:00 – 11:30 DecisiveTree Ensembles (bagging)
12:00 – 13:15 Practice, master class
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 4

10:00 – 11:30 Informational features of signs, SHAP (SHapley Additive exPlanations)
12:00 – 13:15 Practice, master class
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 5

10:00 – 11:30 Decisive Tree Ensembles (bias-variance trade-off, XGBoost)
12:00 – 13:15 Practice, master class
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 6

10:00 – 11:30 Detailed CatBoost Review
12:00 – 13:15 Practice, master class
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

11:30 – 11:45 Coffee break
13:15 – 14:30 Lunch
16:00 – 16:15 Coffee break

30 ALMA UNIVERSITY

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

Programme of Study

Week 4. Independent learning and neural networks

Day 1

10:00 – 11:30 Independent Learning
12:00 – 13:15 Practice, masterclass
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 2

10:00 – 11:30 Independent Learning, part 2
12:00 – 13:15 Practice, master class
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 3

10:00 – 11:30 Introduction to Neural Networks
12:00 – 13:15 Practice, masterclass
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 4

10:00 – 11:30 Complex neural networks
12:00 – 13:15 Practice, masterclass
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 5

10:00 – 11:30 Recurrent Neural Networks
12:00 – 13:15 Practice, masterclass
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

Day 6

10:00 – 11:30 ML applications: RecSys
12:00 – 13:15 Practice, masterclass
14:30 – 16:00 Laboratory work
16:15 – 18:00 Lab, Q & A, discussion of standard errors

11:30 – 11:45 Coffee break
13:15 – 14:30 Lunch
16:00 – 16:15 Coffee break

BIGDATA TEAM

ALMA UNIVERSITY

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

Programme of Study

Week 5. ML Applications

Marina Gorlova
Analyst
Yandex money

Day 1

10:00 – 11:30	Recommender systems. User / Item based. Collaborative filtering
12:00 – 13:15	Recommender systems. Hybrid recommender
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

Day 2

10:00 – 11:30	Image analysis. Solving problems pre-neural networks
12:00 – 13:15	Image analysis. Neural Network Applications
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

Day 3

10:00 – 11:30	Text analysis. Embedding
12:00 – 13:15	Text analysis. Applications
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

Day 4

10:00 – 11:30	Time series. Statistical approaches
12:00 – 13:15	Time series. Neural Network Applications
14:30 – 16:00	Laboratory work
16:15 – 18:00	Laboratory work

Day 5

10:00 – 11:30	Analysis of graphs
12:00 – 13:15	Laboratory work
14:30 – 16:00	Case-studies
16:15 – 18:00	Case-studies

11:30 – 11:45	Coffee break
13:15 – 14:30	Lunch
16:00 – 16:15	Coffee break

30 ALMA UNIVERSITY

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

Programme of Study

Week 6. Big Data - I

Valery Zhuk
Senior Cloud Computing Engineer

clearscale

Day 1

10:00 – 11:30 Business problems
12:00 – 13:15 NoSQL
14:30 – 16:00 Hadoop: Yarn, MapReduce, Hive, Hdfs, Spark
16:15 – 18:00 Continuation

Day 2

10:00 – 11:30 MapReduce Internals + Hive
12:00 – 13:15 HiveQL Basics (UI or CLI?)
14:30 – 16:00 Laboratory work
16:15 – 18:00 Laboratory work

Day 3

10:00 – 11:30 Spark Core
12:00 – 13:15 Practical exercises for pySpark + RDD
14:30 – 16:00 Practical exercises for pySpark + RDD
16:15 – 18:00 Practical exercises for pySpark + RDD

Day 4

10:00 – 11:30 Spark SQL
12:00 – 13:15 Practical exercises for pySpark + DataFrame
14:30 – 16:00 Practical exercises for pySpark + DataFrame
16:15 – 18:00 Practical exercises for pySpark + DataFrame

Day 5

10:00 – 11:30 Difference in the use of the resources of Compute & Storage
12:00 – 13:15 Analogs of Tools
14:30 – 16:00 Case studies
16:15 – 18:00 Case studies

11:30 – 11:45 Coffee break
13:15 – 14:30 Lunch
16:00 – 16:15 Coffee break

ALMA UNIVERSITY

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

Programme of Study

Week 7. Big Data - II

Mikhael Lipkovich
Senior Data Scientist
assaia
The Apron AI

Day 1

10:00 – 11:30	Business problems
12:00 – 13:15	The Problem of Using Data Science on Big Data
14:30 – 16:00	Overview of tools
16:15 – 18:00	Laboratory work

Day 2

10:00 – 11:30	Spark MLlib
12:00 – 13:15	Spark MLlib (continued)
14:30 – 16:00	Practical exercises for pySpark + MLlib
16:15 – 18:00	Practical exercises for pySpark + MLlib

Day 3

10:00 – 11:30	Analysis of data on columns
12:00 – 13:15	GraphX Review
14:30 – 16:00	Practical exercises for pySpark + GraphX
16:15 – 18:00	Practical exercises for pySpark + GraphX

Day 4

10:00 – 11:30	Data Science on the Cloud
12:00 – 13:15	General questions: where to get the data from?
14:30 – 16:00	General questions: implementation solutions
16:15 – 18:00	General questions: work as a data analyst

Day 5

10:00 – 11:30	Case studies
12:00 – 13:15	Case studies
14:30 – 16:00	Case studies
16:15 – 18:00	Case studies

11:30 – 11:45	Coffee break
13:15 – 14:30	Lunch
16:00 – 16:15	Coffee break

30 ALMA UNIVERSITY

SHAKHMARDAN YESSENOV FOUNDATION

YESSENOV DATA LAB

Programme of Study

Week 8. Project Work

Participants are expected to work
6 hours a day during project week

OUR PROGRAMS

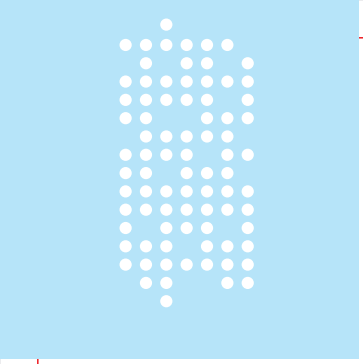
Knowledge	Science	Resources
Yessenov Lectures	Research internships	IT skills
English language	Yessenov scholarship	Komanda SOS
Promotion of science	Graduate studies	Personal performance

Kazakhstan Chess Federation

Almaty Marathon

Almaty Triathlon Federation

STAY: yessenovfoundation.org
IN: info@yessenovfoundation.org
TOUCH: + 7 727 346 92 88



YESSENOV
DATA LAB

SHAKHMARDAN YESSENOV
FOUNDATION

ALMA
UNIVERSITY

In partnership with: Яндекс

BIGDATA TEAM