



**Hotărârea nr. 246/18.12.2019**

*cu privire la aprobarea*

**Raportului privind organizarea Școlii de vară „Data Science”**

În temeiul Hotărârii Consiliului de Administrație nr. 463/27.11.2019 cu privire la înaintarea către Senatul ASE a propunerii de aprobare a raportului privind organizarea Școlii de vară „Data Science”;

În conformitate cu art. 213 alin. 2 lit. n) din Legea nr. 1/2011 a educației naționale, modificată și completată, cu art. 38 alin. 36 din Carta ASE, cu art. 22 alin. 53 din Regulamentul de organizare și funcționare a Senatului ASE, precum și cu Hotărârea Senatului nr. 10/20.01.2016;

Senatul Academiei de Studii Economice din București

**Hotărăște**

**Art. 1.** Aprobarea Raportului privind organizarea Școlii de vară „Data Science”, desfășurată în perioada 16 - 24 august 2019, conform anexei care face parte integrantă din prezenta hotărâre.

**Art. 2.** Consiliul de Administrație va duce la îndeplinire prevederile prezentei hotărâri.

**Președinte Senat,**

**Prof. univ. dr. Pavel NĂSTASE**

**Cancelar Senat,**

**Prof. univ. dr. Paul POCATILU**

**Avizat pentru legalitate,**

**Director Direcția Juridică și Contencios Administrativ**

**Consilier Juridic Elena GĂMAN**



**ACADEMIA DE STUDII ECONOMICE DIN BUCUREȘTI**  
**DIRECȚIA RELAȚII INTERNAȚIONALE**

Piața Romană, Nr. 6, 010374 - București, România  
Tel.: 021.319.19.10, Fax: 021.319.20.23, int. 417 / 232, E-mail: [international@ase.ro](mailto:international@ase.ro)



## Raport de activitate

### Școala internațională de vară Data Science – ediția a 2-a

Școala internațională de vară Data Science s-a desfășurat în perioada 16 – 24.08.2019, la Centrul de Perfecționare Complex Predeal „Ion Gh. Roșca”.

Activitățile au fost dezvoltate la inițiativa Direcției de Relații Internaționale, cu sprijinul Facultății de Cibernetică, Statistică și Informatică Economică și al Facultății de Finanțe, Asigurări, Bănci și Burse de Valori, în cadrul unui proiect finanțat din fondul MEN pentru finanțarea situațiilor speciale.

#### **PREAMBUL**

Proiectul a continuat fructificarea relațiile inter-universitare stabilite în urma unei vizite instituționale la Universitatea din Chicago a conducerii ASE și relațiile cu diaspora românească din SUA și Europa. De asemenea, proiectul a reprezentat a doua ediție a evenimentului care va continua și în anul 2020.

Proiectul a vizat formarea masteranzilor, doctoranzilor și tinerilor profesioniști de către specialiști de top în domeniul „data science”.

Domeniul este unul prioritar pentru universitatea noastră și pentru România și s-a urmărit acoperirea tuturor ariilor majore din cadrul data science printr-un program intensiv, care a combinat aspectele teoretice cu sesiuni aplicative, proiecte și prezentări inspiraționale.

#### **PROGRAM**

Derulat pe parcursul a 7 zile, programul de studiu a acoperit o varietate de aspecte esențiale ale data science: statistică, machine learning, rețele neurale, text mining, data management, data visualisation, aplicații financiare și economice ale data science. Toate activitățile s-au desfășurat în limba engleză.

Participanții au beneficiat de câte șapte ore de activități educaționale zilnic, în regim intensiv (8 ore / zi: 3 ore dimineața, 3 ore după-amiaza și 2 ore de lucru la proiectele individuale și de echipă).

Pe parcursul întregii perioade au fost predate 12 cursuri si sesiuni applicative (cu durata de 3 ore fiecare).

De asemenea, participanții au realizat și prezentat proiecte în echipă si individuale. Programul activităților didactice si de lucru in echipa a acoperit 51 de ore, la care se adaugă activitățile de socializare (reuniunea și cina de bun-venit, evenimentul social, cina de la finalul programului).

Pe parcursul programului, participanții s-au familiarizat cu instrumente software utilizate în analiza datelor: Python, NumPy, SciPy, Matplotlib, nltk, Jupyter Notebook, Ipython, Pandas, Scikit-learn, Seaborn.

Anexăm prezentului raport programul complet al activităților (Anexa 1).

## **LECTORI**

Lectorii recrutați pentru activitățile desfășurate sunt specialiști de top în domeniul data science, profesori și cercetători provenind de la universități de prestigiu și activând în proiecte cu impact major:

- Prof. dr. Dan Nicolae – Universitatea din Chicago, SUA,
- Conf. dr. Răzvan Bunescu – Universitatea din Ohio, SUA,
- Conf. dr. Carlos A. Iglesias – Universidad Politécnica de Madrid, Spania,
- Dr. Mihai Lupu – Research Studios, Austria,
- Conf. dr. Dumitru Roman – SINTEF / Universitatea din Oslo, Norvegia.

Detalii despre lectori se găsesc pe paginile de LinkedIn ale acestora și sunt incluse în programul școlii de vară (Anexa 1).

## **PARTICIPANȚI**

Proiectul a vizat în special consolidarea conexiunilor cu mediul de afaceri și s-a axat pe formarea unei comunități locale de tineri specialiști în data science. Dat fiind caracterul intensiv și foarte specializat al programului și dorința de a propune un nivel de elită au fost selectați 19 participanți care să aibă competențe ridicate în domenii ca algebră liniară, teoria probabilităților, statistică, econometrie, programare și reprezentare de date.

Candidaturile au fost depuse online, pe site-ul școlii de vară ([www.datascience.ase.ro](http://www.datascience.ase.ro)). Cei 19 participanți au fost selectați din 27 de candidați, ținând cont de criteriile enumerate anterior și motivația acestora. Candidaturile au provenit din România, Norvegia, Cehia, Africa de Sud, India.

Selecția candidaților a urmărit crearea unui grup cu competențe compatibile, provenind atât din mediul academic (8 masteranzi, doctoranzi și tinere cadre didactice – 42,1%) cât și din cel profesional (tineri profesioniști – 57,9%), 73,7% români. De asemenea, selecția a încercat să

crească componența internațională a grupului. Această structură a grupului a permis crearea de sinergii atât între participanți, cât și cu profesorii. Lista participanților este prezentată în Anexa 2.

Prin comunitatea creată pe parcursul școlii de vară, estimăm că impactul acesteia va depăși numărul participanților direcți prin includerea cunoștințelor și competențelor dobândite în teze de doctorat / proiecte de masterat, articole științifice și activități de analiză de date în companii care activează în România.

## **ASPECTE FINANCIARE**

Finanțarea școlii de vară a fost din taxele de participare și sponsorizările obținute. Fondurile au fost utilizate pentru acoperirea cheltuielilor de transport pentru lectorii străini, a cheltuielilor cu masa pe parcursul școlii de vară, a cheltuielilor de cazare și a cheltuielilor cu evenimentul social.

3. Bugetul total a fost de 78.159,2 lei

Venituri:

- taxe de participare: 76.550 lei (echivalentul a 400 Euro pentru masteranzii și doctoranzii români, 1000 pentru participanții din mediul de afaceri și din străinătate). Menționăm că taxa de participare a inclus cazarea și masa pentru participanți pe toată durata școlii de vară.
- sponsorizare Ericsson România (cheltuieli achitate direct la furnizori): 16 000 lei + participant pack (1200 lei)
- alte surse (bilete de avion achiziționate de profesori din fonduri proprii): 9500 lei

Cheltuieli:

- transport internațional (bilete de avion pe rute externe – profesori invitați): 18.891,47 lei
- cazare Complex Predeal: 19060 lei
- catering și servirea mesei la restaurant: 25358.52 lei
- eveniment de socializare (excursie): 14.849,21 lei

Sold: 25306.87 lei (pentru finanțarea următoarei ediții)

## **FEEDBACK**

Conform chestionarului de satisfacție aplicat la finalul școlii de vară toți participanții se declară satisfăcuți și foarte satisfăcuți de experiență. De asemenea, majoritatea respondenților au declarat că vor recomanda colegilor participarea la viitoarele ediții și consideră că programul le este util pentru dezvoltarea profesională și progresul în carieră. De asemenea, respondenții se declară mulțumiți și foarte mulțumiți de conținutul sesiunilor de formare în raport cu așteptările lor. Aproape 88% dintre respondenți au menționat că iau în considerare și participarea la ediții viitoare ale școlii de vară.

Participanții au apreciat de asemenea calitatea organizării, condițiile de cazare, mesele oferite, materiale primite și conținutul sesiunilor. Printre sugestiile de îmbunătățire se numără

dezvoltarea programului astfel încât să permită realizarea unor cursuri destinate companiilor, care să permită aprofundarea temelor prezentate.

În concluzie, considerăm obiectivele proiectului realizate. În baza interacțiunii cu participanții și lectorii invitați, considerăm utilă continuarea organizării acestei școli de vară.

Data science este unul dintre domeniile prioritare de dezvoltare în lume și este necesară dezvoltarea unei comunități puternice în jurul acestui domeniu la nivel național pentru a contribui la creșterea competitivității economice și academice de nivel național. Menționăm că școala de vară organizată de universitatea noastră și ajunsă la a doua ediție a fost prima din România pe această temă și considerăm că contribuie într-un mod semnificativ la capitalul de imagine pentru universitatea noastră, plasându-ne printre actorii principali ai domeniului, la nivel național. De asemenea, acest proiect răspunde mai multor obiective de internaționalizare din strategia instituțională și poate constitui fundamentul unei colaborări viitoare de anvergură mai mare cu universitățile de la care provin lectorii și cu companiile (companii multinaționale) de la care provin o parte dintre cursanți.

Prof. univ. dr. Constantin Marius Profiroiu  
Prorector Relații Internaționale



## Anexa 1 – Program

Data Science Summer School  
August 17-23, 2019  
Predeal, Romania

### *Goal:*

The goal of the summer school is to familiarize students with relevant state of the art topics in data science. The program will cover fundamentals of data science and focus on the following key data science topics:

- Data analytics and statistics
- Machine learning
- Text processing
- Data management and visualization

The program will consist of a combination of lecture-style talks introducing various data science paradigms and methods, hands-on sessions, and student projects.

The summer school aims to have a practical orientation, with Python and Jupyter Notebooks being used to exemplify many of the topics covered at the summer school.

At the end of the summer school, the students are expected to have an understanding of key paradigms used in data science and be able to practically apply them in data science projects.

*Prerequisites:* Familiarity with computer programming and basic knowledge about Python, interest in working with data, enthusiasm, and willingness to learn new things!

Basic knowledge of linear algebra, probability theory, knowledge representation would be useful, though not strictly necessary.

### *Speakers:*

- **Dan Nicolae** (University of Chicago, USA), Professor, PhD  
[www.linkedin.com/in/dan-nicolae-221991a](http://www.linkedin.com/in/dan-nicolae-221991a)
- **Razvan Bunescu** (Ohio University, USA), Professor, PhD  
<https://www.linkedin.com/in/razvan-bunescu-8097956>
- **Mihai Lupu** (Research Studios Austria), Data Science Studio Director, PhD  
<https://www.linkedin.com/in/mihailupu>
- **Carlos A. Iglesias** (Universidad Politécnica de Madrid, Spain), Associate Professor, PhD  
<https://www.linkedin.com/in/cif2cif>
- **Dumitru Roman** (SINTEF / University of Oslo, Norway), Senior Research Scientist / Associate Professor, PhD  
<https://www.linkedin.com/in/titiroman>

**SCHEDULE**

		<b>Day 1</b> (17.08)	<b>Day 2</b> (18.08)	<b>Day 3</b> (19.08)	<b>Day 4</b> (20.08)	<b>Day 5</b> (21.08)	<b>Day 6</b> (22.08)	<b>Day 7</b> (23.08)	
<b>Breakfast (8am-9am)</b>									
<b>Morning session (9am-12pm)</b>		Statistics for data science	Statistical learning (hands-on)	Machine Learning (hands-on)	Intro to text processing	Data modeling and databases (incl. hands-on)	Data cleaning and preprocessing (incl. hands-on)	Data visualization (incl. hands-on)	
<b>Lunch (12am-2pm)</b>									
<b>Afternoon session (2pm-5pm)</b>	<b>Arrivals (17.08)</b>	Data science with Python (hands-on) & Projects discussions	Intro to Machine Learning	Deep Learning with Neural Networks (intro and hands-on)	Text processing (hands-on) & Use case: Intellectual Property	<b>Social event</b>	Graph analytics (incl. hands-on)	Individual group/project work & Projects presentations and final discussions	<b>Departures (24.08)</b>
<b>Individual group/project work and/or free time (5pm-7pm)</b>									
<b>Dinner (7pm-9pm)</b>									

### **Statistics for data science**

- A data science pipeline
- Data exploration
- Statistical inference with resampling methods

### **Data science with Python (hands-on) & Projects discussions**

- Intro to Python
- Pandas and data frames
- Probability and simulations

### **Statistical learning (hands-on)**

- Regression models and inference
- Prediction and classification

### **Intro to Machine Learning**

- Feature vector representations
- ML for Classification
  - The Perceptron
  - Logistic Regression
  - Linear vs. Non-linear decision boundaries
- Optimization with Stochastic Gradient Descent
- Clustering with k-Means

### **Machine Learning (hands-on)**

- ML algorithms in Python
  - Implementation using NumPy
  - The *sklearn* library
  - Visualization using Matplotlib
- Experimental evaluation of ML models
  - Perceptron vs. Logistic Regression vs. SVMs
  - Linear vs. non-linear classification

### **Deep Learning with Neural Networks (intro and hands-on)**

- Representation learning
- Deep learning with neural networks
  - Fully connected
  - Convolutional Neural Networks (CNN)
  - Recurrent Neural Networks (RNN)
- Implementation in PyTorch
  - Regression and gradient descent
  - Logistic regression and NNs for non-linear classification
  - RNNs for time series prediction



### **Intro to text processing**

- From text to data
  - Tokenization, Stemming/Lemmatization, Stop-words, Spell-checking
- Finding text (search engines)
  - Inverted index, similarity models
- Learning on text
  - Natural Language Processing – chunking, parsing sentences
  - Latent Semantic Indexing
  - Word2vec and other term representation models
- Combining text and non-text data for ML

### **Text processing (hands-on) & Use case: Intellectual Property**

- Hands-on: Use Spacy and/or genism to process a small amount of text (Abstracts of scientific articles)
- Use-case intellectual property

### **Data modeling and databases (incl. hands-on)**

- Introduction to databases (relational and NoSQL)
- Relational model and data modeling (focus on ER)
- Graph databases (focus on Neo4j)
  - Data model and data modeling
  - Query language
  - Graph algorithms

### **Data cleaning and preprocessing (incl. hands-on)**

- Overview of data quality issues
- Data cleaning and preprocessing with Pandas
  - Importing data, preprocessing, transformations, etc.
  - Visual exploration of data (incl. use of Seaborn)
- Using cleaned data for ML

### **Graph analytics (incl. hands-on)**

- Intro to graph analytics
- Graph analytics with NetworkX
  - Graph manipulation
  - Graph construction
  - Graph analysis metrics
  - Social networks (ego networks, etc.)
  - Clustering and community detection
  - (Link prediction)

### **Data visualization (incl. hands-on)**

- General intro to data visualization
- Data visualization techniques with Seaborn

**Software:** Software tools/services to be used during the sessions and hands-on include:

- Anaconda (<https://www.anaconda.com/>): Installation instructions for various platforms can be found at: <https://docs.anaconda.com/anaconda/install/>
  - Note: For Mac and Linux users, the system PATH must be updated after installation so that 'conda' can be used from the command line.
    - Mac OS X:
      - For bash users: export PATH=~/anaconda3/bin:\$PATH
      - For csh/tcsh users: setenv PATH ~/anaconda3/bin:\$PATH
    - Linux:
      - For bash users: export PATH=~/anaconda3/bin:\$PATH
      - For csh/tcsh users: setenv PATH ~/anaconda3/bin:\$PATH
    - It is recommend the above statement be put in the ~/.bashrc or ~/.cshrc file, so that it is executed every time a new terminal window is open.
    - To check that conda was installed, running "conda list" in the terminal should list all packages that come with Anaconda.
  - A number of tools and libraries that we will use can be configured from Anaconda: Python 3, NumPy, SciPy, Matplotlib, nltk, Jupyter Notebook, Ipython, Pandas, Scikit-learn and Seaborn.
  - For Deep Learning we will use PyTorch (<https://pytorch.org/>): This can be installed from Anaconda, with 'conda' from the command line, the actual command line depends on the platform as follows: using the GUI on pytorch.org, choose the appropriate OS, conda, Python 3.6, CUDA None (for the examples the CPU version should suffice).
  - For Text Processing, SpaCy can be installed with 'conda' from the command line as shown at <https://pypi.org/project/spacy>
- Neo4j (<https://neo4j.com/>): Installation and documentation can be found at <https://neo4j.com/developer/get-started/> (we will likely use the online sandbox service provided at <https://neo4j.com/sandbox-v2/>, so no installation on local machines may be needed for experimenting with Neo4j).

## Anexa 2

Lista participanți:

<i>Nr. crt.</i>	<i>Prenume</i>	<i>Nume</i>	<i>Universitate / companie</i>
3	Eugen	Birsan	Vodafone România
8	Megha	Bolia	Ericsson India Global Pvt
19	Raluca	Căplescu	ASE București
14	Manuela-Simona	Cojocea	Universitatea București
9	Maneesh	Disodia	Ericsson India Global Pvt
15	Ionut	Edu	Universitatea Politehnica București
13	Viviana	Ghizdavin Dragu	Ericsson Romania
7	Sapan Sagar	Gupta	Ericsson India Global Pvt
12	Martin	Hasal	Technical University of Ostrava
11	Ruxandra	Iancu	Ericsson Romania
4	Cristiana	Istrate	Universitatea Politehnica București
6	Codrut	Ivascu	ASE București
5	Petrica	Marian	Universitatea București
17	Vlad	Nedelcu	Ernst & Young Romania
18	Adrian	Otoiu	ASE București
2	Adelina	Sirbu	Ernst & Young Romania
10	Ioan - Gabriel	Stavre	Banca Națională a României
1	Miruna	Ștefan	Ernst & Young Romania
16	Philip	Turk	SINTEF Oslo