BANCO DE PORTUGAL
EUROSYSTEM

Workshop: "Machine Learning" and "Big Data": basic concepts and applications
Hosted by BPLIM

18/19 December, 2018 | Porto

Tuesday, December 18

13:30 – 13:50 – Registration

13:50 – 14:00 – Welcome address and introduction
Paulo Guimarães (BdP)

14:00 – 15:00 – Keynote Speech:
"A common path in innovation: from Academia through Private Companies to Public Administration"
Fabio Fumarola (Italian Digital and Transformation Team)

15:00 – 15:15 – Coffee break

15:15 – 17:00 – Training Session 1: Data Collection
Nelson Areal (U Minho)

17:00 – 17:15 – Coffee break

17:15 – 19:00 – Training Session 2: Big Data and Distributed Computing
Nelson Areal (U Minho)

Wednesday, December 19

08:50 – 09:00 – Welcome

09:00 – 10:45 – Training Session 3: Machine Learning I
João Gama (FEP)

10:45 – 11:15 – Coffee break

11:15 – 13:00 – Training Session 4: Machine Learning II
João Gama (FEP)

13:00 – 14:00 – Lunch

14:00 – 14:15 – Welcome address
Ana Paula Serra (BdP)

14:15 – 15:30 – Session 1

Behavioral models in Banking
Maria R. Sousa (BCP)

How can data be organization's Energy
Liliana Bernardino (Sonae MC)

Decision support systems to relocate vehicles in sharing services
Nuno Oliveira (CEiia Portugal)

15:30 – 15:45 – Coffee break

15:45 – 17:00 – Session 2

Detection of Frauds and Transactions in Local Trade: SAF-T with Big Data Technology
Eurico Lopes (IPCB)

Mining Social Media to Forecast Stock Market Behavior
Paulo Cortez (U Minho)

Detection of outliers for the inspection of potential fraud cases
Rita P. Ribeiro (FCU Porto)
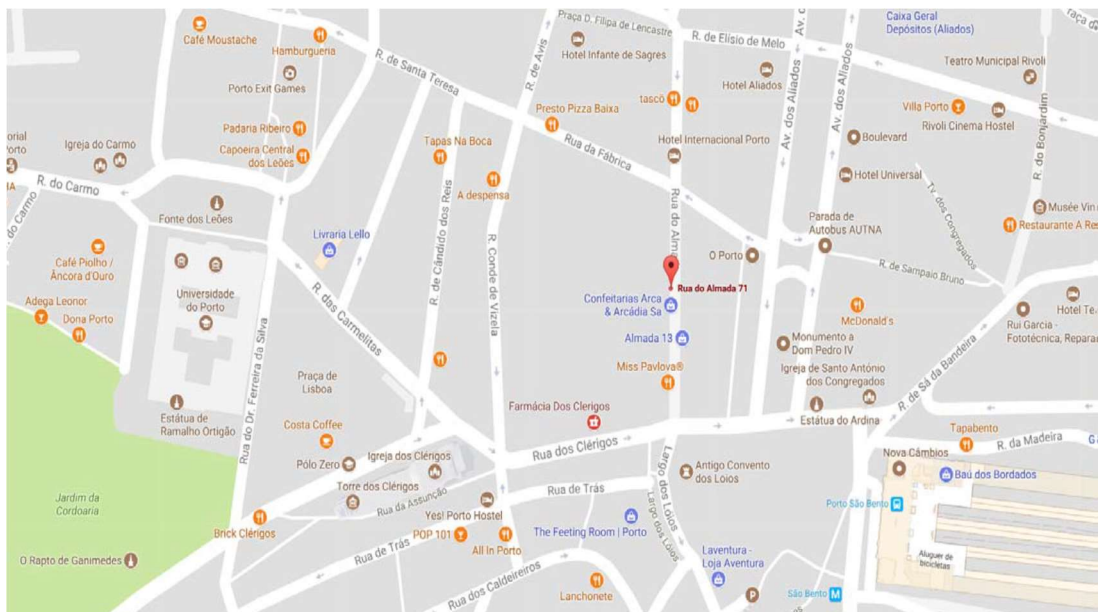
17:00 – End of Workshop

Practical Information:

Location

The workshop will take place at the premises of Banco de Portugal in the city of Porto. This location is right at city center. The address is:

Rua do Almada 71
4000-196 Porto

For any enquiries please contact bplim@bportugal.pt

TRAINING SESSIONS PROGRAM:

TRAINING SESSION 1 (Nelson Areal):

1. Data collection
   - Web data sources (Webscrapping, Using APIs)
   - Data cleaning and merging datasets (Regular expressions, Record Disambiguation)

TRAINING SESSION 2 (Nelson Areal):

2. Working with large datasets:
   - Bigish data:
       - Relational Databases: slqlite, PostgreSQL
       - Column store databases: MonetDB
       - Document database: MongoDB
   - Bigger data: Distributed datastores
3. Distributed computing:
   - Parallel computation
   - Distributed computing (Spark): Using a local cluster / Using the cloud
   - GPU

TRAINING SESSION 3 (João Gama):

   - Supervised Learning:
   - Classifying people or things into groups by recognizing patterns
   - Regression - relationship between a dependent continuous variable and one or more independent variables:
       - Classification and Regression Trees
       - Gradient Descent and Neural Networks
       - Support Vector Machines and Kernel Machines
       - Random Forests
       - Boosting

TRAINING SESSION 4 (João Gama):

   - Unsupervised Learning
   - Associating what events are likely to occur together
   - Social Network analysis
   - Recommender Systems

Biographies:

Fabio Fumarola received a Ph.D. in Data Mining and Machine Learning in May 2011 from the University of Bari (Italy) with supervisor Professor Donato Malerba. During the Ph.D. period, he visited University of Illinois working with the Professor Jiawei Han. After the Ph.D., he got a post doc working on Stream Mining, Web-mining and Big Data. In 2015, he worked as Research Scientist at Unicredit Bank, where he applied Natural Language Processing, Graph Mining and Machine Learning to create innovative products. During 2017 he joined the Italian Digital and Transformation Team as Machine learning Engineer to contribute in changing the way Italian data are used to create public value and innovative services.

Nelson Areal is an Associate Professor at the School of Economics and Management, University of Minho. His research interests are in risk measures and forecasting, option valuation using numerical methods, performance measurement, social responsible investments and text analysis. His work has appeared in various international journals, such as: International Journal of Finance & Economics, Journal of Futures Markets, Journal of Business Ethics, Review of Derivatives Research, Quantitative Finance, Decision Support Systems and Expert Systems with Applications. He obtained his PhD in Accounting and Finance from Lancaster University (UK).

João Gama is an Associate Professor of the Faculty of Economy, University of Porto. He is a researcher and vice-director of LIAAD, a group belonging to INESC TEC. He got the PhD degree from the University of Porto, in 2000. He is senior member of IEEE. He has worked in several National and European projects on Incremental and Adaptive learning systems, Ubiquitous Knowledge Discovery, Learning from Massive, and Structured Data, etc. He served as Co-Program chair of ECML'2005, DS'2009, ADMA'2009, IDA' 2011, and ECML/PKDD'2015. He served as track chair on Data Streams with ACM SAC from 2007 till 2016. He organized a series of Workshops on Knowledge Discovery from Data Streams with ECML/PKDD, and Knowledge Discovery from Sensor Data with ACM SIGKDD. He is author of several books in Data Mining (in Portuguese) and authored a monograph on Knowledge Discovery from Data Streams. He authored more than 250 peer-reviewed papers in areas related to machine learning, data mining, and data streams. He is a member of the editorial board of international journals ML, DMKD, TKDE, IDA, NGC, and KAIS. He (co-)supervised more than 12 PhD students and 50 Msc students.

Maria R. Sousa holds a BSc in Mathematics, 2003, an MSc in Mathematical Engineering, 2007, and a Ph.D. in Management, field of Finance, 2016, all from the University of Porto. Maria is a Credit Risk Modelling Specialist at the Models Validation and Monitoring Office in BCP group, where she has been working in several projects of credit risk modelling (development and coordination) and in regulatory matters (e.g. Basel Accords, ECB TRIM exercises, EBA Benchmarking and IFRS9). Her research is focused in dynamic credit risk modelling, in the context of international financial regulation. Maria has received numerous awards, including the recognition for her influential attitude and outstanding impact to the teams and to the institution, in 2006 and in 2018, and the first place in the BRICS-CCI & CBIC 2013 competition in Data Mining and Finance.

Liliana Bernardino moved from Trás-os-Montes to Porto in 1999 to get Applied Mathematics' Degree in Oporto University. She started her professional journey as a Junior Data Analyst at SONAE in 2004, being focused on Ticket and Consumer Trips Analysis. In 2007 joined Loyalty Card Project Team and in 2011 assumed Customer Intelligence Team' leadership getting in the meanwhile her Masters' Degree focused in Advanced Analytics. Currently she leads Loyalty Card Analytical team with roles both on Campaign Intelligence (CRM) and Advanced Analytics in Sonae MC.

Nuno Oliveira is a Data Scientist working on smart mobility services at CEiiA, Portugal. He completed his PhD (2017) in Information Systems and Technologies at University of Minho. His research interests include: machine learning, time series forecasting, geographical forecasting, natural language processing and behavioral finance.

Eurico Lopes has been a researcher in computer engineering and information systems since the 1990s. Prior to that he worked in the banking sector and in the industry as an engineer, programmer, systems analyst and trainer and latterly as a project leader. Whilst teaching and researching he established various collaborations, including several major projects with local enterprises involving technology transfer across the organization for development and implementing information systems. He is a member of Ordem dos Engenheiros, Senior Member in the IEEE and ACM and works with CCISP as a consultant on ICT and network projects, including work on PRODEP I & II. He was convened as an observer in the users panel on FCCN and member of the Installation Committee Technological and Management School in Castelo Branco and Idanha-a-Nova. Since April 2013 he has been an Expert at the A3ES. His background was in Electrical and Computer Engineering in the IST, Lisbon. He was supervised by Anne de Roeck in the master

dissertation, Master of Science, University of Essex, UK and by A. Bryant for his PhD at Leeds Metropolitan University, UK.

Paulo Cortez is an Associate Professor (with tenure) at the Department of Information Systems, University of Minho. He is also assistant director of the ALGORITMI R&D Centre and coordinator of its Information Systems and Technologies (IST) research group. His research, within the areas of Business Intelligence, Data Science, Modern Optimization and Forecasting, has appeared in Journal of Heuristics, Decision Support Systems, Information Sciences and others. Currently, he is an associate editor of the journals Expert Systems and Decision Support Systems.

Rita P. Ribeiro is an Assistant Professor at the Department of Computer Science of the Faculty of Sciences of the University of Porto and a Researcher at LIAAD-INESCTEC, the Artificial Intelligence and Decision Support Lab of University of Porto. She holds a PhD from the University of Porto in the field of Data Mining, with a thesis centered around utility-based regression and rare extreme values prediction. Her main research topics are in the fields of outlier detection, novelty detection, utility-based learning and evaluation issues on learning tasks. As a member of LIAAD-INESC TEC, she has been involved in several research projects concerning environmental problems, fraud detection and predictive maintenance applications.