

Έντυπο Καταγραφής Πληροφοριών και Συγκέντρωσης Εκπαιδευτικού Υλικού για τα Ανοικτά Μαθήματα

Έκδοση: 1.02, Απρίλιος 2014

Συντάκτης: Δρ. Παντελής Μπαλαούρας, Καθ. Λάζαρος Μεράκος (ΕΚΠΑ)

Προσαρμογή: Αν. Καθ. Γεώργιος Ξυλωμένος (ΟΠΑ)



ανοικτά μαθήματα
opencourses

**Δράση «Ανοικτά Ακαδημαϊκά Μαθήματα στο Οικονομικό
Πανεπιστήμιο Αθηνών»**

Σύνδεσμος: <http://ocw.aueb.gr>



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ, ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Περιεχόμενα

1. Χρήση εντύπου	3
2. Πληροφορίες και εκπαιδευτικό υλικό Ακαδημαϊκού Μαθήματος.....	4
2.1 Πληροφορίες μαθήματος.....	4
2.1.1 Όνομα διδάσκοντος/διδασκόντων (Instructor (s)).	4
2.1.2 Τίτλος Μαθήματος (Course title) όπως αναφέρεται στο πρόγραμμα σπουδών (ΠΣ)	4
2.1.3 Δικτυακός τόπος μαθήματος	4
2.2 Πληροφορίες για τις θεματικές ενότητες ή ενότητες διαλέξεων	9

1. Χρήση εντύπου

Το παρόν έντυπο χρησιμοποιείται για τη συγκέντρωση των πληροφοριών των μαθημάτων κατηγορίας A- και την πρώτη φάση ανάπτυξης των μαθημάτων κατηγορίας A και A+.

Οι πληροφορίες διακρίνονται σε υποχρεωτικές (πράσινοι πίνακες) και προαιρετικές (πορτοκαλί πίνακες). Η συμπλήρωση των υποχρεωτικών στοιχείων δεν απαιτεί ιδιαίτερο χρόνο. Ο όγκος του εντύπου εμφανίζεται μεγάλος καθώς υπάρχουν πολλές προαιρετικές πληροφορίες, όπως πληροφορίες και στην Αγγλική γλώσσα. **Παρακαλούνται τα μέλη ΔΕΠ/ΕΠ να μην αποθαρρύνονται από το όγκο του εντύπου.**

Πολλά από τα στοιχεία υπάρχουν ήδη στο πρόγραμμα σπουδών ή έχουν ήδη συγκεντρωθεί από τη Μονάδα Διασφάλισης Ποιότητας (ΜΟΔΙΠ) του κάθε Ιδρύματος.

Επικοινωνία: opencourses@aueb.gr

2. Πληροφορίες και εκπαιδευτικό υλικό Ακαδημαϊκού Μαθήματος

2.1 Πληροφορίες μαθήματος

2.1.1 Όνομα διδάσκοντος/διδασκόντων (Instructor (s)).

Μιχάλης Βαζιργιάννης

Michalis Vazirgiannis

2.1.2 Τίτλος Μαθήματος (Course title) όπως αναφέρεται στο πρόγραμμα σπουδών (ΠΣ)

Εξόρυξη γνώσης από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό

Data Science and Mining

2.1.3 Δικτυακός τόπος μαθήματος

<https://eclass.aueb.gr/courses/INF131/>

2.1.4 Κωδικός Μαθήματος (Course Code) όπως αναφέρεται στο ΠΣ

3743

2.1.5 Επίπεδο μαθήματος/Κύκλος σπουδών (Course level/cycle).

1. Προπτυχιακό (Undergraduate)/Πρώτος κύκλος σπουδών (First cycle)
2. Μεταπτυχιακό (Graduate)/Δεύτερος κύκλος σπουδών (Second cycle)
3. Διδακτορικό (Doctoral)/ Τρίτος κύκλος σπουδών (Third cycle)

2.1.6 Έτος σπουδών (Year of Study).

Έτος: 1 | 2 | 3 | 4 | 5 | 6

2.1.7 Εξάμηνο (Semester). Επιλέξτε (κάντε bold) από 1 έως 12 όπως αναφέρεται στο ΠΣ.

Εξάμηνο: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12

2.1.8 Τύπος μαθήματος (Type of course).

1. Υποχρεωτικό (compulsory)

2. Επιλογής (optional)

2.1.9 Διδακτικές ώρες στο εξάμηνο. 64

2.1.10 Συνδιδασκαλία. ΟΧΙ

2.1.11 Γλώσσα διδασκαλίας (Course language).

1. Ελληνική
2. Αγγλική
3. Άλλη: (δηλώστε)

2.1.12 Ομάδα στόχος (Target Group)

Προπτυχιακοί φοιτητές του τμήματος Πληροφορικής.

Undergraduate students of the Department of Informatics.

2.1.13 Πιστωτικές μονάδες (Credits.)

Αριθμός μονάδων: 6

2.1.14 Περισσότερα για τον/τους διδάσκοντες (More about instructor).

<http://www.db-net.aueb.gr/michalis/>

<http://www.db-net.aueb.gr/michalis/>

2.1.15 Φωτογραφία διδάσκοντος.



2.1.16 Περιγραφή μαθήματος (Course Overview / Description /Synopsis)

Ανασκόπηση αλγόριθμων επιβλεπόμενης μηχανικής μάθησης, μη επιβλεπόμενη μηχανική μάθηση με έμφαση στη συσταδοποίηση και στους κανόνες συσχέτισης. Βασικές έννοιες αναζήτησης πληροφορίας, τεχνικές αποδόμησης και μείωσης διάστασης των δεδομένων (SVD, PCA, LSI), αλυσίδες Markov. Σημασιολογική ανάλυση με εφαρμογές στην εξόρυξη γνώσης και στην αναζήτηση πληροφορίας. Σημασιολογία

και παγκόσμιος ιστός, εξόρυξη γνώσης στον παγκόσμιο ιστό (περιεχόμενο, δομή, δεδομένα χρήσης), ανάλυση υπερσυνδέσμων, κατάταξη σημαντικότητας, αλγόριθμοι εξατομικευσης (collaborative filtering)

Review of supervised machine learning, unsupervised machine learning with emphasis on clustering and association rules. Basic information retrieval concepts, spectral decomposition and dimension reduction techniques (SVD, PCA, LSI), Markov chains., Knowledge mining the Web (content, structure, usage data), hyperlinks analysis, page ranking, personalization and recommendations(collaborative filtering)

2.1.17 Περιεχόμενα μαθήματος (Course Contents)

1. Εισαγωγή στη διαδικασία Data Science
 - Pre-processing, Exploration, Feature selection, Dimensionality reduction, feature extraction and evaluation
 - Supervised Learning (k-nn, regression, logistic regression, decision trees)
 - Unsupervised learning (Clustering, K-means, EM, Spectral Clustering)
2. Εξόρυξη από Κείμενα και τον Παγκ. Ιστό (Text/Web Mining)
 - Text retrieval metrics, Novelty detection, Graph of Words
 - Recommendations/collaborative filtering
3. Εξόρυξη από Γραφήματα (Graph Mining)
 - Graph Ranking algorithms and evaluation measures
 - Graph clustering and classification
 - Degeneracy (k-core & extensions)
 - Community mining methods & applications in social networks
4. Bigdata
 - Map reduce - distributed processing, technologies (Hadoop, Map Reduce, NoSQL storage)

2.1.18 Μαθησιακοί στόχοι μαθήματος (Course Objectives/Goals)

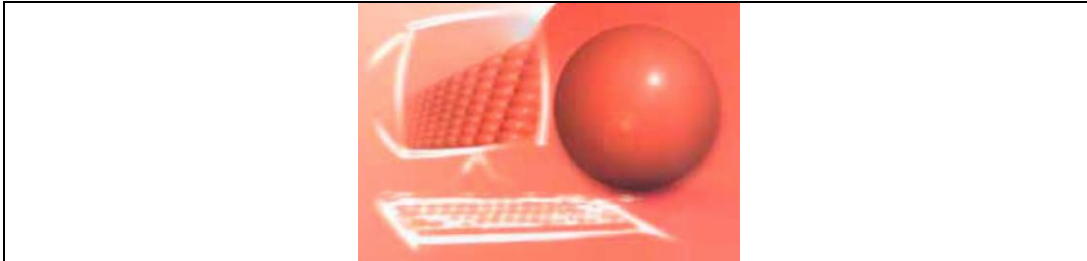
Ο στόχος του μαθήματος είναι η εξοικείωση με τεχνικές / αλγόριθμους ανάλυσης συνόλων δεδομένων και εργαλεία λογισμικού στον χώρο της εξαγωγής γνώσης και μηχανικής μάθησης. Ειδικότερα εισάγεται η διαδικασία της Επιστήμης Δεδομένων (Data Science) με τα στάδια προεπεξεργασία και εξερεύνηση δεδομένων, επιλογής χαρακτηριστικών/μείωσης διάστασης, επιβλεπόμενης/μη επιβλεπόμενης μάθησης και αξιολόγησης/επιλογής μοντέλων. Επίσης δίνεται μια εισαγωγή στις απαιτήσεις και τεχνολογίες μεγάλης κλίμακας δεδομένων (Bigdata).

2.1.19 Λέξεις κλειδιά (Keywords)

Εξόρυξη Δεδομένων, Επιστήμη Δεδομένων, Επιβλεπόμενη Μάθηση, Μη επιβλεπόμενη Μάθηση, Data Exploration, Graph Mining, Text Mining, Web Mining.

Data Science, Supervised Learning, Unsupervised learning, Data Exploration, Graph Mining, Text Mining, Web Mining

2.1.20 Προτεινόμενη φωτογραφία για το μάθημα



2.1.21 Ομάδα ανάπτυξης περιεχομένου (Content Development).

Διαλέξεις: Μιχάλης Βαζιργιάννης, Καθηγητής ΟΠΑ.

2.1.22 Τύποι εκπαιδευτικού υλικού (course format).

- Διαφάνειες
- Σημειώσεις
- Βιντεοδιαλέξεις
- Podcast
- Ήχο
- Πολυμεσικό υλικό
- Διαδραστικές ασκήσεις

2.1.23 Προτεινόμενα συγγράμματα.

Μπορείτε να επιλέξετε ένα από τα ακόλουθα βιβλία:

- Principles of Data Mining , David Hand, Heikki Mannila, and Padhraic Smyth, MIT Press, August 2001.
- Learning from Data – Y. Abu-Mostafa, M. Magdon-Ismail, Hsuan-Tien Lin,
- Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt
- Pattern Recognition and Machine Learning (Information Science and Statistics) Hardcover – October 1, 2007, Christopher M. Bishop
- Hadoop: The Definitive Guide, 3rd Edition, T. WhiteSlides

You can choose one of the following textbooks:

- Principles of Data Mining , David Hand, Heikki Mannila, and Padhraic Smyth, MIT Press, August 2001.
- Learning from Data – Y. Abu-Mostafa, M. Magdon-Ismail, Hsuan-Tien Lin,

- Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt
- Pattern Recognition and Machine Learning (Information Science and Statistics) Hardcover – October 1, 2007, Christopher M. Bishop
- Hadoop: The Definitive Guide, 3rd Edition, T. WhiteSlides

2.1.24 Οργάνωση μαθήματος.

Δομή και συχνότητα διδασκαλίας (Course Meeting Times / Course Structure).

- Διαλέξεις 2 φορές την εβδομάδα (Lectures 2/week)
- Εργαστήριο (1 φορά την εβδομάδα, 2 ώρες)

2.1.25 Μέθοδος διδασκαλίας (teaching method)

- Διδασκαλία καθ' έδρας και συμπληρωματική-ενισχυτική εκπαίδευση μέσω ασύγχρονης πλατφόρμα.
- Εργαστήριο.

2.1.26 Μέθοδοι αξιολόγησης/βαθμολόγησης (Assessment method and criteria).

- Εξετάσεις
- Αξιολόγηση με βάση απόδοση σε project
- Επιτρεπόμενο υλικό κατά τις εξετάσεις: 2 φύλλα χειρόγραφα
- Παλαιότερα θέματα εξετάσεων/προόδων: διαθέσιμα

2.1.27 Προαπαιτούμενα (Expected prior knowledge/ prerequisites and preparation)

- Για προγραμματισμό: Python, Java,
- Βάσεις Δεδομένων,
- Τεχνητή Νοημοσύνη,
- Αλγόριθμοι.

2.1.28 Επιπλέον συνιστώμενη βιβλιογραφία και υλικό προς μελέτη (Literature and study materials / reading list)

Στην ελληνική γλώσσα. Υποχρεωτικό.

Προτείνεται λεκτική περιγραφή που να περιλαμβάνει τα εξής:

- Βιβλία- κείμενα (Text/books)
 - Doing Data Science, Straight Talk from the Frontline, Cathy O'Neil, Rachel Schutt
 - Pattern Recognition and Machine Learning (Information Science and Statistics) Hardcover – October 1, 2007, Christopher M. Bishop
 - Hadoop: The Definitive Guide, 3rd Edition, T. WhiteSlides Βιβλιογραφία
- Online readings
 - Scikit-learn: Machine Learning in Python, Fabian Pedregosa, Gaël

Varoquaux, Alexandre Gramfort et al., Scikit-learn: Machine Learning in Python 2011 Article, The Journal of Machine Learning Research archivVolume 12, 2/1/2011, Pages 2825-2830

- <https://www.kaggle.com/wiki/Tutorials>
- Πηγές στη βιβλιοθήκη του ιδρύματος. Εξόρυξη Γνώσης απο Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό, ΧΑΛΚΙΔΗ ΜΑΡΙΑ, ΒΑΖΙΡΓΙΑΝΝΗΣ ΜΙΧΑΛΗΣ, ISBN: 9604021168
- Άλλα σχετικά ανοικτά μαθήματα άλλων ιδρυμάτων εσωτερικού ή εξωτερικού:
 - Introduction to Data Science
<https://www.coursera.org/course/datasci>
 - Intro to Data Science, Learn What It Takes to Become a Data Scientist
<https://www.udacity.com/course/intro-to-data-science--ud359>

- Άρθρα (Articles)
- Βίντεο: <https://www.youtube.com/watch?v=dMpdoprDEDI>

2.2 Πληροφορίες για τις θεματικές ενότητες ή ενότητες διαλέξεων

2.2.1 Αριθμός Θεματικών Ενοτήτων

8

2.2.2 Τίτλοι Θεματικών Ενοτήτων

1. Εισαγωγή
2. Dimensionality Reduction and Feature Selection
3. Supervised learning
4. Unsupervised Learning (Clustering)
5. Community Detection and Evaluation in Social and Information Networks
6. Web Mining
7. Introduction to Big Data

2.2.3 Αναλυτική περιγραφή ενοτήτων

- 1.Εισαγωγή
- 2.Dimensionality Reduction and Feature Selection
 - Motivation
 - Distance Measures
 - Data Exploration and Preprocessing
 - Feature Ranking and Selection
- 3.Supervised learning
 - Introduction to supervised learning
 - Regression
 - Naïve Bayes
 - K-nn
 - Decision Trees
 - Regression re-visited
- 4.Unsupervised Learning (Clustering)

<ul style="list-style-type: none"> • Clustering • K-means • EM • Spectral Clustering
<p>5. Community Detection and Evaluation in Social and Information Networks</p> <ul style="list-style-type: none"> • Introduction & Motivation • Community evaluation measures • Graph clustering algorithms • Alternative Methods for Community Evaluation • New directions for research in the area of graph mining
<p>6. Web Mining</p> <ul style="list-style-type: none"> • Web personalization and recommendations (collaborative filtering) • Web Advertising
<p>7. Introduction to Big Data</p> <ul style="list-style-type: none"> • Why Big data • Hadoop • MapReduce • HDFS • Hive, HIVE QL

2.2.4 Λέξεις – κλειδιά ανά ενότητα

<p>1. Εισαγωγή</p>
<p>2. Dimensionality Reduction and Feature Selection</p> <p>Distance Measures, Exploratory Data Analysis, Data Quality, Evaluation metrics, Filter methods, Wrapper methods, Embedded methods, Complexity, Data features, k-NN, Linear Algorithms, Matrix Factorization methods, Principal Components Analysis (PCA), Singular Value Decomposition (SVD), Multidimensional Scaling (MDS), Non negative, Matrix Factorization (NMF), Latent Semantic Indexing (LSI)</p>
<p>3. Supervised learning</p> <p>Prediction, Regression, Least Squares method, Bayesian classification, Bayes theorem, Naïve Bayes Classifiers, Linear Discriminant Analysis, k-NN, Boundary, Decision Tree, Entropy, Information gain, Learning – linear regression, LMS</p>
<p>4. Unsupervised Learning (Clustering)</p> <p>Clustering, Partitional method, K-means, time complexity, speed-up tricks, Expectation Maximization, Gaussian Mixture Models, E step, M step, Hierarchical Clustering, Association Mining, Apriori Algorithm, Spectral Clustering, Graph-Cut, Min-Cut, Graph Laplacian, Ratio Cut, Normalized Cut, Algorithm for $k > 2$, Laplacian-Eigenvectors-EigenValues, Modularity function, Newman-Girvan algorithm</p>
<p>5. Community Detection and Evaluation in Social and Information Networks</p> <p>Networks, Social Networks, Community Detection, World Wide Web, Metabolic networks, intra-clusters, inter-clusters, Notation, Evaluation based on internal connectivity, Evaluation based on external connectivity, Evaluation based on network model, Observation, Graph-Cut, Min-Cut, Graph Laplacian, Ratio Cut, Normalized Cut, Modularity, Newman-Girvan algorithm, optimization, Resolution Limit, conductance, Network Community Profile (NCP), Core-periphery, Jellyfish model, Robustness, K-core,</p>

DBLP K-cores, Wikipedia, S-cores, CoreCluster framework, Scalability

6. Web Mining

Personalization, recommendations, Recommender Data Model, Scalability, Privacy and Trust, Robustness, Utilization, User Interests, Collaborative Filtering (CF), User-User Similarity, CF Recommender, Item-to-Item Collaborative filtering , Advertising, Traditional advertising, Online advertising, Cost Per Impression (CPI), Cost Per Click (CPC) , Cost Per Action (CPA), Google AdWords

7. Introduction to Big Data

Big Data, Hadoop, HDFS, MapReduce, Pig, Hive, HBase, MapReduce, Map function, Data Flow, Map task, HDFS, Hive, HiveQL, Querying Data