

# DAMA600 ΠΕΡΙΓΡΑΦΜΑ ΘΕΜΑΤΙΚΗΣ ΕΝΟΤΗΤΑΣ

## 1. ΓΕΝΙΚΑ

<b>ΣΧΟΛΗ</b>	ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ		
<b>ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ</b>	DATA SCIENCE AND MACHINE LEARNING		
<b>ΕΠΙΠΕΔΟ ΣΠΟΥΔΩΝ</b>	ΜΕΤΑΠΤΥΧΙΑΚΟ (επίπεδο 7)		
<b>ΚΩΔΙΚΟΣ ΘΕΜΑΤΙΚΗΣ ΕΝΟΤΗΤΑΣ</b>	DAMA600	<b>ΕΞΑΜΗΝΟ</b>	2/3
<b>ΤΙΤΛΟΣ ΘΕΜΑΤΙΚΗΣ ΕΝΟΤΗΤΑΣ</b>	Mining of Massive Datasets		
<b>ΑΥΤΟΤΕΛΕΙΣ ΕΚΠΑΙΔΕΥΤΙΚΕΣ ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ</b> <i>σε περίπτωση που οι πιστωτικές μονάδες απονέμονται σε διακριτά μέρη του μαθήματος π.χ. Διαλέξεις, Εργαστηριακές Ασκήσεις κ.λπ. Αν οι πιστωτικές μονάδες απονέμονται ενιαία για το σύνολο του μαθήματος αναγράψτε τις εβδομαδιαίες ώρες φόρτου και το σύνολο των πιστωτικών μονάδων</i>		<b>ΩΡΕΣ</b>	<b>ΠΙΣΤΩΤΙΚΕΣ ΜΟΝΑΔΕΣ</b>
Εβδομαδιαίες ώρες φόρτου: 32-33 ώρες x 13 εβδομάδες		420	15 ECTS
<b>ΤΥΠΟΣ ΘΕΜΑΤΙΚΗΣ ΕΝΟΤΗΤΑΣ</b> <i>Υποχρεωτική/Επιλογής/Κατ' επιλογήν υποχρεωτική</i>	Υποχρεωτική		
<b>ΠΡΟΑΠΑΙΤΟΥΜΕΝΕΣ ΘΕΜΑΤΙΚΕΣ ΕΝΟΤΗΤΕΣ:</b>	Για την παρακολούθηση της DAMA600 δεν απαιτείται η ταυτόχρονη παρακολούθηση ή ολοκλήρωση άλλης ΘΕ του DAMA.		
<b>ΓΛΩΣΣΑ ΔΙΔΑΣΚΑΛΙΑΣ και ΕΞΕΤΑΣΕΩΝ:</b>	ΑΓΓΛΙΚΗ		
<b>Η ΘΕΜΑΤΙΚΗ ΕΝΟΤΗΤΑ ΠΡΟΣΦΕΡΕΤΑΙ ΣΕ ΦΟΙΤΗΤΕΣ ERASMUS</b>	ΝΑΙ		
<b>ΗΛΕΚΤΡΟΝΙΚΗ ΣΕΛΙΔΑ ΘΕΜΑΤΙΚΗΣ ΕΝΟΤΗΤΑΣ (URL)</b>	Κάθε ΘΕ έχει τον δικό της χώρο στον ψηφιακό χώρο εκπαίδευσης του ΕΑΠ ( <a href="http://study.eap.gr">http://study.eap.gr</a> , <a href="http://courses.eap.gr">http://courses.eap.gr</a> ), με ελεγχόμενη πρόσβαση (χρήση κωδικού) για φοιτητές και διδακτικό προσωπικό.		

## 2. ΜΑΘΗΣΙΑΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

<p><b>Μαθησιακά Αποτελέσματα</b></p> <p>Περιγράφονται τα μαθησιακά αποτελέσματα της ΘΕ, οι συγκεκριμένες γνώσεις, δεξιότητες και ικανότητες καταλλήλου επιπέδου που θα αποκτήσουν οι φοιτητές μετά την επιτυχή ολοκλήρωση της ΘΕ.</p> <p>Συμβουλευτείτε το Παράρτημα Α</p> <ul style="list-style-type: none"> <li>• Περιγραφή του Επιπέδου των Μαθησιακών Αποτελεσμάτων για κάθε ένα κύκλο σπουδών σύμφωνα με Πλαίσιο Προσόντων του Ευρωπαϊκού Χώρου Ανώτατης Εκπαίδευσης</li> <li>• Περιγραφικοί Δείκτες Επιπέδων 6, 7 &amp; 8 του Ευρωπαϊκού Πλαισίου Προσόντων Διά Βίου Μάθησης και Παράρτημα Β</li> <li>• Περιληπτικός Οδηγός συγγραφής Μαθησιακών Αποτελεσμάτων</li> </ul>
<p><b>Γνώσεις:</b></p> <p>Με την επιτυχή ολοκλήρωση της θεματικής ενότητας, οι φοιτήτριες/φοιτητές θα μπορούν να:</p> <ul style="list-style-type: none"> <li>- Περιγράφουν τις προκλήσεις της εξόρυξης συνόλων δεδομένων μεγάλης κλίμακας και να συζητούν τις αντίστοιχες υπολογιστικές αρχιτεκτονικές.</li> <li>- Ορίζουν και εφαρμόζουν μέτρα ομοιότητας (similarity measures) και να χρησιμοποιούν τεχνικές shingling και minhashing για σύγκριση δεδομένων.</li> <li>- Σχεδιάζουν σχήματα locality-sensitive hashing (LSH) για αποδοτική αναζήτηση ομοιότητας.</li> <li>- Χρησιμοποιούν και αξιολογούν κλιμακώσιμους αλγόριθμους εξόρυξης συχνών συνόλων αντικειμένων (frequent itemset mining).</li> </ul> <p><b>Δεξιότητες:</b></p> <p>Με την επιτυχή ολοκλήρωση της θεματικής ενότητας, οι φοιτήτριες/φοιτητές θα μπορούν να:</p>

- Αναλύουν και μοντελοποιούν δεδομένα μεγάλων γράφων εφαρμόζοντας αλγορίθμους όπως PageRank και ανίχνευση κοινοτήτων (community detection).
- Υλοποιούν αλγορίθμους επεξεργασίας ροών δεδομένων (data stream processing) με μεθόδους δειγματοληψίας και sketching, π.χ. Count-Min Sketch.
- Εφαρμόζουν τεχνικές συσταδοποίησης (clustering) προσαρμοσμένες σε μεγάλα σύνολα δεδομένων, όπως K-means και ιεραρχική συσταδοποίηση (hierarchical clustering).
- Αναπτύσσουν κλιμακώσιμα συστήματα συστάσεων (recommender systems) με collaborative filtering και παραγοντοποίηση πινάκων (matrix factorization).

#### Ικανότητες:

Με την επιτυχή ολοκλήρωση της θεματικής ενότητας, οι φοιτήτριες/φοιτητές θα μπορούν να:

- Κατανοούν μεθόδους μείωσης διαστατικότητας, συμπεριλαμβανομένων των αναλύσεων SVD και CUR.
- Περιγράφουν και υλοποιούν κλιμακώσιμες μεθόδους ταξινόμησης για μεγάλα δεδομένα, π.χ. decision trees, naïve Bayes.
- Εφαρμόζουν αλγορίθμους μηχανικής μάθησης σε κατανεμημένα υπολογιστικά περιβάλλοντα, όπως το πλαίσιο MapReduce.
- Αξιολογούν την αποδοτικότητα, την κλιμακωσιμότητα και την καταλληλότητα τεχνικών εξόρυξης μαζικών δεδομένων σε πραγματικά σενάρια.

#### Γενικές Ικανότητες

Λαμβάνοντας υπόψη τις γενικές ικανότητες που πρέπει να έχει αποκτήσει ο πτυχιούχος (όπως αυτές αναγράφονται στο Παράρτημα Διπλώματος και παρατίθενται ακολούθως) σε ποια / ποιες από αυτές αποσκοπεί η ΘΕ;

Αναζήτηση, ανάλυση και σύνθεση δεδομένων και πληροφοριών, με τη χρήση και των απαραίτητων τεχνολογιών

Προσαρμογή σε νέες καταστάσεις

Λήψη αποφάσεων

Αυτόνομη εργασία

Ομαδική εργασία

Εργασία σε διεθνές περιβάλλον

Εργασία σε διεπιστημονικό περιβάλλον

Παράγωγή νέων ερευνητικών ιδεών

Σχεδιασμός και διαχείριση έργων

Σεβασμός στη διαφορετικότητα και στην πολυπολιτισμικότητα

Σεβασμός στο φυσικό περιβάλλον

Επίδειξη κοινωνικής, επαγγελματικής και ηθικής υπευθυνότητας

και ευαισθησίας σε θέματα φύλου

Άσκηση κριτικής και αυτοκριτικής

Προαγωγή της ελεύθερης, δημιουργικής και επαγωγικής σκέψης

Οι γενικές ικανότητες που οι φοιτητές θα αποκτήσουν είναι οι παρακάτω:

- Αναζήτηση, ανάλυση και σύνθεση δεδομένων και πληροφοριών με χρήση των απαραίτητων τεχνολογιών
- Προσαρμογή σε νέες καταστάσεις
- Λήψη αποφάσεων
- Αυτοδύναμη εργασία
- Ομαδική εργασία
- Σχεδιασμός και διαχείριση έργων
- Επίδειξη κοινωνικής, επαγγελματικής και ηθικής υπευθυνότητας και ευαισθησίας σε θέματα φύλου
- Παράγωγή ελεύθερης, δημιουργικής και επαγωγικής σκέψης

### 3. ΠΕΡΙΕΧΟΜΕΝΟ ΘΕΜΑΤΙΚΗΣ ΕΝΟΤΗΤΑΣ

#### Σκοπός της Ενότητας

Η ενότητα εφοδιάζει τους φοιτητές με εξειδικευμένη γνώση στην εξόρυξη και ανάλυση μαζικών συνόλων δεδομένων, με έμφαση σε κλιμακώσιμους αλγορίθμους και πλαίσια Big Data. Σε αντίθεση με τα παραδοσιακά μαθήματα Επιστήμης Δεδομένων, δίνεται προτεραιότητα σε τεχνικές σχεδιασμένες να χειρίζονται δεδομένα που υπερβαίνουν τη χωρητικότητα της κύριας μνήμης και απαιτούν επεξεργασία σε κατανεμημένα συστήματα. Οι φοιτητές θα εξερευνήσουν την αρχιτεκτονική και τις αρχές συστημάτων όπως τα MapReduce και Spark, τα οποία υποστηρίζουν επεξεργασία δεδομένων μεγάλης κλίμακας.

Θα μάθουν μεθόδους αποδοτικής αναζήτησης ομοιότητας, όπως minhashing και locality-sensitive hashing (LSH), προσαρμοσμένες σε δεδομένα υψηλής διάστασης. Το μάθημα καλύπτει αλγορίθμους εξόρυξης συχνών προτύπων και κανόνων συσχέτισης σε κλίμακα, υπερβαίνοντας τις κλασικές, in-memory προσεγγίσεις.

Στο πλαίσιο των ροών δεδομένων (streaming data), οι φοιτητές θα κατανοήσουν μοντέλα και τεχνικές για επεξεργασία σε πραγματικό χρόνο, όπως sketches και προσεγγιστική καταμέτρηση (approximate counting). Ιδιαίτερη έμφαση δίνεται στην εξόρυξη δομημένων δεδομένων γράφου, όπου θα μελετηθούν οι αλγόριθμοι PageRank, HITS, η ανίχνευση κοινοτήτων (community detection) και η καταμέτρηση τριγώνων (triangle counting)—ιδιαίτερα χρήσιμα στην ανάλυση ιστού και κοινωνικών δικτύων.

Το μάθημα εισάγει επίσης κλιμακώσιμα συστήματα συστάσεων (recommender systems) με collaborative filtering και παραγοντοποίηση πινάκων (matrix factorization). Τεχνικές μείωσης διαστατικότητας, όπως οι αποσυνθέσεις CUR και οι τυχαίες προβολές (random projections), συζητούνται με έμφαση στην κλιμακωσιμότητά τους και την καταλληλότητά τους για μεγάλα σύνολα δεδομένων. Το περιεχόμενο της Μηχανικής Μάθησης επικεντρώνεται στην αποδοτική υλοποίηση αλγορίθμων ταξινόμησης και συσταδοποίησης για μαζικά δεδομένα.

Οι φοιτητές θα εξετάσουν ακόμη τον σχεδιασμό αλγορίθμων υπό περιορισμούς πόρων, και πώς οι συμβιβασμοί ανάμεσα σε προσέγγιση, ταχύτητα και ακρίβεια διαχειρίζονται σε κλίμακα. Σε όλη τη διάρκεια του μαθήματος, τα θεωρητικά θεμέλια συνδυάζονται με πρακτικές εργασίες που αφορούν μεγάλα datasets και κατανεμημένα περιβάλλοντα. Σε αντίθεση με το DAMA510, το οποίο επικεντρώνεται σε στατιστικά μοντέλα και εισαγωγική Μηχανική Μάθηση, η παρούσα ενότητα δίνει προτεραιότητα στις μηχανικές και αλγοριθμικές προκλήσεις της εργασίας με πραγματικά τεράστια δεδομένα. Με την ολοκλήρωση της ΘΕ, οι φοιτητές θα είναι ικανοί να σχεδιάζουν, να υλοποιούν και να αξιολογούν κλιμακώσιμες ροές εξόρυξης δεδομένων (data mining pipelines) χρησιμοποιώντας σύγχρονα πλαίσια.

#### 4. ΔΙΔΑΚΤΙΚΕΣ και ΜΑΘΗΣΙΑΚΕΣ ΜΕΘΟΔΟΙ - ΑΞΙΟΛΟΓΗΣΗ

<p><b>ΤΡΟΠΟΣ ΠΑΡΑΔΟΣΗΣ</b> <i>Πρόσωπο με πρόσωπο, Εξ αποστάσεως εκπαίδευση κ.λπ.</i></p>	<ul style="list-style-type: none"> <li>- Τρόπος παράδοσης: Εξ αποστάσεως εκπαίδευση με τρεις (3) Ομαδικές Συμβουλευτικές Συναντήσεις (ΟΣΣ) διάρκειας 4 ωρών κατά τη διάρκεια του ακαδημαϊκού εξαμήνου, τα Σαββατοκύριακα.</li> <li>- Προσωπική επικοινωνία και ανατροφοδότηση (συμβουλευτικός ρόλος των συνεργαζόμενων διδασκόντων).</li> </ul>
<p><b>ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ</b> <i>Χρήση Τ.Π.Ε. στη Διδασκαλία, στην Εργαστηριακή Εκπαίδευση, στην Επικοινωνία με τους φοιτητές</i></p>	<p>Στις Ομαδικές Συμβουλευτικές Συναντήσεις (ΟΣΣ) ή/και κατά τη διδασκαλία χρησιμοποιούνται:</p> <ul style="list-style-type: none"> <li>- Εργαλεία απομακρυσμένων συναντήσεων (webex, Teams)</li> <li>- Λογισμικό παρουσίασης (PowerPoint, εκπαιδευτικά βίντεο – animations)</li> <li>- Εξειδικευμένο λογισμικό και βάσεις δεδομένων για τα υπό εκπαίδευση αντικείμενα</li> </ul> <p>Επιπλέον, οι φοιτητές χρησιμοποιούν εργαλεία αυτοματισμού γραφείου, φυλλομετρητές ιστού (web browsers) και e-reader για ψηφιακά βιβλία.</p> <p>Η επικοινωνία με τους φοιτητές υποστηρίζεται από:</p> <ul style="list-style-type: none"> <li>- Την ψηφιακή πλατφόρμα του ΕΑΠ (<a href="https://courses.eap.gr/login/index.php/">https://courses.eap.gr/login/index.php/</a>)</li> </ul>

	<p><a href="https://study.eap.gr/login/index.php">https://study.eap.gr/login/index.php</a>) για πληροφορίες ενότητας, αναρτήσεις εκπαιδευτικού υλικού, ανακοινώσεις, μηνύματα, αποτελέσματα εξετάσεων, ομάδες χρηστών, φόρουμ συζητήσεων κ.ά. - Email και μηνύματα (messaging).</p>														
<p><b>ΟΡΓΑΝΩΣΗ ΕΚΠΑΙΔΕΥΣΗΣ</b> Περιγράφονται αναλυτικά ο τρόπος και μέθοδοι εκπαίδευσης. Διαλέξεις, Σεμινάρια, Εργαστηριακή Άσκηση, Άσκηση Πεδίου, Μελέτη &amp; ανάλυση βιβλιογραφίας, Φροντιστήριο, Πρακτική (Τοποθέτηση), Κλινική Άσκηση, Καλλιτεχνικό Εργαστήριο, Διαδραστική διδασκαλία, Εκπαιδευτικές επισκέψεις, Εκπόνηση μελέτης (project), Συγγραφή εργασίας / εργασιών, Καλλιτεχνική δημιουργία, κ.λπ.</p> <p>Αναγράφονται οι ώρες μελέτης του φοιτητή για κάθε μαθησιακή δραστηριότητα καθώς και οι ώρες μη καθοδηγούμενης μελέτης ώστε ο συνολικός φόρτος εργασίας σε επίπεδο εξαμήνου να αντιστοιχεί στα standards του ECTS</p>	<table border="1"> <thead> <tr> <th>Δραστηριότητα</th> <th>Φόρτος Εργασίας Έτους</th> </tr> </thead> <tbody> <tr> <td>3 ΟΣΣ (x 4 ώρες)</td> <td>12</td> </tr> <tr> <td>2 Εκπαιδευτικές δραστηριότητες (x 10 ώρες)</td> <td>20</td> </tr> <tr> <td>2 εξαμηνιαίες εργασίες (x 30 ώρες)</td> <td>60</td> </tr> <tr> <td>Ατομική μελέτη (25 ώρες x 13 εβδομάδες)</td> <td>325</td> </tr> <tr> <td>Τελικές Εξετάσεις</td> <td>3</td> </tr> <tr> <td><b>Σύνολο φόρτου ΘΕ (ώρες)</b></td> <td><b>420</b></td> </tr> </tbody> </table>	Δραστηριότητα	Φόρτος Εργασίας Έτους	3 ΟΣΣ (x 4 ώρες)	12	2 Εκπαιδευτικές δραστηριότητες (x 10 ώρες)	20	2 εξαμηνιαίες εργασίες (x 30 ώρες)	60	Ατομική μελέτη (25 ώρες x 13 εβδομάδες)	325	Τελικές Εξετάσεις	3	<b>Σύνολο φόρτου ΘΕ (ώρες)</b>	<b>420</b>
Δραστηριότητα	Φόρτος Εργασίας Έτους														
3 ΟΣΣ (x 4 ώρες)	12														
2 Εκπαιδευτικές δραστηριότητες (x 10 ώρες)	20														
2 εξαμηνιαίες εργασίες (x 30 ώρες)	60														
Ατομική μελέτη (25 ώρες x 13 εβδομάδες)	325														
Τελικές Εξετάσεις	3														
<b>Σύνολο φόρτου ΘΕ (ώρες)</b>	<b>420</b>														
<p><b>ΑΞΙΟΛΟΓΗΣΗ ΦΟΙΤΗΤΩΝ</b> Περιγραφή της διαδικασίας αξιολόγησης</p> <p>Γλώσσα Αξιολόγησης, Μέθοδοι αξιολόγησης, Διαμορφωτική ή Συμπερασματική, Δοκιμασία Πολλαπλής Επιλογής, Ερωτήσεις Σύντομης Απάντησης, Ερωτήσεις Ανάπτυξης Δοκιμίων, Επίλυση Προβλημάτων, Γραπτή Εργασία, Έκθεση / Αναφορά, Προφορική Εξέταση, Δημόσια Παρουσίαση, Εργαστηριακή Εργασία, Κλινική Εξέταση Ασθενούς, Καλλιτεχνική Ερμηνεία, Άλλη / Άλλες</p> <p>Αναφέρονται ρητά προσδιορισμένα κριτήρια αξιολόγησης και εάν και που είναι προσβάσιμα από τους φοιτητές.</p>	<p><b>Αξιολόγηση Φοιτητών – Διαμόρφωση Τελικής Βαθμολογίας</b></p> <p>α. Δύο (2) Εκπαιδευτικές Δραστηριότητες τύπου quiz πολλαπλής επιλογής (Q), οι οποίες συνεισφέρουν εξίσου στον τελικό βαθμό με ποσοστό 5 % η καθεμία.</p> <p>β. Δύο (2) Εργασίες Εξαμήνου (A), οι οποίες συνεισφέρουν εξίσου στον τελικό βαθμό με ποσοστό 10 % η καθεμία.</p> <p>Η βαθμολόγηση των παραπάνω δραστηριοτήτων ενεργοποιείται μόνο εφόσον ο/η φοιτητής/φοιτήτρια επιτύχει συνολικό ποσοστό <math>\geq 50\%</math> στην τελική ή επαναληπτική εξέταση.</p> <p>γ. Τελική ή επαναληπτική εξέταση (E) με συντελεστή 70%.</p> <p>Η τελική βαθμολογία της ενότητας υπολογίζεται ως εξής (με μέγιστο βαθμό το 10):</p> $\text{Τελικός Βαθμός} = (Q_1 \times 5\%) + (Q_2 \times 5\%) + (A_1 \times 10\%) + (A_2 \times 10\%) + (E \times 70\%)$ <p>Γλώσσα αξιολόγησης: Αγγλικά</p>														

## 5. ΣΥΝΙΣΤΩΜΕΝΗ-ΒΙΒΛΙΟΓΡΑΦΙΑ

-Προτεινόμενη Βιβλιογραφία :

- J. Leskovec, A. Rajaraman & J.D. Ullman (2020). Mining of Massive Datasets (3rd edition). Cambridge University Press
- P.-N. Tan, M. Steinbach, A. Karpatne & V. Kumar (2021). Introduction to Data Mining (2nd Edition). Pearson.

- R. Zafarani, M.A. Abbasi & H. Liu (2014). Social media mining: an introduction. Cambridge University Press.

Επιπλέον ψηφιακό (και πολυμεσικό) υλικό θα διατεθεί online.