

## 2<sup>ο</sup> ΕΞΑΜΗΝΟ:

### ΠΕΡΙΓΡΑΜΜΑ ΜΑΘΗΜΑΤΟΣ «Μηχανική Μάθηση για την Ανάλυση Δεδομένων»

#### 1. ΓΕΝΙΚΑ

|  |   |                           |    |
|--|---|---------------------------|----|
| <b>ΣΧΟΛΗ</b>                                     | ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ          |                           |    |
| <b>ΤΜΗΜΑ</b>                                     | ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ          |                           |    |
| <b>ΕΠΙΠΕΔΟ ΣΠΟΥΔΩΝ</b>                           | Μεταπτυχιακό                              |                           |    |
| <b>ΚΩΔΙΚΟΣ ΜΑΘΗΜΑΤΟΣ</b>                         | 225                                       | <b>ΕΞΑΜΗΝΟ ΣΠΟΥΔΩΝ</b>    | 2ο |
| <b>ΤΙΤΛΟΣ ΜΑΘΗΜΑΤΟΣ</b>                          | Μηχανική Μάθηση για την Ανάλυση Δεδομένων |                           |    |
| <b>ΑΥΤΟΤΕΛΕΙΣ ΔΙΔΑΚΤΙΚΕΣ ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ</b>      | <b>ΕΒΔΟΜΑΔΙΑΙΕΣ ΩΡΕΣ ΔΙΔΑΣΚΑΛΙΑΣ</b>      | <b>ΠΙΣΤΩΤΙΚΕΣ ΜΟΝΑΔΕΣ</b> |    |
| Διαλέξεις  | 2   | 6                         |    |
| Εργαστηριακές Ασκήσεις                           | 1   |                           |    |
| <b>ΤΥΠΟΣ ΜΑΘΗΜΑΤΟΣ</b>                           | Ανάπτυξης δεξιοτήτων                      |                           |    |
| <b>ΠΡΟΑΠΑΙΤΟΥΜΕΝΑ ΜΑΘΗΜΑΤΑ:</b>                  |   |                           |    |
| <b>ΓΛΩΣΣΑ ΔΙΔΑΣΚΑΛΙΑΣ και ΕΞΕΤΑΣΕΩΝ:</b>         | Ελληνική                                  |                           |    |
| <b>ΤΟ ΜΑΘΗΜΑ ΠΡΟΣΦΕΡΕΤΑΙ ΣΕ ΦΟΙΤΗΤΕΣ ERASMUS</b> | ΝΑΙ (Στην Αγγλική)                        |                           |    |
| <b>ΗΛΕΚΤΡΟΝΙΚΗ ΣΕΛΙΔΑ ΜΑΘΗΜΑΤΟΣ (URL)</b>        | <a href="#">xxx</a>                       |                           |    |

#### 2. ΜΑΘΗΣΙΑΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

##### Μαθησιακά Αποτελέσματα

Μετά την επιτυχή ολοκλήρωση του μαθήματος, οι φοιτητές θα είναι σε θέση:

- Να κατανοούν τη στατιστική φύση ενός προβλήματος ανάλυσης δεδομένων τόσο ως προς το είδος των δεδομένων όσο και ως προς την απαιτούμενη στατιστική επεξεργασία
- Να γνωρίζουν τις έννοιες και τις μεθόδους της μηχανικής μάθησης και να κατανοούν τη στατιστική φύση τους και τη σχέση τους με την στατιστική επεξεργασία των δεδομένων
- Να επιλέγουν και να εφαρμόζουν την καταλληλότερη μέθοδο μηχανικής μάθησης για να προσεγγίσουν συστηματικά και να επιλύσουν προβλήματα της ανάλυσης δεδομένων
- Να κατανοούν και να χειρίζονται το περιβάλλον στατιστικής επεξεργασίας R
- Να αντιλαμβάνονται τα ανοιχτά προβλήματα στην περιοχή της μηχανικής μάθησης και της μοντελοποίησης των δεδομένων
- Να διεξάγουν αποτελεσματική βιβλιογραφική αναζήτηση συσχετίζοντας τα ανοιχτά προβλήματα με τη βιβλιογραφία.
- Να αναγνωρίζουν και να διαχωρίζουν τα ερευνητικά ζητήματα από τα προβλήματα εφαρμογής.

##### Γενικές Ικανότητες

- Αναζήτηση, ανάλυση και σύνθεση δεδομένων, τεχνικών και πληροφοριών, με τη χρήση και των απαραίτητων τεχνικών
- Ανάπτυξη και τεκμηρίωση επιχειρημάτων με αξιοποίηση δομημένης μαθηματικής σκέψης
- Συνδυαστική ανάλυση μεθόδων για επίλυση προβλημάτων

- Λήψη αποφάσεων
- Εργασία σε διεπιστημονικό περιβάλλον
- Παραγωγή νέων ερευνητικών ιδεών

### 3. ΠΕΡΙΕΧΟΜΕΝΟ ΜΑΘΗΜΑΤΟΣ

Τύποι δεδομένων και μορφές Ανάλυσης Δεδομένων. Εισαγωγή στην Ανάλυση Δεδομένων με την R. Περιγραφική Ανάλυση. Υπενθύμιση εννοιών της Στατιστικής: Κατανομές Πιθανοτήτων και Στατιστική Συμπερασματολογία (Διαστήματα Εμπιστοσύνης, Παραμετρικοί και Μη Παραμετρικοί Έλεγχοι Υποθέσεων). Εκτίμηση πυκνότητας πιθανότητας, μοντέλα μείξης, εκτίμηση παραμέτρων, επιλογή μοντέλου. Θεωρία αποφάσεων του Bayes. Προετοιμασία δεδομένων (κανονικοποίηση και εξαγωγή χαρακτηριστικών). Μείωση της διάστασης των δεδομένων. Ανάλυση Κυρίων Συνιστωσών (PCA). Ταξινόμηση δεδομένων και ταξινομητές (Μπεϋζιανοί, Τεχνητά Νευρωνικά δίκτυα, Μηχανές διανυσμάτων υποστήριξης (SVM), ταξινομητές ελάχιστης απόστασης, κ.λπ.) Παλινδρόμηση (γραμμική, πολλαπλή, logistic, Μπεϋζιανή). Παλινδρόμηση με Τεχνητά Νευρωνικά Δίκτυα. Προσαρμογή καμπυλών, πρόβλεψη και αναλυση χρονοσειρών. Συσταδοποίηση-ομαδοποίηση δεδομένων. αλγόριθμοι k-means, k-median, DBSCAN, κ.λπ. Δεδομένα υπό αβεβαιότητα, Ασαφή δεδομένα και διαστηματικά δεδομένα. Στατιστική ανάλυση διαστηματικών δεδομένων. Τεχνικές συσταδοποίησης και παλινδρόμησης διαστηματικών δεδομένων.

| Εβδομ. | Τίτλος Ενότητας  | Βιβλιογραφία | e-class |
|--------|--|--------------|---------|
| 1      | Εισαγωγικές έννοιες: <ul style="list-style-type: none"> <li>• Τύποι δεδομένων (στατικά και μεταβαλλόμενα, ντετερμινισμός και αβεβαιότητα)</li> <li>• Φάσεις και μορφές της ανάλυσης δεδομένων</li> <li>• Παραδείγματα προβλημάτων ανάλυσης δεδομένων</li> <li>• Περιγραφική ανάλυση</li> <li>• Το περιβάλλον επεξεργασίας R</li> </ul> |              |         |
| 2      | <ul style="list-style-type: none"> <li>• Υπενθύμιση βασικών εννοιών από τη θεωρία πιθανοτήτων και τη στατιστική.</li> <li>• Στατιστική Συμπερασματολογία (Διαστήματα Εμπιστοσύνης, Παραμετρικοί και Μη Παραμετρικοί Έλεγχοι Υποθέσεων)</li> </ul>  |              |         |
| 3      | Βασικές στατιστικές τεχνικές και μηχανική μάθηση: <ul style="list-style-type: none"> <li>• Εκτίμηση πυκνότητας πιθανότητας, μοντέλα μείξης, εκτίμηση παραμέτρων, επιλογή μοντέλου</li> </ul>   |              |         |
| 4      | <ul style="list-style-type: none"> <li>• Μπεϋζιανή πιθανότητα και στατιστική</li> <li>• Θεωρία αποφάσεων του Bayes</li> <li>• Μπεϋζιανή εκτίμηση πυκνότητας πιθανότητας</li> </ul>   |              |         |
| 5      | <ul style="list-style-type: none"> <li>• Προετοιμασία δεδομένων (κανονικοποίηση και εξαγωγή χαρακτηριστι-</li> </ul>   |              |         |

|    |  |  |  |
|----|--|--|--|
|    | <p>κών)</p> <ul style="list-style-type: none"> <li>• Μείωση της διάστασης των δεδομένων</li> <li>• Ανάλυση Κυρίων Συνιστωσών (PCA)</li> <li>• Πρακτική εξάσκηση στο περιβάλλον R</li> </ul>  |  |  |
| 6  | <p>Ταξινόμηση δεδομένων και ταξινομητές</p> <ul style="list-style-type: none"> <li>• Μπεϋζιανή ταξινόμηση</li> <li>• Τεχνητά Νευρωνικά Δίκτυα</li> <li>• Πρακτική εξάσκηση στο περιβάλλον R</li> </ul>   |  |  |
| 7  | <p>Ταξινόμηση δεδομένων και ταξινομητές</p> <ul style="list-style-type: none"> <li>• Μηχανές διανυσμάτων υποστήριξης (SVM)</li> <li>• Ταξινομητές ελάχιστης απόστασης</li> <li>• Πρακτική εξάσκηση στο περιβάλλον R</li> </ul>                                 |  |  |
| 8  | <ul style="list-style-type: none"> <li>• Παλινδρόμηση (γραμμική, πολλαπλή, logistic)</li> <li>• Μπεϋζιανή παλινδρόμηση</li> <li>• Παλινδρόμηση με Τεχνητά Νευρωνικά Δίκτυα</li> <li>• Πρακτική εξάσκηση στο περιβάλλον R</li> </ul>                            |  |  |
| 9  | <ul style="list-style-type: none"> <li>• Προσαρμογή καμπυλών και πρόβλεψη</li> <li>• Ανάλυση χρονοσειρών</li> <li>• Πρακτική εξάσκηση στο περιβάλλον R</li> </ul>  |  |  |
| 10 | <p>Συσταδοποίηση-Ομαδοποίηση δεδομένων</p> <ul style="list-style-type: none"> <li>• Ομαδοποίηση δεδομένων και μέτρα απόστασης</li> <li>• Οι τεχνικές k-means, k-median</li> <li>• Ο αλγόριθμος DBSCAN</li> <li>• Πρακτική εξάσκηση στο περιβάλλον R</li> </ul> |  |  |
| 11 | <ul style="list-style-type: none"> <li>• Η τυχαιότητα και η αβεβαιότητα στα δεδομένα</li> <li>• Ασαφή δεδομένα και η στατιστική ανάλυσή τους</li> <li>• Διαστηματικά δεδομένα και ανάλυση διαστημάτων</li> <li>• Παραδείγματα και εφαρμογές</li> </ul>         |  |  |
| 12 | <ul style="list-style-type: none"> <li>• Στατιστικές και στατιστική ανάλυση διαστηματικών δεδομένων</li> <li>• Συσταδοποίηση και παλινδρόμηση διαστηματικών δεδομένων</li> <li>• Παραδείγματα και εφαρμογές σε περιβάλλον MATLAB/Octave</li> </ul>             |  |  |
| 13 | <p>Παρουσίαση ατομικών εργασιών ερευνητικού ενδιαφέροντος ή μελέτες εφαρμογής με σύντομη ανάλυση και συζήτηση</p>  |  |  |

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

|   |  |                                 |
|---|--|---------------------------------|
| <b>ΤΡΟΠΟΣ ΠΑΡΑΔΟΣΗΣ.</b>                              | Πρόσωπο με πρόσωπο   |                                 |
| <b>ΧΡΗΣΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ</b> | <input checked="" type="checkbox"/> Χρήση ηλεκτρονικών παρουσιάσεων, αναρτημένων στο e-class<br><input checked="" type="checkbox"/> Χρήση λογισμικού στον υπολογιστή κατά τη διάλεξη<br><input checked="" type="checkbox"/> Χρήση εξειδικευμένου λογισμικού<br><input checked="" type="checkbox"/> Διάθεση εκπαιδευτικού υλικού μέσω e-class<br><input checked="" type="checkbox"/> Διαχείριση εργασιών/ασκήσεων μέσω δικτυακού τόπου.<br><input checked="" type="checkbox"/> Επικοινωνία με φοιτητές μέσω e-mail.<br><input type="checkbox"/> Ηλεκτρονικός χώρος συνομιλιών διδάσκοντος και φοιτητών.                         |                                 |
| <b>ΟΡΓΑΝΩΣΗ ΔΙΔΑΣΚΑΛΙΑΣ</b>                           | <b>Δραστηριότητα</b>   | <b>Φόρτος Εργασίας Εξαμήνου</b> |
|   | Διαλέξεις  | 28 ώρες                         |
|   | Εργαστηριακές Ασκήσεις   | 11 ώρες                         |
|   | Βιβλιογραφική Εργασία  | 20 ώρες                         |
|   | Υλοποίηση Εργασίας   | 50 ώρες                         |
|   | Μη καθοδηγούμενη μελέτη  | 71 ώρες                         |
|   | <b>Σύνολο Μαθήματος</b>  | <b>180 ώρες</b>                 |
| <b>ΑΞΙΟΛΟΓΗΣΗ ΦΟΙΤΗΤΩΝ</b>                            | <ul style="list-style-type: none"> <li>• Οι φοιτητές αναλαμβάνουν να παρουσιάσουν ατομικά μια εργασία (paper) από κάποιο έγκριτο περιοδικό ή συνέδριο που είναι συναφές με το αντικείμενο του μαθήματος (50%).</li> <li>• Γραπτή εξέταση (50%)<br/>ή</li> <li>• Υλοποίηση ατομικής εργασία σε συγκεκριμένο πρόβλημα που δύνανται να επιλέξουν οι φοιτητές σε συνεργασία με τον διδάσκοντα (50%). Η εργασία θα πρέπει να παραδοθεί το αργότερο έως την ημερομηνία που θα οριστεί για τη γραπτή εξέταση.</li> </ul> <p>Τα κριτήρια αξιολόγησης γνωστοποιούνται στους φοιτητές στην πρώτη διάλεξη και αναρτώνται στο e-class.</p> |                                 |

#### 5. ΣΥΝΙΣΤΩΜΕΝΗ-ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] Φουσκάκης, Δ. *Ανάλυση Δεδομένων με Χρήση της R*. Εκδόσεις Τσότρας. Αθήνα, 2013.
- [2] Heumann, Ch. Shalabh, M.S. *Introduction to Statistics and Data Analysis*, Springer, 2016. ISBN 978-3-319-46162-5 (eBook), DOI 10.1007/978-3-319-46162-5
- [3] D. C. Montgomery, G. C. Runger *Applied Statistics and Probability for Engineers*, 6<sup>th</sup> Edition, John Wiley & Sons, 2013
- [4] S. Theodoridis, K. Koutroumbas, *Αναγνώριση Προτύπων*, Εκδόσεις Πασχαλίδης, Αθήνα, 2012.
- [5] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference*,

*and Prediction*, Springer Texts in Statistics, Springer-Verlag, New York, 2001.

- [6] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, Information Science and Statistics Series, 2006.
- [7] R. Neal, *Bayesian Learning for Neural Networks*, Springer, New York, 1996.
- [8] S. Theodoridis, A. Pikrakis, K. Koutroumbas, D. Cavouras, *Εισαγωγή στην Αναγνώριση Προτύπων με MATLAB*, Εκδόσεις Πασχαλίδης, Αθήνα, 2011.