



**ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**ΕΡΓΑΣΙΑ ΣΤΟ ΜΑΘΗΜΑ «ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΣΤΗ  
ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ»**

**ΘΕΜΑ: ΑΝΑΓΝΩΡΙΣΗ ΨΕΥΔΩΝ ΕΙΔΗΣΕΩΝ ΣΕ ΚΕΙΜΕΝΑ ΤΟΥ  
ΚΟΙΝΩΝΙΚΟΥ ΙΣΤΟΥ**

**Στόχος** της παρούσας εργασίας είναι η αναγνώριση ψευδών ειδήσεων (fake news detection) με χρήση τεχνικών μηχανικής μάθησης.

**Βήμα 1 (Προθεσμία 30/10/2019)**

Μελετήστε το άρθρο «A Survey on Natural Language Processing for Fake News Detection» των Oshikawa, Qian και Wang και ετοιμάστε σε ppt διαφάνειες για να παρουσιάσετε το περιεχόμενο του άρθρου, οργανωμένο στις εξής ενότητες:

- Το πρόβλημα της αναγνώρισης ψευδών ειδήσεων (1)
- Τα datasets μιας ή λίγων προτάσεων (2.1.1)
- Τα datasets αναρτήσεων σε κοινωνικά δίκτυα (2.1.2)
- Τα datasets ολόκληρων άρθρων (2.2)
- Τις μορφές εξόδου των εργασιών μάθησης (3.2)
- Τις τεχνικές προεπεξεργασίας των κειμένων (4.1)
- Τις τεχνικές μηχανικής μάθησης (4.4)
- Τα αποτελέσματα (5)
- Τα συμπεράσματα για την είσοδο (6.1)
- Τα συμπεράσματα για τα μοντέλα (6.2)
- Τις σχετικές προκλήσεις (7)

**Βήμα 2 (Προθεσμία 13/11/2019)**

- Μελετήστε την βιβλιογραφία την σχετική με την εφαρμογή. Ψάξτε και για άλλο σχετικό υλικό.
- Κατεβάστε και μελετήστε τα δεδομένα από το λινκ

[https://blog.twitter.com/en\\_us/topics/company/2019/information\\_operations\\_directed\\_at\\_Hong\\_Kong.html](https://blog.twitter.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.html)

Αυτά τα δεδομένα είναι fake news.

- Καθαρίστε τα δεδομένα, ακολουθώντας τεχνικές και χρησιμοποιώντας εργαλεία που προτείνονται στην βιβλιογραφία.

- Θα πρέπει να βρείτε αντίστοιχα δεδομένα, σχετικής θεματολογίας, που δεν είναι fake news και να ξεκινήσετε να τα κατεβάζετε.
- Ανεβάστε η κάθε ομάδα (ένας από κάθε ομάδα) στο eclass μια μικρή παρουσίαση με την σχετική βιβλιογραφία που έχετε βρει και την μεθοδολογία που ακολουθείτε για την συλλογή περαιτέρω δεδομένων.

### **Βήμα 3 (Προθεσμία 18/12/2019)**

Πρέπει να

- έχετε καταλήξει στα χαρακτηριστικά μάθησης που θα χρησιμοποιήσετε
- έχετε συλλέξει, καθαρίσει, προεπεξεργαστεί όλα τα δεδομένα σας, και να τα έχετε μετασχηματίσει σε διανύσματα μάθησης.
- ανεβάσετε η κάθε ομάδα (ένας από κάθε ομάδα) στο eclass μια μικρή παρουσίαση με την δουλειά που έχετε κάνει έως τώρα

### **Βήμα 4 (Προθεσμία: Ημερομηνία Εξετάσεων)**

Πρέπει να

- έχετε ολοκληρώσει τους πειραματισμούς με
  - ο διάφορους αλγορίθμους μηχανικής μάθησης δικής σας επιλογής, βάσει της βιβλιογραφικής σας επισκόπησης
  - ο διάφορα σετ χαρακτηριστικών
- έχετε κάνει σύγκριση των πειραματικών αποτελεσμάτων
- έχετε κάνει ποιοτική ανάλυση των αποτελεσμάτων σας
- έχετε συντάξει επιστημονική εργασία στα Αγγλικά (σαν τις εργασίες της βιβλιογραφίας) μέχρι 10 σελίδες το πολύ, στην οποία θα παρουσιάζετε
  - ο το θεωρητικό υπόβαθρο της δουλειάς σας
  - ο το σετ δεδομένων σας
  - ο την επιλογή των χαρακτηριστικών σας
  - ο την πειραματική σας διαδικασία και τα αποτελέσματά σας (precision και recall για κάθε τιμή της κλάσης ταξινόμησης)
  - ο την αξιολόγηση των αποτελεσμάτων σας
  - ο την σύγκριση των αποτελεσμάτων σας με αυτά άλλων προσεγγίσεων στη βιβλιογραφία
  - ο τα συμπεράσματα στα οποία καταλήξατε και μελλοντικές βελτιώσεις που προτείνετε για την προσέγγισή σας
  - ο την διαφορετικότητα στην προσέγγισή σας (τι διαφορές/καινοτομίες/συνεισφορά έχει η δική σας δουλειά σε σχέση με τις παραπλήσιες της βιβλιογραφίας)
  - ο την λίστα των βιβλιογραφικών σας παραπομπών
- έχετε ανεβάσει το άρθρο της εργασίας σας στο eclass

Η εργασία θα πραγματοποιηθεί σε ομάδες των 3-4 ατόμων.

Η εργασία είναι υποχρεωτική και πιάνει το 60% στον τελικό βαθμό.

Η ομάδα που θα πετύχει το μεγαλύτερο σκορ στα αποτελέσματα θα έχει

- επιπλέον βαθμολογικό bonus
- ευκαιρία να προωθήσει την εργασία της για δημοσίευση σε επιστημονικό συνέδριο