



**ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΜΠΣ «ΨΗΦΙΑΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΣΤΗΝ
ΔΙΟΙΚΗΣΗ ΤΗΣ ΦΙΛΟΞΕΝΙΑΣ ΚΑΙ ΤΟΝ ΤΟΥΡΙΣΜΟ»**

ΕΠΙΧΕΙΡΗΣΙΑΚΗ ΝΟΗΜΟΣΥΝΗ ΣΤΟΝ ΤΟΥΡΙΣΜΟ

ΕΡΓΑΣΤΗΡΙΟ 1: ΠΡΩΤΗ ΕΞΟΙΚΕΙΩΣΗ ΜΕ ΤΟ WEKA

[\(\[HTTP://WWW.CS.WAIKATO.AC.NZ/ML/WEKA\]\(http://www.cs.waikato.ac.nz/ml/weka\)\)](http://www.cs.waikato.ac.nz/ml/weka)

Στόχος αυτού του εργαστηρίου είναι

- η εξοικείωση με το περιβάλλον του Weka
- η εξοικείωση με τεχνικές προεπισκόπησης των δεδομένων
- η εξοικείωση με εργαλεία οπτικοποίησης

Αρχείο Δεδομένων: Tanzania.csv

Το σετ δεδομένων περιλαμβάνει παραδείγματα τουριστών στην Τανζανία. Κάθε τουρίστας αναπαρίσταται μέσω των παρακάτω χαρακτηριστικών:

ΟΝΟΜΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ	ΕΠΕΞΗΓΗΣΗ
id	Unique identifier for each tourist
country	The country a tourist is coming from.
age_group	The age group of a tourist.
travel_with	The relation to the people a tourist is travelling with to Tanzania
total_female	Total number of females
total_male	Total number of males
purpose	The purpose of visiting Tanzania
main_activity	The main activity of tourism in Tanzania
infor_source	The source of information about tourism in Tanzania
tour_arrangment	The arrangement of visiting Tanzania
package_transport_int	If the tour package includes international transportation service
package_accomodation	If the tour package includes accommodation service
package_food	If the tour package includes food service
package_transport_tz	If the tour package includes transport service within Tanzania
package_sightseeing	If the tour package includes sightseeing service
package_guided_tour	If the tour package includes tour guide
package_insurance	if the tour package includes insurance service
night_mainland	Number of nights a tourist spent in Tanzania mainland
night_zanzibar	Number of nights a tourist spent in Zanzibar
first_trip_tz	If it was a first trip to Tanzania
cost_category	The range of expenditures a tourist spends in Tanzania

1. Αποθηκεύστε από τον φάκελο «Εργαστήριο 1» του μαθήματος στο OpenCourses σε έναν φάκελο τοπικά το αρχείο **Tanzania.csv**.
2. Ανοίξτε το αρχείο (δεξί κλικ - Open With) με Wordpad, για να δείτε τι περιέχει.
3. Από το ΕΝΑΡΞΗ επιλέξτε Weka 3.6.6 -> Weka 3.6. Από το παράθυρο του Weka επιλέξτε *Explorer*.
4. Από το μενού επιλέξτε *Preprocess* -> *Open File*. Ανοίξτε το αρχείο **Tanzania.csv**.
5. Στο αριστερό μέρος της οθόνης σας εμφανίζονται πληροφορίες για το σετ δεδομένων σας. Πόσα είναι τα παραδείγματά σας; _____ Πόσα είναι τα χαρακτηριστικά εισόδου σας; _____
6. Τα χαρακτηριστικά σας εμφανίζονται πιο κάτω στα αριστερά της οθόνης. Κάνοντας κλικ πάνω σε ένα χαρακτηριστικό εμφανίζονται στατιστικές πληροφορίες για αυτό στο δεξί μέρος της οθόνης. Τι τιμές παίρνει το καθένα; _____
7. Ποια είναι η έξοδος σας, και τι τιμές παίρνει; _____

Visualize

8. Στην καρτέλα 'Visualize' δείτε διαφορετικά διαγράμματα των δεδομένων σας. Σε κάθε γράφημα μπορούν να συμμετέχουν τρία χαρακτηριστικά: δύο στους δυο άξονες και ένα με χρωματικό κώδικα. Οι ταινίες στα δεξιά είναι μονοδιάστατα γραφήματα των χαρακτηριστικών. Το Jitter είναι τυχαίος θόρυβος που μπορεί να προστεθεί στα δεδομένα για να αποφευχθεί η επικάλυψη στην απεικόνιση.
9. Μπορείτε να επιλέξετε χαρακτηριστικά για τα γραφήματα είτε μέσω των pull-down menus για τα X:, Y: και Colour, είτε με κλικ στα μονοδιάστατα γραφήματα κάθε χαρακτηριστικού (αριστερό κλικ για το X, δεξί για το Y).
10. Προσπαθήστε να βρείτε ένα διάγραμμα στο οποίο οι τιμές της κλάσης ταξινόμησης να διακρίνονται (διαχωρίζονται) καλά. Κρίνοντας από τα διαγράμματα, πρόκειται για δύσκολο πρόβλημα ταξινόμησης;
11. Κάντε κλικ πάνω σε ένα σημείο του γραφήματος. Παρατηρήστε το παράθυρο που ανοίγει με τις πληροφορίες του συγκεκριμένου παραδείγματος.

Διερεύνηση των χαρακτηριστικών

12. Στην καρτέλα Pre-process επιλέξτε με αριστερό κλικ ένα χαρακτηριστικό. Δείτε στο κάτω δεξί μέρος πώς κατανέμονται οι τιμές αυτού του χαρακτηριστικού βάσει της τιμής της κλάσης ταξινόμησης. Τα παραδείγματά σας διακρίνονται καλά ως προς την τιμή της κλάσης ταξινόμησής τους, βάσει των τιμών του χαρακτηριστικού αυτού;
13. Κάντε το βήμα 12 για όλα τα χαρακτηριστικά εισόδου. Ποιό/ά φαίνεται να διακρίνουν καλύτερα τα παραδείγματά σας ως προς την τιμή της κλάσης ταξινόμησής τους;