**ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ *(Υπεύθυνος: Καθ. Σ. Λυκοθανάσης)***

**Ακαδημαϊκό Έτος 2023-24**

**Εργαστηριακή Άσκηση 2: Δένδρα Απόφασης**

**Ημερομηνία Υποβολής (μέσω e-class) μέχρι τις 03-12-2023, στις 23.55.**

**Το όνομα του αρχείου θα έχει τη μορφή: ΧΧ\_ΥΥ\_ΖΖ.pdf, όπου ΧΧ= Επώνυμο, ΥΥ=Αρχικό ονόματος και ΖΖ=ΑΜ. Να δώσετε τις απαντήσεις σας, μετά από κάθε ερώτημα.**

**Θέμα: Δέντρα Απόφασης** [*100 μονάδες*]

Το παρακάτω σύνολο δεδομένων εκπαίδευσης περιγράφει την ιστιοπλοϊκή δραστηριότητα (*Sail*) ενός skipper, με βάση τις καιρικές συνθήκες (*Outlook*), τη δυναμικότητα του επιπλέον πληρώματος (*Company*) και το μέγεθος του σκάφους (*Sailboat*). Θέλουμε να φτιάξουμε ένα δέντρο απόφασης για να προβλέπουμε αν ο skipper θα σαλπάρει με βάση τις τιμές των άλλων ιδιοτήτων του προβλήματος.

(Σημείωση: Χρησιμοποιείστε ακρίβεια 4 δεκαδικών ψηφίων και, αν θέλετε, αξιοποιήστε κάποιον υπολογιστή εντροπίας, όπως είναι ο <https://www.dcode.fr/shannon-index>.)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sail | Outlook | Company | Sailboat |
| 1 | yes | Rainy | big | big |
| 2 | yes | Rainy | big | small |
| 3 | no | rainy | med | big |
| 4 | no | rainy | med | small |
| 5 | yes | sunny | big | big |
| 6 | yes | sunny | big | small |
| 7 | yes | sunny | med | big |
| 8 | yes | sunny | med | big |
| 9 | yes | sunny | med | small |
| 10 | yes | sunny | no | small |
| 11 | no | sunny | no | big |
| 12 | no | rainy | med | big |
| 13 | no | rainy | no | big |
| 14 | no | rainy | no | big |
| 15 | no | rainy | no | small |
| 16 | no | rainy | no | small |
| 17 | yes | sunny | big | big |
| 18 | no | sunny | big | small |
| 19 | no | sunny | med | big |
| 20 | no | sunny | med | big |

**Α.** [*20 μονάδες*] Υποθέτοντας πως έχετε κάνει διαχωρισμό με βάση την ιδιότητα Outlook, υπολογίστε την εντροπία E (Sail | Outlook) για την κλάση του προβλήματος.

**Β.** [*20 μονάδες*] Υπολογίστε το κέρδος πληροφορίας (*IG*) αν ο διαχωρισμός στην κορυφή γίνει με βάση την ιδιότητα Outlook.

**Γ.** [*20 μονάδες*] Χρησιμοποιώντας το κέρδος πληροφορίας ως μετρική, ποια ιδιότητα θα χρησιμοποιηθεί για διαχωρισμό στη ρίζα του δέντρου απόφασης; Δείξτε τους υπολογισμούς.

**Δ.** [*20 μονάδες*] Το παραπάνω σύνολο δεδομένων θα αλλάξει, αν διαγράψουμε ορισμένα αντίγραφα από κάθε μοναδική εγγραφή. Για το αλλαγμένο σύνολο δεδομένων, που φαίνεται παρακάτω, ποια ιδιότητα θα επιλεγόταν στην κορυφή του δέντρου απόφασης;

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Sail | Outlook | Company | Sailboat | | 1 | yes | rainy | big | big | | 2 | yes | rainy | big | small | | 3 | no | rainy | med | big | | 4 | no | rainy | med | small | | 5 | yes | sunny | big | big | | 6 | yes | sunny | big | small | | 7 | yes | sunny | med | big | | 8 | yes | sunny | med | big | | 9 | yes | sunny | med | small | | 10 | yes | sunny | no | small | | 11 | no | sunny | no | big | | 12 | no | rainy | med | big | | 13 | no | rainy | no | big | | 14 | no | rainy | no | big | | 15 | no | rainy | no | small | | 16 | no | rainy | no | small | | 17 | yes | sunny | big | big | | 18 | no | sunny | big | small | | 19 | no | sunny | med | big | | 20 | no | sunny | med | big | |

**Ε.** [*20 μονάδες*] Να κατασκευάσετε 2 αρχεία που να αντιστοιχούν στα παραπάνω σύνολα δεδομένων εκπαίδευσης και να τα τροφοδοτήσετε ως είσοδο στο πρόγραμμα κατασκευής δέντρων απόφασης που βρίσκεται στο AISpace (<http://aispace.org/dTree/>), χρησιμοποιώντας (στο AIspace) το κέρδος πληροφορίας ως μέθοδο επιλογής της διαχωριστικής ιδιότητας σε κάθε κόμβο. Τι παρατηρείτε;

(Υπόδειξη: Να χρησιμοποιήσετε τις default παραμέτρους του AIspace.)