

# Συστήματα Ανακάλυψης Γνώσης από Βάσεις Δεδομένων

## Εργασία 05 – Text Categorization

Ονοματεπώνυμο: Παναγιώτης Γιαννουτάκης

AM: 12/38

Email: [it1238@uom.edu.gr](mailto:it1238@uom.edu.gr)

1.

a.

$$P(\text{ΝΑΙΤηλεόραση}) = 3/5$$

$$P(\text{ΌΧΙΤηλεόραση}) = 2/5$$

$$P(\text{επεισόδιο} \mid \text{ΝΑΙΤηλεόραση}) = (5+1)/(14+8) = 6/22$$

$$P(\text{επεισόδιο} \mid \text{ΌΧΙΤηλεόραση}) = (1+1)/(6+8) = 2/14$$

$$P(\text{γήπεδο} \mid \text{ΝΑΙΤηλεόραση}) = (0+1)/(14+8) = 1/22$$

$$P(\text{γήπεδο} \mid \text{ΌΧΙΤηλεόραση}) = (2+1)/(6+8) = 3/14$$

$$P(\text{ειδήσεις} \mid \text{ΝΑΙΤηλεόραση}) = (1+1)/(14+8) = 2/22$$

$$P(\text{ειδήσεις} \mid \text{ΌΧΙΤηλεόραση}) = (1+1)/(6+8) = 2/14$$

$$\begin{aligned} P(\text{ΝΑΙΤηλεόραση} \mid d6) &= P(\text{ΝΑΙΤηλεόραση}) * P(\text{επεισόδιο} \mid \text{ΝΑΙΤηλεόραση})^2 * \\ &P(\text{γήπεδο} \mid \text{ΝΑΙΤηλεόραση})^3 * P(\text{ειδήσεις} \mid \text{ΝΑΙΤηλεόραση}) = (3/5) * (6/22)^2 * \\ &(1/22)^3 * (2/22) = 0.0000004 \end{aligned}$$

$$P(\text{ΌΧΙΤηλεόραση} \mid d6) = P(\text{ΌΧΙΤηλεόραση}) * P(\text{επεισόδιο} \mid \text{ΌΧΙΤηλεόραση})^2 * \\ P(\text{γήπεδο} \mid \text{ΌΧΙΤηλεόραση})^3 * P(\text{ειδήσεις} \mid \text{ΌΧΙΤηλεόραση}) = (2/5) * (2/14)^2 * \\ (3/14)^3 * (2/14) = 0.00001$$

Οπότε το document 6 ανήκει στην κατηγορία ΌΧΙ Τηλεόραση γιατί  
 $P(\text{ΌΧΙΤηλεόραση} \mid d6) > P(\text{ΝΑΙΤηλεόραση} \mid d6)$ .

**b.**

$$\begin{aligned} P(\text{επεισόδιο} \mid \text{ΝΑΙΤηλεόραση}) &= (3+1) / (3+2) = 4 / 5 \\ P(\text{επεισόδιο} \mid \text{ΌΧΙΤηλεόραση}) &= (1+1) / (2+2) = 2 / 4 \\ P(\text{γήπεδο} \mid \text{ΝΑΙΤηλεόραση}) &= (0+1) / (3+2) = 1 / 6 \\ P(\text{γήπεδο} \mid \text{ΌΧΙΤηλεόραση}) &= (2+1) / (2+2) = 3 / 4 \\ P(\text{ειδήσεις} \mid \text{ΝΑΙΤηλεόραση}) &= (1+1) / (3+2) = 2 / 5 \\ P(\text{ειδήσεις} \mid \text{ΌΧΙΤηλεόραση}) &= (1+1) / (2+2) = 2 / 4 \\ P(\text{πρόγραμμα} \mid \text{ΝΑΙΤηλεόραση}) &= (2+1) / (3+2) = 3 / 5 \\ P(\text{πρόγραμμα} \mid \text{ΌΧΙΤηλεόραση}) &= (0+1) / (2+2) = 1 / 4 \\ P(\text{σειρά} \mid \text{ΝΑΙΤηλεόραση}) &= (1+1) / (3+2) = 2 / 5 \\ P(\text{σειρά} \mid \text{ΌΧΙΤηλεόραση}) &= (0+1) / (2+2) = 1 / 4 \\ P(\text{ομάδα} \mid \text{ΝΑΙΤηλεόραση}) &= (0+1) / (3+2) = 1 / 5 \\ P(\text{ομάδα} \mid \text{ΌΧΙΤηλεόραση}) &= (2+1) / (2+2) = 3 / 4 \\ P(\text{κανάλι} \mid \text{ΝΑΙΤηλεόραση}) &= (3+1) / (3+2) = 4 / 5 \\ P(\text{κανάλι} \mid \text{ΌΧΙΤηλεόραση}) &= (0+1) / (2+2) = 1 / 4 \\ P(\text{ταινία} \mid \text{ΝΑΙΤηλεόραση}) &= (1+1) / (3+2) = 2 / 5 \\ P(\text{ταινία} \mid \text{ΌΧΙΤηλεόραση}) &= (0+1) / (2+2) = 1 / 4 \end{aligned}$$

$$P(\text{ΝΑΙΤηλεόραση} \mid d6) = P(\text{ΝΑΙΤηλεόραση}) * P(\text{επεισόδιο} \mid \text{ΝΑΙΤηλεόραση}) * \\ P(\text{γήπεδο} \mid \text{ΝΑΙΤηλεόραση}) * P(\text{ειδήσεις} \mid \text{ΝΑΙΤηλεόραση}) * (1 - P(\text{πρόγραμμα} \mid \\ \text{ΝΑΙΤηλεόραση})) * (1 - P(\text{σειρά} \mid \text{ΝΑΙΤηλεόραση})) * (1 - P(\text{ομάδα} \mid \text{ΝΑΙΤηλεόραση})) \\ * (1 - P(\text{κανάλι} \mid \text{ΝΑΙΤηλεόραση})) * (1 - P(\text{ταινία} \mid \text{ΝΑΙΤηλεόραση})) = (3/5) * (4/5) * \\ (1/5) * (2/5) * (2/5) * (3/5) * (4/5) * (2/5) * (3/5) = 0.002$$

$$P(\text{ΌΧΙΤηλεόραση} \mid d6) = P(\text{ΌΧΙΤηλεόραση}) * P(\text{επεισόδιο} \mid \\ \text{ΌΧΙΤηλεόραση}) * P(\text{γήπεδο} \mid \text{ΌΧΙΤηλεόραση}) * P(\text{ειδήσεις} \mid \text{ΌΧΙΤηλεόραση}) * (1 -$$

$$\begin{aligned}
 &P(\text{προγραμμα} \mid \text{'ΟΧΙΤηλεόραση})) * (1 - P(\text{σειρά} \mid \text{'ΟΧΙΤηλεόραση})) * (1 - P(\text{ομάδα} \mid \\
 &\text{'ΟΧΙΤηλεόραση})) * (1 - P(\text{κανάλι} \mid \text{'ΟΧΙΤηλεόραση})) * (1 - P(\text{ταινία} \mid \text{'ΟΧΙΤηλεόραση})) \\
 &= (2/5) * (1/2) * (3/4) * (1/2) * (3/4) * (3/4) * (1/4) * (3/4) * (3/4) = 0.006
 \end{aligned}$$

Οπότε φτάνουμε πάλι στο ίδιο συμπέρασμα και με την μέθοδο Bernouli όπως φτάσαμε και με την μέθοδο multinomial Naïve Bayes.