

2ο μέρος εργασίας (Αρχείο weather-nominal)

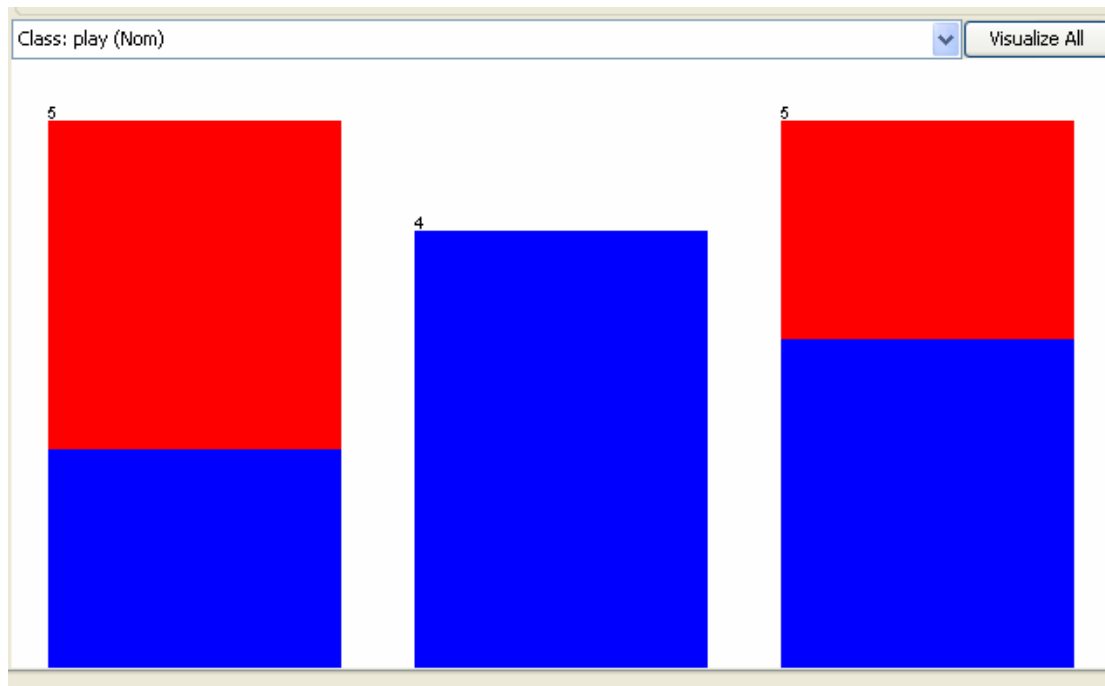


No.	outlook Nominal	temperature Nominal	humidity Nominal	windy Nominal	play Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Στο συγκεκριμένο dataset έχουν καταχωρηθεί οι τιμές από τέσσερα nominal attributes και από ένα class-attribute, επίσης nominal, και από 14 instances αυτών. Εξετάζεται συγκεκριμένα η όποια σχέση υπάρχει μεταξύ των attributes-παραγόντων outlook, temperature, humidity, windy και του class-attribute play (αν θα πραγματοποιηθεί δηλαδή παιχνίδι ή όχι), το οποίο πέρνει τιμές yes και no. Σκοπός είναι η εξαγωγή συσχετίσεων, μέσω οπτικοποίησης του dataset, καθώς και κάποιων γενικών χαρακτηριστικών του συνόλου των δεδομένων μας. Να σημειώσουμε πρώτα, όσον αφορά τα ιστογράμματα που θα ακολουθήσουν, ότι γενικά μας δείχνουν πόσο συχνά κάθε ένα από τα values του class=play (yes και no) συμβαίνουν για κάθε value του attribute που εξετάζεται κάθε φορά.

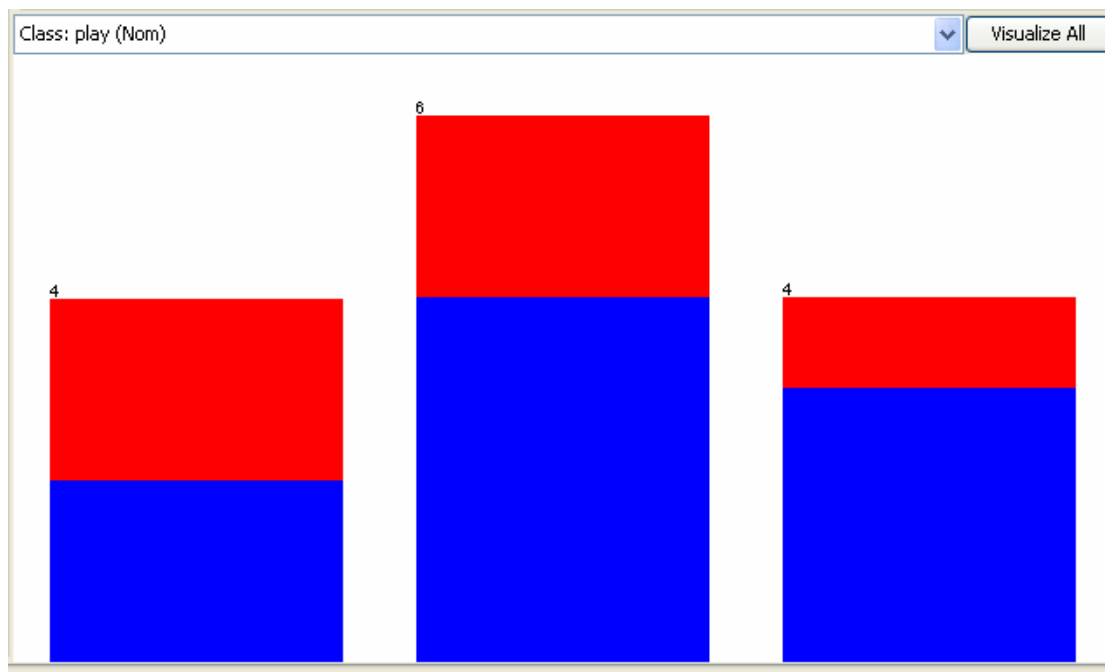
Παρατηρούμε λοιπόν τα εξής:

1) Όσον αφορά το attribute outlook, δεν υπάρχουν καθόλου missing values, έχουμε 3 distinct values (τρεις διαφορετικές nominal τιμές παίρνει-sunny, overcast, rainy) και κανένα unique value. Παραθέτουμε και το ιστόγραμμα του attribute αυτού σε σχέση με το class=play. Γενικά, το κόκκινο χρώμα θα σημαίνει play=no και το μπλέ play=yes:



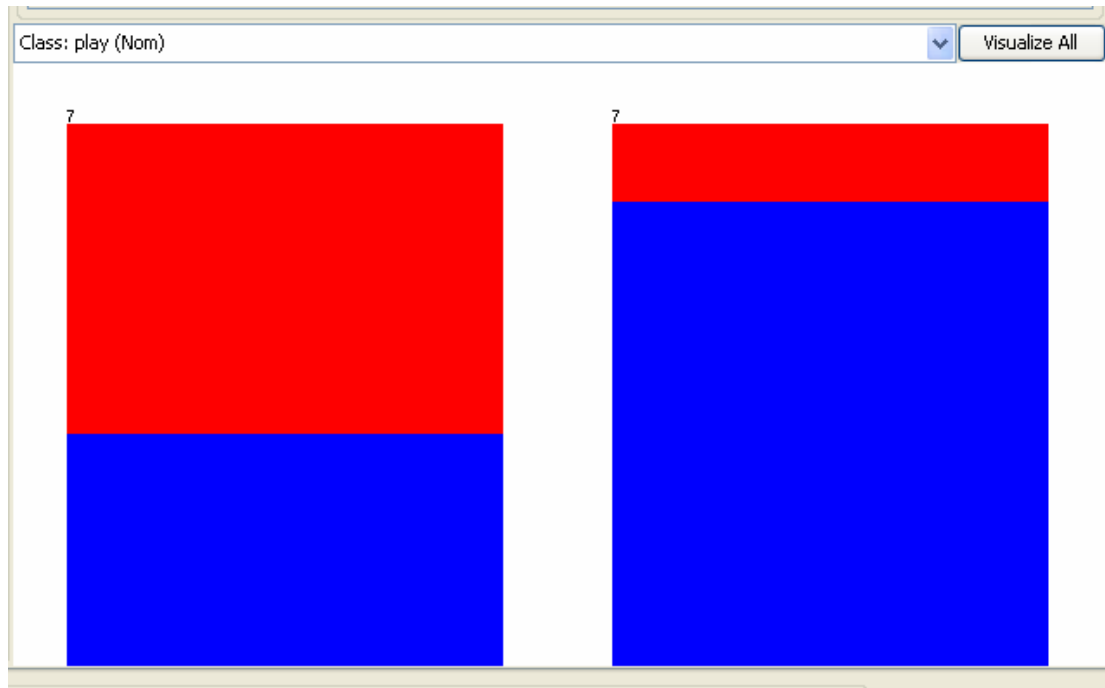
Παρατηρούμε ότι από τις 5 καταχωρημένες τιμές του attribute-oulook ως sunny (πρώτη ράβδος), οι τρεις συνεπάγονται μη πραγματοποίηση του παιχνιδιού-play=no (κόκκινο χρώμα) και οι δύο σε πραγματοποίηση-play=yes (μπλέ χρώμα). Από τις 4 τιμές του, ως overcast (δεύτερη ράβδος), όλες οδηγούν σε πραγματοποίηση του παιχνιδιού και από τις 5 τιμές του ως rainy (τρίτη ράβδος), οι δύο οδηγούν σε play=no και οι υπόλοιπες τρεις σε play=yes.

2) Όσον αφορά το attribute temperature, δεν υπάρχουν καθόλου missing values, έχουμε 3 distinct values (hot, mild, cool) και κανένα unique value.



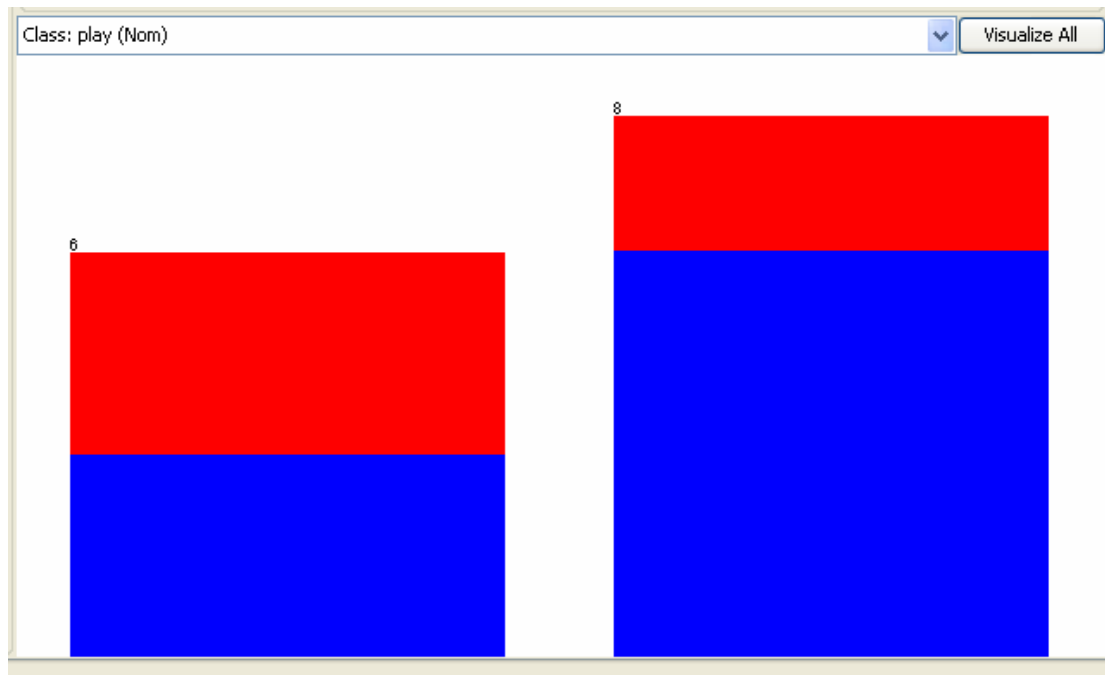
Επίσης από το ιστόγραμμα βλέπουμε ότι από τις 4 καταχωρημένες τιμές του temperature ως hot (πρώτη ράβδος), οι δύο συνεπάγονται play=yes και οι άλλες δύο play=no. Από τις 6 καταχωρήσεις ως mild (δεύτερη ράβδος), οι δύο οδηγούν σε play=no και οι τέσσερις σε play=yes. Από τις 4 καταχωρήσεις ως cool (τρίτη ράβδος), μία συνεπάγεται play=no και τρεις play=yes.

3) Όσον αφορά το attribute humidity, δεν υπάρχουν καθόλου missing values, έχουμε 2 distinct values (high, normal) και κανένα unique value.



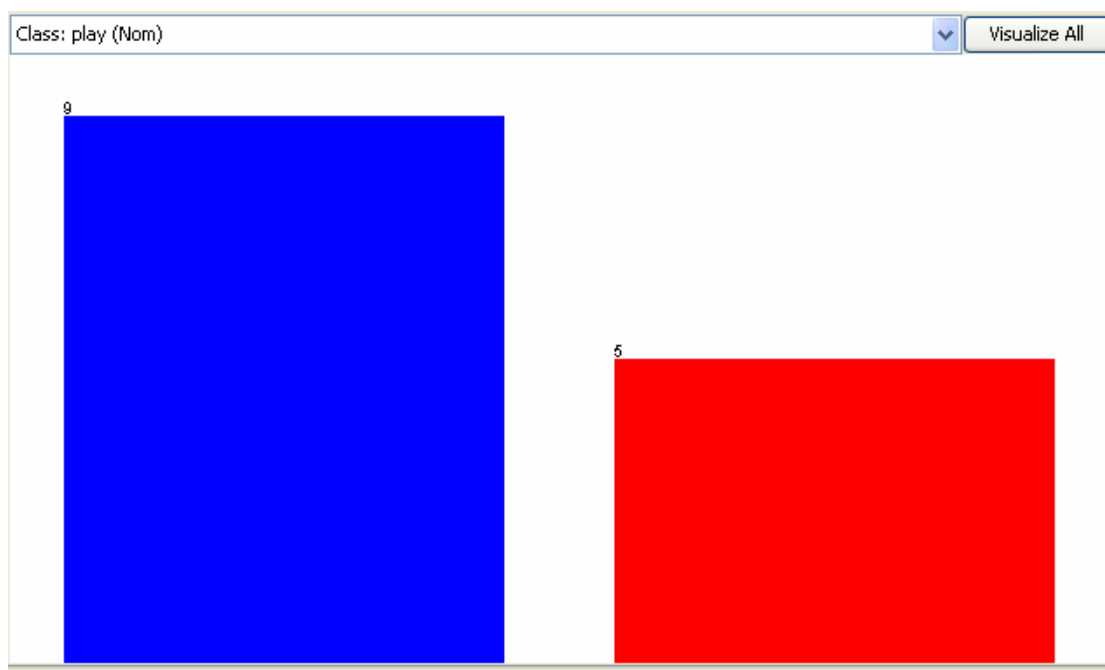
Εδώ έχουμε 7 καταχωρήσεις του attribute-humidity ως high (πρώτη ράβδος), εκ των οποίων οι τέσσερις συνεπάγονται play=no και οι τρεις play=yes. Από τις επίσης 7 καταχωρήσεις του ως normal (δεύτερη ράβδος), μόνο μία οδηγεί σε play=no ενώ για τις υπόλοιπες έξι έχουμε play=yes.

4) Όσον αφορά το attribute windy, δεν υπάρχουν καθόλου missing values, έχουμε 2 distinct values (true, false) και κανένα unique value.



Βάσει του παραπάνω ιστογράμματος, έχουμε 6 καταχωρήσεις του attribute windy ως true (πρώτη ράβδος), εκ των οποίων τρεις συνεπάγονται play=no και οι υπόλοιπες τρεις έχουν ως αποτέλεσμα play=yes. Από τις 8 καταχωρήσεις ως false (δεύτερη ράβδος), για τις δύο έχουμε play=no και για τις έξι έχουμε play=yes.

5) Στο τελευταίο ιστόγραμμα, βλέπουμε ότι στο σύνολο των 14 instances, τα εννιά έχουν ως yes την τιμή του class-attribute play (πρώτη ράβδος) και τα πέντε έχουν no (δεύτερη ράβδος).



Τελειώνοντας, να σημειώσουμε ότι από απλή οπτική διερεύνηση των δεδομένων, προκύπτουν τα εξής associations-συσχετίσεις (προφανώς δεν είναι τα μοναδικά):

- a) outlook=overcast→play=yes.
- b) humidity=normal and windy=false→play=yes.
- c) outlook=sunny and humidity=high→play=no.
- d) outlook=rainy and windy=false→ play=yes.