Εξόρυξη δεδομένων και αλγόριθμοι μάθησης
Εαρινό εξάμηνο 2017 Project RapidMiner (part A)

Exercise 1 –Preprocessing

1. Απόσπασμα του dataset

R...	Cl...	ge...	Na...	Place...	StageID	Gr...	Se...	To...	Sem...	Rel...	rais...	Vis...	Ann...	Dis...	Par...	Par...	Stude...
1	M	M	KW	KuwaIT	lowerlev...	G-04	A	IT	F	Fath...	15	16	2	20	Yes	Good	Under-7
2	M	M	KW	KuwaIT	lowerlev...	G-04	A	IT	F	Fath...	20	20	3	25	Yes	Good	Under-7
3	L	M	KW	KuwaIT	lowerlev...	G-04	A	IT	F	Fath...	10	7	0	30	No	Bad	Above-7
4	L	M	KW	KuwaIT	lowerlev...	G-04	A	IT	F	Fath...	30	25	5	35	No	Bad	Above-7
5	M	M	KW	KuwaIT	lowerlev...	G-04	A	IT	F	Fath...	40	50	12	50	No	Bad	Above-7
6	M	F	KW	KuwaIT	lowerlev...	G-04	A	IT	F	Fath...	42	30	13	70	Yes	Bad	Above-7
7	L	M	KW	KuwaIT	MiddleS...	G-07	A	Math	F	Fath...	35	12	0	17	No	Bad	Above-7
8	M	M	KW	KuwaIT	MiddleS...	G-07	A	Math	F	Fath...	50	10	15	22	Yes	Good	Under-7
9	M	F	KW	KuwaIT	MiddleS...	G-07	A	Math	F	Fath...	12	21	16	50	Yes	Good	Under-7
10	M	F	KW	KuwaIT	MiddleS...	G-07	B	IT	F	Fath...	70	80	25	70	Yes	Good	Under-7
11	H	M	KW	KuwaIT	MiddleS...	G-07	A	Math	F	Fath...	50	88	30	80	Yes	Good	Under-7
12	M	M	KW	KuwaIT	MiddleS...	G-07	B	Math	F	Fath...	19	6	19	12	Yes	Good	Under-7
13	L	M	KW	KuwaIT	lowerlev...	G-04	A	IT	F	Fath...	5	1	0	11	No	Bad	Above-7
14	L	M	leb...	lebanon	MiddleS...	G-08	A	Math	F	Fath...	20	14	12	19	No	Bad	Above-7
15	H	F	KW	KuwaIT	MiddleS...	G-08	A	Math	F	Mum	62	70	44	60	No	Bad	Above-7
16	M	F	KW	KuwaIT	MiddleS...	G-06	A	IT	F	Fath...	30	40	22	66	Yes	Good	Under-7

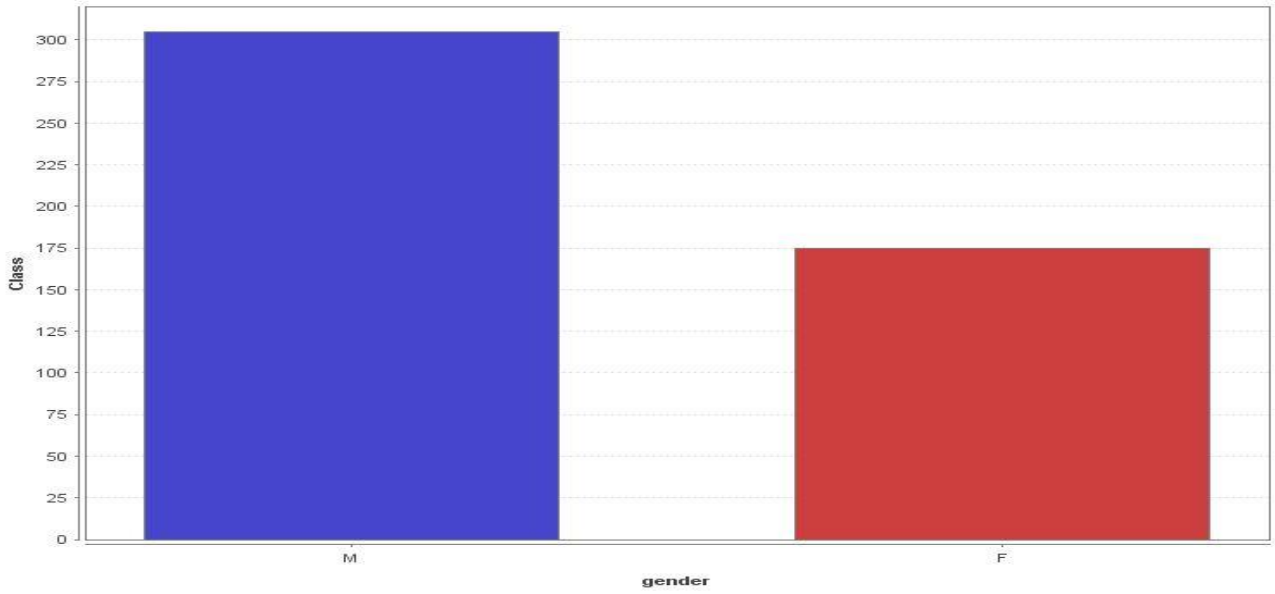
a) (Αναγνώριση dataset) Αναφέρετε τα χαρακτηριστικά του dataset δίνοντας για το καθένα την περιγραφή του, τον τύπο του και τις διακριτές τιμές που λαμβάνει.

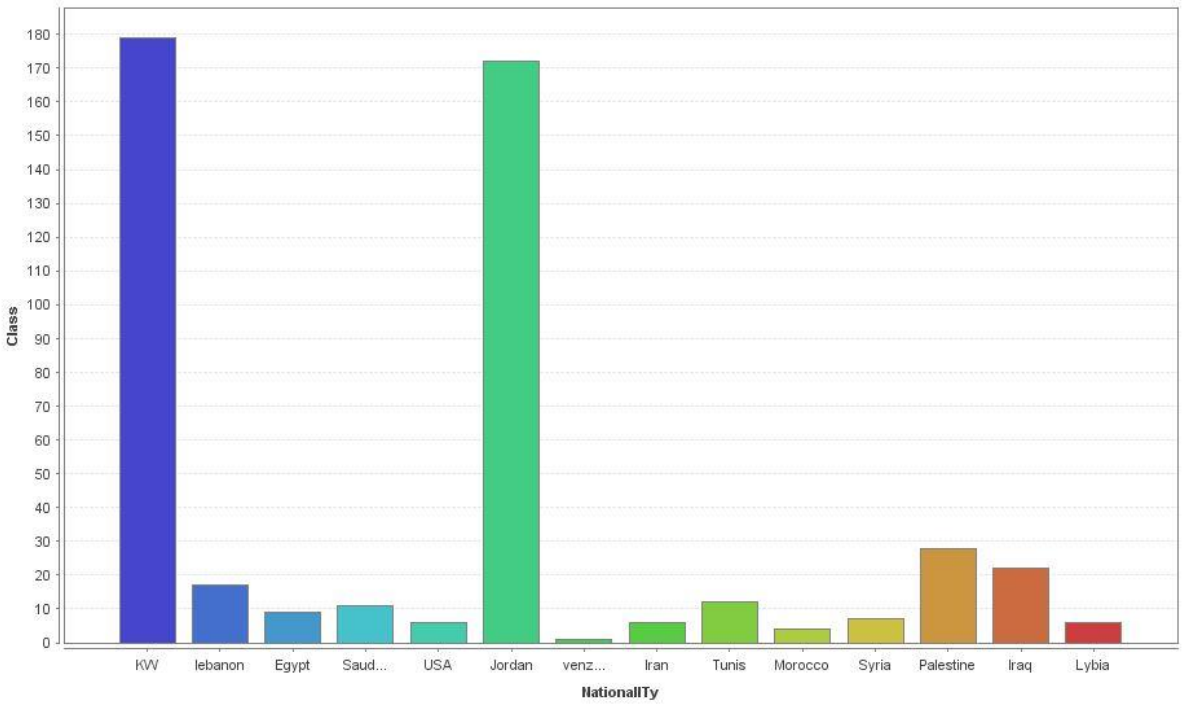
Όπως βλέπουμε έχουμε 16 attributes τα οποία αναλύονται παρακάτω :

- Gender: το φύλο κάθε μαθητή, είναι πολυωνυμικού τύπου χαρακτηριστικό και παίρνει τιμές M για αρσενικό και F για θηλυκο.
- Nationality: αφορά την εθνικότητα κάθε μαθητή, είναι πολυωνυμικού τύπου και για το dataset μας παίρνει μία από τις παρακάτω 14 τιμές(' Kuwait',' Lebanon',' Egypt',' SaudiArabia',' USA',' Jordan',' Venezuela',' Iran',' Tunis',' Morocco',' Syria',' Palestine',' Iraq',' Lybia') .
- PlaceofBirth: αφορά την χώρα γέννησης των μαθητών και ισχύουν τα ίδια με το Nationality.
- StageID: σε ποιο επίπεδο σπουδών ανήκουν οι μαθητές, είναι επίσης είναι πολυωνυμικού τύπου και παίρνει τιμές('lowerlevel','MiddleSchool','HighSchool').
- GradeID: σε ποια τάξη βρίσκεται κάθε μαθητής. Είναι και αυτό πολυωνυμικού τύπου ενώ παίρνει τιμές('G-01', 'G-02', 'G-03', 'G-04', 'G-05', 'G-06', 'G-07', 'G-08', 'G-09', 'G-10', 'G-11', 'G-12 ').
- SectionID: σε ποιο τμήμα ανήκει κάθε μαθητής αφού μια τάξη π.χ. G-02 μπορεί να χωρίζεται σε 3 sections(A,B,C). Είναι πολυωνυμικού τύπου χαρακτηριστικό.
- Topic: αφορά το αντικείμενο του μαθήματος, είναι πολυωνυμικού τύπου και παίρνει μία απ΄τις επόμενες 12 τιμές(' English',' Spanish', 'French',' Arabic',' IT',' Math',' Chemistry', 'Biology', 'Science',' History',' Quran',' Geology').
- Semester: σε ποιο από τα δύο σχολικά εξάμηνα αναφερόμαστε (F για πρώτο, S για δεύτερο), πολυωνυμικού τύπου.
- Relation: ποιος γονιός είναι υπεύθυνος για κάθε μαθητή (Mum για μητέρα, Father για πατέρα) , πολυωνυμικού τύπου.
- Raisedhands: οι φορές που σήκωσε το χέρι του κάποιος μαθητής στην τάξη κατά τη διάρκεια του εξαμήνου, είναι τύπου ακεραίου και παίρνει τιμές από 0-100.
- VisitedResources: πόσες φορές κάθε μαθητής χρησιμοποίησε σχολικούς πόρους, είναι τύπου ακεραίου και παίρνει τιμές από 0-100.
- AnnouncementsView: οι φορές που κάθε μαθητείς ελέγχει τις νέες ανακοινώσεις.Ισχύει ότι και για τα 2 προηγούμενα.
- Discussion: πόσες φορές συμμετείχε ένας μαθητής σε discussion group.Είναι επίσης ακεραίου τύπου και παίρνει τιμές από 0-100.
- ParentAnsweringSurvey: αν ο υπεύθυνος για το παιδί γονιός απάντησε στην έρευνα που του δώσανε απ΄το σχολείο (Yes) ή όχι(No).Είναι πολυωνυμικού τύπου.
- ParentSchoolSatisfaction: αν οι γονείς θεωρούν το σχολείο καλό (Good) ή κακό (Bad). Είναι πολυωνυμικού τύπου.
- StudentAbsenceDays: πολυωνυμικού τύπου χαρακτηριστικό που δείχνει αν κάποιος μαθητής έχει λείψει πάνω (Above-7) ή κάτω(Under-7) από 7 μέρες απ΄το σχολείο.

Τέλος έχουμε το ειδικό χαρακτηριστικό Class(πολυωνυμικού τύπου) που επιθυμούμε να μελετήσουμε στο οποίο κατατάσσονται οι μαθητές με μία από τις 3 τιμές(L,M,H).

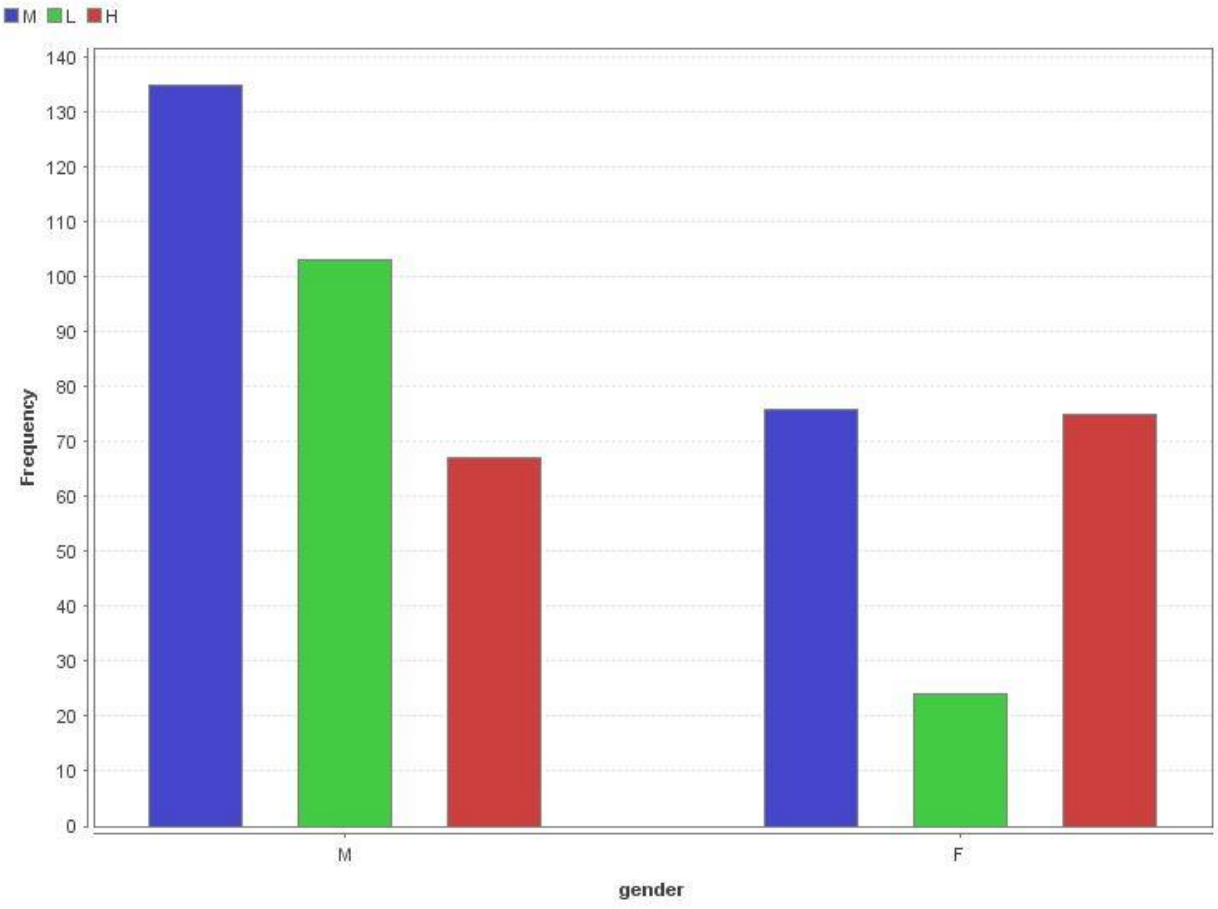
b) Για τα χαρακτηριστικά gender και nationality εισάγετε στην αναφορά σας τα αντίστοιχα charts (bars) που δείχνουν την συμμετοχή (count) των τιμών τους





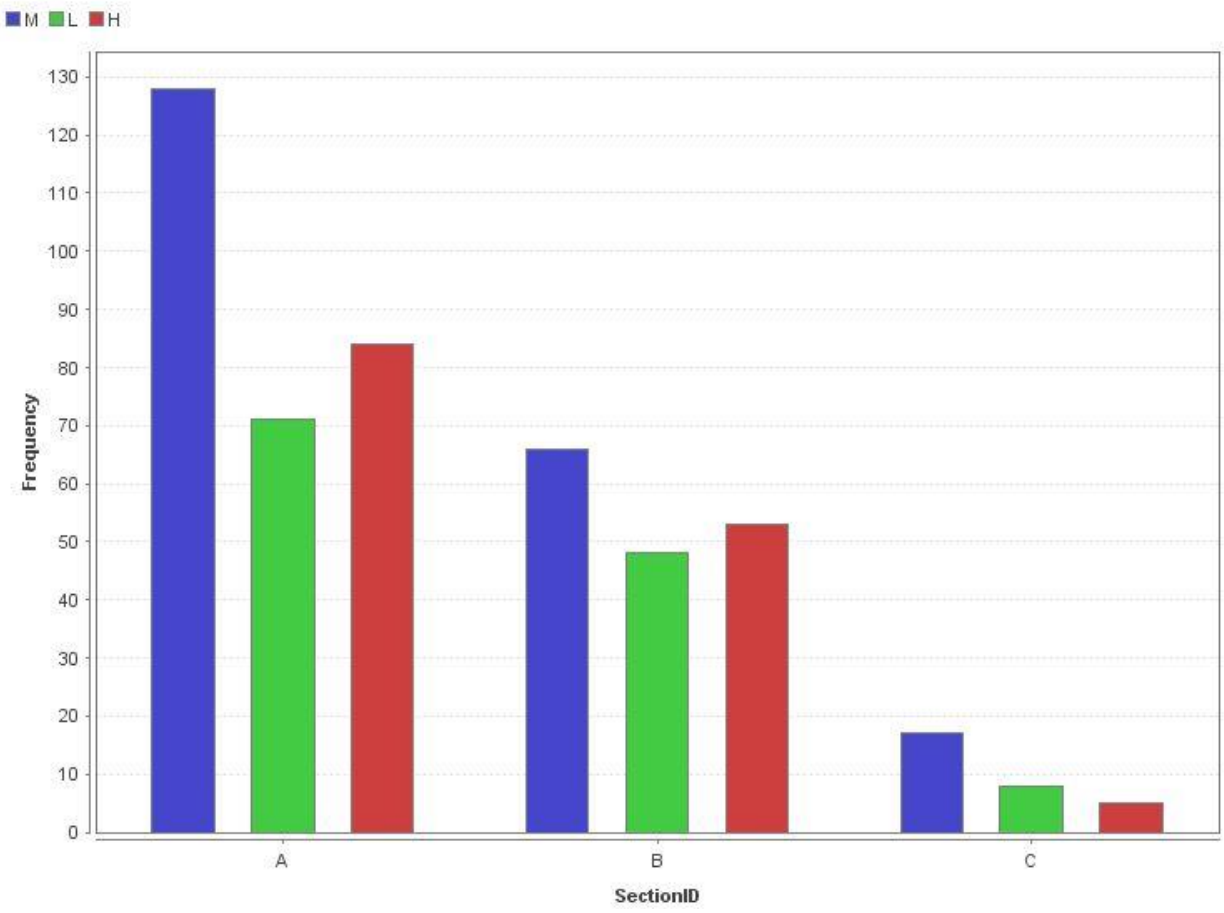
c) Χρησιμοποιώντας το chart Histogram color (όπου color η Class) απαντήστε στα παρακάτω ερωτήματα:

i. Ποιο gender έχει καλύτερη επίδοση;



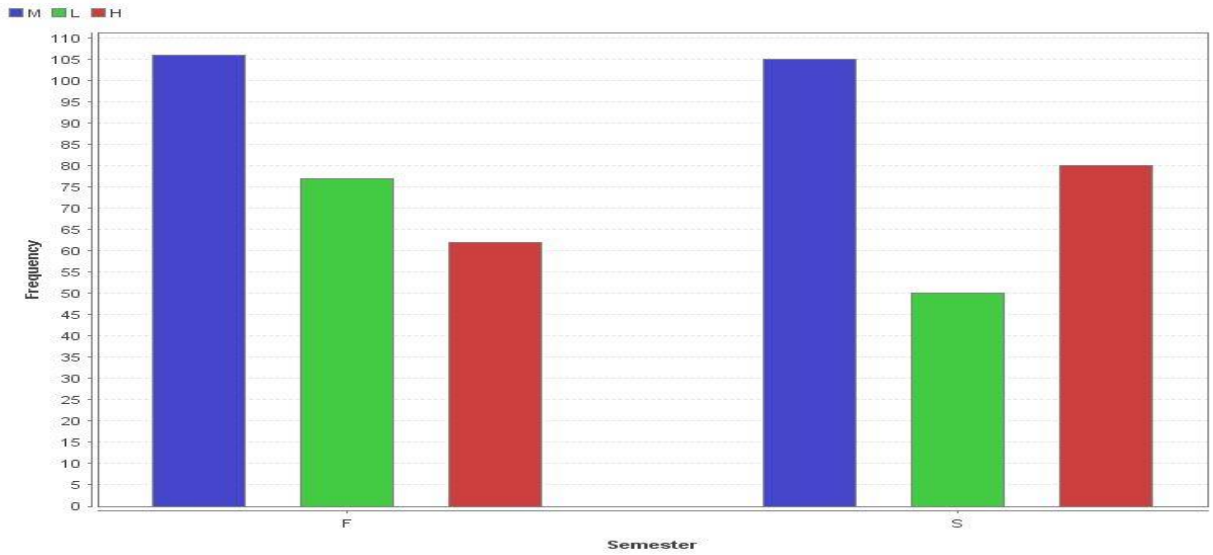
Παρατηρούμε ότι οι γυναίκες που ανήκουν στην High class είναι περισσότερες από τους άντρες στην κλάση αυτή, παρ΄ όλο που το δείγμα περιέχει περισσότερους άντρες. Ακόμα το ποσοστό των γυναικών που ανήκουν στη Low κλάση είναι αρκετά μικρό (25/(75+75+25)=14%).Αντίθετα στους άντρες το ποσοστό αυτό είναι αρκετά μεγαλύτερο(105/(135+67+104)=34%) οδηγώντας έτσι στο συμπέρασμα ότι το γυναικείο φύλο έχει καλύτερο επίδοση.

ii. Ποιο section έχει την χειρότερη επίδοση συγκρινόμενο με τα άλλα δύο;



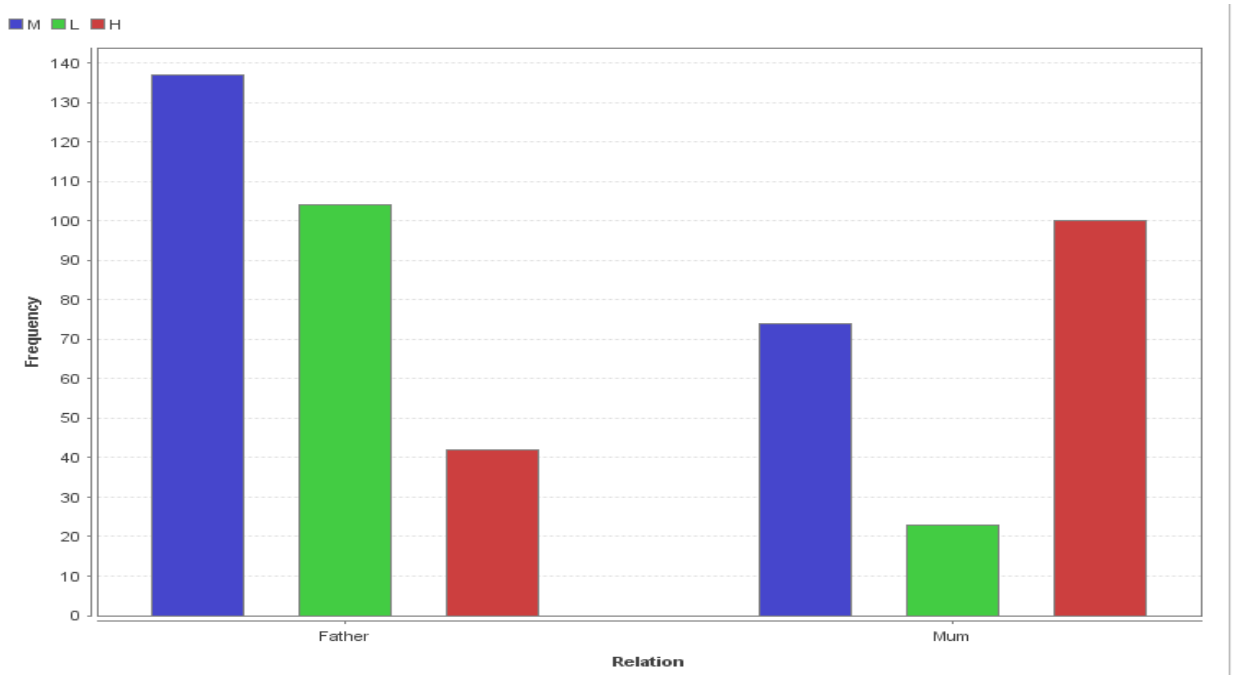
Θα θεωρήσουμε section με χειρότερη επίδοση αυτό που έχει μεγαλύτερο ποσοστό μαθητών στην κλάση L. Το section A έχει: 71/(128+71+84)=25% , section B: 48/(66+48+53)=28.7%, section C: 8/(17+8+4)=27.6% .Επειδή βλέπουμε ότι το ποσοστό της κλάσης L είναι αρκετα κοντά για τα section B και C θα επιλέξουμε σαν χειρότερο αυτο με τους λιγότερους στην κλάση H,δηλαδή το section B.

iii. Ποιό semester έχει τους περισσότερους φοιτητές με class L;



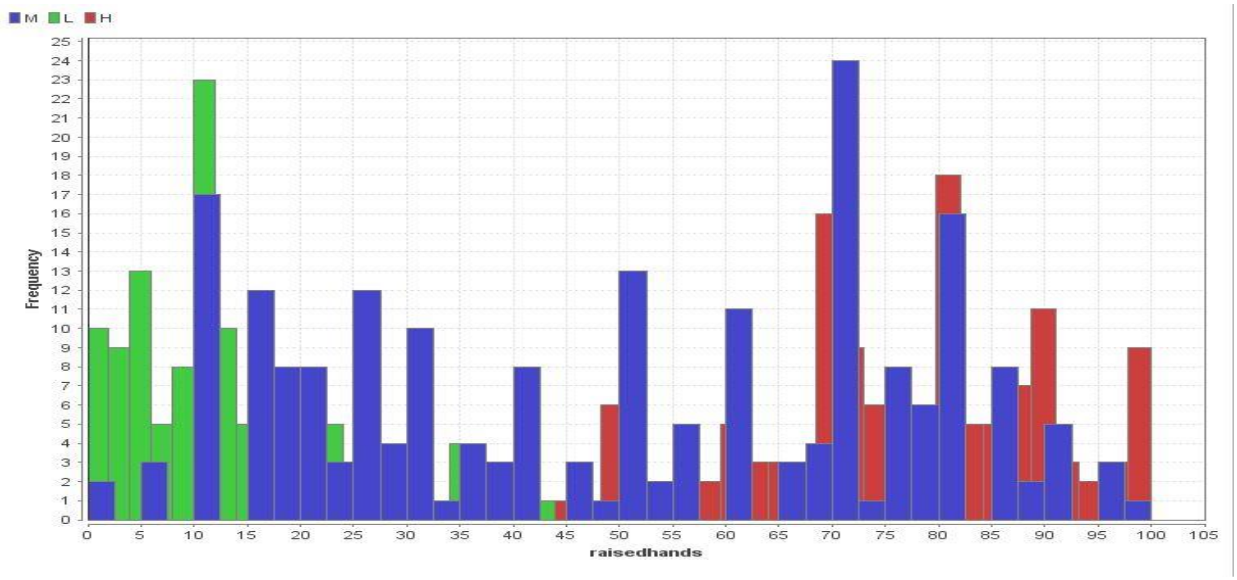
Όπως βλέπουμε το πρώτο εξάμηνο έχει περισσότερους μαθητές με κλάση L.

iv. Επηρεάζει το χαρακτηριστικό relation την απόδοση των φοιτητών; Αν ναι, ποιος γονέας φέρνει καλύτερα αποτελέσματα



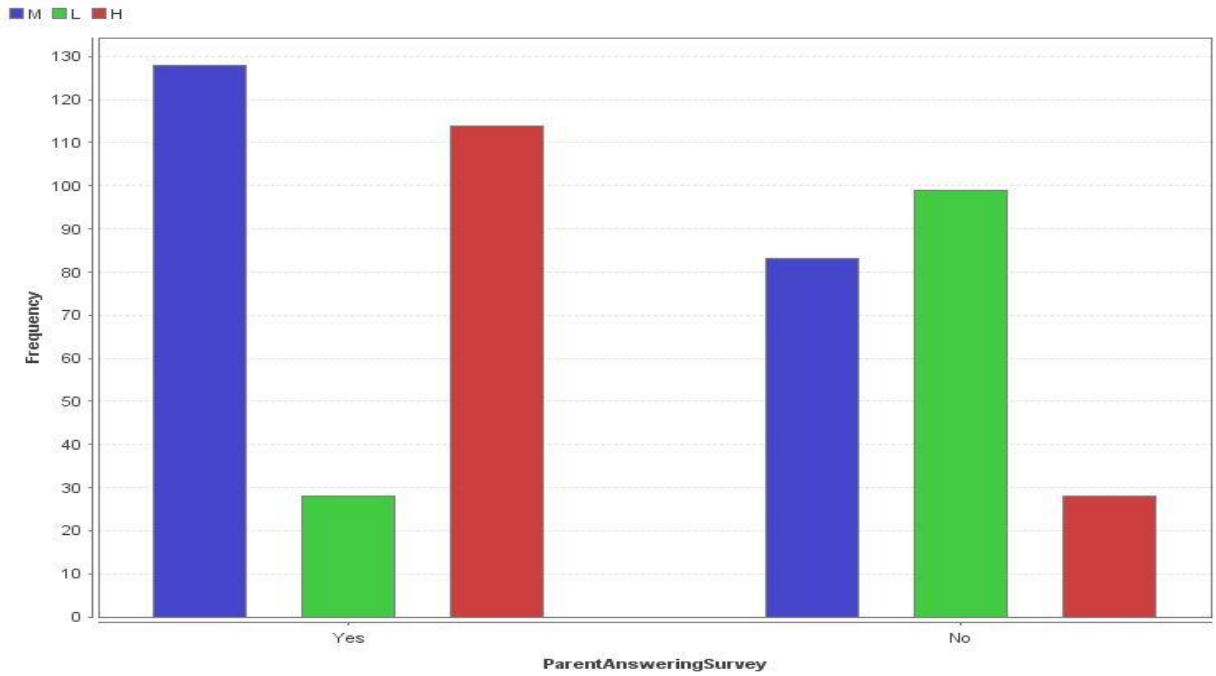
Βλέπουμε ότι το χαρακτηριστικό relation επηρεάζει αρκετά την απόδοση των μαθητών αφού φαίνεται ότι οι μαμάδες τα πηγαίνουν πολύ καλύτερα.

v. Σε ποια κλάση ανήκουν οι φοιτητές που έχουν περισσότερες απορίες (raisedhands);



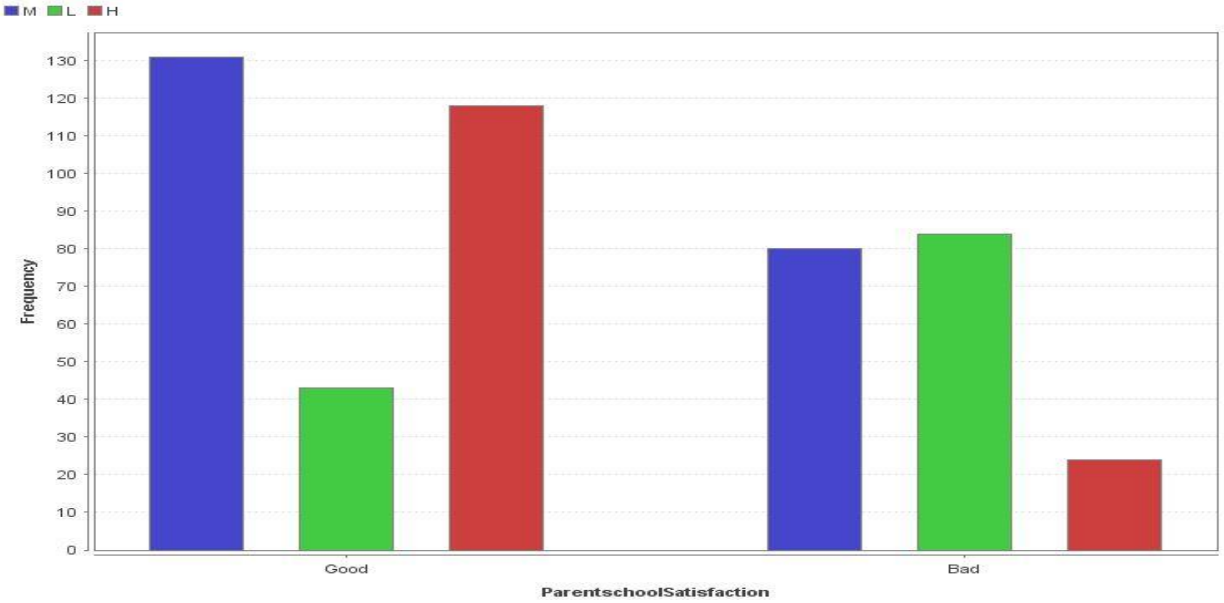
Από το παραπάνω ιστόγραμμα συμπεραίνουμε ότι οι μαθητές που σηκώνουν τις περισσότερες φορές το χέρι τους ανήκουν στην κλάση H.

vi. Οι γονείς που απάντησαν στο ερωτηματολόγιο έχουν φοιτητές με καλύτερη απόδοση;



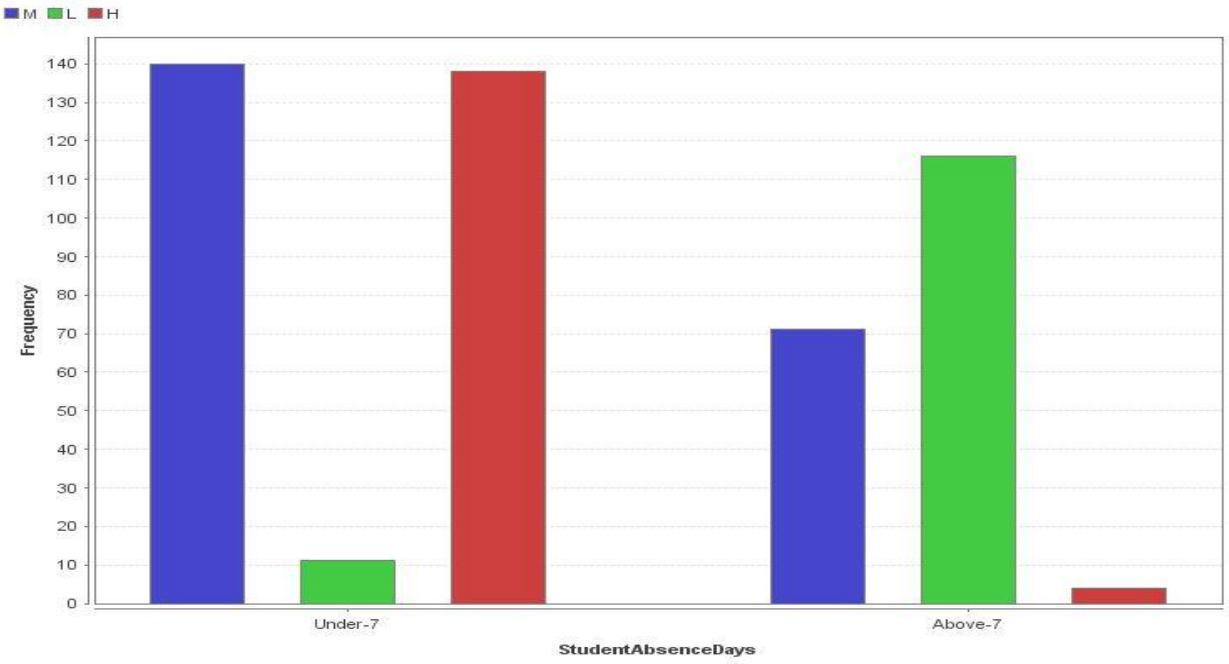
Είναι εμφανές ότι οι γονείς που απάντησαν στην έρευνα του σχολείου έχουν μαθητές με καλύτερη απόδοση αφού η αναλογία των κλάσεων H και L είναι αντιστρόφως ανάλογη για τις 2 περιπτώσεις.

vii. Υπάρχει συνάφεια (ομοιότητα) μεταξύ των χαρακτηριστικών ParentAnsweringSurvey και ParentsSchoolSatisfaction;



Μπορούμε να πούμε ότι υπάρχει συνάφεια μεταξύ των δύο αυτών χαρακτηριστικών όπως αυτό φαίνεται και από τα δύο ιστογράμματα.

viii. Το χαρακτηριστικό StudentAbsenceDays συμβάλει στην κλάση και σε ποιο βαθμό (ελάχιστα, αρκετά, σημαντικά).



Βλέποντας την μεγάλη και αντίστροφα ανάλογη απόκλιση μεταξύ των κλάσεων H και L ανάλογα με τον αριθμό απουσιών καταλαβαίνουμε πως το χαρακτηριστικό StudentAbsenceDays συμβάλει σημαντικά στην κατηγοριοποίηση των μαθητών.

2. Εντοπίστε τα 5 σημαντικότερα χαρακτηριστικά.

   Τα 5 σημαντικότερα χαρακτηριστικά όπως αυτά προέκυψαν από την 1.2-Feature Extraction process είναι:
   VisitedResources,StudentAbsenceDays,Relation,raisedhands,ParentAnsweringSurvey

3. Δημιουργία μειωμένου dataset με βάση τα παραπάνω 5 χαρακτηριστικά.

4. Δημιουργία dataset το οποίο πέρα από τα 5 σημαντικά χαρακτηριστικά έχει ακόμα 2 δικής μας επιλογής.Τα χαρακτηριστικά αυτά είναι τα AnnouncementsView και Discussion τα οποία επιλέχθηκαν λόγω της συχνής συμμετοχής τους στο Decision Tree, αλλά και πρατηρώντας στα ιστογράμματα τους ότι η τιμή τους είναι αντιπροσωπευτική της κλάσης των μαθητών(π.χ μεγάλη τιμή-> High class).

Exercise 2 –Decision Tree

1. 2.1-(Decision Tree) split

accuracy: 67.36%
ConfusionMatrix:

True:	M	L	H
M:	27	5	6
L:	7	33	0
H:	29	0	37

2. 2.2-(Decision Tree) cross

accuracy: 68.54% +/- 7.24% (mikro: 68.54%)
ConfusionMatrix:

True:	M	L	H
M:	111	17	32
L:	25	110	2
H:	75	0	108

3. 2.3-(Decision Tree) cross a5

accuracy: 67.29% +/- 5.67% (mikro: 67.29%)
ConfusionMatrix:

True:	M	L	H
M:	115	23	37
L:	30	103	0
H:	66	1	105

4. 2.4-(Decision Tree) cross a7

accuracy: 67.29% +/- 7.74% (mikro: 67.29%)
ConfusionMatrix:

True:	M	L	H
M:	108	20	33
L:	29	107	1
H:	74	0	108

5. Συγκριτικός πίνακας

	Split	Cross	Reduced 5	Reduced 7
Decision tree accuracy	67.36%	68.54% +/- 7.24%	67.29% +/- 5.67%	67.29% +/- 7.74%

Exercise 3 – Naive Bayes

1. 3.1-(Naive Bayes) split

accuracy: 65.28%
ConfusionMatrix:

True:	M	L	H
M:	32	7	12
L:	14	31	0
H:	17	0	31

2. 3.2-(Naive Bayes) cross

accuracy: 65.42% +/- 9.23% (mikro: 65.42%)
ConfusionMatrix:

True:	M	L	H
M:	108	20	40
L:	40	106	2
H:	63	1	100

3. 3.3-(Naive Bayes) cross a5

accuracy: 73.75% +/- 5.61% (mikro: 73.75%)
ConfusionMatrix:

True:	M	L	H
M:	131	15	28
L:	32	111	2
H:	48	1	112

4. 3.4-(Naive Bayes) cross a7

accuracy: 70.00% +/- 6.60% (mikro: 70.00%)
ConfusionMatrix:

True:	M	L	H
M:	116	16	30
L:	40	111	3
H:	55	0	109

5. Συγκριτικός πίνακας

	Split	Cross	Reduced 5	Reduced 7
Naive Bayes accuracy	65.28%	65.42% +/- 9.23%	73.75% +/- 5.61%	70.00% +/- 6.60%

Exercise 4 – Rule Induction

1. 4.1-(Rule Induction) split

accuracy: 70.83%
ConfusionMatrix:

True:	M	L	H
M:	43	9	10
L:	3	26	0
H:	17	3	33

2. 4.2-(Rule Induction) cross

accuracy: 69.17% +/- 7.95% (mikro: 69.17%)
ConfusionMatrix:

True:	M	L	H
M:	155	41	49
L:	14	84	0
H:	42	2	93

3. 4.3-(Rule Induction) cross a5

accuracy: 68.54% +/- 7.88% (mikro: 68.54%)
ConfusionMatrix:

True:	M	L	H
M:	148	33	53
L:	20	92	0
H:	43	2	89

4. 4.4-(Rule Induction) cross a7

accuracy: 69.17% +/- 7.61% (mikro: 69.17%)
ConfusionMatrix:

True:	M	L	H
M:	152	40	48
L:	26	86	0
H:	33	1	94

5. Συγκριτικός πίνακας

	Split	Cross	Reduced 5	Reduced 7
Rule Induction accuracy	70.83%	69.17% +/- 7.95%	68.54% +/- 7.88%	69.17% +/- 7.61%

Exercise 5 – Neural Net

1. 5.1-(Neural Net) split

accuracy: 77.08%
ConfusionMatrix:

True:	M	L	H
M:	52	10	11
L:	6	27	0
H:	5	1	32

2. 5.2-(Neural Net) cross

accuracy: 78.75% +/- 3.70% (mikro: 78.75%)
ConfusionMatrix:

True:	M	L	H
M:	161	19	31
L:	18	107	1
H:	32	1	110

3. 5.3-(Neural Net) cross a5

accuracy: 75.00% +/- 4.17% (mikro: 75.00%)
ConfusionMatrix:

True:	M	L	H
M:	145	13	41
L:	26	114	0
H:	40	0	101

4. 5.4-(Neural Net) cross a7

accuracy: 75.00% +/- 5.74% (mikro: 75.00%)
ConfusionMatrix:

True:	M	L	H
M:	150	19	40
L:	23	108	0
H:	38	0	102

5. Συγκριτικός πίνακας

	Split	Cross	Reduced 5	Reduced 7
Neural Net accuracy	77.08%	78.75% +/- 3.70%	75.00% +/- 4.17%	75.00% +/- 5.74%

Exercise 6 – k-NN

Σημείωση: Χρησιμοποιήθηκε Discretize by Binning χωρίζοντας το διάστημα 0-100 των αριθμητικών μας χαρακτηριστικών σε 8 bins.

1. 6.1-(k-NN) split

accuracy: 72.92%
ConfusionMatrix:

True:	M	L	H
M:	47	9	13
L:	5	28	0
H:	11	1	30

2. 6.2-(k-NN) cross

accuracy: 70.00% +/- 6.54% (mikro: 70.00%)
ConfusionMatrix:

True:	M	L	H
M:	133	17	46
L:	28	109	2
H:	50	1	94

3. 6.3-(k-NN) cross a5

accuracy: 68.33% +/- 6.31% (mikro: 68.33%)
ConfusionMatrix:

True:	M	L	H
M:	121	22	40
L:	32	105	0
H:	58	0	102

4. 6.4-(k-NN) cross a7

accuracy: 72.92% +/- 5.82% (mikro: 72.92%)
ConfusionMatrix:

True:	M	L	H
M:	135	17	36
L:	29	110	1
H:	47	0	105

5. Συγκριτικός πίνακας

	Split	Cross	Reduced 5	Reduced 7
k-NN accuracy	72.92%	70.00% +/- 6.54%	68.33% +/- 6.31%	72.92% +/- 5.82%

Exercise 7 – SVM

1. 7.1-(SVM) split

accuracy: 65.97%
ConfusionMatrix:

True:	M	L	H
M:	44	9	17
L:	9	27	2
H:	10	2	24

2. 7.2-(SVM) cross

accuracy: 67.29% +/- 4.76% (mikro: 67.29%)
ConfusionMatrix:

True:	M	L	H
M:	153	31	63
L:	28	95	4
H:	30	1	75

3. 7.3-(SVM) cross a5

accuracy: 63.12% +/- 6.32% (mikro: 63.12%)
ConfusionMatrix:

True:	M	L	H
M:	139	26	75
L:	37	100	3
H:	35	1	64

4. 7.4-(SVM) cross a7

accuracy: 66.88% +/- 4.79% (mikro: 66.88%)
ConfusionMatrix:

True:	M	L	H
M:	151	31	63
L:	28	95	4
H:	32	1	75

5. Συγκριτικός πίνακας

	Split	Cross	Reduced 5	Reduced 7
SVM accuracy	65.97%	67.29% +/- 4.76%	63.12% +/- 6.32%	66.88% +/- 4.79%

Exercise 8 – Compare models

1.

	split	cross	Reduced 5	Reduced 7
Decision tree	67.36%	68.54% +/- 7.24%	67.29% +/- 5.67%	67.29% +/- 7.74%
Naive Bayes	65.28%	65.42% +/- 9.23%	73.75% +/- 5.61%	70.00% +/- 6.60%
Rule Induction	70.83% 77.78%(discrete)	69.17% +/- 7.95% 71.25% +/- 4.35%(discrete)	68.54% +/- 7.88% 71.88% +/- 5.37%(discrete)	69.17% +/- 7.61% 71.67% +/- 4.95% (discrete)
Neural Net	77.08%	78.75% +/- 3.70%	75.00% +/- 4.17%	75.00% +/- 5.74%
k-NN	72.92%	70.00% +/- 6.54%	68.33% +/- 6.31%	72.92% +/- 5.82%
SVM	65.97%	67.29% +/- 4.76%	63.12% +/- 6.32%	66.88% +/- 4.79%

2. Ποιος αλγόριθμος δίνει τα καλύτερα αποτελέσματα. Σε ποιο dataset κάνετε την αξιολόγησή σας;

Ο αλγόριθμος που τρέχει καλύτερα για όλα τα είδη dataset βλέπουμε ότι είναι ο Neural Net.

3. Σε τι είδος dataset αποδίδει καλύτερα ο κάθε αλγόριθμος. Αναφέρετε τους περιορισμούς που υφίστανται στον καθένα από τους άνω αλγορίθμους ως προς τον τύπο των χαρακτηριστικών και της κλάσης.

- Decision Tree: Η απόδοση του αλγορίθμου αυτή είναι σχεδόν ίδια για όλα τα dataset ενώ δεν υπήρξε κάποιος περιορισμός που χρειάστηκε να εφαρμοστεί.
- Naive Bayes: Ο αλγόριθμος αυτός αποδίδει καλύτερα στα dataset που έχουν μόνο τα σημαντικότερα χαρακτηριστικά, δηλαδή στα reduced 5 και 7. Δεν υπήρχε ούτε εδώ κάποιος περιορισμός αναφορικά με τον τύπο των χαρακτηριστικών.
- Rule Induction: Βλέπουμε ότι και αυτός ο αλγόριθμος αποδίδει περίπου το ίδιο για όλα τα dataset. Χωρίς να αποτελεί περιορισμό παρατηρούμε ότι αν διακριτοποιήσουμε τα δεδομένα η απόδοση του αλγορίθμου αυξάνεται.
- Neural Net: Δεν μπορούμε να πούμε ότι ο αλγόριθμος αποδίδει καλύτερα σε κάποιο συγκεκριμένο dataset. Επειδή ο operator αυτός δεν μπορεί να χειριστεί πολυωνυμικού τύπου χαρακτηριστικά χρειάστηκε να τα μετατρέψουμε σε αριθμητικού τύπου.
- K-NN: Ούτε για αυτόν τον αλγόριθμο μπορούμε να αποφανθούμε ότι τρέχει καλύτερα σε κάποιο είδος dataset. Ενώ δεν είναι δεσμευτική, η διακριτοποίση των αριθμητικών attribute βελτίωσε αρκετά την απόδοση του αλγορίθμου.
- SVM: Για τον αλγόριθμο αυτό βλέπουμε ότι η απόδοση του κυμαίνεται στα ίδια περίπου επίπεδα για όλα τα dataset.Όπως και με τα νευρωνικά δίκτυα χρειάστηκε να μετατρέψουμε τα polynominal χαρακτηριστικά σε integer τύπου.

4. Δώστε την ερμηνεία σας για την απόδοση του κάθε αλγορίθμου στο συγκεκριμένο dataset.

- Decision Tree: Τρέχοντας τα processes για τον συγκεκριμένο αλγόριθμο επιβεβαιώνονται τα θεωρητικά αποτελέσματα που περιμέναμε, δηλαδή μικροί χρόνοι εκπαίδευσης και εκτέλεσης καθώς και μέτρια ακρίβεια.
- Naive Bayes: Παρατηρήσαμε ότι ο αλγόριθμος αυτός αποδίδει καλύτερα στα 2 τελευταία dataset με τα μειωμένα χαρακτηριστικά.Αυτό γίνεται γιατί ο Naive Bayes υποθέτει ότι τα χαρακτηριστικά είναι ανεξάρτητα μεταξύ τους επομένως όταν λαμβάνει υπόψιν μόνο τα σημαντικότερα χαρακτηριστικά του dataset δίνει καλύτερα αποτελέσματα.
- Rule Induction: Βλέπουμε ότι και αυτός ο αλγόριθμος δίνει μια μέτρια ακρίβεια που βελτιώνεται λίγο αφού διακριτοποιήσουμε τα δεδομένα πάνω στα οποία δουλεύουν καλύτερα τα rule-based συστήματα.
- Neural Net: Ο αλγόριθμος που μας δίνει τα καλύτερα αποτελέσματα για το dataset μας αλλά χρειάζεται και τον μεγαλύτερο χρόνο εκπαίδευσης.Ακόμα επιβεβαιώνεται ότι αποδίδει καλύτερα σε μεγαλύτερα dataset αφού όταν τον τρέξαμε για μικρότερο δείγμα η ακρίβεια του έπεσε.

- K-NN: Ο αλγόριθμος αυτός μας δίνει μια μέτρια ακρίβεια για το dataset μας.Ωστόσο, ιδιαίτερη σημασία έχουν τα χαρακτηριστικά που επιλέγονται καθώς και η παράμετρος κ (οι κ πλησιέστεροι γείτονες).
  Τα συγκεκριμένα αποτελέσματα προέκυψαν με κ=3.
- SVM: Παρατηρούμε ότι ο SVM αλγόριθμος έχει χαμηλότερη απόδοση από ότι θα περιμέναμε από αυτόν.Ακόμα βλέπουμε ότι χρειάζεται κάποιον σημαντικό χρόνο εκπαίδευσης.Ο συγκεκριμένος αλγόριθμος χρησιμοποιείται συνήθως σε εφαρμογές αναγνώρισης χειρόγραφων ,ομιλίας ή εικόνας γεγονός το οποίο δικαιολογεί την μέτρια απόδοση του στο δικό μας dataset.