MELANOMA MACHINE LEARNING

Αποτελεσματα απο την εκπαίδευση διαφορων clasifiers που προβλεπουν αν καποιος εχει μελανομα η όχι. Οι classifier που χρησιμοποιουμε είναι οι Random Forest, Linear Regression,SVM,Linear Logistic Regression,OLS και GaussianNB.

Για το παρακατω πειραμα κρατήσαμε ολα τα snp με **pvalue <= p = 0.001.**
Οι ασθενεις μας ειναι στο συνολο 4980. Απο αυτούς οι 3962 δεν εχουν την ασθενεια**(control = 0)** και οι υπόλοιποι 1018 εχουν την ασθενεια**(case = 1)**.

Τα snp με **pvalue <= p = 0.001** είναι 5415 για να αποφύγουμε το **overfitting** μειώνουμε τα snps με τη μεθοδο του **random forest.** Το ποσο μειώνοντε τα χαρακτηριστικα εξαρτάται απο το **train data**.

Οι classifiers Random Forest, OLS και Linear Regression οι τιμες που κάνουν predict δεν τις επιστρεφουν σε 0 και 1 αλλα σε δεκαδικη μορφή, οπότε για τα παρακάτω πειράματα γι αυτους τους classifiers υπολογίζουμε το μεσο όρο των τιμών που έκαναν predict και οποια τιμή ειναι κατω απο το μεσο ορο την κανουμε μηδεν αλλιως ενα.

<u>1ο Πειραμα:</u>
Καναμε ενα τυχαίο sample του παραπανω δειγματος όπου το 90% ειναι το trainData και το 10% ειναι το testData. Μετα το sample στο 90% του δειγματος οι 3572 ασθενεις δεν εχουνε την ασθενεια(και 910 ασθενεις εχουν την ασθενεια. Στο 10% του δειγματος οι 390 ασθενεις δεν εχουν την ασθενεια και οι 108 εχουν την ασθενεια. Θα παρουιασουμε τα confusion matrix απο τις παραπανω τεχνικες που αναφεραμε.

Random Forest: [[291 99]
[7 101]]

Linear Regression: [[389   1]
[108   0]]

cros validation = {1: 0.79317269076305219, 2: 0.77911646586345384, 3:0.18674698795180722, 4: 0.89959839357429716, 5: 0.89959839357429716, 6: 0.19678714859437751, 7:0.88353413654618473, 8: 0.79116465863453811, 9: 0.91365461847389562, 10:0.8875502008032129}

SVM: [[390   0]
[108   0]]

cros validation = {1: 0.23694779116465864, 2: 0.20281124497991967, 3: 0.79919678714859432, 4: 0.81927710843373491, 5: 0.80522088353413657, 6: 0.20281124497991967, 7:0.21686746987951808, 8: 0.80522088353413657, 9: 0.19879518072289157, 10:0.8012048192771084}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[386 4]**
[41 67]]
cros validation = {1: 0.95783132530120485, 2: 0.94779116465863456, 3:0.95180722891566261, 4: 0.95582329317269077, 5: 0.96586345381526106, 6: 0.9618473895582329, 7:0.94779116465863456, 8: 0.94176706827309242, 9: 0.95381526104417669, 10:0.93975903614457834}

LINEAR LOGISTIC REGRESSION: [[388 2]
[44 64]]

cros validation = {1: 0.91967871485943775, 2: 0.94176706827309242, 3:0.94779116465863456, 4: 0.93975903614457834, 5: 0.95180722891566261, 6: 0.94377510040160639, 7:0.94377510040160639, 8: 0.94176706827309242, 9: 0.96586345381526106, 10:0.95582329317269077}

OLS: [[315 75]
[18 90]]

cros validation = {1: 0.8775100401606426, 2: 0.89156626506024095, 3: 0.91767068273092367, 4: 0.89156626506024095, 5: 0.89759036144578308, 6: 0.90963855421686746, 7:0.87951807228915657, 8: 0.89759036144578308, 9: 0.87951807228915657, 10:0.92168674698795183}

GaussianNB: [[378 12]
[49 59]]

cros validation = {1: 0.96586345381526106, 2: 0.95783132530120485, 3:0.94779116465863456, 4: 0.95983935742971882, 5: 0.94779116465863456, 6: 0.94979919678714864, 7:0.9618473895582329, 8: 0.94779116465863456, 9: 0.94377510040160639, 10:0.95381526104417669}

<u>2ο Πειραμα:</u>
Σε αυτο το πειραμα κάναμε balance το train data. Δηλαδη το train data αποτελειται απο 1018 ασθενεις που δεν εχουνε την ασθενεια και 1018 ασθενεις εχουν την ασθενεια. Οι υπόλοιποι 2944 που δεν εχουν την ασθενεια αποτελουν το test data.

Random Forest: [[1525 1419]
[0 0]]

Linear Regression: [[2927 17]
[0 0]]
cros validation = {1: 0.18473895582329317, 2: 0.20682730923694778, 3: 0.7831325301204819, 4: 0.19477911646586346, 5: 0.20682730923694778, 6: 0.1465863453815261, 7:0.20682730923694778, 8: 0.79116465863453811, 9: 0.23092369477911648, 10:0.75903614457831325}

SVM: [[2209 735]
[0   0]]

cros validation = {1: 0.84738955823293172, 2: 0.81325301204819278, 3:0.82530120481927716, 4: 0.78514056224899598, 5: 0.80923694779116462, 6: 0.83534136546184734, 7:0.80522088353413657, 8: 0.77911646586345384, 9: 0.82730923694779113, 10:0.82329317269076308}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[2424 520]**
[0   0]]
cros validation = {1: 0.89156626506024095, 2: 0.88554216867469882, 3: 0.8775100401606426, 4: 0.91967871485943775, 5: 0.8875502008032129, 6: 0.90361445783132532, 7:0.91164658634538154, 8: 0.88955823293172687, 9: 0.88152610441767065, 10:0.90361445783132532}

LINEAR LOGISTIC REGRESSION: [[2453 491]
[0   0]]
cros validation = {1: 0.91767068273092367, 2: 0.89357429718875503, 3:0.88152610441767065, 4: 0.88554216867469882, 5: 0.91365461847389562, 6: 0.91566265060240959, 7:0.91164658634538154, 8: 0.90963855421686746, 9: 0.89156626506024095, 10:0.90963855421686746}

OLS: [[1507 1437]
[0   0]]

cros validation = {1: 0.75502008032128509, 2: 0.74497991967871491, 3:0.75100401606425704, 4: 0.76907630522088355, 5: 0.76706827309236947, 6: 0.77510040160642568, 7:0.77911646586345384, 8: 0.76706827309236947, 9: 0.71285140562248994, 10:0.74497991967871491}

GaussianNB: [[2427 517]
[0   0]]

cros validation = {1: 0.81726907630522083, 2: 0.83534136546184734, 3:0.83534136546184734, 4: 0.82931726907630521, 5: 0.86144578313253017, 6: 0.8493975903614458, 7:0.8393574297188755, 8: 0.8112449799196787, 9: 0.83132530120481929, 10:0.8393574297188755}

<u>3ο Πειραμα:</u>
Σε αυτο το πειραμα κάναμε το train data αποτελειται απο 1068 ασθενεις που δεν εχουνε την ασθενεια και 712 ασθενεις που εχουν την ασθενεια.Το test data αποτελειται απο 2894 ασθενεις που δεν εχουνε την ασθενεια και 306 ασθενεις που εχουν την ασθενεια

Random Forest: [[1655 1239]
[19 287]]

Linear Regression: [[12 2882]
[2 304]]

cros validation = {1: 0.76706827309236947, 2: 0.17670682730923695, 3:0.21485943775100402, 4: 0.76907630522088355, 5: 0.82329317269076308, 6: 0.81726907630522083, 7:0.20281124497991967, 8: 0.22088353413654618, 9: 0.78714859437751006, 10:0.77911646586345384}

SVM: [[2510 384]
[116 190]]

cros validation = {1: 0.79718875502008035, 2: 0.82530120481927716, 3: 0.83734939759036142, 4: 0.80923694779116462, 5: 0.80522088353413657, 6: 0.78112449799196793, 7: 0.7831325301204819, 8: 0.82730923694779113, 9: 0.84136546184738958, 10: 0.81927710843373491}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[2510 384]**
[116 190]]
cros validation = {1: 0.89156626506024095, 2: 0.88554216867469882, 3: 0.8775100401606426, 4: 0.91967871485943775, 5: 0.8875502008032129, 6: 0.90361445783132532, 7:0.91164658634538154, 8: 0.88955823293172687, 9: 0.88152610441767065, 10:0.90361445783132532}

LINEAR LOGISTIC REGRESSION: [[2573 321]
[76 230]]

cros validation = {1: 0.90160642570281124, 2: 0.88353413654618473, 3:0.90562248995983941, 4: 0.88955823293172687, 5: 0.92369477911646591, 6: 0.90160642570281124, 7:0.92971887550200805, 8: 0.9337349397590361, 9: 0.90562248995983941, 10:0.91566265060240959}

OLS: [[1639 1255]
[42 264]]

cros validation = {1: 0.77911646586345384, 2: 0.81325301204819278, 3:0.79518072289156627, 4: 0.79919678714859432, 5: 0.78714859437751006, 6: 0.8112449799196787, 7:0.79919678714859432, 8: 0.75100401606425704, 9: 0.8012048192771084, 10:0.77710843373493976}

GaussianNB: [[2587 307]
[158 148]]

cros validation = {1: 0.89959839357429716, 2: 0.86947791164658639, 3:0.86546184738955823, 4: 0.86345381526104414, 5: 0.85742971887550201, 6: 0.86947791164658639, 7:0.89357429718875503, 8: 0.88152610441767065, 9: 0.87349397590361444, 10:0.84538152610441764}

4ο Πειραμα:
Σε αυτο το πειραμα κάναμε το train data αποτελειται απο 1068 ασθενεις που δεν εχουνε την ασθενεια και 712 ασθενεις που εχουν την ασθενεια.Το test data αποτελειται απο 306 ασθενεις που δεν εχουνε την ασθενεια και 306 ασθενεις που εχουν την ασθενεια

Random Forest: [[235 71]
[64 242]]

Linear Regression: [[304 2]
[303 3]]

cros validation = {1: 0.82845188284518834, 2: 0.43514644351464438, 3:0.41004184100418412, 4: 0.56485355648535562, 5: 0.61924686192468614, 6: 0.54811715481171552, 7:0.44351464435146443, 8: 0.53138075313807531, 9: 0.41841004184100417, 10:0.59751037344398339}

SVM: [[272 34]
[118 188]]

cros validation = {1: 0.7489539748953975, 2: 0.79079497907949792, 3: 0.75732217573221761, 4: 0.81589958158995812, 5: 0.82426778242677823, 6: 0.79497907949790791, 7:0.82845188284518834, 8: 0.80753138075313813, 9: 0.79497907949790791, 10:0.82572614107883813}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[274 32]**
[75 231]]
cros validation = {1: 0.85355648535564854, 2: 0.84937238493723854, 3: 0.83263598326359833, 4: 0.86192468619246865, 5: 0.85774058577405854, 6: 0.84937238493723854, 7:0.82426778242677823, 8: 0.87866108786610875, 9: 0.87029288702928875, 10:0.85477178423236511}

LINEAR LOGISTIC REGRESSION: [[271 35]
[74 232]]

cros validation = {1: 0.84100418410041844, 2: 0.83263598326359833, 3: 0.89539748953974896, 4: 0.91213389121338917, 5: 0.84518828451882844, 6: 0.83263598326359833, 7: 0.87447698744769875, 8: 0.82426778242677823, 9: 0.87029288702928875, 10: 0.86721991701244816}

OLS: [[246 60]
[88 218]]

cros validation = {1: 0.81171548117154813, 2: 0.81171548117154813, 3: 0.85355648535564854, 4: 0.87029288702928875, 5: 0.82008368200836823, 6: 0.86610878661087864, 7: 0.83263598326359833, 8: 0.82008368200836823, 9: 0.84937238493723854, 10: 0.82572614107883813}

GaussianNB:[[271 35]
[155 151]]

cros validation = {1: 0.84100418410041844, 2: 0.83263598326359833, 3: 0.80334728033472802, 4: 0.79497907949790791, 5: 0.80334728033472802, 6: 0.85774058577405854, 7: 0.82845188284518834, 8: 0.80753138075313813, 9: 0.82008368200836823, 10: 0.82572614107883813}

<u>5ο Πειραμα:</u>
Σε αυτο το πειραμα κάναμε το train data αποτελειται απο 800 ασθενεις που δεν εχουνε την ασθενεια και 712 ασθενεις που εχουν την ασθενεια.Το test data αποτελειται απο 306 ασθενεις που δεν εχουνε την ασθενεια και 306 ασθενεις που εχουν την ασθενεια.

Random Forest: [[234 72]
[68 238]]

Linear Regression: [[302   4]
[302   4]]

cros validation = {1: 0.53773584905660377, 2: 0.5, 3: 0.51886792452830188, 4:0.81603773584905659, 5: 0.51415094339622647, 6: 0.51415094339622647, 7:0.47169811320754718, 8: 0.45283018867924529, 9: 0.49528301886792453, 10:0.54166666666666663}

SVM: [[248 58]
[91 215]]
cros validation = {1: 0.81132075471698117, 2: 0.78773584905660377, 3:0.74528301886792447, 4: 0.79716981132075471, 5: 0.78301886792452835, 6:0.80188679245283023, 7:0.79716981132075471, 8: 0.76415094339622647, 9:0.79245283018867929, 10:0.74537037037037035}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[266 40]**
[64 242]]

cros validation = {1: 0.78301886792452835, 2: 0.88207547169811318, 3: 0.84433962264150941, 4: 0.83490566037735847, 5: 0.83018867924528306, 6: 0.83018867924528306, 7:0.83018867924528306, 8: 0.81603773584905659, 9: 0.84433962264150941, 10:0.83333333333333337}

LINEAR LOGISTIC REGRESSION:[[266 40]
[67 239]]

cros validation = {1: 0.84433962264150941, 2: 0.839622641509434, 3: 0.8632075471698113, 4: \0.84905660377358494, 5: 0.83490566037735847, 6: 0.80660377358490565, 7:0.839622641509434, 8: 0.85849056603773588, 9: 0.8867924528301887, 10:0.85648148148148151}

OLS: [[229 77]
[100 206]]
cros validation = {1: 0.80660377358490565, 2: 0.81603773584905659, 3: 0.81132075471698117, 4: 0.78773584905660377, 5: 0.77358490566037741, 6: 0.76886792452830188, 7:0.82075471698113212, 8: 0.75471698113207553, 9: 0.80660377358490565, 10:0.83333333333333337}

GaussianNB:[[260 46]
[132 174]]

cros validation = {1: 0.81132075471698117, 2: 0.84433962264150941, 3: 0.84905660377358494, 4: 0.79716981132075471, 5: 0.75, 6: 0.75471698113207553, 7: 0.84433962264150941, 8:0.81603773584905659, 9: 0.82075471698113212, 10: 0.83333333333333337}

<u>6ο Πειραμα:</u>
Σε αυτο το πειραμα κάναμε το train data αποτελειται απο 3662 ασθενεις που δεν εχουνε την ασθενεια και 818 ασθενεις που εχουν την ασθενεια.Το test data αποτελειται απο 300 ασθενεις που δεν εχουνε την ασθενεια και 200 ασθενεις που εχουν την ασθενεια.

Random Forest: [[262 38]
[30 170]]

Linear Regression: [[2 298]
[1 199]]

cros validation = {1: 0.91365461847389562, 2: 0.21887550200803213, 3:0.84136546184738958, 4: 0.19277108433734941, 5: 0.82128514056224899, 6: 0.86345381526104414, 7:0.79317269076305219, 8: 0.20481927710843373, 9: 0.23493975903614459, 10:0.21887550200803213}

SVM:[[300 0]
[200 0]]

cros validation = {1: 0.22690763052208834, 2: 0.18072289156626506, 3:0.22489959839357429, 4: 0.18674698795180722, 5: 0.22088353413654618, 6: 0.79317269076305219, 7:0.19277108433734941, 8: 0.18875502008032127, 9: 0.21887550200803213, 10:0.19477911646586346}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[293 7]**
[103 97]]
cros validation = {1: 0.93172690763052213, 2: 0.94979919678714864, 3:0.93574297188755018, 4: 0.92771084337349397, 5: 0.94779116465863456, 6: 0.94176706827309242, 7:0.94779116465863456, 8: 0.93775100401606426, 9: 0.94779116465863456, 10:0.94578313253012047}

LINEAR LOGISTIC REGRESSION:[[296 4]
[102 98]]

cros validation = {1: 0.93574297188755018, 2: 0.94779116465863456, 3: 0.9337349397590361, 4: 0.93574297188755018, 5: 0.94578313253012047, 6: 0.92971887550200805, 7:0.94377510040160639, 8: 0.94779116465863456, 9: 0.94979919678714864, 10:0.92771084337349397}

OLS: [[287 13]
[87 113]]

cros validation = {1: 0.8875502008032129, 2: 0.85742971887550201, 3: 0.81726907630522083, 4: 0.88152610441767065, 5: 0.85542168674698793, 6: 0.82730923694779113, 7:0.88152610441767065, 8: 0.87148594377510036, 9: 0.88353413654618473, 10:0.88955823293172687}

GaussianNB:[[286 14]
[61 139]]

cros validation = {1: 0.94779116465863456, 2: 0.94979919678714864, 3:0.95381526104417669, 4: 0.94979919678714864, 5: 0.96787148594377514, 6: 0.97590361445783136, 7:0.94979919678714864, 8: 0.93172690763052213, 9: 0.9618473895582329, 10:0.97389558232931728}

7ο Πειραμα:

Σε αυτο το πειραμα κάναμε το train data αποτελειται απο 1250 ασθενεις που δεν εχουνε την ασθενεια και 818 ασθενεις που εχουν την ασθενεια.Το test data αποτελειται απο 300 ασθενεις που δεν εχουνε την ασθενεια και 200 ασθενεις που εχουν την ασθενεια.

Random Forest: [[215 85]
[34 166]]

Linear Regression: [[1 299]
[2 198]]

cros validation = {1: 0.40625, 2: 0.6484375, 3: 0.3828125, 4: 0.5625, 5: 0.359375, 6: 0.4140625, 7: 0.83984375, 8: 0.62109375, 9: 0.3984375, 10: 0.5757575757575758}

SVM: [[264 36]
[76 124]]

cros validation = {1: 0.79296875, 2: 0.80859375, 3: 0.80859375, 4: 0.83984375, 5: 0.78515625, 6: 0.796875, 7: 0.80859375, 8: 0.77734375, 9: 0.82421875, 10: 0.82196969696969702}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[266 34]**
[46 154]]
cros validation = {1: 0.8828125, 2: 0.89453125, 3: 0.8671875, 4: 0.84765625, 5: 0.84765625, 6:0.828125, 7: 0.8828125, 8: 0.86328125, 9: 0.8515625, 10: 0.86742424242424243}

LINEAR LOGISTIC REGRESSION:[[266 34]
[44 156]]

cros validation = {1: 0.88671875, 2: 0.90625, 3: 0.875, 4: 0.90234375, 5: 0.89453125, 6:0.87890625, 7: 0.88671875, 8: 0.8515625, 9: 0.890625, 10: 0.84848484848484851}

OLS: [[229 71]
[42 158]]

cros validation = {1: 0.84765625, 2: 0.84765625, 3: 0.8125, 4: 0.85546875, 5: 0.8515625, 6:0.890625, 7: 0.8828125, 8: 0.83203125, 9: 0.8828125, 10: 0.85606060606060608}

GaussianNB:[[259 41]
[110 90]]

cros validation = {1: 0.82421875, 2: 0.78125, 3: 0.796875, 4: 0.82421875, 5: 0.796875, 6:0.83203125, 7: 0.859375, 8: 0.8359375, 9: 0.83984375, 10: 0.8257575757575758}

8ο Πειραμα:

Σε αυτο το πειραμα κάναμε το train data αποτελειται απο 1000 ασθενεις που δεν εχουνε την ασθενεια και 818 ασθενεις που εχουν την ασθενεια.Το test data αποτελειται απο 300 ασθενεις που δεν εχουνε την ασθενεια και 200 ασθενεις που εχουν την ασθενεια.

Random Forest: [[216 84]
[31 169]]

Linear Regression:[[296 4]
[196 4]]

cros validation = {1: 0.58008658008658009, 2: 0.53246753246753242, 3:0.53679653679653683, 4: 0.44155844155844154, 5: 0.59740259740259738, 6: 0.52380952380952384, 7:0.42424242424242425, 8: 0.4329004329004329, 9: 0.56277056277056281, 10:0.43933054393305437}

SVM: [[239 61]
[58 142]]

cros validation = {1: 0.80952380952380953, 2: 0.80519480519480524, 3:0.80519480519480524, 4: 0.79220779220779225, 5: 0.81818181818181823, 6: 0.78354978354978355, 7:0.77056277056277056, 8: 0.74025974025974028, 9: 0.83982683982683981, 10:0.77824267782426781}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[260 40]**
[36 164]]

cros validation = {1: 0.83549783549783552, 2: 0.8571428571428571, 3: 0.81385281385281383, 4: 0.8614718614718615, 5: 0.8528138528138528, 6: 0.81385281385281383, 7:0.83116883116883122, 8: 0.87012987012987009, 9: 0.83116883116883122, 10:0.84100418410041844}

LINEAR LOGISTIC REGRESSION:[[255 45]
[37 163]]

cros validation = {1: 0.86580086580086579, 2: 0.80519480519480524, 3:0.84848484848484851, 4: 0.89610389610389607, 5: 0.8528138528138528, 6: 0.8528138528138528, 7:0.83549783549783552, 8: 0.8571428571428571, 9: 0.82251082251082253, 10:0.85355648535564854}

OLS:[[221 79]
[45 155]]

cros validation = {1: 0.77922077922077926, 2: 0.77056277056277056, 3:0.82251082251082253, 4: 0.81385281385281383, 5: 0.8441558441558441, 6: 0.80952380952380953, 7:0.83549783549783552, 8: 0.83549783549783552, 9: 0.8571428571428571, 10:0.82426778242677823}

GaussianNB:[[257 43]
[83 117]]

cros validation = {1: 0.80519480519480524, 2: 0.87012987012987009, 3:0.83549783549783552, 4: 0.80952380952380953, 5: 0.80519480519480524, 6: 0.83116883116883122, 7:0.8528138528138528, 8: 0.80952380952380953, 9: 0.89177489177489178, 10:0.81589958158995812}

<u>9ο Πειραμα:</u>
Σε αυτο το πειραμα κάναμε το train data αποτελειται απο 1000 ασθενεις που δεν εχουνε την ασθενεια και 818 ασθενεις που εχουν την ασθενεια.Το test data αποτελειται απο 245 ασθενεις που δεν εχουνε την ασθενεια και 200 ασθενεις που εχουν την ασθενεια.

Random Forest: [[185 60]
[34 166]]

Linear Regression:[[242   3]
[196   4]]

cros validation = {1: 0.43805309734513276, 2: 0.53097345132743368, 3:0.56194690265486724, 4: 0.43805309734513276, 5: 0.54424778761061943, 6: 0.58407079646017701, 7:0.59734513274336287, 8: 0.44247787610619471, 9: 0.46017699115044247, 10:0.37991266375545851}

SVM:[[194 51]
[58 142]]

cros validation = {1: 0.77876106194690264, 2: 0.81415929203539827, 3:0.83185840707964598, 4: 0.82300884955752207, 5: 0.79203539823008851, 6: 0.77433628318584069, 7:0.72566371681415931, 8: 0.78761061946902655, 9: 0.80088495575221241, 10:0.78165938864628826}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[212 33]**
[36 164]]

cros validation = {1: 0.8584070796460177, 2: 0.87168141592920356, 3: 0.85398230088495575, 4: 0.82743362831858402, 5: 0.82300884955752207, 6: 0.87610619469026552, 7:0.82743362831858402, 8: 0.83185840707964598, 9: 0.83628318584070793, 10:0.81659388646288211}

LINEAR LOGISTIC REGRESSION:[[206 39]
[37 163]]

cros validation = {1: 0.84513274336283184, 2: 0.81415929203539827, 3:0.86283185840707965, 4: 0.83185840707964598, 5: 0.8584070796460177, 6: 0.88053097345132747, 7:0.85398230088495575, 8: 0.82743362831858402, 9: 0.87610619469026552, 10:0.8733624454148472}

OLS:[[187 58]
[50 150]]

cros validation = {1: 0.82743362831858402, 2: 0.82300884955752207, 3: 0.7831858407079646, 4: 0.83628318584070793, 5: 0.79646017699115046, 6: 0.80973451327433632, 7:0.81415929203539827, 8: 0.84070796460176989, 9: 0.83628318584070793, 10:0.79039301310043664}

GaussianNB:[[214 31]
[83 117]]

cros validation = {1: 0.84070796460176989, 2: 0.84513274336283184, 3:0.84070796460176989, 4: 0.82300884955752207, 5: 0.80973451327433632, 6: 0.84513274336283184, 7:0.82743362831858402, 8: 0.79646017699115046, 9: 0.86725663716814161, 10:0.79912663755458513}

<u>10ο Πειραμα:</u>
Σε αυτο το πειραμα κάναμε το train data αποτελειται απο 1000 ασθενεις που δεν εχουνε την ασθενεια και 818 ασθενεις που εχουν την ασθενεια.Το test data αποτελειται απο 306 ασθενεις που δεν εχουνε την ασθενεια και 200 ασθενεις που εχουν την ασθενεια.

Random Forest: [[219 87]
[30 170]]

Linear Regression:[[302   4]
[196   4]]

cros validation = {1: 0.38793103448275862, 2: 0.45689655172413796, 3:0.43965517241379309, 4: 0.84051724137931039, 5: 0.57327586206896552, 6: 0.43534482758620691, 7:0.53879310344827591, 8: 0.56034482758620685, 9: 0.57758620689655171, 10:0.46610169491525422}

SVM:[[244 62]
[58 142]]

cros validation = {1: 0.79741379310344829, 2: 0.77155172413793105, 3:0.77155172413793105, 4: 0.78017241379310343, 5: 0.81034482758620685, 6: 0.75862068965517238, 7:0.81034482758620685, 8: 0.81034482758620685, 9: 0.81034482758620685, 10:0.79661016949152541}

SVM εδω περα δωσαμε την παραμετρο kernel =’linear’ **:[[266 40]**
[36 164]]

cros validation = {1: 0.82758620689655171, 2: 0.83620689655172409, 3:0.81465517241379315, 4: 0.85344827586206895, 5: 0.88362068965517238, 6: 0.81034482758620685, 7:0.84051724137931039, 8: 0.83620689655172409, 9: 0.82758620689655171, 10:0.85169491525423724}

LINEAR LOGISTIC REGRESSION:[[261 45]
[37 163]]

cros validation = {1: 0.84913793103448276, 2: 0.84482758620689657, 3:0.89655172413793105, 4: 0.85775862068965514, 5: 0.86206896551724133, 6: 0.89224137931034486, 7:0.88362068965517238, 8: 0.86637931034482762, 9: 0.84913793103448276, 10:0.83050847457627119}

OLS:[[[225 81]
[45 155]]

cros validation = {1: 0.81465517241379315, 2: 0.80172413793103448, 3:0.83620689655172409, 4: 0.82758620689655171, 5: 0.84913793103448276, 6: 0.82327586206896552, 7:0.88793103448275867, 8: 0.81034482758620685, 9: 0.81034482758620685, 10:0.83898305084745761}

GaussianNB: [[263 43]
[83 117]]

cros validation = {1: 0.85344827586206895, 2: 0.85775862068965514, 3:0.80603448275862066, 4: 0.84482758620689657, 5: 0.84913793103448276, 6: 0.79741379310344829, 7:0.85775862068965514, 8: 0.80172413793103448, 9: 0.81896551724137934, 10:0.8347457627118644}