



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Σχολή Θετικών Επιστημών

Τμήμα Υπολογιστικής Φυσικής

Υπολογιστική Κβαντική Φυσική και Εφαρμογές

**Εφαρμογή μεθόδων μηχανικής μάθησης για
διαχωρισμό σήματος και υποβάθρου σε εξωτικό
σενάριο Higgs**

Κουφετίδης Γιώργος

AEM:4432

Οι αλγόριθμοι που χρησιμοποιήθηκαν είναι ο Knn (K- Nearest Neighbors) και Random Forest.

Για τον Knn και Random Forest αλγόριθμο έγινε διαχωρισμός των δεδομένων σε high και low level quantities. Για την κάθε κατηγορία δεδομένων διαχωρίζονται σε δεδομένα που είναι γνωστό το αποτέλεσμα και στην συνέχεια χρησιμοποιώντας τα υπόλοιπα δεδομένα διαμορφώνεται το μοντέλο που σε συνέχεια χρησιμοποιείται για έλεγχο.

Αποτελέσματα του Knn:

Low quantities

Accuracy = 56.6%

	precision	recall	f1-score	support
0.0	0.58	0.45	0.51	992
1.0	0.56	0.68	0.61	1008
accuracy			0.57	2000
macro avg	0.57	0.57	0.56	2000
weighted avg	0.57	0.57	0.56	2000

High quantities

Accuracy = 65.7%

	precision	recall	f1-score	support
0.0	0.67	0.60	0.64	992
1.0	0.65	0.71	0.68	1008
accuracy			0.66	2000
macro avg	0.66	0.66	0.66	2000
weighted avg	0.66	0.66	0.66	2000

High and Low quantities

Accuracy = 60.4%

	precision	recall	f1-score	support
0.0	0.62	0.51	0.56	992
1.0	0.59	0.70	0.64	1008
accuracy			0.60	2000
macro avg	0.61	0.60	0.60	2000
weighted avg	0.61	0.60	0.60	2000

Αποτελέσματα Random Forest:

Accuracy = 58.4%

	precision	recall	f1-score	support
0.0	0.61	0.45	0.52	992
1.0	0.57	0.71	0.63	1008
accuracy			0.58	2000
macro avg	0.59	0.58	0.58	2000
weighted avg	0.59	0.58	0.58	2000

High quantities

Accuracy = 68.5%

	precision	recall	f1-score	support
0.0	0.61	0.45	0.52	992
1.0	0.57	0.71	0.63	1008
accuracy			0.58	2000
macro avg	0.59	0.58	0.58	2000
weighted avg	0.59	0.58	0.58	2000

High and Low quantities

Accuracy = 70.3%

	precision	recall	f1-score	support
0.0	0.61	0.45	0.52	992
1.0	0.57	0.71	0.63	1008
accuracy			0.58	2000
macro avg	0.59	0.58	0.58	2000
weighted avg	0.59	0.58	0.58	2000

Αποτελέσματα ANN:

Low quantities

Accuracy= 56.6%

High quantities

Accuracy=69.5%

Συμπεράσματα

Για κάθε αλγόριθμο παρατηρείται μεγαλύτερη ακρίβεια των high quantities δεδομένων σε αντίθεση με τα δεδομένα low quantities. Κάτι τέτοιο έχει ως συνέπεια την μεγαλύτερη συσχέτιση των μεταβλητών αυτών με την ταξινόμηση των συμβάντων σήματος ή υποβάθρου. Στον Knn αλγόριθμο παρατηρήθηκε ακρίβεια 56.6 για low quantities, 65.7 % για high quantities και για όλα 60.4% Ο αλγόριθμος του Random Forest παρουσίασε καλύτερη ακρίβεια με low quantities στα 58.4%, high quantities 68.5% και όλα μαζί 70.3%. Υπολογίζοντας την ακρίβεια με την βοήθεια νευρωνικών δικτύων επιτευχθεί ακρίβεια για low quantities 56.6% και αντίστοιχα για high quantities 69.4% και για όλα 62.7%.