

# Μηχανική Μάθηση: Μέθοδοι και Αλγόριθμοι

Μεταπτυχιακή απαλλακτική εργασία ΦΕΒ. 2023-24

Καθηγητής:  
Η. Μαγκλογιάννης

Καμπάσης Παναγιώτης  
Πανεπιστήμιο Πειραιώς  
Τμήμα Ψηφιακών Συστημάτων  
Μεγάλα Δεδομένα και Αναλυτική  
8 Μαρτίου 2024

# Μηχανική Μάθηση: Μέθοδοι και Αλγόριθμοι

Πάνος Καμπάσης  
27/01/2024

## Περιεχόμενα

<b>I</b>	Εισαγωγή	2
I-A'	Η βάση δεδομένων . . . . .	2
I-B'	<i>features</i> . . . . .	2
<b>II</b>	Στατιστικά	3
II-A'	Κατανομή των χαρακτηριστικών . . . . .	3
II-B'	Σύγκριση Μοντέλων και Αποτελέσματα . . . . .	3
II-Γ'	Βήματα . . . . .	3
<b>III</b>	Παλινδρομητές	4
III-A'	Πολυωνυμική Παλινδρόμηση . . . . .	4
<b>IV</b>	Ταξινομητές	4
IV-A'	<i>Random Forest Classifier</i> . . . . .	4
<b>V</b>	Επιστροφή στην παλινδρόμηση	4
V-A'	<i>Random Forest, Linear and Gradient Boosting Regression</i> . . . . .	4
<b>VI</b>	Συσταδοποίηση	6
VI-A'	Πίνακας συσχέτισης . . . . .	6
VI-B'	<i>K – means</i> . . . . .	6
VI-Γ'	Χαρακτηριστικά που συνδέονται περισσότερο με το <i>area</i> . . . . .	6
VI-Δ'	Αναδρομή . . . . .	7
	Αναφορές	8

Περίληψη—Στο παρόν θα εξετάσουμε την σχέση μεταξύ διαφόρων δεικτών του μετεωρολογικού δείκτη πυρκαγιάς και δοσμένης έκτασης επιρρεπή σε πυρκαγιά. Θα επιχειρήσουμε να εξάγουμε αποτελέσματα από τεχνικές μηχανικής μάθησης με σκοπό να εξετάσουμε πόσο αποδοτικό είναι το εκάστοτε μοντέλο στην πρόβλεψη ανάδευσης πυρκαγιάς.

Λέξεις κλειδιά:

Δασική Πυρκαγιά, Εκτίμηση Δασικής Έκτασης, Μηχανική Μάθηση, Δένδρα απόφασης, Βαθιά Μάθηση, *Random Forest*, *Gradient Boost*, Παλινδρόμηση, Συσταδοποίηση.

## I. Εισαγωγή

Τα δεδομένα μας έχουν δεχτεί ήδη μια προετοιμασία και ως εκ τούτου δεν εμπεριέχουν κενές τιμές ή ιδιαίτερους θορύβους. Έτσι μπορούμε να εξετάσουμε απ' ευθείας τα διαφορετικά χαρακτηριστικά που προσδίδουν διαφορετικοί αλγόριθμοι μηχανικής μάθησης για να συμπεράνουμε ποιος μας δίνει τις καλύτερες συνθήκες για την πιο ακριβή κατηγοριοποίηση.

A'. Η βάση δεδομένων

*UC Irvine : Machine Learning Repository*

Η βάση δεδομένων ονομάζεται *Forest Fires* και αφορά μετεωρολογικές μετρήσεις από την βορειο-ανατολική περιοχή της Πορτογαλίας (*Cortez and Morais, 2007*). Όπως και οι προ - επιμελητές της για το χαρακτηριστικό 'area' που περιγράφει την έκταση της πυρκαγιάς στο *Montesinho park* επιχειρούμε να διορθώσουμε την έντονη λοξότητα του στο 0 εφαρμόζοντας λογαριθμοποίηση της με την  $\log(x + 1)$

B'. *features*

Τα χαρακτηριστικά που διαθέτουμε περιγράφονται παρακάτω:

- "X": Συντεταγμένες του πάρκου, έχει τιμές από 1 έως 9.
- "Y": Συντεταγμένες του πάρκου, έχει τιμές από 1 έως 9.
- "FFMC" (*Fine Fuel Moisture Code*) : Αντιπροσωπεύει την υγρασία καυσίμου από απορρίματα και διάφορα άλλα στοιχεία. Αποτυπώνει έτσι μια μορφή δείκτη υγρασίας ανάλογο μιας απόκλισης χρόνου 16 ωρών. Πεδίο τιμών 0-101. Παρατηρείται μια λοξότητα μικρής τάξης στα αριστερά ενώ υπάρχει συγκέντρωση γύρω από τις τιμές 90-95.

- "DMC" (*Duff Moisture Code*) : Δείκτης υγρασίας. Παίρνει αριθμητικές τιμές από την μέση τιμή της υγρασίας από ασθενώς συμπαγής δεσμούς από οργανικά στρώματα στο υπέδαφος. Χρησιμεύει στην αποτύπωση της ξηρασίας. Παρατηρούμε λοξότητα στα δεξιά και μακριά ουρά κάτι που υποδηλώνει μεγάλο εύρος υγρασίας στα οργανικά υποστρώματα.
- "DC" (*Drought Code*): Δείκτης ξηρασίας με αριθμητικές τιμές. Παρουσιάζει ομοιότητα με το παραπάνω αλλά περιγράφει την υγρασία στα βαθιά και συμπαγή οργανικά στρώματα στο υπέδαφος. Επίσης λοξό προς τα δεξιά που σημαίνει ότι έχουμε έντονη διακύμανση μεταξύ της βάσης δεδομένων μας με κάποιους μήνες να έχουν μεγαλύτερα ποσοστά υγρασίας από άλλους.
- "ISI" (*Initial Spread Index*) : Δείκτης εξάπλωσης της πυρκαγιάς. Επίσης παρατηρούμε λοξότητα στα δεξιά με βαριά ουρά επομένως κάποιες περιοχές έχουν σημαντική διαφορά στο ρυθμό εξάπλωσης.
- "temp" (*temperature*) : Δείκτης θερμοκρασίας σε βαθμούς Κελσίου. Δέχεται τιμές από 15 - 25. Εμφανίζει κανονική κατανομή.
- "RH" (*Relative Humidity*): Σχετική υγρασία. Η λοξότητα προς τα αριστερά συνιστά πως τα χαμηλά επίπεδα υγρασίας εμφανίζονται συχνότερα στη βάση δεδομένων μας.
- "wind" : Μέτρηση του αέρα σε *km/h*. Οι περισσότερες τιμές είναι μεταξύ 2-6.
- "rain" : Ισχυρή λοξότητα στα δεξιά που σημαίνει ότι στις περισσότερες περιπτώσεις είχαν λίγα περιστατικά βροχής. Κάποια στιγμιότυπα ωστόσο παρουσιάστηκε έντονη βροχόπτωση.
- "area" : Είναι η μεταβλητή στόχος. Δέχεται τιμές από 0.0 έως 1090. Αποτυπώνει έκταση σε εκτάρια.

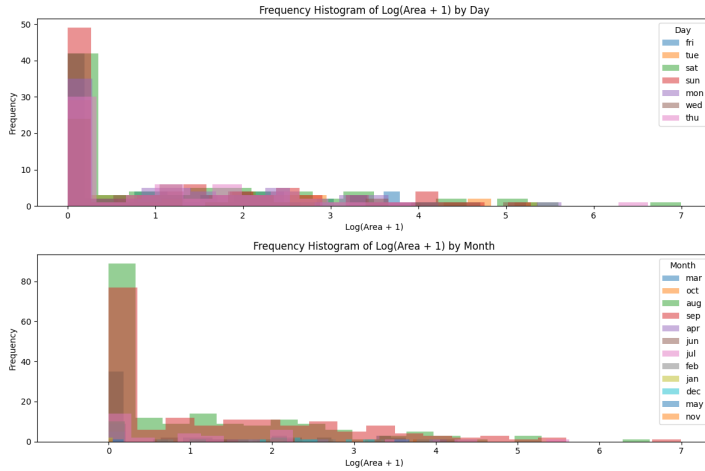
X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0.0	0.0
7	4	oct	tue	90.6	35.4	669.1	6.7	18.0	33	0.9	0.0	0.0
7	4	oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0.0	0.0
8	6	mar	fri	91.7	33.3	77.5	9.0	8.3	97	4.0	0.2	0.0
8	6	mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0.0	0.0

Η Δομή της Βάσης Δεδομένων μας. Έχουμε συνδιασμό από κατηγορικές και αριθμητικές τιμές.

## II. Στατιστικά

### Α'. Κατανομή των χαρακτηριστικών

Παραθέτουμε κάποια στατιστικά για τα δεδομένα μας. Στα γραφήματα των χαρακτηριστικών, μας ενδιαφέρει εάν εμφανίζεται κάποια συσχέτιση στις περιοχές της πυρκαγιάς με τις μέρες και μήνες του χρόνου. Εκ πρώτης όψης δεν φαίνεται να υπάρχει σχέση μεταξύ των κατηγορικών "day", "month" με το χαρακτηριστικό στόχο "area" ωστόσο παρατηρούμε κάποια συσχέτιση με τις ημέρες, τα ιστογράμματα για τις διαφορετικές μέρες φαίνεται να έχουν κοινές κατανομές και σχήμα.



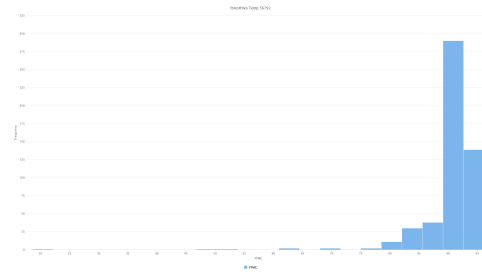
Σχήμα 2: Σχέση του χαρακτηριστικού "area" με "day", "month"

### Β'. Σύγκριση Μοντέλων και Αποτελέσματα

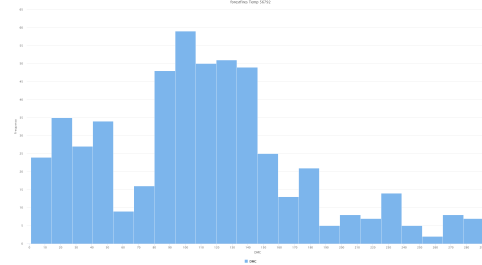
Αναλύουμε τις μεθόδους που ακολουθήσαμε για την εκπαίδευση και εκμάθηση των μοντέλων μας καθώς και τα αποτελέσματα του κάθε μοντέλου. Ξεκινάμε από απλές στατιστικές αναλύσεις και καταλήγουμε σε μεθόδους μηχανικής μάθησης. Κύρια εργαλεία στην ανάλυση μας: *python & RapidMiner*.

### Γ'. Βήματα

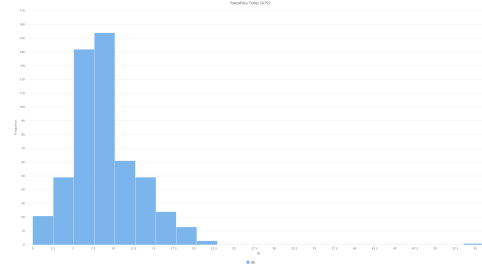
- Ως κύρια μετρική σύγκρισης παρατηρούμε το μέσο τετραγωνικό σφάλμα ( $MSE, R^2$ )
- Οι μέθοδοι μηχανικής μάθησης που θα εξετάσουμε είναι οι Γραμμική και Πολυωνυμική Παλινδρόμηση *SVR (Support Vector Reggresion)*, *Random Forest*, *GradientBoost* και έπειτα *K – means*, *t – SNE*.
- Η αριθμητική φύση του χαρακτηριστικού "area" μας ωθεί να δοκιμάσουμε διάφορα μοντέλα παλινδρόμησης. Δοκιμάζουμε το απλό πολυωνυμικό μοντέλο, *Gradient Boosting Regressor* και *Random Forest Regression*
- Παρατηρώντας το εύρος των τιμών του χαρακτηριστικού στόχου παρατηρούμε ότι υπάρχει μια έντονη διαφορά στις τιμές. Αυτό φαίνεται ευκολότερα στο αντίστοιχο γράφημα (Σχήμα 2). Είναι εύλογο να



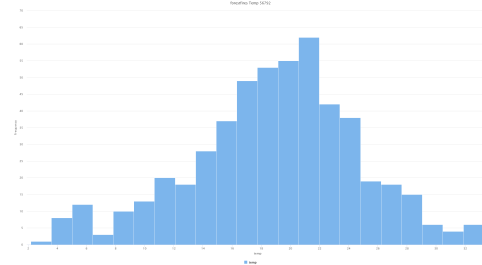
(α') *FFMC*



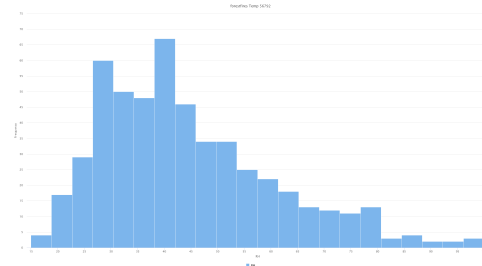
(β') *DMC*



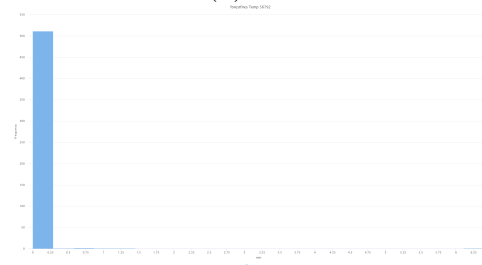
(γ') *ISI*



(δ') *Temp*



(ε') *RH*



(ς') *Rain*

να είναι δύσκολη η πρόβλεψη για πραγματικούς αριθμούς. Εάν πράγματι δεν πετύχουμε τα επιθυμητά αποτελέσματα με τις μεθόδους παλινδρόμησης θα χωρίσουμε το χαρακτηριστικό "area" σε κατηγορίες και θα πραγματοποιήσουμε ταξινομήσεις.

### III. Παλινδρόμητες

#### A'. Πολυωνυμική Παλινδρόμηση

Ξεκινάμε δοκιμάζοντας την Πολυωνυμική Παλινδρόμηση. Ο λόγος που αποφεύγουμε το απλό γραμμικό μοντέλο απ' ευθείας είναι επειδή στη βάση δεδομένων εμφανίζεται πολλές φορές το 0, σημαντικά λιγότερες αριθμοί κοντά στο διάστημα 1-7 και ακόμη λιγότεροι έχουν ακραίες τιμές της τάξης του 90. Κάτι τέτοιο είναι σημαντικά δύσκολο να προβλεφθεί από την ευθεία της γραμμικής παλινδρόμησης. Τα αποτελέσματα που δίνει η Πολυωνυμική Παλινδρόμηση είναι τα εξής:

*Mean Squared Error (Polynomial Regression) : 13757.87*

$R^2 - \text{Score(Polynomial Regression)} : -0.17$

Η αρνητική τιμή στο συντελεστή συσχέτισης  $R^2$  μας δείχνει αμέσως ότι το μοντέλο μας έχει εκπαιδευτεί με λάθος τρόπο καθώς το δυνατό εύρος έχει τιμές από 0 έως 1. Γνωρίζουμε εξάλλου και από τα παραπάνω γραφήματα πως η λοξότητα των δεδομένων (ειδικά του χαρακτηριστικού-στόχο area) είναι ιδιαίτερα σημαντική και έτσι αν δεν βρούμε το καταλληλό μοντέλο είναι εύλογο να μην μπορεί να επιφέρει ακριβή αποτελέσματα.

### IV. Ταξινομητές

Εφόσον οι Παλινδρομήσεις δεν επέφεραν σημαντικά αποτελέσματα δοκιμάζουμε αν διαχωρίσουμε το χαρακτηριστικό στόχο σε 3 κατηγορίες:

$low \in (-\infty, 5], \text{ medium} \in (5, 10], \text{ high} \in (10, +\infty)$

αυτό θα μας επιτρέψει να χρησιμοποιήσουμε τον εξής ταξινομητή:

#### A'. Random Forest Classifier

Το μοντέλο αυτό σημειώνει σημαντικά καλύτερα αποτελέσματα από τους παλινδρομητές. Παραθέτουμε μερικές μετρικές:

*Accuracy of RandomForestClassifier : 0.71*

	<i>precision</i>	<i>recall</i>	<i>f1 - score</i>	<i>support</i>
<i>high</i>	0.50	0.10	0.16	21
<i>low</i>	0.73	0.97	0.83	74
<i>medium</i>	0.00	0.00	0.00	9
<i>accuracy</i>			0.71	104
<i>m.avg</i>	0.41	0.36	0.33	104
<i>w.avg</i>	0.62	0.71	0.62	104

Επομένως θα προσθέσουμε μερικούς ακόμα κατηγοριοποιητές και θα συγκρίνουμε τα αποτελέσματα τους.

### V. Επιστροφή στην παλινδρόμηση

Έχοντας τα παραπάνω ως οδηγό επιχειρούμε να βελτιώσουμε τους παλινδρομητές με τον εξής τρόπο: Εφόσον οι πρώτες παλινδρομήσεις μας δίνουν κακό *fit* του μοντέλου γνωρίζουμε ότι η μη κανονικότητα της βάσης μας κατά πάσα πιθανότητα ευθύνεται για αυτό το γεγονός. Προς έρευνας αυτού εξετάζουμε το *Infogain* των χαρακτηριστικών μας. Αυτό προέκυψε από τα καλά αποτελέσματα του *Random Forest* που χρησιμοποιεί *Infogain* για τη δημιουργία των δέντρων απόφασης. Έτσι έχουμε τον πίνακα:

<i>Feature</i>	<i>InformationGain</i>
<i>DC</i>	4.027724879894656
<i>DMC</i>	3.9949672455241574
<i>temp</i>	3.9463476908780755
<i>ISI</i>	3.380162144839595
<i>FFMC</i>	3.1761051682235064
<i>RH</i>	2.9696910128037732
<i>wind</i>	2.017453378139807
<i>X</i>	1.5251384010655311
<i>Y</i>	1.0712557687478865
<i>month_sep</i>	0.47129042246419406
<i>month_aug</i>	0.46311900065577394
<i>day_sun</i>	0.333234739813058
<i>day_sat</i>	0.3051304180164003

Καλύτερα χαρακτηριστικά σύμφωνα με το *Infogain*:

$[DC, DMC, temp, ISI, FFMC, RH, wind,$

$X, Y, month\_sep, month\_aug, day\_sun, day\_sat]$

#### A'. Random Forest, Linear and Gradient Boosting Regression

Δοκιμάζουμε τα αποτελέσματα των παραπάνω μοντέλων μηχανικής μάθησης αφαιρώντας χαρακτηριστικά από την παραπάνω λίστα, προσμετρούμε επιδόσεις και συζητάμε τα αποτελέσματα:

Όλα τα χαρακτηριστικά

	<i>Random Forest</i>	<i>Linear</i>	<i>Gradient Boost</i>
$R^2 \text{ Score}$	-0.05	-0.01	-0.07
<i>MSE</i>	2.31	2.21	2.36

Αφαιρώντας τα τελευταία 4 χαρακτηριστικά, μέρες και μήνες

	<i>Random Forest</i>	<i>Linear</i>	<i>Gradient Boost</i>
$R^2 \text{ Score}$	-0.03	0.01	-0.11
<i>MSE</i>	2.27	2.191	2.44

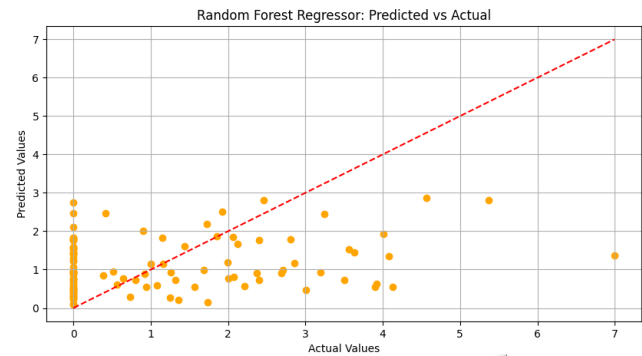
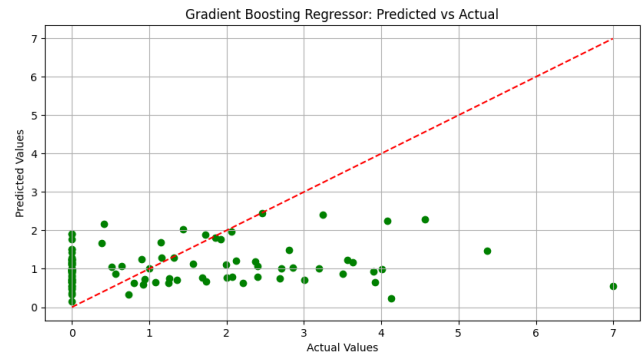
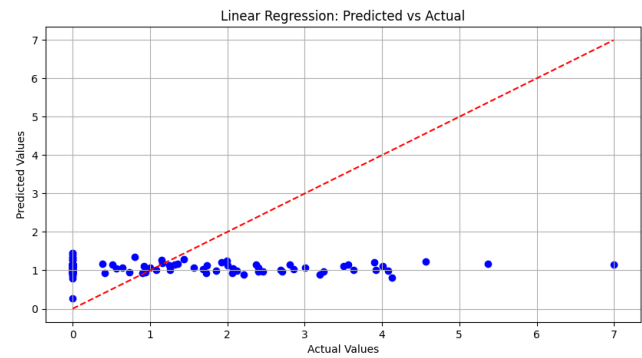
Τα αποτελέσματα δεν είναι θετικά και αυτό φαίνεται όπως και πριν από το  $R^2$ . Όμως διαθέτουμε τα χαρακτηριστικά που περιγράφουν καλύτερα τη βάση επομένως συνεχίζουμε την αφαίρεση.

Τα 4 πρώτα χαρακτηριστικά

	<i>Random Forest</i>	<i>Linear</i>	<i>Gradient Boost</i>
$R^2$ Score	0.02	-0.01	0.00
MSE	2.16	2.23	2.19

Επομένως με αυτά τα χαρακτηριστικά πετυχαίνουμε τα καλύτερα αποτελέσματα όσον αφορά το μέσο τετραγωνικό σφάλμα το οποίο είναι πολύ μικρό. Όπως φαίνεται στο  $R^2$  δεν εξηγείται καθόλου το ποσοστό της ολικής μεταβλητότητας και έτσι χάνουμε από την συνολική διακύμανση. Με άλλα λόγια τα μοντέλα αυτά σχεδόν αγνοούν τις πολύ μεγάλες φωτιές. Γεγονός εν μέρη λογικό από την άποψη ότι τα δεδομένα μας περιέχουν πολύ λίγες περιπτώσεις με μεγάλες πυρκαγές. Ωστόσο αξίζει να σημειώσουμε ολοκληρώνοντας την ανάλυση μας για τους παλινδρομητές ότι δεν εξηγούν καλά τα δεδομένα μας. Πετυχαίνουν πολύ κοντινές προβλέψεις τοπικά δηλαδή στις μικρές πυρκαγές και χάνουν εντελώς τις περιπτώσεις που βρίσκονται δεξιότερα στην ουρά της κατανομής (τις λίγες περιπτώσεις πολύ μεγάλων πυρκαγιών).

Ολοκληρώνουμε με τα πιο αναλυτικά αποτελέσματα του κάθε παλινδρομητή με *cross-validation* (σαφώς το *cross-validation* δεν θα βελτιώσει την συμπερίληψη των μεγάλων πυρκαγιών, αφού κάνοντας το *re-sampling* σε δεδομένα που ήδη έχουν λίγο πλήθος στις μεγάλους τιμές εν τέλει συλλογικά πέφτει ακόμα περισσότερο η διακύμανση που διατηρείται):



Στο σημείο αυτό αντί να περάσουμε σε *re-enforcement training* θα δώσουμε έμφαση σε άλλο τρόπο συμπερίληψης των ακραίων τιμών. Αντί λοιπόν να ενισχύσουμε την ικανότητα των μοντέλων να αντιλαμβάνεται τις λίγες ακραίες τιμές περνάμε σε οπτικές αναπαραστάσεις των ομάδων αυτών.

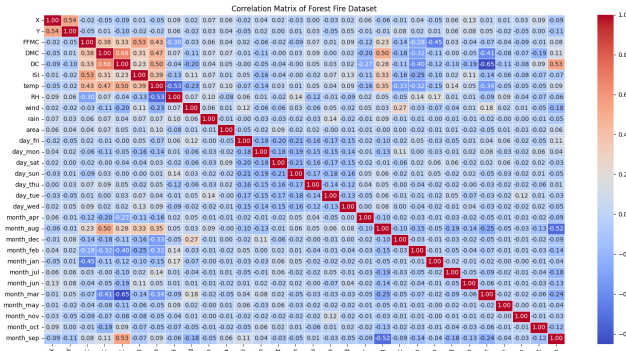
	<i>LinearRegression</i>	<i>GradientBoosting</i>	<i>RandomForest</i>
MSE	1.91	2.26	2.31
MAE	1.15	1.20	1.21
ExplainedVariance	-0.01	-0.19	-0.21
MedianAbsoluteError	1.05	0.99	1.06

## VI. Συσταδοποίηση

Για να έχουμε μια πιο παραστατική εικόνα του αντικειμένου θα δημιουργήσουμε μια συσταδοποίηση για να μπορέσουμε να δούμε τα πιο σημαντικά χαρακτηριστικά και το πόσο καλά μπορούμε να διαχωρίσουμε τα δεδομένα σε ομάδες.

### Α'. Πίνακας συσχέτισης

Ο παρακάτω αποτελεί τον πίνακα συσχέτισης της βάσης μας. Προσθήσαμε και τις μέρες και μήνες με *one – hot encoding* καθώς υπάρχουν σημαντικές συσχετίσεις.

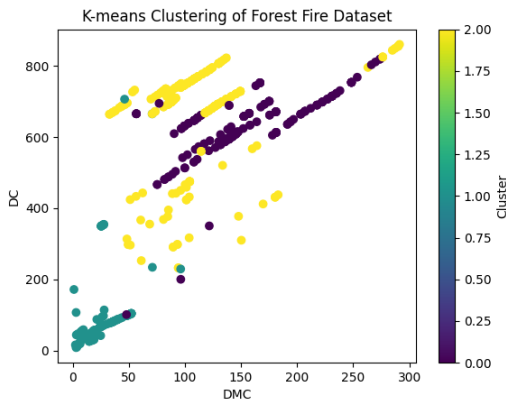


Σχήμα 3: *Correlation Matrix*

Έτσι επιλέγουμε τα δύο χαρακτηριστικά που έχουν την υψηλότερη συσχέτιση μεταξύ τους *DMC*, *DC*.

### Β'. *K – means*

Εφαρμόζοντας τον αλγόριθμο έχουμε:



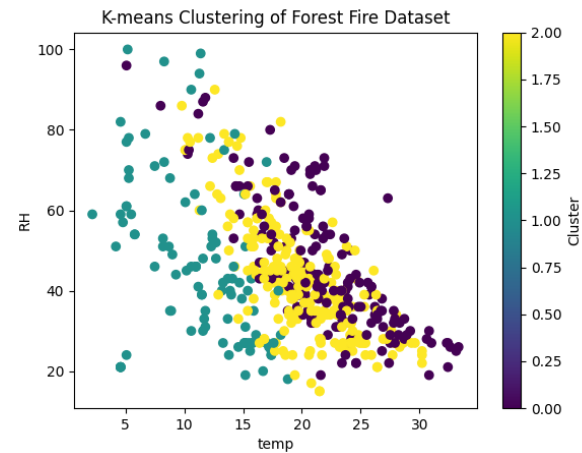
Σχήμα 4: *Clustering με DMC, DC*

Εφόσον τα χαρακτηριστικά μας είναι δείκτες ξηρασίας παίρνουμε μια αρκετά διαυγή εικόνα από 3 ομάδες. Κοντά στο 0 είναι οι μικρές φωτιές και όσο απομακρυνόμαστε βρίσκουμε τις μεγαλύτερες που είναι λιγότερες σε πλήθος. Από αυτό το γράφημα έχουμε μια αναπαράσταση από 2 μόνο χαρακτηριστικά της βάσης που διαχωρίζουν σε ένα βαθμό την ένταση της φωτιάς που θα έχουμε. Σε νέα

δεδομένα, μελετώντας μόνο το *DMC*, *DC* θα μπορούσαμε να έχουμε μια ιδέα για το τι έντασης φωτιά θα μπορούσαμε να έχουμε.

Γ'. Χαρακτηριστικά που συνδέονται περισσότερο με το *area*

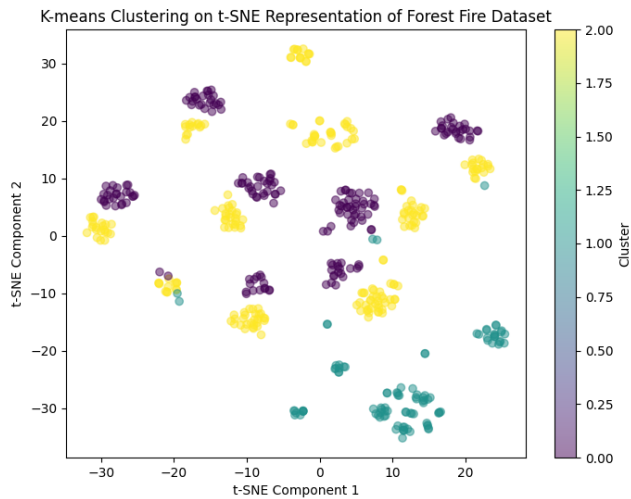
Κοιτώντας τον πίνακα συσχέτισης βλέπουμε την μεγαλύτερη κατά απόλυση τιμή συσχέτιση των χαρακτηριστικών με τον *area* αυτά είναι το *temp* και η ημέρα Σάββατο. Δεν θα προχωρήσουμε σε οπτικοποίηση όμως καθώς το χαρακτηριστικό *day\_sat* έχει τιμές 0 και 1 επομένως θα μας δώσει πρακτικά 2 γραμμές και δεν θα βοηθήσει στην δημιουργία συστάδων. Επιλέγουμε το αμέσως επόμενο μεγαλύτερο κατά απόλυτη τιμή που είναι το *RH*. Αρνητικά συσχετισμένο με την *area* σαφώς αφού είναι δείκτης υγρασίας. Έτσι έχουμε:



Σχήμα 5: *Clustering με temp, RH*

Βλέπουμε και πάλι μια καλή διαφοροποίηση και συμπεραίνουμε ότι η λωρίδα κοντά στο 0 χρώματος μπλε είναι η κατηγορία των μικρών πυρκαγιών και προχωρώντας στο κίτρινο θα είναι οι μεσαίες φωτιές και μωβ οι μεγάλες (καθώς ανέβηκε θερμοκρασία και έπεσε η υγρασία). Έτσι έχουμε μια καλή πρώτη εικόνα των ομάδων.

Για να κάνουμε την παραπάνω εικόνα πιο βοηθητική και να διαχωρίζει ο αλγόριθμος *K – means* καλύτερα τα δεδομένα εφαρμόζουμε μέθοδο μείωσης διαστάσεων *t – SNE* ώστε να πάρουμε πλέον 2 νέες συνιστώσες που θα διαχωρίζουν τα δεδομένα καλύτερα με χρήση πιθανοτήτων. Αυτό κάνει την εικόνα σαφώς καλύτερη αλλά πλέον για την διεξαγωγή συμπερασμάτων από νέα δεδομένα θα χρειαστούμε κάθε χαρακτηριστικό, όχι μόνο 2.



Σχήμα 6: *Clustering*

Κρίνοντας από τα πρώτο γραφήματα ξέρουμε ότι η συστάδα με μπλε χρώμα θα είναι η μικρή φωτιά, η κίτρινη η μεσαία και η μωβ η μεγάλη. Αυτά σαφώς δεν είναι ρεαλιστικές αποδόσεις απλώς σχετίζονται με τα χαρακτηριστικά που είναι κατά μεγαλύτερο τρόπο συσχετισμένα μεταξύ τους αλλά και με το χαρακτηριστικό που μας ενδιαφέρει. Άρα αποτελούν αρκετά καλό τρόπο να χαρτογραφήσουμε την μορφή της βάσης μας.

Δ'. Αναδρομή

Επομένως οι παραπάνω απεικονίσεις μας δίνουν μια αρκετά σαφή εικόνα για την βάση μας χωρίς να θυσιάζουν την ικανότητα τους για γενίκευση. Επομένως σε νέα δεδομένα ελέγχουμε την θέση τους σύμφωνα με τα χαρακτηριστικά  $DMC, DC$  και έπειτα με  $temp, RH$ . Εάν έχουμε μεγάλη προσθήκη από νέα δεδομένα, πραγματοποιούμε και πάλι  $t-SNE$  και παρατηρούμε σε ποια συστάδα προσπίπτουν για να εξάγουμε συμπέρασμα.



## Αναφορές

- [1] Ramez Alkhatib, Wahib Sahwan, Anas Alkhatieb and Brigitta Schtt, ‘A Brief Review of Machine Learning Algorithms in Forest Fires Science ’ (Published: 17 July 2023)
- [2] A review of machine learning applications in wildfire science and management: Piyush Jain [piyush.jain@canada.ca](mailto:piyush.jain@canada.ca), Sean C.P. Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D. Flannigan, (28 July 2020)
- [3] Alexandrov, D., Pertseva, E., Berman, I., Pantiukhin, I., and Kapitonov, A. 2019. Analysis of machine learning methods for wildfire security monitoring with an unmanned aerial vehicles. In *Conference of Open Innovation Association (FRUCT)*, 8–12 April 2019. *IEEE Computer Society*. pp. 3–9.
- [4] *Predicting the area burned by forest fires with weather data*, Manuele Macchia, September 2021