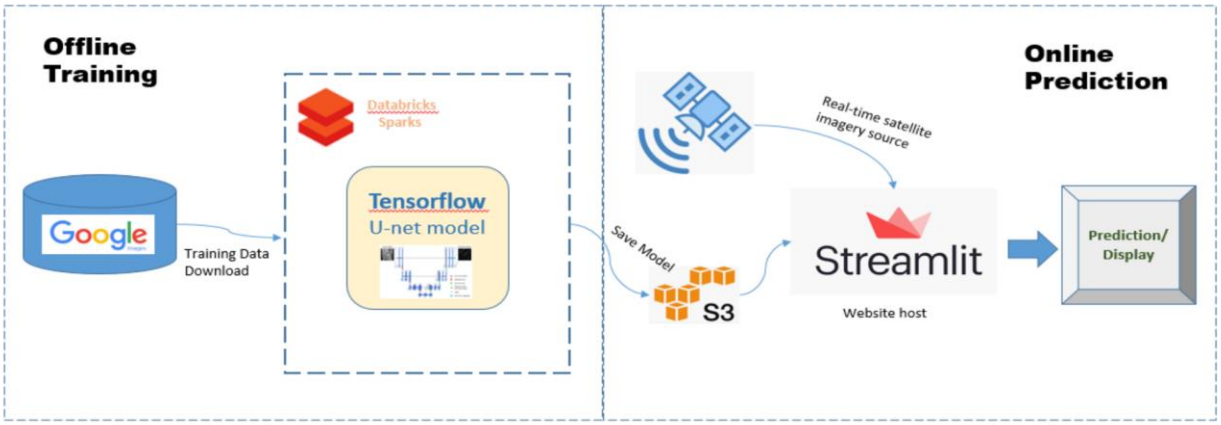
Forest Fire Prediction - Assignment 01

Paper 1: “Forest Fire Prediction Using Image Processing And Machine Learning” - Mohana Kumar S, Sowmya B J, Pryanka S, Ruchita Sharma, Shivank Tej, Spoorthi Ashok Karani. Nat. Volatiles & Essent. Oils, 2021; 8(4): 13116-13134

Set de date folosit: Datele de intrare sunt in format de imagini satelitare cu paduri, autorii nu mentioneaza direct un set de date, doar spun ca folosesc “Google Images API” pentru a face rost de imagini, ceea ce ne face sa intelegem setul de date a fost creat manual. Autorii nu ne ofera setul de date in vreun format, deci setul de date folosit de acestia nu este disponibil, ca date de iesire avem masca de segmentare unde este posibil sa existe foc. Modelul mai poate prezice si dimensiunea in km^2 si nivelul total de dioxid de carbon emis, dar acest lucru posibil sa fie scos din masca de segmentare (nu este menționat).

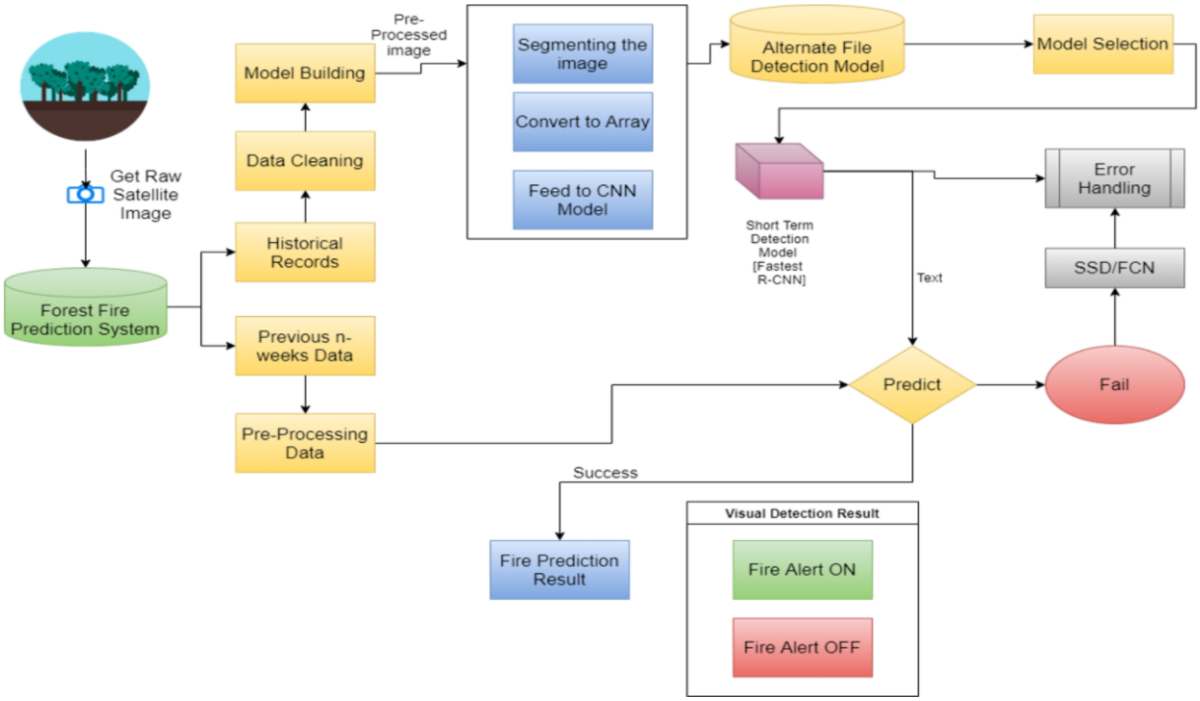
Arhitectura antrenare:



Algoritmi folositi:
- Faster R-CNN (object detection model, cu mentiunea ca se poate folosi SSMD - single shot multibox detector in cazul in care Faster R-CNN esueaza): folosit pentru “fire prediction result”, in care sunt detectate zonele in care ar fi foc.
- SVM (masina cu suport vectorial): pentru prezicerea daca este un foc real sau nu.

Autorii nu mentioneaza dimensiunea modelului / modul de hypertune a parametrilor.

Arhitectura model:



Metrici calculate si rezultate:
- Matricea de confuzie: pentru a identifica unde se “incurca” modelul.
- Acuratete: (total right predictions / total predictions) * 100, rezultat: 92%
- Recall: TP / (TP + FN), rezultat: 97.5%
- Precizie: TP/(TP + FP), rezultat: 84.78%
- F-Measure: (2 * Recall * Precision) / (Recall + Precision), rezultat: 90.7%

Review la paper: un paper bun pentru a intelege approach-urile uzuale (are un review bun la state of the art), dar datele nu sunt clare, si nici arhitectura modelului.

Paper 2: “Prediction and data mining of burned areas of forest fires: Optimized data matching and mining algorithm provides valuable insight”, David A. Wood

Algoritm folosit: TOB (Transparent Open Box) network

(Sursa: https://www.sciencedirect.com/science/article/pii/S2589721721000118).

Este o metodă bazată pe învățare automată folosită pentru a prezice suprafața arsă în funcție de variabile forestiere, meteorologice și de mediu, evitând folosirea de corelații, regresii sau relații statistice între variabile.

Algoritmul presupune două etape (1 și 2) și furnizează două estimări, a doua fiind optimizată. Compararea estimărilor în două etape cu subseturile de date de antrenare și testare ajută la identificarea și respingerea soluțiilor optimizate care se adaptează prea bine variabilelor de date subadicente (evită overfitting-ul). TOB stabilește cele mai bune ($Q \leq 10$) potriviri de înregistrări de date într-un subset mare de date de antrenare (pentru fiecare înregistrare de date specifică în subseturi de date relativ mici de ajustare - între aproximativ 100 și 150 de înregistrări de date, bazate pe analiza sensibilității). Se compară sumele diferențelor de pătrate ale variabilelor (VSD – variable squared differences) pentru toate variabilele de intrare între înregistrările de date specifice din subsetul de ajustare cu toate înregistrările de date din baza de date mai mare a subsetului de date de antrenare.

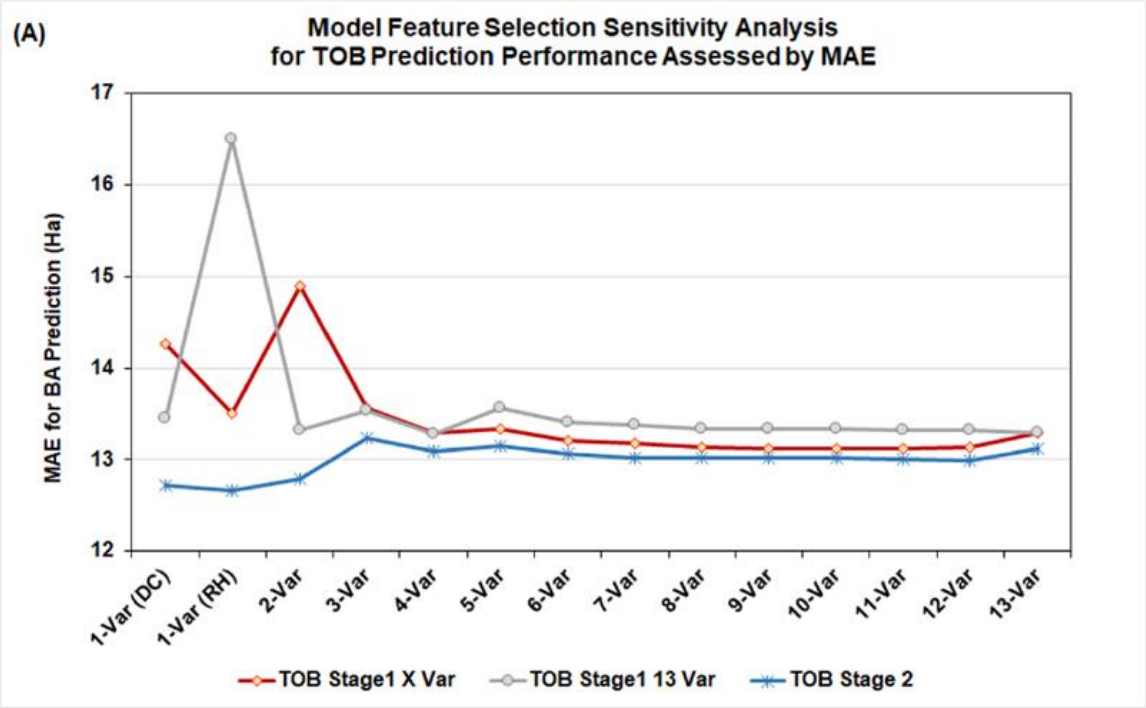
Dataset: forest-fire dataset from Portugal Montesinho Natural Park (https://archive.ics.uci.edu/dataset/162/forest+fires)

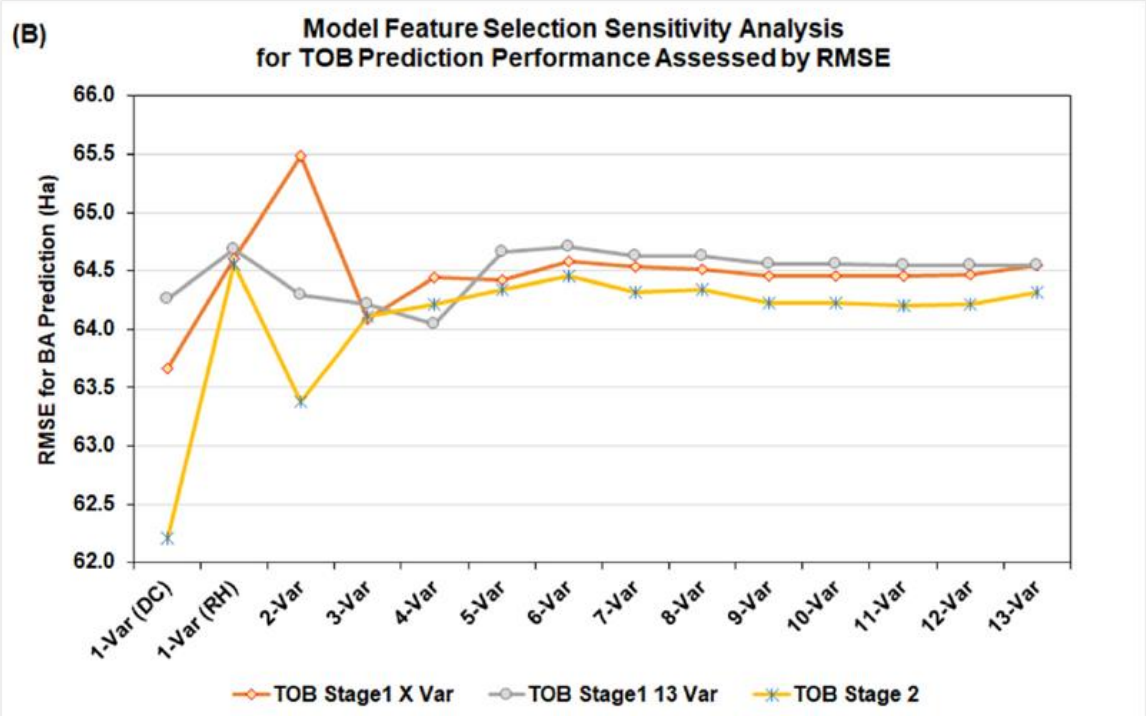
Conține 13 variabile ca input (conțin factori spațiali, temporali și meteorologici) adunate în urma analizării a 517 incendii din Portugal Montesinho Natural Park.

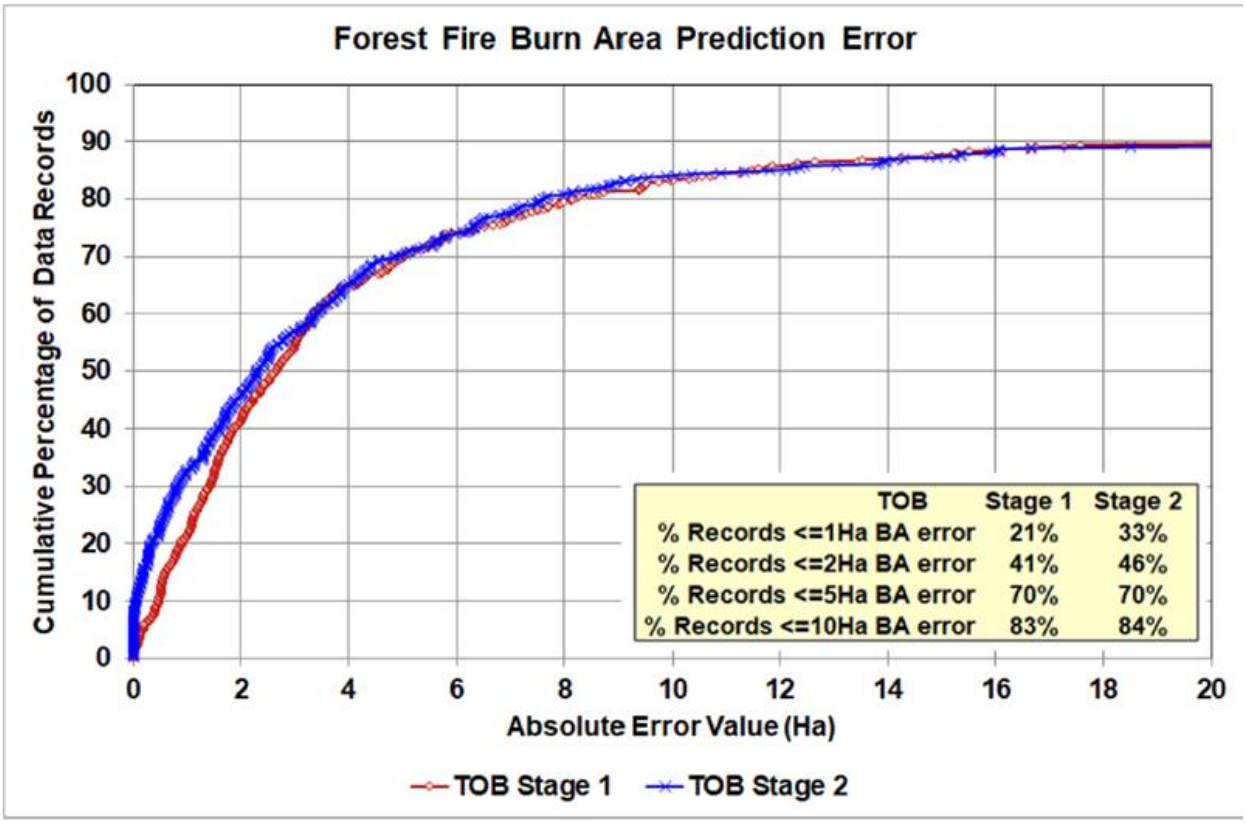
Metricile calculate:

- MSE (mean squared error)
- MAE (mean absolute error)
- RMSE (root mean squared error)
- PD (Procent deviation)
- APD (Average Procent deviation)
- AAPD (Absolute Average Procent deviation)
- SD (Standard Deviation)
- R (Correlation coefficient)

Rezultate:


(A)
Model Feature Selection Sensitivity Analysis
for TOB Prediction Performance Assessed by MAE
MAE for BA Prediction (Ha)
17
16
15
14
13
12
1-Var (DC)
1-Var (RH)
2-Var
3-Var
4-Var
5-Var
6-Var
7-Var
8-Var
9-Var
10-Var
11-Var
12-Var
13-Var
TOB Stage1 X Var
TOB Stage1 13 Var
TOB Stage 2


(B)
Model Feature Selection Sensitivity Analysis
for TOB Prediction Performance Assessed by RMSE
RMSE for BA Prediction (Ha)
66.0
65.5
65.0
64.5
64.0
63.5
63.0
62.5
62.0
1-Var (DC)
1-Var (RH)
2-Var
3-Var
4-Var
5-Var
6-Var
7-Var
8-Var
9-Var
10-Var
11-Var
12-Var
13-Var
TOB Stage1 X Var
TOB Stage1 13 Var
TOB Stage 2



Prediction accuracies achieved for independent testing subsets						
Statistics for solutions from 50 (100-record) tuned subsets	TOB Stage 2 solution: Optimized for RMSE		TOB stage 1 solution (1-Var (DC only) mean		TOB stage 2 solution: optimized for MAE	
	MAE (Ha)	RMSE (Ha)	MAE (Ha)	RMSE (Ha)	MAE (Ha)	RMSE (Ha)
			5.992	**11.75**		
Minimum	**4.603**	**9.11**	**← Best Case →**		**4.433**	**9.06**
25 Percentile	4.854	9.86	All solutions tested with 100-data record independent testing subsets		5.317	11.57
50 Percentile	5.597	11.95			6.344	12.72
Mean	5.758	11.79			6.037	13.08
75 Percentile	6.378	13.42			6.856	15.27
Maximum	7.576	17.92			7.623	19.89
Standard deviation	0.849	2.097			0.965	2.763

Dependent Variable Prediction Value Calculated for Data Record # 369						
Rank of Top-Matching Data Records	Top-ranking Matched Records in Training Subset	Dependent Variable Normalized Value (BA Ln)	Sum of Weighted Squared Errors (SumE)	Relative Magnitude of Contribution to Prediction	Relative Contribution fraction	Contributions to Prediction Components
TOB Stage 1 Equal Weights & Q=10		W_N =0.5	**TOB Stage 1 Prediction for Test Record # 369**			
Testing Subset Record Case #1A:	**369**	**-0.2530**	**SumE**	**Y = Total (X)/ Sum E**	**F= Y/ Total (Z)**	**Predicted Value F*BA(Ln)**
1 (1st ranking match)	506	-0.4520	1.000E-07	158.9626	0.3017	-0.1364
2	424	-0.8450	1.000E-07	158.9626	0.3017	-0.2550
3	376	0.0571	1.000E-07	158.9626	0.3017	0.0172
4	348	-1.0000	2.228E-06	7.1346	0.0135	-0.0135
5	349	-1.0000	2.228E-06	7.1346	0.0135	-0.0135
6	357	-0.7669	2.228E-06	7.1346	0.0135	-0.0104
7	350	-0.7225	2.228E-06	7.1346	0.0135	-0.0098
8	354	-0.7139	2.228E-06	7.1346	0.0135	-0.0097
9	353	-0.6831	2.228E-06	7.1346	0.0135	-0.0093
10 (10th rankingMatch)	351	-0.5570	2.228E-06	7.1346	0.0135	-0.0075
The top three ranking matches contribute 91% to the TOB Stage 1 predicted value			**1.590E-05** **Total (X)**	**526.8304** **Total (Z)**	**1.0000** **Sum of F**	**-0.4478** **Normalized Prediction = Sum (F*BA Ln)**
					Min Dependent Variable Actual Value (BA Ln):	0.0000
					Min Dependent Variable Actual Value (BA Ln):	6.9956
					Stage 1 Provisional Prediction of Dependent Variable (BA Ln) Value for Data Record: #369	**1.9315**
					Actual Recorded Dependent Variable Value (BA Ln) for Data Record: #369	2.6130
					Difference between Actual and TOB Stage 1 Predicted BA (Ln) Value:	0.6815
TOB Stage 2 with Optimized Weights						
Stage 2 TOB Optimized Q = 6		**BA (Ln)**	**SumE**	**Y = Total (X)/ Sum E**	**F= Y/ Total (Z)**	**Predicted Value F*BA(Ln)**
6	506	-0.4520	8.232E-01	2.0910	0.0184	-0.0083
5	424	-0.8450	3.319E-01	5.1867	0.0458	-0.0387
1 (Best)	376	0.0571	2.339E-02	73.5913	0.6492	0.0371
4	348	-1.0000	2.819E-01	6.1067	0.0539	-0.0539
3	349	-1.0000	1.322E-01	13.0179	0.1148	-0.1148
2	357	-0.7669	1.288E-01	13.3687	0.1179	-0.0904
N/A	350	-0.7225	0.000E+00	0.0000	0.0000	0.0000
N/A	354	-0.7139	0.000E+00	0.0000	0.0000	0.0000
N/A	353	-0.6831	0.000E+00	0.0000	0.0000	0.0000
N/A	351	-0.5570	0.000E+00	0.0000	0.0000	0.0000
The three records with the lowest SumE contribute 88% to the TOB Stage 2 predicted value		**Solution 1-Var (DC) All 517 Records Applied**	**1.721E+00** **Total (X)**	**113.3623** **Total (Z)**	**1.0000** **Sum of F**	**-0.2691** **Normalized Prediction = Sum (F*BA Ln)**
					Min Dependent Variable Actual Value (BA Ln):	0.0000
					Min Dependent Variable Actual Value (BA Ln):	6.9956
					Stage 2 Optimized Prediction of Dependent Variable Value (BA Ln) for Data Record: #369	**2.5567**
					Actual Recorded Dependent Variable Value (BA Ln) for Data Record: #369	2.6130
					Difference between Actual and TOB Stage 2 Predicted BA(Ln) Value:	0.0563
					Actual Recorded Dependent Variable Value (BA) for Data Record: #369 = $(e^{(BALn)})$-1:	**12.64**
					Predicted TOB Stage 1 Dependent Variable Value (BA) for Data Record: #369 = $(e^{(BALn)})$-1:	**5.90**
					Predicted TOB Stage 2 Dependent Variable Value (BA) for Data Record: #369 = $(e^{(BALn)})$-1:	**11.89**