20th International Sunflower Conference, June 20 to 23, 2022, Novi Sad, Serbia

FEATURE SELECTION AND PERFORMANCE ASSESSMENT OF MACHINE LEARNING ALGORITHMS FOR SUNFLOWER OIL YIELD PREDICTION

Sandra Cvejić, Olivera Hrnjaković, Milan Jocković, Aleksandar Kupusinac, Ksenija Doroslovački, Ilija Radeka, Siniša Jocić, Dragana Miladinović, Vladimir Miklič

Institute of Field and Vegetable Crops, Novi Sad, Serbia
Faculty of Technical Sciences, University of Novi Sad, Serbia

AIM:

(1) to investigate the potential use of ML algorithms for predicting sunflower oil yield
(2) to compare two different subsets to identify important features for the prediction
(3) to assess the performance of ML algorithms applied to sunflower oil yield prediction
(4) to visually analyze the prediction models' results to guide applications in practice

2 DATA SUBSETS:

First dataset description	Feature
Seed yield	Predictor
Hybrid type (OIL, IMI, SU, HO, CON)	Predictor
Resistance to broomrape	Predictor
Resistance to downy mildew	Predictor
Maturity	Predictor
Seed oil yield	**Target**

Second dataset description	Feature
Seed yield	Predictor
Locality	Predictor
Weather parameters	Predictor
Seed oil yield	**Target**

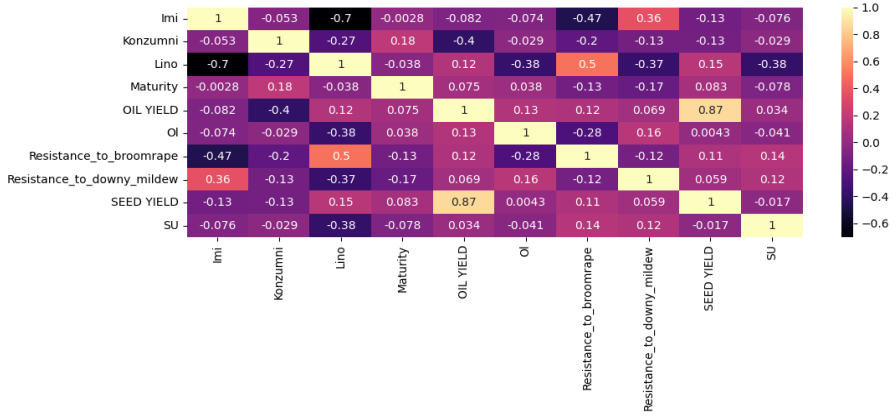
ML ALGORITMS:

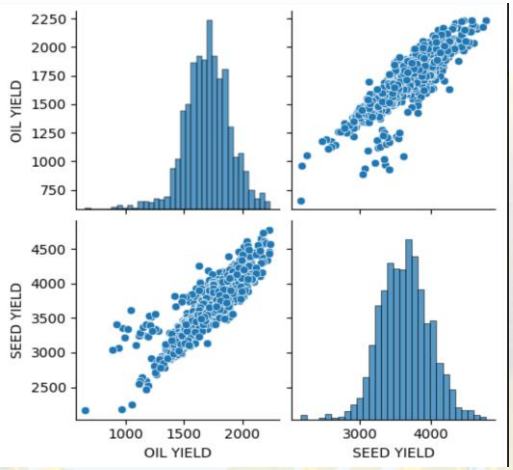
Artificial Neural Network (ANN)
Support Vector Regression (SVR)
K-Nearest Neighbour (KNN)
Random Forest Regressor (RF)

1250 sunflower hybrid samples

FEATURES ANALYSIS:

- The data in both subsets was split to train (70%) and test set (30%)
- There is a significant positive correlation between seed yield and oil yield as well as a small positive correlation between resistance to broomrape, ol, oil and oil yield



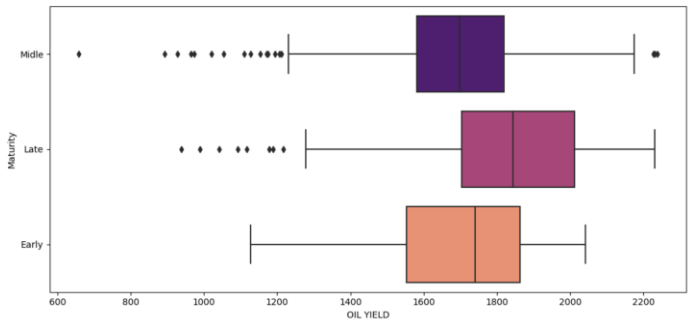


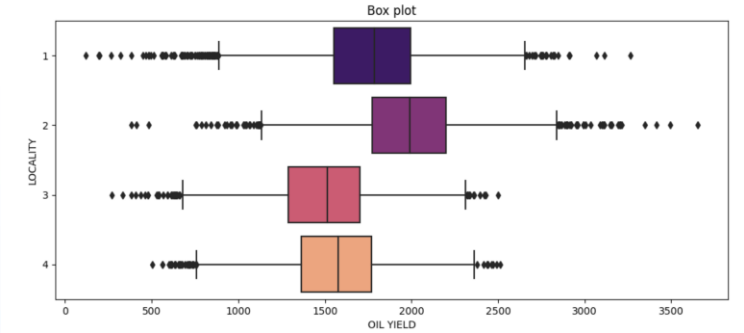
Parameters	All (n=1250)	Early (n=62)	Middle (n=1123)	Late (n=65)	p-value
Seed yield	3632 (2161-4781)	3694 (2593-4341)	3615 (2161-4744)	3872 (3008-4781)	<0.001*
Oil yield	1700 (658-2238)	1691 (1127-2042)	1695 (658-2238)	1783 (938-2231)	0.003

Parameters	All (n=15000)	Locality 1 (n=3750)	Locality 2 (n=3750)	Locality 3 (n=3750)	Locality 4 (n=3750)	p-value
Seed yield	3632 (214-7447)	3842 (214-6567)	4103 (779-7447)	3234 (797-5195)	3350 (1212-5383)	<0.001*
Oil yield	1700 (120-3651)	1763 (120-3262)	1987 (384-3651)	1489 (267-2500)	1561 (501-2512)	<0.001*

STATISTICAL RESULTS:

- **Oil yield in relation to maturity and locality (weather conditions)**

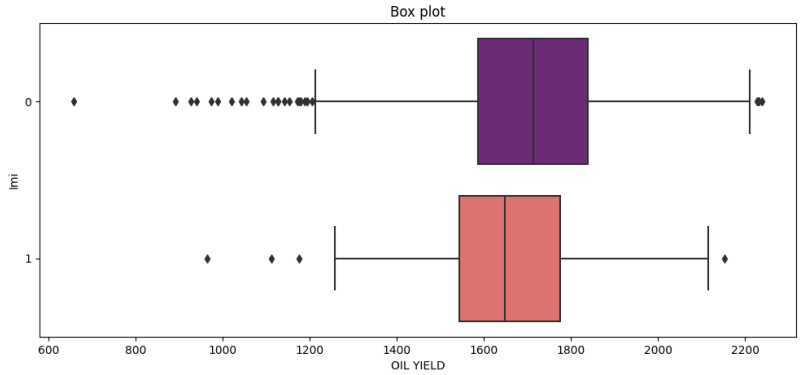




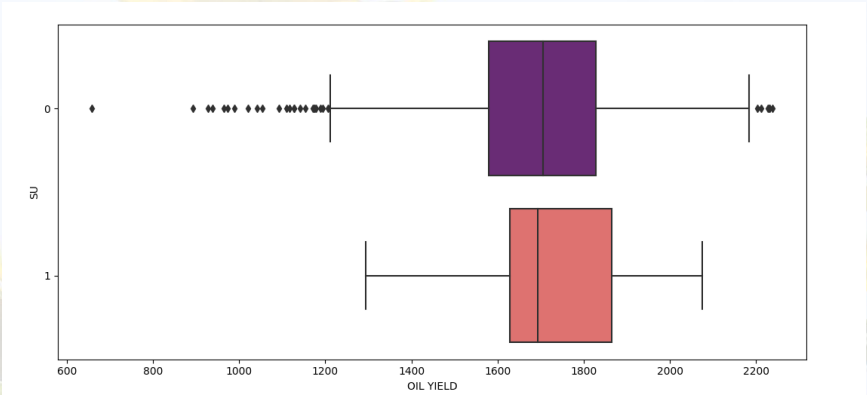
Locality	Koordinates	Soil type	T_{avr} Apr 2018 (°C)	T_{avr} May 2018 (°C)	T_{avr} Jun 2018 (°C)	T_{avr} July 2018 (°C)	T_{avr} Avg 2018 (°C)	T_{avr} Sep 2018 (°C)	T_{avr} 2018 (°C)	Deviation T_{avr} perennial (°C)	No days T_{max} >20°C	No days T_{max} >30°C	No days T_{max} >35° C	No. Rainy days	Realized vegetation rainfalls in mm	Realized vegetation rainfalls in %
1 - Subotica	46°05′53″N; 19°40′16″E	Sandy soil	16.5	20.4	21.6	23.0	24.3	18.0	20.63	+2.1	169	35	0	51	312	93
2 - Novi Sad	45°19′51″N; 19°50′59 E	Chernozem	17.2	20.4	21.4	21.9	24.0	18.5	20.57	+2.3	173	37	0	51	436	121
3 - Kikinda	45°43′11″N; 20°18′07″E	Humogley + Salty soil	16.7	20.6	21.4	22.4	24.2	18.6	20.65	+2.1	176	43	0	55	409	124
4 - Vršac	44°58′25"N, 21°13′17" E	Chernozem	17.3	20.3	20.9	22.1	24.1	18.4	20.52	+2.0	171	48	0	45	447	115

STATISTICAL RESULTS:

- **Oil yield distribution in relation to IMI (statistically significant - p-value <0.001)**

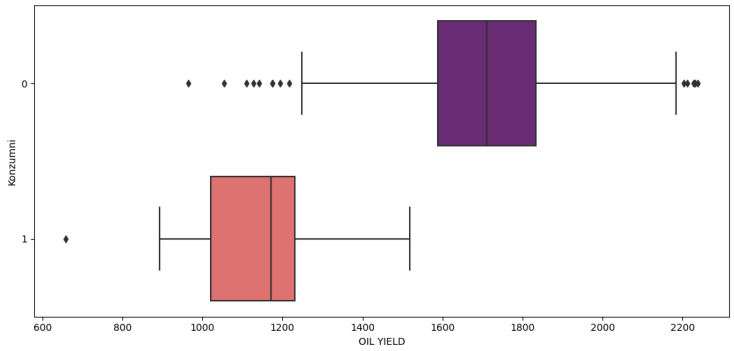


- **Oil yield distribution in relation to SU (statistically significant - p-value <0.001)**

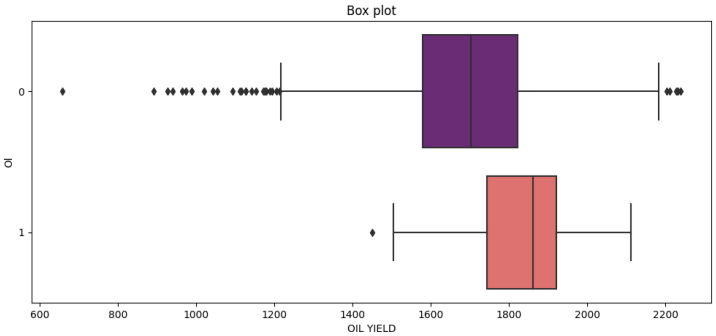


STATISTICAL RESULTS:

- Oil yield distribution in relation to CON (statistically significant - p-value <0.001)



- Oil yield distribution in relation to OI (statistically significant - p-value < 0.001)



RESULTS:

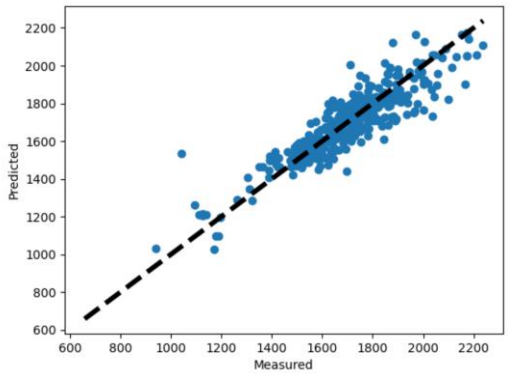
- **RMSE (Root Mean Squared Error) was used as the evaluation metric**

First dataset results	
Algorithm	**RMSE**
SVR	189
RF	**39**
NN	200
KNN	65

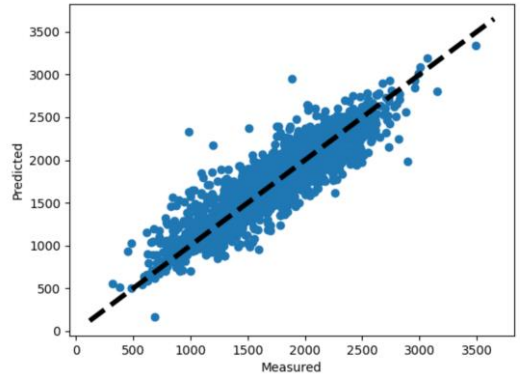
Second dataset results	
Algorithm	**RMSE**
SVR	162
RF	**66**
NN	390
KNN	98

- **Graphs below show the best model's performance (RF), i.e. the relationship between actual values and predicted oil yield values**

First dataset



Second dataset



ACKNOWLEDGEMENT:

Science Fund of the Republic of Serbia, through IDEAS project “Creating climate smart sunflower for future challenges” (SMARTSUN) grant number 7732457

Република Србија
МИНИСТАРСТВО ПРОСВЕТЕ, НАУКЕ И ТЕХНОЛОШКОГ РАЗВОЈА

Ministry of Education, Science and Technological Development of the Republic of Serbia, grant number: 451-03-68/2022-14/ 200032

Center of Excellence for Innovations in Breeding of Climate-Resilient Crops - Climate Crops, Institute of Field and Vegetable Crops, Novi Sad, Serbia