

Agri-yield Prediction using machine learning

Existing Research

- 2 different approaches have been widely adopted for weather and yield prediction
 - 1) Process based simulation methods
 - 2) Statistical methods

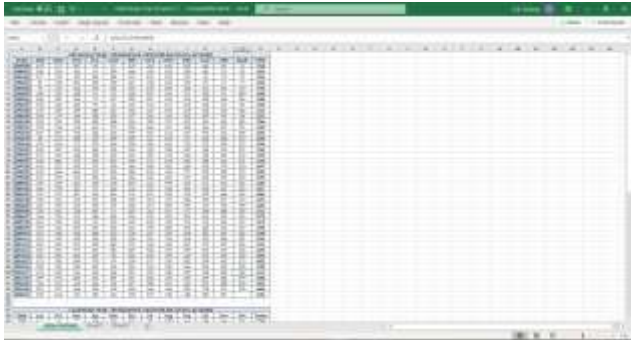
- Several advantages presented by each model:
 - 1) Process based simulation models - utilise decades of research on crop culture and soil science.
 - 2) Statistical methods - backed by solid statistical inferences and mathematics.

Disadvantages and Gaps

- Process based simulation methods
 - i. Extensive architecture development.
 - ii. Highly specific nature.
 - iii. Limited extension capabilities.
- Statistical methods
 - i. Non-linear relationships between data.
 - ii. Lack of stochastic component following normal distribution.

Machine Learning Models

- Machine learning models address the above shortcomings of statistical models and process-based simulation models.
- Reflect non linear dependencies between data.
- No requirement for extensive parameterisation or architecture development.
- Relatively more accessible than traditional models.

A screenshot of an Excel spreadsheet with multiple data tables. Four red markers are placed on the first row of each of the four tables, indicating specific data points or headers.A screenshot of an Excel spreadsheet showing a single data table. The table is transposed, with data organized in rows and columns.

Methodology

- Datasets were obtained from UPASI Tea Research Foundation and KDHP Co Pvt. Ltd.
- Climatic data was non-contiguous present in 828 different tables across 7 different sheets.
- Yield data was present in transposed format in separate tables.
- Null and Garbage values were present.

Models

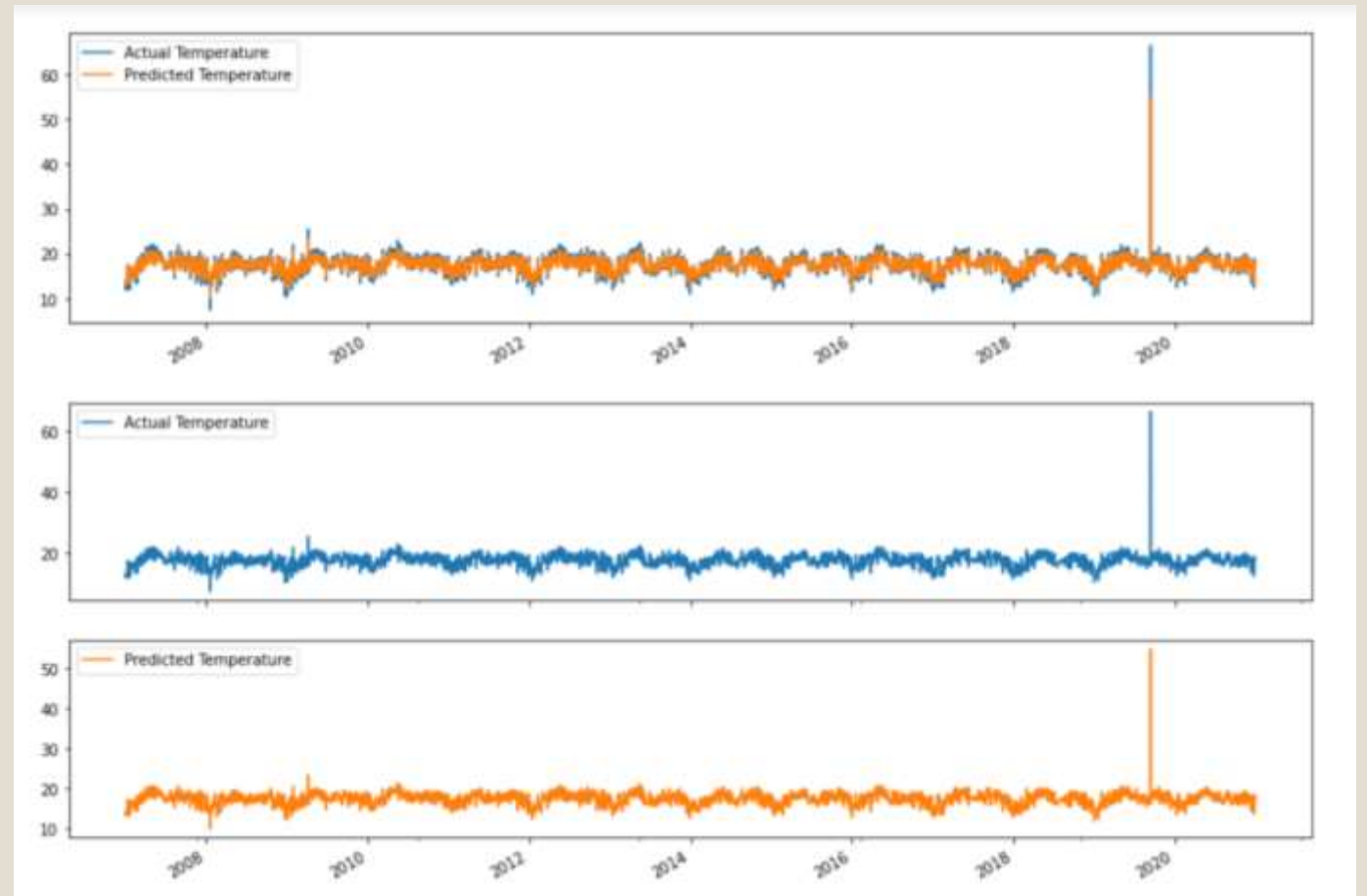
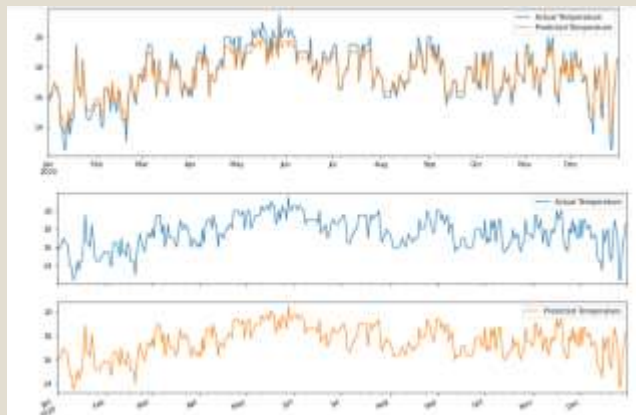
- Python programming language used for training and developing the model.
- Four machine learning models build for each climatic parameter, i.e. rainfall, temperature and humidity.
- The algorithms used are:
 - SVR (Support-Vector Machine)
 - SVM (Support-Vector Regression)
 - MLP (Multilayer Perceptron)
 - KNN (K-nearest Neighbour)

Results for Temperature Prediction

	SVR	SVM	MLP	KNN
R^2	0.9978	0.982	0.9567	0.7227

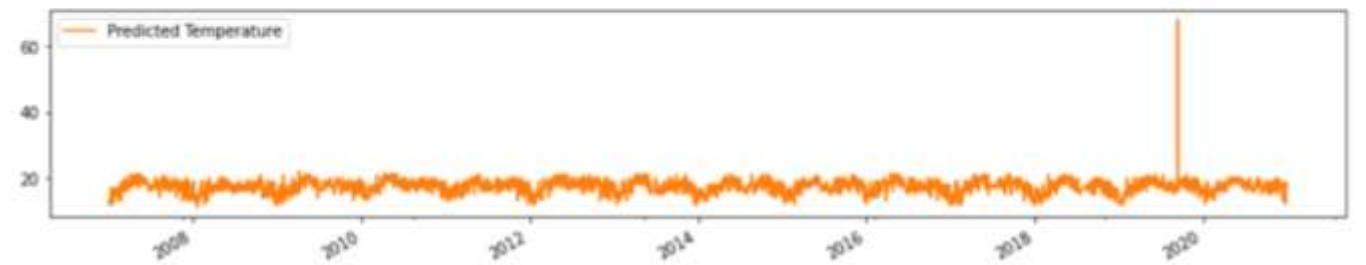
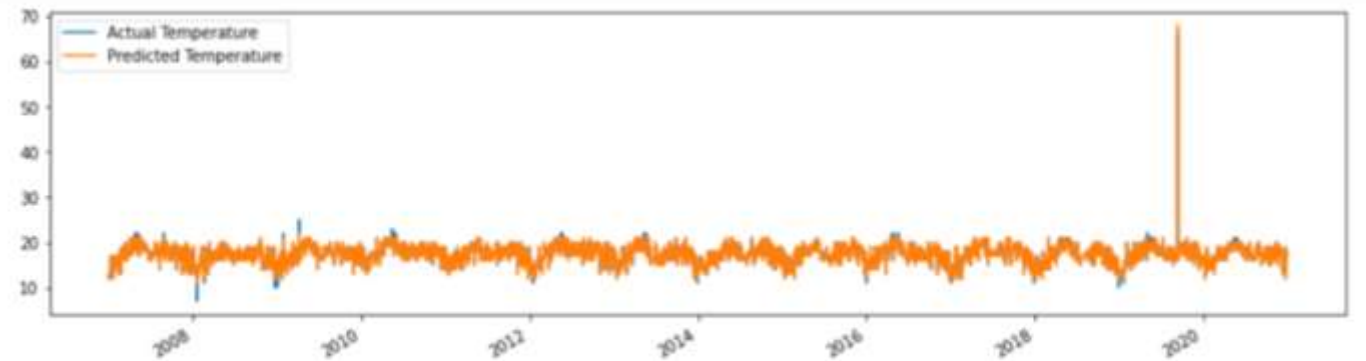
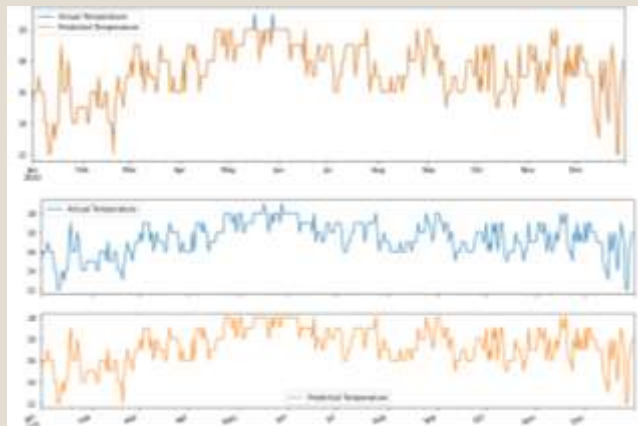
SVR

	Actual Temperature	Predicted Temperature
2020-01-01	15.5	15.845132
2020-01-02	16.0	16.227578
2020-01-03	16.0	16.227531
2020-01-04	16.5	16.610022
2020-01-05	17.0	16.992546
2020-01-06	16.5	16.609950
2020-01-07	16.5	16.609914
2020-01-08	16.0	16.227299
2020-01-09	14.0	14.697015
2020-01-10	14.0	14.696927



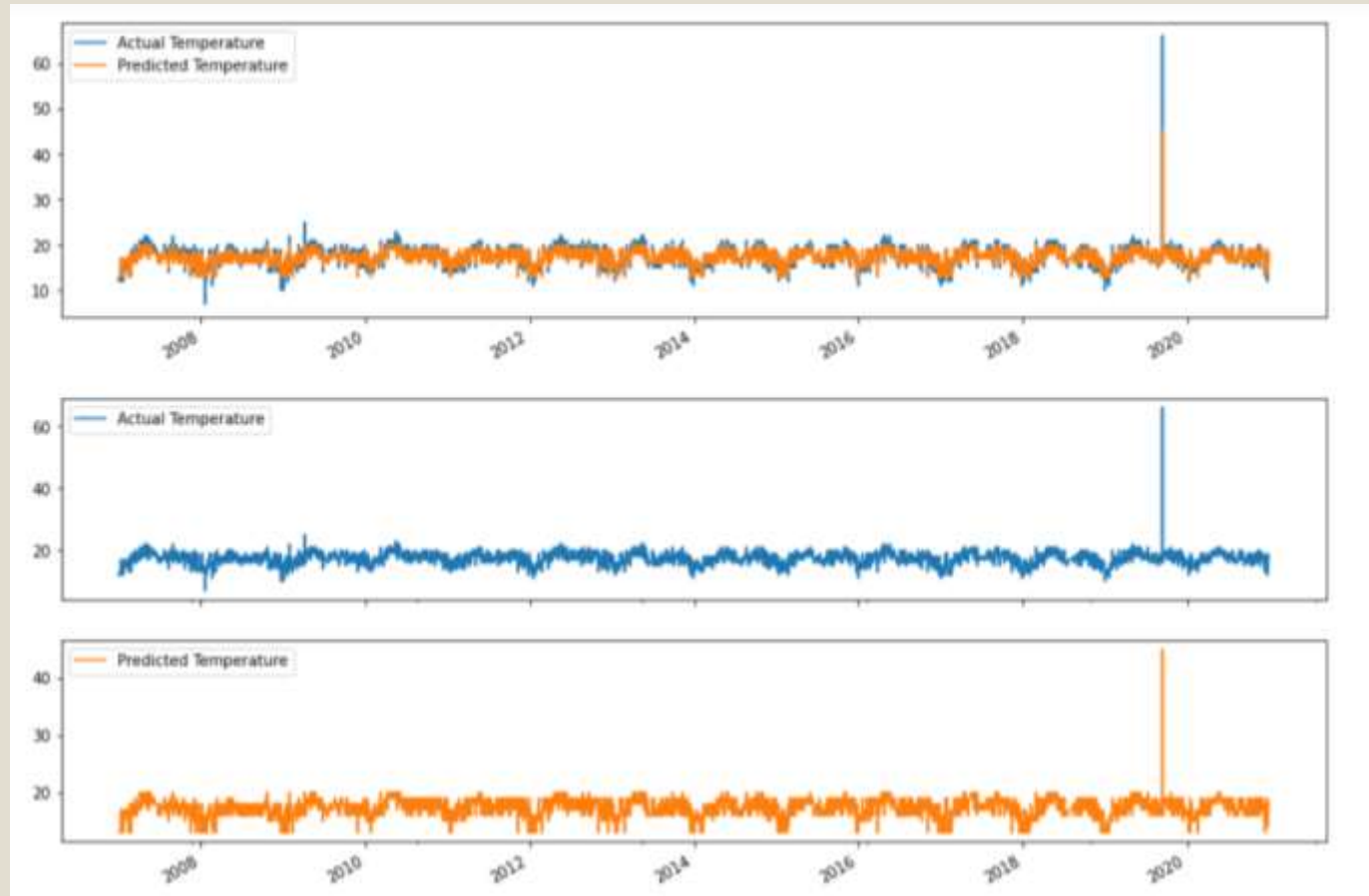
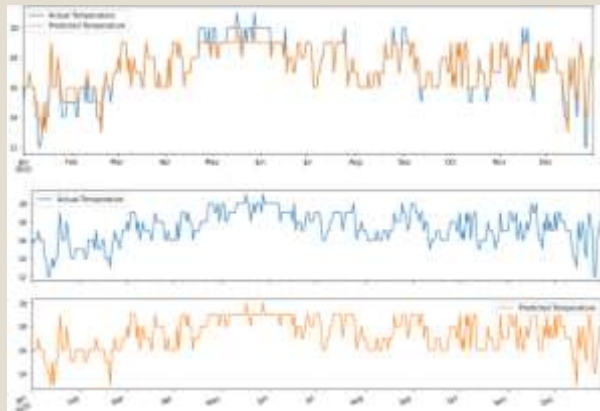
MLP

	Actual Temperature	Predicted Temperature
2007-01-01	12	12
2007-01-02	12	12
2007-01-03	13	13
2007-01-04	13	13
2007-01-05	13	13
2007-01-06	13	13
2007-01-07	12	12
2007-01-08	14	14
2007-01-09	15	15
2007-01-10	17	17



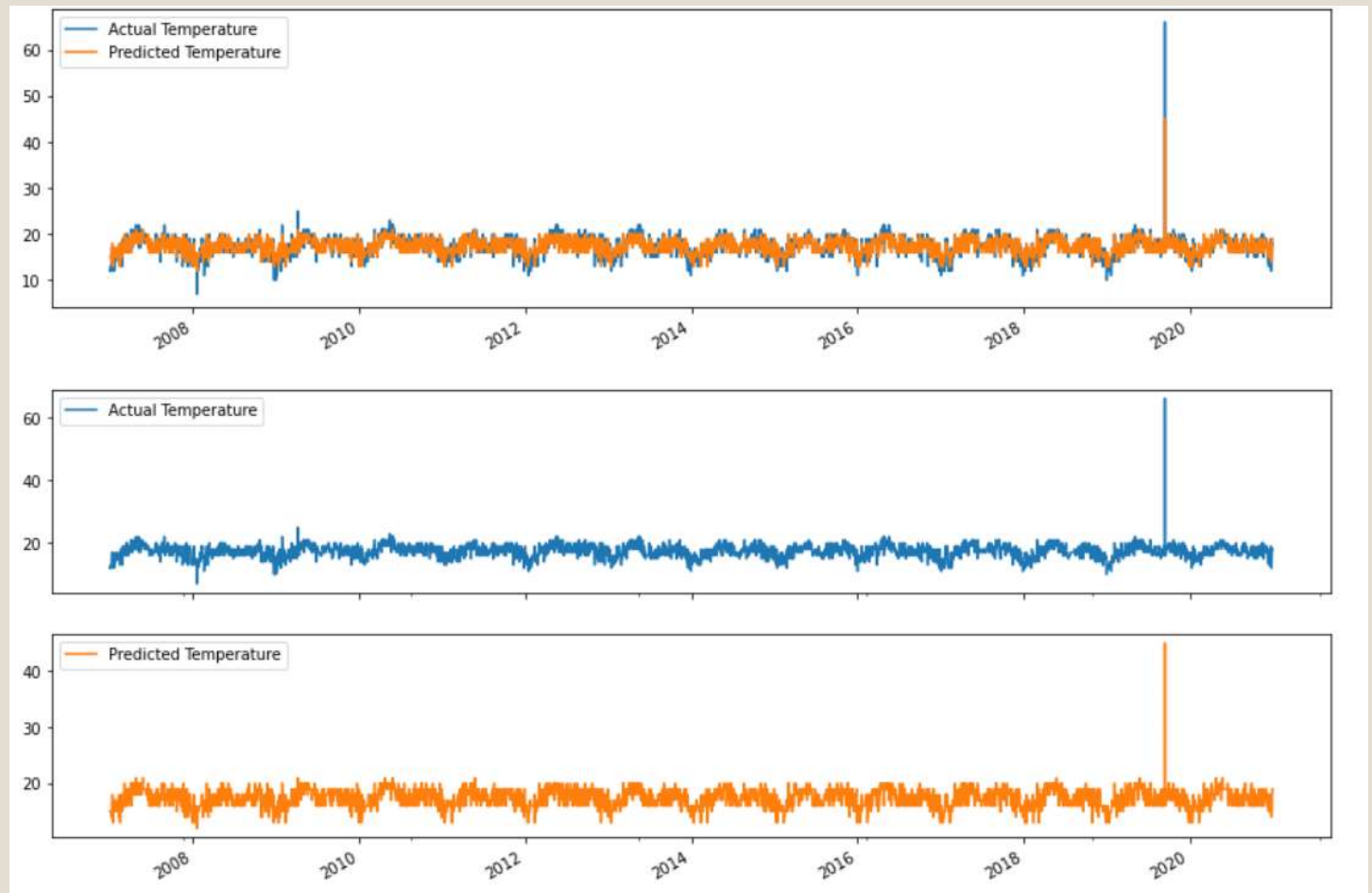
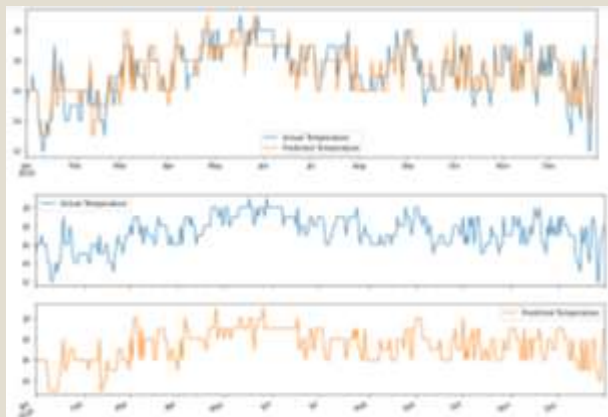
SVM

	Actual Temperature	Predicted Temperature
2020-01-01	15	16
2020-01-02	16	16
2020-01-03	16	16
2020-01-04	16	16
2020-01-05	17	17
2020-01-06	16	16
2020-01-07	16	16
2020-01-08	16	16
2020-01-09	14	15
2020-01-10	14	15



KNN

	Actual Temperature	Predicted Temperature
2007-01-01	12	15
2007-01-02	12	15
2007-01-03	13	15
2007-01-04	13	15
2007-01-05	13	15
2007-01-06	13	15
2007-01-07	12	15
2007-01-08	14	14
2007-01-09	15	14
2007-01-10	17	17

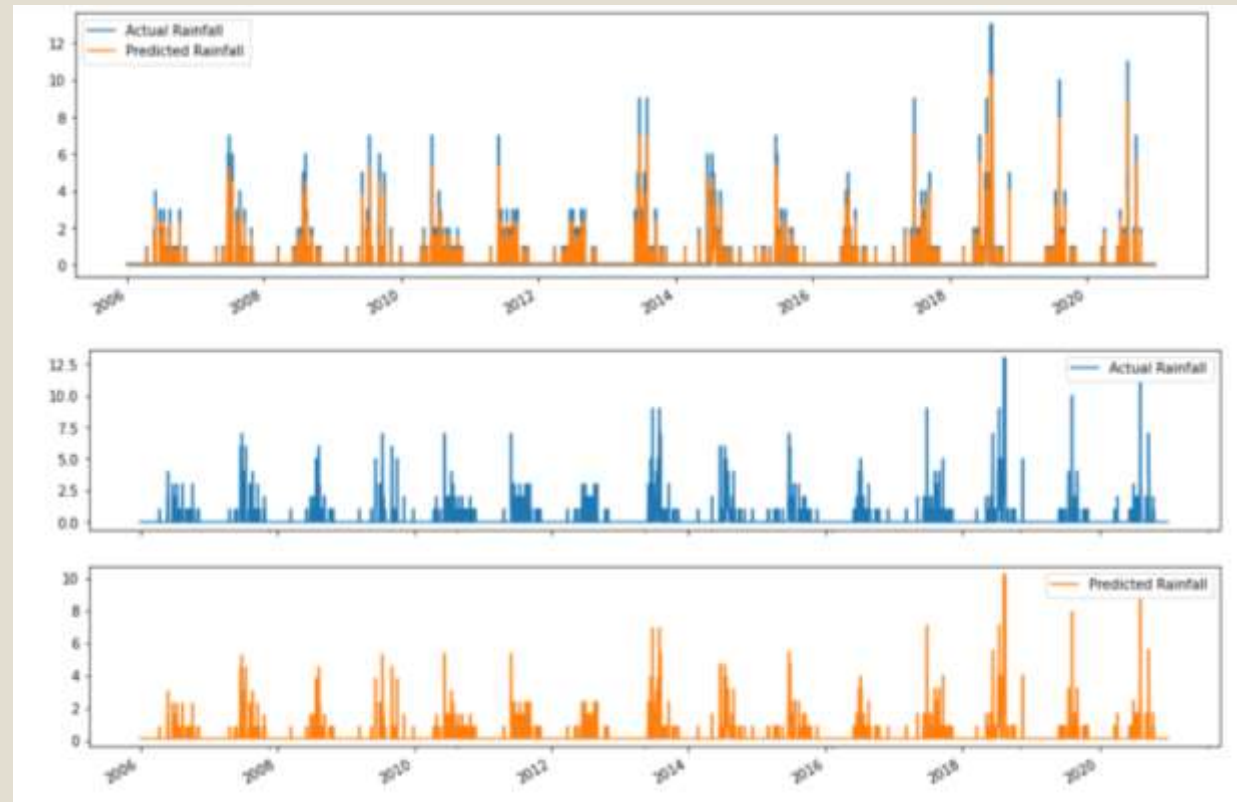
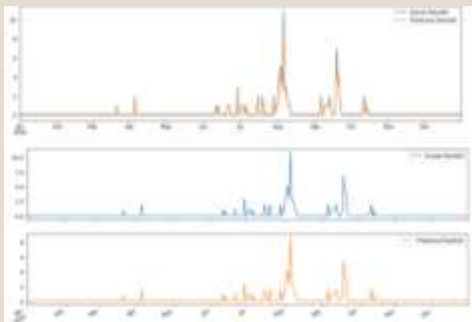


Results for Rainfall Prediction

	SVR	SVM	MLP	KNN
R^2	0.9926	0.9969	0.9878	0.8757

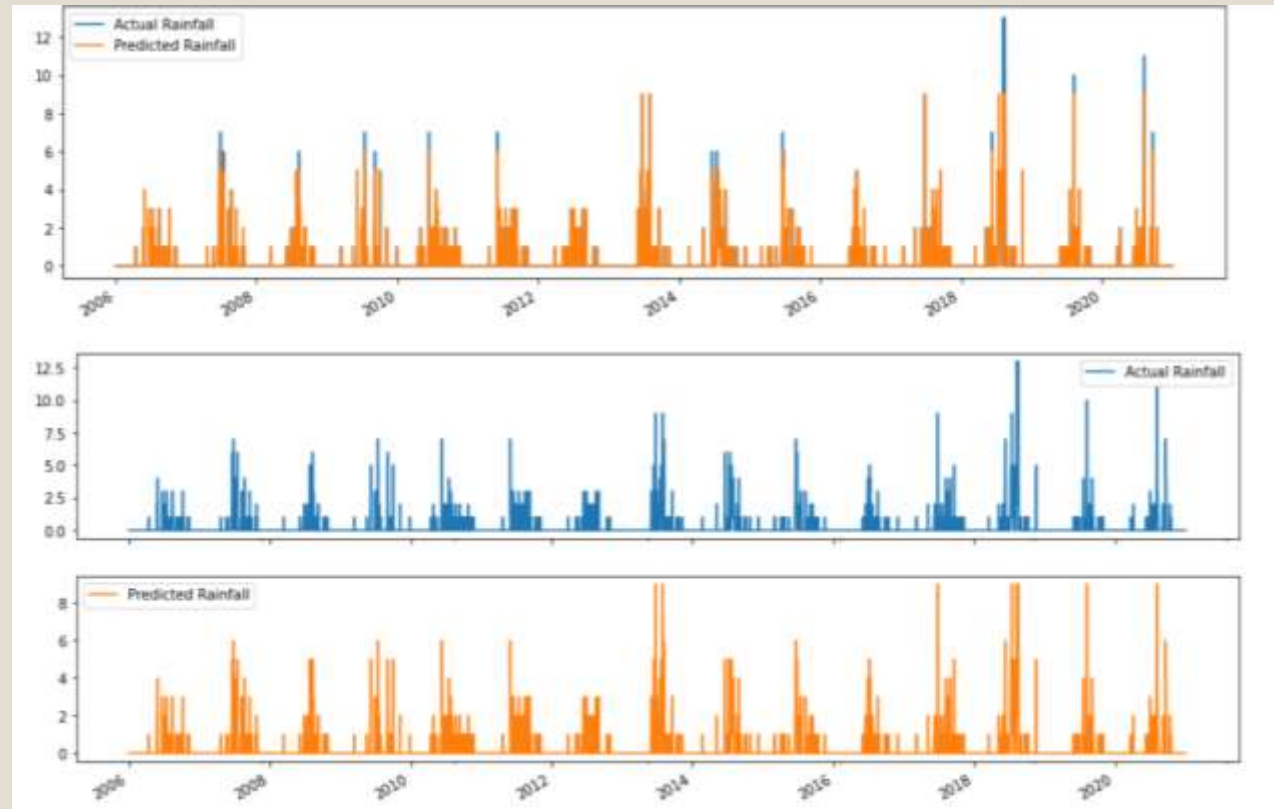
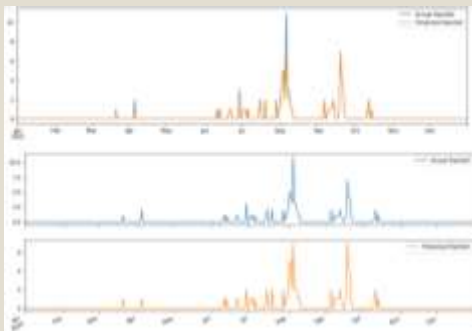
SVR

	Actual Rainfall	Predicted Rainfall
2020-04-05	2	1.675888
2020-06-29	3	2.466086
2020-07-16	2	1.676511
2020-07-20	2	1.676755
2020-07-29	2	1.677369
2020-08-02	3	2.463632
2020-08-03	5	4.039452
2020-08-04	5	4.039599
2020-08-05	3	2.463869
2020-08-06	11	8.767915



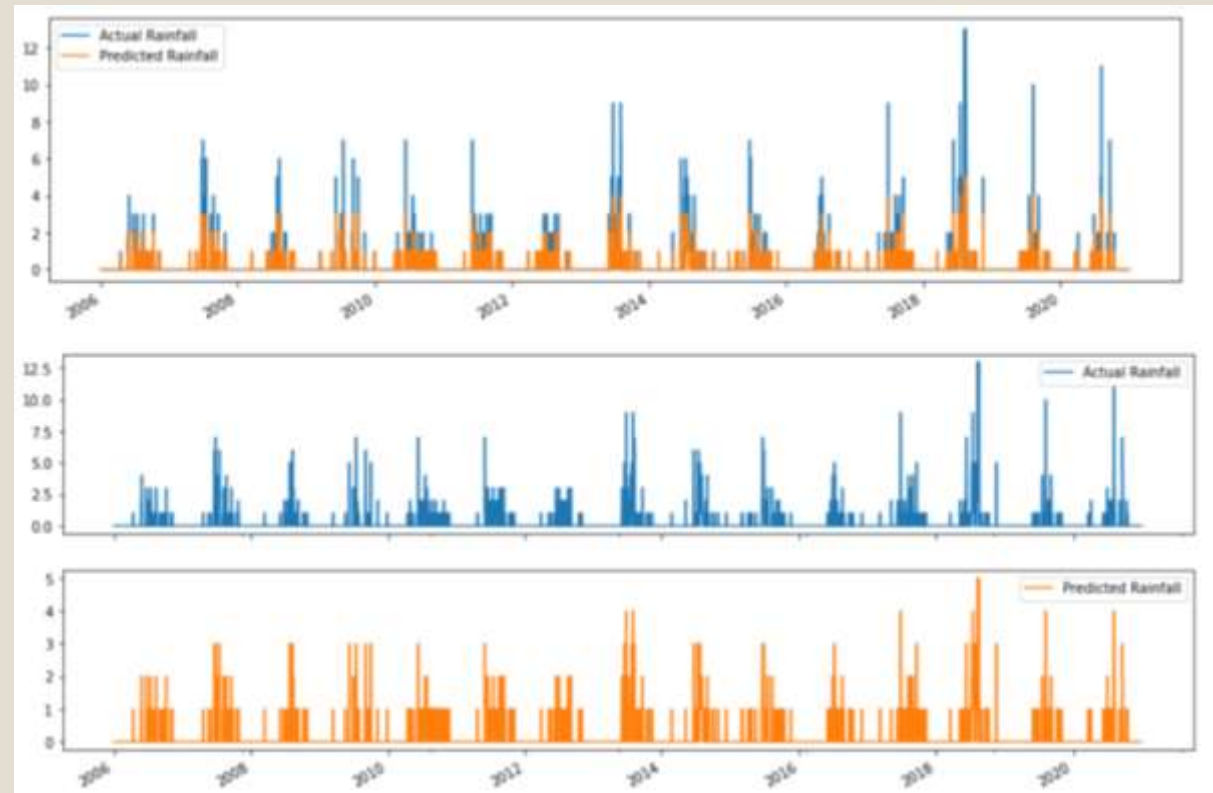
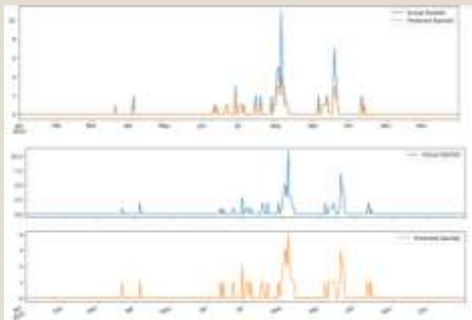
MLP

	Actual Rainfall	Predicted Rainfall
2020-03-21	1	1
2020-04-05	2	1
2020-06-11	1	1
2020-06-13	1	1
2020-06-21	1	1
2020-06-22	1	1
2020-06-29	3	2
2020-07-02	1	1
2020-07-03	1	1
2020-07-04	1	1



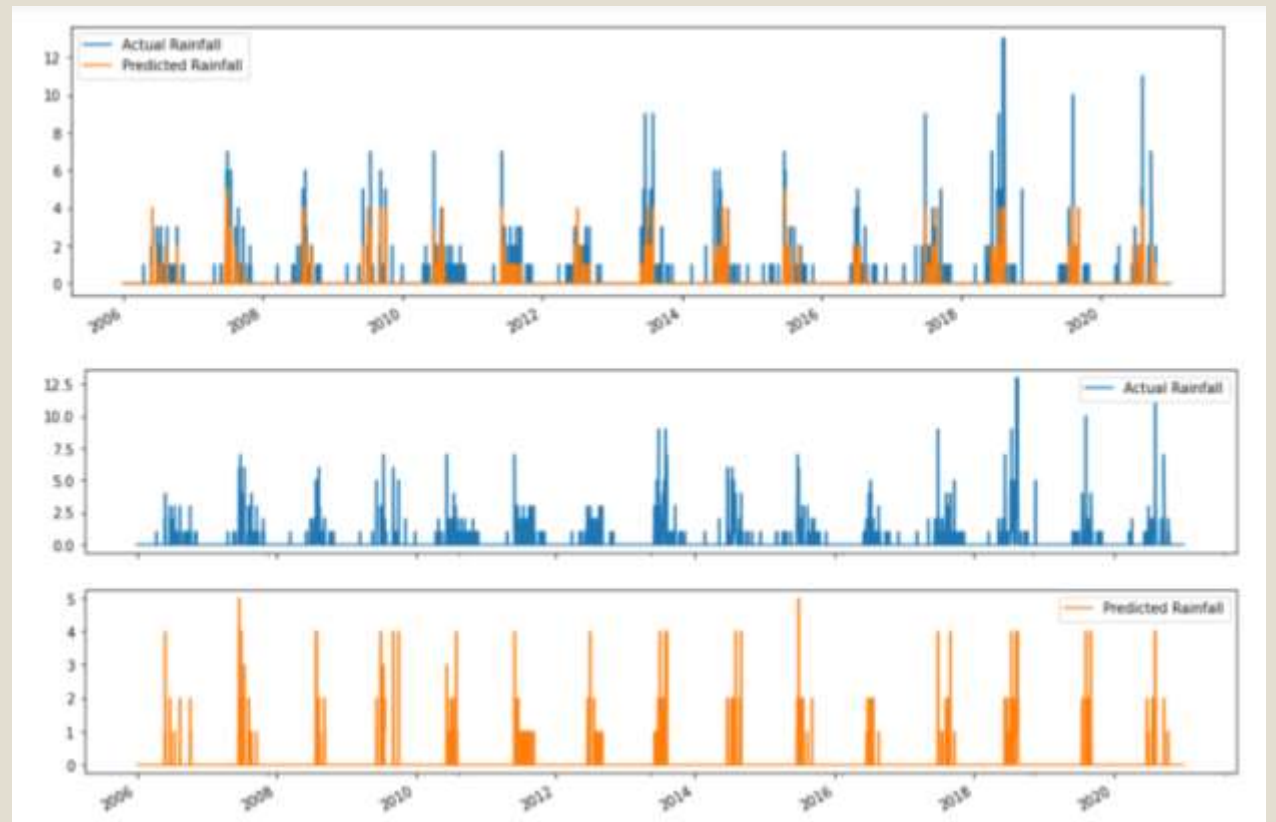
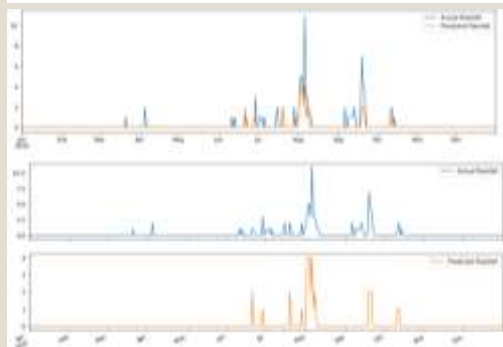
SVM

	Actual Rainfall	Predicted Rainfall
2006-04-16	1	1
2006-05-26	1	1
2006-05-27	1	1
2006-05-28	2	1
2006-05-29	1	1
2006-05-30	2	1
2006-05-31	4	2
2006-06-01	2	1
2006-06-23	2	1
2006-06-24	2	1



KNN

	Actual Rainfall	Predicted Rainfall
2006-05-28	2	1
2006-05-31	4	4
2006-06-23	2	2
2006-06-24	2	1
2006-06-25	3	1
2006-07-15	3	1
2006-07-17	2	1
2006-07-18	2	1
2006-08-13	2	1
2006-08-16	3	2

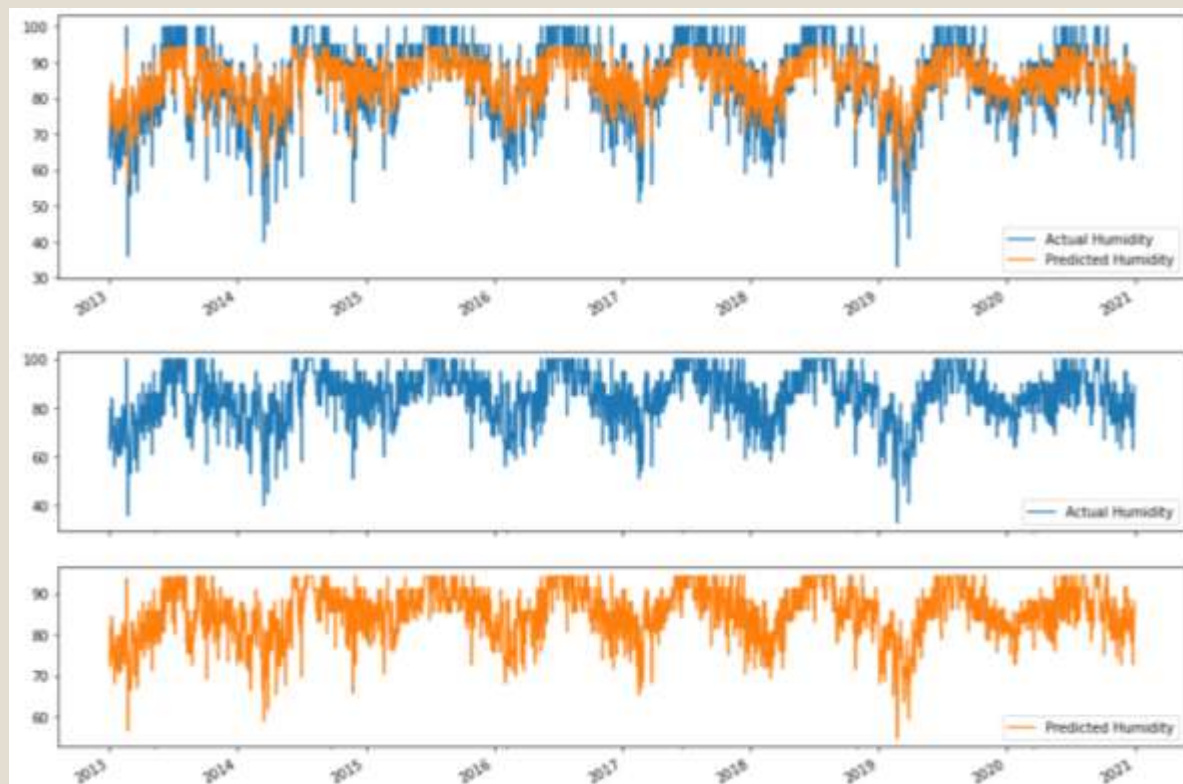
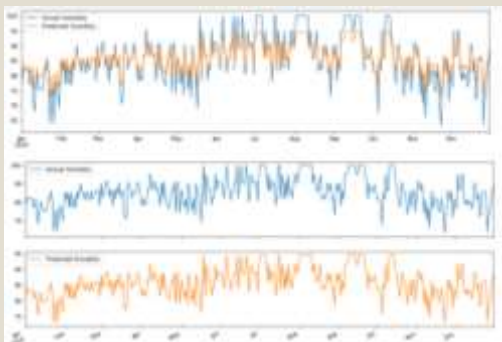


Results for Humidity Prediction

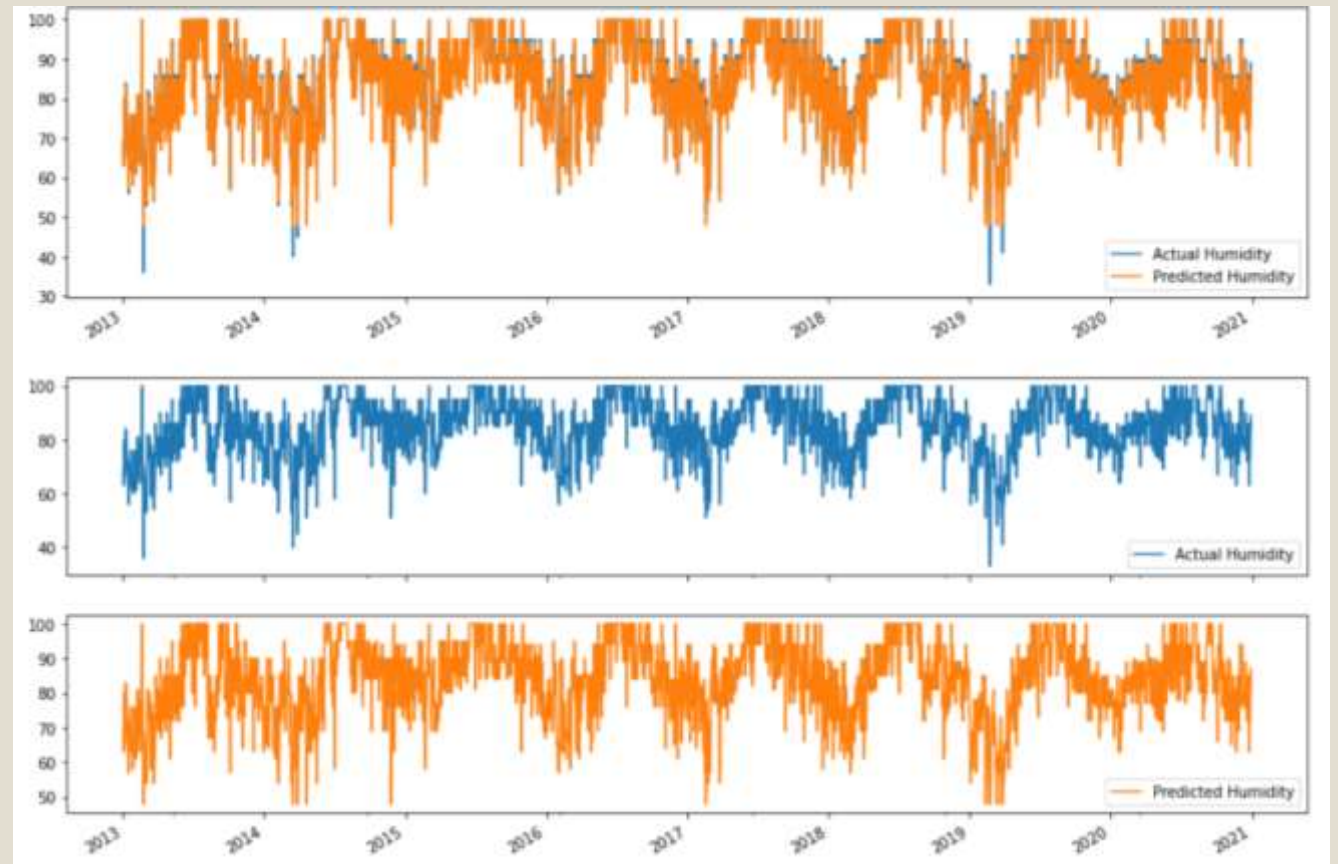
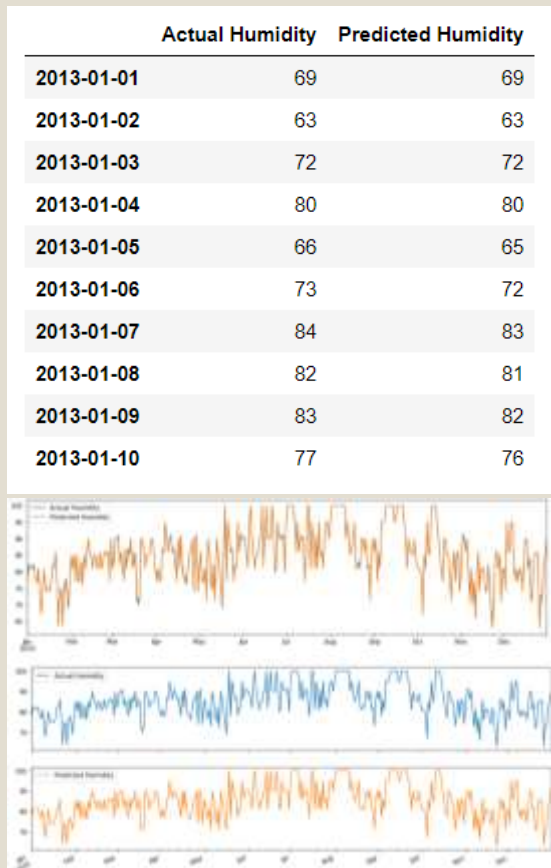
	SVR	SVM	MLP	KNN
R^2	0.9968	0.8842	0.996	0.9846

SVR

	Actual Humidity	Predicted Humidity
2013-01-01	89	75.691184
2013-01-02	63	72.206552
2013-01-03	72	77.433226
2013-01-04	80	82.085684
2013-01-05	66	73.946146
2013-01-06	73	78.012727
2013-01-07	84	84.412999
2013-01-08	82	83.247834
2013-01-09	83	83.829612
2013-01-10	77	80.336706

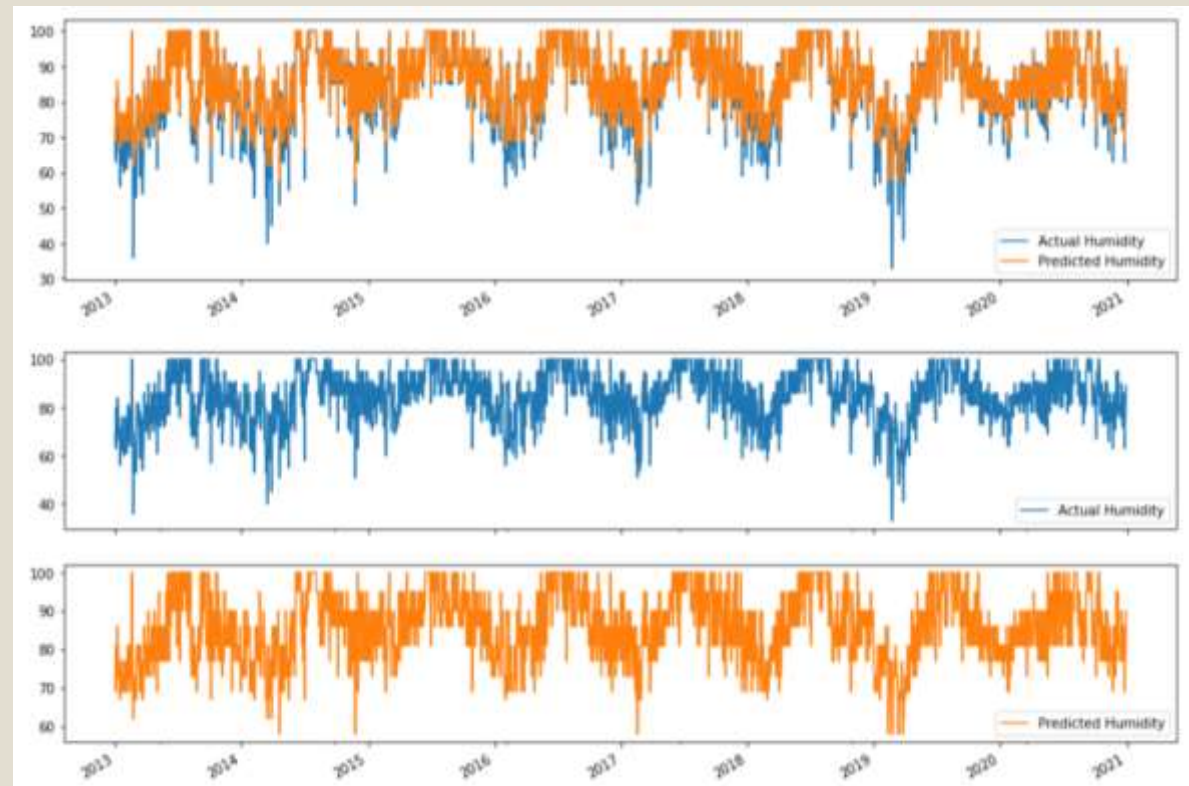
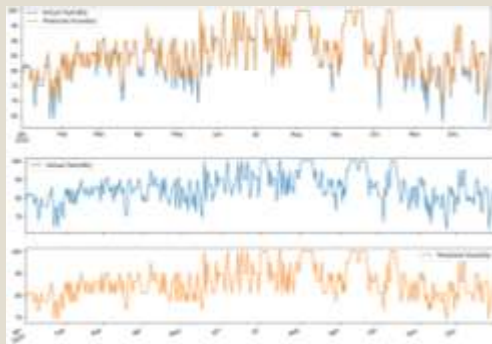


MLP

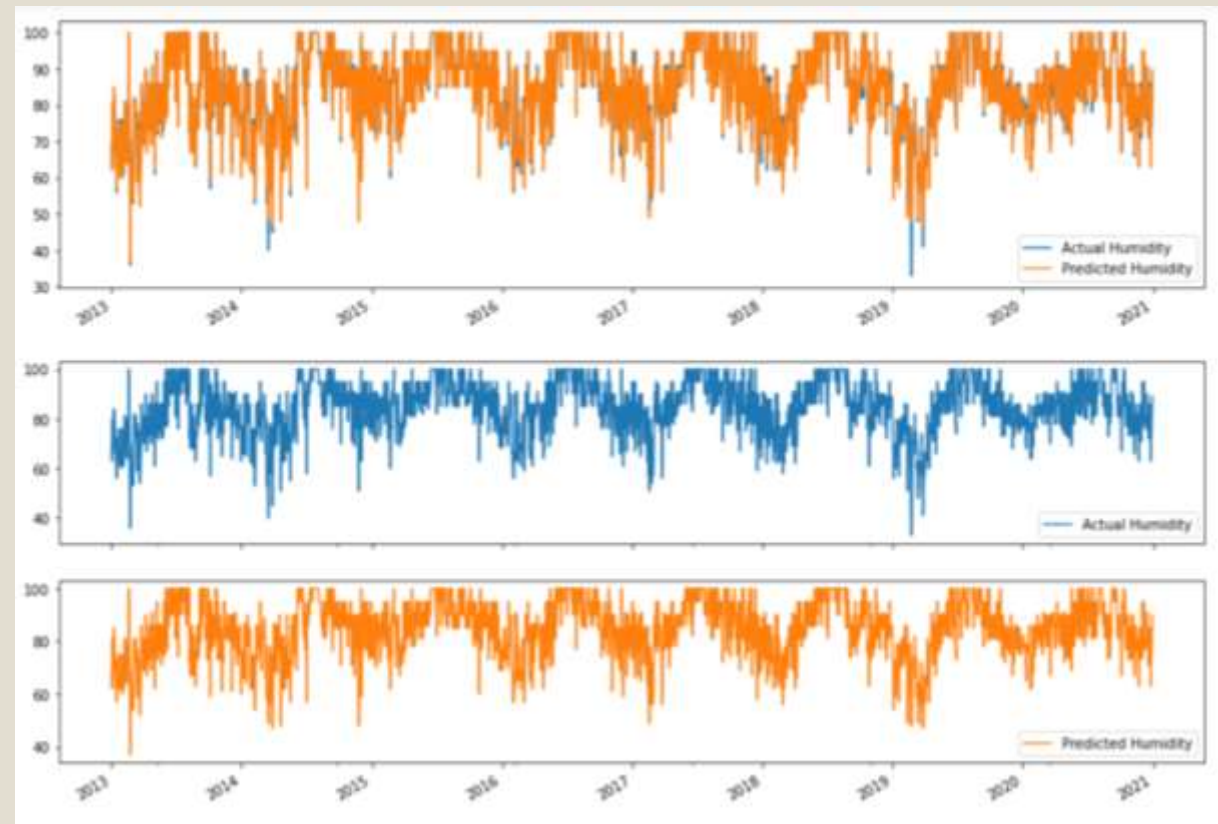
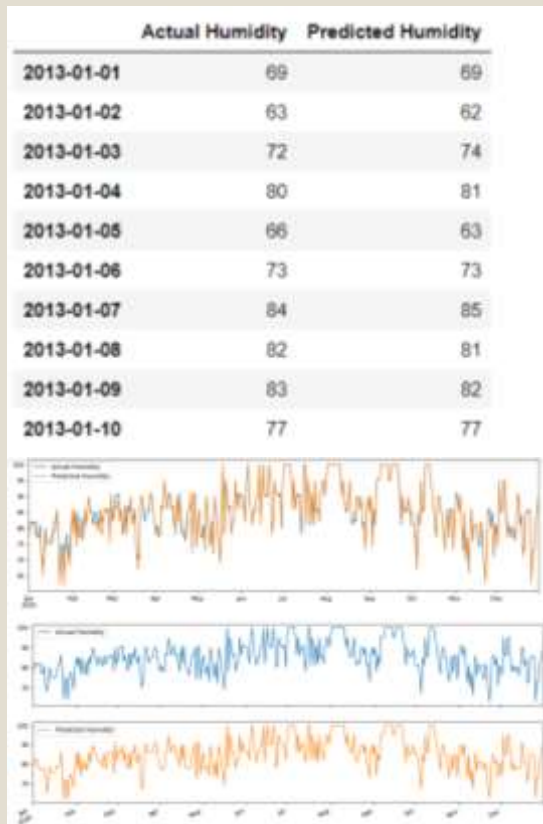


SVM

	Actual Humidity	Predicted Humidity
2013-01-01	69	73
2013-01-02	63	69
2013-01-03	72	77
2013-01-04	80	81
2013-01-06	66	73
2013-01-08	73	77
2013-01-07	84	86
2013-01-08	82	81
2013-01-09	83	81
2013-01-10	77	77



KNN



YIELD

- Model to predict model was trained using two different sets of parameters
 - 1) Parameters of model 1: Rainfall, Humidity and Temperature
 - Efficiency of SVR: 25.77%
 - Efficiency of MLP: 67.19%
 - 2) Parameters of model 2: Rainfall, Humidity, Temperature and Historical yield
 - Efficiency of SVR: 96.69%
 - Efficiency of MLP: 94.93%

Model 1

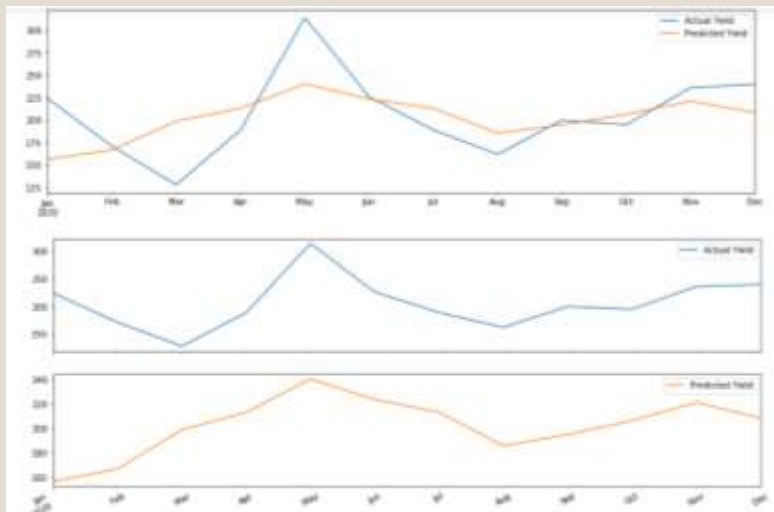
	Actual Yield	Predicted Yield
2020-01-01	224.0	156.408011
2020-02-01	171.0	166.843335
2020-03-01	128.0	198.867367
2020-04-01	189.0	213.103956
2020-05-01	314.0	240.422385
2020-06-01	226.0	223.476042
2020-07-01	189.0	213.134619
2020-08-01	162.0	185.644642
2020-09-01	200.0	194.807987
2020-10-01	195.0	206.337181

SVR

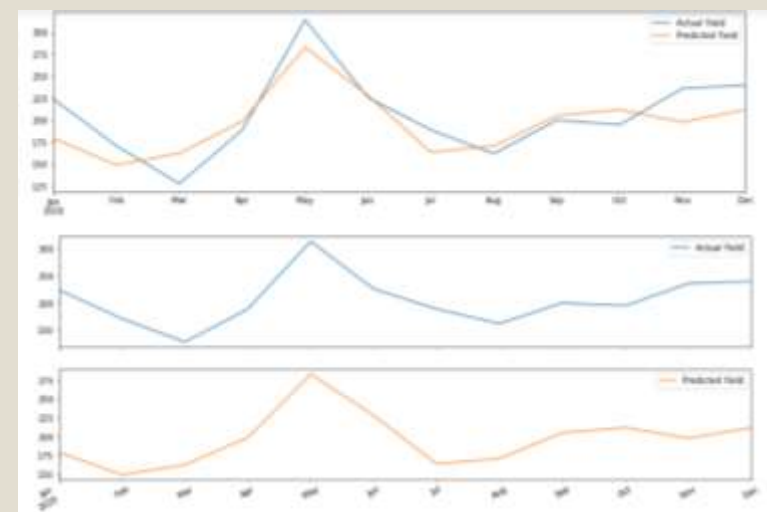
	Actual Yield	Predicted Yield
2020-01-01	224	179.197661
2020-02-01	171	149.181311
2020-03-01	128	162.571922
2020-04-01	189	198.721428
2020-05-01	314	282.845096
2020-06-01	226	228.327316
2020-07-01	189	163.732088
2020-08-01	162	170.698845
2020-09-01	200	205.064914
2020-10-01	195	211.969376
2020-11-01	236	197.891108
2020-12-01	240	211.873551

MLP

Model 1



SVR



MLP

Model 2

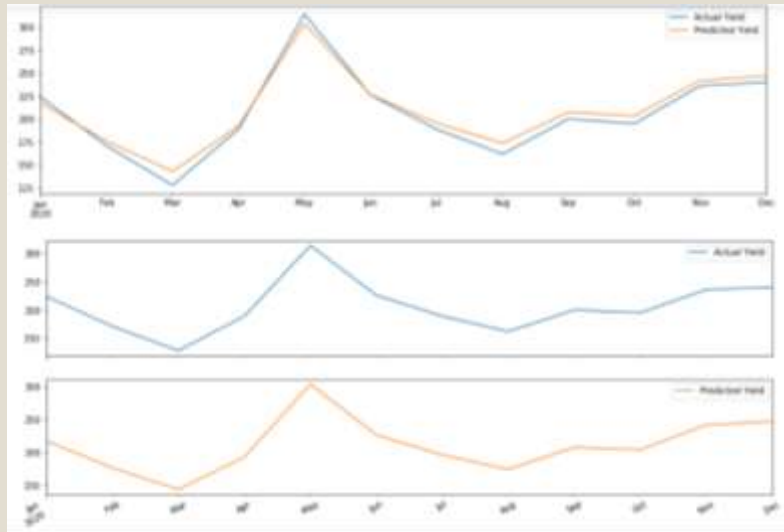
	Actual Yield	Predicted Yield
2020-01-01	224.0	217.133723
2020-02-01	171.0	175.980269
2020-03-01	128.0	143.249043
2020-04-01	189.0	192.408056
2020-05-01	314.0	303.550794
2020-06-01	226.0	226.350917
2020-07-01	189.0	195.704231
2020-08-01	162.0	173.697388
2020-09-01	200.0	207.341321
2020-10-01	195.0	203.539403

SVR

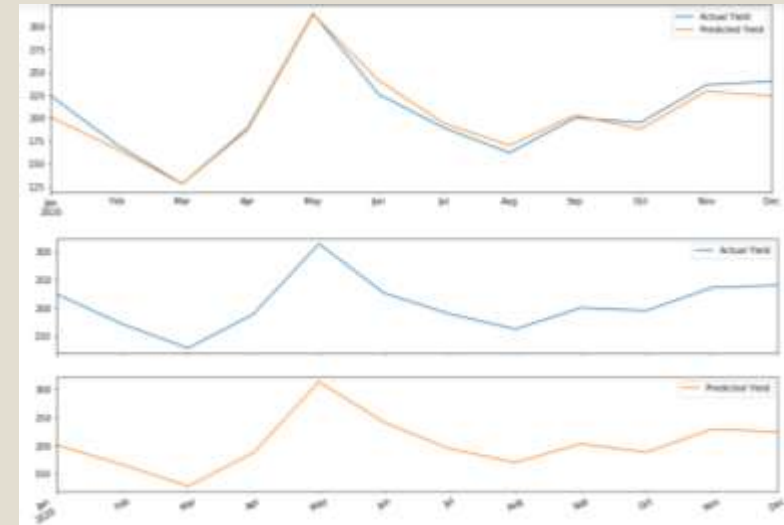
	Actual Yield	Predicted Yield
2020-01-01	224	201
2020-02-01	171	166
2020-03-01	128	128
2020-04-01	189	187
2020-05-01	314	313
2020-06-01	226	241
2020-07-01	189	194
2020-08-01	162	170
2020-09-01	200	203
2020-10-01	195	188

MLP

Model 1



SVR



MLP