Research project

A LARGE-SCALE COMPARISON OF ARTIFICIAL INTELLIGENCE AND DATA MINING (AI&DM) TECHNIQUES IN SIMULATING RESERVOIR RELEASES OVER THE UPPER COLORADO REGION

Presentation by
Sitthatka Jarussang

INTRODUCTION

METHODOLOGY

CONCLUSION

INTRODUCTION

INTRODUCTION

- Reservoirs are man-made for a wide variety of uses.
- In the past, The weather has changed more and more resulting in many disasters Including dam break.
- In the case of a dam failure This indicates the need to develop efficient and flexible tools.

INTRODUCTION

- It has previously developed innovative models using linear programming methods for large scale.
- However, under the limitations set in the model, it is currently not suitable.

INTRODUCTION

- Under limited supplies, the method was changed to an AI&DM model.
- The recommended and suitable model for prediction is regression.
- There are many AI&DM models, each with different advantages and disadvantages.

INTRODUCTION

- However, this research group therefore simulates every model that they study and uses statistical measures to decide for each model which model can make the best predictions under statistical measures.

METHODOLOGY

METHODOLOGY

THIS PAPER

MY STUDY

METHODOLOGY THIS JOURNAL

1. AI&DM

In this Journal, they start with select AI&DM model that they study.

- Linear regression
- The Linear Ridge regression
- support vector regression (SVR)
- Support Vector Machine (SVM)
- KNN-Regression (K=3)
- KNN-Regression (K=10)
- CART
- Random forest
- XGBoost
- Multiple Layer Perceptron (MLP)
- Artificial Neural Network (ANN)
- Long short term Memory model

METHODOLOGY THIS JOURNAL

2. Statistic measurement

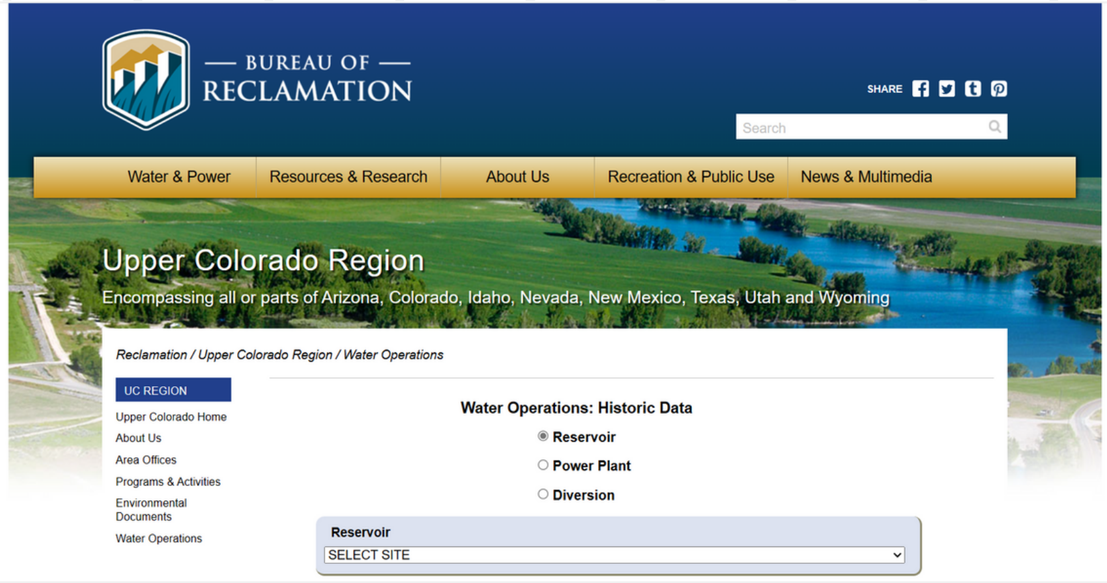
Next Step, they will choose statistic measurement.

- Corralation Coefficient
- Nash-Sutcliffe Model Efficiency Coefficient
- RMSE-observation standard deviation ratio
- Root mean square error
- Kling-Gupta efficiency
- Percent bias

METHODOLOGY THIS JOURNAL

3. Data case

In This cases, they will use data from the Upper Colorado region under the jurisdiction of the U.S. Bureau of Reclamation (USBR), choose 33 reservoir. To select variable for study, They choose Strorage, inflow and outflow.



METHODOLOGY THIS JOURNAL

3. Data case

Initials	Names	Lat	Lon	Data Start Date	Data Length (Years)	Elevation (Meters)
BSR	Big Sandy Reservoir	42.24923	−109.429	1/1/1990	30	2060
CAU	Causey Reservoir	41.29019	−111.583	1/1/1999	21	1745
CRY	Crystal Reservoir	38.45359	−107.335	1/1/1978	42	2251
DCR	Deer Creek Reservoir	40.40667	−111.527	1/1/1987	33	1653
DIL	Dillon Reservoir	39.6074	−106.055	1/1/1985	35	2751
ECH	Echo Reservoir	40.96486	−111.432	1/1/1967	53	1691
ECR	East Canyon Reservoir	40.92053	−111.601	1/1/1992	28	1749
EBR	Elephant Butte Reservoir	33.15349	−107.191	1/1/2007	13	1323
FGR	Flaming Gorge Reservoir	40.91499	−109.422	1/1/1963	57	1828
FON	Fontenelle Reservoir	42.0283	−110.061	1/1/1990	30	1976
GMR	Green Mountain Reservoir	39.8783	−106.33	1/1/1977	43	2406
HNR	Huntington North Reservoir	39.34173	−110.947	1/1/1999	21	1774
HYR	Hyrum Reservoir	41.62663	−111.872	1/1/1999	21	1427
JOR	Jordanelle Reservoir	40.40729	−111.528	1/1/1997	23	1636
JVR	Joes Valley Reservoir	39.28848	−111.269	1/1/1996	24	2129
LCR	Lost Creek Reservoir	41.18417	−111.399	1/1/1998	22	1824
LEM	Lemon Reservoir	37.38171	−107.661	1/1/1965	55	2478
MCP	Mcphee Reservoir	37.57651	−108.572	1/1/1991	29	2073
MCR	Meeks Cabin Reservoir	41.02533	−110.58	1/1/1998	22	2647
MPR	Morrow Point Reservoir	38.4518	−107.538	1/1/1977	43	2184
NAV	Navajo Reservoir	36.80237	−107.613	1/1/1986	34	1801
PIN	Pineview Reservoir	41.25402	−111.843	1/1/1990	30	1495
RFR	Red Fleet Reservoir	40.58028	−109.442	1/1/1989	31	1721
RID	Ridgway Reservoir	38.19918	−107.742	1/1/1990	30	2101
ROC	Rockport Reservoir	40.79	−111.404	1/1/1969	51	1807
RUE	Ruedi Reservoir	39.3631	−106.818	1/1/1980	40	2349
SCO	Scofield Reservoir	39.7862	−111.119	1/1/1996	24	2338
SJR	Silver Jack Reservoir	38.23207	−107.543	1/1/1992	28	2725
STA	Starvation Reservoir	40.18876	−110.444	1/1/1981	39	1700
STE	Steinaker Reservoir	40.50567	−109.531	1/1/1976	44	1655
TPR	Taylor Park Reservoir	38.818	−106.607	1/1/1963	57	2847
USR	Upper Stillwater Reservoir	40.56	−110.699	1/1/1991	29	2445
VAL	Vallecito Reservoir	37.37775	−107.575	1/1/1986	34	2318

METHODOLOGY THIS JOURNAL

3. Data case

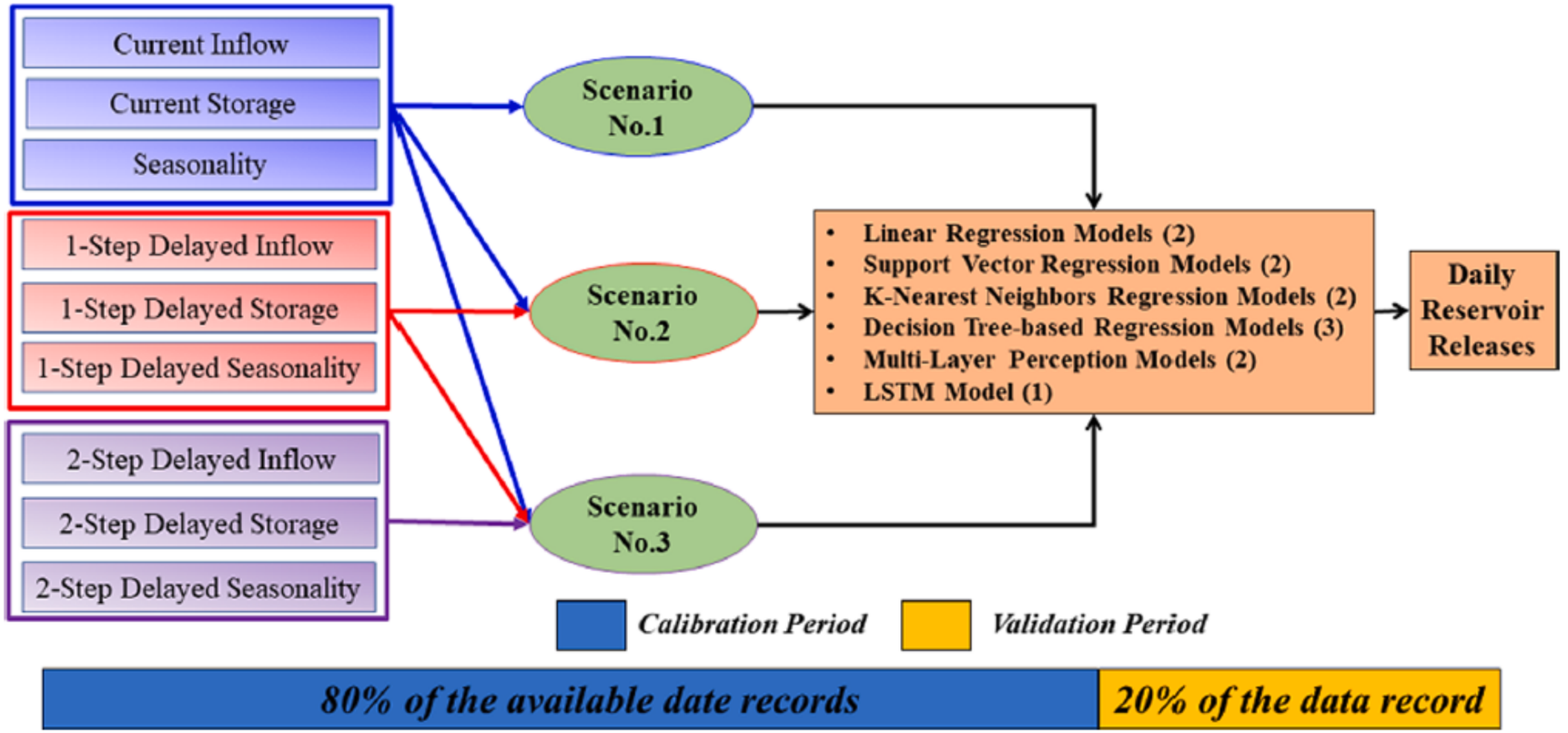


Fig. 2. Experiment Design.

METHODOLOGY THIS JOURNAL

4.Result report

Reservoir	Statistic measure 1	Statistic measure 2	...	Statistic measure nth
Reservoir 1	Linear	CART	...	...
Reservoir 2	KNN	KNN	...	...
.	...	...	...	...
.	...	...	...	...
.	...	...	...	...
Reservoir nth	...	...	...	...

To decision rule, The largest number of methods in the table is the best method.

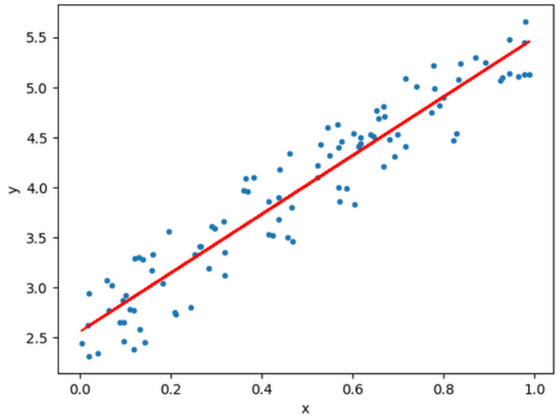
METHODOLOGY

METHODOLOGY MY STUDY

1. AI&DM

For Example in my study, I will show 4 methods including:

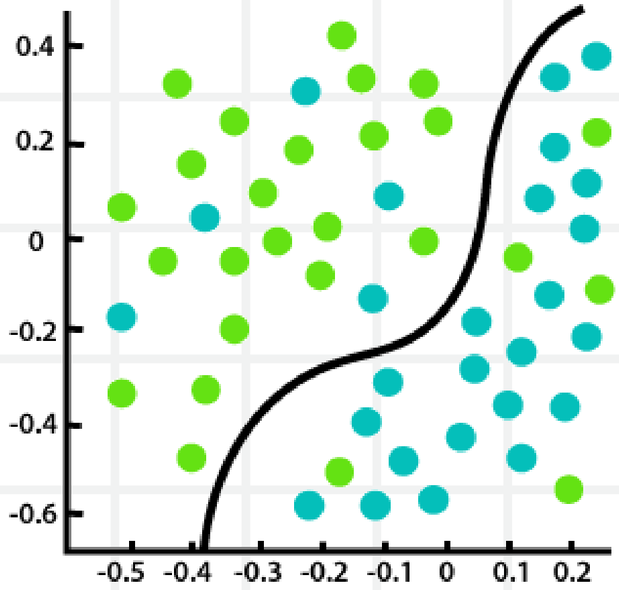
- Linear regression
- KNN-Regression (K=3)
- CART
- support vector regression (SVR)



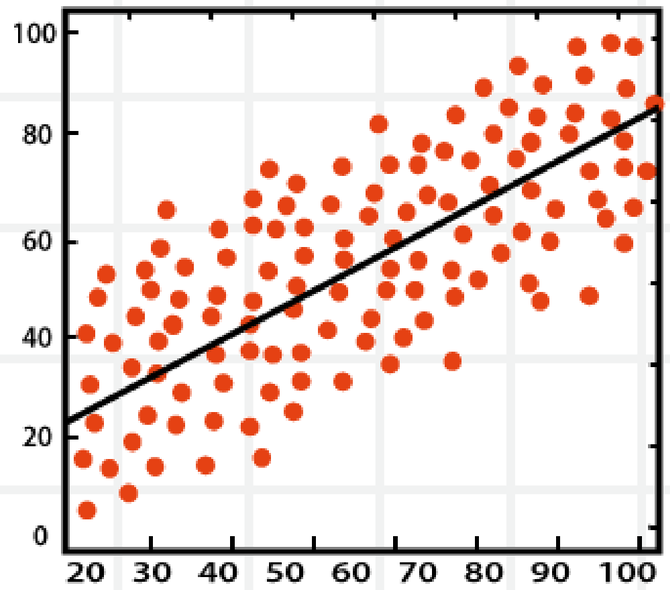
METHODOLOGY MY STUDY

1. AI&DM

Classification vs Regression



Classification



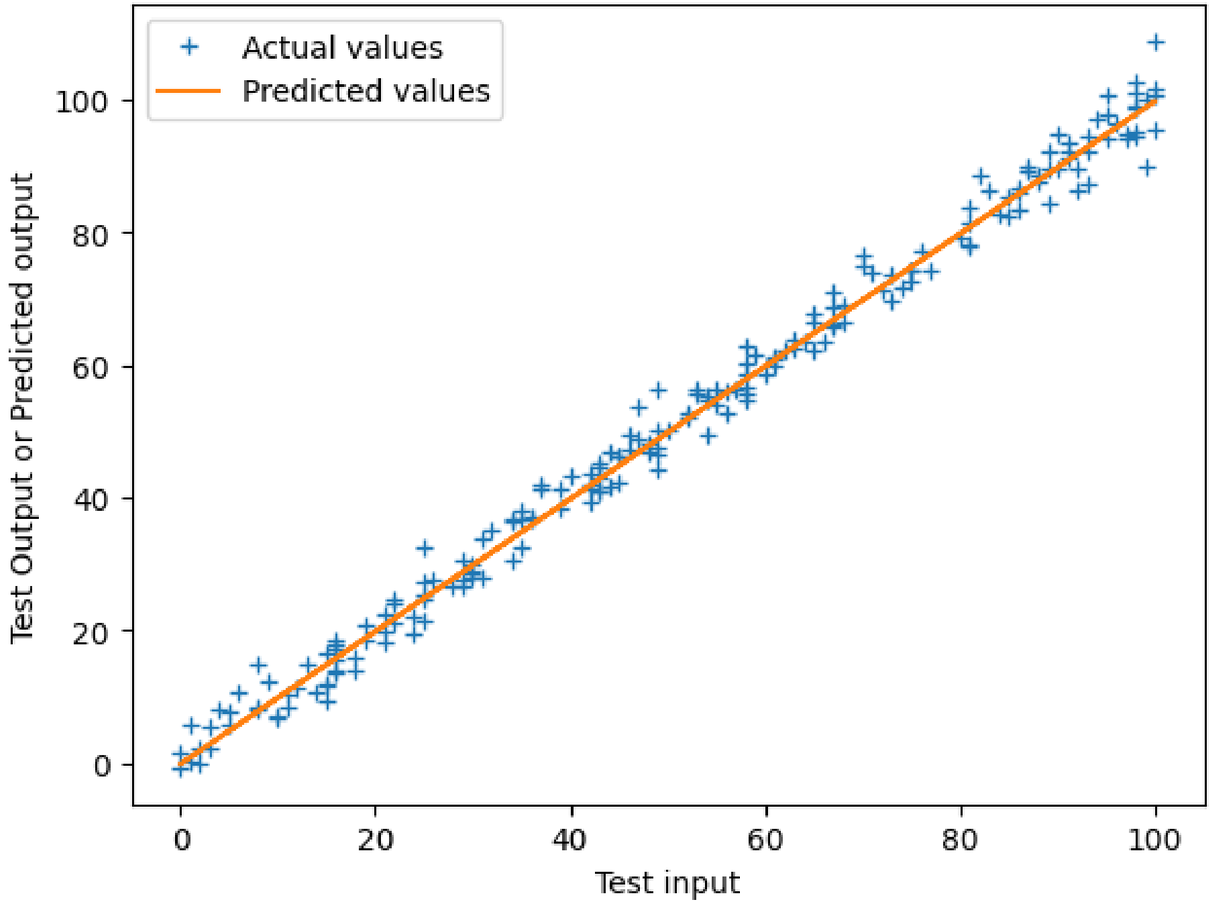
Regression

METHODOLOGY MY STUDY

1. AI&DM
Linear regression

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data.

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_n x_{n,i} + \varepsilon$$



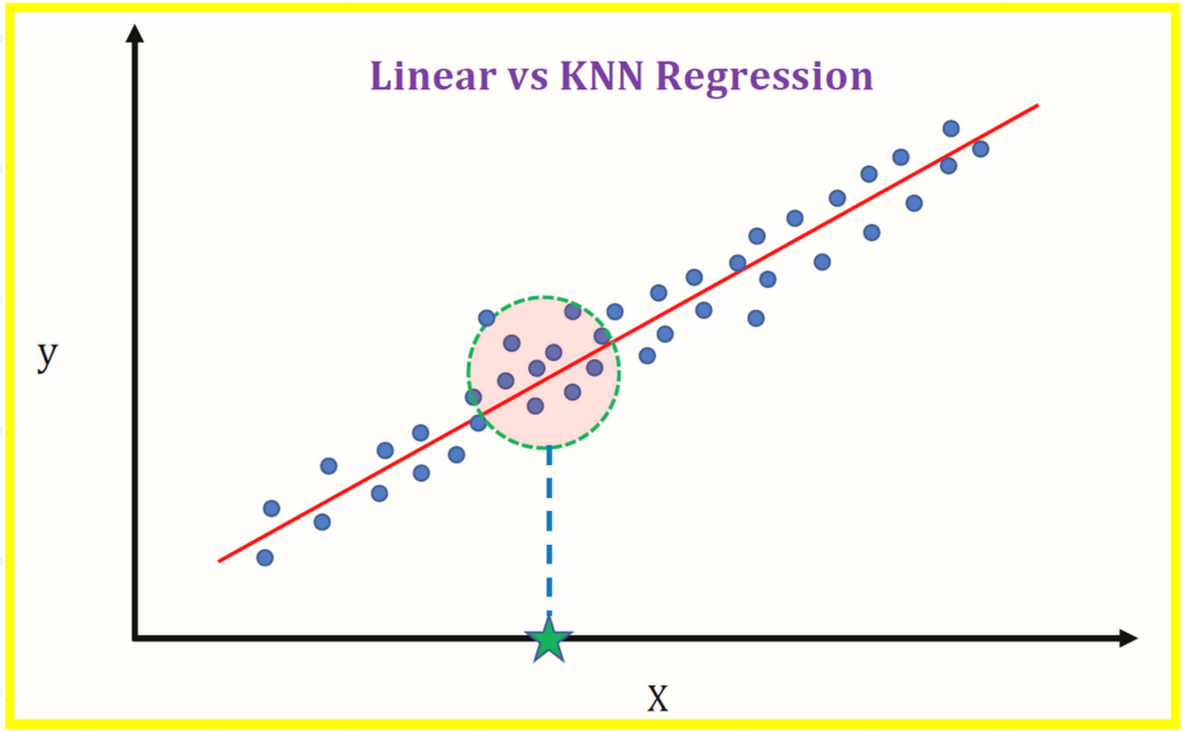
METHODOLOGY

MY STUDY

1. AI&DM
KNN regression

In KNN regression, the algorithm works by finding the K-nearest data points in the training dataset to a given input (the data point you want to make a prediction for).

To finding nearest data, it will be considering number of k.



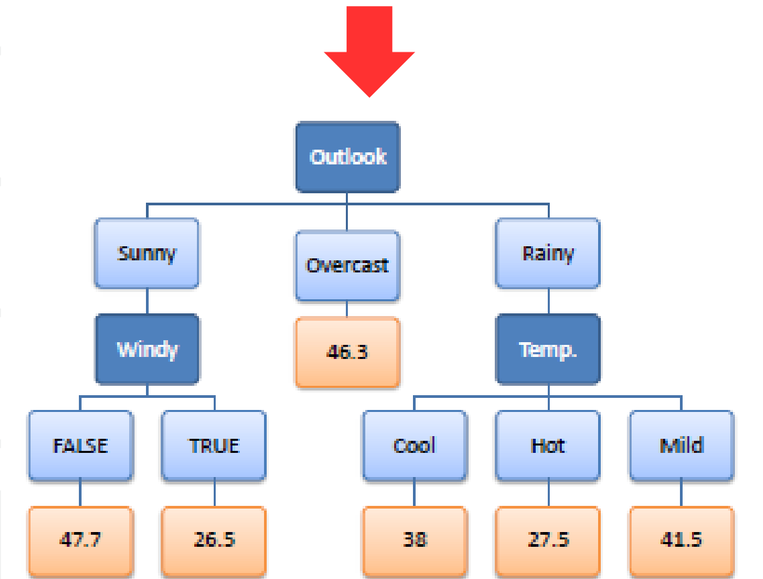
METHODOLOGY

MY STUDY

1. AI&DM
CART

CART, which stands for "Classification and Regression Trees," is a versatile and widely used machine learning algorithm that is used for both classification and regression tasks.

Predictors				Target
Outlook	Temp.	Humidity	Windy	Hours Played
Rainy	Hot	High	False	25
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	False	62
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	36
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	48
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	62
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30



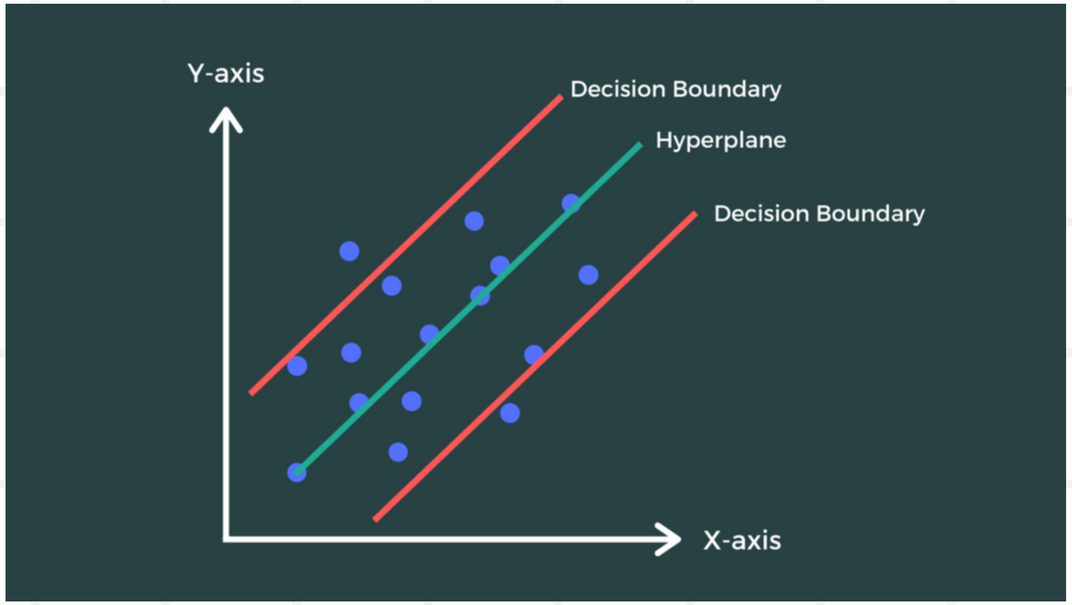
METHODOLOGY

MY STUDY

1. AI&DM

support vector regression (SVR)

SVR is designed to predict continuous (numerical) values rather than class labels and is particularly useful when the relationship between input features and the target variable is non-linear.



METHODOLOGY MY STUDY

2. Statistic measurement

For Example in my study, I will choose statistic measurement Including.

Corralation Coefficient

Root mean square error

RMSE-observation standard deviation ratio

METHODOLOGY MY STUDY

2. Statistic measurement
Classification vs Regression

Regression

- Mean absolute error (MAE)
- Mean squared error (MSE)
- R squared (R^2)
- Adjusted R squared (Adj-R^2)

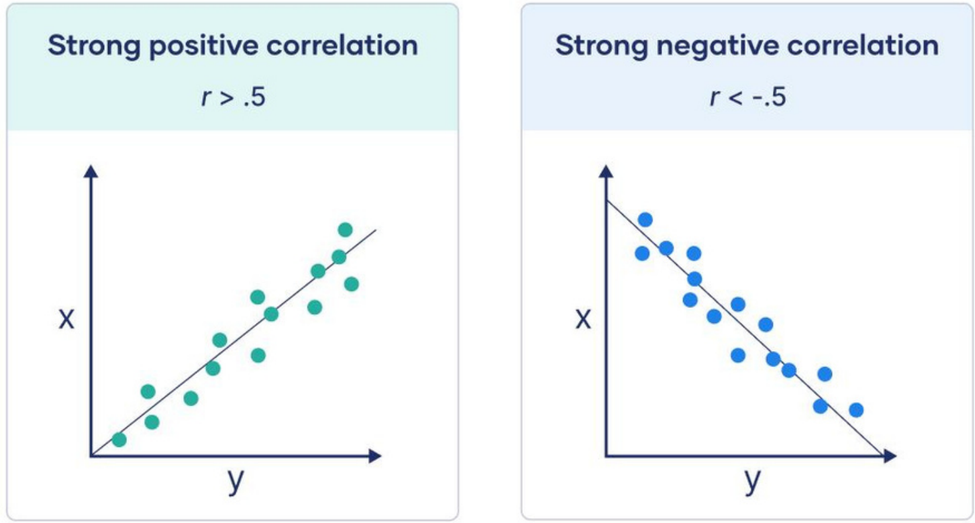
Classification

- Accuracy
- Precision
- Recall
- Area under curve (AUC)
- Confusion matrix

METHODOLOGY MY STUDY

2. Statistic measurement
Correlation Coefficient

The correlation coefficient is a statistical measure that quantifies the degree to which two variables are related or associated with each other. It assesses the strength and direction of the linear relationship between two variables.

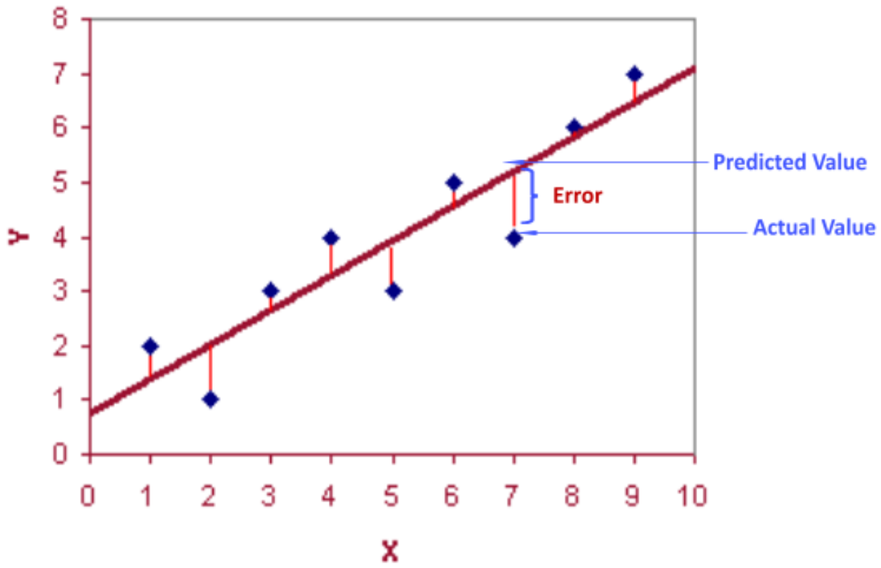


The good model with correlation should be highest value with ignore symbol.

METHODOLOGY MY STUDY

2. Statistic measurement
Root mean square error (RMSE)

RMSE stands for Root Mean Square Error, and it is a widely used metric in statistics and machine learning to measure the accuracy of a predictive model, particularly in regression tasks.



The good model with RMSE should be lowest value .

METHODOLOGY MY STUDY

2. Statistic measurement
RMSE-observation
standard deviation ratio(RSR)

RMSE-observation standard deviation ratio is a base-on RMSE considering the standard deviation of observed data.

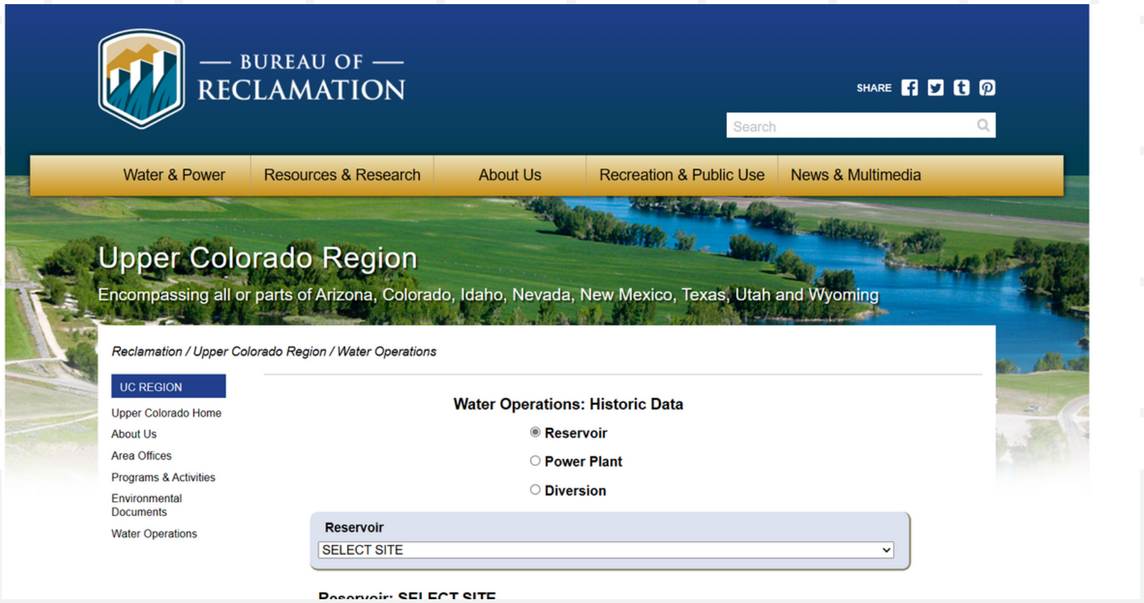
Evaluation Level	RSR
Very good	$0.00 \leq RSR \leq 0.50$
Good	$0.50 < RSR \leq 0.60$
Satisfactory	$0.60 < RSR \leq 0.70$
Unsatisfactory	$RSR > 0.70$

The good model with RSR should be lowest value .

METHODOLOGY MY STUDY

3. Data case

For Example in my study, I will choose data from the Upper Colorado region under the jurisdiction of the U.S. Bureau of Reclamation (USBR), choose 5 reservoir. To select variable for study, They choose Strorage, inflow and outflow.



METHODOLOGY MY STUDY

3. Data case

Initials	Names	Lat	Lon	Data Start Date	Data Length (Years)	Elevation (Meters)
BSR	Big Sandy Reservoir	42.24923	−109.429	1/1/1990	30	2060
CAU	Causey Reservoir	41.29019	−111.583	1/1/1999	21	1745
CRY	Crystal Reservoir	38.45359	−107.335	1/1/1978	42	2251
DCR	Deer Creek Reservoir	40.40667	−111.527	1/1/1987	33	1653
DIL	Dillon Reservoir	39.6074	−106.055	1/1/1985	35	2751
ECH	Echo Reservoir	40.96486	−111.432	1/1/1967	53	1691
ECR	East Canyon Reservoir	40.92053	−111.601	1/1/1992	28	1749
EBR	Elephant Butte Reservoir	33.15349	−107.191	1/1/2007	13	1323
FGR	Flaming Gorge Reservoir	40.91499	−109.422	1/1/1963	57	1828
FON	Fontenelle Reservoir	42.0283	−110.061	1/1/1990	30	1976
GMR	Green Mountain Reservoir	39.8783	−106.33	1/1/1977	43	2406
HNR	Huntington North Reservoir	39.34173	−110.947	1/1/1999	21	1774
HYR	Hyrum Reservoir	41.62663	−111.872	1/1/1999	21	1427
JOR	Jordanelle Reservoir	40.40729	−111.528	1/1/1997	23	1636
JVR	Joes Valley Reservoir	39.28848	−111.269	1/1/1996	24	2129
LCR	Lost Creek Reservoir	41.18417	−111.399	1/1/1998	22	1824
LEM	Lemon Reservoir	37.38171	−107.661	1/1/1965	55	2478
MCP	Mcphee Reservoir	37.57651	−108.572	1/1/1991	29	2073
MCR	Meeks Cabin Reservoir	41.02533	−110.58	1/1/1998	22	2647
MPR	Morrow Point Reservoir	38.4518	−107.538	1/1/1977	43	2184
NAV	Navajo Reservoir	36.80237	−107.613	1/1/1986	34	1801
PIN	Pineview Reservoir	41.25402	−111.843	1/1/1990	30	1495
RFR	Red Fleet Reservoir	40.58028	−109.442	1/1/1989	31	1721
RID	Ridgway Reservoir	38.19918	−107.742	1/1/1990	30	2101
ROC	Rockport Reservoir	40.79	−111.404	1/1/1969	51	1807
RUE	Ruedi Reservoir	39.3631	−106.818	1/1/1980	40	2349
SCO	Scofield Reservoir	39.7862	−111.119	1/1/1996	24	2338
SJR	Silver Jack Reservoir	38.23207	−107.543	1/1/1992	28	2725
STA	Starvation Reservoir	40.18876	−110.444	1/1/1981	39	1700
STE	Steinaker Reservoir	40.50567	−109.531	1/1/1976	44	1655
TPR	Taylor Park Reservoir	38.818	−106.607	1/1/1963	57	2847
USR	Upper Stillwater Reservoir	40.56	−110.699	1/1/1991	29	2445
VAL	Vallecito Reservoir	37.37775	−107.575	1/1/1986	34	2318

4. step to get result

I will show in python

CONCLUSION

CONCLUSION

comclusion

	Reservoir	Good Method consider Correlation	Good Method consider RMSE	Good Method consider RSR
0	Fontenelle	CART	CART	CART
0	Crystal	KNN	KNN	KNN
0	Lemon	CART	CART	CART
0	Rockport	Linear	Linear	Linear
0	Upper Stillwater	CART	CART	CART

For this study, We can conclude that by the 4 methods and 3 statistic measurements, the good prediction is CART

CONCLUSION

Discussion

1. If we have many reservoirs, More Models, or More Statistic measurement. it can be not same result.
2. We can see the pattern each reservoirs is same model. you can use that model to predict outflow from each reservoirs.

	Reservoir	Good Method consider Correlation	Good Method consider RMSE	Good Method consider RSR
0	Fontenelle	CART	CART	CART
0	Crystal	KNN	KNN	KNN
0	Lemon	CART	CART	CART
0	Rockport	Linear	Linear	Linear
0	Upper Stillwater	CART	CART	CART

THANK YOU