Prediction of RPO by Model Tree

Saleh Shahinfar
Department of Dairy Science
University of Wisconsin-Madison
JAM2013

Introduction:

- Replacement decisions have a major impact on dairy farm profitability.

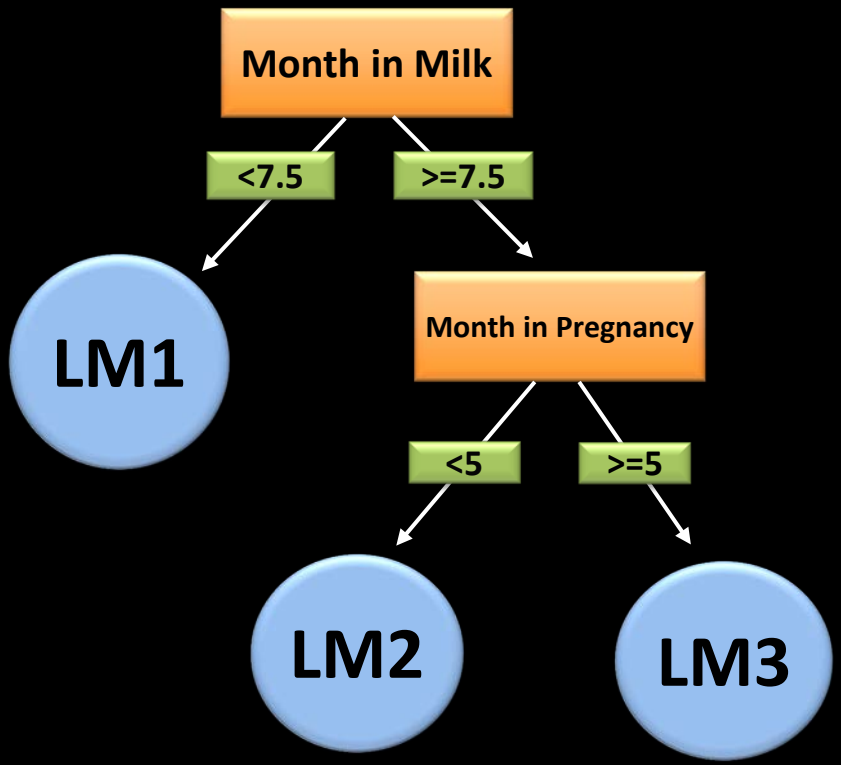
- Retention Pay-Off (RPO): the expected profit from keeping the cow compared with immediate replacement. Rank animals and make the final replacement decisions (De Vries, 2004).

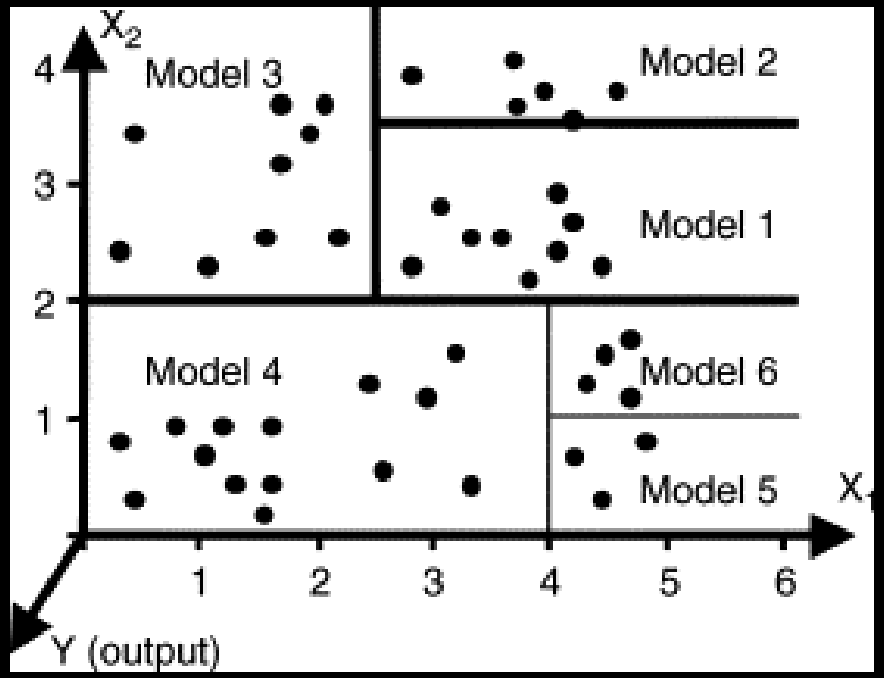
- Dynamic Programming (DP): uses the *divide and conquer* algorithm to find the optimal replacement decisions. The technique breaks a large multi-stage problem into series of single-stage problems and solves them one by one.

- DP has been widely studied to find the optimal replacement policies in dairy cattle, but they are complex and therefore not extensively applied in farm decision-making.

- Machine Learning (ML): Provide fast and accurate predictions of non-linear and inter-correlated variables, and therefore a valuable method for both researchers and producers.
- ML can mimic the behavior of a dynamic programming model, and predict RPO very accurately.

- Model Tree is a decision tree that are used for numeric predictions. They are similar to decision trees except in each leaf they store a linear model that predicts the class value of the instances that reach the leaf.





Data:

- 122716 simulated records as follows.

Features
Milk production (1, 2, 3)
Milk price (1, 2, 3)
Replacement cost (1, 2, 3)
Milk class (1-5)
Lactation number (1-9)
Month in milk (1-20)
Pregnancy status (0-9)
Output
Retention Pay Off ($)

Levels	1	2	3
Variation	-20%	Base	+20%
Milk Production	18000	22000	26000
Milk Price	0.12	0.16	0.2
Replacement costs	1000	1300	1600

- %66 data was used as Train set and %34 as Test set.

Relative Absolute Error:

- 1- Measures absolute error and is not affected by outliers.
- 2- It considers the relative magnitude of the error compared with the prediction.

- $$Relative\ Absolute\ Error = \sum_{i=1}^{n} \frac{|p_i - a_i|}{|a_i - \bar{a}|}$$

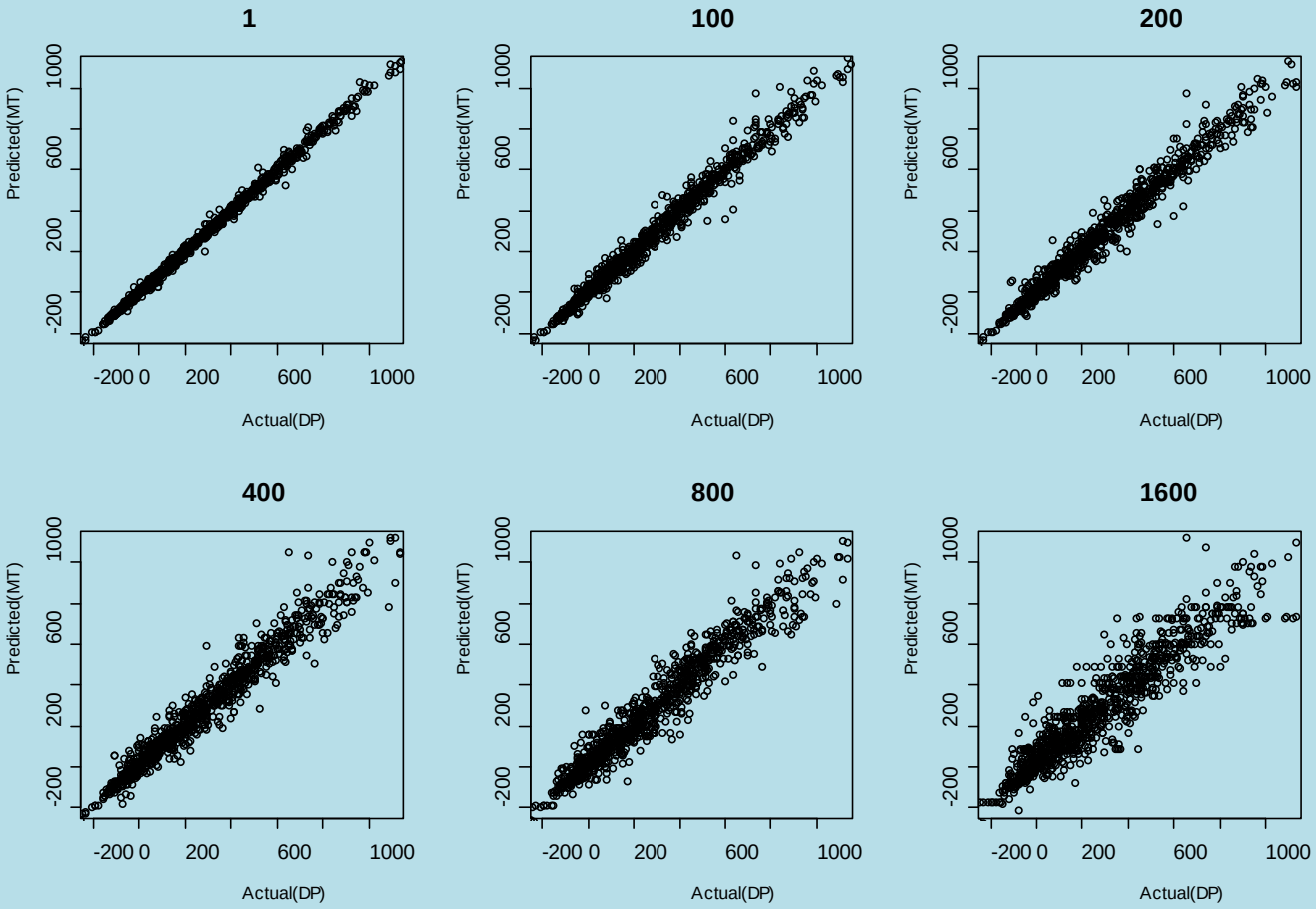
- Where p_i is predicted value, a_i is actual value and $\bar{a}$ is the prediction by an arbitrary predictor, here average of actual values (Witten and Frank, 2005).

Results:

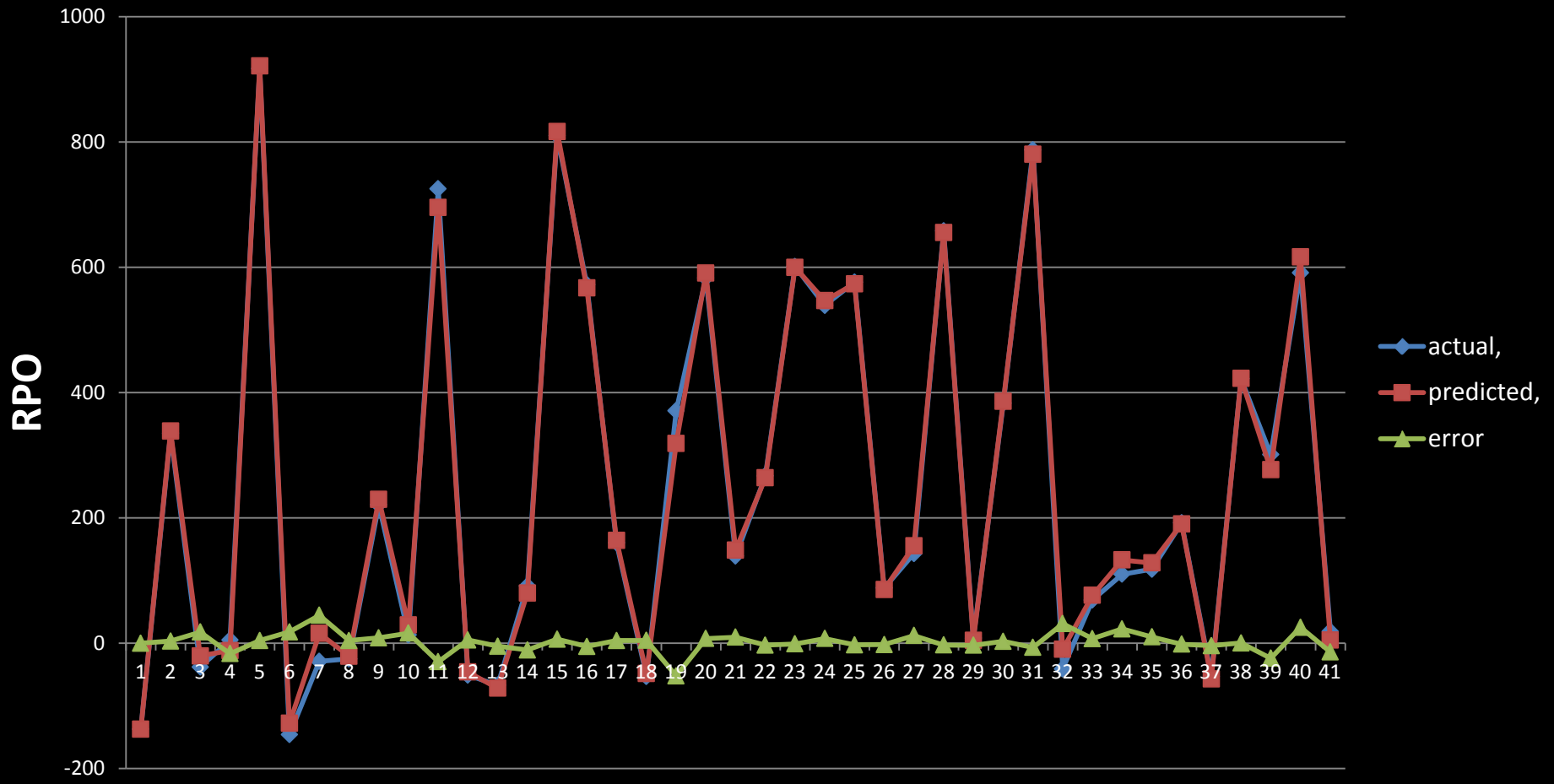
Performance of Model Trees in prediction of RPO with different constraints (minimum numbers of instances per Linear Model).

Min Numbers of Instances per LM	Number of Rules	Correlation Coefficient	Root Mean Square error	Relative Absolute Error %
1	2447	0.9978	19.48	5.08
100	**204**	**0.9907**	**39.88**	**10.93**
200	152	0.9848	50.85	14.15
400	89	0.9763	63.45	17.91
800	47	0.9658	76.02	22.34
1600	28	0.9364	102.81	30.60

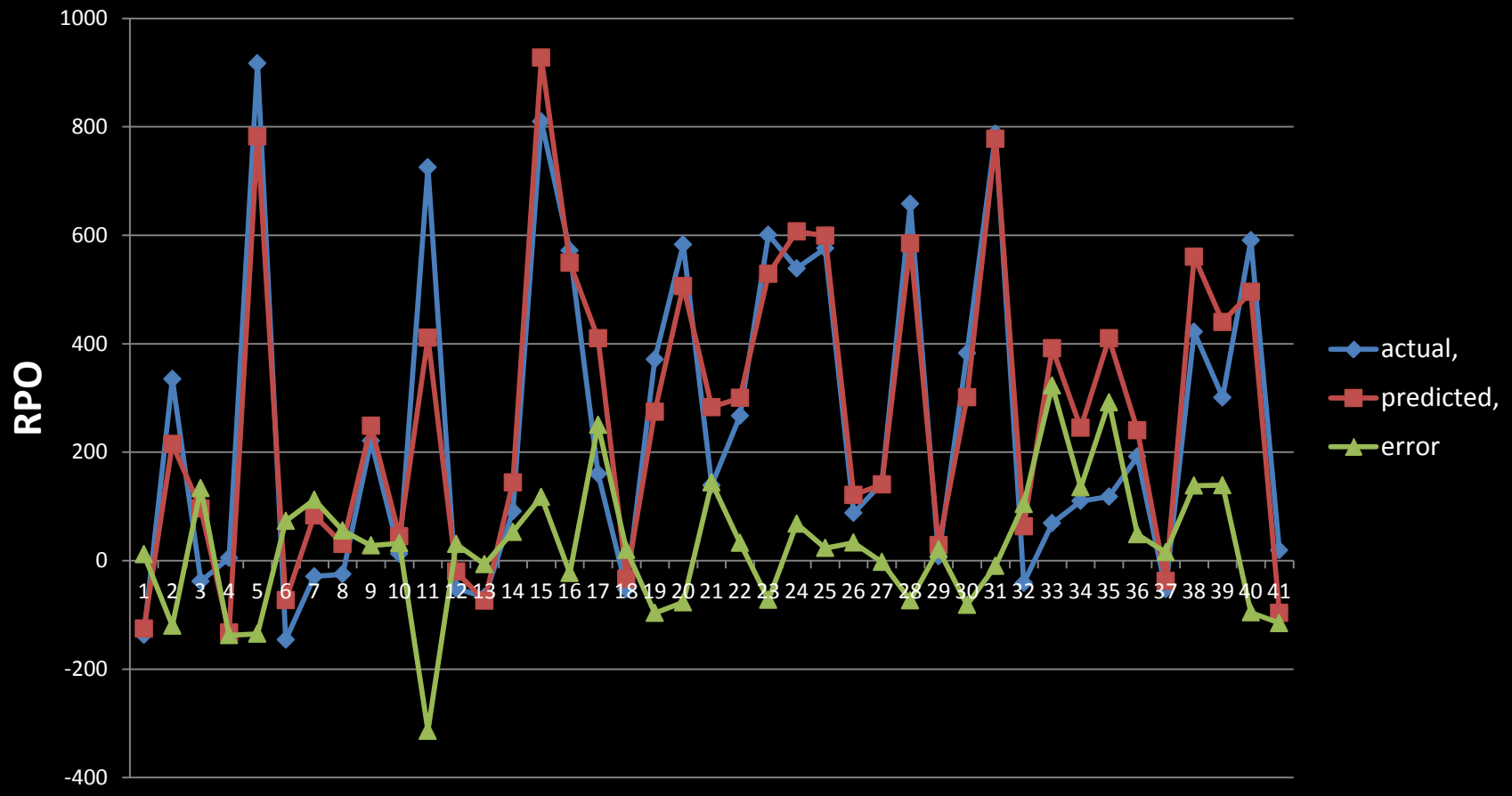
Figure 2: Actual (predicted by DP) vs. Predicted (predicted by MT) values of RPO for six different scenarios showed in table 1, for the simulated test set .



Actual (predicted by DP), Predicted (predicted by MT) values of RPO and error, for 40 randomly selected instances from the test set, predicted by 1 instance per LM constraints.



Actual (predicted by DP), Predicted (predicted by MT) values of RPO and error, for 40 randomly selected instances from the test set, predicted by 1600 instance per LM constraints.



Predicted (MT) vs. Actual (DP) deciles of RPO3 for 41,723 test cases.

		Actual Deciles									
		1	2	3	4	5	6	7	8	9	10
Predicted Deciles	1	**3,776**	375	21	0	0	0	0	0	0	0
	2	385	**3,257**	491	34	5	0	0	0	0	0
	3	10	528	**3,057**	535	30	10	2	0	0	0
	4	1	12	584	**2,893**	650	31	1	0	0	0
	5	0	0	18	689	**2,816**	636	11	2	0	0
	6	0	0	1	20	653	**2,891**	576	30	1	0
	7	0	0	0	1	17	576	**3,004**	545	29	0
	8	0	0	0	0	1	23	550	**3,101**	487	10
	9	0	0	0	0	0	5	27	485	**3,316**	339
	10	0	0	0	0	0	0	1	9	339	**3,826**

Confusion matrix for percentage of binary prediction of RPO for 41,723 test cases.

		Actual (Dynamic Programming)	
		Positive	Negative
Predicted (Model Tree)	Positive	0.91	0.09
	Negative	0.01	0.99

Results from Real data :

- A correlation coefficient of **0.994** and relative absolute error rate of **0.10** was obtained by applying the trained model for *prediction of RPO* to 102 actual voluntary culling records from UW-Madison dairy herd, which indicates its accuracy and relevance of this method in real-life replacement decisions.

- An executable version of the program soon will be available on-line at the University Of Wisconsin Dairy Management Website (DairyMGT.info) which can be used to evaluate animals on commercial herds.

- Questions!
- Comments!
- Discussions!