An Interval Tree Approach to Predict Forest Fires using Meteorological Data

Dima Alberg
Department of Industrial
Engineering and Management,
SCE - Shamoon College of
Engineering,
Beer-Sheva, Israel

ABSTRACT
Interval prediction can be more useful than single value prediction in many continuous data streams. This paper introduces a novel Interval Prediction Tree IP3 algorithm for interval prediction of numerical target variables from temporal mean-variance aggregated continuous data. This algorithm characterized by: processing incoming mean-variance aggregated multivariate temporal data, splitting each of the continuous features of the input according to the best mean-variance and making stable interval predictions of a target numerical variable with a given degree of statistical confidence. As shown by empirical evaluations in forest fires data set the proposed method provides better performance than existing regression tree models.

General Terms
Data Mining, Regression Tree, Prediction Algorithms.

Keywords
Interval Prediction, Mean-Variance Aggregation, Prediction Tree, Forest Fires.

1. INTRODUCTION
In many data streams, the data is available as time-continuous statistical moments such as mean or variance that are calculated over pre-defined measurement intervals, rather than as raw values sampled at discrete points in time. Examples of such aggregated data streams include meteorological data, financial data, process control systems, and sensor networks. For example, a meteorological station may be continuously storing aggregated mean and variance estimators for a large number of meteorological attributes at predefined time intervals (such as every 10 minutes). However, reporting single predicted values for the mean response values of new measurement intervals can be misleading. The reason is that due to a large unexplained variance of the target variable, in many intervals the actual mean values may be very different from any specific point estimation. In this paper, the attention was shifted from predicting a single mean value to predicting intervals, which are expected to contain the actual mean values with a given probability. The above considerations cause a need for a stable algorithm that can process incoming mean-variance aggregated multivariate temporal data and makes stable interval predictions of a target numerical variable, with a given degree of statistical confidence. This work contributes to the field of mining massive temporal data sets and continuous data streams by introducing Interval Prediction Tree IP3 algorithm, which builds compact and stable interval-based prediction tree models of numerical output variables using aggregated statistical moments of numerical input attributes. The paper is organized as follows: the related work is described in the next section, while in Section 3 the IP3 algorithm methodology is introduced and represented. In Section 4, experimental results are reported for forest fires data set. Finally, Conclusion section discusses the main features of the proposed algorithm and summarizes the main experimental findings.

2. RELATED WORK
Most batch regression and tree models for predicting numerical variables, such as MARS [7], CART [3], RETIS [10], M5 [12], M5P[14], SMOTI [5], MAUVE [13], MOPT [1], GUIDE [11], and FIMT [9], are not designed for aggregated temporal data. Hence, they cannot utilize the relationship between multiple statistical moments (such as mean and standard deviation) of aggregated numerical attributes. Actually, most of the regression tree algorithms apply binary recursive partitioning binary, since the nodes are always split into two child nodes, and recursive, because the process is repeated at every node. It is also possible to split the data into three or more subsets or child nodes. Regression trees provide quite simple and easily interpreted regression models with reasonable accuracy. However, according to Breiman et al. [4], these methods are known for their split instability. Finally, the interested reader may find a more detailed survey of regression tree methods in [1].

The IP3 Interval Prediction Tree methodology presented in this paper extends the main idea of the traditional interval regression trees algorithms and includes the following principal enhancements: First, in the case of numerical attributes traditional interval regression trees algorithms calculates every possible splitting point by using the recursive least squares algorithm. This task is computationally expensive and it has a negative effect on the scalability of the algorithm. The proposed IP3 algorithm avoids the computationally intensive and memory exhaustive sorting operation by a more simple and accurate non-sorting procedure. The IP3 tree stopping criterion is significantly extended by two stopping rules. The first stopping rule is applied by the algorithm when the selected terminal node instances are normally distributed. The second stopping rule is applied when the terminal node instances are within a constant unbiased prediction interval.

3. INTERVAL PREDICTION TREE
The proposed method is based on the assumption that input and output variables in an aggregated data stream are characterized by linear or nonlinear dependencies (or both), which can be represented using the proposed IP3 model. The proposed algorithm differs from currently described state-of-the-art regression tree algorithms such as CART, RETIS, M5,

M5P, SMOTI, MAUVE, GUIDE and FIMT by the following characteristics:

- The use of synchronous mean and variance unbiased estimators of numerical features.
- Node splitting based on the Mahalanobis distance between the two statistical estimators.
- Novel representation of prediction intervals at the tree leaves.

Both statistics can be used as candidate predictive features by a prediction tree induction algorithm. Therefore, if the two represented statistics indeed exhibit independent and identical behavior, then the aggregated input variable can be represented within a robust two-tail prediction interval at a user-defined confidence level, such as 95%. An IP3 tree segment example is displayed in Figure 1.

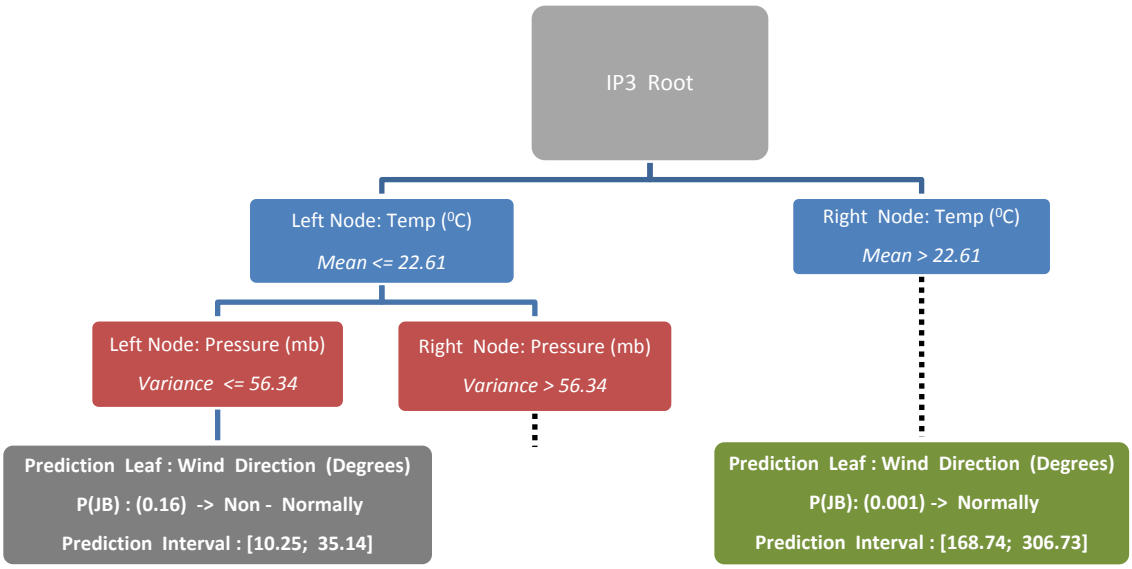


Figure 1: The IP3 tree representation

Finally, he suggested approach enables one to utilize predictive feature information obtained from mean and variance of temporally aggregated instances This approach also enables one to achieve a considerable reduction in the depth of the induced prediction tree by using interval prediction tree leaves.

3.1 The IP3 Construction Procedure

The pseudocode in Figure 2 identifies the best split for predicting the mean and variance of a numerical target variable. This procedure applies to splitting the values of a bivariate numerical input variable X with respect to the target variable Y, where both variables X and Y are represented by the sample mean and variance, according to the predefined temporal resolution r. Generally, the splitting procedure includes five principal steps, where the first three steps perform Mahalanobis distance calculations between bivariate mean-variance input and target variables, while the last two steps identify the best splitting feature (mean or variance) and the corresponding splitting threshold.

The first step consists of the Mahalanobis distance calculation M(X) for the numerical input variable in each aggregated instance using AVG(X) and VAR(X) and the chi - square outlier test detection procedure. Outlying instances are ignored by the algorithm in order to split the regression tree only on values generated by stable data points. In IP3 algorithm, data point stability is measured by MXMY, which is the Mahalanobis distance between input and target variables. A high Mahalanobis distance is an indication of instability, and vice versa. Thus, the best splitting aggregated instance should minimize Mahalanobis distance MXMY. The best (most stable) predictive feature (sample mean or variance) for the selected instance should have a minimal contribution to the value of the Mahalanobis distance MXMY. This means that it should maximize the absolute difference between the value of MXMY and the values of input estimators (mean and variance of X).

The second step consists of the Mahalanobis distance calculation M(Y) for the target numerical variable in every aggregated instance using AVG(Y).

The third step of the algorithm is the evaluation of all candidate splits (represented by the values of the input variable X in aggregated instances, which are not outliers) and selecting the best splitting aggregated instance (having the minimum Mahalanobis distance MXMY).

The fourth step is aimed at selecting the best predictive feature (sample mean or variance) of a given input variable in the selected splitting instance. In this step, according to the previously identified non-outlying instances of MXMY, the algorithm recalculates the value of the target numerical variable M(Y) in the best splitting instance of X Temp_Min.

Input:	The mean-variance aggregated input variable, X Target variable, Y
Output:	The best mean or variance contributor for input attribute X The best split point for input attribute X

```
Begin
For every instance with distinct values in input X Do                                   (1)
    Calculate the Mahalanobis distances vector MX, where MX = M(AVG(X),VAR(X))
Next
For every instance in target Y Do                                                       (2)
    Calculate the Mahalanobis distances vector MY, where MY = M(AVG(Y),VAR(Y))  Next
If MX is not outlier                                                                    (3)
   Calculate the Mahalanobis distances vector MXMY, where MXMY = M(M(X), M(Y))
   If MXMY is not outlier
      Temp_Min=Min {MXMY}
   End If
End If
For every non-outlying instances MXMY in target Y Do                                    (4)
    Recalculate the MY vector value in Temp_Min instance
    [MY = M(AVG(Y),VAR(Y))]
Next
With Temp_Min instance                                                                  (5)
     MAVG = M(M(Y), AVG (X))
     MVAR = M(M(Y), VAR (X))
     Best_Contributor = (Max(|MXMY - MAVG |, |MXMY - MVAR|))
End With
Return IP3 Best_Split for Best_Contributor
End
```

Figure 2: IP3 splitting criterion pseudo-code

Finally, in the final fifth step, the algorithm calculates the absolute differences between the value of MXMY and the values of the input estimators (mean and variance of X) in the best splitting instance of X Temp_Min, and chooses the best node estimator Best_Contributor that maximizes that difference. If the number of outliers is equal to the number of aggregated instances in the training set, the algorithm ignores a given input variable and shifts to the next variable, or stops the tree construction if there are no remaining input variables.

This subroutine can mitigate the shortcomings of classical regression tree split methods because it is based on correlations between bivariate mean-variance input and target variables, by which different patterns can be identified and analyzed. Furthermore, this subroutine is a useful way to determine similarity between temporally aggregated data sets. It differs from the existing regression tree split methods in being scale-invariant and by taking into account the correlations of the temporally bivariate mean-variance aggregated data sets, subsequently choosing the best node estimator.

4. FOREST FIRES DATA SET

The Forest Fire Data Set from the Montesinho Natural Park in northern Portugal is available at the UCI KDD Archive (http://www.ics.uci.edu). The final data set contains 517 numerical instances collected by Cortez and Morais [6] from January 2000 to December 2003, and is integrated from two different databases. The data in the first database, which contains 517 instances, was collected on a daily basis every time a forest fire occurred. Each instance in the first database represents a forest fire occurrence that has following numerical Fire Weather Index (FWI) attributes: Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC) and Initial Spread Index (ISI). The first three indices are related to fuel codes: the FFMC denotes the moisture content of surface litter that influences ignition and fire spreading, while the DMC and DC represent the moisture content of shallow and deep organic layers, which affect the fire's intensity. The ISI represents a score that correlates with the velocity of fire spreading. The timestamp in the first database represents the day of a specific forest fire occurrence. The XC and YC attributes represent the X and Y spatial coordinates.

The target (prediction) attribute in the Forest Fire Data Set is the total burned forest area (AREA). To reduce skewness and improve symmetry, the logarithm function $y = \ln(x + 1)$, which is a common transformation that tends to improve regression results for right-skewed targets, was applied by

Cortez and Morais [6] to the numerical total burned area (AREA) target attribute.

Finally, in order to compare the obtained results with results presented by the Cortez and Morais [6], and to draw inferences about the impact of fire and weather attributes, four distinct feature sets were evaluated: STFWI - using spatial, temporal, and the four FWI components (TST, XC, YC, FFMC, DMC, DC, and ISI); STM - with the spatial, temporal, and four weather variables (TST, XC, YC, TMP, RH, WS, and RN); FWI - using only the four FWI components (FFMC, DMC, DC, and ISI); and M - with the four weather conditions (TMP, RH, WS, RN).

The results in Table 1 show that under RMSE and RMAE criteria the SVM, M5P, Bagging M5P, MOPT and IP3 models statistically outperform Additive Regression, B-RepTree, M5Rules, NN-MLP, MOPT, RepTree, and Retis-M algorithms in the STFWI and STM data sets.

Table 1: STFWI and STM data set learners comparison (10 time 10 fold cross validation)

Learner	STFWI Data Set				STM Data Set			
	TS	RMAE	RMSE	CCM	TS	RMAE	RMSE	CCM
Add. Reg.	-	1.42±0.10	1.88±0.16	-	-	1.47±0.11	1.97±0.16	-
B-M5P	164	1.16±0.10	1.45±0.15	2.18±0.15	170	1.16±0.11	1.44±0.16	2.25±0.16
B-RepTree	383	1.21±0.12	1.57±0.16	3.28±0.16	359	1.21±0.12	1.56±0.17	3.26±0.17
M5 Rules	138	1.28±0.12	1.63±0.15	2.25±0.15	140	1.30±0.13	1.64±0.20	2.3±0.2
M5P	191	1.16±0.11	1.46±0.15	2.31±0.15	203	1.20±0.11	1.49±0.19	2.45±0.19
NN-MLP	-	1.97±0.10	2.56±0.18	-	-	2.13±0.11	2.72±0.17	-
IP3	29	1.15±0.15	1.43±0.12	1.56±0.12	29	1.15±0.14	1.43±0.18	1.57±0.18
MOPT	110	1.56±0.10	1.84±0.18	2.12±0.12	95	1.48±0.14	1.79±0.17	2.06±0.15
RepTree	320	1.26±0.15	1.71±0.18	3.14±0.18	299	1.29±0.14	1.70±0.19	3.12±0.19
RETIS-M	193	1.25±0.09	1.63±0.19	2.49±0.19	193	1.23±0.11	1.59±0.19	2.5±0.19
SVM RBF	-	1.16±0.10	1.41±0.16	-	-	1.16±0.12	1.41±0.17	-

An interesting result that might be inferred from the Table 1 tree size measure TS is that it provides information indicating that the proposed IP3 model builds more compact and accurate prediction trees than other state-of-the-art regression tree models. The MOPT algorithm induced relatively compact prediction tree models but also demonstrated low prediction accuracies. This result can be explained by the global MOPT split point calculation procedure and an unavailable subroutine for the removal of the outliers of two statistical moments which was successfully implemented in corresponding IP3 algorithm.

Figure 3.a and Figure 3.b show that in STFWI and STM data sets under the RMSE measure the SVM, IP3, B-M5P, and M5P models significantly outperformed the other state-of-the-art models with a 90% confidence level. In the case to combine the accuracy RMSE with the tree size TS in the Cost Complexity Measure CCM [3] the IP3 model significantly outperformed other state-of-the-art models with corresponding values of 1.56 and 1.57 in STFWI and STM data sets.

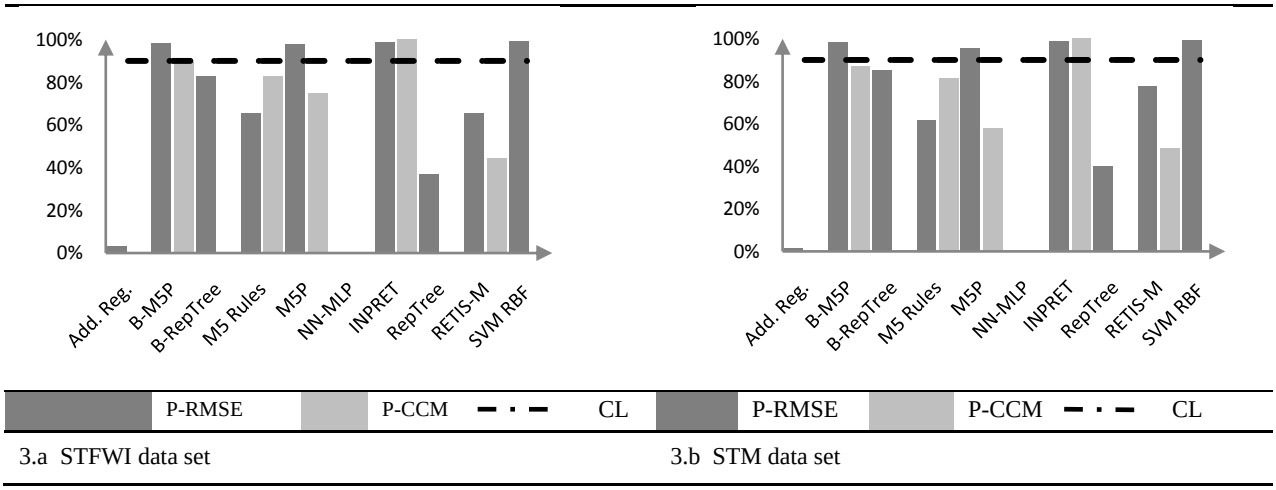


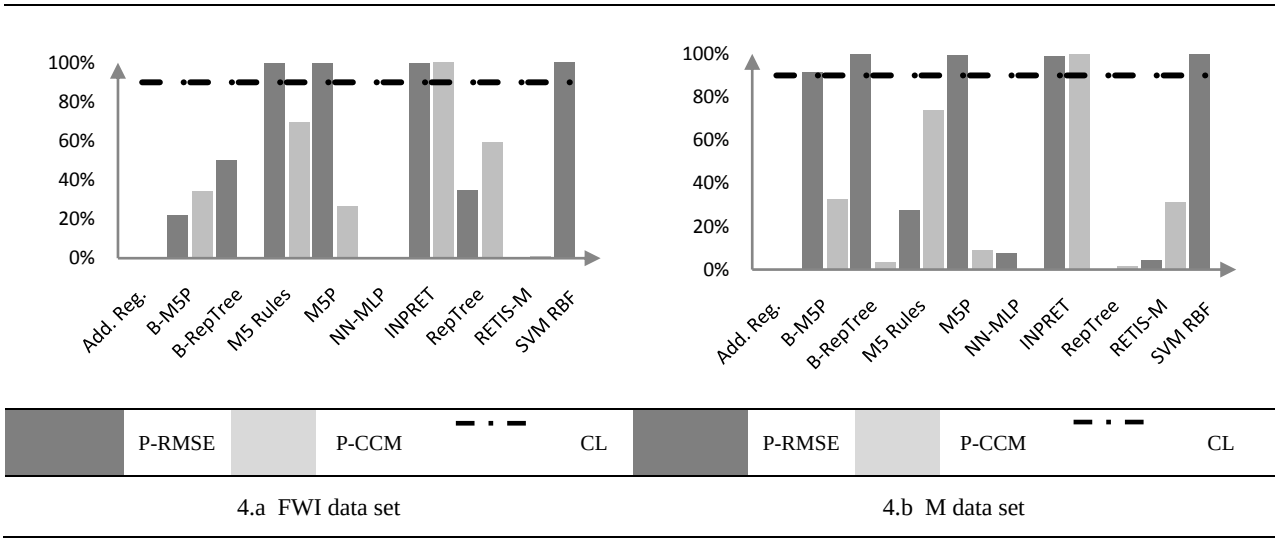
Figure 3: STFWI and STM data sets learners (%) of confidence level (CL) comparison

The results in Table 2 show that in the FWI data set under RMSE and RMAE criterion, the SVM, M5P, IP3, and M5 Rules models significantly outperformed Additive Regression, B-RepTree, NN-MLP, MOPT, RepTree and RETIS algorithms. Respectively, in the M data set under RMSE and RMAE criterion, the SVM, B - REPTree, B - M5P, and IP3 models significantly outperform Additive Regression, M5Rules, M5P, NN-MLP, MOPT, RepTree and RETIS algorithms.

Table 2: FWI and M data sets learners comparison (10 time 10 fold cross validation)

Learner	FWI Data Set				M Data Set			
	TS	RMAE	RMSE	CCM	TS	RMAE	RMSE	CCM
Add. Reg.	-	1.28±0.10	1.65±0.15	-	-	1.32±0.10	1.76±0.15	-
B-M5P	122	1.20±0.09	1.54±0.14	2.7±0.14	175	1.19±0.10	1.50±0.16	3.16±0.16
B-RepTree	169	1.20±0.11	1.52±0.15	3.12±0.15	213	1.44±0.11	1.44±0.16	3.46±0.16
M5 Rules	114	1.20±0.10	1.45±0.14	2.53±0.14	143	1.25±0.11	1.57±0.16	2.92±0.16
M5P	137	1.18±0.11	1.44±0.22	2.74±0.22	201	1.19±0.11	1.46±0.16	3.36±0.16
NN-MLP	-	1.21±0.15	1.61±0.18	-	-	1.20±0.14	1.6±0.17	-
IP3	19	1.19±15	1.45±0.18	1.63±0.18	37	1.17±0.16	1.47±0.20	1.82±0.2
MOPT	87	1.25±17	1.58±0.17	1.75±0.18	112	1.24±0.16	1.52±0.18	1.72±0.2
RepTree	111	1.20±0.12	1.53±0.16	2.58±0.16	197	1.27±0.16	1.67±0.20	3.54±0.2
RETIS-M	155	1.27±0.16	1.60±0.19	3.07±0.19	165	1.22±0.15	1.61±0.17	3.17±0.17
SVM RBF	-	1.16±0.12	1.41±14	-	-	1.16±0.11	1.41±0.14	-

Finally, Figure 4.a and 4.b show that the proposed IP3 algorithm outperforms state-of-the-art tree algorithms in terms of the CCM, in both FWI and M data sets. Also, here the MOPT algorithm demonstrates relatively compact and balanced trees but gives way to IP3 and other state-of-the-art tree algorithms in terms of prediction accuracy.



4.a FWI data set

4.b M data set

Figure 4: FWI and M data sets learners (%) of confidence level (CL) comparison

Additional interesting result that can serve as an important added value to Cortez and Morais [6] work is the irrelevance of the spatial and temporal variables, because when they are removed, the performance of the IP3 and SVM models does not decrease significantly. In fact, a tree with the best cost-complexity configuration ($\alpha \geq 0.01$) is yielded by the FWI setup and IP3 model, which statistically confirms the significance of this result. For the IP3 method, it is better to use spatio-temporal components in conjunction with FWI or weather conditions, rather than FWI or weather variables only. However, from the RMSE or RMAE point of view, the best option is the SVM RBF model predictor.

5. CONCLUSIONS

The results presented in each segment of the data set in Table 3 display the comparative analysis of the forest fires experiment described in previous section 4. They reflect the relative preference of some methods that are significantly better than the worst model at the 90% of confidence level. Each cell should be viewed clockwise starting with the top-left corner. For example, in STFWI data set, the SVM RBF, IP3, B-M5P, and M5P models significantly outperform other state-of-the-art tree algorithms such as Neural Network MLP [2] and kernel-based Additive Regression [8] in terms of RMSE measure. In case of CCM, for the same data set, only the IP3 and the B-M5P models are significantly more accurate than other corresponding state-of-the-art models.

Summarizing the results presented in the Table 3 it can concluded that the principal advantage of IP3 is that it provides a more compact representation of prediction tree size TS and statistically significant prediction accuracy in comparison to other regression tree algorithms. According to the demonstrated experiments results the IP3 models achieved on average a 31% reduction in interval prediction tree size TS and an 11% improvement in CCM accuracy, without significant loss in RMSE accuracy. These results were achieved due to the use of mean-variance predictors, outlier detection and removal, and implementation of new node splitting techniques that were implemented in the proposed IP3 algorithm.

Table 3: Experiments comparing between IP3 and state-of-the-art methods

Data Set	RMSE		CCM		TS	
STFWI	SVM RBF[1]	**IP3 [2]**	**IP3[1]**	B-M5P[2]	**IP3[1]**	M5 Rules[2]
	M5P[4]	B-M5P[3]	-	-	M5P[4]	B-M5P[3]
STM	SVM RBF[1]	**IP3[2]**	**IP3[1]**	-	**IP3[1]**	M5 Rules[2]
	M5P[4]	B-M5P[3]	-	-	RETIS-M[4]	B-M5P[3]
FWI	SVM RBF[1]	M5P[2]	**IP3[1]**	-	**IP3[1]**	RepTree[2]
	M5 Rules[4]	**IP3[3]**	-	-	B-M5P[4]	M5 Rules[3]
M	SVM RBF [1]	B-RepTree[2]	**IP3[1]**	-	**IP3[1]**	M5 Rules[2]
	IP3[4]	B-M5P[3]	-	-	B-M5P[4]	RETIS-M[3]

6. REFERENCES

[1] Alberg, D., Last, M., & Kandel, A. (2012). Knowledge Discovery in Data Streams with Regression Tree Methods. WIREs Data Mining Knowledge Discovery 2012 , 2, 69-78.

[2] Bishop, C. (2006). Pattern Recognition and Machine Learning. New York: Springer.

[3] Breiman, L., & Friedman, J. (1985). Estimating Optimal Transformations for Multiple Regression and Correlation. Journal of American Statistic Association , 80, 580 - 597.

[4] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees. Belmont: Wadsworth & Brooks/Cole. Pacific Grove.

[5] Ceci, M., Appice, A., & Malerba, D. (2003). Comparing Simplification Methods for Model Trees with Regression and Splitting Nodes. Foundations of Intelligent Systems, 14th International LNAI Symposium, 2871, pp. 49 - 56.

[6] Cortez, P., & Morais, A. (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data. In M. F. In J. Neves (Ed.), New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence, (pp. 512-523). Guimaraes, Portugal.

[7] Friedman, J. (1991). Multivariate Adaptative Regression Splines. In Annals of Statistics , 1 - 19.

[8] Friedman, J. (2002). Stochastic Gradient Boosting. Computational Statistics & Data Analysis , 38 (4), 367 - 378.

[9] Ikonomovska, E., Gama, J., Sebastiao, R., & Gjorgjevik, D. (2009). Regression Trees from Data Streams with Drift Detection. Discovery Science, (pp. 121 - 135).

[10] Karalic, A. (1992). Linear Regression in Regression Tree Leaves. In Proceedings of International School for Synthesis of Expert Knowledge, 10(3), pp. 151 - 162.

[11] Loh, W. (2009). Regression by Parts: Fitting Visually Interpretable Models with GUIDE. (W. H. in C. Chen, Ed.)

[12] Quinlan, J. (1992). Learning with Continuous Classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence. World Scientific.

[13] Vens, C., & Blockeel, H. (2006). A Simple Regression Based Heuristic for Learning Model Trees. Intelligent Data Analysis , 10(3), 215 - 236.

[14] Wang, Y., & Witten, I. (1997). Inducing of Model Trees for Predicting Continuous Classes. In Proceedings of the 9th European Conference on Machine Learning (pp. 128 - 137). Springer-Verlag.