ISLIQ: Improved Supervised Learning in Quest to Nowcast Snow/No-Snow

KISHOR KUMAR REDDY C[1], RUPA C H[2] AND VIJAYA BABU[3]

[1]Research Scholar, Department of Computer Science and Engineering, K L University, Guntur
Kishoar23@gmail.com

[2]Professor, Department of Computer Science and Engineering, V R Siddhartha Engineering College,
Vijayawada, rupamtech@gmail.com

[3]Professor, Department of Computer Science and Engineering, K L University, Guntur
Vijaymtech28@gmail.com

Abstract: - Nowcasting the presence of snow/no-snow is a major problem to most of the researchers, academicians, scientists and so on, as it would affects the lives of humans, animals and aquatic life, vegetation, tourism sector to a greater extent across the globe. Previously, many of the scientists, researchers, academicians and so on provided solutions but limited to the usage of Satellite imagery, Radar imagery and so on. We provided approaches related to the same, by making use of decision trees, but the main drawback with them is computational complexity during the evaluation of split points. With this focus, in this paper we are providing Improved Supervised Learning in Quest (ISLIQ) decision tree algorithm for the nowcasting of snow/no-snow, the evaluation of splitting criterion is based on interval range instead of computing whenever there is a change in the class label, which drastically reduces the number of split points. Further, we also evaluated the performance measures such as specificity, sensitivity, precision, dice, accuracy and error rate and compared the results with various decision tree algorithms.

Key-Words: - Classification, Decision Tree, ISLIQ, Nowcasting, Snow, Split Points

1 Introduction

From the past few decades, most of the researchers are continuously working on the weather nowcasting, as this is the serious natural calamity. Prediction of snow/no-snow is one among them, affects the lives of humans, animals and aquatic life, vegetation, tourism sector to a greater extent across the globe [1]. A reliable nowcasting of critical weather (snow/no-snow) is required for an optimal planning of snow or ice removal on the roads. The weather forecasts are divided into the following categories.

a) Now casting-in which the details about the current weather and forecasts up to a few hours ahead are given [3].

b) Short range forecasts (1 to 3 days) -in which the weather in each successive 24 hr intervals may be predicted up to 3 days. This forecast range is mainly concerned with the weather systems observed in the latest weather charts, although generation of new systems is also considered [3].

c) Medium range forecasts (4 to 10 days)- Average weather conditions and the weather on each day may be prescribed with progressively lesser details and accuracy than that for short range forecasts [3].

d) Long range /Extended Range forecasts (more than 10 days to a season). There is no rigid definition for Long Range Forecasting, which may range from a monthly to a seasonal forecast [3].

The researchers, attempted to now cast the weather informally for millennia, and formally since the nineteenth century. Earlier, weather predictions were mainly based upon changes in current weather conditions, barometric pressure, and the condition of the sky [3]. In 650 BC, the Babylonians predicted the weather based on cloud patterns and also from astrology.Later, in about 340 BC, Aristotle described weather patterns in Meteorologica [3]. Later, Theophrastus compiled a book on forecasting of weather, called the Book of Signs [3]. By 300 B.C., Chinese astronomers developed a calendar, in which they divided a year into 24 festivals and each festival is allied with a different type of weather [3]. In 904 AD, IbnWahshiyya's Nabatean discussed about the forecasting of weather from changes in the atmosphere and also signs from the planetary astral alterations; signs of rain based on observation of the lunar phases; and weather forecasts based on the movement of winds [3]. All the ancient weather forecasting methods, usually based on observed patterns of events. For instance, if the sunset was red, the particular day often brought fair weather.

However, not all of these predictions on weather prove reliable, and many of the methods found that not accurate. This increased the interest in weather forecasting to most of the academicians, researchers, scientists and so on.

From the last few decades, the availability of climate data increases makes it important to find effective and accurate tools to analyze and extract hidden knowledge from this huge data. Meteorological data mining is a form of data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge [5]. Useful knowledge can play important role in understanding the climate variability and climate nowcasting.

A wide range of snow/no-snow presence now cast methods are employed in weather nowcasting at regional and national levels [2][4][6-25][38-39] [41-43]. Fundamentally there are two approaches to now cast the presence of snow/no-snow. They are Empirical and Dynamical Methods. The Empirical approach is based on analysis of past historical data of weather and its relationship to a variety of atmospheric variables. The most widely use empirical approaches used for climate prediction are decision trees, regression, artificial neural network, fuzzy logic and group method of data handling [26-35] [44]. The dynamical approach, predictions are generated by physical models based on system of equations that predict the future snowfall. The forecasting of weather by computer using equations are known as numerical weather prediction. To predict the weather by numeric means, meteorologist has develop atmospheric models that approximate the change in temperature, pressure, humidity, wind speed, dew-point etc using mathematical equations.

In the present research, the presence of snow/no-snow prediction is implemented with the use of empirical statistical technique, decision trees [35]. A decision tree is a classification scheme which generates a tree and a set of rules, representing the model of different classes, from a given dataset. The set of records available for developing classification methods is generally divided into two disjoint subsets -a training set and a test set. The former is used for deriving the classifier, while the latter is used to measure the accuracy of the classifier [35].

We have laid out the rest of this paper as follows: Section 2 provides the related work. Section 3 introduces decision tree algorithm. Section 4 provides the results obtained on twenty one

international locations, and these results have been compared with other models. Section 5 concludes the paper followed by references.

2 Related Works

Till date, many of the researchers, academicians, scientists across the globe proposed many methodologies in the prediction of snow using satellite imagery, radar imagery, physical instruments, various algorithms, models and so on, adding to it some researchers estimated the amount of snow while some researchers detected the density of snow and few discriminated the differences between wet snow and dry snow [2] [6-25]. Though the approaches presented by various researchers were presumed to be conditionally fruitful. But a conditional approach would though yield a good result, but there is a requirement to have an approach that predicts the worse case weather conditions with a model so as to predict any sporadic snowfall at an initial stage in order to take precautionary measure at worse conditions.

Irene Y.H Gu et al. [6], put forward a full automatic image analysis system for detection and analysis of snow/ice coverage on electric insulators of power lines using images which were captured by visual cameras in a remote outdoor laboratory test bed. Jinmei Pan et al. [7], put forth a passive microwave remote sensing techniques that detected wet snow in the south of china. Yajaira Mejia et al. [8], gave an approach for estimating the snowfall using neural networks on multi source remote sensing observations and ground based meteorological measurements. Melanie Wetzel et al. [9], projected a technique that supports the snowfall forecast and for the verification of radar limited mountainous terrain that includes matching the output parameters and graphics from high resolution mesoscale models to surface mesonets. Pascal Sirguey et al. [10], made use of ASTER and MODIS sensors, both on the TERRA platform by implementing the ARSIS concept so as to fuse the high spatial content of the two 250m spectral bands of MODIS into five 500m bands using wavelet based multi resolution analysis in the mountainous environment.

Michael A. Rotondi [11] illustrated a Markov chain models across eight national weather stations using historical data from the global historical climatology network to predict a 'snow day'. Gail M. Skofronick Jackson et al. [12], in their research interpreted how instruments like the W-band radar of Cloudsat, Global Precipitation Measurement Dual-Frequency Precipitation Radar ku- and Ka-bands, and the Microwave Imager can be used in the

simulations of lake effect and synoptic snow events in order to determine the minimum amount of snow. Gail M. Skofronick Jackson et al. [13], demonstrated thresholds for detecting falling snow from satellite-borne active and passive sensors.

Andrea Spisni et al. [14], presented an operational chain developed in the Emilia-Romagna region to monitor snow cover and snow water equivalent over the area managed by the Regional Catchment Technical Service. Alberto Martinez Vazquez et al. [15] presented an algorithm using GB-SAR imagery for the automatic recognition and classification of snow avalanches. Jeremie Bossu et al. [16], made use of a structure, based on computer vision which detects the presence of snow or rain. Noel Dacruz Evora et al. [17], used brightness temperature data, provided by seven channels SSM/I aboard the Defense Meteorological Satellite Program F-11 and F-13 spacecrafts. Using which a modelling framework was put forth by combining passive microwave data, neural network based models and geostatistics for snow water equivalent retrieval and mapping. Hossein Zeinivand et al. [18], enforced a spatially distributed physically based model to detect snow and melting in the Latyan dam watershed in Iran.

Xiaolan Xu et al. [19], developed a model that can be used for both active and passive microwave remote sensing of snow. B.B Fitzharris et al. [20], presented three case studies on the usage of satellite imagery for mapping seasonal snow cover in New Zealand, and also explored the effectiveness of using AVHRR imagery in order to obtain the presence of snow, snow covered area and snow line elevation on the mountain ranges of New Zealand. Ashok N.Srivastava et al. [21], in their research discussed the results based on kernel methods for unsupervised discovery of snow, ice, clouds and other geophysical processes based on data from the MODIS instrument.

G.Singh et al. [22], developed a Radar Snow Index model to identify snow using SAR polarimetry techniques. In their research, full polarimetric L-band ALSOS-PALSAR data of snow cover area in Himalayan region have been analyzed based on various component scattering mechanism models and all model results are compared. Fan Ke et al. [23], developed a model to identify winter time heavy snow over Northeast China by using a inter annual increment prediction approach. Folorunsho Olaiya [24] investigated the use of artificial neural networks and decision tree algorithms in forecasting maximum temperature, rainfall, evaporation and wind speed using meteorological data collected from the city of Ibadan, Nigeria through Nigerian Meteorological Agency, Oyo state office. Manjeet Singh et al. [25] forwarded an attempt to develop an automatic technique for avalanche area identification and also its severity index.

For the detailed relevant work refer our earlier papers [38] [39] [41] [45].

3 Decision Tree Algorithm

Classification is the task of learning a target function f that maps each attribute set x to one of the pre-defined class labels y. The input for the classification is the training dataset, whose class labels are already known. Classification analyzes the training dataset and constructs a model based on the class label, and aims to assign a class label to the future unlabelled records. A set of classification rules are generated by such a classification process, which can be used to classify future data and develop a better understanding of each class in the database. Some of the classification models are decision trees, neural networks, genetic algorithms, statistical models like linear/geometric discriminates [35]. In the present research, we are proposing ISLIQ decision tree [40] for the nowcasting of snow/no-snow.

3.1 Procedure for evaluating the split points and decision tree generation

Step 1: Read the dataset and generate a sorted attribute list along with its associated class label.

Step 2: Evaluate the interval range, shown in equation 1.

$$Interval\,Range = \frac{a\max - a\min}{Group\;size} \tag{1}$$

Where amax is the maximum value for the particular attribute, amin is the minimum value for the particular attribute and Group size is to be fixed by user. Upon the experimentation 2 is giving as maximum accuracy levels when compared with other levels. Based on the interval range, evaluate the split points and it is shown in equation (2) [38].

$$Split\,po\text{int} = Midpo\text{int}(Changed\;class\;label) \tag{2}$$

Step 3: Calculate the G_Info value for the class label by using equation 3 [35].

$$G\_in\,fo = 1 - \sum_{i=1}^{M} P_i^2 \tag{3}$$

Where P_i is the probability that a relative tuple in given dataset belongs to class C_i, M is the number of class labels.

Step 4: Calculate the G_Info_D value for every attribute by using equation 4 [35].

$$G\_\inf o_D = \sum_{j=1}^{N} P_j \left[1 - \sum_{i=1}^{M} P_i^2 \right]$$

(4)

Step 5: The Gini Index is obtained by finding the difference between G_Info and G_Info$_D$ values by using equation 5 [35].

$$GiniIndex = G\_in fo - G\_in fo_D$$

(5)

Step 6: The maximum Gini Index value is considered as the best split point and is the root node, as shown in equation 6 [35].

$$Best\ Split\ po\mathrm{int} = Maximum\ (Gini\ Index)$$

(6)

Repeat this procedure until every node ends with a unique class label.

3.1 Illustration

The steps in the proposed algorithm are explained using a sample dataset, shown in Table 1. In further Tables, T indicates Temperature, D indicates Dew Point, H Indicates Humidity, P indicates Pressure, V indicates Visibility, W indicates Wind Speed, S indicates Snow; NS indicate No Snow and SP indicate Split Point.

The evaluated split points after pre-sorting for temperature attribute is shown in Table 2. In Table2, amax is 36 and amin is 9, group size is assumed as 2. By applying the values in equation (1), the obtained interval range value is 13.5. As the range value is always to be in integer format we are making it floor and hence the interval range is taken as 13. The first record is S and the next 13th record is also S and hence no change in the class label. Similarly the next 13th record is NS. There is a change in the class label hence evaluating the split points using equation (2) and the obtained split point is 19.5. Similarly repeat the same procedures for the rest attributes also.

Similarly for the rest attributes, compute the split points. The computed split points for Temperature is 19.5, Dew point are 0.0, 0.5 and 1.0, Humidity 35.0, Pressure 1006.5, Visibility 1.0, 1.5, 2.0, 2.5, 3.0 and 3.5, Wind 7.5 respectively. Initially consider Temperature attribute and its split point is 19.5. Consider the split point 19.5 and assume it as root node and construct a binary tree i.e. keep the values which are less than to 19.5 at left side and the rest values at right side, by usiing the proposed algorithm. Similarly compute for all the attributes.

The computed Gini Index [38] values for iteration 1 are shown in Table 3. It is very clear that among all the attributes the attribute V is having maximum

gini Index value of 0.10 at split position 3.0. hence, it is chosen as the best split point for iteration 1 and the same procedure is repeated till it ends with a unique class label and the values for all the iterations is shown in Table 4. The decision tree [35] obtained for the sample dataset shown in Table 1 is shown in Figure .1

Table 1. Sample weather dataset

T	D	H	P	V	W	C
11	0	52	1016	2	3	S
14	0	45	1014	4	13	NS
24	0	20	1012	3	14	NS
27	0	17	1008	4	10	NS
10	1	57	1018	2	10	S
10	1	55	1018	1	3	S
11	1	53	1014	2	5	S
11	1	52	1016	3	8	NS
12	1	46	1019	3	10	S
12	1	51	1016	3	10	S
13	1	49	1020	2	5	NS
13	1	50	1017	2	2	NS
13	1	43	1018	2	13	NS
13	1	46	1013	3	14	NS
13	1	44	1017	3	10	NS
14	1	44	1016	3	2	NS
14	1	41	1015	4	11	NS
16	1	41	1018	3	6	NS
16	1	38	1014	2	19	NS
27	1	18	1008	4	10	NS
30	1	16	1008	4	11	NS
31	1	16	1005	3	10	NS
36	1	12	999	1	13	NS
9	2	63	1018	1	5	S
11	2	58	1014	2	8	S
11	2	61	1015	3	13	S
11	2	60	1015	1	5	S
12	2	51	1016	2	8	S
12	2	53	1014	2	5	S
12	2	59	1016	2	11	S

Decision trees are especially attractive in the data mining environment as they represent rules. Rules can be readily being expressed in natural language and are easily comprehensible.

Table 2. Pre-sorting and evaluation of split points For Temperature attribute

T	C	Split Points
9	S	
10	S	
10	S	
11	NS	
11	S	
11	S	
11	S	
11	S	
11	S	No change in the class label
12	S	
12	S	
12	S	
12	S	
12	S	
13	NS	
13	NS	
13	NS	
13	NS	
13	NS	
14	NS	
14	NS	
14	NS	19.5
16	NS	
16	NS	
24	NS	
27	NS	
27	NS	
30	NS	
31	NS	
36	NS	
Group Size = 2: Interval range = 13		

Table 3. Gini Index values for iteration 1

Attribute	Split Point	Gini Index
T	19.5	0.09
D	1.0	0.01
H	35.0	0.09
P	1006.5	0.02
V	**3.0**	**0.10**
W	7.5	0.02

The classification rules [35] obtained for the sample dataset are shown below:

1. If [(V < 3.0) and (P < 1006.5)] Then (prediction = nosnow)
2. If [(V < 3.0) and (P >= 1006.5) and (V < 2.0)] Then (prediction = snow)
3. If [(V < 3.0) and (P >= 1006.5) and (V >= 2.0) and (D < 1.0)] Then (prediction = snow)
4. If [(V < 3.0) and (P >= 1006.5) and (V >= 2.0) and (D >= 1.0) and (W < 3.5)] Then (prediction = nosnow)
5. If [(V < 3.0) and (P >= 1006.5) and (V >= 2.0) and (D >= 1.0) and (W >= 3.5)] Then (prediction = snow)
6. If [(V >= 3.0) and (V < 3.5) and (W < 7.5)] Then (prediction = nosnow)
7. If [(V >= 3.0) and (V < 3.5) and (W >= 7.5) and (H < 35.0)] Then (prediction = nosnow)
8. If [(V >= 3.0) and (V < 3.5) and (W >= 7.5) and (H >= 35.0)] Then (prediction = snow)
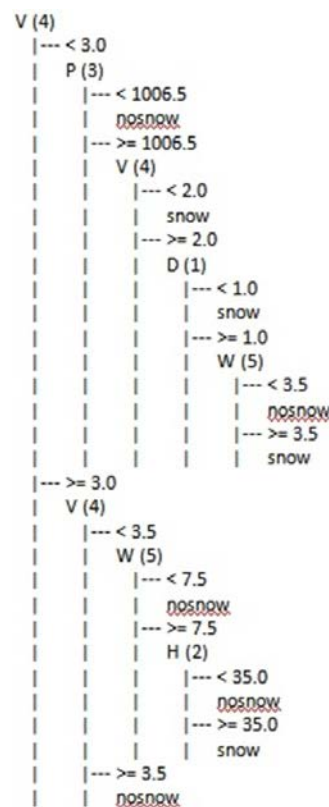9. If [(V >= 3.0) and (V >= 3.5)] Then (prediction = nosnow)

```
V (4)
|--- < 3.0
|   P (3)
|   |--- < 1006.5
|   |   nosnow
|   |--- >= 1006.5
|   |   V (4)
|   |   |--- < 2.0
|   |   |   snow
|   |   |--- >= 2.0
|   |   |   D (1)
|   |   |   |--- < 1.0
|   |   |   |   snow
|   |   |   |--- >= 1.0
|   |   |   |   W (5)
|   |   |   |   |--- < 3.5
|   |   |   |   |   nosnow
|   |   |   |   |--- >= 3.5
|   |   |   |   |   snow
|--- >= 3.0
|   V (4)
|   |--- < 3.5
|   |   W (5)
|   |   |--- < 7.5
|   |   |   nosnow
|   |   |--- >= 7.5
|   |   |   H (2)
|   |   |   |--- < 35.0
|   |   |   |   nosnow
|   |   |   |--- >= 35.0
|   |   |   |   snow
|   |--- >= 3.5
|   |   nosnow
```

Figure 1. Decision tree for the data shown in Table 1.

4 Results and Discussions

The proposed method has been tested on 21 international locations historical datasets of snow/no-snow, collected from various meteorological departments [36]. We conducted experiments by implementing our proposed algorithm in Java Net Beans IDE 7.2. All experiments were performed on intel i3 core

Table 4. Computed Gini Index values for all iterations

T		D		H		P		V		W	
S	GI	S	GI	S	GI	S	GI	S	GI	S	GI
19.5	0.09	1.0	0.01	35.0	0.09	1006.5	0.02	3.0	0.10	7.5	0.02
19.5	0.06	1.0	0.01	35.0	0.06	1006.5	0.06	2.0	0.00	7.5	0.01
19.5	0.0	1.0	0.01	35.0	0.0	1006.5	0.0	2.0	0.04	7.5	0.00
19.5	0.0	1.0	0.02	35.0	0.0	1006.5	0.0	1.0	0.0	3.5	0.00
19.5	0.0	1.0	0.0	35.0	0.0	1006.5	0.0	1.0	0.0	3.5	0.07
19.5	0.0	1.0	0.0	35.0	0.0	1006.5	0.0	1.0	0.0	3.5	0.00
19.5	0.03	1.0	0.01	35.0	0.03	1006.5	0.00	3.5	0.03	7.5	0.01
19.5	0.04	1.0	0.02	35.0	0.04	1006.5	0.02	1.0	0.0	7.5	0.04
19.5	0.09	1.0	0.04	35.0	0.09	1006.5	0.04	1.0	0.0	3.5	0.00
							0.0				

processor and 4 GB RAM with windows 7 operating system. We also divided our data set in to two parts: training set (75%), which is used to create the model, and a test set (25%), which is used to verify that the model is accurate and not over fitted. Table 5 summarizes the characteristics of the datasets, arranged in alphabetical order, presenting the number of instances, training instances, testing instances, and classes.

The comparison in terms of split points is presented in Table 6. Apparently, almost all results for ISLIQ are better than that of SLIQ. The results clearly show that the proposed ISLIQ reduces nearly 51.48 % when compared with SLIQ decision tree algorithms Figure 2 presents the number of split points of the SLIQ and proposed ISLIQ algorithms graphically.

A common but poorly motivated way of evaluating results of Machine Learning Experiments is using specificity, accuracy, precision, dice and error rate. Specificity relates to the test's ability to exclude a condition correctly [37]. Precision is defined as the proportion of the true positives against all the positive results. Dice is defined as proportion of two time true positives and the combination of true positives, false positives and false negatives [37]. Prediction error is a measure of the performance of a model to predict the correct output, given future observations used as predictors. In order to reveal the performance of our proposed

ISLIQ algorithm, we presented comparisons between SLIQ, SPM, SLGAS and ISLIQ in terms of accuracy, specificity, precision, dice and error rate.

Table 5. Dataset Description

City Name	Training	Testing	Attributes	Classes
Aberdeen	4750	1583	6	2
Bangkok	4305	1435	6	2
Barcelona	4510	1504	6	2
Benton	17281	5761	6	2
Botswana	4535	1512	6	2
Brazil	4775	1592	6	2
Cairo	4607	1536	6	2
Chennai	4525	1508	6	2
Delhi	4511	1504	6	2
Eglinton	4738	1580	6	2
Humberside	777	259	6	2
Hyderabad	4387	1462	6	2
Iceland	2634	878	6	2
Lahore	3665	1222	6	2
Manchester	4753	1585	6	2
Norway	4579	1526	6	2
Olympia	17281	5761	6	2
Perth	4636	1546	6	2
Sellaness	4059	1353	6	2
Tiruptahi	4529	1510	6	2
Valley	4561	1521	6	2

Table 6. Comparison of split points

City Name	SLIQ	ISLIQ
Aberden	740	210
Bangkok	112	50
Barcelona	174	41
Benton	449	208
Botswana	195	88
Brazil	460	299
Cairo	165	162
Chennai	130	64
Delhi	281	162
Eglinton	360	43
Humbersid	171	34
Hyderabad	116	72
Iceland	385	207
Lahore	190	51
Mancheste	499	211
Norway	765	560
Olympia	449	349
Perth	246	136
Sellaness	391	154
Tirupathi	154	108
Valley	706	200
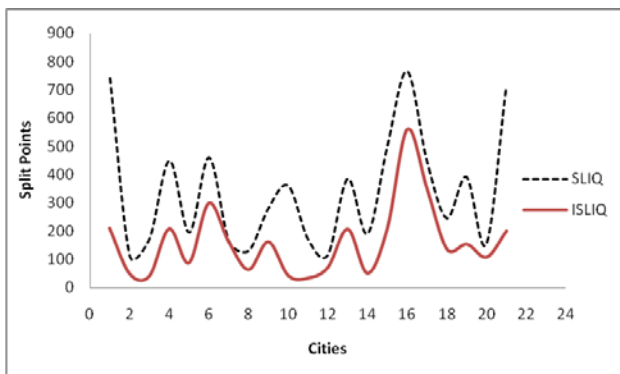
Figure 2. Split Points Comparison

The comparison in terms of classification accuracy is presented in Table 7. Apparently, almost all accuracy results for ISLIQ are better than those of SLIQ, SPM and SLGAS. For some of the cities, the accuracy levels are more for other algorithms when compared with ISLIQ. But, on an average the ISLIQ model outperforms when compared with other algorithms. Figure 3 presents the classification accuracy of the SLIQ, SPM, SLGAS and proposed ISLIQ algorithms graphically.

The comparison in terms of specificity is presented in Table 8. Apparently, almost all specificity results for ISLIQ are better than those of SLIQ, SPM and SLGAS. For some of the cities, the

specificity levels are more for other algorithms when compared with ISLIQ. But, on an average the ISLIQ model outperforms when compared with other algorithms. Figure 4 presents the classification specificity of the SLIQ, SPM, SLGAS and proposed ISLIQ algorithms graphically.

Table 7. Comparison of accuracy

| City	SLIQ	SPM	SLGAS	ISLIQ
Aberden	87.3	85.47	85.97	87.61
Bangkok	96.09	94.49	95.19	98.11
Barcelona	95.8	95.14	95.67	96.07
Benton	70.05	70.14	72.03	70.12
Botswana	93.78	96.16	93.58	95.43
Brazil	75.5	73.05	75.75	73.36
Cairo	88.99	89.7	89.77	89.98
Chennai	76.65	76.12	77.51	76.35
Delhi	96.14	94.94	96.8	93.15
Eglinton	89.24	90.06	90.06	89.56
Humberside	93.05	94.59	94.98	93.82
Hyderabad	96.5	97.8	94.79	96.4
Iceland	89.17	88.49	90.2	88.49
Lahore	84.82	86.05	85.89	84.65
Manchester	92.74	92.87	89.58	93.43
Norway	88.99	90.89	90.62	90.89
Olympia	70.05	70.14	72.03	70.12
Perth	94.3	94.43	96.31	94.43
Sellaness	75.9	77.67	79.45	84.4
Tirupathi	97.54	97.41	97.41	97.48
Valley	90	91.38	90.32	91.3
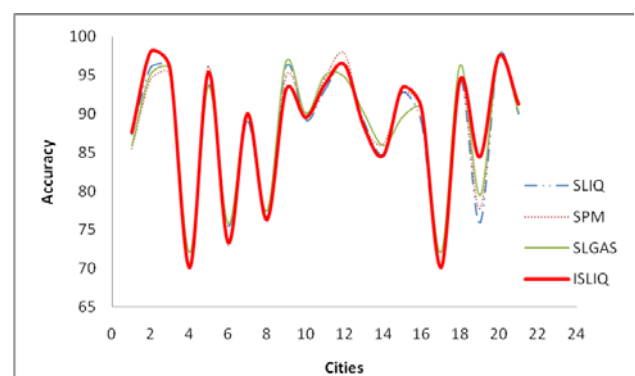
Figure 3. Accuracy comparison

The comparison in terms of precision is presented in Table 9. Apparently, almost all precision results for ISLIQ are better than those of SLIQ, SPM and SLGAS. For some of the cities, the precision levels are more for other algorithms when compared with ISLIQ. But, on an average the ISLIQ model outperforms when compared with other algorithms.

Figure 5 presents the classification precision of the SLIQ, SPM, SLGAS and proposed ISLIQ algorithms graphically.

Table 8. Comparison of specificity

City Name	SLIQ	SPM	SLGAS	ISLIQ
Aberdeen	93.24	91.22	92.19	93.69
Bangkok	96.36	94.75	95.45	98.53
Barcelona	97.59	97.8	97.87	98.83
Benton	71.82	72.96	75.08	73.07
Botswana	94.68	97.24	94.41	96.56
Brazil	79.52	77.08	80.7	77.53
Cairo	95.53	96.95	96.74	96.43
Chennai	82.36	81.03	82.36	75.27
Delhi	97.83	96.34	98.44	94.58
Eglinton	97.52	97.87	97.66	98.44
Humberside	95.49	98.36	98.36	97.13
Hyderabad	97.77	99.16	95.97	98.16
Iceland	96.51	93.29	95.48	93.67
Lahore	86.93	89.5	90.59	89.7
Manchester	95.69	95.9	92.94	96.9
Norway	95.92	98.14	97.78	97.92
Olympia	71.82	72.96	75.08	73.07
Perth	96.73	96.87	98.8	96.87
Sellaness	78.74	82.96	85.45	93.02
Tiruptahi	98.06	97.86	98.06	97.93
Valley	93.86	95.31	93.72	93.91
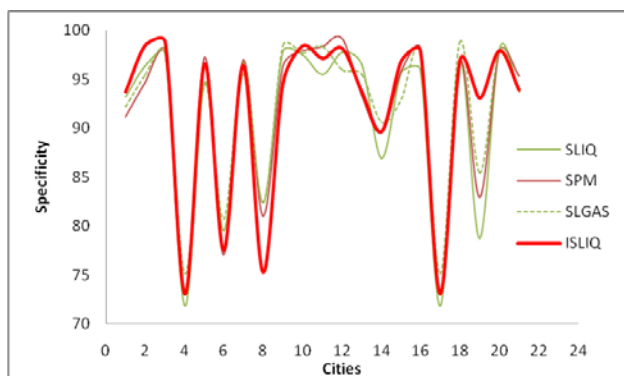
Figure 4. Comparison of specificity

The comparison in terms of dice is presented in Table 10. Apparently, almost all dice results for ISLIQ are better than those of SLIQ, SPM and SLGAS. For some of the cities, the dice are more for other algorithms when compared with ISLIQ. Figure 6 presents the classification dice of the SLIQ, SPM, SLGAS and proposed ISLIQ tree algorithms graphically.

Table 9. Comparison of precision

| City Name	SLIQ	SPM	SLGAS	ISLIQ
Aberdeen	60.69	53.93	55.93	62.16
Bangkok	3.7	2.59	2.98	0
Barcelona	35.18	15.78	29.54	22.72
Benton	52.38	52.59	55.15	52.57
Botswana	12.22	18	12.63	13.55
Brazil	31.26	27.9	30.4	28.13
Cairo	22.22	17.3	22.03	22.12
Chennai	14.33	16.28	18.15	14.32
Delhi	3.03	8.47	8	4.76
Eglinton	46.96	56.52	56	51.11
Humberside	42.1	55.55	60	46.15
Hyderabad	3.03	0	3.33	14.12
Iceland	56.45	50.94	59.77	51
Lahore	54.16	57.76	58.14	54.78
Manchester	41.81	42.45	26.05	45.88
Norway	20.83	33.33	31.11	35.55
Olympia	52.38	52.59	55.15	52.57
Perth	9.25	9.61	21.73	9.61
Sellaness	31.19	30.52	32.66	42.95
Tiruptahi	6.45	8.57	0	8.82
Valley	6.31	8.1	12.5	11.42
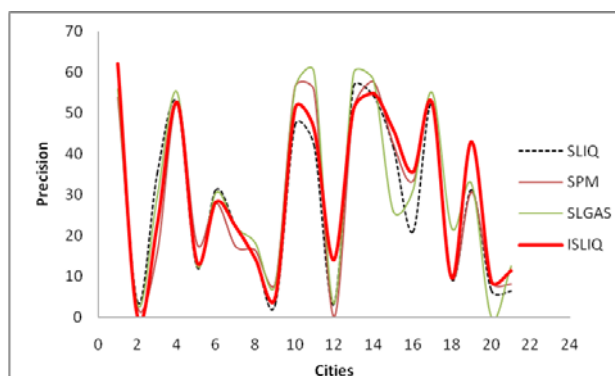
Figure 5. Comparison of precision
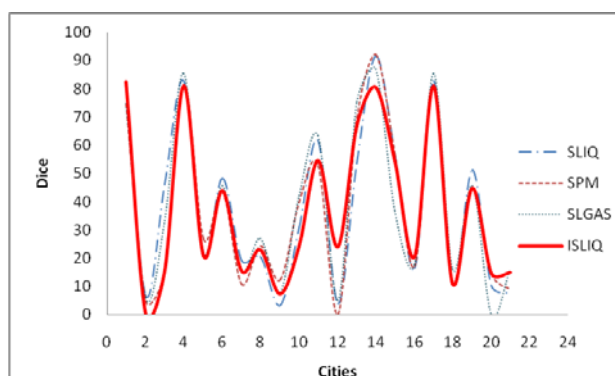
Figure 6. Comparison of dice

The comparison in terms of error rate is presented in Table 11. Apparently, almost all error rate results for ISLIQ are better than those of SLIQ, SPM and

SLGAS. Figure 7 presents the error rate of the SLIQ, SPM, SLGAS and proposed ISLIQ algorithms graphically. As can be observed, ISLIQ obtained better results than SLIQ, SPM and SLGAS i.e. lesser error rate.

Table 10. Comparison of dice

City Name	SLIQ	SPM	SLGAS	ISLIQ
Aberdeen	81.76	74.65	74.57	82.63
Bangkok	6.89	4.93	5.63	0
Barcelona	46.34	15.18	33.33	15.62
Benton	82.73	81.25	85.5	81.02
Botswana	20.95	26.86	22.01	20.77
Brazil	48.33	43.71	45.6	43.83
Cairo	19.25	10.77	15.29	15.63
Chennai	20.4	24.39	27.04	22.95
Delhi	3.38	12.34	8	7.47
Eglinton	30.84	39.79	42.21	24.46
Humberside	61.53	52.63	63.15	54.54
Hyderabad	3.84	0	5.12	24.11
Iceland	53.84	69.67	75.36	67.1
Lahore	91.49	92.06	86.84	80.51
Manchester	57.14	56.96	36.63	54.54
Norway	16.39	17.1	17.83	20.64
Olympia	82.73	81.25	85.5	81.02
Perth	10.75	10.98	16.12	10.98
Sellaness	51.14	44.73	45.55	44.85
Tiruptahi	10.25	14.28	0	14.63
Valley	7.59	8.75	16.25	14.91

Table 11. Comparison of error rate

| City Name	SLIQ	SPM	SLGAS	ISLIQ
Aberdeen	12.70	14.53	14.03	12.39
Bangkok	3.91	5.51	4.81	1.89
Barcelona	4.20	4.86	4.33	3.93
Benton	29.95	29.86	27.97	29.88
Botswana	6.22	3.84	6.42	4.57
Brazil	24.50	26.95	24.25	26.64
Cairo	11.01	10.30	10.23	10.02
Chennai	23.35	23.88	22.49	23.65
Delhi	3.86	5.06	3.20	6.85
Eglinton	10.76	9.94	9.94	10.44
Humberside	6.95	5.41	5.02	6.18
Hyderabad	3.50	2.20	5.21	3.60
Iceland	10.83	11.51	9.80	11.51
Lahore	15.18	13.95	14.11	15.35
Manchester	7.26	7.13	10.42	6.57
Norway	11.01	9.11	9.38	9.11
Olympia	29.95	29.86	27.97	29.88
Perth	5.70	5.57	3.69	5.57
Sellaness	24.10	22.33	20.55	15.60
Tiruptahi	2.46	2.59	2.59	2.52
Valley	10.00	8.62	9.68	8.70
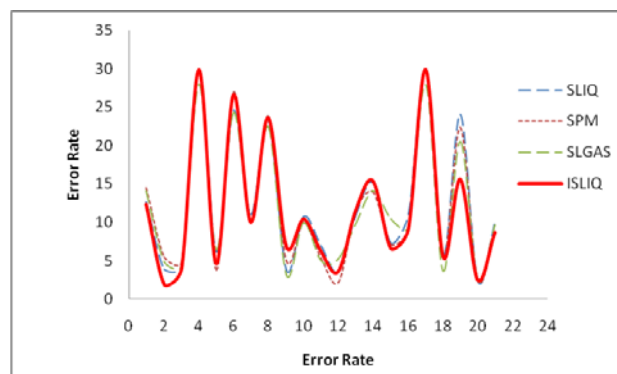
Figure 7. Comparison of Error rate

5. Conclusion

Nowcasting of snow/no-snow more accurately is the major research problem to most of the academicians, researchers, and scientists across the globe. In this paper, we proposed ISLIQ decision tree for the nowcasting of snow/no-snow. The proposed method out performs in terms of various performance measures like sensitivity, specificity, precision, dice and accuracy when compared with other models. The proposed method provides less computational complexity by evaluating the interval range, which significantly decreases the number of split points. Experimental results show that the ISLIQ algorithm scales up well to both large and small datasets with large number of attributes and class labels.

In future, we will adopt other attribute selection measures like entropy and gain ratio in order to increase the performance in the prediction. Further, we will try to decrease the number of split points, which decreases computational complexity to a greater extent by adopting clustering mechanisms.

Acknowledgments

References:

[1] Robert A. Houze: Cloud Dynamics. Academic Press (1994).

[2] Nabilah Filzah Mohd Radzuan, Andi Putra, Zalinda Othman, Azuraliza Abu Bakar and Abdul Razak Hamdan: Comparative Study – Three Artificial Intelligence Techniques for Rain Domain in Precipitation Forecast. International Journal of Environmental, Ecological, Geological and Mining Engineering (2013) 898-903.

[3] http://en.wikipedia.org/wiki/Weather_forecasting.

[4] http://en.wikipedia.org/wiki/Quantitative_precipitation_forecast.

[5] J. Han: Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers. (2001)

[6] Irene Y.H. Gu, Unai Sistiaga, Sonja M. Berlijn and Anders Fahlstrom: Online Detection of Snow Coverage and Swing Angles of Electrical Insulators on Power Transmission Lines Using Videos. IEEE International Conference on Image Processing (2009) 3249 – 3252.

[7] Jinmei Pan, Lingmei Jiang and Lixin Zhang: Wet Snow Detection in the South of China by Passive Microwave Remote Sensing. IEEE International Geo Science and Remote Sensing Symposium (2012) 4863-4866.

[8] Yajaira Mejia, Hosni Ghedira, Shayesteh Mahani and Reza Khanbilvardi: A Neural Network Based Approach for Multi-Spectral Snowfall Detection and Estimation, IEEE International Geo Science and Remote Sensing Symposium (2007) 2276 – 2279.

[9] Melanie Wetzel, Michael Meyers, Randolph Borys, Ray Mcanelly, William Cotton, Andrew Rossi, Paul Frisbie, David Nadler, Douglas Lowenthal, Stephen Cohn, and William Brown: Mesoscale Snowfall Prediction And Verification In Mountainous Terrain. AMS Journal Of Weather And Forecasting (2004) 806-828.

[10] Pascal Sirguey, Renaud Mathieu Yves Arnaud, Muhammad M. Khan and Jocelyn Chanussot: Improved Resolution For The Detection Of Snow With Modis Using Wavelet Fusion. IEEE International Geo Science and Remote Sensing Symposium (2007) 3975-3978.

[11] Michael A. Rotondi: Estimating Transition Matrices to Predict Tomorrow's Snowfall Using Real Data. Journal of Statistics Education (2010) 1-14.

[12] Gail M. Sko Fronick Jackson, Benjamin T. Johnson, and S. Joseph Munchak: Detection Thresholds of Falling Snow From Satellite-Borne Active And Passive Sensors, IEEE Transactions On Geo Science and Remote Sensing (2013) 4177-4189.

[13] Gail M. Skofronick Jackson and Benjamin T. Johnson: Thresholds Of Detection For Falling Snow From Satellite-Borne Active And Passive Sensors, IEEE International Geo Science And Remote Sensing Symposium (2007) 2637-2640.

[14] Andrea Spisni, Fausto Tomei, Sara Pignone, Enrico Muzzi, Alessandro Panzacchi, Gabriele Antolini, Giulia Villani, Michele Di Lorenzo, Rosanna Foraci1, Marco Bittelli and Erin S. Brooks: Snow Cover Analysis in Emilia-Romagna. European Journal of Remote Sensing (2011) 59-73.

[15] Alberto Martinez Vazquez and Joaquim Fortuny Guasch: Snow Avalanche Detection and Classification Algorithm for GB-SAR Imagery, IEEE International Geo Science and Remote Sensing Symposium (2007) 3740-3743.

[16] Jeremie Bossu, Nicolas Hautière and Jean Philippe Tarel: Rain or Snow Detection in Image Sequences through Use of a Histogram of Orientation of Streaks, International Journal of Computer Vision (2011) 348-367.

[17] Noel Dacruz Evora, Ddominique Tapsoba and Danielle De Seve: Combining Artificial Neural Network Models, Geo statistics, and Passive Microwave Data for Snow Water Equivalent Retrieval and Mapping. IEEE Transactions on Geo Science and Remote Sensing (2008) 1925-1939.

[18] Hossein Zeinivand and Florimond De SmISLIQ: Simulation of Snow Covers area by a Physical Based Model. World Academy of Science, Engineering and Technology (2009) 428-433.

[19] Xiaolan Xu, Ding Liang, Leung Tsang, Konstantinos M. Andreadis, Edward G. Josberger, Dennis P. Lettenmaier, Donald W. Cline and Simon H. Yueh: Active Remote Sensing of Snow Using NMM3D/DMRT and Comparison With Clpx Ii Airborne Data, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2010) 689-697.

[20] B. B. Fitzharris and B. P. Mcalevey: Remote Sensing of Seasonal Snow Cover in the Mountains of New Zealand Using Satellite Imagery. Taylor and Francis Geocarto International (1999) 35-44.

[21] Ashok N. Srivastava and Julienne Stroeve: Onboard Detection of Snow, Ice, Clouds and Other Geophysical Processes Using Kernel Methods. Proceedings of the ICML 2003 Workshop on Machine Learning Technologies for Autonomous Space Applications (2003) 1-5.

[22] G. Singh, Y. Yamaguchi, S. E. Park and G. Venkataraman: Identification of Snow Using SAR Polarimetry Techniques. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science (2010) 146-149.

[23] Fan Ke and Tian Baoqiang: Prediction of Wintertime Heavy Snow Activity in North East China. Springer Chinese Science Bulletin (2013) 1420-1426.

[24] Folorunsho Olaiya: Application of Data Mining Techniques in Weather Prediction and Climate Change Studies, International Journal of Information Engineering and Electronic Business (2012) 51-59.

[25] Manjeet Singh, V. D. Mishra, N. K. Hakur and Jyoti Dhar Sharma: Remote Sensing GIS Based Statistical Modelling for the Prediction of Natural Hazards. International Journal of Engineering Research and Technology, (2012) 1-7.

[26] J.R. Quinlan: Induction of Decision Trees. Journal of Machine Learning (1986) 81-106.

[27] B. Chandra and P. Paul Varghese: Fuzzy Sliq Decision Tree Algorithm: IEEE Transactions on Systems, Man and Cybernetics (2008) 1294-1301.

[28] Anuja Priyama, Abhijeeta, Rahul Guptaa, Anju Ratheeb and Saurabh Srivastavab: Comparative Analysis of Decision Tree Classification Algorithms, International Journal of Current Engineering and Technology (2013) 334-337.

[29] Masud Karim and Rashedur M. Rahman: Decision Tree and Naive Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing. Journal of Software Engineering and Applications (2013) 196-206.

[30] Pedro Domingos and Michael J. Pazzani: Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier. International Conference on Machine Learning (1996) 105-112.

[31] Manish Mehta, Rakesh Agarwal and Jorma Rissanen: SLIQ: A Fast Scalable Classifier for Data Mining. International Conference on Extending Database Technology (1996) 18-32.

[32] Rodrigo Coelho Barros, Marcio Porto Basgalupp, Andre C.P.L.F. De Carvalho and Alex A. Freitas: A Survey of Evolutionary Algorithms for Decision Tree Induction, IEEE Transactions on Systems, Man and Cybernetics (2012) 291-312.

[33] J.R. Quinlan: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993).

[34] S. Safavian and D. Landgrebe: A Survey of Decision Tree Classifier Methodology, IEEE Transactions on Systems, Man And Cybernetics (1991) 660 -674.

[35] Arun k pujari: Data Mining Techniques. Universities Press (2004).

[36] http://www.wunderground.com/.

[37] Powers and M. W. David: Evolution: From Precision, Recall and F-Measure to Roc, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies (2011) 37-63.

[38] Kishor Kumar Reddy C, Vijaya Babu B, Rupa C H: SLEAS: Supervised Learning using Entropy as Attribute Selection Measure. International Journal of Engineering and Technology (2014) 2053-2060.

[39] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu B: A Pragmatic Methodology to Predict the Presence of Snow/No-Snow using Supervised Learning Methodologies. International Journal of Applied Engineering Research (2014) 11381-11394.

[40]] Chandra, B., Mazumdar, S., Vincent, A. and Parimi, N: ISLIQ Decision Tree Algorithm for Classification in Data Mining. Proceedings of the Third International Conference on Web Information Systems Engineering (2002) 160−169.

[41] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu: SPM: A Fast and Scalable Model for Predicting Snow/No-Snow. World Applied Sciences Journal (2014) 1561-1570.

[42] Pramote Luenam, Supawadee Ingsriswang, Lily Ingsrisawang, Prasert Aungsuratana, and Warawut Khantiyanan: A Neuro-Fuzzy Approach for Daily Rainfall Prediction over the Central Region of Thailand. International Multi Conference of Engineers and Computer Scientists (2010).

[43] Lily Ingsrisawang, Supawadee Ingsriswang, Pramote Luenam, Premjai Trisaranuwatana, Song Klinpratoom, Prasert Aungsuratana, and Warawut Khantiyanan: Applications of Statistical Methods for Rainfall Prediction over the Eastern Thailand. International Multi

Conference of Engineers and Computer Scientists (2010).

[44] Thair Nu Phyu: Survey of Classification Techniques in Data Mining. International Multi Conference of Engineers and Computer Scientists (2009).

[45] Kishor Kumar Reddy C, Rupa C H and Vijaya Babu: SLGAS: Supervised Learning using Gain Ratio as Attribute Selection Measure to Nowcast Snow/No-Snow. International Review on Computers and Software (2015).