Employing Data Mining Techniques to Predict Occurrence of Thunderstorm Using Hourly Weather Datasets :In The Case Of Gondar Control Zone

Abebe Mulu
Department of Information Technology
Faculty of Informatics
University of Gondar, Ethiopia

Belay Enyew
Department of Information Technology
Faculty of Informatics
University of Gondar, Ethiopia

Abstract—**Thunderstorms has meaningfully effects on both terminal and en route flights and reduce airspace capacity that results delays and have increased substantially en-route congestion. Current technology cannot provide reliable long-term prediction of thunderstorms for aviation operation. The objective of this study was to apply the data mining techniques to predict the occurrence of thunderstorms using 10 years NMA's synoptic dataset of Gondar station using design science research method. From collected data sets seven important attributes (cloud amount, cloud type, temperature, pressure, wind speed, rain fall and humidity) was selected from other variables or attributes to build the model. The experiments have been conducted using the six-step hybrid process model using four selected modeling algorithms. After performing an experiment using classification algorithms decision tree and rule induction, the models is evaluated based on their prediction accuracy in classifying the instances of the data set into thundered and non-thundered situations. From those classifier PART is selected by having best classifying accuracy that can classify 10718 or 99.70% instances as correct out of 10750 instances which is processed from Gondar aeronautics and Synoptic station)**

Keywords— Thunderstorm, synoptic data, Data mining, PART classifier, predictive model.

I.   INTRODUCTION

Thunderstorm is one of the most spectacular weather phenomena in the atmosphere. It is the towering cumulus or the cumulonimbus clouds of the convective origin and high vertical extent that are capable of producing lightning and thunder. Usually, these thunderstorms have the spatial extent of a few kilometers and life span less than an hour. However multi-cell thunderstorms developed due to organized intense convection may have a life span of several hours and may travel over a few hundreds of kilometers [2]. Rasika Kalbende and Nitin Shelke [12], define the development and occurrence of thunderstorm; when a layer of warm and moist air rises to a larger extent, and updrafts to the cooler regions of the atmosphere the updraft that contains moisture condenses in order to form massive cumulonimbus clouds and eventually leads to the development of precipitation. Almost all thunderstorms develop under atmospheric conditions of low static stability, with abundant heat and moisture at low levels. The formation begins when dense (sinking) cold air overlies less dense, warm and moist (rising) air. Development is greatly enhanced when a catalyst such as strong heating and/or a trough is present. Strong up-

draughts then gradually form and the heat energy in the air and water vapors gets converted to wind and electrical energy. When the atmosphere is sufficiently unstable and the immediate surrounding promotes continuous contribution of energy into a growing cloud, a severe thunderstorm then develops, Strong up- and down-draughts characterize a well-developed thunderstorm [3]. Mark Weber and Dimitris Bertsimas [14], describes Thunderstorms has significantly effects on both terminal and en route fights and reduce both terminal and en route airspace capacity that Results delays and have increased substantially in the past decade due to increase en route congestion. Current technology cannot provide reliable long-term forecasts of the aviation impact of thunderstorms. Even when good short-term forecasts are available, the current air traffic management system often cannot effectively exploit them to improve network flow because of workload and airspace management difficulties. Generally thunderstorm have the following risks such as severe turbulence, severe clear icing, large hail, heavy precipitation, low visibility, Gust Front, Downburst, Macro burst, Microburst and electrical discharges within and near the cell. Data mining (DM) techniques are very popular for solving various problems. As a brief description, data mining is a mechanism for obtaining patterns from an existing dataset. Those extracted patterns are used to interpret the new or existing data into useful information [5]. DM has a potential to identify hidden knowledge from huge datasets. Many researchers apply data mining to explore hidden pattern from met record data. This study will use data mining technique for developing best model for predicting the Thunderstorm using spatiotemporal and synoptic data to improve the prediction of thunder storms.

II.   PROBLEM STATEMENT

Weather forecasting or predicting is one of the most technical and technological problems around the world [3]. Prediction of significant weather components such as tornedo, thunderstorm and tunnel clouds has important effects on different economic and social activities of human being that helps to adjust him/herself with the event and to protect themselves from those weather effects. Especially forecasting thunderstorm is one of the most difficult tasks in weather prediction, due to their relatively small spatial and temporal extension and the inherent non-linearity of their dynamics and physics [2].as discussed in related works different Scholars studies about the

thunderstorm, its properties, different effect on biosphere and how to predict or forecast the occurrences using different variables and different machine learning algorithms, But the maximum accuracy from the previous study is 98% which is studied by Himadri Chakrabarty and Sonia Bhattacharya [1]. Even it predicts the occurrence with in 12 hour using one time observation data of a day. But as an Ethiopia especially in Gondar there are money occurrences are observed with in a 12 hour. The values of attributes that is registered at 00:00 is completely different values at 06:00. However these and other scholars' tries to predict the occurrence of thunderstorm using different techniques still it continuous as a challenge in aviation industries and other sectors. This study tries to increase the accuracy of the predictive model by using other data mining algorithm such as decision tree and other rule based algorithms and also consider effects of other additional attributes or variables which is not tested before such as spatiotemporal dataset which has its own role for occurrence of thunderstorm. The data set that all previous studies used were one time per day observation which is not contain every events that is occurred in a day while This study incorporates three observation time data to develop a model. Finally the study attempt to answer the following research questions:

What are the most determinant attributes that uses for occurrence of Thunderstorm?

Which mining algorithm produces best Thunderstorm prediction model?

A. Objective of the study

General objectives: The general objective of this study developing Data Mining Predictive model that used to predict occurrence of thunderstorm in the case of Gondar Control Zone.

Specific objectives: The specific objective of the study are:

- To understand the problem domain by reviewing literatures on DM technology and their application in the prediction of thunderstorm.
- To identify the determinant attributes (features) that has great role for the formation of thunderstorm
- To prepare the data for analysis, to apply classification algorithms, to train, test and build the models using synoptic dataset.
- To compare the models based on their performance and select the best model.

B. Scope of study

The aim of this study was determining and predicting the occurrence of thunderstorm using of DM techniques on spatiotemporal datasets and ten years synoptic data from 2007 to 2016 that is stored in Ethiopian national meteorology agency data base having three observation time (06:00, 09:00 and 12:00). The model predicts the occurrence of thunderstorm before three hour of the event.

C. Research methodology

According to Smolander et al. (1990) [41], a method can be considered as a predefined and organized collection of techniques and a set of rules which state by whom, in what order, and in what way the techniques are used to achieve or maintain some objectives. For this study a design science

research methodology which is developed by P. Ken, T. Tuure, A. R. Marcus and C. Samir (2007) [50], is used. DSRM provides a mechanism through which design, testing, and implementation of an IT artifact can be improved to the extent that it represents and the principle of what the artifact must to be [51].

Research design: This research was design to identify the determinate factors for the formation of thunderstorm and predicting the occurrence in in Gondar control zone. To explore the application of data mining on this particular research, hybrid (Ciso.et al) data mining methodology was employed. Because this model is both academic and industrial. Ciso.et al involves sixth iterative process or steps including: Understanding the problem domain, Understanding of the data, Preparation of the data, Data mining, Evaluation of the discovered knowledge, and Use of the discovered knowledge steps [3].

Literature review: In this study Relevant literatures related to Thunder storms and its significant effect, the overview of data mining, knowledge discovery process model, data mining tasks, knowledge base system, and related works and Various books, journals, magazines, articles manuals, conference papers and related works and resources were reviewed.

Implementation Tools: The following tools and application software were used to accomplish the research process: WEKA 3.7.5 data mining tools: WEKA stands for Waikato Environment for Knowledge Learning. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. Based on the prior knowledge of the researcher on the applicability of the tools for the purpose and freely availability of the software made the researcher to use the tools to build analysis and evaluate the model being developed for research goal. Ms-Excel: it will used for data preparation, pre-processing and analysis task because it has the capability of filtering attribute with different values. Besides, it is a very important application software to make ready the data and easily convert into of the file. Dataset preparation: method of Data collection & preprocessing.

Evaluation mechanisms and testing procedure: The data set collected from ENMA was preprocessed and divided into two data groups called training set and test set data. Using training set data the model was built using two decision tree and rule induction algorithms and the model tested using 10-fold cross validation test mode. These developed models in this research is compared using their classification accuracy and different confusion matrices (True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), False Negative Rate (FNR), precision, recall F-measure, Relative Operating Characteristics (ROC), the number of correctly classified instances, and number of leaves and the size of the trees, execution time.

D. Significance of the study

Weather forecasting is a vital application in meteorology and has been one of the most scientifically and technologically challenging problems around the world in the last century. Generally the study has the following major advantages in different fields of study.

ENMA: Ethiopian National Meteorology Agency (ENMA) forecast or predict the weather components using four categories. That is now casting- (current weather and forecasts up to a few hours ahead), Short range forecasts (1 to 3 days), Medium range forecasts (4 to 10 days) and   Long range /Extended Range forecasts (more than 10 days to a season). All types of forecasting or prediction is depend on the previous data and present observation of observers which is exposed for human and technical errors. So the develop model help for NMA employee or professionals to easily and accurately Predict and identify special weather elements basically for thunderstorm which has short duration and great impacts on nature.

Aviation industries: aviation operations are constrained by different weather elements and all air navigation operators needs weather information for their safe operations. The result of these study helps for operators to design their routes and to reduce numbers of delays that is occurred due to thunderstorm and other related weather elements.

others: the study also have significant advantages for Air traffic controllers, maritime navigators and emergency and rescue handling organizations to make decisions during their day to day activities.

II.   RELATED WORKS (STUDIES)

Different scholars or authors were tries to develop different predictive models using different techniques and uses different data sets having different attributes to predict the occurrence of thunderstorm. Himadri Chakrabarty and Sonia Bhattacharya [1], tries to predict the occurrences of thunderstorms by K-Nearest Neighbor Technique using three types of weather variables such as moisture difference, adiabatic lapse rate (temperature), and wind-shear. The model that is developed by Himadri Chakrabarty and Sonia Bhattacharya [1], can classify with 82% accuracy of occurrences and non-occurrences of thunderstorm in 12 hour. Litta A. J et al [2], also develops Artificial Neural Network Model that used to predict the occurrences of thunderstorm using surface temperature. Their result clearly indicated that overall accuracy is 76%.

Himadri Chakrabarty et al [3], uses Artificial Neural Network to Predict Squall-Thunderstorms Using RAWIND Data and develop a model that used to forecast 98% 'squall-storm days' and 'no storm days'. Himadri Chakrabarty and

Sonia Bhattacharya [4], in 2014 also tries to Predict Severe Thunderstorms using artificial neural network technique. Multilayer Perceptron (MLP) has been applied on the weather parameters of moisture difference, adiabatic lapse rate and vertical wind shear which were recorded by the radiosonderawind (RSRW) in the early morning at 06.00 am local time. MLP classified and predicted "severe storm" and "no storm" days nearly up to 70% having around 12 hours lead time.

Waylon Collins and Philippe Tissot [6], uses artificial neural network to forecast location of thunderstorm and develop a model having 50% accuracy.

III.   DATA PREPROCESSING

Preprocessing the data includes multiple steps to assure the highest possible data quality, thus efforts are made to detect and remove errors, resolve data redundancies, and to handle errors. Different techniques were used in data preprocessing of this study.

A.   Data Cleaning

Data cleaning (or data cleansing) routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data [14]. The data mining process get a confusion with the unclean data of real world database that contains incomplete, inconsistent and noisy data. Thus, data cleaning is mandatory in order to improve the quality of data that improves the performance of the data mining techniques.

Handling Missing Value: There are two methods to handle missing attribute values belong either to sequential methods (called also preprocessing methods) or to parallel methods (methods in which missing attribute values are taken into account during the main process of acquiring knowledge) [38]. This study used Sequential methods to handle missed values of the attributes from the collected data. For numeric attributes mean or average values was used and for nominal attributes highly frequent value or mode was used. The following table 1 below shows detail descriptions of attributes, total numbers of missed values, techniques that used to replaced missed values and the replaced values.

Table 1 attribute missing values, replacement techniques and replaced values

s. no	Attribute name	Number of Missed values	Data type	Method or techniques	Replaced with
1.	Cloud amount	450	Numeric	MODE	CU
2.	Temperature	110	Numeric	AVERAGE	28.25
3.	Rain Fall	102	Numeric	AVERAGE	32.06
4.	Relative Humidity	430	Numeric	AVERAGE	68
5.	Wind Speed	420	Numeric	AVERAGE	4.37
6.	Wind Direction	328	Numeric	MODE	270
7.	Pressure	320	Numeric	AVERAGE	1023.67
8.	Observation Time	-	Numeric		
9.	Elevation	-	Numeric		
10.	Latitude	-	Numeric		
11.	Longitude	-	Numeric		
12.	Sun shine	320	Numeric	MODE	9.8

B. Data reduction

Data reduction is one of the tasks in data mining that needs to be done before the actual mining task is takes place. Although, 11400 instances are collected, after preprocessing the redundant instance that is measured in similar observation time as weather report and special weather report were selected one and rejected or deleted the other and finally the researcher re-sampled the dataset into 10750 instances for data mining. After formulating the required sample dataset, the data is converted in to comma delimited Excel file (CSV format), then to the arff file which is suitable for mining using WEKA 3.7

C. Data Integration

Even if most data set of the study was collected from on data base some attributes such as Pressure and special weather components were collected from hourly registration log book which has different format and different data representation (i.e. symbolic and abbreviated forms of data). When matching those attributes from the database special attention was paid to integrate with similar event of other attributes structure of the data. This is to ensure that any attribute functional dependencies and referential constraints in the source system to match those in the target system.

D. Data Transformation

Among those techniques of transformation, data discretization / binning is selected for this study. Binning is used to reduce data size by dividing the range of a continuous attribute in to interval. Interval labels can then be used to replace the actual data values [30]. The collected data was transformed into a format appropriate for further Data Mining process by dimension reduction (such as feature selection and extraction, and record sampling), and attribute transformation (such as discretization of numerical attributes and functional transformation). The Divided range of continuous values was assigned into N intervals of equal sizes and labeled based on standards that is seated by WMO and discussion with domain expert.

IV. DATA MINING EXPERIMENTATION

Creating a Model is one of the major tasks which is undertaken under the phase of data mining in hybrid methodology. In this phase several data mining techniques are applied and their parameters are adjusted to optimal values. Some of the tasks include: - experimental setup or design, selecting the modeling technique, building a model and evaluating the model.

A. Experimental setup

In any data mining process before building a model, we need to generate a procedure or mechanism to test the model's quality and validity. In this research 10750 datasets are used for training and testing the model. WEKA 3.6.9 software has used to set up and measure the quality, validity and test of the selected model. K-fold (10-folds) cross validation was used because of relatively its preference and low variations [30]

B. Attribute selection

To select the best attribute subset selector, the investigator uses information gain attribute evaluator with ranker search method. Information gain attribute evaluator works by evaluating the worth of an attribute by measuring the information gain with respect to the class. As shown in figure 1 below, information gain attribute subset evaluator algorithm ranks the attributes based on the information gain with respect to class. The researcher selected 7 best attributes according to their rank from 12 independent attributes.

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 8 Thunder):
        Information Gain Ranking Filter

Ranked attributes:
  0.24792   7 Rainfall
  0.24269   2 lowor_CLD_TYPE
  0.14459   1 cloud_ammount
  0.12146   3 tempMax
  0.07483   4 RH
  0.0097    6 windspeed
  0.00202   5 pressure

Selected attributes: 7,2,1,3,4,6,5 : 7
```

Figure 1 Attribute selection table

C. Selecting modeling technique

Four predictive models involving J48, REPTree, PART and JRip classifier algorithms are constructed. J48, and REPTree are tree based classifiers in WEKA whereas PART and JRip are rule based classifiers. Four mining algorithms (J48, REPTree, PART and JRip) which were used to build the models yields for different models having different performance i.e. these algorithms yield a model having 99.53%, 99.25%, 99.70% and 99.57% accuracy respectively and all the modeling techniques have greater than 99% Recall, precision and F-measure. From all modeling techniques PART algorithm yield best performed model compared to all tested algorithms.

D. Evaluation of the developed models

In this study as discussed in table 2, the performance of the models is evaluated based on their prediction accuracy in classifying the instances of the data set into thundered and non-thundered situations. As we can see from the result of the experiments in the above scenarios, there is a slight difference between the results of each classifier. From those classifier PART is selected by having best classifying accuracy that can classify 10718 or 99.70% instances out of 10750 correctly. While other classifiers Results J48, JRip and REPTree show nearly equal number of incorrectly classified instances. The highest incorrect classification is scored by REPTree.

Addition to prediction accuracy, classifiers are also evaluated to measure how they correctly classified each class to their correct class or incorrectly classified to another class. Hence, to evaluate the performance of the classifiers used in this study True Positive rate, Precision, Recall and F- measure are used.

From all four experiments which was conducted using 10-fold cross validation PART rule induction classifier list is best based on accuracy which registered 99.70% and correctly classified instance which is accounts 10718 out of 107500 instances.

Considering the best learning models built by sing the four modeling techniques, all the modeling techniques have greater than 99% Recall, precision and F-measure. From all modeling techniques listed above table PART algorithm shows slightly high difference in all values as compared to the other techniques.

From the above four modeling technique PART gives the best results in predicting the occurrence of thunderstorm class as it can be seen its F-Measure value (0.997) is the highest as compared to the others. The J48 modeling technique, which shows almost equivalent predicting performance with PART having 0.996 F-measure values is the second best modeling technique for predicting the thundered class. Furthermore, based on criteria of minimum time taken for building model, experiment-2 (building model REPTree algorithm) is best since it built in 0.06 seconds.

Comparison of the models using classifying accuracy, ROC, Precision and time execution

As we show from the following figure 2 below and table 2 from four experiments PART has best classifying accuracy, ROC and precision compared to other classifiers.
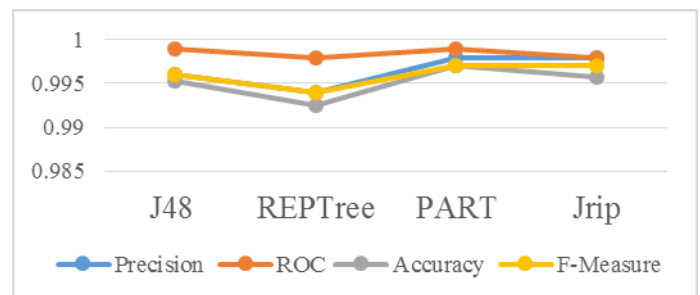

Figure 2: Comparison of the Algorithms Using Classifying Accuracy, ROC, and Precision.

PART has best performance among the four classifiers. As shown in figure 2, PART has preferred prediction accuracy, ROC value and precision. On the other hand PART has registered the least FP rate (0.3%) compared to the other three algorithms. However, as shown in figure 3. PART classifier takes much time than others. PART classifier algorithm has generated 81 rules.

to evaluate the performance of the classifiers used in this study

Classifier	TP RATE	FP RATE	INSTANCES				Precision	RECALL	F-Measure	ROC	Accuracy	T-execution	Class
			CCI		ICI								
			NO	%	NO	%							
J48	0.996	0.005	10699	99.53	51	0.47	0.996	0.996	0.996	0.999	99.53	0.39	No-thunder
	0.995	0.004					0.994	0.995	0.994				Thundered
REP Tree	0.993	0.008	10669	99.25	81	0.75	0.994	0.993	0.994	0.998	99.25	0.06	No-thunder
	0.992	0.007					0.99	0.992	0.991				Thundered
PART	0.997	0.003	10718	99.70	32	0.3	0.998	0.997	0.997	0.999	99.70	1.98	No-thunder
	0.995	0.003					0.997	0.997	0.996				Thundered
JRip	0.994	0.002	10704	99.57	46	0.43	0.998	0.994	0.996	0.998	99.57	15.74	No-thunder
	0.998	0.006					0.992	0.998	0.995				Thundered

True Positive rate, Precision, Recall and F- measure are used.

Table 2 TPR, Precision, Recall, F-Measure, ROC, Accuracy and time execution of the algorithms
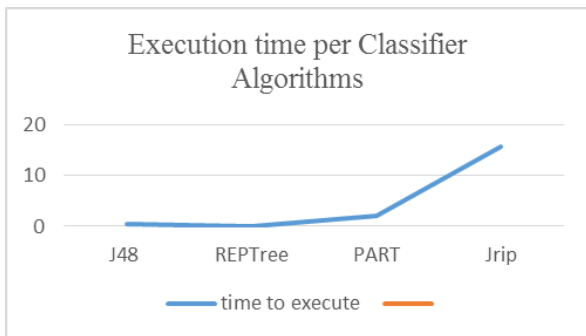

Figure 3. The Execution Time to build The Model

The researcher exhaustively discussed based on the result of the model with domain expert to determine variable that used to predict the occurrence of thunderstorms.

CCI: correctly classified instances TPR: True Positive Rate. FPR: False Positives Rate, ROC: Relative operating character curve, PR: precision rate, RR: Recall rate, ICI: Incorrectly classified Instances

situations.   As we can see from the result of the experiments in the above00 scenarios, there is a slight difference between the results of each classifier. From those classifier PART is selected by having best classifying accuracy that can classify 10718 or 99.70% instances out of 10750 correctly. While other classifiers Results J48, JRip and REPTree show nearly equal number of incorrectly classified instances. The highest incorrect classification is scored by REPTree.

Addition to prediction accuracy, classifiers are also evaluated to measure how they correctly classified each class to their correct class or incorrectly classified to another class. Hence,

In the other word the performance of the classifier is alsoevaluated in terms of different confusion matrices (True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), False Negative Rate (FNR), precision, recall F-measure, Relative.

Operating Characteristics (ROC), the number of correctly classified instances, and number of leaves and the size of the trees and execution time.

Based on the above result in experiment 3 (PART rule induction classifier list) is selected. Because from all four experiments which was conducted using 10-fold cross validation as we show from table 2 Experiment 3 (PART rule induction classifier list) is best based on accuracy which registered 99.7 % and correctly classified instance which is accounts 10718 out of 107500 instances.

V.   CONCLUSION AND RECOMMENDATION

Thunderstorms has significantly effects on both terminal and en route fights and reduce airspace capacity that Results delays aviation impact of thunderstorms. The goal of this research was to apply the data mining techniques to predict the occurrence of thunderstorms using synoptic dataset. The research uses design science research methodology and follow the six-step Cios et AL. (2000) hybrid data mining process model and 10750 datasets are used for training and testing the model using WEKA 3.7 Four mining algorithms (J48, REPTree, PART and JRip) were used to build the models gets for different models having different performance i.e. these algorithms yield a model having 99.53%, 99.25%, 99.70% and 99.57% accuracy respectively and all the modeling techniques have greater than 99% Recall, precision and F-measure. From all modeling techniques PART algorithm yield best performed model compared to all tested

algorithms. The rule induced from PART algorithm is used to develop the knowledgebase system using Swing prolog that is used to assist air traffic controllers to make decision. The results obtained from this study are very promising and the model also have better performance related to previous studies. The researchers thought that considering the special attributes such as cloud type, measured rainfall, cloud amount and current temperature improves the performance of the model. The results obtained from this research indicate that data mining is useful in bringing relevant information to the service providers (NMA) as well as decision makers (duty air traffic controllers). This research is mainly conducted for an academic purpose. However, the results of this study are found to be hopeful to be applied to address practical problems that is observed real life activity in aviation industries. This research work can contribute a lot towards a comprehensive study in this area in the future, in the context of our country. The results of this study have also shown that the uses of DM technologies with knowledge base system are well applicable in other weather element prediction. Hence, based on the result of this study the researcher believes further researches have to be done to increase the benefits of the developed model the following are recommended for future study

1. Using Satellite data addition to synoptic datasets to build the model: The Agency have geospatial database which recorded events throughout the day using satellite. Constructing the model using those images and satellite data is complete and more reliable than those lese quality data in that synoptic database.

2. Building a model by considering Topography, plants Distribution (vegetation index) and water surface of the earth which has significant role for development of thunder bring clouds. So considering these variables may give significant change on rules gain from the model and gives better prediction model

3. Integrating Data mining model with the Knowledge Based System.

REFERENCES

[1] H. C. a. S. Bhattacharya, "Application of K-Nearest Neighbor Technique to Predict Severe Thunderstorms," International Journal of Computer Applications (0975 – 8887), vol. Volume 110 1, no. 10, , pp. 1-4, January 2015.

[2] L. A. J. S. M. I. a. C. N. Francis, "Artificial Neural Network Model for the Prediction of Thunderstorms over Kolkata," International Journal of Computer Applications (0975 – 8887), vol. Volume 50 – , no. 11,, pp. 50-55, July 2012.

[3] C. A. M. S. B. a. A. D. G. Himadri Chakrabarty, "Application of Artificial Neural Network to Predict Squall-Thunderstorms Using RAWIND Data," International Journal of Scientific & Engineering Research, vol. Volume 4, no. 5, pp. 1313-1317, May-2013.

[4] F. A. a. U. Jr, " Introduction to Knowledge Management. ,," Jakarta, Indonesia: Asean Foundation, 2008.

[5] H. C. a. S. Bhattacharya, "Prediction of Severe Thunderstorms applying Neural Network using RSRW Data," International Journal of Computer Applications (0975 – 8887), vol. 89 –, no. 16,, pp. 1-3, March 2014.

[6] U. D. O. C. N. O. A. A. A. N. W. S. U. D. O. T. F. A. ADMINISTRATION, AVIATION WEATHER SERVICES Advisory Circular, AC 00-45G, Change 2, U.S.: Aviation Weather Services, Advisory Circular 00-45G Change 2 (October 2014), October 2014.

[7] W. C. a. P. Tissot**, "USE OF AN ARTIFICIAL NEURAL NETWORK TO FORECAST THUNDERSTORM LOCATION," Texas A&M University – Corpus Christi, Texas, 2008..

[8] M. E. S. et al, "Thunderstorm characteristics associated with RHESSI identified terrestrial gamma ray flashes," JOURNAL OF GEOPHYSICAL RESEARCH, VOL. , ,, vol. 115, no. A00E38, doi:10.1029/2009JA014622, pp. 1-10, 2010.

[9] D. O. T. F. A. A. F. i. S. Service, "AVIATION WEATHER For Pil sand Flight perations Personnel," asa PUBLICATIONS, Washington, D.C., Revised 1975.

[10] Aviation Weather Hazards,LAKP-Prairies NAV CANADA, 2009.

[11] M. J. B. A. G. S. Linoff, Data Mining Techniques For Marketing, Sales, and Customer Relationship Management Second Edition, Indiana: Wiley Publishing, Inc., 2004.

[12] M. N. A. a. S. G. Shiva, "Comparative Analysis of Serial Decision Tree Classification," International Journal of computer science and security, vol. 3, no. 3, pp. 230-238, 1995.

[13] e. a. Fayyad Usama, "From Data Mining to Knowledge Discovery in Databases," AI Magazine, vol. 17, no. 3, pp. 37-50, 1996.

[14] G. M. a. J. S. Óscar Marbán, "A Data Mining & Knowledge Discovery Process Model," Data Mining and Knowledge Discovery in Real Life Applications, , vol. I, no. ISBN 978-3-902613-53-0,, pp. 1-9, 2009.

[15] J. H. a. M. Kamber, Data Mining: Concepts and Techniques Second Edition, AMS T E RDAM BOS TON: by Elsevier Inc., 2006 .

[16] M. Mary K. Obenshain, "Application of Data Mining Techniques to Healthcare Data," chicago journals, vol. 25, no. 8, pp. 690-695, August 2004.

[17] M. S. P. D. a. D. V. M. Thakare, "DATA MINING SYSTEM AND APPLICATIONS:," International Journal of Distributed and Parallel systems (IJDPS), vol. I, no. I, pp. 32-41, September 2010.

[18] M. F. S. Ana Azevedo, "KDD, SEMMA AND CRISP-DM A PARALLEL OVERVIEW," in IADIS European Conference, Portugal, 2008.

[19] C.-S. Chang, "Knowledge Discovery from Dynamic Data on a Nonlinear System," Open Journal of Applied Science scintific research publishing, , vol. 5, pp. 576-585, 2015.

[20] L. A. K. a. P. MUSILEK, "A survey of Knowledge Discovery and Data Mining process models," The Knowledge Engineering Review, Vol. :1, 1–24. 2006, Cambridge University Press, vol. 21, no. 10.1017/S0269888906000737, pp. 1-22, 2006.

[21] M. M. N. J. Sara khan, "A Critical Review of Data Mining Techniques in Weather Forecasting Weather Forecasting," International Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no. 4, pp. 1-4, April 2016.

[22] S A F E T Y A D V I S O R Weather No. 4 Thunderstorms and ATC,AOPA Air Safety Foundation, 2008,.

[23] R. K. N. Shelke, "A Novel Approach for Thunderstorm and Lightning Detection System," International Journal of Science and Research (IJSR), no. ISSN (Online): 2319-7064, pp. 1-5, 2015):.

[24] D. B. D. Bharati M. Ramageri, "ROLE OF DATA MINING IN RETAIL SECTOR," International Journal on Computer Science and Engineering (IJCSE), vol. 5, no. 01, pp. 47-51, Jan 2013.

[25] D. B. Mark Weber, "IMPROVING AIR TRAFFIC MANAGEMENT DURING THUNDERSTORMS," in Avionics Systems Conference, Cambridge,, 2005.

[26] I. U. a. N. K. Enete, "IMPACTS OF THUNDERSTORM ON FLIGHT OPERATIONS IN PORT-HARCOURT INTERNATIONAL AIRPORT OMAGWA, RIVER STATE, NIGERIA," International Journal of Weather, Climate Change and Conservation Research, vol. 1, no. 1, pp. 1-10,, March 2015.

[27] D. O. C. USA, "Thunderstorm Conditions Affecting Flight Operations," THUNDERSTORM PROJECT, CHICAGO, March 1949.

[28] A. M. S. &. S. workshop, "Integrating Space Weather Observations & Forecasts into Aviation Operations," American Meteorological Society (AMS), USA, March 2007.

[29] LAW ON CIVIL AVIATION, 26 March 2007.

[30] A. CHALI, "An Integration of Prediction Model with Knowledge Base System for Motor Insurance Fraud Detection:," ADDIS ABABA UNIVERSITY, ADISS ABABAETHIOPIA, February 2016.

[31]   E. F. Ian H. Witten, Practical Machine Learning Tools and Techniques,Second Edition, SAN FRANCISCO: Elsevier Inc, 2005 .

[32]   A. Azevedo, "KDD, semma and CRISP-DM: A parallel overview," ResearchGate, no. ISBN: 978-972-8924-63-8, pp. 1-5, 02 June 2014.

[33]   Daniel T. Larose, Data Mining Methods and Models, John Wiley and Sons, Inc., 2006.

[34]   M. S. Dr. Sudhir B. Jagtap, "Census Data Mining and Data Analysis using WEKA," (ICETSTM – 2013) International Conference in "Emerging Trends in Science, Technology and Management-2013, Singapore, pp. 35-38, 2013.

[35]   T. C. Corporation, Introduction to Data Mining and Knowledge Discovery, U.S.A.: Two Crows Corporation, 1999.

[36]   L. Rokach, "DECISION TREES," ResearchGate, pp. 165-187, 2005.

[37]   K. K. a. N. Kerdprasop, "Bridging Data Mining Model to the Automated Knowledge Base of Biomedical Informatics," International Journal of Bio-Science and Bio-Technology, vol. 4, no. 1, pp. 13-28, 2012.

[38]   H. M. M. e. al, "INFLUENCE OF MODELS AND SCALES ON THE RANKING OF MULTIATTRIBUTE ALTERNATIVES," Brazilian Operations Research Society, pp. 524-538, September 19, 2011.

[39]   O. M. ·. L. Rokach, Data Mining and Knowledge Discovery Handbook, New York: Springer Science+Business Media, 2010.

[40]   K. J. C. W. P. R. W. S. L. A. Kurgan, Data Mining A Knowledge Discovery Approach, USA),: Springer Science+Business Media, 2007 .

[41]   "National Metreology Agency," ENAMA, [Online]. Available: https://public.wmo.int/en/media/news-from-members/ethiopia-national-meteorology-agency. [Accessed 23 Dcember 2017].

[42]   ICAO, Meteorological Service for International Air Navigation, Annex 3, International Civil Aviation, on 7 November 200 ,.

[43]   ICAO, Safety Management Manual (SMS), Safety Management, Third Edition — 2012.

[44]   Procedures for Airnavigation scervice (DOC-4444)-Air Traffic Management, International Civil Aviation Organization, Fifteenth Edition — 2007.

[45]   An Introduction to the WEKA Data Mining System, May 2005.

[46]   RobertGrossman, "Synthesis Lectures on Data Mining and Knowledge Discovery," in Synthesis Lectures on Data Synthesis Lectures on Data Discovery , Chicago, Morgan & Claypool Publishers series, 2010.

[47]   K. P. Tripathi, "A Review on Knowledge-based Expert System:Concept and Architecture," IJCA Special Issue on "Artificial Intelligence Techniques - Novel Approaches & Practical Applications", pp. 1-5, 2011.

[48]   D. S. a. S. J. F. M. L. Maher, "Tools and Techniques for Knowledge Based Expert Systems for Engineering Design," Carnegie Mellon University Research Showcase @ CMU, no. DRC-12-22^84, December, 1984.

[49]   B. Aebissa, "developing A Knowledge based ssystem for coffee disease diagnosis and treatment," AAu, 2012.

[50]   "Matthew Huntbach,"," Dept of Computer Science, Queen Mary and Westfield College, AI 1 Notes on semantic nets and frames," London , , 1996..

[51]   Ton de Jong, ""Types and Qualities of knowledge," EDUCATIONAL PSYCHOLOGIST, vol. 31, no. (2),, pp. 105-113, , 1996..

[52]   T. T. A. R. M. a. C. S. P. Ken, " A Design Science Research Methodology for Information Systems Research," , vol. 24, no. 3, pp. 5 - 78,, 2007.

[53]   J. L. Bond, "An Investigation into the Benefits of Design Science Research for the Development of Wicked Educational Information Systems: A Case Study," ., 2014.