A Hybrid Data Model for Prediction of Disaster using Data Mining Approaches

Arjun Singh[#1], Prof. AmitSaxena[*2]

Department of Computer Science Engineering [#1],
Truba Institute of Engineering and Information Technology, Bhopal
Address-290/2, Shakti nagar near bhagatsinghchowk
Bhopal (M.P.),INDIA
*Head of Department of Computer Science Engineering [*2],*
Truba Institute of Engineering and Information Technology, Bhopal
Address-Truba Institute of Engineering and Information Technology, Bhopal
Bhopal (M.P.),INDIA

Abstract**—***The disasters are the unwanted event which affect the human life and human not able to recover form it's effects by available resources. Therefore these events are classified according to the natural or the human initiated. The human initiated disasters are controllable but the natural events are unpredictable in nature. In the literature the different data models are available that claims to understand and predict the much nearer data for any unpredictable events such as stock market, frauds and others. All these techniques are developed with the help of data mining techniques. That involves the computational algorithms for learning with historical patterns and predicts the data according to current situations. By the motivations of these techniques in this presented work a new data model is proposed which collect the real world knowledge from the web data sources and use with the data mining algorithms for predicting the unpredictable natural disasters. In order to perform this K-means clustering and HMM based hybrid model is proposed. The k-means clustering algorithm helps to prepare the disaster observations and the HMM algorithm is used for training and prediction of the disaster events. The implementation of this model is performed using the JAVA technology. After implementation the results are considered in conditions of exactness, error rate, memory expenditure and the time consumption. The results show the presentation of the futuremethod is accurate and efficient for prediction with any kind of other data also.*

Keywords**—** *natural disaster, prediction, data mining, HMM, K-means, Hybrid data mining*

I. INTRODUCTION

Data mining is a club of techniques for performing analysis of raw data. The analysis of the data is performed for finding the application centric data extraction, grouping of data, classification of similar patterns, object recognition or prediction and decision making. For performing such complicated task the computational algorithms are applied on the data to refine them, computationally analyses them and provide the meaningful patterns from the data. Therefore the methods and techniques have the potential to understand the pattern of available data and approximate the upcoming pattern. Such kind of data modelling is termed as the predictive data modelling.

In most of the cases for predictive techniques implementation the supervised learning techniques are used. The supervised learning functions in two major modules training and testing. Training process evaluates the data pattern and during the testing the algorithm recognize the similar pattern data. The main advantage to use the supervised technique is their performance and accurate outcomes as compared to unsupervised approaches of learning. In this presented work the predictive data modelling is the key area of interest. Therefore a supervised learning model is proposed for design and implementation for predicting the natural events. The key idea behind the concept of disaster prediction is taken from the different natural event prediction techniques such as weather prediction or stock market forecasting. Thus a new model for prediction of disasters according to the places proposed to obtain.

The disasters are the unwanted natural events that produce a significant amount of losses in terms of human life and the revenue. In this context the disaster management techniques are deployed to reduce or control the loss. According to the disaster management techniques the entire process is handled in three major phases before disaster preparedness that helps to aware about the target disasters and their complexities. Secondly during the disaster relief, which provides the help to handle current situations of disasters finally the recovery in this phase the social, economic and support based recovery is performed. According to proposed hypothesis if the disasters are predicted before it's happening then a lot of losses can be prevented. Therefore in this presented work a new data mining based model is presented that helps to predict the unwanted natural events. Thus the following intermediate objectives are included for work.

1. Study about the different kinds of disasters in India: in this phase the different natural disasters are studied that frequently affect the human life much frequently.
2. Study the predictive data mining techniques: in this phase the different data mining and

prediction techniques are studied that are help to design an accurate data model.

3. Design and implementation of new data model by which prediction of disaster becomes feasible: in this phase a new data model using the experience of the study is proposed and implemented for prediction of the disasters prediction system is performed.
4. Performance analysis of the proposed data model: in this phase the proposed data model is evaluated on different performance parameters. Therefore the accuracy and the resource consumption is measured as performance parameters.

II. PROPOSED WORK

The prediction is a task of historical data analysis and approximation of new predictive values as predictive results. In this chapter the futureprognostic model and their functional aspects are described. That accepts the historical news data and process the data for finding the future patterns of natural disasters.

A. Domain overview

Data mining techniques applied for finding the different data patterns over the set of historical data. This concluded patterns are help to recognize the patterns of data or help to predict the similar outcomes based on the learning algorithm's analysis. In this context learning term is used for supervised learning. In this technique first the computational algorithm computes the patterns for learning and then applied the algorithm on fresh data for recognizing the similar patterns of data. This kind of computing is termed as prediction task or classification task. Not only the algorithms are supervised some algorithms also works on the technique of unsupervised learning patterns. In this work the supervised approach is followed, which is accurate and efficient than the unsupervised learning techniques.

The main aim of the work is to extract the web data and NEWS information from web. After data extraction need to refine data using pre-processing techniques, cluster them in similar data groups and train the predictive algorithm. After training the subsequent patterns are produced as input to approximate the new upcoming natural event prediction. In order to perform the entire task first the Google News API is used to extract the data from web data source. In further the pre-processing techniques are employed on the raw data for cleaning and filtering. In next step the K-means algorithm is implemented for clustering the data and recognizing the observation matrix for the prediction algorithm. Finally the hidden Markov model is applied on the new trend for finding the most likely natural event among possible one. This section provides the overview of the proposed work involved in this study and the details of the entire model design are demonstrated in the next section.

B. Methodology

The proposed methodology contains the four key stages for process the data and obtaining predictive patterns using the proposed data model as discussed in figure 2.1.According to the above given figure 3.1, the entire system contains four individual components. That accepts the outcome from the previous step and produces the enhanced data to use with the learning task. Additionally individual component has its own advantage. That is also responsible to complete the aim a time to pass out come to next phase.

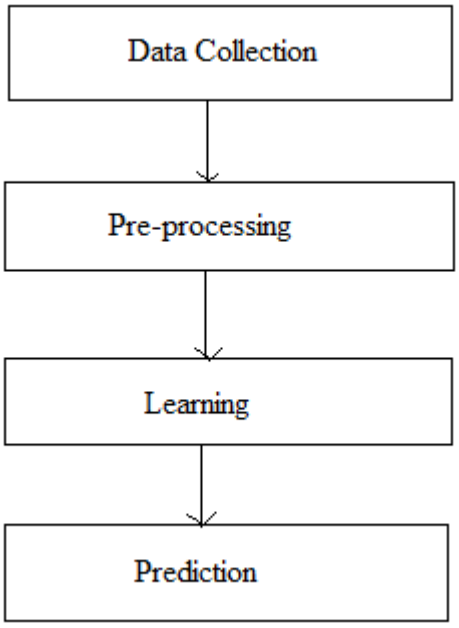

Figure 2.1 methodology phases

Data collection

Aim: the main aim is to design this module, is to find the text data from the web data source. This web data help us to maintain the fresh knowledge base from real world data.

Component model: the proposed data model for collection of data from real world and for creating a strong database to manage different patterns of disasters a continuously growing database is prepared. Additionally it reduces the time by utilizing data loader at the same time. To provide the functional aspects and the responsibilities the individual components are described as:

1. ***User interface:*** That is the GUI (graphical user interface) based user interface for consuming the services provided by the solution. Here user can navigate with the system to regulate the system

processes according to their requirements on the different other kinds of data.

2. ***Array of disasters:*** using the provided user interface, the provision is made to provide initial input to the system. Therefore user here put come common names of disasters. This disaster list is variable according to the needs. For example Droughts, flash floods, cyclones, avalanches, landslides.

3. ***Data extractor:*** that component is developed using the Google search API (application programming interface). That perform two important task:

a. Extract the data from keyword database and put them into a search engine

b. Execute the data from search engine, and collect the results obtained by the Google search API.

4. ***Cleaning of data:*** the data extracted is in HTML format therefore here HTML data filtered for extracting the text from HTML pages. The HTML to text conversion is the main aim of this phase.

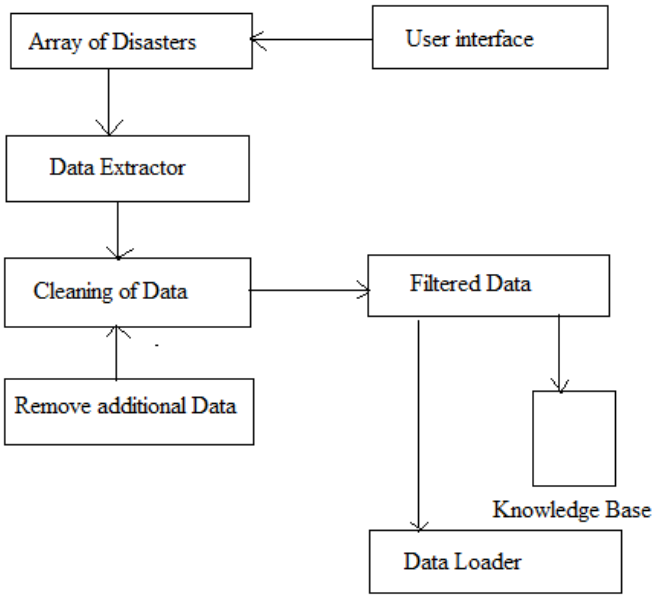

Figure 2.2 data collection

5. ***Remove Additional Data:*** The extracted data form HTML has also various noisy contents as the special characters. This in this phase additional data is removed from the extracted text and the data is maintained in a three column data structure. In first column the title of search, contents of search (only text), and source URL is stored.

6. ***Refined data:*** after completing the complete process of data refinement here the outcome available as the three column data structure.

7. ***Knowledge Base:*** that is the storage unit of the system which contains some data tables. This knowledge base is prepared in order to store newly arrived data or fresh data from the internet and also used load the previous data during new experiments.

8. ***Data loader:*** in this step the previous database and current extracted data is combined. This combined data forwarded to next phase for pre-processing and feature extraction.

Pre-processing

Aim: the main aim of this module design is to reduce the addition text or reduce the dimension of data. This help to reduce the amount of storage overhead of the system.

Description: the data obtained from the previous phase is obtained here in the following form as demonstrated in table 2.1.

Title	Content	URL
Natural disasters in India - Wikipedia, the free encyclopedia	Droughts, flash floods, cyclones, avalanches, landslides brought on by torrential rains, and snowstorms pose the greatest threats. A natural disaster might be caused by earthquakes, flooding, volcanic eruption, landslides, hurricanes etc.	https://en.wikipedia.org/wiki/_Natural_disasters_in_India
….	…..	….

Table 2.1 Previous phase outcome

The filter is prepared here to clean the additional data from the exiting data as demonstrated in table 2.1. After applying the filter the refined data is provided using the table 2.2

Title	Content	1. URL
Natural disasters in India Wikipedia the free encyclopedia	Droughts flash floods cyclones avalanches landslides torrential rains snowstorms greatest threats natural disaster caused earthquakes flooding volcanic eruption landslides hurricanes	en.wikipedia.org/wiki/Natural_disasters_in_India
….	…..	….

Table 2.2 pre-processed data

This refined data is produced to classify the data according to the knowledge exit on the database according to their relevancy of the keywords database. The bays classification algorithm is described as:

The Naive Bayes classification algorithmic rule is a probabilistic classifier. It is based on probability models that incorporate robust independence assumptions. The independence assumptions usually don't have an effect on reality. So they're thought of as naive. You can derive probability models by using Bayes' theorem (proposed by Thomas Bayes). Based on the nature of the probability model, you'll train the Naive Bayes algorithm program in a much supervised learning setting. In straightforward terms, a naive Bayes classifier assumes that the value of a specific feature is unrelated to the presence or absence of the other feature, given the category variable. There are two kind of likelihood as follows [26]:

- Posterior Probability [P (H/X)]
- Prior Probability [P (H)]

Where, X is data tuple and H is some hypothesis. According to Baye's Theorem

$$P\left(\frac{H}{X}\right) = \frac{P\left(\frac{X}{H}\right)P(H)}{P(X)}$$

The refined outcome of this stage is produced into next stage for performing the learning the next section describes the learning process of the system.

Learning

To demonstrate the learning process a hybrid techniques using two different algorithms are provided namely K-means algorithm and HMM algorithm. Therefore first both the algorithms are described first. After that the proposed learning model is provided.

K-means clustering

The K-Means clustering algorithm is a partition-based cluster assessmentmethod [27]. According to the algorithm we initially select k objects as initial cluster centers, after that calculate the distance amongevery object and every cluster center and allocate it to the next cluster, update the averages of all clusters, replicate this process until the criterion function congregate. Square error criterion for clustering

$$E = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left\|x_{ij} - m_i\right\|^2$$

$x_{ij}$is the sample j of i-class, $m_i$is the center of i-class, n_i i is the number of samples of i-class. K-means clustering algorithm is basicallyexplain as

Input: N objects to be cluster (xj, Xz … xn), the number of clusters k;
Output: k clusters and the sum of differenceamongevery object and its next cluster center is the smallest;
Process: 1. Arbitrarily select k objects as initial cluster centers$(m_1, m_2, \dots, m_k)$; 2. Calculate the detachmentamongevery object Xi and each cluster center, then dispenseevery object to the nearest cluster, formula for calculating distance as: $d(x_i, m_i) = \sqrt{\sum_{j=1}^{d}\left(x_i - m_{j1}\right)^2}, i = 1 \dots N, j = 1 \dots k$ $d(x_i, m_i)$is the distance among data i and cluster j. 3. Calculate the mean of objects in every cluster as the new cluster centers, $m_i = \frac{1}{N}\sum_{j-1}^{n_i} x_{ij}, i = 1,2, \dots, K$ N_i is the number of samples of current cluster i; 4. Repeat 2) 3) until the criterion function E converged, return $(m_1, m_2, \dots, m_k)$ Algorithm terminates.

Table 2.3 K-mean algorithm steps

Hidden Markov Model

An HMM is a double entrenched stochastic process with two hierarchy levels. It can be used to model much furthermultifaceted stochastic procedure as evaluate to a conventional Markov model. In anexact state, surveillance can be produce according to aconnectedlikelihoodallotment. It is only the surveillance and not the state that is discernible to an external spectator. An HMM can be characterized by the following [28]:

1. N is the number of states in the model. We denote the set of states' $S = \{S_1, S_2, \dots, S_N\}$, where S_i, $i = 1,2,..N$ is ahuman being state. The state at time instant t is signifying by qt.
2. M is the number of distinct observation symbols per state. We denote the set of symbols V = {V_1; V_2; ...V_M}, where V_i, I = 1;2;...; M is an individual symbol.
3. The state transition probability matrix A = [a_{ij}], where

$$a_{ij} = P\left(q_{t+1} = S_j \middle| q_t = S_i\right), 1 \le i \le N, 1 \le j \le N; t = 1,2 \dots$$

Here $a_{ij} > 0$ for all i, j. Also,

$$\sum_{j=1}^{N} a_{ij} = 1, 1 \le i \le N$$

4. The remark symbol probability matrix B = $\{b_j(k)\}$, where

$$b_j(k) = P(V_k|S_j), 1 \le j \le N, 1 \le k \le M \text{ and}$$

$$\sum_{k=1}^{M} b_j(k) = 1, 1 \le j \le N$$

5. The initial state probability vector r= πi , where

$$\pi_i = P(q_1 = S_i), 1 \le i \le N$$

Such that

$$\sum_{i=1}^{N} \pi_i = 1$$

6. The remark sequence $O = O_1, O_2, \dots, O_R$, where each remark O_t is one of the symbols from V, and R is the number of remarks in the sequence.

It is apparent that a complete requirement of an HMM wants the estimate of two model parameters, N and M, and three likelihood distributions A, B, andπ. We use the notation λ = (A; B; π) to identify the absolute set of parameters of the model, where A, B absolutely contain N and M.

Asurveillance sequence O, as mentioned over, can be generated by variousprobable state sequences. Observe as one such particular sequence $Q = Q_1, Q_2, \dots, Q_R$; where Q_1 is the initial state. The probability that O is generated from this state sequence is given by

$$P(O|Q, \lambda) = \prod_{t=1}^{R} P(O_t|q_t, \lambda)$$

Where statistical independence of observations is assumed Above Equation can be expanded as

$$P(O|Q, \lambda) = b_{q1}(O_1) b_{q2}(O_2) \dots \dots b_{qR}(O_R)$$

The probability of the state sequence Q is given as

$$P(Q|\lambda) = \pi_{q1} a_{q1q2} a_{q2q3} \dots \dots a_{qR-1qR}$$

Thus, the likelihood of production of the surveillance sequence O by the HMM specified by can be written as follows:

$$P(O|\lambda) = \sum_{all\ Q} P(O|Q, \lambda)\, P(Q|\lambda)$$

Deriving the value of $P(O|\lambda)$ using the direct definition of is computationally intensive. Hence, a process named as Forward-Backward process is used to compute$P(O|\lambda)$.

Aim of Learning: the main goal of this process is to learn with the sequences of disasters occurred in a place and use the learning to predict according to the current scenarios.

Component Description: the proposed learning model is described using figure 3.2. In this diagram the outcome of the previous step is produced as input that input is ***pre-processed*** data. This data is first used with the k-means clustering algorithm. The clustering algorithm groups the similar data for creating observations additionally the disaster events are recognized as the states. Using filtered data the surveillance matrix and evolution matrix is prepared. Finally these prepared matrixes are used with the hidden Markov model. Using this prepared data models using transition and observation matrix the learning process is performed.

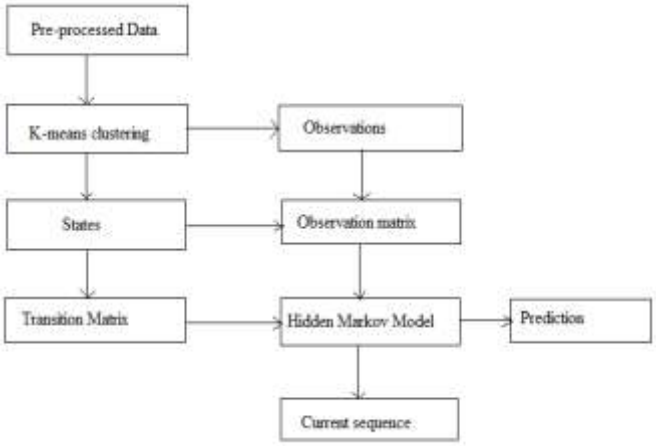

Figure 2.3 learning model

Prediction

In this phase the figure 2.3 is used to predict the data. Thus here the current input sequence of disasters is produced and the upcoming next event is predicted.

C. Proposed algorithm

This section includes the four different phases (i.e. data collection, pre-processing, learning and prediction) into a single process steps. This process steps are termed here as the algorithm of the proposed system.

Input: List of disasters D_l, current sequence S Output: Predicted event P
Process: 1. $R_l = ReadDisasterList(D_l)$ 2. $for(i = 0; i \le l; i++)$ a. $NewSR =$

$GoogleAPI.Search(R_i)$

b. $SR = SR + NewSR$

3. End for
4. $PRO = Preprocess(SR)$
5. $[RL, IR] = Bays.classify(PRO)$
6. $[ST, O] = Kmeans.DoCluster(RL)$
7. $T_m = HMM.Train(ST, O)$
8. $P = T_m.predict(S)$

Table 2.4 proposed algorithm

III. RESULTS ANALYSIS

After successfully implementation of the proposed hazard prediction model the performance of the technique is measured. Additionally the consideredpresentation of the implemented system is demonstrated using the line graph. The details of obtained performance are reported through this chapter.

A. Accuracy

Accuracy is the measurement of the correctness of a predictive or classification data model. According to the accuracy evaluation the total correctly predicted values are results as the accuracy of the data model. The accuracy of the system can be computed using the following formula.

$$accuracy = \frac{total\ correctly\ classified\ patterns}{total\ patterns\ produced\ to\ classify} X100$$

Dataset Size	Accuracy %
30	78.52
70	84.28
160	82.51
240	86.26
432	88.73
672	84.61
857	90.42

Table 3.1 accuracy

The computation of the performance is performed during the experimentation with the different size of datasets. The presentation of the futuremethod in terms of accuracy percentage is demonstrated using line graph in figure 3.1 additionally the corresponding values of the accuracy according to the size of data set is give using table 3.1. In this diagram to represent the performance X axis contains the amount of data produced for training and testing and the Y axis demonstrate the gainpresentation in conditions of proportion. According to the acquire results the performance of the system is improved as the amount of good feature data is captured. Thus the proposed data model is adoptable and efficient for prediction of natural hazards.

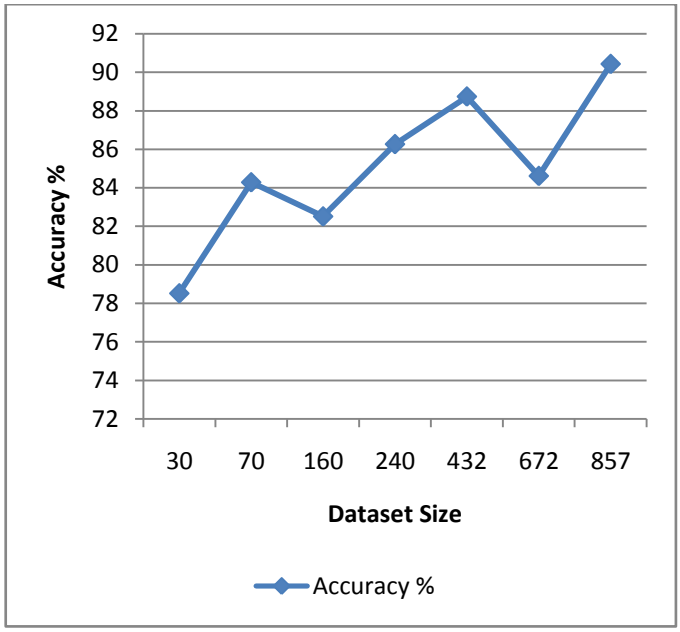

Figure 3.1 accuracy

B. Error rate

The error rate of the data mining technique provides the information about the incorrectness of the developed data model. Therefore the error rate is the amount of data which not properly classified or predicted with the help of trained learning algorithm. That is computed using the following formula.

$$error\ rate = 100 - accuracy$$

Or

$$error\ rate = \frac{total\ misclassified\ data}{total\ data\ to\ classify} X100$$

Dataset Size	Error rate %
30	21.48
70	15.72
160	17.49
240	13.74
432	11.27
672	15.39
857	9.58

Table 3.2 error rate

Figure 3.2 error rate %

The performance of the implemented system is evaluated in terms of the error rate percentage. The table 3.2 shows the error rate values corresponding to the dataset instances produced for training and testing. Additionally their graphical representation in terms of line graph is given using figure 3.2. In this diagram the amount of dataset size in terms of number of instances (dataset rows) are given in X axis and the obtained percentage error rate is given using the Y axis. According to the established results the error rate of the system is reducing and becomes consistent after 2000 rows of the data. Therefore the proposed data model is efficient and accurate with consistent performance.

C. Memory usages

In computation and the algorithm analysis the amount of main memory is required to execute the algorithm with the inputs. The amount of main memory requirement is termed as the memory usages or the space complexity of the system. The main memory requirement of the proposed system is also computed and reported in this section. For implementation of the proposed system the JAVA technology is used therefore to compute the required memory utilization by the algorithm the following formula is used.

$$memory\ usage = total\ memory - free\ memory$$

The memory requirement of the system according to the amount of data size and their obtained values are demonstrated using table 3.3 and the figure 3.3 contains the graphical representation of the memory usages data. According to the diagram the X axis contains the amount of data to be procedure and the Y axis demonstrates the consumed memory in terms of KB (kilobytes). According to the performance analysis the proposed technique consumes adoptable or wearable amount of memory resource for computing the accurate outcomes.

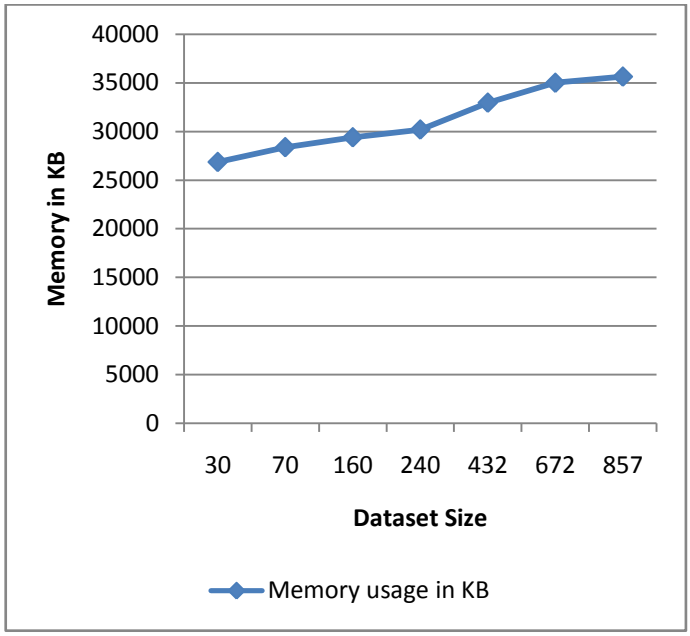

Figure 3.3 memory usages

Dataset Size	Memory usage in KB
30	26847
70	28371
160	29387
240	30183
432	32947
672	34992
857	35628

Table 3.3 memory usage

D. Time consumption

The amount of consumed time for processing the input amount of data is termed as time expenditure of the system. That is also termed as the time complexity of the learning algorithm. The computation of the time consumption can be performed using the following formula.

$$time\ consumed = computing\ end\ time - computing\ start\ time$$

Dataset Size	Time consumed MS
30	42
70	79
160	154
240	236
432	442
672	697
857	905

Table 3.4 time consumed

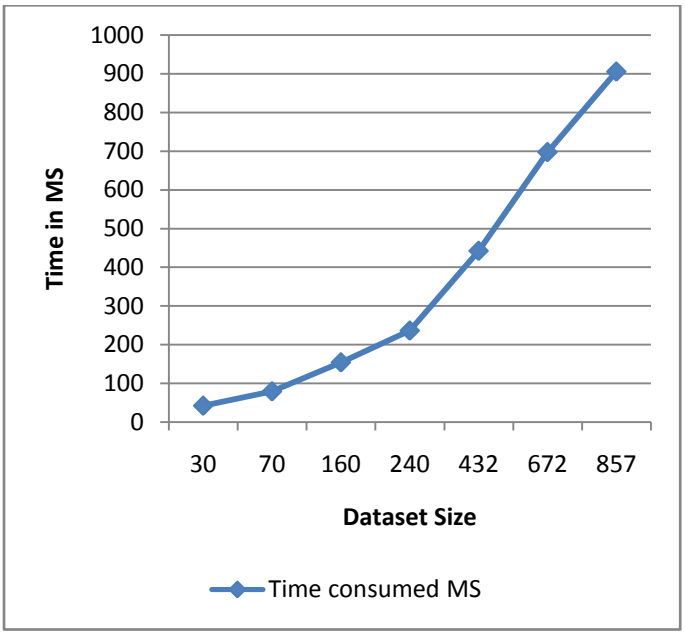

Figure 3.4 time consumed

The presentation of the system in terms of time consumption is demonstrated using table 3.4 and figure 3.4. In this diagram the X axis shows the amount of data to be procedure in terms of number of instances and the Y axis shows the amount of time consumed in terms of milliseconds. According to the find results the appearance of the prospecttechnique is adoptable and less time overriding as evaluate to the conventional systems.

IV. Conclusions

This chapter provides the summary of the proposed work performed for designing and implementation of the predictive data model. This model helps to predict the disaster events using the current patterns input. The summary of entire work performed is given as the conclusion and the future extension of the work is also suggested with the work.

A. Conclusion

The proposed model is a hybrid data model for proving the accurate data analysis and similar patterns prediction. Therefore first the real world data is computed with the help of Google search API and then the pre-processing is performed for refinement of the obtained data. Finally the pre-processed data is used with the Bays classifier to classify the text data according to their relevancy of search outcomes. Finally the data is used with the learning model, the learning model compile the data using two algorithms namely K-means clustering and the HMM algorithm. The K-means clustering is used here to prepare the observations from the data and the states are the disaster events. Finally by providing the current sequences the upcoming disaster events are predicted.

The implementation of the proposed data model is performed with the help of JAVA technology. Additionally after implementation the presentation of the system is also computed. The measured performance is given in terms of percentage accuracy and error rate produced by the system and for measuring the efficiency the time and space complexity is computed. The obtained performance is summarized using the table 4.1.

S. No.	Parameters	Remark
1	Accuracy	The accuracy of the data model is enhances with the large amount of data but fluctuating with the noisy data
2	Error rate	The error rate reduces with the less complex datasets and increases with noisy data patterns
3	Memory usages	The memory consumption is depends on the amount of data produced for training and testing
4	Time consumption	Time is also in similar ratio affected as the amount of data increases or decreases

Table 4.1 performance summary

According to the obtained performance summary the proposed model is accurate and efficient both. Therefore that is suitable to use for high accurate prediction requirements. Additionally that consumes fewer resources during predictions.

B. Future work

The main aim to develop a hybrid classification technique for predicting the disaster event is developed successfully. Additionally that is an accurate and efficient data model with less resource consumption. In near future the following extensions of the work are possible.

1. Extend the proposed model with the season based observations by which the precise outcomes are promises to achieve
2. Implement the system with the professional Google API which produces more authentic outcomes for preserving with the database.

REFERENCES

[1] Li Zheng, Chao Shen, Liang Tang, ChunqiuZeng, Tao Li, Steve Luis, and Shu-Ching Chen, "Data Mining Meets the Needs of Disaster Information Management", IEEE Transactions on Human-Machine Systems, Vol. 43, No. 5, September 2013 451

[2] Disaster Risk and Resilience, UN System task team on the post 2015 UN development agenda,http://www.un.org/en/development/desa/policy/untaskteam_undf/thinkpieces/3_disaster_risk_resilience.pdf

[3] Safer Homes, StrongerCommunities: A Hand Book For Reconstructing After Natural Disasters, http://www.gfdrr.org/sites/gfdrr.org/files/Disaster_Types_and-Impacts.pdf

[4] UCL, "EM-DAT: The OFDA/CRED International Disaster Database," UCL, http://www.emdat.be.

[5] NATURAL HAZARDS AND DISASTER MANAGEMENT, CENTRAL BOARD OF SECONDARY EDUCATION PREET VIHAR, DELHI – 110092

[6] T. Tingsanchali, "Urban flood disaster management", Procedia Engineering 32 (2012) 25 – 37, Published by Elsevier Ltd.

[7] Shitangsu Kumar Paul and Jayant K. Routray, "An Analysis of the Causes of Non-Responses to Cyclone Warnings and the Use of Indigenous Knowledge for Cyclone Forecasting in Bangladesh", Climate Change and Disaster Risk Management, Climate Change Management, DOI: 10.1007/978-3-642-31110-9_2, Springer-Verlag Berlin Heidelberg 2013

[8] Tsun-Hua Yang, Sheng-Chi Yang, Jui-Yi Ho, Gwo-Fong Lin, Gong-Do Hwang, Cheng-Shang Lee, "Flash flood warnings using the ensemble precipitation forecasting technique: A case study on forecasting floods in Taiwan caused by typhoons", Journal of Hydrology, 520 (2015) 367–378

[9] Roger S. Pulwarty, Mannava V.K. Sivakumar, "Information systems in a changing climate: Early warnings and drought risk management", Weather and Climate Extremes, 3 (2014) 14–21

[10] Lorenzo Alfieri, Peter Salamon, Florian Pappenberger, Fredrik Wetterhall, JuttaThielen, "Operational early warning systems for water-related hazards in Europe", environmental science & policy 21 (2012) 35 – 49

[11] Muhammad Atiq Ur Rehman Tariq, Nick van de Giesen, "Floods and flood management in Pakistan", Physics and Chemistry of the Earth, 47–48 (2012) 11–20

[12] Hans J.P. Marvin, Gijs A. Kleter, H.J. (Ine) Van der Fels-Klerx, Maryvon Y. Noordam, Eelco Franz, Don J.M. Willems, Alistair Boxall, "Proactive systems for early warning of potential impacts of natural disasters on food safety: Climate-change-induced extreme events as case in point", Food Control 34 (2013) 444e456

[13] Florian Pappenberger, Hannah L. Cloke, Dennis J. Parker, Fredrik Wetterhall, David S. Richardson, JuttaThielen, "The monetary benefit of early flood warnings in Europe", environmental science & policy 51 (2015) 278 – 291

[14] Emma E.H. Doyle, John McClure, David M. Johnston, Douglas Paton, "Communicating likelihoods and probabilities in forecasts of volcanic eruptions", Journal of Volcanology and Geothermal Research 272 (2014) 1–15

[15] Jean Dominique Creutin, Marco Borga, Eve Gruntfest, Céline Lutoff, DavideZoccatelli, Isabelle Ruin, "A space and time framework for analyzing human anticipation of flash floods", Journal of Hydrology 482 (2013) 14–24

[16] JACK HERRMANN, "Disaster Response Planning & Preparedness: PHASES OF DISASTER", Copyright 2007 – New York Disaster Interfaith Services (NYDIS)

[17] KainazBomiSheriwala, "Data Mining Techniques in Stock Market", INDIAN JOURNAL OF APPLIED RESEARCH, Volume 4, August 2014.

[18] S. L. Pandhripande and Aasheesh Dixit, "Prediction of 2 Scrip Listed in NSE using Artificial Neural Network", International Journal of Computer Applications (IJCA), Volume 134, No.2, January 2016.

[19] Data Mining - Classification & Prediction, available online: http://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm, [accessed 27 April 2016].

[20] Samuel OdeiDanso, "An Exploration of Classification Prediction Techniques in Data Mining: The insurance domain", Master's Degree in Advanced Software Engineering, School of Design, Engineering, and Computing. Bournemouth University, September, 2006.

[21] Dr. B. Srinivasan and K. Pavya, "A STUDY ON DATA MINING PREDICTION TECHNIQUES IN HEALTHCARE SECTOR", International Research Journal of Engineering and Technology (IRJET), PP. 552-556, Volume 3, Mar-2016

[22] Vipin Kumar, JoydeepGhosh and • David J. Hand, "Top 10 algorithms in data mining", Knowledge and Information System, PP. 1–37, (2008).

[23] Vapnik V (1995), the nature of statistical learning theory. Springer, New York.

[24] Kavitha G, Udhayakumar A and Nagarajan D, "Stock Market Trend Analysis Using Hidden Markov Models", available online: https://arxiv.org/ftp/arxiv/papers/1311/1311.4771.pdf.

[25] Haiqin Yang, Laiwan Chan, and Irwin King, "Support Vector Machine Regression for Volatile Stock Market Prediction", Intelligent Data Engineering and Automated Learning IDEAL, PP. 391- 396, Springer-Verlag Berlin Heidelberg 2002

[26] RoshaniChoudhary, JagdishRaikwal, "An Ensemble Approach to Enhance Performance of Webpage Classification", International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5614-5619

[27] Juntao Wang, Xiaolong Su, "An improved K-Means clustering algorithm", 978-1-61284-486-2/111$26.00 ©2011 IEEE

[28] ShwetaJaiswal, Atish Mishra, Praveen Bhanodia," Grid Host Load Prediction Using GridSim Simulation and Hidden Markov Model", International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 7, July 2014).