

An Optimization of Feature Selection for Classification using Modified Bat Algorithm

V. Ysaswini

Research Scholar, Computer Science and Engineering Department, Pondicherry Engineering College, Puducherry, India
E-mail: yasaswini2907@pec.edu

Santhi Baskaran

Professor, Information Technology Department, Pondicherry Engineering College, Puducherry, India
E-mail: santhibaskaran@pec.edu

Received: 05 February 2021; Revised: 03 April 2021; Accepted: 22 April 2021; Published: 08 August 2021

Abstract: Data mining is the action of searching the large existing database in order to get new and best information. It plays a major and vital role now-a-days in all sorts of fields like Medical, Engineering, Banking, Education and Fraud detection. In this paper Feature selection, which is a part of Data mining is performed to do classification. The role of feature selection is in the context of deep learning and how it is related to feature engineering. Feature selection is a preprocessing technique which selects the appropriate features from the data set to get the accurate result and outcome for the classification. Nature-inspired Optimization algorithms like Ant colony, Firefly, Cuckoo Search and Harmony Search showed better performance by giving the best accuracy rate with a smaller number of features selected and also fine f-Measure value is noted. These algorithms are used to perform classification that accurately predicts the target class for each case in the data set. We propose a technique to get the optimized feature selection to perform classification using Meta Heuristic algorithms. We applied new and recent advanced optimized algorithm named Modified Bat algorithm on University of California Irvine datasets that showed comparatively equal results with best performed existing firefly but with a smaller number of features selected. The work is implemented using JAVA and the Medical dataset has been used. These datasets were chosen due to nominal class features. The number of attributes, instances and classes varies from chosen dataset to represent different combinations. Classification is done using J48 classifier in WEKA tool. We demonstrate the comparative results of the presently used algorithms with the existing algorithms thoroughly. The significance of this research is it will show a great impact in selecting the best features out of all the existing features which gives best accuracy rates which helps in extracting the information from raw data in Data Mining Domain. The Value of this research is it will manage main fields like medical and banking which gives exact and proper results in their respective field. The best quality of the research is to optimize the selection of features to achieve maximum predictive accuracy of the data sets which solves both single variable and multi-variable functions through the generation of binary structuring of features in the dataset and to increase the performance of classification by using nature inspired and Meta Heuristic algorithms

Index Terms: Optimization, Meta-heuristic, Feature Extraction, Deep learning, Firefly, Cuckoo, Harmony, Bat Algorithm.

1. Introduction

Data Mining [1] is the way of searching important information from the huge present all over in the repository. Data Mining falls in to two ways namely Association and Classification analyzing methods.

Data mining process involves the following stages.

a) Problem Definition. In this stage the analysis of the problem in the business problem is done and tries to get the clear idea of the problem to be solved. This takes some time to make an exact definition of the problem and it does not require any data tools.

b) Exploration of Data. In this stage data is explored by identifying quality problem to understand the metadata meaning. It is next level of problem definition stage which frequently exchanges the data. **c) Preparation of Data.** In this stage data model is built after the exploration of data. Data is collects, clear the unwanted data and arrange the data in a format like tables and records.

d) Data Modeling. At this stage after preparation of information, different mining functions are applied to the same kind of data. A high quality of mining model is prepared based on the changes in the parameters until we get optimal data model. Finally, the good quality model is built and evaluated.

e) Evaluation of the Model. In this stage the evaluated model is checked and tested whether the quality is good or not and objective is satisfied or not?

f) Deployment. In this stage after the evaluation of data, the exporting of the data is done and the results are checked into database tables.

Optimization algorithm provides a systematic way of developing and leveling new solutions to gain an optimal result. The optimization process must only be used in those problems where there is a specific need of accomplishing a quality or a competitive work. It is expected that the solution obtained through an optimization method is better than other results in terms of the selected objective.

The Research objective is to obtain a framework to optimize the selection of features computing the best accuracy of data sets and to increase the performance of classification using Meta-Heuristic algorithms:

- Bat Algorithm(Movement of Virtual Bats, Loudness and Pulse Emission)
- Modified Bat Algorithm(Modification in Pulse emission and Loudness of Bats)

The Fundamental problems solved are:

1. Accuracy: Guessing value of predicted attributes. It is the degree to which the result of a calculation or specification conforms to the correct value.
2. Speed: Time to construct the model (training time) time to use the model (classification/prediction time).
3. Robustness: -Ability to handle noise and missing values.
4. Feature Selection: No. of. Features selected.
5. Scalability: Efficiency in disk-resident databases.

This paper shows the Bat algorithm and Modified Bat algorithm accuracy rates when compared to existing algorithms namely Firefly, Cuckoo search and Harmony Search algorithms that showed almost equal results of the best accuracy rates in existing work. In existing work all the above mentioned three algorithms have been used in feature selection problem and gave good accuracy rate when compared to ABC algorithm. Out of these three algorithms, Firefly algorithm-Comparatively Brightest Firefly algorithm gave best results with less time and less number of features selected. The search space has been increased now in proposed system from local space to Global space and in Exhaustive searches method which is combination of all the possible selections in the set and also the selected algorithms turned from exploitation phase to exploration phase of selecting the features in their own way.

There are various applications with respect to data mining and optimization techniques in different fields. This method proves the better analysis which gives the best results and improved accuracy.

The following are the different field of applications.

1. Network Security
2. Computer Vision and Processing
3. Nature Inspired fields.
4. Medical Fields
5. Transition Probabilities for Radio Systems
6. Intrusion Detection
7. Education
8. Financial Banking

2. Literature Review

Huan Liu et.al, [17] Proposed Feature selection is a process that selects a subset of original features. The optimality of a feature subset is measured by an evaluation criterion. As the dimensionality of a domain expands, the number of features N increases. Finding an optimal feature subset is usually intractable and many problems related to feature selection have been shown to be NP-hard. A typical feature selection process consists of four basic steps, namely, subset generation, subset evaluation, stopping criterion, and result validation. Subset generation is a search procedure that produces candidate feature subsets for evaluation based on a certain search strategy. Each candidate subset is evaluated and compared with the previous best one according to a certain evaluation criterion. If the new subset turns out to be better, it replaces the previous best subset. The process of subset generation and evaluation is repeated until a given stopping criterion is satisfied. Then, the selected best subset usually needs to be validated by prior knowledge or different tests via synthetic and/or real world data sets. Feature selection can be found in many areas of data mining such as classification, clustering, association rules, and regression. For example, feature selection is called subset or variable selection in Statistics. A number of approaches to variable selection and coefficient shrinkage for regression are summarized. In this survey, we focus on feature selection algorithms for classification and clustering. Early research

efforts mainly focus on feature selection for classification with labeled data (supervised feature selection) where class information is available. Latest developments, however, show that the above general procedure can be well adopted to feature selection for clustering with unlabeled data (or unsupervised feature selection) where data is unlabeled.

Fourteen standard datasets drawn from the UCI collection proposed by B.Kalpana et.al [13] were used in the experiments. These datasets were chosen due to nominal class features. The number of instances, attributes and number of classes vary in the chosen dataset to represent different combinations. The learning algorithm chosen for classifying are Naïve Bayes, K-NN (k=10) and C4.5 tree. All datasets were run on Pentium machine with Java 6 and 3 GB RAM. The result shows that C4.5 tree performs well in most cases. Naïve Bayes is a true predictor. As the performance by Naïve Bayes and C4.5 tree is better when compared to KNN. KNN algorithm with k=10 has an overall average good performance on all types of datasets.

Dr. T.Sai Durga et.al [14] proposed that the classifier used for evaluating the feature subsets generated are J48, NaïveBayes and Logistic. Feature subset generation by Firefly Algorithm, Cuckoo Search Algorithm and Harmony Search has been implemented using Net Beans IDE. These algorithms gave different results by selecting different features out of all the features present in the dataset. The different algorithms selected are Artificial Bee Colony optimization algorithm (ABC), Firefly algorithm (FA), Cuckoo search algorithm (CSA), Harmony search algorithm (HSA). After processing of all the features in 10 fold cross validation and through Weka tool, the final result is that, Firefly algorithm in Feature selection gave the best results by giving best accuracy with correct selection of attributes.

Thair Nu Phyu [15] proposed different Classification Techniques in data mining. Decision trees and Bayesian Network (BN) generally have different operational profiles, when one is very accurate the other is not and vice versa. On the contrary, decision trees and rule classifiers have a similar operational profile. The goal of classification result integration algorithms is to generate more certain, precise and accurate system results. Numerous methods have been suggested for the creation of ensemble of classifiers. Although or perhaps because many methods of ensemble creation have been proposed, there is as yet no clear picture of which method is best. Classification methods are typically strong in modeling interactions. Several of the classification methods produce a set of interacting loci that best predict the phenotype. However, a straightforward application of classification methods to large numbers of markers has a potential risk, picking up randomly associated markers.

Samina Khalid et.al [16], proposed comparison of Feature Selection Algorithms and extraction. The objective of both selection and extraction methods concerns the reduction of feature space in order to improve data analysis. This aspect becomes more important when real world datasets are considered, which can contain hundreds or thousands features. The main difference between feature selection and extraction is that the first performs the reduction by selecting a subset of features without transforming them, while feature extraction reduces dimensionality by computing a transformation of the original features to create other features that should be more significant. Traditional methods and their recent enhancements as well as some interesting applications concerning feature selection are presented. Feature selection improves knowledge of the process under consideration, as it points out the features that mostly affect the considered phenomenon. Moreover, the computation time of the adopted learning machine and its accuracy need to be considered as they are crucial in machine and data mining applications.

3. Description of Algorithms

3.1. Existing Work Algorithms-Firefly Algorithm:

Firefly Algorithm (FA)[9] being a Nature Inspired algorithm works based on flashing nature of the fireflies. The main reason for its flash is to move as a system which provides signal to absorb the other fireflies towards itself.

The algorithm was implemented to perform Feature Selection (FS) for Image Processing and Eigen Value Optimization problem along with other related domains and has been performed so that better results are obtained. In order to achieve best optimal feature subset increases the predictive accuracy of the classifier. In this algorithm, along with considering the brighter firefly to obtain the predictive accuracy of the dataset we have also considered a comparatively brighter firefly and the predictive accuracy of that method have also been calculated.

- Choosing a Brightest Firefly
- Choosing a Comparatively Brighter Firefly

It includes previously chosen firefly solution and the newly selected brighter firefly solution in our computation. The latest solution is found by solving the below formulation

$$New X_i = X_i + \beta_0 \exp(-\gamma(r_{ij})^2)(X_i - X_j) + \alpha \varepsilon_i \quad (1)$$

Where,

Table 1. Parameter Settings of Firefly Algorithm

X_i	The solution pointed by the current firefly (Classification Accuracy)
X_j	The solution pointed by the brightest firefly (First Method) The solution pointed by the comparatively brighter firefly (Second Method)
β_0	Between 0 and 1
γ	Between 0.1 and 10
r_{ij}	Distance is fixed to 1
α	Parameters selected within the range [0,1] randomly

3.2. Existing Work Algorithm-Cuckoo Search Algorithm

Cuckoo Search Algorithm (CSA) [10] being idealized by its breeding behavior was tested on engineering optimization and embedded design problems. The results being obtained by this algorithm for engineering optimization problems are quite convincing in its results. Hence forth, Cuckoo Search Algorithm for FS is implemented. In this Cuckoo Search Algorithm addition of three constraints such as Eviction, Abandon and Survival is done. This type of method uses its historical memories for the location and status of the eggs being laid by the cuckoos.

This Algorithm Mainly Concentrates on Replacing Not Good Nests with The Potentially Good Nests. The position of the egg replaces the position of the random new eggs in another nest in three cases. if the eggs have been evicted, if the cuckoos abandon the nest or if the eggs have been hatched resulting to its survival.

- Hatching of eggs – Survival of the Cuckoo
- Abandoning the nest – Host bird abandons its nest and migrates to some other place to build another nest.
- Evicting the eggs – The host bird throws the cuckoo bird's eggs.

Equation has been used as the main formula for the computation of Cuckoo Search Algorithm.

$$F_{ij} = \frac{\{(\alpha[I_i(next) - I_i(bp)]) * iter\}}{maxcuckoo} \quad (2)$$

Where,

Table 2. Parameter settings for Cuckoo Search Algorithm

F_{ij}	Fitness function used to find the alpha value of the cuckoo
$I_i(next)$	Accuracy of the cuckoo bird being selected
$I_i(bp)$	Accuracy of the host cuckoo bird
$Iter$	Fixed to 1
α	Between 0 and 4
Maxcuckoo	Maximum number of cuckoos in a particular dataset

3.3. Existing Work Algorithm-Harmony Search Algorithm:

The underlying principle behind this HSA algorithm [11] is has been that this algorithm might face the search on the grounds of Pitch, Amplitude and Timbre producing a perfect harmony. The initial random solutions that are considered may be far away from the feasible solutions. To get closer to the feasible and promising solutions, the exact solutions can be obtained by choosing only the amplitude of the tone.

In this method, the best accuracy of Tunes is found with the frequency value of Tunes which varies with time (t) and computation is done for the new solution.

In HSA, Individual music player (variable which takes decision) runs a bit of music (value) which finds the best harmony (global optimum) at the last. Based on the Lambda value two constraints will be considered. Noise (high pitch) and Melody (low pitch).

Equation has been used as the main formula for the computation of Amplitude in Harmony Search Algorithm

$$\lambda = \frac{c}{h_i[next] * h_i[music]} \quad (3)$$

Where,

Table 3. Parameter Settings for Harmony Search Algorithm

λ	Lambda Value (Amplitude)
C	Velocity of Light
$h_i[next]$	Accuracy of newly selected tune
$h_i[music]$	Accuracy of the existing tune that needs to be compared with the new tune

4. Proposed Work Algorithm- Modified Bat Algorithm Applied in Feature Selection.

In Proposed work, Modified Bat algorithm is used to check the results compared to the existing three algorithms namely Firefly algorithm, Cuckoo Search algorithm, and Harmony Search algorithm. The advanced features of proposed work is that the search capacity of Modified Bat algorithm is far better than existing algorithm as it checks both Exploration and Exploitation search and selection is based on random features with high pulse and frequency. So, the result varies from existing to proposed work in a better way.

Objective Function: Count of Bats $X_i = X_{i1}$ to X_{iD} to the power of T. where I belongs to the range $[1, Np]$ [13]

1. Initialize the count of Bats in the Search space X_i and V_i
2. Initialize the frequencies (f_i), Pulses (r_i) and the Loudness (A_i)
3. while ($t < \text{Max number of iterations}$)
Generate new solutions by adjusting frequency,
Update velocities and locations/solutions
4. if ($\text{rand} > r_i$)
Select a solution among the best solutions
Generate a local solution around the selected best solution
end if
Generate a new solution by flying randomly
5. if ($\text{rand} < A_i \ \& \ f(x_i) < f(x_*)$)
Accept the new solutions
Increase r_i and reduce A_i
end if
6. Rank the bats and find the current best x_*
end while

In this algorithm [14] a little change in the computation of emission of pulse rate and loudness of the bats is improved which showed a superior results when evaluated with the Bat algorithm.

Following pseudo code of the Modified Bat algorithm is explained.

1. Firstly, initialize the objective function of the algorithm and assign the best solution x_{best} in the count of bats.
2. In the second new solutions are generated according to the movement of the virtual bats in the search space.
3. In the third step best solution is determined by using random walks in local search.
4. In forth step Evaluation of the new solution is moved out.
5. In fifth step, bank the current best solution.
6. In the last step, compute the best solution and update it.

The echolocation behavior of micro bats is explained in this algorithm. The primary use of Bat Algorithm (BA) [12] is variant behavior of the micro bats based up on the frequency tuning, velocity v_i^t and its location x_i^t and calculates based on the iteration t with d-dimensional search or solution space. According to Yang, the following mathematical equation is updated and written as

$$f_i = f_{\min} + (f_{\max} - f_{\min}) \beta \quad (4)$$

$$v_i^t = v_i^{t-1} + (x_i^t - x_*) f_i \quad (5)$$

$$x_i^t = x_i^{t-1} + v_i^t \quad (6)$$

where $\beta \in [0,1]$ which uniformly distributes the random vector.is a random vector drawn from a uniform distribution. A direct exploitation for searching local solutions which modifies the present good results.

$$x_{next} = x_{prev} + \varepsilon A^t \quad (7)$$

From the above equation, ε refers to range of $[-1, 1]$ which can be any number between -1 and 1. A^t refers to the calculation of overall best loudness mean. During the iterations the emission of pulse from bats and its loudness is variant which is calculated based on the following equation,

$$A_i^{t+1} = \alpha A_i^t \quad (8)$$

$$\gamma_i^{t+1} = \gamma_i^0 [1 - \exp(-\gamma t)] \quad (9)$$

Where $0 < \alpha < 1$ and $\gamma > 0$ are constants.

The main modifications in this algorithm is changing the ranges of the parameters defining $A_0 = 1$ and $A_{min} = 0$, $f_{min} = 0$ and $f_{max} = 2$, $\alpha = \gamma = 0.9$ to 0.975

5. Experimental Setup and Result Analysis

Fourteen standard datasets drawn from the UCI collection were used in the experiments. These datasets were chosen due to nominal class features. The number of attributes, instances and number of classes vary in the chosen dataset to represent different combinations. All the features in 10 fold cross validation is run through Weka tool to get classification accuracy. The classifier used for evaluating the feature subsets generated is J48, Naïve Bayes and Logistic. Feature subset (FS) generation by Firefly Algorithm (FA), Cuckoo Search Algorithm (CSA) and Harmony Search Algorithm (HSA) is implemented using Net Beans IDE in the existing work. Feature subset generation by Nature-inspired algorithms named Modified Bat algorithm (MBA) has been implemented using Net Beans IDE while the UCI dataset of medical field is run through WEKA tool to get the classification rate which is processed through J48 classifier in the proposed work.

Table.4 describes the different types of diseases in medical field and instances, Features and Class of each disease. For example, in this dataset Iris disease can be diagnosed by 4 features and it falls in to 3 classes.

Table 4. Dataset Description

Datasets	No. of Instances	No. of Features	No. of Class
Heart-c	303	13	2
Dermatology	366	34	6
Hepatitis	155	19	2
Lung Cancer	32	56	2
Pima Indian Diabetes	768	8	2
Iris	150	4	3
Wisconsin	699	10	2
Lymphography	148	18	4
Diabetes	768	8	2
Sonar	208	60	2
Heart-Stalog	270	13	2
Audiology	226	69	10

Table 5. Comparison of all the accuracy of Existing with Proposed Algorithms

Datasets	FA-FS FA1 (%)	FA-FS FA2 (%)	CSA – FS (%)	HSA – FS (%)	MBA-FS (%)
Heart-C	83.15	83.07	78.217	79.53	83.191
Hepatitis	69.03	67.00	64.516	67.74	69.072
Lung Cancer	89.50	89.33	84.375	87.37	89.443
Pima	78.70	76.10	75.651	77.47	79.981
Iris	96.00	96.00	96.00	96.00	96.00
Lymphography	84.10	82.25	78.378	84.43	82.295
Diabetes	77.47	76.00	75.65	77.21	78.583
Dermatology	96.72	96.17	95.901	96.07	96.210
Heart-Stalog	84.44	83.39	80.74	84.81	84.345
Audiology	79.051	79.20	79.646	77.87	79.942

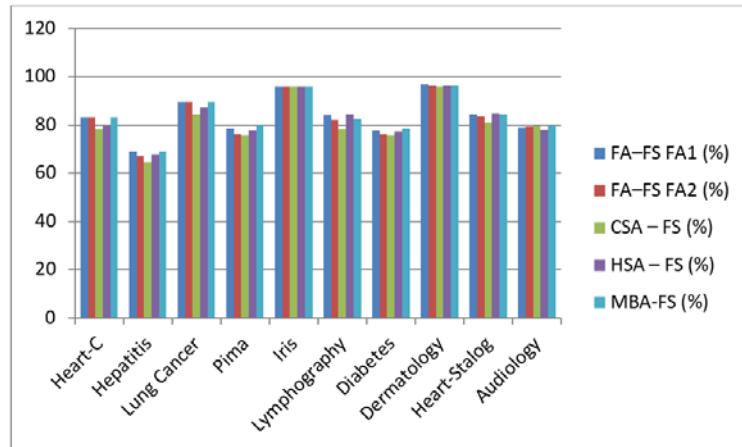


Fig.1. Graphical Representation of Accuracy of all these algorithms

From Fig.1, we can observe that the Modified Bat algorithm which is obtained for 10 UCI datasets showed better accuracy when compared to other algorithms.

Table 6. Comparison of all the features selected in Existing with Proposed Algorithms

Datasets	FA-FS	CSA - FS	HSA - FS	BA-FS	MBA-FS
Heart-C	6	7	5	6	5
Hepatitis	8	10	9	8	9
Lung Cancer	14	20	18	16	18
Pima	3	3	4	4	3
Iris	2	3	2	2	2
Lymphography	8	7	10	8	7
Diabetes	4	5	6	5	4
Dermatology	23	22	25	21	23
Heart-Stalog	7	6	7	8	7
Audiology	53	55	49	56	51

In this Table 6, it is clear that the number of features selected for Modified Bat algorithm is best when compared to other existing and the proposed Bat algorithm.

6. Inference

From the table.4, We can infer that Modified Bat Algorithm applied for UCI dataset for FS with respect to Accuracy gives better as well equal accuracies

By comparing Firefly Algorithm, Cuckoo Search Algorithm and Harmony Search Algorithm for FS with respect to Features: we can infer that the features for Modified Algorithm gets reduced as well gets increased in some of the datasets in other Algorithms.

7. Conclusion

With the emerging trends of Information Technology, Data Mining throws out its purpose to each individual spotting out only the relevant features by the Feature Selection process along with Classification from eons of database to reach the final outcome. The proposed system for FS Optimization to perform Classification is applied and the results have been obtained using UCI datasets. The datasets are taken from UCI repository and Table 1 describes the 10 datasets that we have used. The UCI datasets are used in continuing works in literature; such as Classification, FS and Classifier Ensemble, Stacking Ensemble, so we have adopted these datasets for our work. This paper has attempted to give accuracy results of the several popular Meta-Heuristic algorithms. Finally analyzing the proposed algorithms compared to existing algorithms we found that Modified Bat algorithms showed better results in some particular data in the dataset with a minimum of 0.3% ahead compared to the Existing algorithm and for few diseases in the Dataset there is little more difference of about 1% ahead to the existing algorithm. This work can be extended with some other Nature inspired algorithms like Lion, Squirrel etc to check whether these algorithms shows better results compared to present used algorithms and the search space can also be increased with change in the range of the exploration phase. There is a possibility of using different feature selection technique to compare with presently used Wrapper based Technique.

References

- [1] Tan, Steinbach, Kumar. (2005). "Introduction to Data Mining".
- [2] Hassan AbouEisha et.al, (2018) "Extensions of Dynamic Programming for Combinatorial Optimization and Data Mining"
- [3] Sunil Kawale, "Datamining and Optimization Techniques" International Journal of Statistika and Matematika", Volume 6, Issue 2, 2013 pp 70-72.
- [4] Nidhi Tomar and Prof. Amit Kumar Manjhvar, "A Survey on Data mining optimization Techniques", International Journal of Science Technology & Engineering, Volume 2, Issue 06, December 2015.
- [5] Basturk B, Karaboga D (2006) "An artificial bee colony (ABC) algorithm for numeric function optimization". IEEE Swarm Intelligence Symposium, 12–14 May, Indianapolis.
- [6] Bergh F, Engelbrecht AP (2006) "A study of particle swarm optimization particle trajectories". Inf Sci 176. 937–971.
- [7] Rao, R. Venkata. "Teaching Learning Based Optimization Algorithm. And Its Engineering Applications". Springer, 2015.
- [8] Rao, R. Venkata, and V. D. Kalyankar. "Parameter optimization of modern machining processes using teaching–learning-based optimization algorithm". Engineering Applications of Artificial Intelligence 26, no. 1 (2013). 524-531.
- [9] Shunmugapriya .P and Kanmani S, P.Sindhuja, G.Koperundevi, V.Yasaswini, "Firefly Algorithm Approach for the Optimization of Feature Selection to Perform Classification", International Conference on Advances in Engineering & Technology, IEEE-ICAET 2014.
- [10] Xin-She Yang, Suash Deb, "Cuckoo Search Via Levy Flights", World Congress On Nature and Biologically Inspired Computing (NaBIC 2009) .
- [11] Xin-She Yang and X. He. "Bat algorithm: Literature review and applications". International Journal of Bio-Inspired Computation, 5(3):141-149, 2013.
- [12] Richardson, P.: The secret life of bats. <http://www.nhm.ac>.
- [13] B.Kalpna, Dr.V.Saravanan and Dr.K. Vivekanandan. "A survey of feature Selection models on Classification", Vol 3, No.1, Jan-Feb. 2012.
- [14] T.Sai Durga, V.Yasaswini. "An Enhancement for the optimization of feature selection to perform Classification Using Meta Heuristic Algorithms", In International Journal of Latest Engineering Research and Applications (IJLERA), Volume – 01, Issue – 09, December – 2016, PP – 64-70.
- [15] Thair Nu Phyu. "Survey of Classification Techniques in Data Mining", International Multi Conference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong.
- [16] Samina Khalid, Tehmina Khalil, Shamila Nasreen. "A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning", Science and Information Conference 2014, August 27-29, 2014.
- [17] Huan liu and lei yu. "Toward integrating Feature Selection algorithms for Classification and Clustering". IEEE transactions on Knowledge and Data Engineering, vol 17. No 4, April 2005.
- [18] Lakshmi, T. M., Martin, A., Begum, R. M., & Venkatesan, V. P. (2013). An analysis on performance of decision tree algorithms using student's qualitative data. International Journal of Modern Education and Computer Science, 5(5), 18.
- [19] Ogunde A. O and Ajibade D. A., " A Data Mining System for Predicting University Students' Graduation Grades Using ID3 Decision Tree Algorithm ". Journal of Computer Science and Information Technology
- [20] Shivam Goyal, Jaskirat Singh, "Two-Level Alloyed Branch Predictor based on Genetic Algorithm for Deep Pipelining Processors", International Journal of Modern Education and Computer Science, Vol.9, No.5, pp.27-33, 2017.
- [21] Zahid Ullah, Muhammad Fayaz, Asif Iqbal, "Critical Analysis of Data Mining Techniques on Medical Data", International Journal of Modern Education and Computer Science, Vol.8, No.2, pp.42-48, 2016.
- [22] N. Shamli, B. Sathiyabhama, "Parkinson's Brain Disease Prediction Using Big Data Analytics", International Journal of Information Technology and Computer Science, Vol.8, No.6, pp.73-84, 2016.

Authors' Profiles



Author: **V Yasaswini**, Research Scholar in the department of Computer Science and Engineering, Pondicherry Engineering College.



Co-Author: **Santhi Baskaran**, Professor in the Department of Information Technology, Pondicherry Engineering College and specialized in Real Time Systems, Cloud Computing domains.

How to cite this paper: V. Yasaswini, Santhi Baskaran, "An Optimization of Feature Selection for Classification using Modified Bat Algorithm", International Journal of Information Technology and Computer Science(IJITCS), Vol.13, No.4, pp.38-46, 2021. DOI: 10.5815/ijitcs.2021.04.04