International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 2, pp 80-85, February-2023
https://doi.org/10.47001/IRJIET/2023.702011

Feature Selection in Genetic Algorithm Based Sentiment Classification Used Five Layered Artificial Neural Network for Cross Domain Opinion Mining

[1]Dr. S.Gnanapriya, [2]D.Umanandhini

[1]Hod & Assistant Professor, Department of Computer Applications, Kovai Kalaimagal college of Arts and science, Narasipuram, Coimbatore-641109, Tamilnadu, India
[2]Hod & Assistant Professor, Department of Computer Science, Kovai Kalaimagal college of Arts and science, Narasipuram, Coimbatore-641109, Tamilnadu, India

***Abstract* - In this paper, we present a review of Natural Language Processing (NLP) techniques for opinion mining. First, we introduce genetic algorithm for feature selection ANN used for classifier many researchers are proposing new ideas, models, applying machine learning algorithms, and more as a result of web mining and web usage mining. Internet use has expanded to practically all types of applications, including e-commerce. E-commerce enables consumers/customers to purchase things online, while web analytics enables website administrators to determine which products sell the most. In many e-commerce decision-making jobs, opinion mining is the key to analytics. A dataset of product reviews, such as books, DVDs, electronics, and kitchen appliances, is obtained. Genetic Algorithm is used to identify the features to perform opinion mining. The measures for measuring the performance of the proposed work are accuracy and F-measure. There is a comparison between domain-specific and domain-dependent words. The suggested work outperforms the existing work in terms of the chosen performance indicators, according to the findings.**

Keywords: Feature Selection, Opinion Mining, Genetic Algorithm.

I. INTRODUCTION

1.1 Opinion Mining

Opinioned text provided a new research area for the study of texts. Traditionally, the facts and information-centred view of the text has been extended to allow applications that are conscious of the sentiment. Increased use of the Internet and electronic activities such as booking tickets, online purchases, e-commerce, social media interactions, blogging, etc. have led to the need to collect, turn and evaluate enormous quantities of information. New methods to interpret and summarize the details therefore need to be implemented. Organizations take the product review provided by consumers seriously, as it adversely affects the company's revenue. Organizations are now making an effort to respond to the feedback and track the efficacy of their advertisement campaigns. Sentiment analysis, a popular method, is used in this regard to isolate and analyze feelings.

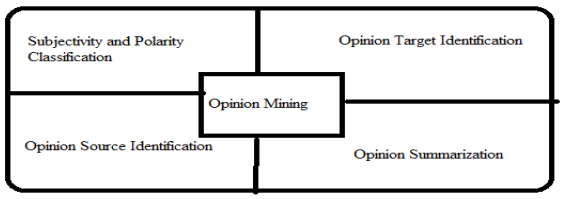


Figure 1: Tasks of Opinion Mining

1.2 Natural Language Processing

Natural Language Processing as well as Opinion Mining are done as sub-domains of web content mining, with product input from customers 'polarities. Opinion Mining's model of the object. Machines should be both accurate and efficient in their ability to perceive and understand human emotions and feelings. Analyzes of emotions as well as mining views are ways of carrying it out. Problems with sensitivity analyzes can be adequately solved by manual instruction. Completely automated systems for evaluating feelings that do not require human interaction have not yet been developed, and this is primarily due to the many challenges in the domain itself.

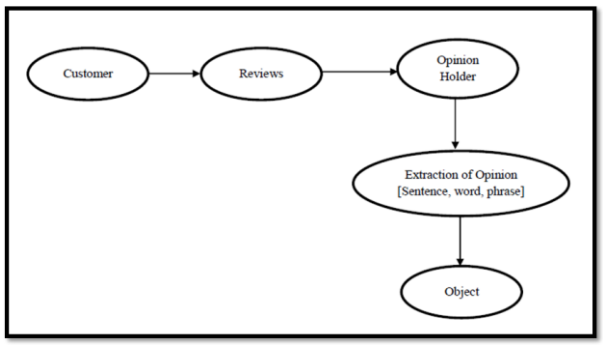


Figure 2: Model of Opinion Mining

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 2, pp 80-85, February-2023

https://doi.org/10.47001/IRJIET/2023.702011

The first difficulty is a term that in some circumstances is considered positive and in certain cases is perceived as negative. The second is that not all people are sharing their views in the same way. Nearly all modern text processing relies on the assumption that minute differences between two data sets do not over-alter the context. Although "the image is good" is in no way similar to "the image is not good" in sentiment analyzes. Within their sentences human beings can be inconsistent. Usually, comments will have both positive and negative remarks which are typically addressed one sentence at a time by evaluating the text.

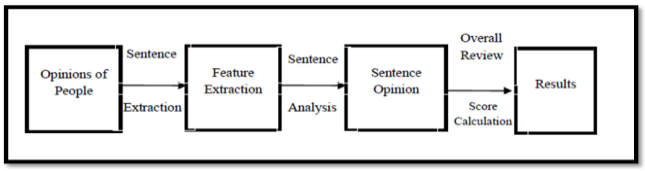


Figure 3: Opinion mining and sentiment analysis process

II. GENETIC ALGORITHM

Some of the fundamental terms that can aid in our understanding of genetic algorithms include the following:

- **Population:** This is a subset of all potential answers to the problem at hand.
- **Chromosomes:** One of the components of the population is a chromosome.
- **Gene:** This is a chromosomal component.
- **Allele:** The value assigned to a gene on a particular chromosome.
- **Fitness function:** This is a function that improves an output by using a certain input. The input is the solution, and the output is the appropriateness of the solution.
- **Genetic operators:** In genetic algorithms, the most advantageous people mate to generate children who are superior to their parents.

Genetic operators can change a person's genetic makeup. A heuristic search algorithm called a genetic algorithm (GA) is used to tackle search and optimization issues. A subset of evolutionary algorithms, which are utilised in computation, includes this algorithm.

The idea of genetics and natural selection is applied by genetic algorithms to offer answers to issues.

These algorithms are more intelligent than random search algorithms because they direct the search to the best performing area of the solution space using historical data.

GAs is also based on chromosomal behavior and genetic structure. Every chromosome plays a part in offering a potential answer. The fitness function assists in supplying the traits of each person inside the population. The solution is better the larger the function.

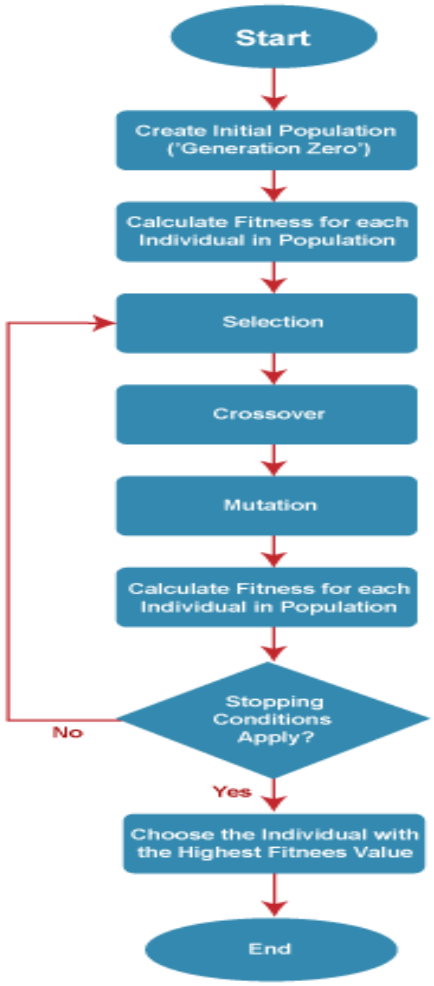


Figure 4: General Workflow of a Simple Genetic Algorithm

2.1 Five Layered Artificial Neural Network (FLANN) Classifier

The classification undertaking is performed by FLANN. FL-ANN is a Five Layered Outspread Premise Function (RBF) based classifier neural network that makes use of influenced use of descent approach and regression based classification. It enhances smoothing parameter of RBF part through descent approach. It enhances smoothing parameter of RBF part through slant plummet approach. It comprises of five layers named as info, pattern (or design), summation, institutionalization and yield and is depicted the Fig.5.

Connected info vector is transmitted to design layer through information layer. Design layer incorporates each preparation datum with RBF part. Squared Euclidean separation between input vector and preparing information vector is ascertained as in (1) where p means aggregate of preparing information at design layer.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 2, pp 80-85, February-2023
https://doi.org/10.47001/IRJIET/2023.702011

$$dist(j)=\left\|x-t_j\right\|^2, 1\le j\le p \quad (1)$$

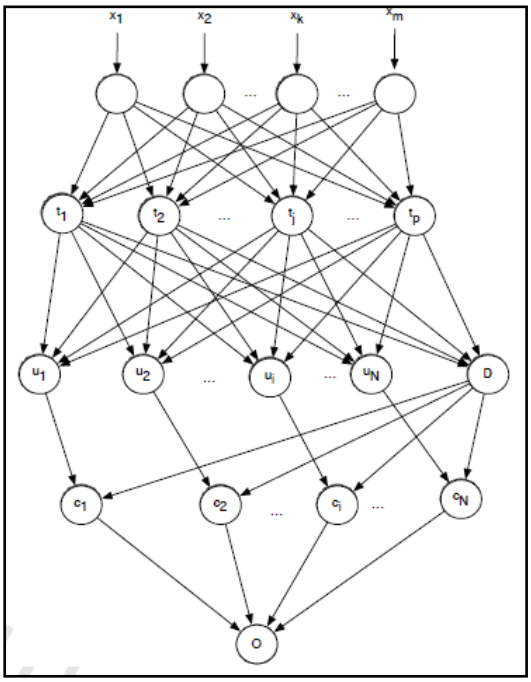


Figure 5: Five Layered Artificial Neural Network

Computed squared Euclidean separations are utilized as a piece of RBF divide fill in as in *r(j)* indicates yield of j^{th} prepared information and speaks to straightening parameter. Yields of RBF piece work are the yield estimations of example layer neurons. What's more, this layer incorporates ***N*** target estimations of each preparation datum controlled by contrasting class.

III. FEATURE SELECTION OBJECTIVES

There are several goals to feature selection.

1) It filters out irrelevant and noisy features, leaving just those with the least amount of redundancy and the greatest relevance to the target variable.
2) It shortens the time and effort required to train and evaluate a classifier, resulting in more cost-effective models.
3) It increases the effectiveness of learning algorithms, prevents over fitting, and aids in the creation of more general models.

The framework for feature selection performs the following steps:

Step 1: Generating a new feature;
Step 2: Determining whether adding the newly generated feature to the set of currently selected features;
Step 3: Determining whether removing features from the set of currently selected features;
Step 4: Repeat Step 1 to Step 3.

3.1 The Three Methods of Feature Selection

Filter, wrapper, and embedding techniques are the three types of feature selection methods, depending on how they interact with the classifier.

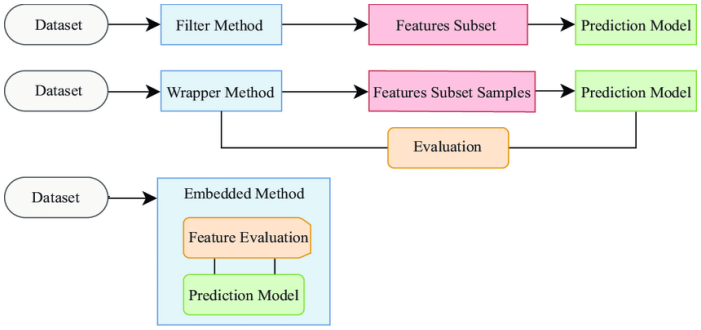


Figure 6: Feature Selection Methods for Machine learning

3.2 Feature selection methods

Some popular techniques of feature selection in machine learning are:

- Filter methods
- Wrapper methods
- Embedded methods

3.2.1 Filter methods

These methods are frequently applied in the pre-processing phase. These approaches choose traits from the dataset, whether or not a machine learning technique is applied.

They are great at eliminating redundant, correlated, and duplicate features and are quick and affordable to compute, but they do not get rid of multi-co linearity.

This can be beneficial when qualities are assessed independently of other features, but it falls short when a combination of features can enhance the performance of the model as a whole.

Set of all features → Selecting the best subset → Learning algorithm → Performance

Filter Methods Implementation

3.2.2 Wrapper methods

Wrapper methods, also referred to as greedy algorithms, train the algorithm iteratively using a subset of features. The model's characteristics are added and eliminated in accordance

International Research Journal of Innovations in Engineering and Technology (IRJIET)

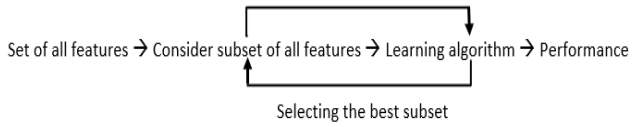
ISSN (online): 2581-3048

Volume 7, Issue 2, pp 80-85, February-2023

https://doi.org/10.47001/IRJIET/2023.702011

with the conclusions reached during earlier training. Typically, stopping criteria, such as when the model's performance declines or when a particular number of features are attained, are pre-defined by the person training the model.

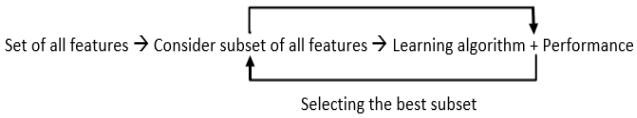
Wrapper strategies have an advantage over filter methods in that they offer the model a perfect set of features to train on, giving the model greater accuracy at a higher computational cost than filter methods.



Wrapper Methods Implementation

3.2.3 Embedded techniques

The feature selection technique is blended into the learning process in embedded methods, giving it its own built-in feature selection methods. Embedded methods overcome the disadvantages of filter and wrapper methods by combining their benefits. These approaches are similar to filter methods in that they are faster and more accurate than filter methods, and they take into account a mixture of features.



Embedded Methods Implementation

IV. IMPLEMENTATION: FEATURE SELECTION IN GENETIC ALGORITHMS

Feature selection using algorithm deals with isolating the very few relevant features from the large set. This is not exactly the classical feature selection problem known in Data mining. Among the various categories of feature selection algorithms, Genetic algorithm to solve real world optimization problems. There are many applications used for genetic algorithm in data mining. Robotics is used to solve real word problems among those. Advantages of genetics algorithm:

- Perform better than common feature selection techniques.
- Genetic algorithms can manage data sets with many features and it can be easily parallelized in computer clusters.

One of the issues using Genetic algorithm for feature selection is that the optimization process can be very attractive and there is a possibility for the GA to overfill to the predictors. One of the feature selections for most algorithms is the genetic algorithm. This is a stochastic method for function optimization based on the mechanics of natural genetics and biological evolution.

In The Genetic Algorithm is an optimization method inspired by the procedures of natural evolution. In feature selection, the function to optimize is the generalization performance of a predictive model. The combination of features selection using a genetic algorithm and corresponding sentiment words can help produce accurate, meaningful, and high-quality sentiment analysis results. We present here a genetic algorithm dedicated for a particular feature selection problem encountered in genetic analysis of Cross domain opinion mining.

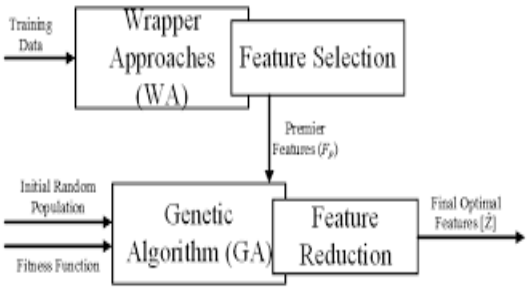


Figure 7: Workflow of feature selection in GA

```
START:
Create_Initial_Population
WHILE iteration_number < max_number_of_iterations
    FOR each chromosome, Evaluate_Fitness:
                Create_Neural_Network
                Network_Train
                Network_Validate
                Network_Test
                Fitness = Classification_Accuracy
    END_FOR
    Parent_Select
    Crossover
    Mutation
    Elitism
END_WHILE
Output_Fittest_Chromosome
END
```

Figure 8: Source code feature selection in GA

4.1 Dataset and Methods

The Proposed techniques contain a collection of item reviews from Amazon.com. This dataset contains three types of files positive, negative and unlabelled in XML format. These files were extracted using XML file splitter and reviews were converted into the text file. The dataset contains 500 positive files and 500 negative files for each domain. The reviews are about four-item domains: Books (B), DVDs (D), Electronics (E) and Kitchen appliances (K) and are written in the English language. For the experiment, labelled dataset of 500 positive and 500 negative files was used. An instance in each domain is recorded in Table 1. From this dataset, 12 cross-domain sentiment classification errands were

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 2, pp 80-85, February-2023

https://doi.org/10.47001/IRJIET/2023.702011

constructed: B → D; B → E; B → K; D → B; D → E; D → K; E → B; E → D; E → K; K → B; K → D; K → E, where the word before arrow corresponds to the source domain and the word after an arrow corresponds to the target domain.

This research work uses the below-mentioned performance metrics for the performance measure:

- Accuracy is used as an evaluation measure. Accuracy is the extent of correctly classified examples to the aggregate number of examples; then again, error rate refers to incorrectly classified examples to correctly classified examples. F-measure or precision and recall can be used as evaluation measures.
- F-measure is just defined in terms of true positive (TP), false positive (FP) and false-negative (FN), while a true negative (TN) isn't considered. Accuracy and F-measure are compared for a proposed approach which demonstrates that, in general, F-measure is like accuracy.

4.2 Accuracy Analysis

Accuracy indicates the algorithms ability to make differentiation among the better classes and correct classes. In order to estimate the accuracy, it is necessary to calculate the proportion from TP, TN, FP and FN. It is mathematically expressed as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

True Positive (TP): Number of instances with positive results that are correctly classified.

False Positive (FP): Number of instances with positive results that are wrongly classified.

True Negative (TN): Number of instances with adverse positive results that are wrongly classified.

False Negative (FN): Number of instances with adverse positive results that are wrongly correctly classified.

4.3 F-Measure Analysis

F-Measure is utilized to measure how far the accuracy is correct. It is also described as the weighted harmonic mean of precision and recall. In order to estimate the f-measure, it is necessary to calculate the proportion from precision and recall. It is mathematically expressed as

$$F - Measure = 2 \times \left(\frac{Precision \times Recall}{Precision + Recal} \right)$$

Classification Accuracy Analysis of GA

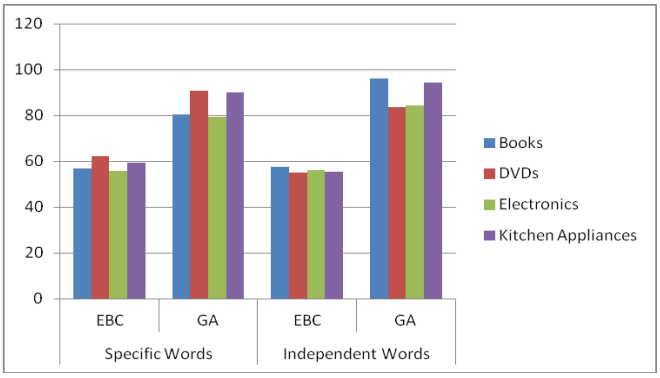


Figure 9: GA Classification Accuracy Analysis

Table 1: Numerical Values of GA

Domains	Specific Words		Independent Words	
	EBC	GA	EBC	GA
Books	56.71	80.35	57.63	96.11
DVDs	62.37	90.62	54.98	83.62
Electronics	55.76	79.21	56.02	84.20
Kitchen Appliances	59.45	89.88	55.29	94.30

V. CONCLUSION

This article reviews feature selection studies for reducing high dimensional data from a variety of disciplines. Subset formation, evaluation, and stopping criteria are part of the feature selection process.

The filter, wrapper, and embedding methods for feature selection as well as three general approaches are also included. Many typical applications of predictive analytics rely heavily on feature selection in genetic algorithms. One must choose the advanced genetic algorithm approaches because it contains a lot of features.

Genetic algorithm has high performance compared to the other feature selection algorithms with different classification techniques. It is the best feature selection algorithm for Cross domain opinion mining. In this paper show how genetic algorithms can be applied to optimize the performance of a cross domain by selecting the most relevant features selection. Genetic algorithms will be useful tools for data mining in many situations.

The paper shows that feature selection based on genetic algorithm along with an ensemble approach outperformed the other approaches. We conducted a comparative study experiments on multi domain dataset and Books, DVD Electronics review dataset in opinion mining.

ISSN (online): 2581-3048
Volume 7, Issue 2, pp 80-85, February-2023
https://doi.org/10.47001/IRJIET/2023.702011

This research article aims to propose Feature Selection using Genetic Algorithm Approach based Five Layered Artificial Neural Network (GA-FLANN) for Cross Domain Opinion Mining.

Develop optimal parameter values and select parameter values of GA that optimize the performance of Chromosomes populations' distribution, Processing of classification. GA has attained 90% classification accuracy in domain-specific words and 96% classification accuracy in domain-independent words.

REFERENCES

[1] Ali, Daehan Kwak, Pervez Khan, Shaker El-Sappagh, Amjad Ali, Sana Ullah, Kye Hyun Kim, Kyung-Sup Kwak, Transportation sentiment analysis using word embedding and ontology-based topic modeling, Knowledge-Based Systems, Volume 174, 2019, Pages 27-42.

[2] Chihli Hung, Shiuan-Jeng Chen, Word sense disambiguation based sentiment lexicons for sentiment classification, Knowledge-Based Systems, Volume 110, 2016, Pages 224-232.

[3] Sixing Wu, Yuanfan Xu, Fangzhao Wu, Zhigang Yuan, Yongfeng Huang, Xing Li, Aspect-based sentiment analysis via fusing multiple sources of textual knowledge, Knowledge-Based Systems, Volume 183, 2019, 104868. https://doi.org/10.1016/j.knosys.2019.104868.

[4] Pinlong Zhao, Linlin Hou, Ou Wu, Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification, Knowledge-Based Systems, Volume 193, 2020, 105443.

[5] María Lucía Barrón Estrada, Ramón Zatarain Cabada, Raúl Oramas Bustillos, Mario Graff, Opinion mining and emotion recognition applied to learning environments, Expert Systems with Applications, Volume 150, 2020, 113265.

[6] Aminu Da'u, Naomie Salim, Idris Rabiu, Akram Osman, Recommendation system exploiting aspect-based opinion mining with deep learning method, Information Sciences, Volume 512, 2020, Pages 1279-1292.

[7] Madhu Bala Myneni, Rohit Dandamudi, Harvesting railway passenger opinions on multi themes by using social graph clustering, Journal of Rail Transport Planning & Management, 2019, 100151.

[8] Aitor García-Pablos, Montse Cuadros, German Rigau, W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis, Expert Systems with Applications, Volume 91, 2018, Pages 127-137.

[9] Nagendra Kumar, Rakshita Nagalla, Tanya Marwah, Manish Singh, Sentiment dynamics in social media news channels, Online Social Networks and Media, Volume 8, 2018, Pages 42-54.

[10] Vivian Lay Shan Lee, Keng Hoon Gan, Tien Ping Tan, Rosni Abdullah, Semi-supervised Learning for Sentiment Classification using Small Number of Labeled Data, Procedia Computer Science, Volume 161, 2019, Pages 577-584.

[11] Vinodhini Gopalakrishnan, Chandrasekaran Ramaswamy, Patient opinion mining to analyze drugs satisfaction using supervised learning, Journal of Applied Research and Technology, Volume 15, Issue 4, 2017, Pages 311-319.

[12] Mu-Yen Chen, Ting-Hsuan Chen, Modeling public mood and emotion: Blog and news sentiment and socio-economic phenomena, Future Generation Computer Systems, Volume 96, 2019, Pages 692-699.

[13] Feilong Tang, Luoyi Fu, Bin Yao, Wenchao Xu, Aspect based fine-grained sentiment analysis for online reviews, Information Sciences, Volume 488, 2019, Pages 190-204.

Citation of this Article:

Dr. S.Gnanapriya, D.Umanandhini, "Feature Selection in Genetic Algorithm Based Sentiment Classification Used Five Layered Artificial Neural Network for Cross Domain Opinion Mining" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 2, pp 80-85, February 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.702011

****

