7135CEM - Modelling and Optimisation Under Uncertainty

Mansoor Syed, SID:11211802, Email: syedm35@uni.coventry.ac.uk

Marium. SID:11387853, Email: mariumm@uni.coventry.ac.uk

Faculty of Engineering Environment and Computing,

MSc. Data Science and Computational Intelligence,

Coventry University,

Coventry. United Kingdom

Table of Contents

Topic Modelling To Solve Hidden Relations In Unsupervised Text Data

I. ABSTRACT

Topic modeling is one of the best techniques for finding a collection of words in unsupervised text data. Comparing many techniques LSA and LDA have a prominent role to detect hidden patterns in a topic and understand a bag of words and making required documents. This is a vital factor to understand data to end-users in a readable format and recognize latent patterns in the unlabelled data and improve those areas. In this research paper, we analyzed by using the latest text data of vaccine myth where peoples have a different opinion on it, this paper evaluates and clean this text data with pre-processing step and shows documents and topic which discussed most in this unsupervised data by using LDA and LSA recognize top problems in the dataset to improve. In addition, we also evaluate which model performs best for this given data by using coherence parameters in LSA and LDA. To evaluate this process this research paper used advanced python programming language which has required inbuilt libraries for Topic Modelling.

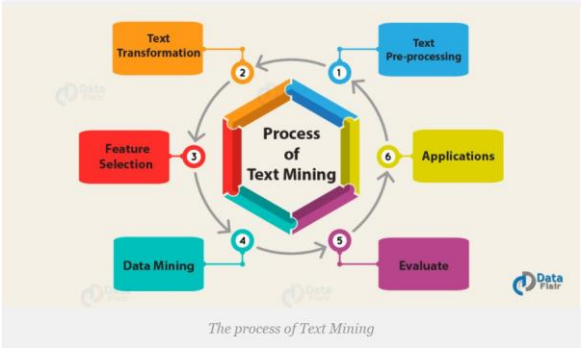
Keywords - LSA, LDA, Topic Modelling, Python.

II. INTRODUCTION

Nowadays people purely rely on a digital platform from starting of the day to the end of the day. Accumulation of electronic data is rapidly increasing every second, so to handle and understand those unsupervised data and get precise customer feedback and deliver efficient service through only with human knowledge is not possible if it is possible so companies required more people means more expenses and time taking but this problem will be solved through best Statical methods, so many researchers do analysis on machine learning and Natural language processing and understand Topic modeling is one of the best techniques for finding a collection of words in an unsupervised dataset. In this project to find word collection most powerful and precise statistical tools, we are used LDA (Latent Dirichlet Allocation) and LSA (Latent Semantic Analysis) These are Natural language processing tools that were executed accurately results in Unsupervised datasets. To make an analysis of unsupervised dataset it need to change into supervised by using text mining technique which is known as text data mining which is a process of altering unstructured data into a structured format. To make data in a particular format there was a clear process which is shown in figure 1.

Figure 1

Text Data Mining Steps



Note. From Data Mining, by Data Flair.

(https://data-flair.training/blogs/text-mining/)

In this Project before performing LSA And LDA pre-processing step was taken which was discussed in the coming section of the background study.

Let's discuss briefly how LSA and LDA models work in past research in topic modeling, there was vast research was taken place on those statistical tools from past decades, (Blei et al., 2003, p. 993) Blei and his colleges made analysis on LDA which is data modeling method which was put forward for the purpose of Topic model of data information with latent topic. Later this method used many researchers to show topic modeling on different sentimental analyses. (Steinberger et al., 2004, p. 93) group of researchers made analysis on LSA for summarizing the task on Topic on Novelty attributes when reconstructing diagonal matrix S and they show positive results with the LSA algorithm.(Kalepalli et al., 2020) kalepali and colleges made a comparison of LSA and LDA in their research paper they conclude there were different overcomes according to iteration, authors mentioned in the research paper there was less divergence in LSA after 250 iterations as opposed to LDA having better divergence after 250 iterations. The latest research made by Gupta and Patel (Gupta et al., 2021) on LSA and LDA on Topic modeling their results shown LSA overcomes LDA they made analysis on text summarization. Like many researchers made analysis on a different dataset with different problems a new problem arises when new dataset was analyzing. even after much research work made on these algorithms using this same algorithm with a new dataset is crucial to solve a new problem. In this research paper we are used latest dataset which was updated on December in the year 2021 which was vaccine myth dataset many researchers evaluated analysis on vaccine myth latest analysis done by Indian researchers in the year 2021 on covid vaccine myth, for that analysis 200 participants were used for survey in it 32 were doctors they made sentiment analysis on participants asking question and reasons which stopped them by not taking the vaccine, they executed their analysis effectively. But they made analysis only on the Indian vaccine myth with only 200 participates but this project analyzed various vaccine myth data around the world which was extracted from thousands of tweets and posts of people with sentimental analysis of text data.

III. DATASET

Dataset which was using in this project was latest Kaggle dataset of Reddit Vaccine Myth which has 8 columns and 1597 entries which is small text unsupervised dataset.

Figure 2

Sample Dataset

	title	score	id	url	comms_num	created	body	timestamp
0	Health Canada approves AstraZeneca COVID-19 va...	7	lt74vw	https://www.canadaforums.ca/2021/02/health-can...	0	1.614400e+09	NaN	2021-02-27 06:33:45
1	COVID-19 in Canada: 'Vaccination passports' a ...	2	lsh0ij	https://www.canadaforums.ca/2021/02/covid-19-i...	1	1.614316e+09	NaN	2021-02-26 07:11:07
2	Coronavirus variants could fuel Canada's third...	6	lohlle	https://www.canadaforums.ca/2021/02/coronaviru...	0	1.613887e+09	NaN	2021-02-21 07:50:08
3	Canadian government to extend COVID-19 emergen...	1	lnptv8	https://www.canadaforums.ca/2021/02/canadian-g...	0	1.613796e+09	NaN	2021-02-20 06:35:13
4	Canada: Pfizer is 'extremely committed' to mee...	6	lkslm6	https://www.canadaforums.ca/2021/02/canada-pfi...	0	1.613468e+09	NaN	2021-02-16 11:36:28
5	Canada: Oxford-AstraZeneca vaccine approval ex...	5	lftbji	https://www.canadaforums.ca/2021/02/canada-oxf...	0	1.612869e+09	NaN	2021-02-09 13:17:11
6	Comment	1	ej9x066	NaN	0	1.553474e+09	Your OP. It's not a myth. Only one vaccine con...	2019-03-25 02:34:53

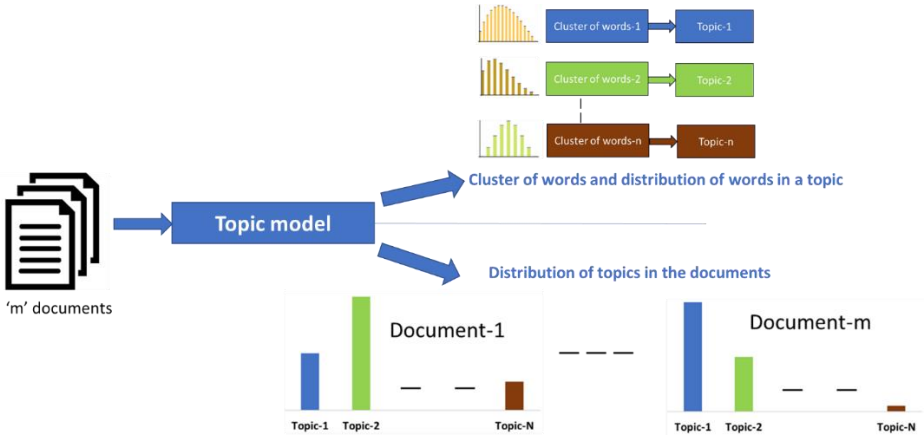
the dataset which was using in this project was unsupervised dataset which do not have proper labels to understand and it was cleaned by using pre-process and made clusters to the dataset and shows which data belong to in which topics by analysing documents. In addition, shows proper label for each topic, the process which extract data by topic was done by using latest NLP tools of LDA and LSA so this project can easily evaluate thousands of entries in seconds which gets more time to human to read and analysis of those topics. This is fastest process which get results as fast as possible according to results people need to work on related problem to solve the issues.

IV. BACKGROUND STUDY

a. Topic Modelling

Figure 3

Topic Modelling



Note. From Analytics Vidhya, by Chirag Goyal, 2021.

(https://www.analyticsvidhya.com/blog/2021/06/part-16-step-by-step-guide-to-master-nlp-topic-modelling-using-lsa/)

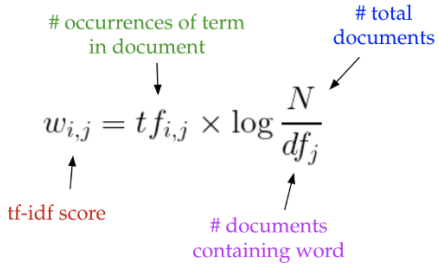
The above figure clearly shows how topic modeling works by filtering 'm' number of documents on 'm' number of topics by clustering of words and distribution of words in a topic.

b. Latent Semantic Analysis (LSA)

Latent Semantic Analysis is one of the fundamental techniques in topic modeling which is used for text summarization and text classification and dimension reductions. Due to many meanings of slightly different text data was only recognized by humans, not with the machine where the same text has different hidden concepts or topics around so in this situation LSA plays a major role to understand those latent concepts of a topic. In LSA we generate Matrix the word present in the paragraph of the document in the corpus. The row shows each paragraph word that was present in documents' unique values and the column matrix indicates each word in which documents it was present. For that TF-IDF is used to calculate the occurrence of a word in a corpus.

Figure 3

TF-IDF



Note. From Analytics Vidhya, by Chirag Goyal, 2021.

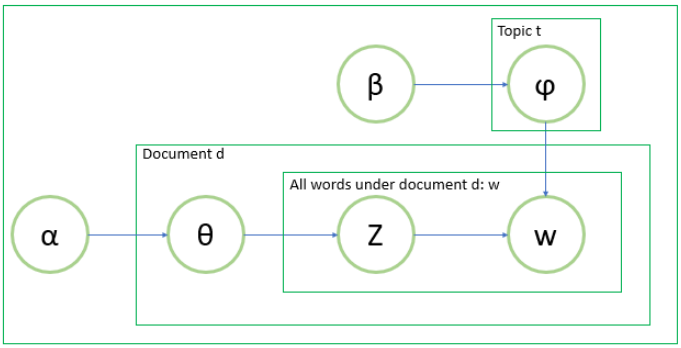
(https://www.analyticsvidhya.com/blog/2021/06/part-16-step-by-step-guide-to-master-nlp-topic-modelling-using-lsa/)

c. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation algorithm was best for the purpose of text modelling, the word Latent means which recognize hidden patterns in the documents and Dirichlet indicates which distributes topics in documents or words in a topic called Dirichlet distribution.

Figure 4

LDA Algorithm



Note. From Towards Data Science, by Edward Ma, 2018.

(https://towardsdatascience.com/2-latent-methods-for-dimension-reduction-and-topic-modeling-20ff6d7d547)

V. EXPRIMENTAL SETUP

This research paper analysis is done through python programming language for that this project used Jupiter Notebook from anaconda 3 which has the latest python version of 3.8 and imported and installed required libraries. Before performing LSA and LDA there was an important step needed to perform which is called pre-processing.

Pre-Processing

In this project we created a dictionary and corpus which was base to the topic modeling and created a unique id for each word in the documents with the help of genism. LSA and LDA algorithm recognize meaningful words but it does not recognize stop words like a, an, and so on. To remove those words from the text dataset there was a process called pre-processing by downloading and importing Natural Language Processing Tools (NLTK) and performing stemming and tokenization effectively. There was a stop word inbuilt function present in python which was performed and executed below figure 5.

Figure 5

Stop Words

```
326
{'off', 'except', 'must', 'was', 'say', 'while', 'five', 'unless', 'she', 'about', 'an', '‘ve', 'out', 'of', 'against', 'made', 'whom', 'along', 'our', 'that', 'somewhere',
'ca', 'whereby', 'name', 'when', 'a', 'n‘t', 'another', 'more', 'rather', "'re", 'whence', 'onto', 'ourselves', 'with', 'show', 'thereupon', 'these', "'m", 'their', 'whereve
r', 'all', 'very', 'bottom', "'ll", 'eleven', 'your', 'since', 'otherwise', 'her', 'because', 'many', 'below', 'even', 'never', 'yourselves', 'therefore', 'one', 'last', 'ev
ery', 'its', 'three', 'still', 'part', 'ours', 'from', 'here', 'amongst', 'used', 'whoever', 'but', 'sometimes', 'they', 'if', 'full', 'own', 'get', 'anyhow', 'did', 'fort
y', 'yours', 'someone', 'thereby', 'any', 'yet', 'four', 'hundred', 'mostly', 'seem', 'to', 'throughout', 'seeming', 'beyond', "'ve", 'are', 'everything', 'whatever', 'sixt
y', 'keep', 'through', 'seemed', 'using', 'is', 'afterwards', 'anyone', 'anything', 'nine', 'together', 'for', 'without', 'always', '’ve', 'moreover', 'during', 'upon', 'giv
e', 'have', 'down', 'both', 'around', 'first', 'front', 're', 'cannot', 'done', 'formerly', 'up', 'under', 'once', 'whereupon', 'put', "n't", 'whither', 'amount', 'also',
'‘m', 'among', 'besides', 'fifty', 'should', 'becoming', 'thus', 'were', 'six', 'where', 'his', 'noone', 'various', 'anywhere', 'such', 'third', 'therein', 'no', 'seems', 'l
ess', 'whereas', 'than', 'move', 'which', 'in', 'perhaps', 'everyone', 'him', 'others', 'twelve', 'several', 'before', 'neither', 'may', 'those', 'via', 'whereafter', 'int
o', 'call', 'few', 'by', 'much', 'on', 'me', 'other', 'hereupon', 'you', "'s", 'nobody', 'alone', 'or', 'take', 'hereafter', 'nevertheless', 'thru', 'nor', '‘s', 'becomes',
'towards', 'i', 'whose', 'meanwhile', 'enough', 'sometime', 'regarding', 'two', '’re', 'whether', 'became', 'herein', 'please', 'and', 'doing', '‘re', 'empty', 'who', 'every
where', 'make', 'although', 'latterly', 'too', "'d", 'thereafter', 'ever', 'themselves', 'almost', 'serious', 'we', 'how', 'within', 'hence', 'same', 'it', 'namely', 'mine',
'until', 'been', 'again', 'what', 'behind', 'further', 'do', 'does', 'wherein', 'something', 'already', 'above', 'itself', 'least', 'himself', 'myself', 'being', 'really',
'why', 'us', '’ll', '’d', 'ten', 'be', 'due', 'quite', 'become', 'them', 'this', 'herself', 'however', 'top', 'has', 'latter', 'over', 'whole', 'n’t', '’s', 'eight', 'most',
'none', 'after', 'might', 'thence', '‘d', 'would', 'hereby', 'so', 'fifteen', 'beside', 'somehow', 'either', 'toward', 'next', 'else', 'only', 'whenever', 'not', 'yourself',
'just', 'elsewhere', 'he', '‘ll', '’m', 'go', 'anyway', 'at', 'former', 'nothing', 'will', 'as', 'indeed', 'there', 'twenty', 'now', 'see', 'then', 'between', 'back', 'acros
s', 'hers', 'am', 'had', 'each', 'side', 'beforehand', 'nowhere', 'can', 'often', 'my', 'the', 'though', 'well', 'could', 'per', 'some'}
```

Figure 6

original dataset

```
0                                Health Canada approves AstraZeneca COVID-19 vaccine
1    COVID-19 in Canada: 'Vaccination passports' a near certainty says bio-ethicist
2                                Coronavirus variants could fuel Canada's third wave
3                          Canadian government to extend COVID-19 emergency benefits
4       Canada: Pfizer is 'extremely committed' to meeting vaccine delivery targets
5                    Canada: Oxford-AstraZeneca vaccine approval expected this week
6                                                  Fuck you anti-vaxxing retards
7                                       COVID-19: Músicos que han recibido la vacuna
8                                       Anti-Vaccine Points Refuted A Thousand Times
9                                Does the Vitamin K Shot Contain 100mcg of Aluminum?
Name: title, dtype: object
```

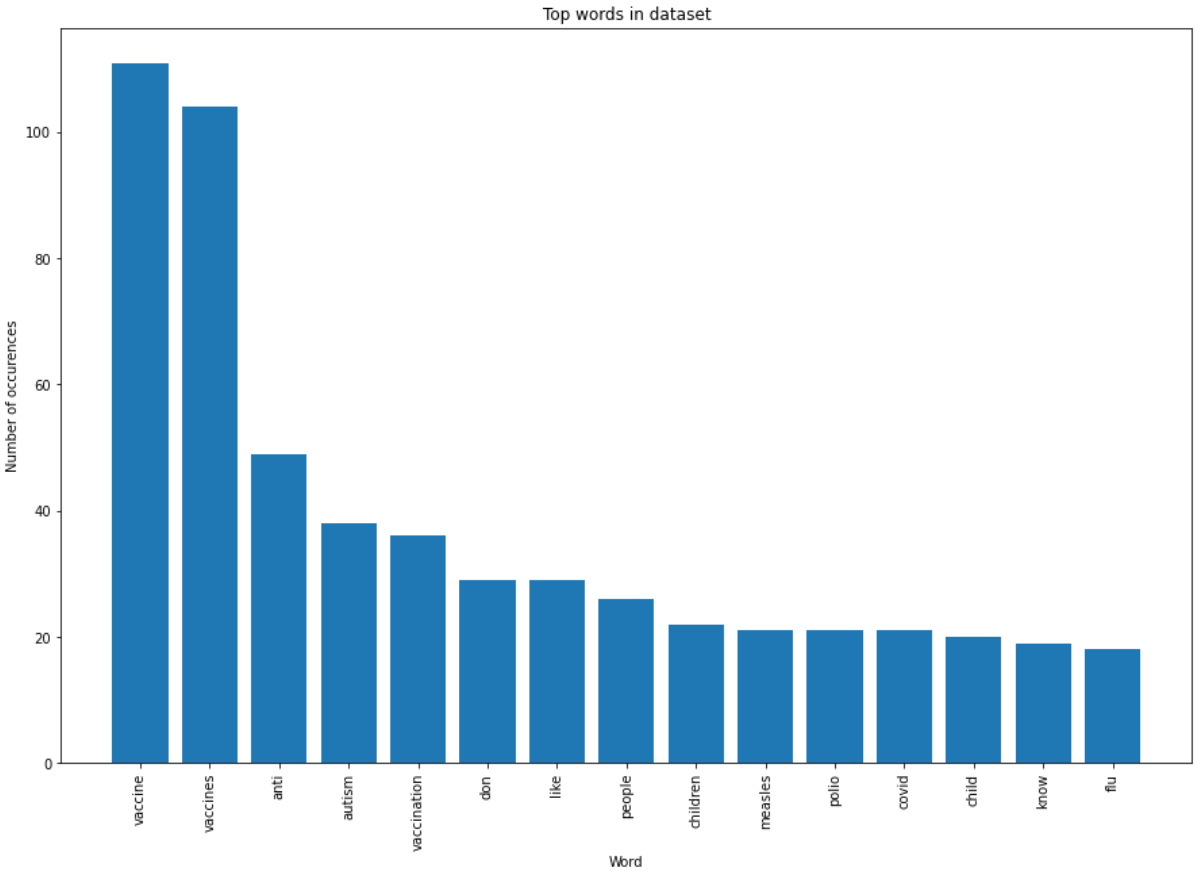
Figure 7

Data after cleaning

```
[['health', 'canada', 'approves', 'astrazeneca', 'covid19', 'vaccine'],
 ['covid19',
  'canada',
  'vaccination',
  'passport',
  'near',
  'certainty',
  'say',
  'bioethicist'],
 ['coronavirus', 'variant', 'fuel', 'canada', 'wave'],
 ['canadian', 'government', 'extend', 'covid19', 'emergency', 'benefit'],
 ['canada',
  'pfizer',
  'extremely',
  'committed',
  'meeting',
  'vaccine',
  'delivery',
```

Figure 8

Top Words in Dataset after cleaning



The above bar graph shows words of text which were present in our dataset, the main topic people talk about in this text dataset is about vaccine, and the least talk about flu.

Figure 9

Total Number of Words in the dataset after cleaning

```
Total number of words:  7860
Mean number of words per post:  16.83083511777302
```

The total number of words that were present in our dataset was 7860 after removing stop words and the mean value of each post was 16.83 words.

Figure 10

Post Showing Before and After Vectorization

```
Before Vectorization: "You believe what you want, but I know what happened to my child. I don't care what studies you have posted"
After Vectorization:
  (0, 63)       1
  (0, 615)      1
  (0, 310)      1
  (0, 253)      1
  (0, 114)      1
  (0, 186)      1
  (0, 95)       1
  (0, 544)      1
  (0, 424)      1
```

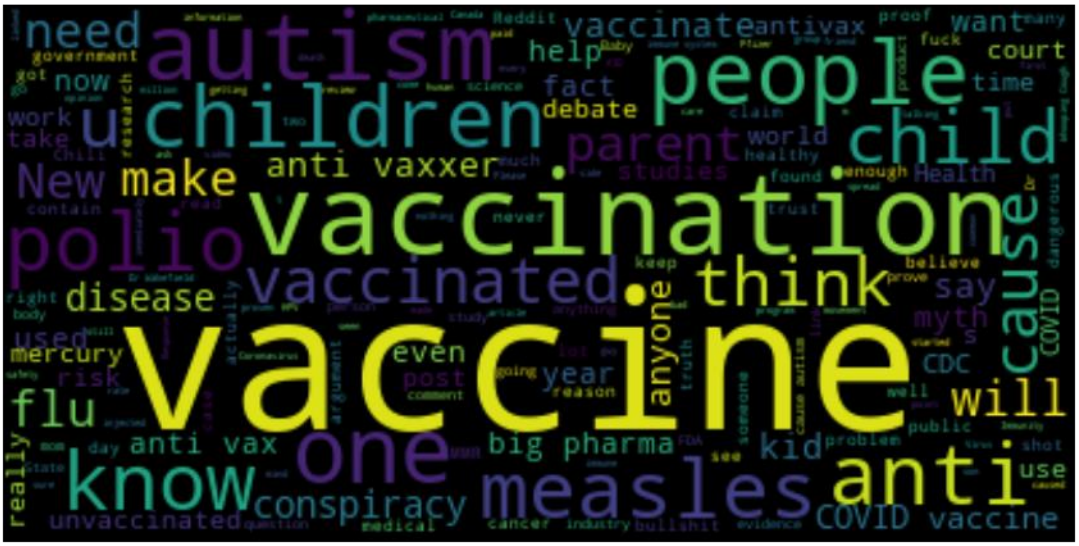
To apply LDA and LSA analysis we need to convert text data into the vectorized form so the above results clearly show how data was changed to vector form and displayed 10 lines of data as shown above in figure 10.

Exploratory analysis

To verify whether the pre-processing has been done effectively we installed word cloud in Jupiter platform and imported word cloud, stop words and image color generator from installed word cloud to see top words in image format and conforms pre-processing process has done successfully or whether needed before training process. But our results show perfect top words which were shown the same as in bar graph figure 8. In this research paper pre-processing was efficiently completed and also we can quickly see in cloud top words which are Vaccine, vaccination, and anti.

Figure 11

Top Words with Exploratory analysis



VI. RESULTS

LSA Model Analysis

In this project, we performed the LSA model and were assigned to show the top five topics in the dataset by giving a command number of commands equal to five. Below figure 12 shows the LSA model's top five topics.

Figure 12

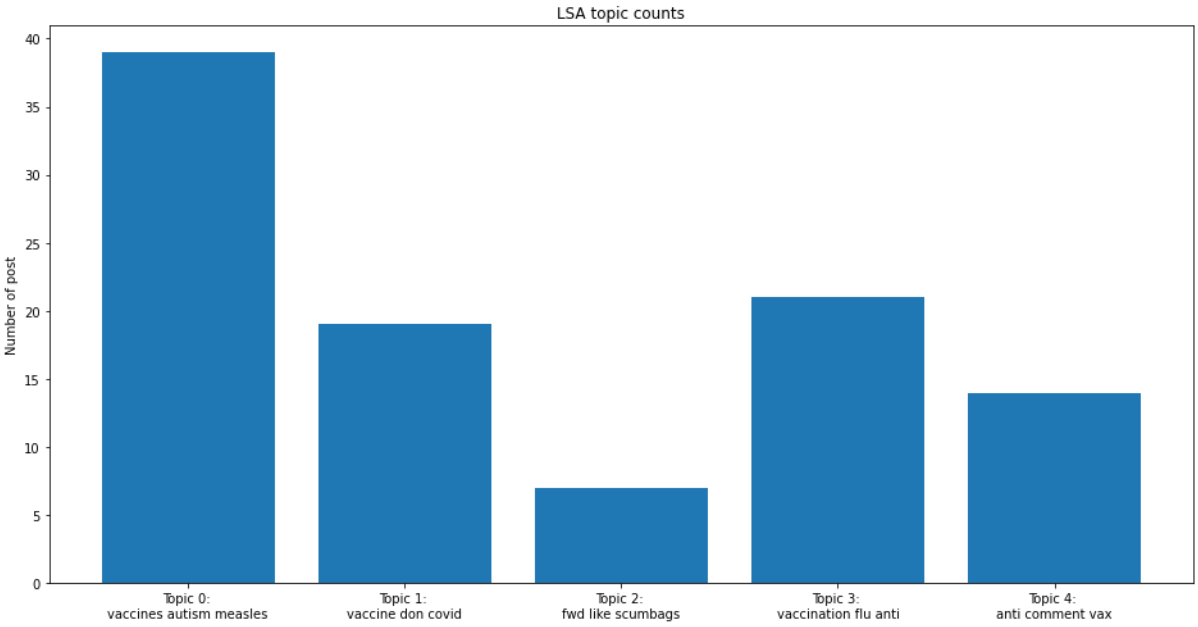
LSA Model Top 5 Topics

```
Topic 1:  vaccines autism measles like really proof contain vaccinated big flu non
Topic 2:  vaccine don covid dr medical history children people til news new
Topic 3:  like fwd industry scumbags course use heads murdering arose majority running
Topic 4:  vaccination anti flu vaccinations don mercury need argument fuck protect immunity
Topic 5:  anti comment vax paid bots article zero map opinion older object
```

Figure 12 shows the top five topics of our dataset and displayed 11 words of each topic.

Figure 13

LSA Model Top 5 Words in Topics



The above figure shows the top five topics which were present in our dataset extracted through the LSA model are

Topic 1: Vaccines Autism Measles

Topic 2: Vaccine don covid

Topic 3: fwd like scumbags

Topic 4: vaccination flu anti

Topic 5: anti comment vax

Figure 14

T-distributed Stochastic Neighbour Embedding

```
[t-SNE] Computing 99 nearest neighbors...
[t-SNE] Indexed 100 samples in 0.000s...
[t-SNE] Computed neighbors for 100 samples in 0.004s...
[t-SNE] Computed conditional probabilities for sample 100 / 100
[t-SNE] Mean sigma: 0.575917
[t-SNE] KL divergence after 250 iterations with early exaggeration: 47.027943
[t-SNE] KL divergence after 1050 iterations: 0.056431
```

t-SNE is a dimensional method to visualize high dimensional data it was in this project because of unstructured text data which was trained to analyze the LSA model, parameters that were given to fit the data in LSA are given below.

perplexity=50

number of components = 2

learning rate=100

number of iterations = 2000

verbose=1

random state=0 (42)

angle=0.75

if we see the above results of t-SNE KL (Kullback–Leibler) divergence after 250 iterations with early exaggeration was 47.02 and after 1050 iterations it was 0.05 which was extremely low.

Figure 15

t-SNE Clustering of Five LSA Topics

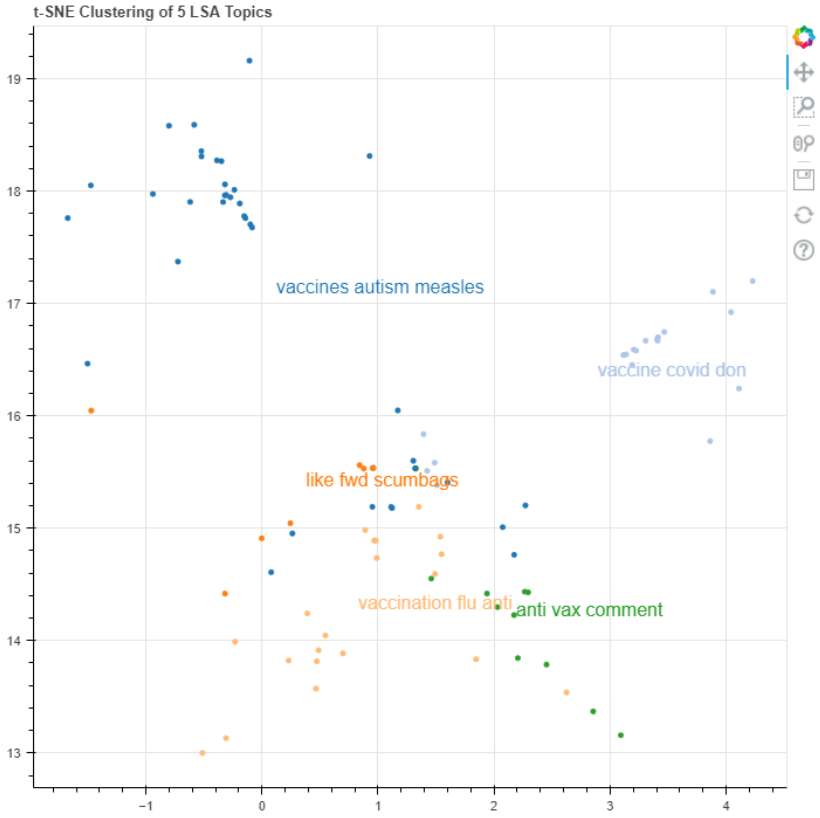


Figure 16

Coherence score of Latent Semantic Indexing

```
Coherence Score for LSI using c_v:  0.45936408787172345

Coherence Score for LSI using u_mass:  -14.416571139438503
```

The coherence score of the LSA model of C_V and U_mass are 0.45 and -14.41 respectively.

LDA Model Analysis

The second algorithm which was used in this project was the LDA model which is the best model for Topic modeling. we assign 5 topics to show through the LDA model which were extracted as shown in below figure 17.

Figure 17

LDA Model Top 5 Topics

```
Topic 1:  vaccines anti measles truth dihydrogen blog check reviews contain april
Topic 2:  vaccine dr health work autism kids til polio 80 don
Topic 3:  vaccines comment industry said make having section scumbags bots murdering
Topic 4:  vaccines flu non children vaccination review medical measles person isn
Topic 5:  vaccines vaccine don like anti vaccination people vaccinations help fwd
```

Figure 18

LDA Model Top 5 words in Topics

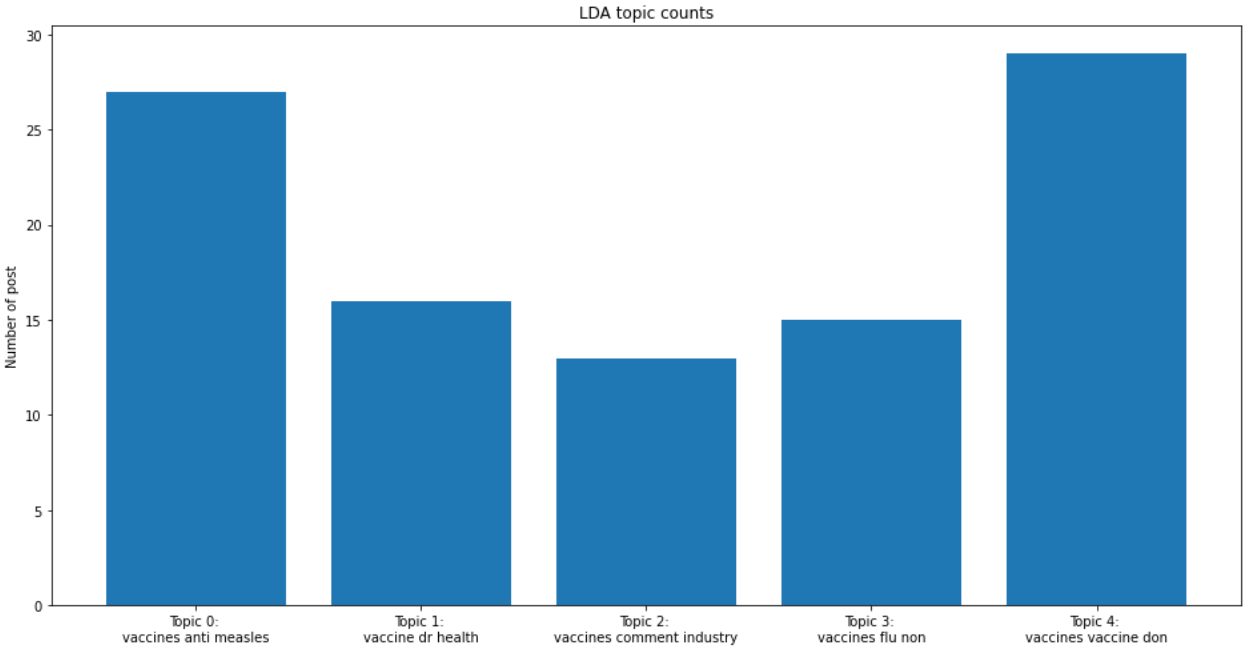


Figure 19

T-distributed Stochastic Neighbour Embedding

```
[t-SNE] Computing 99 nearest neighbors...
[t-SNE] Indexed 100 samples in 0.000s...
[t-SNE] Computed neighbors for 100 samples in 0.002s...
[t-SNE] Computed conditional probabilities for sample 100 / 100
[t-SNE] Mean sigma: 0.736287
[t-SNE] KL divergence after 250 iterations with early exaggeration: 47.017986
[t-SNE] KL divergence after 1450 iterations: 0.032563
```

Almost similar results occur in LDA in KL divergence after 250 iterations with early exaggeration was 47.01 and after 1450 iterations it shows with 0.03.

Figure 20

Coherence score of LDA

```
Perplexity:  -7.593914392239438

LDA Coherence Score using c_v:  0.506756675497816

LDA Coherence Score using u_mass:  -14.045941582608602
```

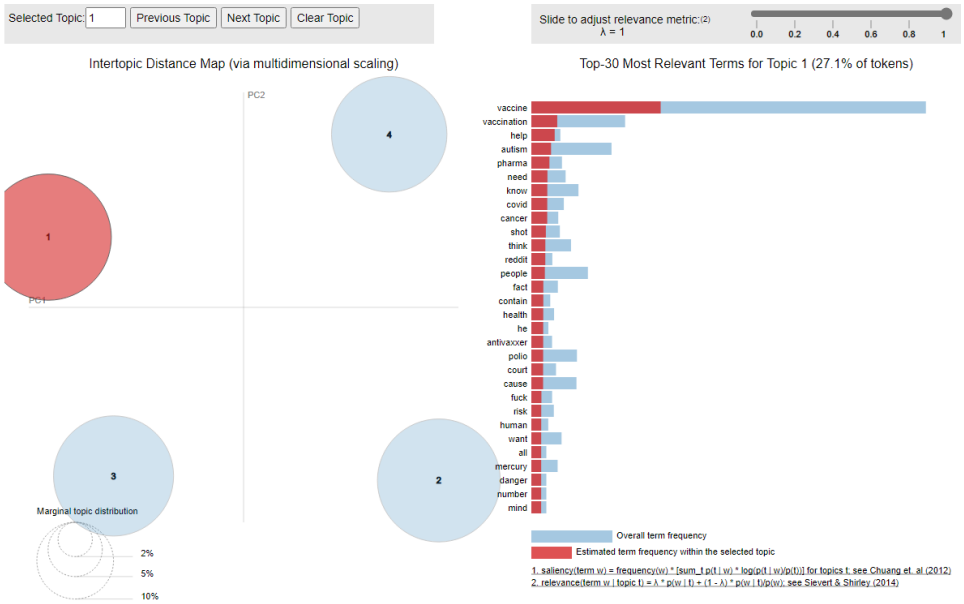
The coherence score of LDA using C_V and U_Mass is 0.50 and -14 respectively.

Figure 21

t-SNE Clustering of Five LSA Topics

Figure 22

PylLDA Topic 1 Results



In Topic 1 with help of pylLDA shown top 30 term frequency words. But we only look top 5 words to demonstrate data clearly which are vaccine, vaccination, help, autism , and pharma which shown in above figure 22. At same way this project shown top words from other topic model which were shown below figure 23,24 and 25 respectively.

Figure 23

PylLDA Topic 2 Results

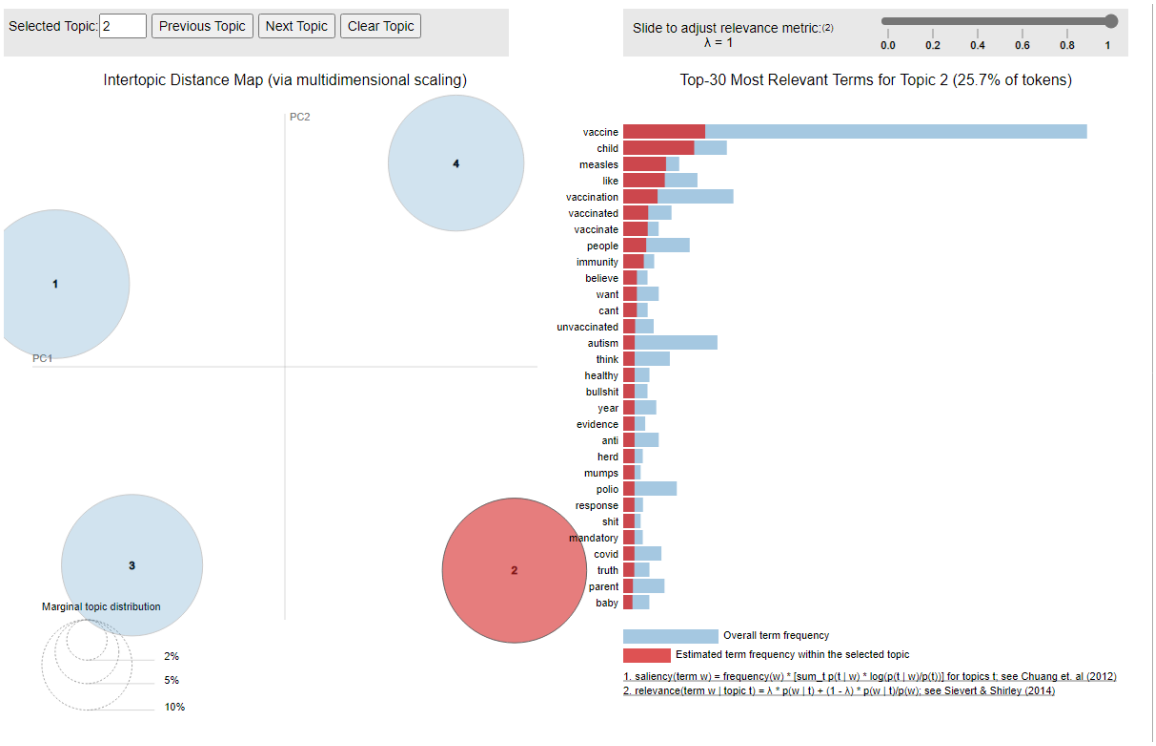


Figure 24

PylLDA Topic 3 Results

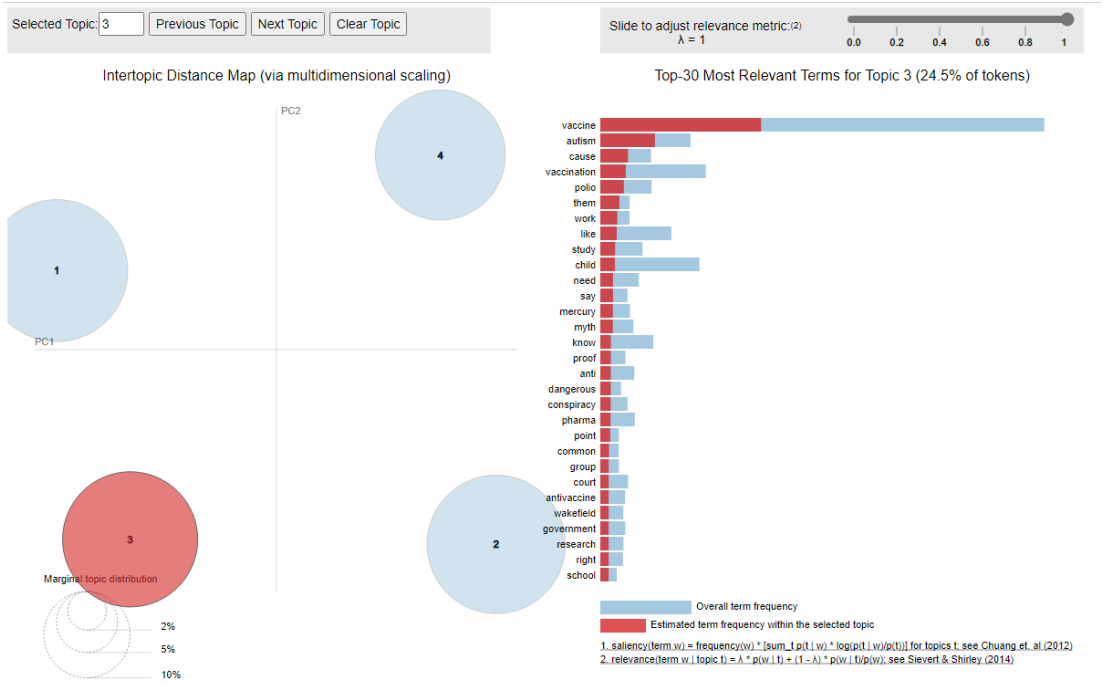
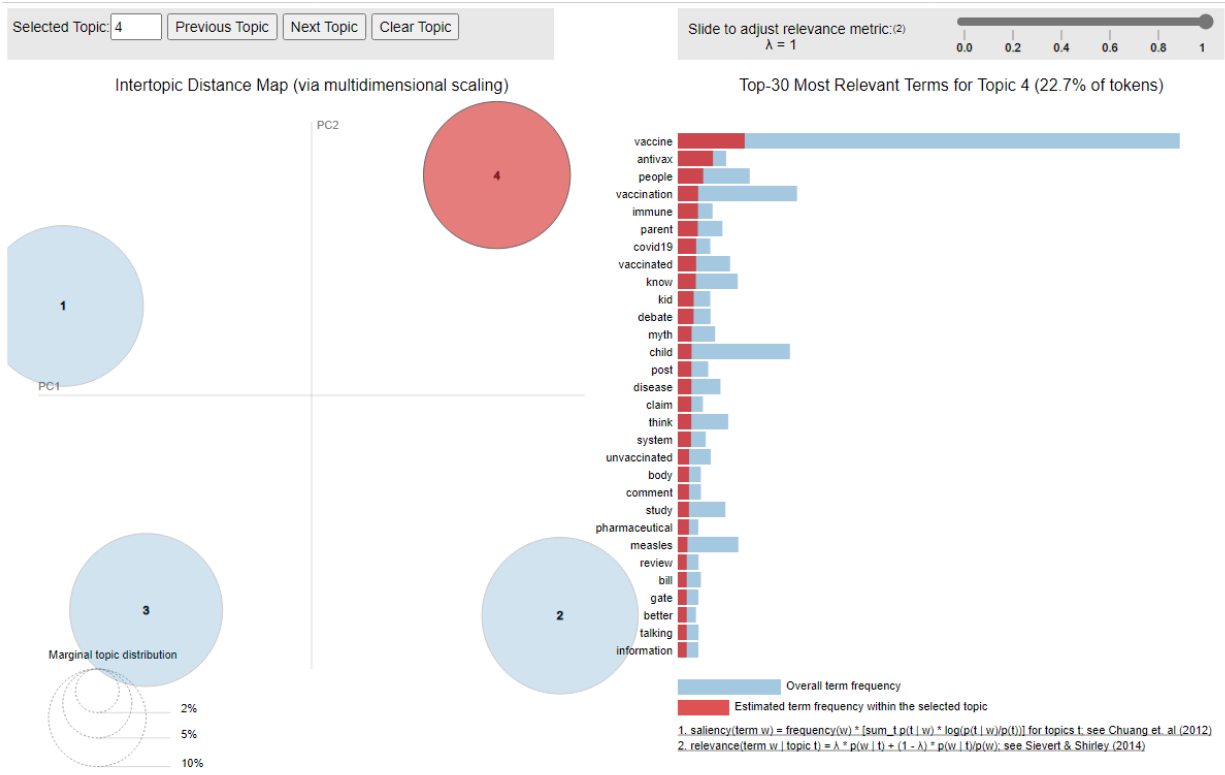


Figure 25

PylLDA Topic 4 Results



VII. DISCUSSIONS

By observing two algorithms LSA and LDA have almost similar exaggerations results after 250 iterations with 47.02 and 47.01 respectively. But changes seen in coherence results LSA has 0.45 and LDA has 0.50 in C_V value in U_Mass almost similar results with -14. These results clearly show LDA outperformed LSA. If we see top topic models in our data which are mainly talk about vaccines, autism, antivax, measles, and polio. This is the most frequently spoken topic in this text data, in these areas government or any organization needs to improve to take steps and clear people myths in this field.

VIII. CONCLUSION

In conclusion, Unstructured data become most common today's world due to so many social media platforms using people. Government or any organization to understand peoples problems and solve their issues through unstructured data is easily possible but understanding unsupervised data is a tough task to many organizations due to the unlabelled data structure but this problem was solved in this project through LSA and LDA by pre-processing the data in this process LDA overcome LSA and shown frequently speaking problems about the vaccines, polio, autism, anti-vax, and measles this are the main topic people has myth on it, so holding campaigns get familiar to people on those topics and eradicate there fear with help of good solutions is necessary.

IX. REFERENCES

Data Flair. *What is Text Mining in Data Mining – Process & Applications.*

https://data-flair.training/blogs/text-mining/

Blei, M., Jordan, I., & Ng, A.(2003). Latent Dirichlet Allocation.

Journal of Machine Learning Research (3rd Vol.), 993-1022.

Steinberger,J., & Jezek,k.(2004). *Using latent semantic analysis in text summarization and summary Evaluation.* 93-100.

Kalepalli, Y., Tasneem, S., Pasupuleti, D., & Suneetha, M. (2020). *Effective Comparison of LDA with LSA for Topic Modelling*. IEEE Publications. *https://ieeexplore.ieee.org/document/9120888*.

Hritvik, G., & Mayank, P. (2021). *METHOD OF TEXT SUMMARIZATION USING LSA AND SENTENCE BASED TOPIC MODELLING WITH BERT.* IEEE Publications. *https://ieeexplore.ieee.org/document/9395976*.

Rutvik, D., Kopal, S., & Karthik, K.(2021) .*Statistical and Sentimental Analysis on Vaccination against COVID-19 in India*. IEEE Publications. *https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9510179.*

Edward, M.(2018).2 latent methods for dimension reduction and topic modeling.

Towards Data Science. https://towardsdatascience.com/2-latent-methods-for-dimension-reduction-and-topic-modeling-20ff6d7d547.

Appendix

Python Code:

https://github.com/mansoorsyed11/7135CEM---Modelling-and-Optimisation-Under-Uncertainty-/blob/9f99e31e621d52c21e42c874c944f9aa8a793711/My%20OP%20Complete%20Code.ipynb

PROPOSAL

Topic Modelling To Solve Hidden Relations In Unsupervised Text Data

Name: Mansoor Syed, Student ID: 11211802, Mail ID: syedm35@uni.coventry.ac.uk

Name: Marium, Student ID: 11387853, Mail ID: mariumm@uni.coventry.ac.uk

Project Proposal:

Nowadays people purely reply on digital platform from starting of the day to end of the day. Accumulation of electronic data is rapidly increasing in every second so to handle and understand those unsupervised data and get precise customer feedback and deliver efficient service through only with human knowledge is not possible if it is possible so companies required more people means more expanses and time taking but this problem will be solved through best Statical methods, so many researchers do analysis on machine learning and Natural language processing and understand Topic modelling is one of the best techniques for finding collection of words in unsupervised dataset. In this project to find word collection most powerful and precise statistical tools we are going to use LDA (Latent Dirichlet Allocation) and LSA (Latent Semantic Analysis) this are Natural language processing tools which were execute accurate results in Unsupervised datasets. If requires more statistical tools to be compared additionally while doing this project. To do this project we need unsupervised text data due to lack of time this project was extracting dataset from Kaggle which was recent 2021 dataset of ABC News which has 1195191 unique values, and two column features which are Published Date and Headline Text. To execute this project this research paper going to use advanced high level python programming language which has required machine learning libraries to solve this problem. In conclude this project creates clusters to extracted data made analysis shows which algorithms gives best results.

Dataset Link:

https://www.kaggle.com/hamditarek/topic-modeling-lsa-and-lda/data

Project Plan:

The estimated time of this Project is 30 hours but due to new algorithms and advanced programming language we work on this project more hours as we required until to finish this project before deadline.

Step 1: Understands the Topics of LSA and LDA on Topic Modelling analysis difference between Topic Classification and Topic modelling to have clear idea on Topic modelling.

Step2: analysis of dataset features and prepare and code python programming language to present the results.

Step3: according to results we obtained need to write clear and precise report by using required sections.

Step4: clearly recheck our report and programming code and submit before deadline.

Sample data:

	publish_date	headline_text
0	20030219	aba decides against community broadcasting lic...
1	20030219	act fire witnesses must be aware of defamation
2	20030219	a g calls for infrastructure protection summit
3	20030219	air nz staff in aust strike for pay rise
4	20030219	air nz strike to affect australian travellers

Changed dataset due to many people used same dataset.

NEW DATASET: https://www.kaggle.com/gpreda/reddit-vaccine-myths

Task 2: Evolutionary and Fuzzy Systems

Fuzzy Logic Optimized Controller (FLC) for an Intelligent Assistive Care Environment

Fuzzy Logic controllers are widely used in industries for various purposes. Many situations demand change in the how It works and addition of some components in it. Care homes are always on the lookout for workers and caregivers to care for the elderly, and there has been a drop in the number of persons applying for the position, as well as the risks connected with hiring anyone, given that the elderly are frail and any disease or infection might be fatal to them. In light of this, all care homes experienced a personnel shortage in 2020 and 2021 as a result of the Corona virus epidemic, and faced numerous challenges. The demand for assisted living technologies is greater than ever before. The design and implementation of a fuzzy logic controller for assisted living for disabled residents is discussed in this study. A system where Clients' activity can be monitored by assisted care systems, which can notify ambulance service if they become hurt. They can also remind individuals of possible dangers around in the house, or by a stove that has been left on. Sensation and audio feedback from engaging robots could be used to emulate a human or animal partner. Many ways exist to govern this system, but fuzzy logic was chosen because it provides a human-like sensibility to the control approach and is self-tolerant to less accurate data. Fuzzification, Defuzzification, and Fuzzy Rule inference are all components of the Fuzzy Logic Controller. The goal is to learn how to use the Fuzzy Rule base and inference methods to regulate environmental parameters as well as other things like prescription reminders and temperature. When the level of precision necessary isn't too high, it's a great option. It's a strong, easy-to-manage method. It has the ability to process several inputs and produce a variety of outputs.

Fuzzy Logic

Fuzzy logic is a means of modelling logical thinking in which a statement's truth isn't binary. The fuzzy inference system is an artificial intelligence system that mimics how humans solve problems. "I'm pretty weary, but it's just about gym time," for example. Pretty weary and just about gym time are two terms with a lot of ambiguity. What does pretty weary mean, and how much time left until gym time? Without fuzzy logic, it would be difficult for a machine to make conclusions based on human fuzzy notions. The concept of fuzzy logic is analogous to how people experience and infer things. “Brian is old (fluent, muscular, needy, etc.)," for example, is designed to mimic logical reasoning using ambiguous or inadequate assertions. It's a type of many-valued logic in which the values are expressed as levels of truth. By assigning levels of truth to statements, fuzzy logic emerges. Fuzzy logic is a method of encoding experience-based information in the form of logical principles that computers can grasp.

We are going to build a smart home for a disabled residence that can help them to do daily activates easily in an automatic and simple manner. To build a smart home we need required input and output parameters that can be suitable to make an intelligent home. The plan was designed in MATLAB by using Fuzzy logic Designer. Every input and output parameters membership function was assigned as required to smart home and rules were created and displayed surface of the parameters to understand input and output results in visualized manner.

Inputs of FLC system:

Temperature, Humidity, Micro Phone, Presence Sensor, Corporal Sensor, Environmental Sensor, CO2 sensor, and Automatic Robot switch.

Outputs of FLC system:

HVAC, TV, Light, Medication Control, Blind, Fire Alarm, and Cleaning Robot.

Defuzzification: Centroid

Membership Function Types:

In this project, we used only two types of membership function

i. Gaussian Membership Function

$$f(x;\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}}$$

Standard deviation 'σ' and *Mean* 'c' with Two Parameters

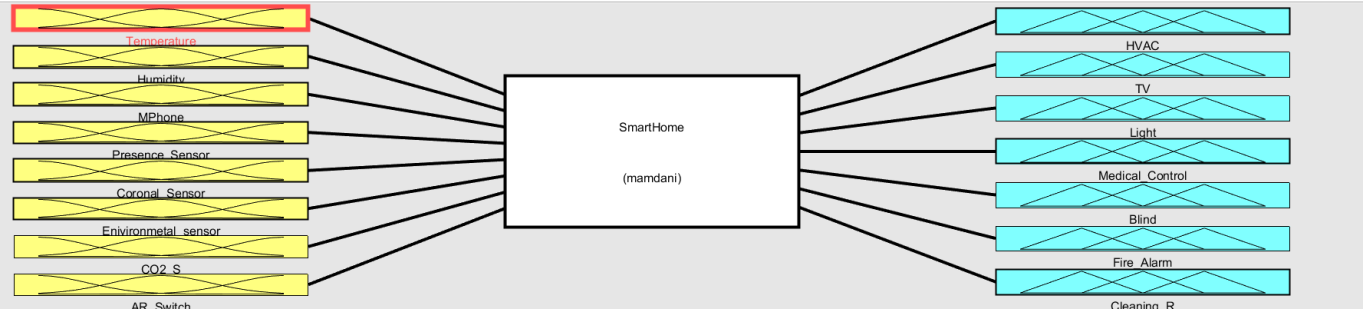
ii. Triangular Membership Function

$$f(x;a,b,c) = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ \frac{c-x}{c-b}, & b \le x \le c \\ 0, & c \le x \end{cases}$$

In triangular membership, the function has three parameters as a, b, and c.

Figure 1

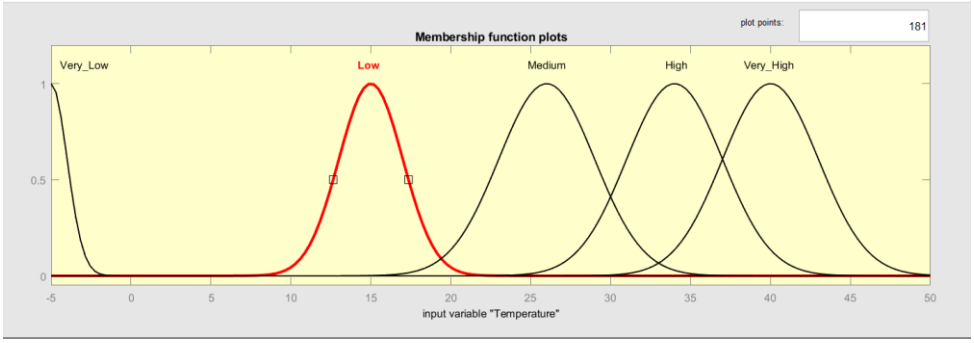
Design of Smart Home with FLC



Membership Functions of Inputs Variable

Figure 1.1

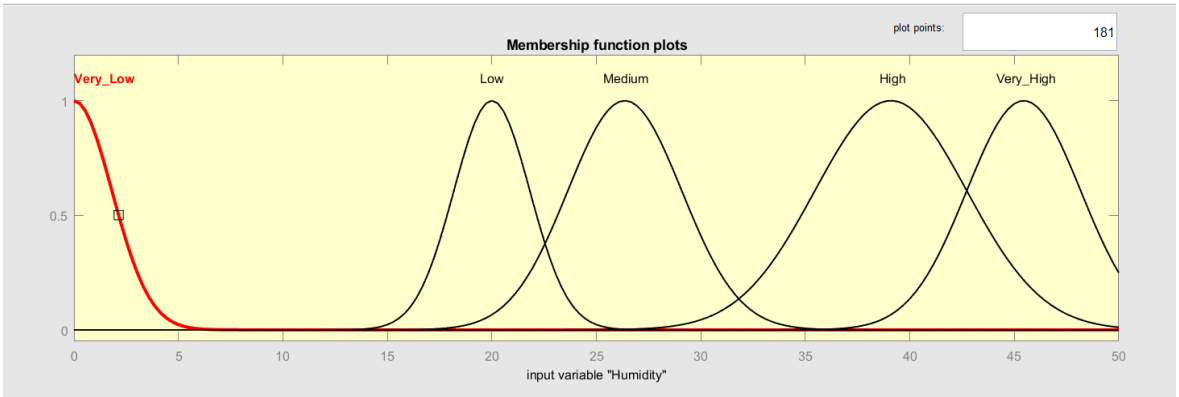
Membership Function Plots of Input Variable Temperature



In the above-shown Five Membership Plots of Temperature Variable range are Very Low, Low, Medium, High, and Very High. The Temperature units are ‘Celsius’.

Figure 1.2

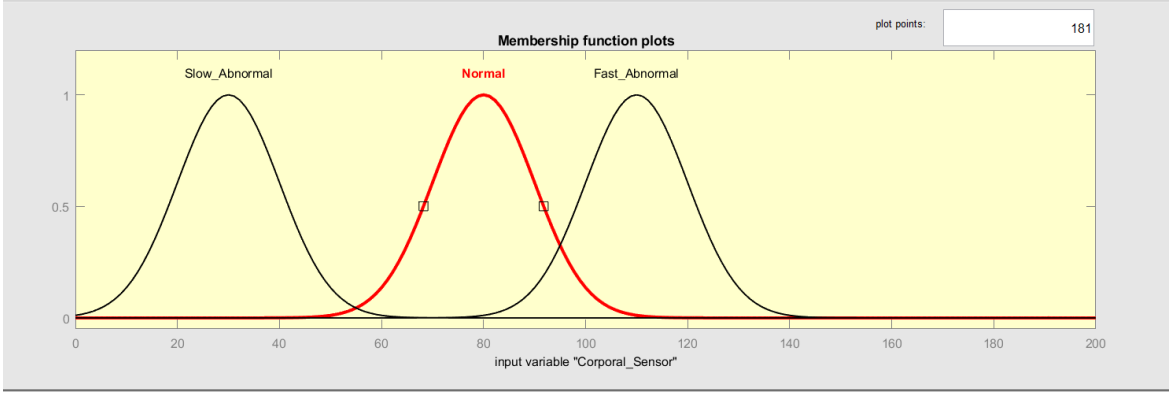
Membership Function Plots of Input Variable Humidity



In the above-shown Five Membership Plots of Humidity Variable range are Very Low, Low, Medium, High, and Very High. The Humidity units are '**gm -3**'

Figure 1.3

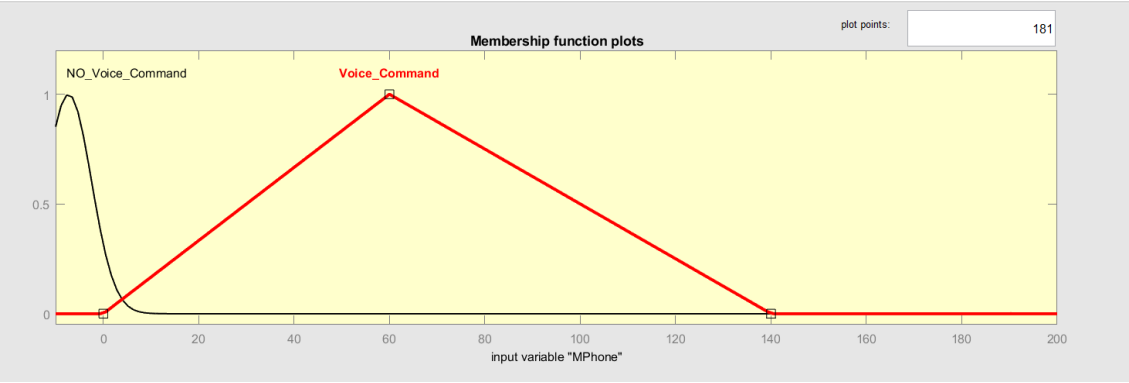
Membership Function Plots of Input Variable Corporal Sensor



In the above-shown Three Membership Plots of Corporal Sensor Variable are abnormal, Normal, Abnormal. The units are heartbeat rate per minute**.**

Figure 1.4

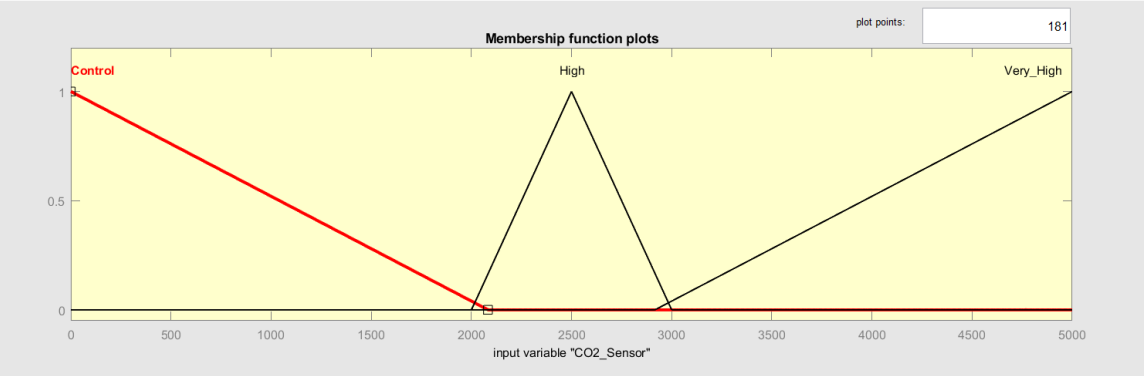
Membership Function Plots of Input Variable Micro Phone



There are only two membership function presents in Micro Phone which is Voice command and No Voice command. This microphone responds when sound is in between 0 to 140 decibels. Units are in dB

Figure 1.5

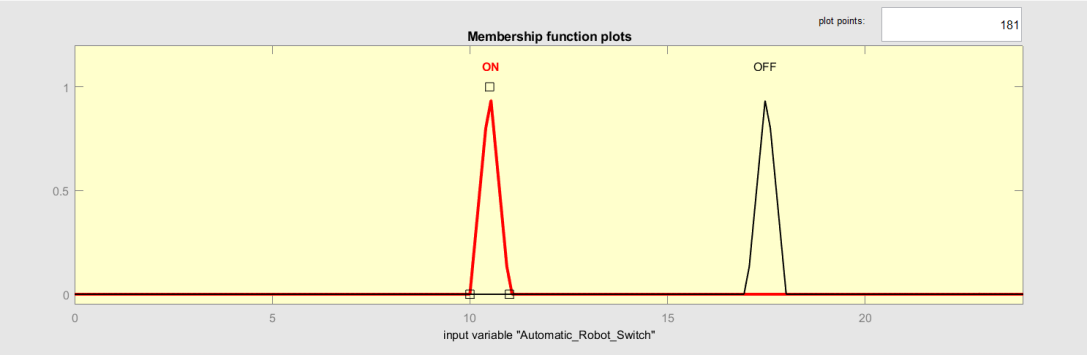
Membership Function Plots of Input Variable CO2 Sensor



There were Three Membership Variables for the above co2 sensor input variables with a range of Control, High, and Very High. The units are ppm.

Figure 1.6

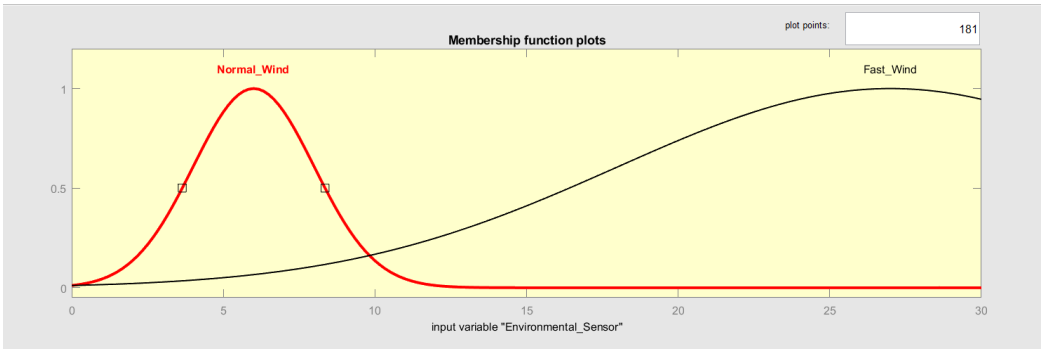
Membership Function Plots of Input Variable Automatic Robot Switch



There are only two membership function presents in Automatic Robotic Switch which is ON and OFF with certain time period the automatic robot switch turn off after one hour time every day it starts at 10 am and end at 11 am, it function twice a day morning and evening at 6 to 7 pm. The units are time 24 hours per day.

Figure 1.7

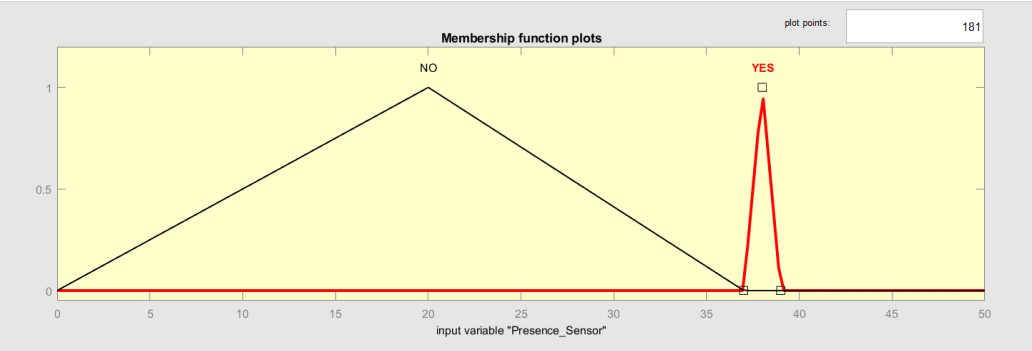
Membership Function Plots of Input Variable Environmental Sensor



There were Two Membership Variables for the above Environmental sensor input variables with a range of Normal Wind, and Fast Wind. The units are KM/PH.

Figure 1.8

Membership Function Plots of Input Variable Presence Sensor

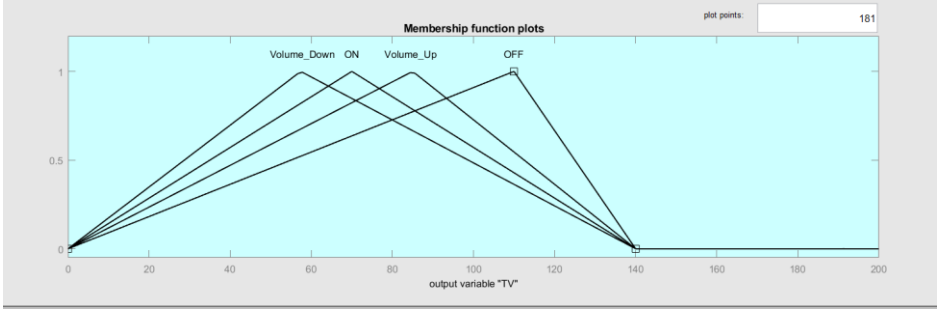


There is only two membership function presents in Presence Sensor which is YES and NO. it works only when human temperature matches between 36 to 39 degree Celsius.

Membership Functions of Outputs Variable

Figure 1.9

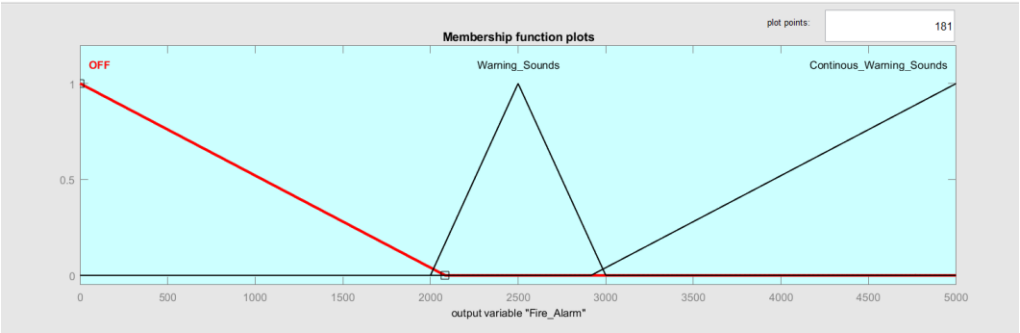
Membership Function Plots of Output Variable TV



There were Four Membership Functions present in the Output variable of TV which controls ON, Volume Down, Volume Up, and OFF. These commands are connected to an IoT device of Micro Phone where we can command our voice to handle those tasks in a smart manner without the use of any physical items. The sound should be 0 to 140 dB to get a response.

Figure 1.10

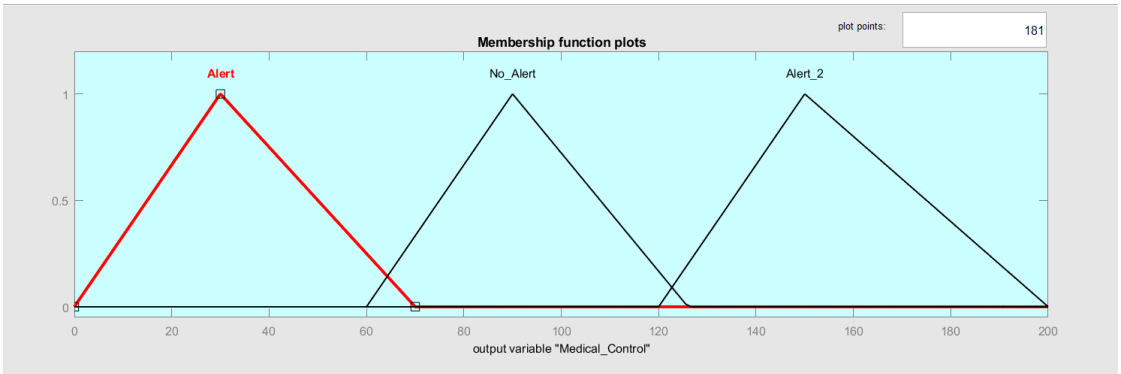
Membership Function Plots of Output Variable Fire Alarm



In this output parameter, there was three membership function present which is OFF, Warning Sounds, Continuous Warning sounds this was connected to CO2 sensor input according to CO2 level it will be responded.

Figure 1.11

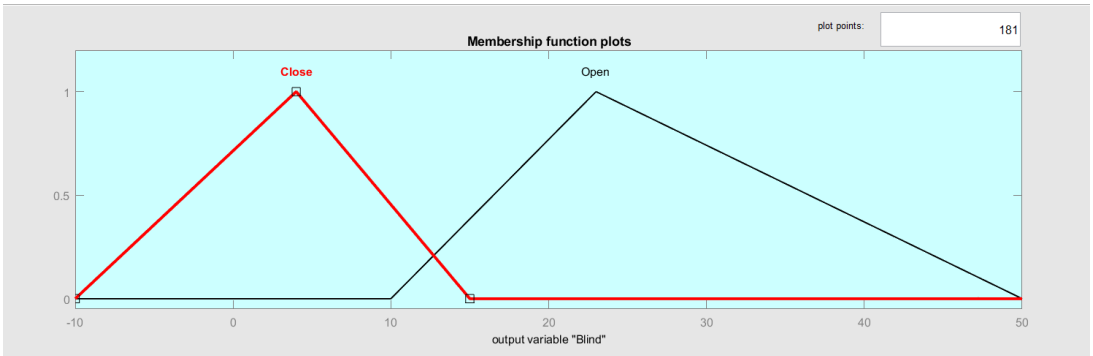
Membership Function Plots of Output Variable Medication Control



There were Three Membership Functions present in the Output variable of Medication Control which makes actions like Alert when there was an abnormal seen in the coronal sensor.

Figure 1.12

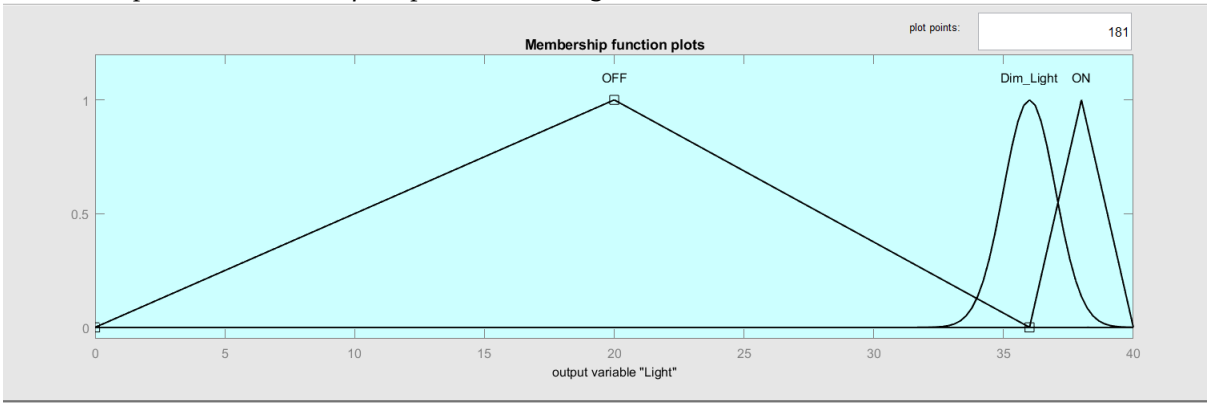
Membership Function Plots of Output Variable Blind



There were Two Membership Functions present in the Output variable of Blind which is Open and Close. It will be responded to according to Environmental Sensor.

Figure 1.13

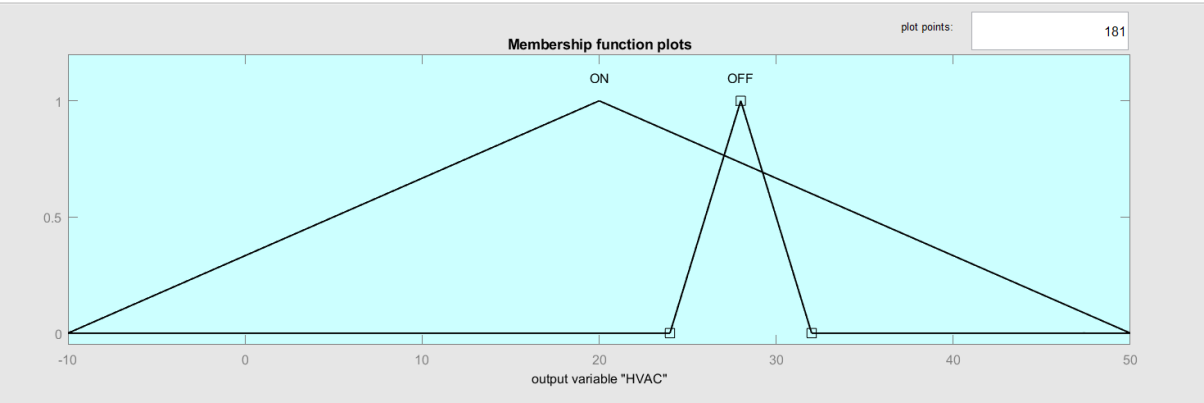
Membership Function Plots of Output Variable Light



There were Three Membership Functions present in the Output variable of Light which are OFF, ON, and Dim. Which were connected to IoT device of Microphone and works through commands as well as human body temperature when some enter in the smart home according to body temperature it Turns ON the light if it is no human it Turns off.

Figure 1.14

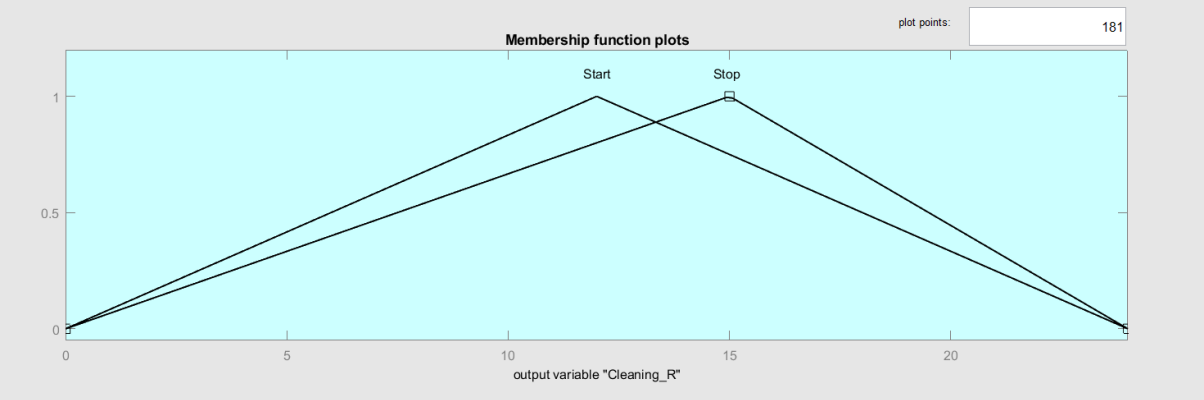
Membership Function Plots of Output Variable HVAC



There were Three Membership Functions present in the Output variable of HVAC which are ON, OFF, and Turn ON which works based on Temperature and Humidity. This plays a major role in smart homes the functions of Heating, ventilation, and air conditioning, and many more functions.

Figure 1.15

Membership Function Plots of Output Variable Cleaning Robot



There were Two Membership Functions present in the Output variable of the Cleaning Robot which is Start and Stops. Which were connected to Robot Switch.

Fuzzy Rules

we have created 42 rules of Fuzzy Logic Control which are listed below.

Figure 1.16

Fuzzy Logic Rules

1. If (Temperature is Very_Low) and (Humidity is Very_Low) then (HVAC is ON) (1)
2. If (Temperature is Very_Low) and (Humidity is High) then (HVAC is ON) (1)
3. If (Temperature is Very_High) or (Humidity is Very_High) then (HVAC is ON) (1)
4. If (Temperature is Low) and (Humidity is Very_High) then (HVAC is ON) (1)
5. If (Temperature is Medium) and (Humidity is Very_High) then (HVAC is ON) (1)
6. If (Temperature is High) and (Humidity is Very_High) then (HVAC is ON) (1)
7. If (Temperature is Medium) and (Humidity is Medium) then (HVAC is OFF) (1)
8. If (Temperature is Medium) and (Humidity is Low) then (HVAC is OFF) (1)
9. If (Temperature is Medium) and (Humidity is Very_Low) then (HVAC is OFF) (1)
10. If (Temperature is Medium) and (Humidity is Very_Low) and (Environmental_Sensor is Normal_Wind) then (Blind is Open) (1)
11. If (Temperature is High) and (Humidity is Very_Low) and (Environmental_Sensor is Normal_Wind) then (Blind is Open) (1)
12. If (Temperature is Low) and (Humidity is Very_Low) and (Environmental_Sensor is Normal_Wind) then (Blind is Open) (1)
13. If (Temperature is Very_High) and (Humidity is Medium) and (Environmental_Sensor is Normal_Wind) then (Blind is Open) (1)
14. If (Temperature is Very_High) and (Humidity is High) and (Environmental_Sensor is Normal_Wind) then (Blind is Close) (1)
15. If (Temperature is Very_High) and (Humidity is Very_High) and (Environmental_Sensor is Normal_Wind) then (Blind is Close) (1)
16. If (Temperature is Medium) and (Humidity is Medium) and (Environmental_Sensor is Fast_Wind) then (Blind is Close) (1)

17. If (Temperature is High) and (Humidity is Low) and (Environmental_Sensor is Fast_Wind) then (Blind is Close) (1)
18. If (Temperature is High) and (Humidity is Very_Low) and (Environmental_Sensor is Fast_Wind) then (Blind is Close) (1)
19. If (Environmental_Sensor is Fast_Wind) then (Blind is Close) (1)
20. If (Micro_Phone is Voice_Command) or (Presence_Sensor is YES) then (Light is ON) (1)
21. If (Presence_Sensor is YES) then (Light is ON) (1)
22. If (Micro_Phone is NO_Voice_Command) and (Presence_Sensor is YES) then (Light is ON) (1)
23. If (Micro_Phone is NO_Voice_Command) and (Presence_Sensor is NO) then (Light is OFF) (1)
24. If (Micro_Phone is NO_Voice_Command) or (Presence_Sensor is NO) then (Light is OFF) (1)
25. If (Micro_Phone is Voice_Command) then (Light is ON) (1)
26. If (Micro_Phone is Voice_Command) then (Light is OFF) (1)
27. If (Micro_Phone is Voice_Command) then (Light is Dim_Light) (1)
28. If (Micro_Phone is Voice_Command) then (TV is ON) (1)
29. If (Micro_Phone is Voice_Command) then (TV is OFF) (1)
30. If (Micro_Phone is Voice_Command) then (TV is Volume_Up) (1)
31. If (Micro_Phone is Voice_Command) then (TV is Volume_Down) (1)
32. If (Coronal_Sensor is Normal) then (Medical_Control is No_Alert) (1)
33. If (Coronal_Sensor is Slow_Abnormal) then (Medical_Control is Alert) (1)
34. If (Coronal_Sensor is Fast_Abnormal) then (Medical_Control is Alert_2) (1)
35. If (CO2_Sensor is Control) then (Fire_Alarm is OFF) (1)
36. If (CO2_Sensor is High) then (Fire_Alarm is Warning_Sounds) (1)
37. If (CO2_Sensor is Very_High) then (Fire_Alarm is Continous_Warning_Sounds) (1)
38. If (CO2_Sensor is Very_High) then (Blind is Open)(Fire_Alarm is Continous_Warning_Sounds) (1)
39. If (Micro_Phone is Voice_Command) then (HVAC is ON) (1)
40. If (Micro_Phone is Voice_Command) then (HVAC is OFF) (1)
41. If (AR_Switch is ON) then (Cleaning_R is Start) (1)
42. If (AR_Switch is OFF) then (Cleaning_R is Stop) (1)

Surface area Visualization

Figure 1.17

Surface area diagram of Humidity, Temperature, and HVAC

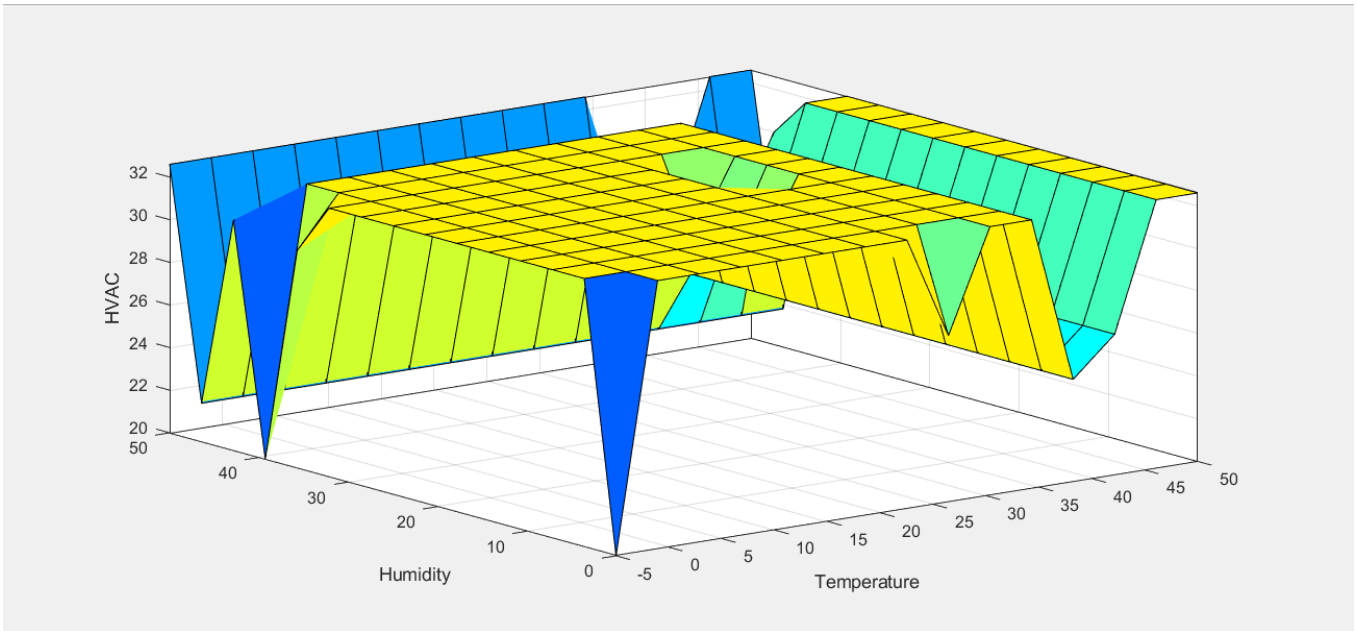


Figure 1.18

Surface area diagram of Temperature, Environmental Sensor and HVAC

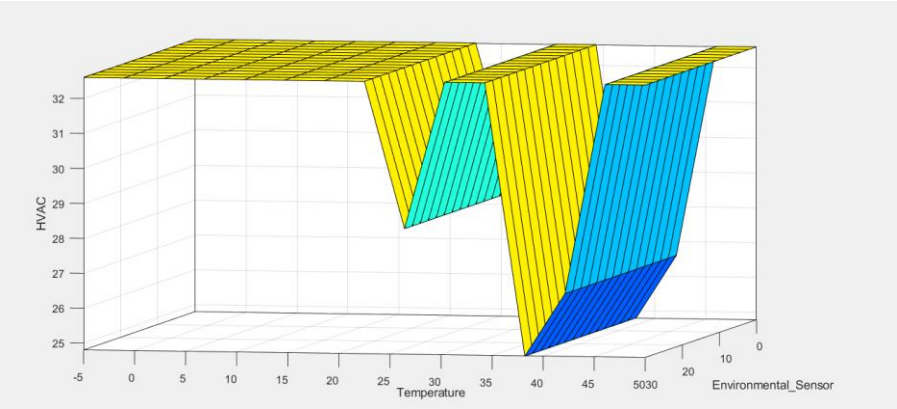


Figure 1.19

Surface area diagram of Physical Sensor, Micro Phone, and HVAC

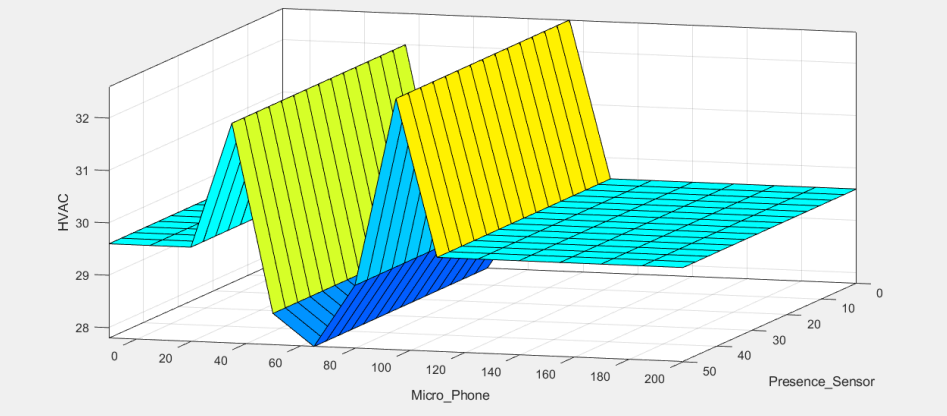


Figure 1.20

Surface area diagram of Micro Phone, Presence Sensor, and Light.

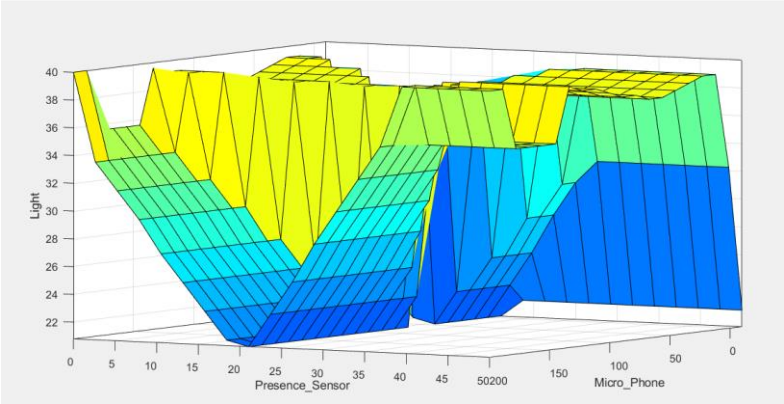


Figure 1.21

Surface area diagram of CO2 Sensor, Temperature, and Fire Alarm.

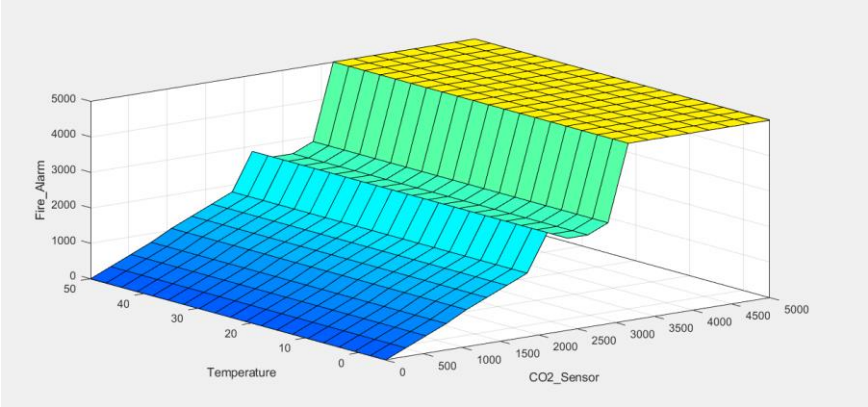


Figure 1.22

Surface area diagram of Temperature, Environmental sensor, and Blind.

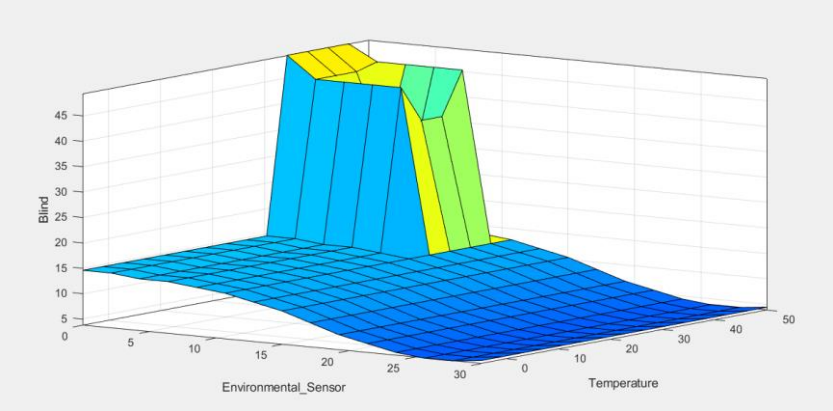


Figure 1.23

Surface area diagram of Presence Sensor, Micro Phone, and TV.

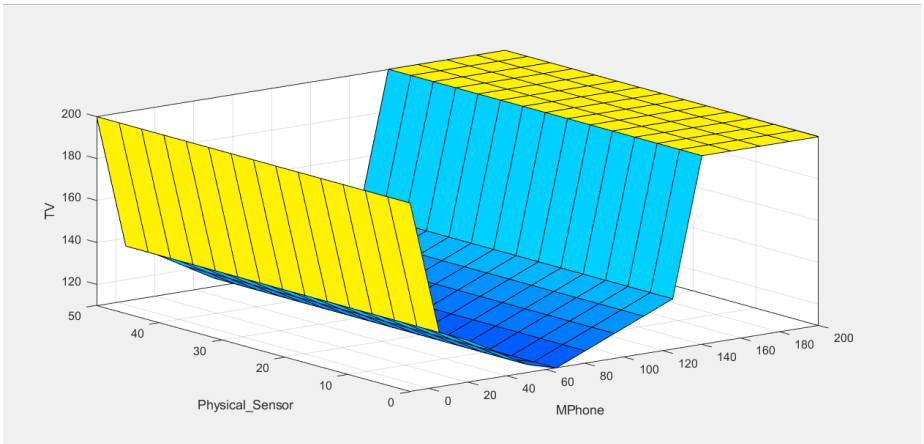


Figure 1.24

Surface Plot of CO2 Sensor and Fire Alarm.

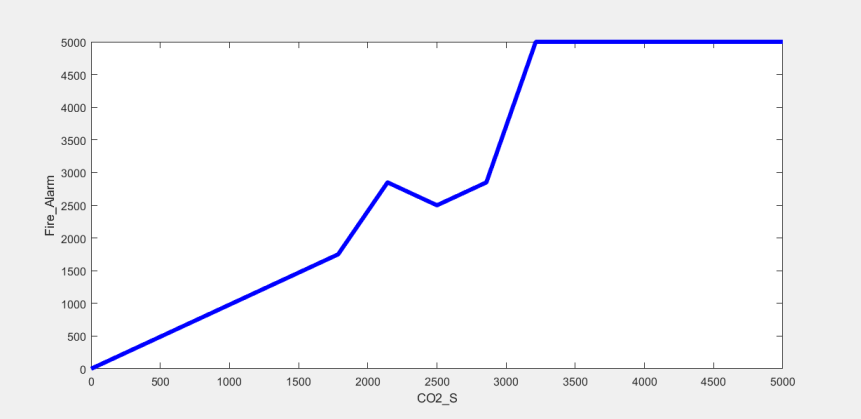


Figure 1.25

Surface Plot of Light and Presence Sensor.

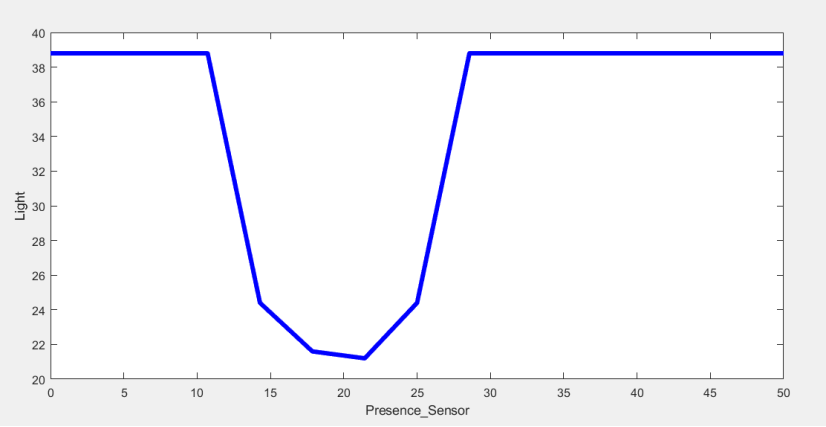


Figure 1.26

Surface Plot of Corporal and Medication Control.

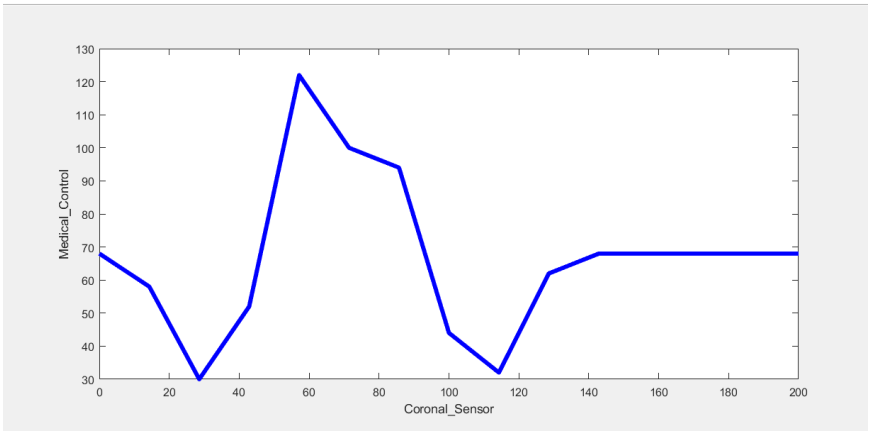


Figure 1.27

Surface Plot of Blind and Environmental Sensor.

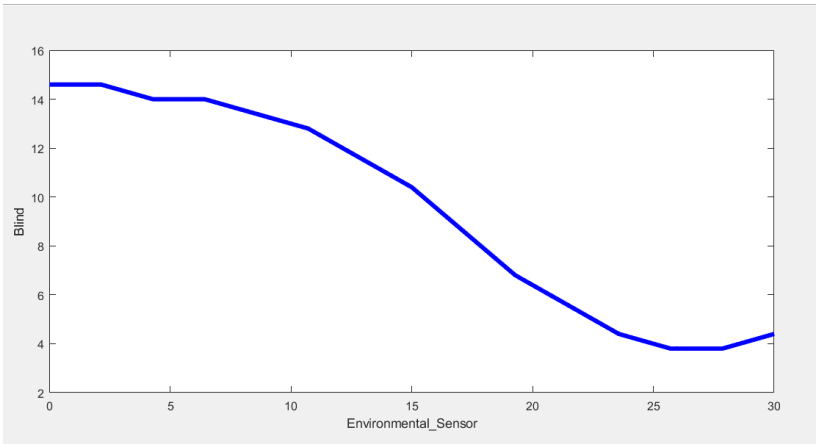


Figure 1.28

Surface Plot of HVAC and Micro Phone.

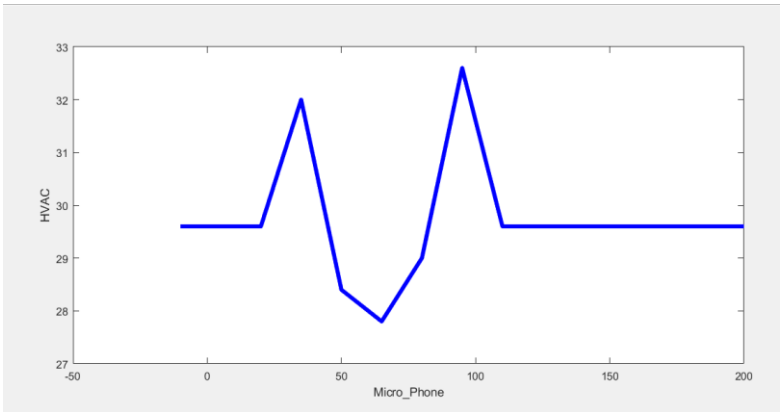


Figure 1.29

Surface Plot of Micro Phone and TV

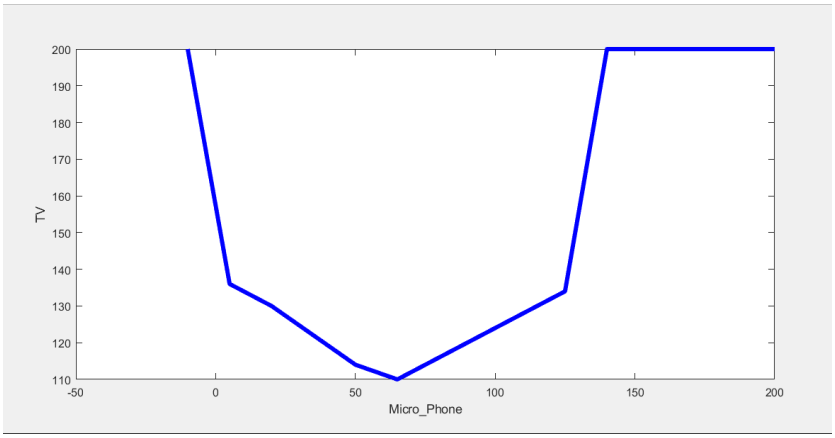


Figure 1.30

Surface Plot of Micro Phone and Light

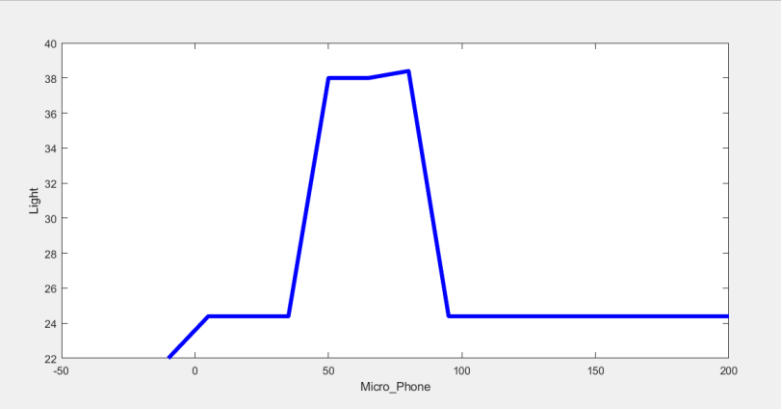
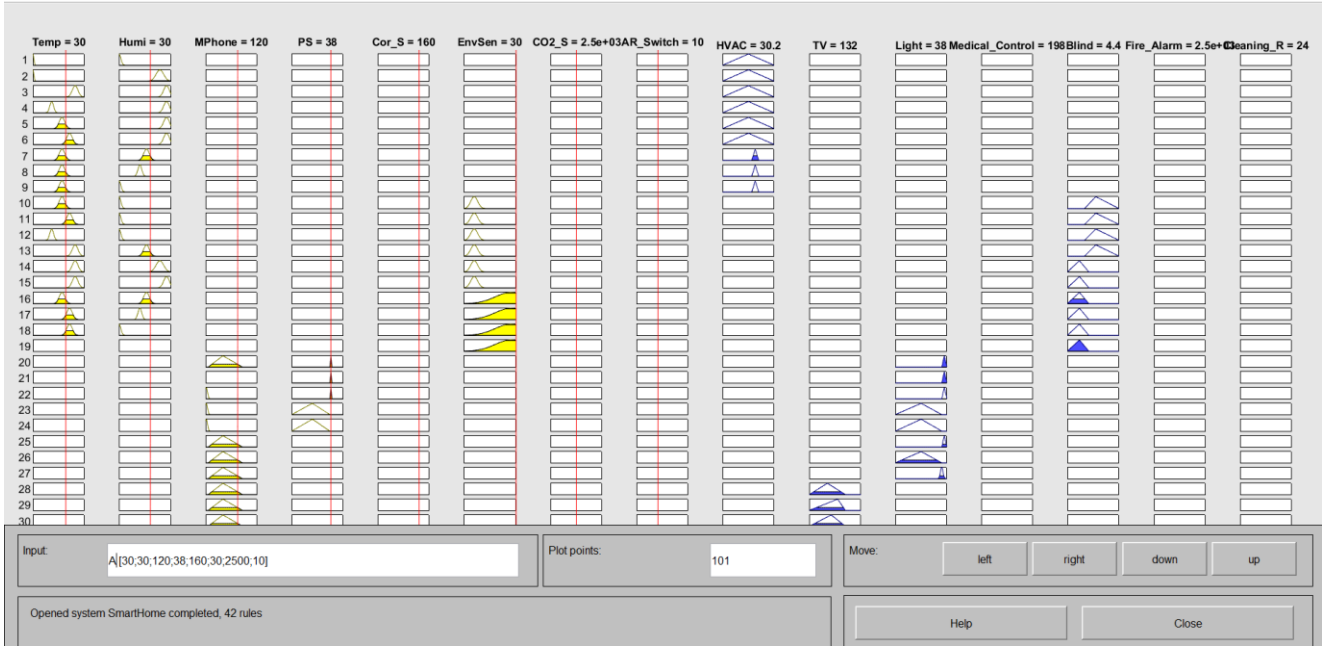


Figure 1.31

First Scenario Checking



In this project, we created 42 rules it can work according to different input scenarios lets check some of the scenarios by changing inputs parameters.

First Scenario

Inputs:

Temperature = 30 Celsius

Humidity = 30 gm/m3

Micro Phone = 120 disables

Presence Sensor = 38 Celsius

Corporal Sensor = 160 Per Minute

Environmental Sensor = 30

CO2 Sensor = 2500

Automatic Robot Switch = 10

Outputs:

HVAC = 30.2 Celsius

TV = 132

Light = 38 Celsius

Medical Control = 198 Per Min (heart rate)

Blind = 4.4

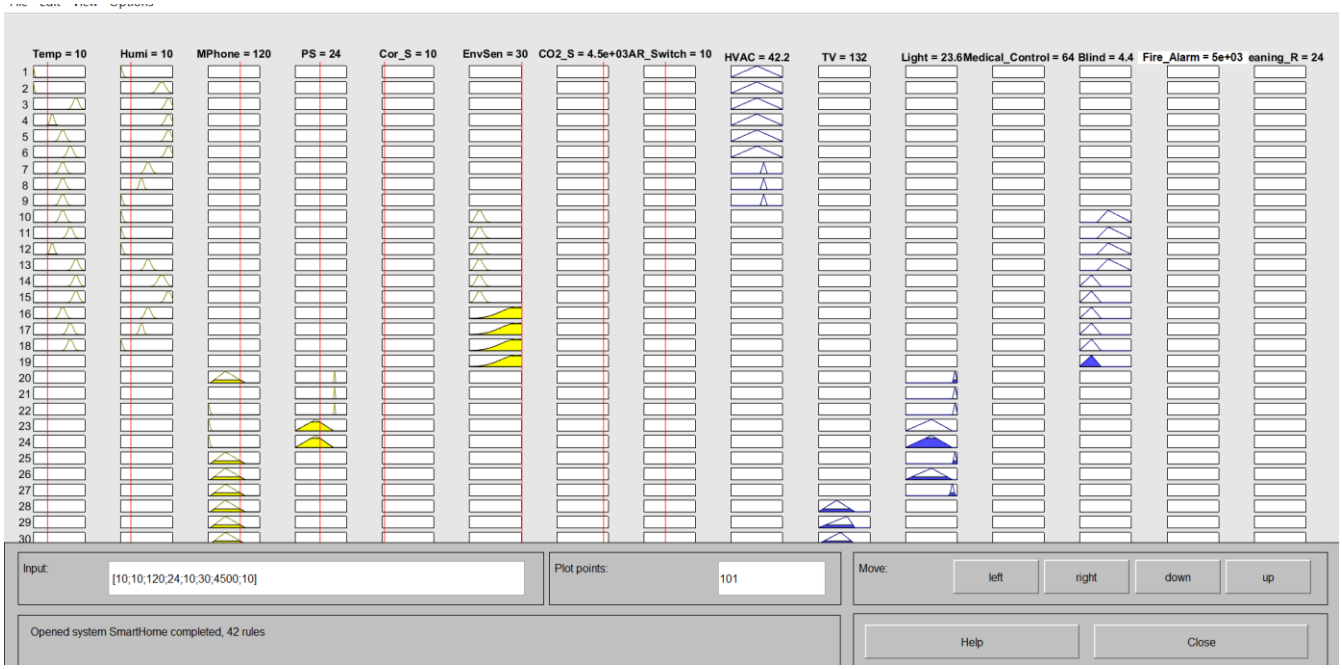
Fire Alarm = 2500

Cleaning Robot = 24

If we see our first scenario outputs are satisfied with our rules for example as our rules indicates if CO2 is 2500 which is a Warning sounds indication so our output should be approximately equal to that number so the fire alarm has 2500 which matches our rules, as we have seen in Output parameter of alarm figure 1.10. other input parameter presence sensor which helps to show human presence in that area according to it lights need to be function has seen output parameter light has 38 Celsius which clearly shows human presence in this area automatically lights will be ON. Our rule has in between 37 to 39 degrees Celsius shows human presence which satisfies given output.

Figure 1.32

Second Scenario Checking



Second Scenario

Inputs:

Temperature = 10 Celsius

Humidity = 10 gm/m3

Micro Phone = 120 disables

Presence Sensor = 24 Celsius

Corporal Sensor = 10 Per Minute

Environmental Sensor = 30

CO2 Sensor = 4500

Automatic Robot Switch = 10

Outputs:

HVAC = 42.2 Celsius

TV = 132

Light = 23.6 Celsius

Medical Control = 64 Per Min (heart rate)

Blind = 4.4

Fire Alarm = 5000

Cleaning Robot = 24

In the second scenario, I changed only some input parameters which were temperature, humidity, CO2, Corporal Sensor, and Presence sensor. According to there was the change in HVAC, Fire Alarm, Medical Alert and Lights in the smart home.

Part 2 – Optimize the FLC designed by Mamdani Interference

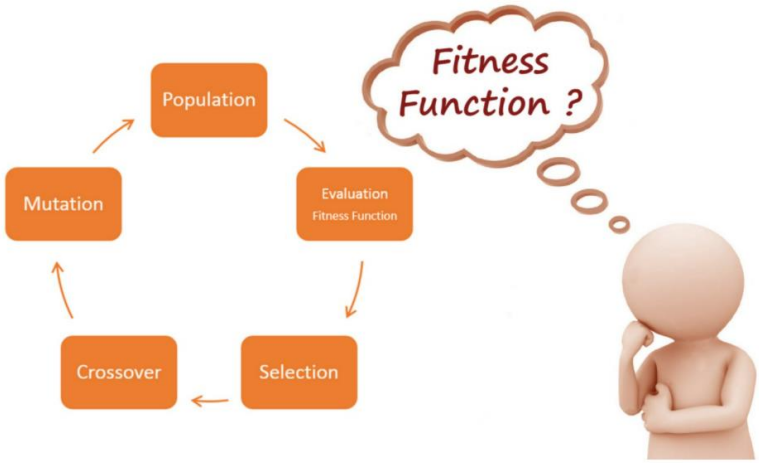
There are many algorithms to optimize the Fuzzy interference system but, in this project, we used a genetic algorithm which is an accurate model to evaluate the process step by step.

Genetic Algorithm

To evaluate fitness function, we need to follow the cyclic process which has mainly five steps starting from a population of chromosomes evaluated and finding fitness and making selection and do crossover and performing mutation.

Figure 2

Evolution of Fitness Function in Cyclic Process



Note: From Towards Data science, by Vijini Mallawaarachchi, 2017.

(https://towardsdatascience.com/how-to-define-a-fitness-function-in-a-genetic-algorithm-be572b9ea3b4)

STEP 1: Initial Population

Population means a set of chromosomes. The length of chromosomes is defined by calculating the parameters of our inputs we used two types of Membership Functions which are Gaussian and triangular methods.

Temperature has 10 Parameters

Humidity has 10 Parameters

Micro Phone has 5 Parameters

Presence Sensor has 6 Parameters

Coronal Sensor has 6 Parameters

Environmental Sensor has 6 Parameters

C02 sensor has 9 Parameters

Automatic Robot Switch has 6 Parameters

Total Chromosome Length is (10+10+5+6+6+6+9+6) = 58

Binary Form of Input Membership Functions.

Temperature									
VL		L		M		H		VH	
-15	-5	0	15	20	27	29	35	36	50
1111111111110000	1111111111110101	0000000000000	0000000000001111	0000000000000100	0000000000001011	0000000000001101	0000000000100011	0000000000100100	0000000000110010

Table1. Temperature Inputs in Binary Form.

Humidity									
VL		L		M		H		VH	
-5	-2	0	20	40	45	60	70	80	100
1111111111111011	1111111111111110	0000000000000	0000000000010100	0000000000101000	0000000000101101	0000000000111101	0000000001000110	0000000001010000	0000000001100100

Table2. Humidity Inputs in Binary Form.

CO2 Sensor								
Control			High			Very High		
600	800	1000	2000	3000	7000	10000	15000	20000
0000001001011000	0000001100100000	0000001111101000	0000011111010000	0000101110111000	0001101101011000	0001001110001000	0011101010011000	0100111000100000

Table3. C02 Sensor Inputs in Binary Form.

Microphone				
No Voice Command		Voice Command		
0	0	40	60	120
0000000000000000	0000000000000000	0000000001010000	0000000001111100	0000000001111000

Table4. Microphone Inputs in Binary Form.

Environmental Sensor			
Normal Wind		Fast Wind	
15	25	40	120
0000000000001111	0000000000001101	0000000000101000	0000000000110010

Table 5. Microphone Inputs in Binary Form.

Robot Switch					
On			Off		
10	11	6	9	8	12
0000000000001010	0000000000001011	0000000000000110	0000000000001001	0000000000001000	0000000000001100

Table 6. Robot Switch Inputs in Binary Form.

Coronal Sensor					
Slow Abnormal		Normal		Fast Abnormal	
50	55	70	75	110	115
0000000000110010	0000000000110111	0000000001000110	0000000001001011	0000000000110110	0000000000100011

Table 7. Coronal Sensor Inputs in Binary Form.

Presence Sensor					
No			Yes		
10	15	20	36	37	39
00000000000001010	00000000000001111	00000000000010100	00000000000100100	00000000000100101	00000000000100111

Table 8. Presence Sensor Inputs in Binary Form.

Binary Form of Output Membership Functions.

HVAC								
OFF			Turn On			ON		
10	15	5	10	15	20	25	30	35
11111111111111010	11111111111110000	11111111111111011	00000000000001010	00000000000001111	00000000000000100	00000000000101001	00000000000111110	00000000000100011

Table 9. HVAC Outputs in Binary Form.

Light								
Off			Dim Light			On		
0	20	30	35	36	37	35	36	37
00000000000000000	00000000000010100	00000000000011110	00000000000100011	00000000000100100	00000000000100101	00000000000100011	00000000000100100	00000000000100101

Table 10. Light Outputs in Binary Form.

Medication Control								
No Alert			Alert			Alert 2		
70	75	78	40	50	55	110	115	130
00000000001000110	00000000001001011	00000000001001110	00000000001010000	00000000001110010	00000000001110111	00000000001101110	00000000001110011	00000000010000010

Table 11. Medication Control outputs in Binary Form.

TV											
Volume Down			ON			Volume Up			OFF		
50	60	70	80	90	100	110	115	130	20	30	40
00000000000110010	00000000000111100	00000000001000110	00000000001010000	00000000001011010	00000000001100100	00000000001101110	00000000001110011	00000000010000010	00000000000010100	00000000000011110	00000000000101000

Table 12. TV outputs in Binary Form.

Blind					
Close			Open		
-10	-5	-2	0	10	15
1111111111111010	1111111111110101	1111111111111110	0000000000000000	0000000000001010	0000000000001111

Table 13. Blind outputs in Binary Form.

Cleaning Robot					
Start			Stop		
10	11	6	2	8	9
0000000000001010	0000000000001011	0000000000000110	0000000000000010	0000000000001000	0000000000001001

Table 14. Cleaning Robot outputs in Binary Form.

Fire Alarm								
Off			Warning Sounds			Continuous Warning		
600	800	1000	2000	3000	7000	10000	15000	20000
0000001001011000	0000001100100000	0000001111101000	0000111110100000	0001011101111000	0001101101011000	0001001110001000	0011101010011000	0100111000100000

Table 15. Fire Alarm outputs in Binary Form.

STEP2: Fitness Function

Fitness function is one of the crucial tasks we performed in this project which calculates the fitness of the chromosomes. To calculate the fitness there were many platforms present but the global optimum toolbox is easy and gives accurate results.

STEP3: Selection Process

in this step, our main objective is selecting two chromosomes from the total population for mating. So that we can get offspring with higher fitness values. There are many methods present for the selection of pair of chromosomes which are Roulette Wheel, Boltzmann selection, Tournament Selection, Uniform, and Stochastic Uniform method. But for this project, we are selecting Roulette Wheel Selection which is best among all due to its effectiveness.

STEP 4: Crossover

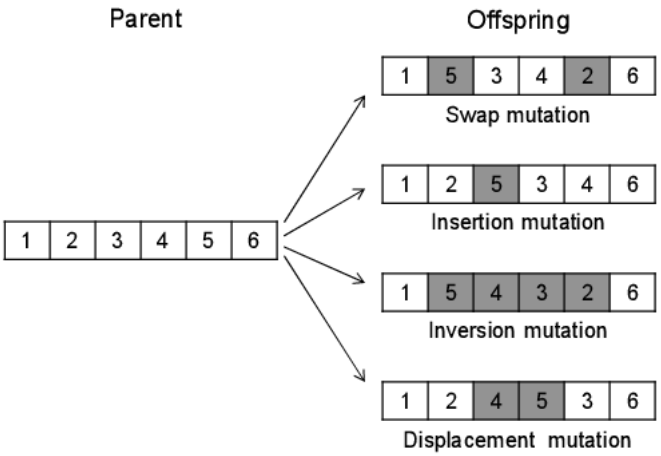
Cross-over was performed between two chromosomes by selecting genes randomly. In global optimal default value is 0.8 which can be changed according to usage and fitness function.

STEP 5: Mutation

The mutation is the most significant process in the Genetic algorithm, mutation was applied randomly for randomly selected offspring and calculates the fitness and compare with previous fitness values. This process happens in cyclic until to optimize the fitness values.

Figure 2.1

Different Types of Mutations.



Note: From Research gate.

(https://www.researchgate.net/figure/Mutation-operators-applied-to-chromosomes-in-the-proposed-genetic-algorithm_fig8_272093243)

Sugeno Fuzzy Method

In this project, we showed Mamdani interference method, there is another different method present to solve fuzzy logic which was Sugeno. In Sugeno, if a = x and b = y, then output is c = f(a,b), whereas the Mamdani method output of each rule again a fuzzy logic set.

For instance, in case our project, inputs are Temperature, Humidity, Micro Phone, Presence sensor, Corporal sensors, Environmental sensor, CO2 sensor, and Automatic Robot switches. When applying the Sugeno method, the outcome of fuzzy logic is a function of eight inputs. In Mamdani, the output function is discrete whereas in Sugeno it is continuous. Sugeno does not have output rules as Mamdani has so it's less interpretable than Mamdani.

Part 3 – Compare different optimization techniques on CEC'2005 functions

There are number of algorithms used for optimization that have been developed to solve real world problems relating to optimization. there are many methods like particle swarm optimization, evolution strategies, simulated annealing, QN method and others. Every category has its own set of rules and operators or parameters that its working on. There has been discussion regarding how some issues are well suited for particular algorithms but not for others, as well as the need of testing them under various settings, initializing processes, size, and scalability. Several benchmark functions were

introduced in CEC''05, a special session on the application of optimization methods to real-world issues. There were 25 benchmark functions introduced, with five Unimodal and twenty-five Multimodal functions.

We've chosen two functions from the document and decided to run the test on two algorithms, namely

Particle Swarm Optimization

It's a technique which is bio-inspired algorithm that searches for the best solution in the solution space and It differs from other techniques in that it simply requires the objective function and is unaffected by the gradient or any differential form of the objective and has fewer hyperparameter.

Genetic Algorithm

It's a query optimization technique derived from natural selection process' s. The primary idea behind this algorithm is to emulate the 'survival of the fittest' notion; it mimics the processes that take place in biological ecosystems, where the powerful adjust and thrive whereas the weakest dies.

CE'2005 Function that we were to Two functions from CE'05 were chosen to evaluate the performance of the algorithms discussed above:

F7: Griewank's Shifted Rotated Function Without Bounds

$$F_7(\mathbf{x}) = \sum_{i=1}^{D} \frac{z_i^2}{4000} - \prod_{i=1}^{D} \cos\left(\frac{z_i}{\sqrt{i}}\right) + 1 + f\_bias_7 \ , \ \mathbf{z} = (\mathbf{x} - \mathbf{o}) * \mathbf{M} \ , \ \mathbf{x} = [x_1, x_2, ..., x_D]$$

D: dimensions

$\mathbf{o} = [o_1, o_2, ..., o_D]$: the shifted global optimum

M': linear transformation matrix, condition number=3

M =**M'**(1+0.3|N(0,1)|)

F14: Shifted Rotated Expanded Scaffer's F6

$$F(x, y) = 0.5 + \frac{(\sin^2(\sqrt{x^2 + y^2}) - 0.5)}{(1 + 0.001(x^2 + y^2))^2}$$

Expanded to

$$F_{14}(\mathbf{x}) = EF(z_1, z_2, ..., z_D) = F(z_1, z_2) + F(z_2, z_3) + ... + F(z_{D-1}, z_D) + F(z_D, z_1) + f\_bias_{14},$$

$$\mathbf{z} = (\mathbf{x} - \mathbf{o}) * \mathbf{M}, \mathbf{x} = [x_1, x_2, ..., x_D]$$

D: dimensions

$\mathbf{o} = [o_1, o_2, ..., o_D]$: the shifted global optimum

M: linear transformation matrix, condition number=3

Testing Function 7 on PSO & GA for Dimension = 2

Genetic Algorithm for Dimension 10

Iteration	Minima for each Iteration
1	-179.8523
2	-179.9941
3	-179.9557
4	-179.5914
5	-179.7234
6	-179.4687
7	-179.7532
8	-179.9834
9	-179.4391
10	-179.5917
11	-179.5934
12	-179.8252
13	-179.5971
14	-179.4162
15	-179.5489

Std Dev	0.2112
Max	-179.1393
Min	-179.7563
Mean	-179.5414

Testing Function 7 on GA for Dimension = 2

Iteration	Minima for each Iteration
1	-179.9398
2	-179.9777
3	-179.9884
4	-179.9624
5	-179.9467
6	-179.9997
7	-179.9452
8	-179.9854
9	-179.9842
10	-179.8563
11	-179.9523
12	-179.9620
13	-179.9685
14	-179.9594
15	-179.9400

Std Dev	0.0167
Max	-179.9398
Min	-179.9997
Mean	-179.964

Testing Function 7 on PSO Dimension = 2 & Dimension = 10

Particle Swarm Optimization for D = 2

Iteration	**Minima for each Iteration**
1	713.9777
2	392.2797
3	417.5872
4	518.9847
5	449.1822
6	437.1947
7	397.2002
8	395.1998
9	393.7942
10	724.9162
11	703.868
12	682.1971
13	729.6906
14	391.4477
15	729.3453

Std Dev	327.20
Max	724.91
Min	391.44
Mean	512.06

Testing Function 7 on PSO for Dimension = 10

Std Dev	0.000679
Max	-169.8210
Min	-169.8220
Mean	-169.8215

Testing Function 14 for GA and PSO for Dimension 2 and 10

Genetic Algorithm for Dimension 2.

Std Dev	**0.0184333801**
Median	**-289.89631368**

Genetic Algorithm for Dimension 10.

Std Dev	**0.424796890**
Median	**-295.7689065**

Testing Function 14 for PSO Dimension 2.

Std Dev	**0.0277464590**
Median	**-289.89777098**

Testing Function 14 for PSO Dimension 10.

Std Dev	**0.046738976**
Median	**-295.39985789**

Conclusion:

We were able to view the big picture thanks to the functions we utilised to analyse both optimization strategies. We tested both GA and PSO for function 7, and as shown in the results, PSO performed better for both dimension 2 and 10 for function 7, whereas Genetic Algorithm took the lead when we tested them using function 14 because it performed better with dimension 2 than Particle Swarm Optimization, but PSO performed better when the dimensions were changed to 10. Both optimization strategies performed admirably in the end.

References

MathWorks. *Gaussmf.* *https://uk.mathworks.com/help/fuzzy/gaussmf.html*.

Chun Liu. (2014). *Multi-Robot Task Allocation for Inspection Problems with Cooperative Tasks Using Hybrid Genetic Algorithms.* *https://www.researchgate.net/figure/Mutation-operators-applied-to-chromosomes-in-the-proposed-genetic-algorithm_fig8_272093243*.

Vijini Mallawaarachchi. *(2017). How to define a Fitness Function in a Genetic Algorithm.*

Towards Data Science.https://towardsdatascience.com/how-to-define-a-fitness-function-in-a-genetic-algorithm-be572b9ea3b4.