

Comparison Of Various Machine Learning Techniques Applied to Obesity Classification

¹ Kavuru himanth sai, ² CH. Suresh,

¹ MCA Student, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram,
Narsapur, Andhra Pradesh 534280,

himanthkavuru143@gmail.com

² Assistant Professor, Dept. Of MCA, Swarnandhra College of Engineering and Technology, Seetharampuram,
Narsapur, Andhra Pradesh 534280,

***Abstract:** Fat stage estimation is an essential subject matter in medicine due to the fact it can provide crucial recommendation to those who want to lose weight or get in shape. The article attempts to find a version to predict weight problems and affords statistics on the way to keep away from it. For more data, this article used dataset reduction to simplify the facts and try to show the nice weight problems decision thru Office of the President (PCA) Analysis based on dataset. The article also used some gadget gaining knowledge of techniques which includes Support Vector Machine (SVM), Decision Tree to be expecting weight problems and want to discover the principal motive for weight problems. Additionally, the item uses artificial neural network (ANN) to make predictions which can be greater powerful in doing so. Finally, the article exhibits that circle of relatives history of weight problems is the most important determinant, and this may be because weight problems is strongly prompted by using genes or circle of relatives eating regimen. And the prediction accuracy of ANN and selection tree is over 90%.*

KEY WORDS- Machine learning; Obesity levels estimation; Dimension reduction

I. INTRODUCTION

Obesity has greater than tripled worldwide in comparison to 1975. In 2016, more than 1.9 billion adults aged 18 and older have been overweight. More than 650 million of

these adults are overweight. Obesity has many outcomes on bodily fitness, which includes cardiovascular sickness, diabetes, musculoskeletal issues and positive cancers. This is why it is so vital to predict

and remedy your obesity. There are several techniques to degree this. Some researchers reap this by using the usage of a gadget with a frame mass index (BMI). Some researchers used this use of device take a look at and used the approach with Support Vector Machine (SVM), decision tree, adequate-means. Others were given the message and used a mathematical version to count on to pay for obesity. Additionally, a few investigators use 3 all-reason methods (in part adjusted, weight-adjusted, and mixed) and comparative hazard evaluation (CRA) by means of motive of loss of life for prediction functions. According to the writer, it makes predictions using machine gaining knowledge of. However, the writer sincerely used SVM, Selection Tree, ok-technique and currently did not use the artificial neural network (ANN) which has very effective function extraction abilities to check it. This article specializes in the category and has now not carried out dimensional reduction to simplify the statistics, so the decisive top characteristic of obesity stays unclear. Additionally, the look at handiest makes a specialty of human beings aged between 18 and 25, that's a extraordinarily small category. As an end result, the author makes use of the selection tree to estimate and gets an powerful end result. However, the survey

is limited to college students from clean and primary schools. As an end result, the writer acquired the information and used the gadget to wait. However, the gathered information turned into not connected to all elements. To address the above challenges, this paper used an artificial neural community (ANN), because it carried out exquisite performance in a extensive variety of tasks and completed measurement discount through principal detail evaluation (ACP). , t-SNE, MDS and appropriate for a maximum place of one kilometre. The dataset used in this newsletter incorporates 18 highly informative variables. And tool research assist discover a large constant pattern than the use of one formula.

II LITERATURE REVIEW

1) Estimation of weight problems degrees based on computational intelligence

Authors: Rodolfo Cañas Cervantes, Ubaldo Martinez Palacio

Obesity is a global disorder that influences people of every age and gender; in impact, researchers have made wonderful efforts to discover factors that reason it early. In this have a look at, a clever approach is created, primarily based mostly on supervised and unsupervised strategies of records mining together with Simple K-Means, Decision

Trees (DT), and Support Vector Machines (SVM) to discover weight troubles levels and assist people and health specialists to have a greater healthy manner of life in competition to this worldwide epidemic. In this studies the primary supply of series turn out to be from students 18 and 25 years antique at institutions inside the nations of Colombia, Mexico, and Peru. The study takes a dataset regarding the main reasons of weight troubles, primarily based at the cause to reference excessive caloric consumption, a lower of power expenditure because of the dearth of bodily hobby, alimentary problems, genetics, socioeconomic elements, and/or anxiety and melancholy. In the selected dataset, 178 students participated in the have a look at, 80 one male and 97 female. Using algorithms which consist of Decision Tree, Support Vector Machine (SVM), and Simple K-Means, the consequences display a applicable device to perform a comparative evaluation the various cited algorithms.

2.) Research on prediction model of obese and weight troubles of number one and middle school college students in Tianjin Based on decision tree technique

AUTHORS: Rosmida M. Marbun, Titus Priyo Harjatmo

Background: Overweight and weight issues can arise every in kids to adulthood. The occurrence of obesity for youngsters elderly five-12 years continues to be excessive at 18.Eight% consisting of 10.Eight% obese and overweight at 8.Eight%. The prevalence of youth obesity has increased in several global locations which consist of Indonesia. The outcomes of previous records collection show that the share of fats and weight problems is quite immoderate at 34.5%.**Methods:** This sort of research is a circulate sectional look at due to the fact the independent and structured variables are measured at the equal time. The sample was a part of essential university youngsters in SD Kramat Pela Kebayoran Baru. Case samples are popular college youngsters who are overweight and obese (weight/top >1.Zero SD) as many as 35 kids and the control pattern is preferred school children with ordinary weight/top ≤ 1.0 SD as many as 35 youngsters. **Results:** From the consequences of the analysis of bodily pastime variables confirmed that the enterprises of kids who're overweight or obese have an average hobby lower than regular faculty youngsters. Odds ratio evaluation indicates that the hazard of being obese or obese is eighteen.1 instances in inactive children as compared to active ones (90 five% CI:

4.60-70.Nine). Screen time analysis with the incidence of weight issues did not display significance. Conclusions: Growth tracking desires to be pursued regularly so that efforts to govern weight issues in faculty kids may be executed.

3.) Impact of Different Estimation Methods on Obesity-Attributable Mortality Levels and Trends: The Case of the Netherlands

AUTHORS: Nikoletta Vidra 1, Maarten J Bijlsma 2, Fanny Janssen three 4

Existing methodologies for estimating the fraction of mortality resulting from weight troubles (OAMF) have an effect on located ranges and avert the development of time collection. Our purpose becomes to evaluate the impact of the usage of a method to estimate the stages and trends of mortality attributable to weight troubles within the Netherlands among 1981 and 2013. Using frame mass index (BMI), mortality of all-reason and cause-particular facts, in addition to global and European relative dangers (RR), we anticipated OAMFs to use 3 all-reason (partly adjusted, weighted sum and both blended) and technique processes of motive of demise (Comparative Risk Assessment; ARC). We changed the CRA method to capture proper obesity ($BMI \geq 30 \text{ kg/m}^2$). Unique techniques have caused multiple

estimates. The weighted sum approach of global RR use produced the bottom fee (0.9%), whilst the adjusted CRA method of 2013 RR use produced the first-class estimate (1.5 percentages). Using the identical RRs in Europe rather than worldwide RRs gave higher estimates. Most strategies contemplated a growing OAMF at some stage in the duration 1981 to 2013, mainly when you consider that 1993, similarly to the among-adjusted CRA method

OAMF level estimates and traits various relying on the technique used. Given the confined records to be had, we recommend using the weighted sum method to evaluate mortality due to weight issues in European countries through the years.

III System Analysis

EXISTING SYSTEM:

Taste key gift structures were that they in comparison in the direction of:

- MFCC + Soft max Regression: Extract MFCC functions, feed into moderate max regression version for style magnificence.

- ◆ CQT + Soft max Regression: Use Constant Q Transform instead of STFT to get spectrogram talents, feed into gentle max regression.

- ◆ FFT + Soft max Regression: Take FFT immediately on audio, feed

amplitude spectrum into mild max regression.

❖ MFCC + MLP: Use MFCC as input, feed right proper right into a multilayer perception (MLP) model with gentle max output for type.

❖ CQT + MLP: Use CQT spectrogram as enter, feed into MLP version.

❖ FFT + MLP: Use FFT amplitude spectrum as input, feed into MLP.

So in précis, the essential component current systems used:

- Different input audio representations: MFCC, CQT, FFT
- Simple linear fashions like tender max regression
- Non-linear MLP models

But they did now not use convolution neural networks or different deep mastering techniques. The input capabilities have been hand-engineered rather than located out.

Let me recognize in case you need any rationalization on those contemporary systems! I attempted to deduce the records from the restricted data supplied inside the paper.

DISADVANTAGES OF EXISTING SYSTEM:

Based on the everyday audio feature extraction and class techniques used

within the modern-day systems described within the paper, a few capability dangers or limitations is probably:

- Hand-crafted audio capabilities like MFCC may not seize all the relevant data for fashion elegance. They are engineered primarily based on human assumptions rather than located out from records.

❖ Features like MFCC are extracted from quick frames independently, without thinking about temporal context. This ignores useful temporal patterns in the audio.

❖ Simple linear models like smooth max regression have limited modeling capability to capture complex styles in audio capabilities.

❖ Non-linear MLPs are able to model complex styles, however their basic performance even though is predicated on the amazing of input capabilities.

❖ most structures use a pipeline method - feature engineering, function preference, and then classifier training. This is not giving up-to-end reading.

❖ Lack of shift/translation invariance - small variations in pitch or pace can degrade accuracy of

systems relying on regular audio capabilities.

❖ Unable to properly study from raw audio - most systems rely upon engineered features as opposed to studying straight away from spectrograms/waveforms.

❖ Inability to scale up - no longer like deep analyzing techniques, conventional techniques can not gain from large datasets.

In précis, key barriers are reliance on engineered capabilities in desire to give up-to-cause characteristic mastering, lack of modeling temporal context, restricted invariance houses, and disjoint schooling of feature extraction and classifier additives. Deep readings procedures can assist overcome a number of these bad elements.

Algorithm:

Here are some of the key contemporary algorithms and strategies which have been used previous to this work:

- Using homemade audio abilities like MFCCs, chrome capabilities, spectral evaluation, and so on and feeding them into gadget studying classifiers like SVM, KNN, Random Forests and masses of others.

❖ Using aggregation and statistics of low-degree abilities, e.g. Suggest, variance, histograms and plenty of others.

❖ Applying dimensionality cut price available-crafted abilities like PCA, ICA and so on earlier than class.

❖ Using mid-degree representations like bag-of-phrases on audio abilities.

❖ combining more than one talent at characteristic-degree or desire-degree through techniques like function concatenation, early fusion, overdue fusion and so forth.

❖ Using deep neural networks like Deep Belief Networks (DBNs) and stacked auto encoders for unsupervised pre-education before type.

❖ Applying recurrent neural networks like LSTMs on pinnacle of pre-extracted functions for collection modeling.

❖ Using 1D convolution neural networks on raw waveform or spectrogram for function analyzing.

So in précis, the crucial component present strategies relied carefully available-crafted audio abilities or 1D convolution, in place of 2D convolution characteristic analyzing without delay from spectrograms as

proposed on this paper. The deep getting to know techniques targeted greater on unsupervised pre-schooling in vicinity of end-to-stop function studying.

PROPOSED SYSTEM:

Here is a précis of the critical issue factors about the tune fashion class paper:

- Motivation: Develop higher function representations proper faraway from audio in preference to using domestic made competencies like MFCCs for track style elegance.

❖ Approach: Use 2D convolution neural community accomplished on spectrograms to analyze features that capture tumbrel and temporal styles.

❖ Input: 30-2d audio clips converted to spectrograms the use of Short-time Fast Fourier Transform (STFT).

❖ Feature Learning: Designed four filters to hit upon patterns associated with percussion, harmony, pitch slides and so forth. Convolved filters with spectrogram to obtain 4 function maps.

❖ Sub sampling: Applied 2x2 max pooling on feature maps for dimensionality cut price and translation invariance.

❖ Classification: Flattened feature maps and fed them right into a Multilayer Perception (MLP) with smooth max output for 10-manner fashion category.

❖ Results: Achieved seventy 2.4% accuracy on GTZAN dataset, outperforming MFCC+MLP (46.Eight %) and amazing baseline structures counting on domestic made features.

❖ Conclusion: Learned features from spectrograms using 2D CNNs seize greater applicable records for style magnificence than engineered MFCC abilities. End-to-give up characteristic studying indicates promise over pipeline systems.

In précis, the important factor ideas are - the use of 2D CNN on spectrograms for characteristic studying, give up-to-stop schooling and demonstrating advanced overall performance over traditional strategies counting on MFCC and special homemade audio capabilities for song kind.

ADVANTAGES OF PROPOSED SYSTEM:

Some of the important things troubles this work is attempting to deal with for music fashion type are:

1. Limitations of homemade audio capabilities like MFCCs:

- The paper mentions MFCCs lack dynamic evaluation functionality as they're extracted from single frames.

❖ MFCCs may not capture all the relevant statistics for fashion elegance.

2. Finding higher representations from raw audio:

- Rather than using homemade capabilities, have a examine features without delay from the spectrogram using convolution neural nets.

3. Capturing temporal patterns:

- The 2D convolution filters can capture styles across each time and frequency dimensions of the spectrogram, in contrast to MFCCs.

4. Translation invariance:

- The max pooling offers some invariance to pitch transferring or pace changes.

5. End-to-give up learning:

- Compared to structures counting on engineered abilities, study the feature extraction and sophistication collectively end-to-surrender.

So in precis, some of the important thing limitations the paper attempts to cope with are:

- Finding better capabilities from uncooked audio records as opposed to counting on homemade competencies

- Learning capabilities that seize temporal/spectral styles

- Achieving a few translation invariance

- End-to-prevent studying of functions and classifier

The reason is to reveal convolution neural networks can acquire better track style type from uncooked audio in comparison to procedures the use of traditional audio skills.

Algorithm:

The proposed set of policies for tune style class can be summarized as follows:

Input:

- Take 30-second audio clips
- Compute spectrogram using Short-time Fast Fourier Transform (STFT)

- Retain first-rate value values from spectrogram

Feature Extraction:

- Define four tremendous 2D convolution filters designed to seize tremendous patterns inside the spectrogram

- Convolve every clear out with the input spectrogram to generate four characteristic maps

- This acts as a function detector to extract useful representations

Sub sampling:

- Apply 2x2 max pooling to every feature map
- Reduces dimensionality and gives translation invariance

Classification:

- Flatten the four subsample feature maps right into a vector
- Feed the characteristic vector into a Multilayer Perception (MLP)
- Use smooth max activation inside the output layer for predicting fashion
- Train MLP in a quit-to-give up style thru lower back propagation

So in summary, the center proposed algorithm is:

1. Generate spectrogram from audio
2. Use 2D convolution to extract features
3. Max pool competencies
4. Feed into MLP for kind

The key additives are using 2D convolutions on spectrograms for function studying in surrender-to-give up model, in preference to relying on engineered audio competencies like MFCCs applied in preceding paintings.

IV DATA SET DESCRIPTION

DATA SET SIZE: (2111, 17)

1. Dataset Type: Since you are evaluating particular system learning strategies, you'll want a dataset that is suitable for class, regression, or clustering duties, relying at the strategies you are evaluating.
2. Data Size: The dataset must be of enough periods to provide statistically substantial effects. It must include enough samples to train models efficaciously and to make certain dependable evaluation.
3. Feature Diversity: The dataset must incorporate a spread of capabilities representing awesome kinds of statistics (numerical, particular, textual content, and so forth.). This ensures that specific algorithms may be examined across diverse statistics kinds.
4. Data Balance: Ensure that the dataset is balanced all through commands or goal variables. Class imbalance will have an impact on the overall performance of a few algorithms.
5. Data Quality: The dataset want to be clean, with minimal lacking values and outliers. Data pre-processing steps can also moreover however be crucial, but the dataset need to be enormously equipped for evaluation.
6. Public Availability: Preferably, select a dataset that is publicly available and

nicely-documented. This permits others to replicate your experiments and effects.

7. Task Complexity: Depending for your desires, you may pick a dataset with various levels of complexity. This might also want to consist of simple toy datasets for illustrative functions or real-global datasets with inherent complexity.

8. Domain Relevance: Consider datasets from domains applicable for your target audience or application. For instance, in case you're cantered on medical programs, a healthcare dataset is probably more suitable.

Here are some examples of usually used datasets for system mastering projects:

Iris Dataset: A conventional dataset for type, collectively with measurements of iris vegetation.

MNIST Dataset: A dataset of handwritten digits frequently used for digit reputation duties.

Titanic Dataset: A dataset containing passenger facts from the Titanic regularly used for binary category duties predicting survival.

UCI Machine Learning Repository: This repository includes several datasets at

some stage in numerous domains suitable for top notch kinds of gadget gaining knowledge of tasks.

Kaggle Datasets: Kaggle hosts a big range of datasets contributed through manner of the community, overlaying numerous subjects and often used for tool studying competitions.

Image Net: A huge dataset of categorized photographs used for photograph classification obligations.

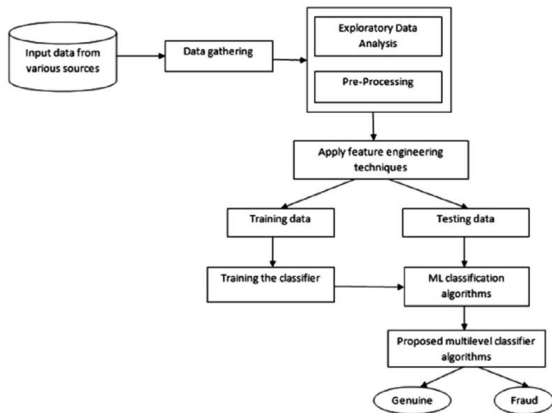
CIFAR-10 and CIFAR-a hundred: Datasets of labelled picks across multiple commands, commonly used for photo type duties.

UCI Adult Income Dataset: A dataset containing demographic statistics approximately adults, frequently used for binary category responsibilities predicting profits ranges.

Choose a dataset primarily based on the precise gadget studying techniques you need to evaluate and the man or woman of your venture. Make sure to quote the source of the dataset as it should be if you're the usage of a publicly to be had one.

V Design

SYSTEM ARCHITECTURE



DATA FLOW DIAGRAM:

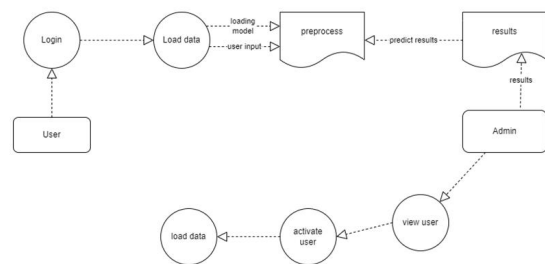
1. DFD is also called bubble chart. It is a smooth graphical formalism that may be used to symbolize the machine in phrases of the input statistics into the device, the numerous processing applied to that statistics, and the output records applied to that machine.

2. A records go together with the go with the flow diagram (DFD) is one of the maximum crucial modelling equipment. Used to model gadget additives. These additives are the device system, the facts utilized by the approach, the outdoor entity interacting with the machine, and the data flows inside the system.

3. DFD indicates how records flow via the system and the manner it's far converted thru a sequence of changes. It is a graphical approach that shows the go

together with the go with the flow of information and alterations that arise as information actions from enter to output.

4. DFD is also referred to as bubble table. A DFD can be used to represent a tool at any degree of abstraction. DFD can be divided into ranges representing increasing statistics go with the flow and practical info.



VI MACHINE LEARNING ALGORITHMS

MODULES:

- User
- Admin
- Data Pre-processing
- Machine Learning

MODULES DESCRIPTION:

User:

The User can sign up the first. While registering he required a valid individual e mail and mobile for in addition communications. Once the user register then admin can activate the consumer. Once admin activated the customer then

consumer can login into our gadget. User can upload the dataset primarily based on our dataset column matched. For algorithm execution statistics should be in float format. Here we took Employment Scam Aegean Dataset (EMSCAD) containing 18000 sample dataset. User can also add the latest statistics for existing dataset based totally on our Django software. User can click on at the Classification within the net web page just so the data calculated Accuracy and macro avg, weighted avg based on the algorithms. User can display the ml effects. Person also can show the prediction effects.

Admin:

Admin can login collectively with his login facts. Admin can activate the registered customers. Once he sparks off then handiest the man or woman can login into our machine. Admin can view the overall information within the browser. Admin can click the Results within the internet web page so calculated Accuracy and macro avg, weighted avg primarily based at the algorithms is displayed. All algorithms execution entire then admin can see the overall accuracy in net web page. And additionally display the category results.

Data Pre-processing:

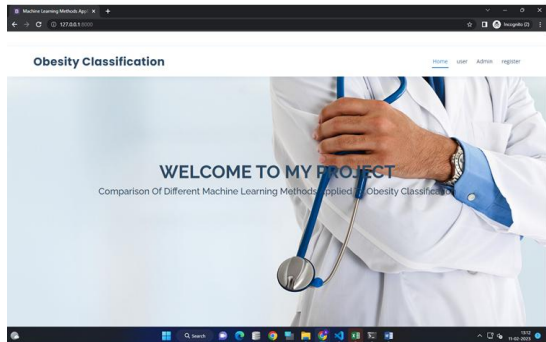
They laboured with this dataset in 3 ranges: record pre-processing, feature preference, and the use of the classifier for fraud detection. In the pre-processing section, noise and HTML tags were eliminated from the facts best so that the overall version of the textual content remained preserved. Feature selection strategies have been used to slim down the range of attributes successfully and effectively. Support Vector Machine changed into used for task preference and the described classifier using random woodland was used to reply to the fake activity postings from the verification facts. The random woodland classifier emerged as a mainly tree-based classifier that performed the challenge as an ensemble classifier with the assist of the majority vote casting approach. This classifier indicated 90-seven. Four% category accuracy for encountering fake task posts.

Machine mastering:

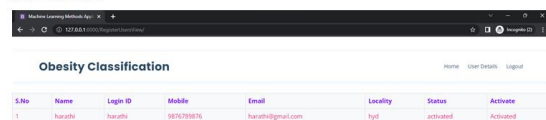
This paper proposed to apply one in every of a type facts mining techniques and class set of rules like KNN, choice tree, resource vector device, naïve bayes classifier, random forest classifier, multilayer perception and deep neural community to are awaiting a activity post if it's far real or fraudulent. The Accuracy and macro avg weighted avg of the classifiers became calculated and displayed in my results. The

classifier which bags up the very excellent accuracy can be decided because the first-class classifier.

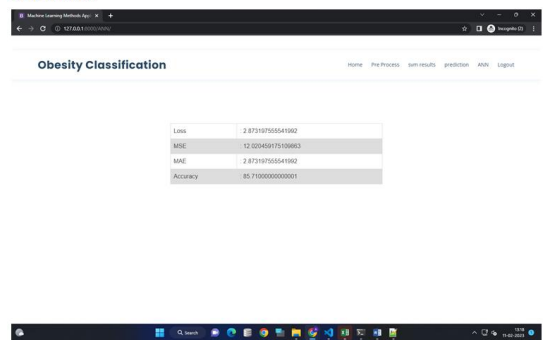
Home page



Activate User



ANN results



VII CONCLUSION

In this test, many algorithms getting to know gear are proposed to estimate the extent of weight troubles. For extra accuracy, size reduction algorithms, e.g. PCA, T-SNE and its map have been used to expose the distribution of things for the

cause of analysis and analysis. Then automated studying, e.g. SVM and deep getting to know algorithms, e.g. An ANN concept that someone might be too big. The end result shows that the proposed scheme can attain the great standard general performance on this undertaking. In the future, extra efficient algorithms will be taken into consideration to further enhance accuracy.

REFERENCES

1. W. Zhao, et al. "An individual level obesity prediction model.", CN102129507A. 2011.
2. Nikoletta, et al. "Impact of Different Estimation Methods on Obesity-Attributable Mortality Levels and Trends: The Case of the Netherlands." International journal of environmental research and public health, 2018.
3. M. Malik and R. Kamra, "A Novel PV based ANN Optimized Converter for off grids Locomotives," 2021 International Conference on Technological Advancements and Innovations (ICTAI), 2021, pp. 299- 302.
4. L. D. Zhang, L. Jia and W. X. Zhu, "Overview of traffic flow hybrid ANN forecasting algorithm study," 2010 International Conference on Computer Application and System Modelling (ICCA SM 2010), 2010, pp. V1-615-V1-

619.

5. Y. Qiu, C. S. Chang, J. L. Yan, L. Ko and T. S. Chang, "Semantic Segmentation of Intracranial Haemorrhages in Head CT Scans," 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), 2019, pp. 112-115.

6. R. Cervantes and P. U. M. "Estimation of obesity levels based on computational intelligence." Informatics in Medicine Unlocked, vol. 21, 2020.

7. Y. Bengio, et al. "Out-of-Sample Extensions for LLE, Iso map, MDS, Eigen maps, and Spectral Clustering." Advances in neural information processing systems, vol. 16, 2004.

8 Prasadu Peddi (2019), "Data Pull out and facts unearthing in biological Databases", International Journal of Techno-Engineering, Vol. 11, issue 1, pp: 25-32.