ISSN No:-2456-2165

UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

Methods and Applications of Unsupervised Learning Machines

Dipartimento Di
Ingegneria
Corso Di Laurea in
Ingegneria Gestionale
Classe n. L-9

Candidato:
Luca Daniele
Vailati
Matricola n.
1073650

Relatore:
Chiar.ma Prof.ssa Francesca
Maggioni
Anno Accademico

ISSN No:-2456-2165

ABSTRACT

This article offers a comprehensive introduction to the key concepts of unsupervised machine learning, a type of machine learning where models are trained using data that has not been labeled or categorized. The primary goal of unsupervised learning is to identify hidden patterns or structures within the data without the need for predefined labels. It explores various unsupervised learning techniques, such as clustering, dimensionality reduction, and anomaly detection, highlighting their potential applications in fields like data mining, pattern recognition, and market segmentation. In addition to the theoretical framework, the article delves into a practical example by focusing on one of the most commonly used unsupervised learning algorithms: k-means clustering. The k-means algorithm is a popular method for partitioning data into distinct groups (clusters) based on similarities. It is especially useful for discovering patterns in large datasets, where the objective is to assign data points to clusters based on their proximity to a centroids. The article further demonstrates the implementation of this algorithm within the GAMS (General Algebraic Modeling System) environment, showing how it can be used to perform clustering tasks and interpret the results within a computational framework designed for optimization and mathematical modeling. This hands-on example serves as a practical guide for readers looking to apply unsupervised machine learning techniques to real-world problems.

ISSN No:-2456-2165

TABLE OF CONTENT

ISSN No:-2456-2165

CHAPTER ONE
INTRODUCTION TO UNSUPERVISED MACHINE LEARNING

➢ *What is Unsupervised Learning?*

Unsupervised learning, alternatively referred to as unsupervised machine learning, employs machine learning algorithms to examine and categorize datasets that lack labels. These algorithms unveil concealed patterns or groupings within the data, all without any human involvement. The capacity of unsupervised learning to identify similarities and disparities in information renders it an optimal solution for exploratory data analysis, cross-selling tactics, customer segmentation, and Fig recognition.

➢ *Unsupervised vs. Supervised vs. Semi-Supervised Learning*

Unsupervised learning and supervised learning are often discussed in conjunction with one another. Unlike unsupervised learning algorithms, supervised learning algorithms utilize labeled

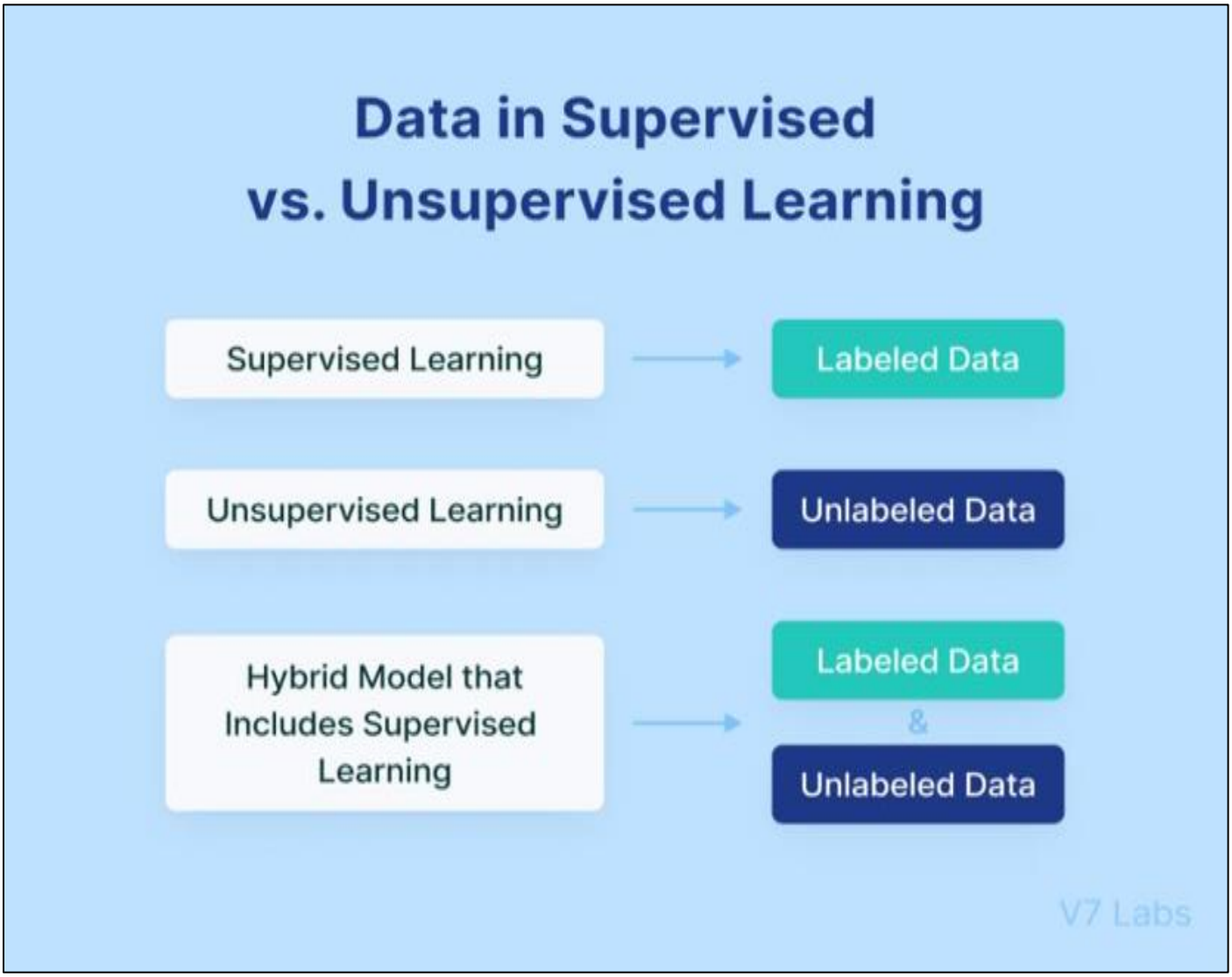

Fig 1 Scheme of Differences between Data in Supervised and Unsupervised Learning
Source: Pragati Baheti, Supervised and Unsupervised learning [Differences & Examples], v7labs.com

Data to predict future outcomes or assign data to specific categories based on the regression or classification problem they are attempting to solve. While supervised learning algorithms tend to be more accurate than unsupervised learning models, they require upfront human intervention to appropriately label the data. However, these labeled datasets allow supervised learning algorithms to avoid computational complexity as they do not require a large training set to produce intended outcomes. Common regression and classification techniques include linear and logistic regression, naïve Bayes, KNN algorithm, and random forest. Semi-supervised learning occurs when only a portion of the input data has been labeled. Unsupervised and semi-supervised learning can be more appealing alternatives as it can be time-consuming and costly to rely on domain expertise to appropriately label data for supervised learning.

ISSN No:-2456-2165

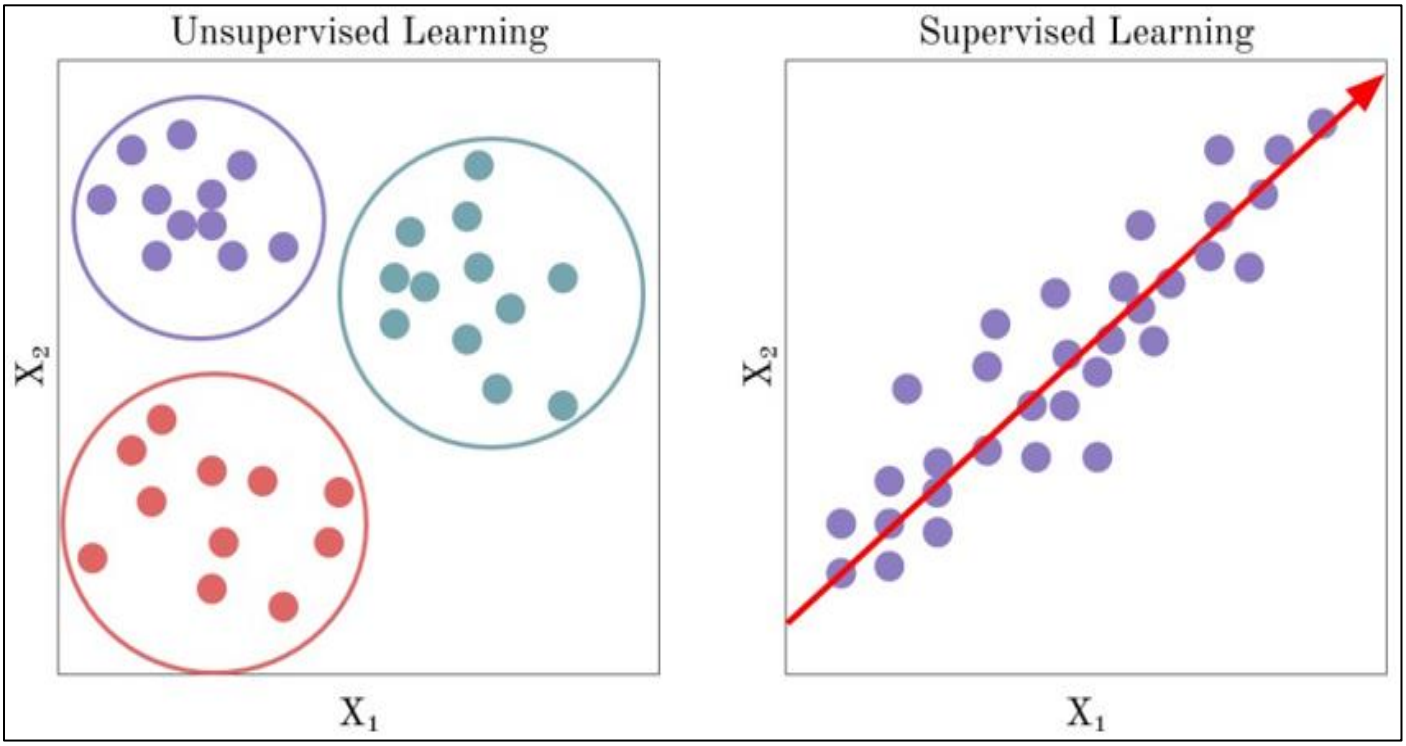

Fig 2 Graphs of Unsupervised and Supervised Learning
Source: Aidan Wilson, A brief introduction to Unsupervised Learning towardsdatascience.com

The primary distinction between supervised and unsupervised learning lies in the requirement for labeled training data. Supervised machine learning relies on input and output data that has been labeled, while unsupervised learning processes raw or unlabeled data. In supervised machine learning, the model learns the relationship between the labeled input and output data. The models are refined until they can accurately predict the outcomes of unseen data. However, the creation of labeled training data often requires significant resources. On the other hand, unsupervised machine learning learns from unlabeled raw training data. An unsupervised model discovers relationships and patterns within this unlabeled dataset, making it useful for identifying inherent trends in a given dataset. Overall, supervised and unsupervised machine learning differ in their approach to training and the type of data the model learns from. Consequently, they also differ in their final application and specific strengths.

Supervised machine learning models are typically employed to predict outcomes for unseen data, such as predicting fluctuations in house prices or understanding the sentiment of a message. These models are also used to classify unseen data based on learned patterns. In contrast, unsupervised machine learning techniques are generally used to comprehend patterns and trends within unlabeled data. This can involve clustering data based on similarities or differences, or identifying underlying patterns within datasets. Unsupervised machine learning can be utilized to cluster customer data in marketing campaigns or to detect anomalies and outliers.

Source: Pragati Baheti, Supervised and Unsupervised learning [Differences & Examples], v7labs.com

- *"The main Differences of Supervised vs Unsupervised Learning include:*

✓ The need for labelled data in supervised machine learning.
✓ The problem the model is deployed to solve. Supervised machine learning is generally used to classify data or make predictions, whereas unsupervised learning is generally used to understand relationships within datasets.
✓ Supervised machine learning is much more resource-intensive because of the need for labelled data.
✓ In unsupervised machine learning it can be more difficult to reach adequate levels of explainability because of less human oversight."

Quote from: Julianna Delua, Supervised vs unsupervised learning, What's the difference?, IBM

A primary distinction between supervised and unsupervised learning lies in the problems that the final models are employed to address. Both types of machine learning models acquire knowledge from training data, but each approach possesses distinct strengths for various applications. Supervised machine learning, for instance, acquires knowledge of the relationship between input and output through labeled training data. Consequently, it is utilized to classify new data by employing these learned patterns or to predict outputs.

ISSN No:-2456-2165

On the other hand, unsupervised machine learning is valuable for identifying underlying patterns and relationships within unlabeled, raw data. This renders it particularly useful for exploratory data analysis, segmentation or clustering of datasets, or projects aimed at comprehending how data features interconnect with other features for automated recommendation systems.

- *Examples of Supervised Machine Learning include:*

✓ Classification and identification of the input data as part of a learned group.
✓ Regression and prediction of the outcomes from continuously changing data.

- *Examples of Unsupervised Machine Learning include:*

✓ Clustering and then grouping together data points with similar data.
✓ Association and understanding on how certain data features are connected with other features.

➢ *Applications of Unsupervised Machine Learning*

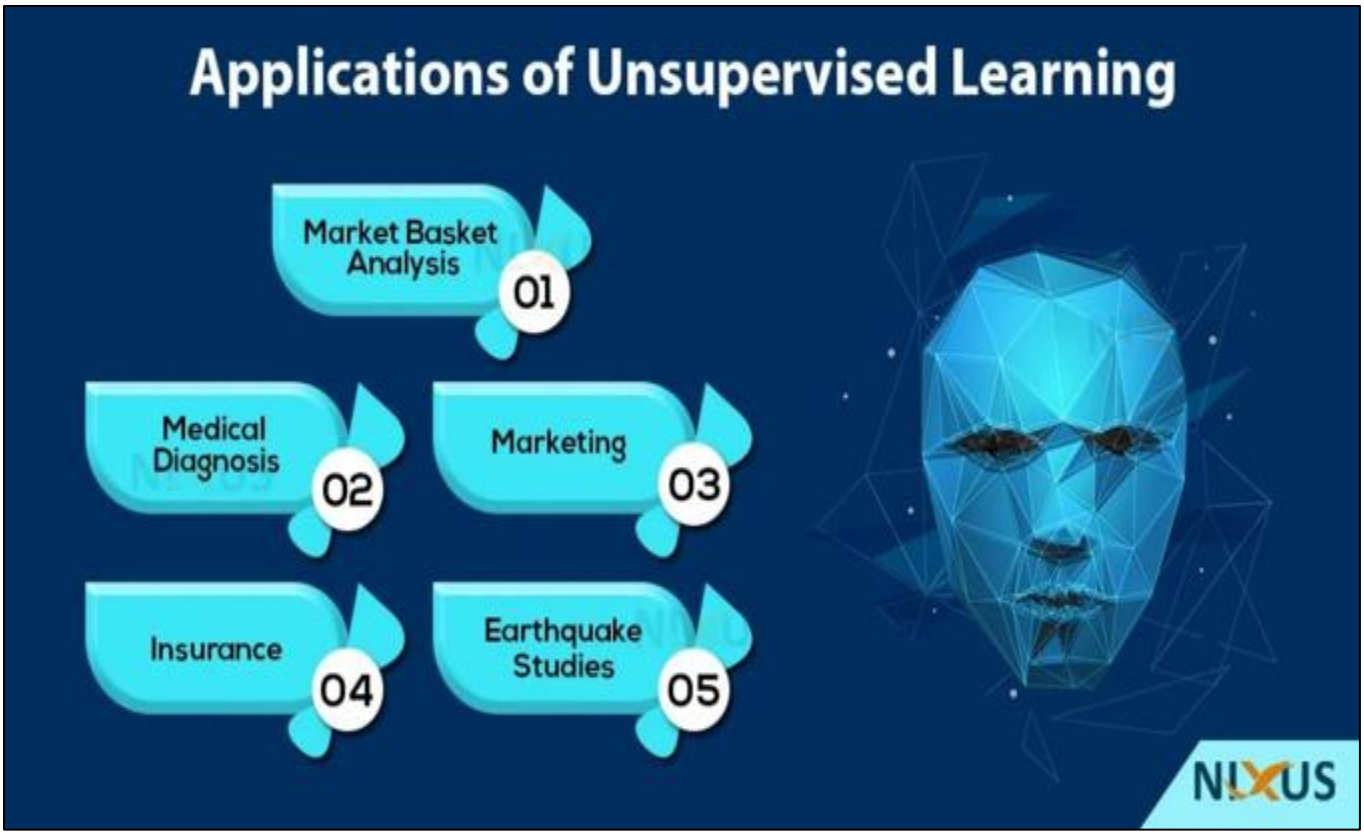

Fig 3 Displayed the Applications of Unsupervised Learning
Source: Unsupervised learning types, algorythms and applications, Nixus.com

"Machine learning techniques have become a common method to improve a product user experience and to test systems for quality assurance. Unsupervised learning provides an exploratory path to view data, allowing businesses to identify patterns in large volumes of data more quickly when compared to manual observation."

Source: Seldon, Supervised vs unsupervised learning compared, Seldon.com

- *The most Common Real-World Applications of Unsupervised Machine Learning are:*

✓ *Market Analysis:*
Market analysis is widely recognized as one of the most prominent examples and applications of unsupervised learning. This approach is frequently utilized by major retailers to uncover the associations between various products.

✓ *Medical Diagnosis:*
Patients are treated quickly using predetermined association rules, because they assist in determining the likelihood of sickness for a certain ailment or product.

✓ *Marketing:*
The identification of consumer groups that exhibit similar behavior, through the analysis of a large database of client data that encompasses their attributes and past purchases, serves to enhance the efficacy of marketing endeavors.

✓ *Insurance:*
The identification of policyholder groups exhibiting a high average claim cost is instrumental in the detection of fraudulent activities.

✓ *Earthquake Studies:*
The identification of hazardous areas is achieved through the grouping of observed earthquake epicenters.

Source: Unsupervised learning types, algorythms and applications, Nixus.com

➢ *Unsupervised Machine Learning in Anatomic Pathology*

- *Objective*

The development of precise supervised machine learning algorithms is impeded by the absence of representative annotated datasets. In the field of anatomic pathology, a majority of data is unlabeled, and the creation of extensive, annotated datasets is a time- consuming and arduous task. Unsupervised learning, which does not necessitate annotated data, has the potential to aid in overcoming this challenge. The purpose of this review is to introduce the concept of unsupervised learning and demonstrate how clustering, generative adversarial networks, and autoencoders can potentially address the lack of annotated data in anatomic pathology.

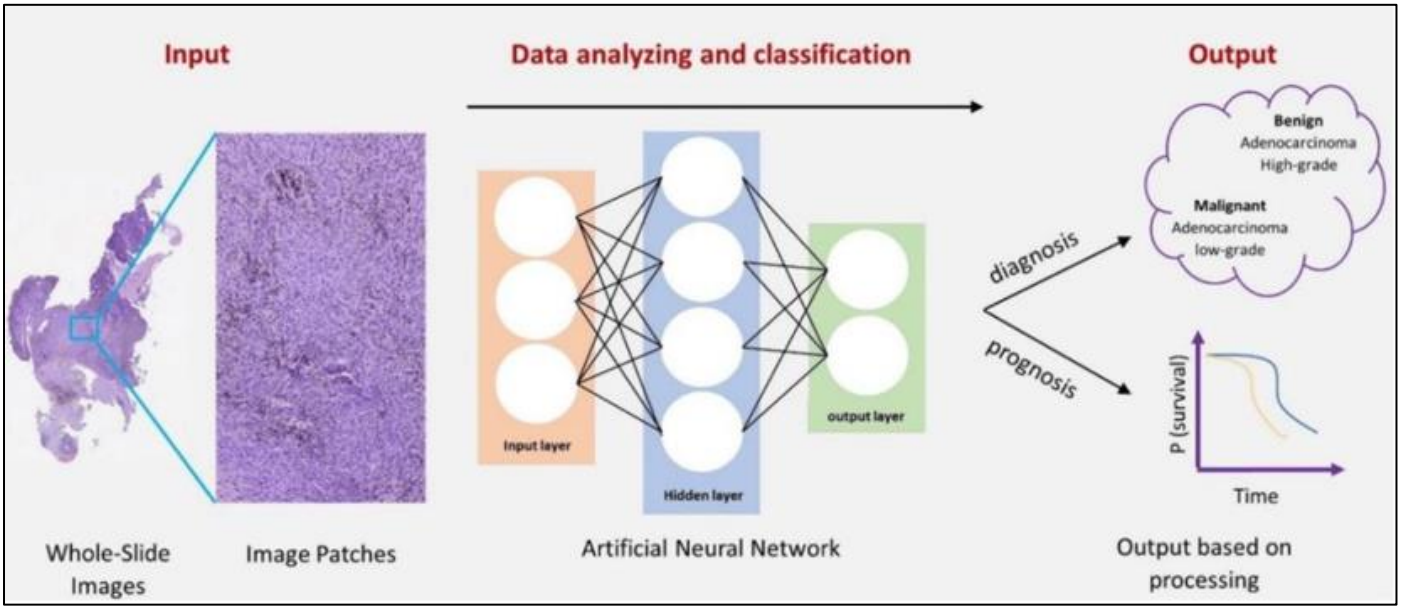

Fig 4 Explanation of a Data analysis for Pathologies, Structure and Development
Source: Dr. Alhassan Ali Ahmed, Dr. Mohamed Abouzid, Prof. Elzbieta Kaczmarek, 2022,
Deep Learning Approaches in Histopathology.

➢ *Results*

Clustering can be employed as a component of semisupervised learning, wherein labels are disseminated from a subset of annotated data points to the remaining unlabeled data points within a dataset. Generative Adversarial Networks (GANs) can aid in this process by generating substantial quantities of synthetic data and conducting data normalization.

Autoencoders enable the training of a network on a vast, unlabeled dataset and the subsequent transfer of acquired representations to a classifier using a smaller, labeled subset (unsupervised pretraining).

➢ *Conclusions*

Unsupervised machine learning methodologies, including clustering, Generative Adversarial Networks, and autoencoders, whether employed independently or in conjunction, have the potential to mitigate the scarcity of annotated data in the field of pathology and enhance the efficacy of constructing supervised learning models.

Source: Ewen D. McApline, Pamela Michelow, Turgay Celik, The Utility of Unsupervised Machine Learning in Anatomic Pathology

ISSN No:-2456-2165

➢ *Machine Learning and Advertising*

"The job of marketers is to get the right product in front of the right consumer at the moment that they're most likely to buy that product. The ability for marketers to do just that and to drill down to more and more specific niches of customers has grown exponentially over time. AI technology is now being used to help marketers get even more specific with predictive targeting and personalization." Appen.com, Quote 1

- *Machine Learning used in Advertising*

With the increasing emphasis on data-driven marketing and the reliance on machine learning algorithms, enterprises are actively seeking avenues to leverage these technologies in order to enhance their advertising endeavors. Recent progress in artificial intelligence (AI) has facilitated machines in acquiring knowledge from extensive data sets and subsequently making informed decisions. Consequently, a multitude of machine learning models, including predictive modeling, natural language processing, and deep learning, have been developed for utilization in marketing.

- *The Forecast*

With the increasing emphasis on data-driven marketing and the reliance on machine learning algorithms, enterprises are actively seeking avenues to leverage these technologies in order to enhance their advertising endeavors. Recent progress in artificial intelligence (AI) has facilitated machines in acquiring knowledge from extensive data sets and subsequently making informed decisions. Consequently, a multitude of machine learning models, including predictive modeling, natural language processing (NLP), and deep learning, have been developed for utilization in marketing.

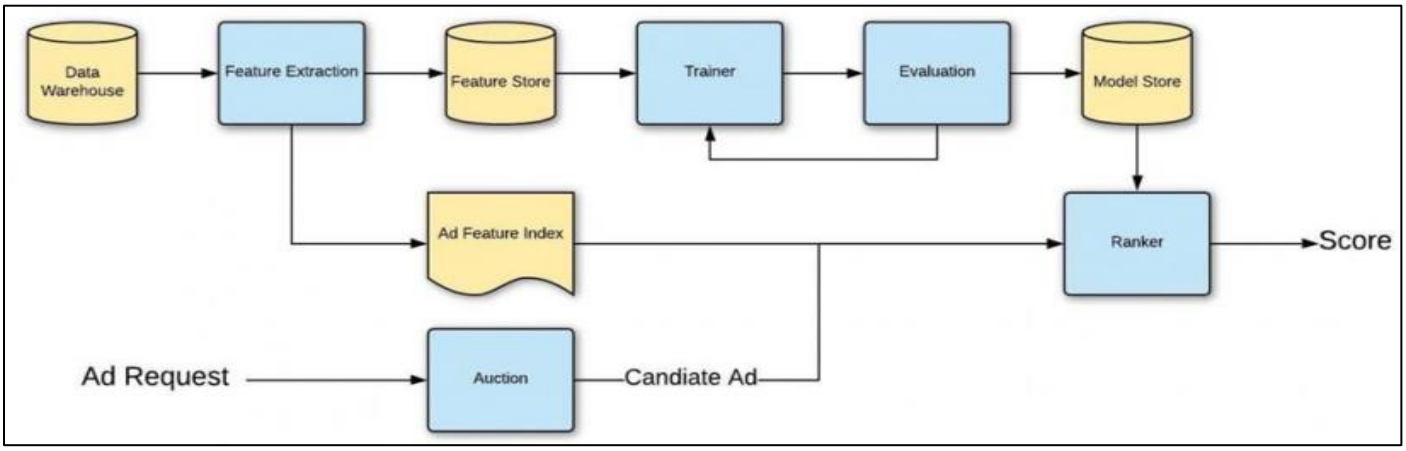

Fig 5 Graphics Displaying an Advertisement Clustering Process

Source: Cứ Máu Là Xong, Advertising system architecture: How to build a production Ad platform, Khamd.com

- *Some Recommendations*

As previously mentioned, *Google* employs a recommendations-based machine learning methodology for predictive text. Similarly, ecommerce advertisers worldwide utilize a recommendations-based machine learning algorithm to effectively deliver product advertisements. By acquiring knowledge and making predictions about the preferences of various user segments, the product ads can be tailored to showcase items that align with the individual's interests. It would be futile, for instance, to present Joe the painter with a set of golf clubs.

- *What is the Targeting*

One of the most crucial elements of marketing is having a comprehensive understanding of one's target audience. The acquisition of extensive data on individuals enables marketers to tailor their messages more effectively. Machine learning has proven to be a valuable tool for marketers in this regard, as it offers novel methods for identifying and targeting audiences with the highest likelihood of converting into customers. An example of how machine learning is employed in the realm of advertising and marketing is through the utilization of targeted ads based on the probability of a purchase. By leveraging machine learning, advertisers can select targeting preferences based on the probability that an individual will make a purchase within a specified timeframe. However, it is important to acknowledge that this approach can potentially encroach upon user privacy. An illustrative case is the controversy surrounding Target, wherein their machine learning algorithm purportedly predicted a teenager's pregnancy. While the veracity of this story remains uncertain, it raises pertinent questions regarding the extent to which our data should be shared or utilized for marketing purposes. It is plausible that the young girl may have purchased a pregnancy test and subsequently bought nausea medication, and the machine learning algorithm connected these two data points, leading to her inclusion in a maternity mailer.

- *Specific Targeting*

The algorithms gather data regarding the content and available ad spaces, including the video's subject matter, the objects featured in the video, and the emotional response it may evoke. Through the utilization of machine learning, *CatapultX's* exclusive

algorithms are capable of aligning the specific moment in the video with an advertisement that is relevant to the content and more likely to generate engagement or ad-recall. As the learning process is ongoing, the algorithm's accuracy improves with each exposure to impressions.

➢ *Future for Machine Learning in Advertising*

In the future, the most skilled marketers and advertisers will be able to generate insights more efficiently through the use of machine learning. This will enable marketers to conduct tests and make improvements at a faster pace. Advertisers and agencies should therefore prepare for multiple versions of creative content early on in a campaign, and allocate sufficient resources for side-by-side testing. This will allow brands to benefit from AI insights more rapidly. We are eagerly anticipating the continued integration of machine learning into marketing and advertising strategies in the coming years.

One of the greatest advantages of machine learning and AI is their ability to analyze vast amounts of data and make accurate predictions. This capability can be harnessed in advertising and marketing to identify relevant signals and present the appropriate advertisements to the right individuals, at the optimal time, without relying on personally identifiable information. With the assistance of machine learning models and AI technology, advertisements are becoming highly targeted and capable of delivering the return on investment that marketers anticipate from digital campaigns.

- *Predictive Target and Test*

Predictive targeting is a marketing strategy that leverages artificial intelligence and machine learning algorithms to anticipate forthcoming customer choices by analyzing behavioral patterns and historical data. This data is utilized to forecast the probability or likelihood of a particular individual undertaking a specific action, such as making a purchase, interacting with products, or converting in alternative manners.

Moreover, predictive targeting tools can assist brands in developing more refined customer personas, enabling them to determine the appropriate target audience for each campaign. This ensures that potential customers receive the most pertinent advertisements possible.

- *The AI Product Recommendations and their Hyper-Relevance:*

One of the most effective and efficient methods to guide a customer through the buyer's journey is by utilizing a product recommendation advertisement. However, the success of this approach is contingent upon the relevance of the advertisement to the individual at any given moment. This is where an AI-powered recommendation model can prove invaluable in eliminating guesswork from the process.

Recommendation models are typically constructed based on established customer attributes and behaviors. This enables the model to suggest products to a new customer based on the information it has gathered about them. Examples of recommendation models in action can be observed on platforms such as *Netflix* and *Amazon,* where they suggest shows to watch and products to purchase. Additionally, popular search engines and social networks that rely on advertising for revenue streams also utilize recommendation models.

Fig 6 Improvements in sales, satisfaction of the consumer and marketing with machine leaning
Source: Louis Columbus, 10 Ways AI and Machine Learning Are Improving Marketing in 2021, business2community.com

ISSN No:-2456-2165

- *The Advanced Product*

One of the significant advancements in recommendation models in recent years has been the transition from utilizing explicit feedback to implicit feedback. Initially, recommendation models relied on explicit feedback provided by customers, such as their preferred product categories. However, more recent models now rely on implicit feedback, analyzing behavioral signals to discern customer intent.

Furthermore, advanced recommendation models have become increasingly precise and specific. Instead of solely relying on product categories, these models now utilize SKU numbers and focus on individual products.

This shift towards specificity in recommendations has transformed the advertising landscape. It is no longer solely centered around products and product categories; rather, it revolves around understanding customers and their purchasing journey. Advertisers are now able to anticipate customers' desires even before the customers themselves are aware of them.

Moreover, recommendation models are not limited to these advancements. The future of AI and machine learning models for advertising will not only incorporate historical user data but also integrate a user's response to advertisements. Recommendations will be continuously updated in real-time. Naturally, effectively managing this vast and diverse dataset necessitates a data partner with a profound understanding of the associated challenges, such as *Appen*.

- *Personalized Ad Targeting:*

Individuals and organizations frequently alter their preferences, making it crucial for advertisers to promptly identify and adjust to these changes. Furthermore, personalized advertisements have become an essential necessity rather than a mere luxury, as people are more inclined to make purchases when the ad aligns with their specific journey. There are various approaches and perspectives through which ads can be personalized, including but not limited to seasonality, weather conditions, geographical location, unique characteristics, personal interests, cultural background, and past buying behavior.

- *Better Brand Safety and better Alignment:*

The primary aim of advertising personalization and predictive targeting is to ensure that the appropriate advertisement is presented to the correct customer. However, this is not the only advantage. The use of artificial intelligence and machine learning to provide product recommendations to customers can also contribute to the establishment of a stronger customer relationship and the effective management of brand safety and alignment. The placement of your advertisement reflects upon your company, and if it is positioned alongside biased, negative, or non-factual content, it can lead to a decline in brand trust and favorability. The context surrounding your advertisement plays a crucial role in defining your brand identity and alignment. By utilizing AI tools and machine learning algorithms, you can ensure that your advertisements are exclusively displayed in locations that align with your brand and are deemed safe.

- *How to make Better Advertising Decisions:*

One of the primary advantages that artificial intelligence (AI) and machine learning can bring to advertising is the capacity to enhance brand and advertising decision-making.

Through the utilization of AI, brands are able to make advertising decisions that are grounded in data, rather than relying on guesswork. This enables brands to accurately identify the target audience for their ads and determine the optimal timing for their delivery. Additionally, machine learning algorithms can assist in selecting the most suitable platforms for ad placement, ensuring the maintenance of a secure and consistent brand Fig . Unlike humans, who may be influenced by personal biases, machine learning models are not susceptible to such limitations. Consequently, decisions made using AI are based on objective data rather than subjective intuition.

Source: Appen, Machine learning in advertising – Predictive targeting and moderation, Appen.com

ISSN No:-2456-2165

CHAPTER TWO
UNSUPERVISED MACHINE LEARNING ALGORITHMS

A. What Means Clustering

Clustering can be regarded as the paramount unsupervised learning quandary; thus, akin to any other quandary of this nature, it pertains to the identification of a structure within an assortment of unlabeled data. A broad definition of clustering could be delineated as "the procedure of categorizing entities into groups whose constituents exhibit similarity in some manner." Consequently, a cluster denotes a compilation of entities that are "similar" amongst themselves and are dissimilar to entities belonging to other clusters.

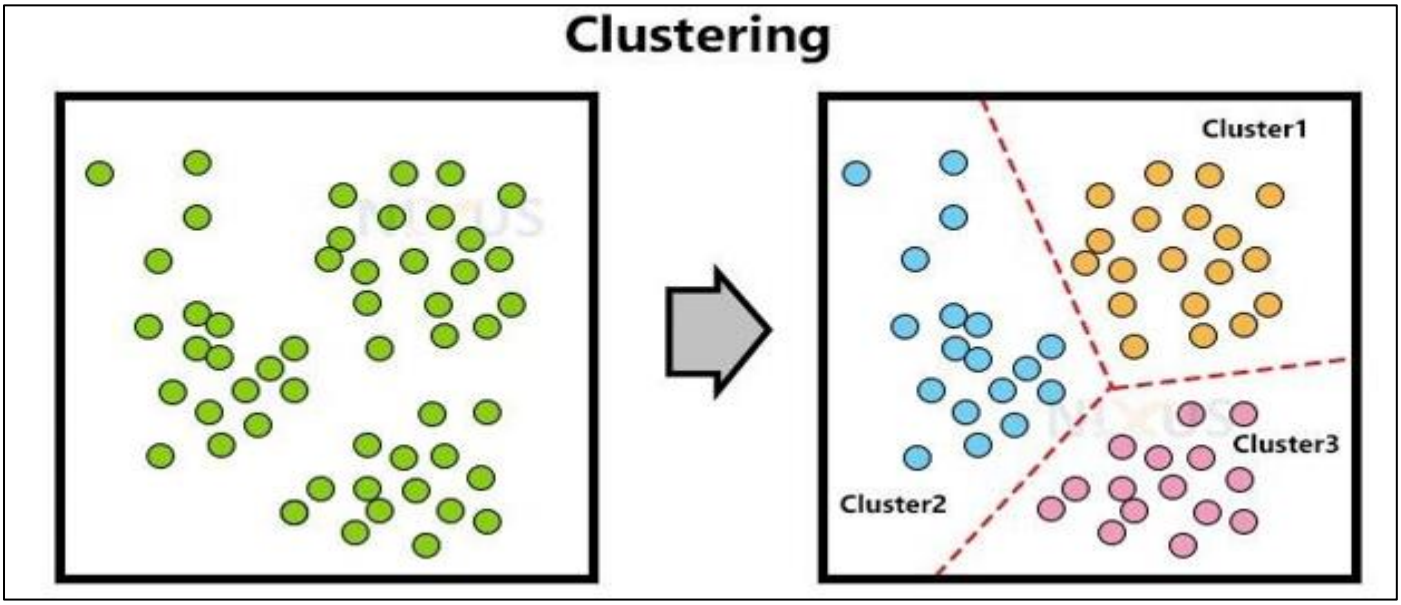

Fig 7 Difference in Pre and Post Clusering Process
Source: Unsupervised learning types, algorythms and applications, Nixus.com

➢ *Goals of Clustering*

The objective of clustering is to ascertain the internal grouping within a collection of unlabeled data. However, the determination of what constitutes a good clustering is a matter of inquiry. It has been demonstrated that there is no definitive "best" criterion that is detached from the ultimate objective of the clustering. As a result, it is incumbent upon the user to provide this criterion in a manner that aligns with their requirements, ensuring that the outcome of the clustering is tailored to their needs.

➢ *Types of Clustering in Unsupervised Learning Machines:*

- *Exclusive Clustering:*

Exclusive clustering is a form of clustering wherein a point is restricted to belonging to only one cluster at any given time. This particular method of grouping is commonly referred to as "hard" clustering. Exclusive clustering encompasses the K-means clustering technique.

- *Hierarchical Clustering:*

Hierarchical clustering, also referred to as hierarchical cluster analysis, is an unsupervised clustering methodology that can be categorized into two distinct types: agglomerative and divisive clustering.

- *Agglomerative Clustering:*

Agglomerative clustering is a technique for clustering that follows a "bottoms-up" approach. Initially, the data points are identified as independent groups and subsequently merged together repeatedly based on their similarity until a single cluster is formed.

ISSN No:-2456-2165

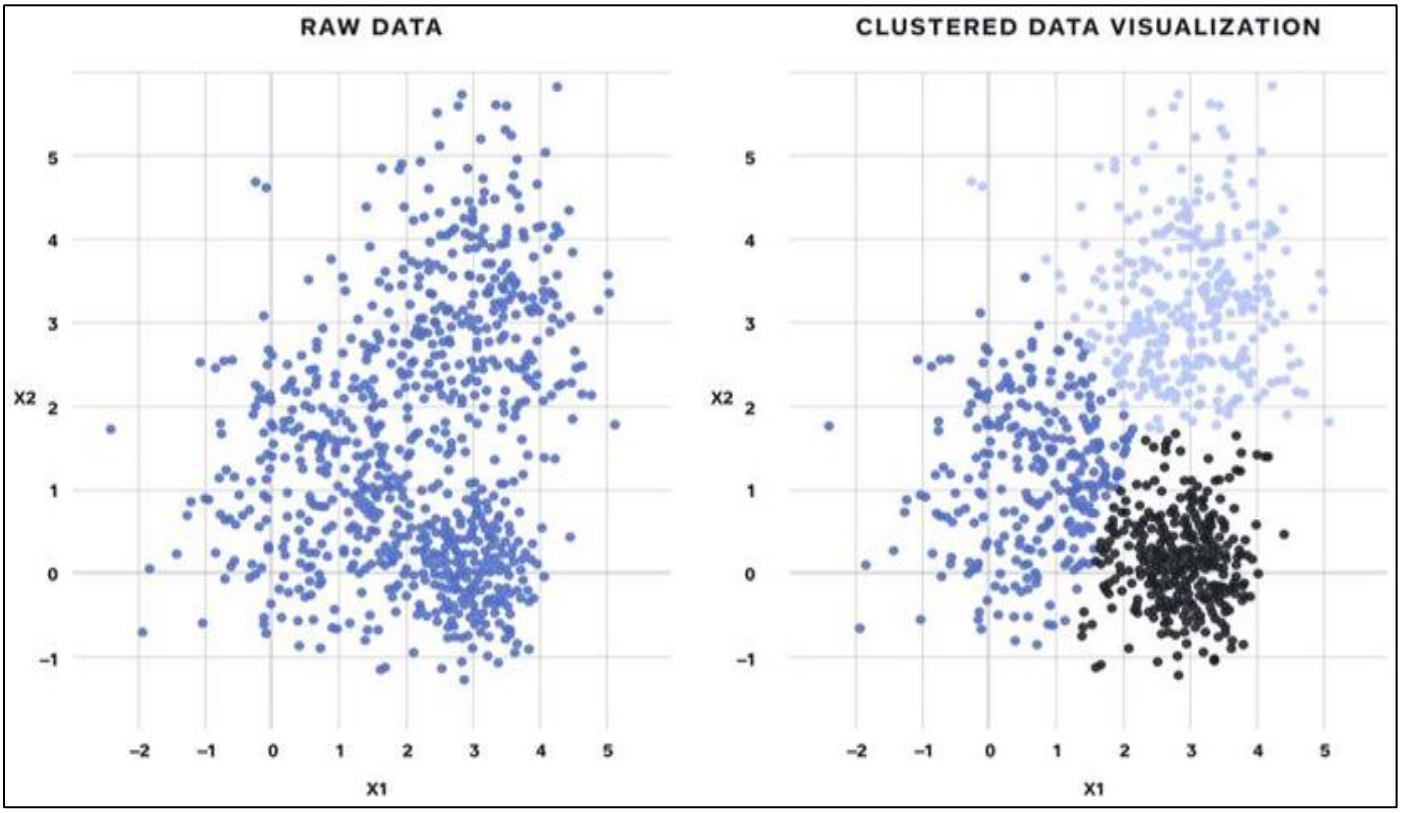

Fig 8 Showed a Set of Data before Clustering and after, with Relative Classification
Source: January 2019, Abhishek Kumar, A Clustering Approach for Customer Billing Prediction in Mall: A Machine Learning Mechanism - Scientific Figure

- *Probabilistic Clustering:*

A probabilistic model is an unsupervised technique employed to address density estimation and "soft" clustering problems. Data points are probabilistically grouped in clustering based on the likelihood of belonging to a specific distribution. The Gaussian Mixture Model is widely recognized as one of the most frequently utilized models in this context.

Source: Unsupervised learning types, algorythms and applications, Nixus.com

B. K-Means Models

The objective of the K-means algorithm is to divide the data into *K* separate clusters, with the intention of minimizing the distance within each cluster and maximizing the distance between clusters.

Let S be a set such that:

$$\sum_{k=1}^{K} \sum_{p \in S_k} ||x^p - c^k||_2^2 \qquad (2.1)$$

$$S = \{S_1, S_2, \ldots, S_K\} \qquad c^k = \frac{1}{|S_k|} \sum_{p \in S_k} x^p$$

ISSN No:-2456-2165

To be more precise, when provided with a Training Set (S), the aim is to allocate all elements to one of the sets S1, S2, ..., Sk in a manner that minimizes the sum of squares within each cluster.

$$\sum_{p\in S_k} ||x^p - c^k||_2^2 = \sum_{p\in S_k} (x^p - c^k)^T(x^p - c^k)$$

$$= \sum_{p\in S^k} x^{pT} x^p - 2 \sum_{p\in S^k} c^{kT} x^p + |S_k| c^{kT} c^k$$

$$= \sum_{p\in S^k} x^{pT} x^p - 2 |c^{kT} \left(\frac{1}{|S_k|} \sum_{p\in S_k} x^p\right) + |S^k| c^{kT} c^k$$

$$= \sum_{p\in S_k} x^{pT} x^p - |S_k| c^{kT} c^k.$$

This means that:

$$\frac{1}{2|S^k|} \sum_{p\in S_k} \sum_{q\in S_k} ||x^p - x^q||_2^2 = \frac{1}{2|S^k|} \sum_{p\in S_k} \sum_{q\in S_k} (x^p - x^q)^T(x^p - x^q) =$$

$$\frac{1}{2|S^k|} \sum_{p\in S_k} \sum_{q\in S_k} x^{pT} x^p + \frac{1}{2|S^k|} \sum_{p\in S_k} \sum_{q\in S_k} x^{qT} x^q - \frac{2}{2|S^k|} \sum_{p\in S_k} \sum_{q\in S_k} x^{pT} x^q =$$

$$\frac{2}{2|S^k|} |S^k| \sum_{p\in S_k} x^{pT} - \frac{1}{|S^k|^2} |S^k| \left(\sum_{p\in S_k} x^p\right)^T \left(\sum_{q\in S_k} x^q\right) =$$

$$\sum_{p\in S_k} x^{pT} x^p - |S_k| c^{kT} c^k.$$

The conclusions follows as:

$$\sum_{p\in S_k} ||x^p - c^k||_2^2 = \frac{1}{2|S^k|} \sum_{p\in S_k} \sum_{q\in S_k} ||x^p - x^q||_2^2$$

$$min_S \sum_{K=1}^{K} \frac{1}{2|S^k|} \sum_{p\in S_k} \sum_{q\in S_k} ||x^p - x^q||_2^2 \qquad (2.2)$$

Now are expressed the following concepts:

- Within cluster sum of squares or WCSS = $\sum_{K=1}^{K} \sum_{p \in S_k} ||x^p - c^k||_2^2$
- Between cluster sum of squares or BCSS = $\sum_{K=1}^{K} |S_k||c^k - c||_2^2$
- Total sum of squares or TSS = $\sum_{p=1}^{P} ||x^p - c||_2^2$

Where

$$c = \frac{1}{p} \sum_{p=1}^{p} x^p$$

Then

$$\text{TSS} = \text{WCSS} + \text{BCSS} \qquad (2.3)$$

In fact, it's possible to write the following equations:

$$\sum_{p=1}^{P} ||x^p - c||_2^2 = \sum_{k=1}^{K} \sum_{p \in S_k} ||x^p - c||_2^2 = \sum_{k=1}^{K} \sum_{p \in S_k} ||x^p + c^k - c^k - c||_2^2 =$$

$$\sum_{k=1}^{K} \sum_{p \in S_k} ||x^p + c^k||_2^2 + \sum_{k=1}^{K} \sum_{p \in S_k} ||c^k - c||_2^2 + 2 \sum_{k=1}^{K} \sum_{p \in S_k} (x^p - c^k)^T (c^k - c) =$$

$$\text{WCSS} + \text{BCSS} + 2 \sum_{k=1}^{K} \sum_{p \in S_k} (x^p - c^k)^T (c^k - c) =$$

But

$$\sum_{k=1}^{K} \sum_{p \in S_k} (x^p - c^k)^T (c^k - c) = \sum_{k=1}^{K} (c^k - c)^T \left(\sum_{p \in S^k} (x^p - c^k) \right) =$$

$$\sum_{k=1}^{K} (c^k - c)^T \left(\sum_{p \in S^k} x^p - |S_k| c^k \right) = 0$$

Therefore:

$$\sum_{p=1}^{P} ||x^p - c||_2^2 = \sum_{\substack{k=1 \\ p \in S_k}}^{K} \sum ||x^p - c^k||_2^2 \sum_{k=1}^{K} |S_k| \, || \, c^k - c||_2^2$$

ISSN No:-2456-2165

Based on the aforementioned equality, given that the quantity TSS remains constant, it can be deduced that the minimization of WCSS is tantamount to the maximization of BCSS.

$$\min_{r \in \{0,1\}^{PxK}} \sum_{K=1}^{K} \frac{\sum_{p=1}^{P-1} \sum_{q=p+1}^{P} ||x^p - x^q||_2^2 r_{pk} r_{qk}}{\sum_{p=1}^{P} r_{pk}} \qquad (2.4)$$

$$\text{Subjected to} \quad \sum_{k=1}^{K} r_{pk} = 1, \qquad \text{for all } p = 1, \ldots, P$$

$$\text{Where} \quad r_{pk} = \begin{cases} 1 & \text{if } p \in S_k \\ 0 & \text{otherwise} \end{cases}$$

The issue at hand is of NP-hard nature, as it has been demonstrated that the K-means problem is NP-hard, even when dealing with instances in a two-dimensional plane. This outcome has been achieved through a reduction process originating from the planar 3SAT problem. The NP-hardness is established through a reduction from the Exact Cover by 3- Sets problem.

Source: Renato De Leone, 3 February 17 2023, Machine Learning, Course notes - Version 4.1, Cap 5: Unsupervised Learning Techniques, pages (51 - 60), RDL.

C. K-means Algorithm

The K-means clustering algorithm calculates the centroids and continues iterating until it discovers the optimal centroid. It operates under the assumption that the number of clusters is already known and is commonly referred to as a flat clustering algorithm. The number of clusters identified by the algorithm from the data is denoted by 'K' in K-means.

Given an initial set of K centers, c1, c2, ..., cK, the algorithm proceeds with the following two steps until the centers cease to change:

First, in the assignment step, each point in the dataset is assigned to the closest center, resulting in the determination of clusters S1, S2, ..., SK, such that for each k = 1, ..., K.

$$S_k = \{p \in \{1, \ldots, P\} : ||x^p - c^k||_2^2 \leq ||x^p - c^i||_2^2, \; i = 1, \ldots, K\},$$

Second assignment step: Centers recalculation

$$c^k = \frac{1}{|S_k|} \sum_{p \in S_k} x^p$$

The algorithm achieves convergence towards a local minimum, which is contingent upon the initial selection of the centers. The most straightforward initialization procedure entails randomly selecting K elements from the dataset to serve as centers.

➢ *Algorithm (1): Pseudocode for k-means algorithm*

Choose K initial centers $c^1, c^2, \ldots, c^K$;
while *not convergence* **do**
  $S_1 = S_2 = \ldots S_K = \emptyset$;
  forall $p = 1, \ldots, P$ **do**
    Determine k such that

$$\left\|x^p - c^k\right\|_2^2 \leq \left\|x^p - c^i\right\|_2^2, i = 1, \ldots, K$$

    Set $S_k = S_k \cup \{p\}$
  forall $k = 1, \ldots, K$ **do**
    Set

$$c^k = \frac{1}{|S_k|} \sum_{p \in S_k} x^p$$

➢ *Examples of Data Sets analyzed using the k-means Algorithm:*

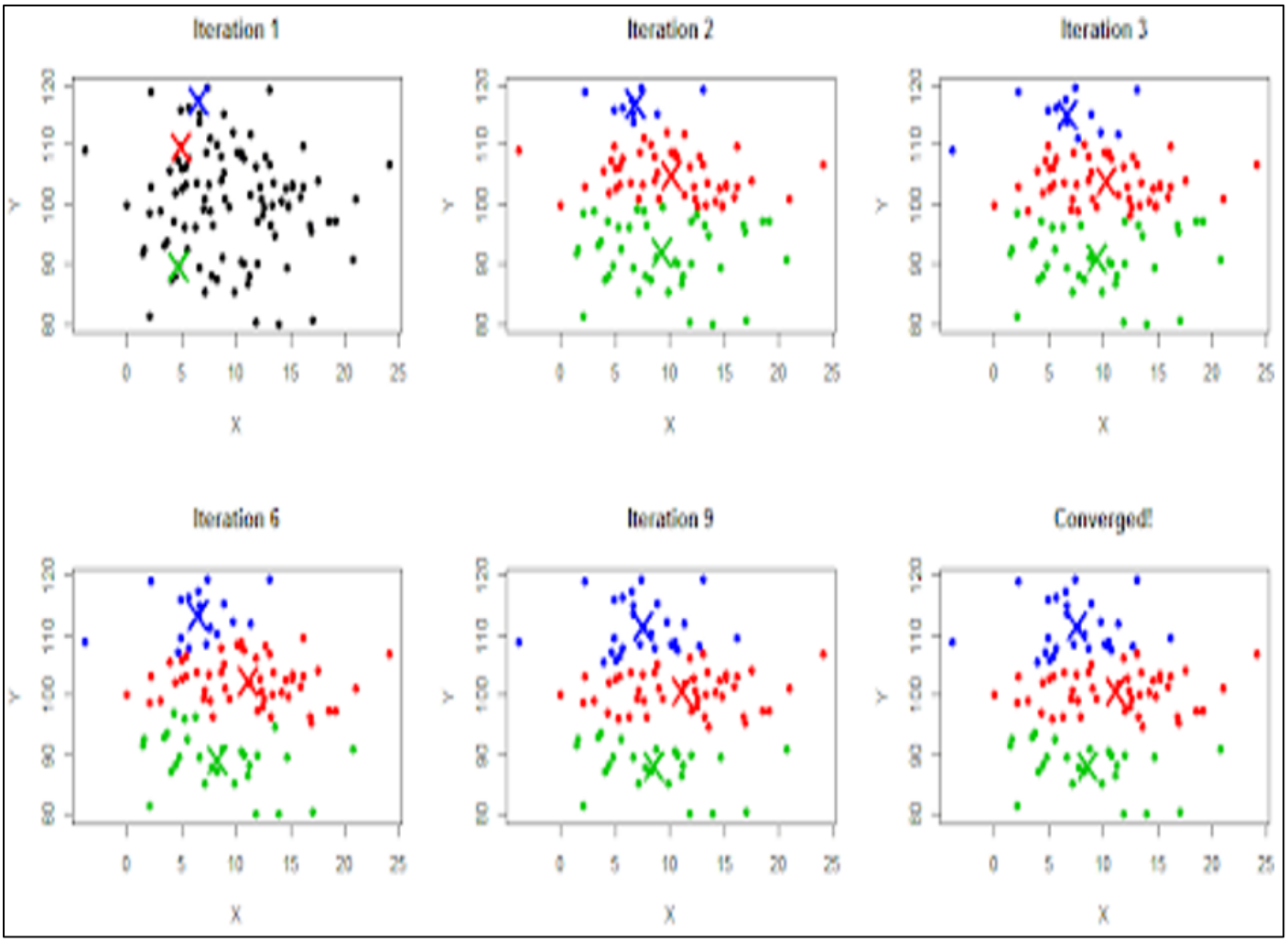

Fig 9 Showing Three Data Types, with 6 Interactions
Source:Dayanithi, K-Means Clustering, Medium, levelup.gitconnected.com

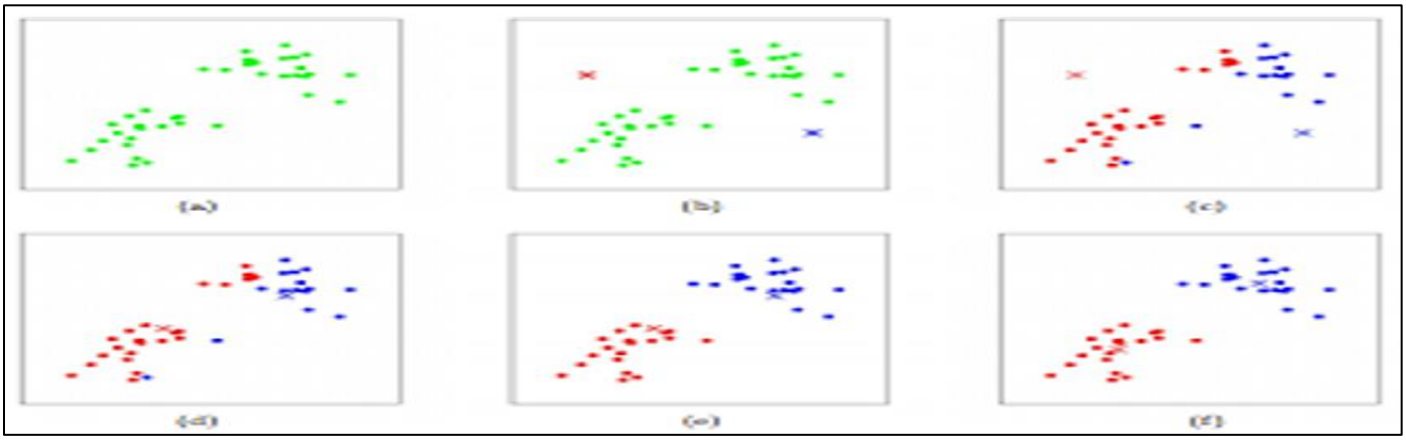

Fig 10 Displaying an analysis of Two Different Data Types with 6 Interactions
Source: Sara Wendte, K-Means Clustering Applications in Digital Signal Processing

Source: Renato De Leone, 3 February 17 2023, Machine Learning, Course notes - Version 4.1, Cap 5: Unsupervised Learning Techniques, pages (51 – 60), RDL.

D. Hierarchical Clustering

Hierarchical clustering is widely utilized as a method for categorizing entities. It facilitates the formation of groups wherein the entities within a group exhibit similarities amongst themselves while being distinct from entities in other groups. These clusters are visually depicted in a hierarchical tree structure known as a dendrogram.

➢ *Hierarchical Clustering Offers Several Notable Advantages:*

- It is unnecessary to pre-determine the quantity of clusters. Instead, the dendrogram can be severed at the suitable level to achieve the intended number of clusters.

- Data can be conveniently summarized and organized into a hierarchical structure through the use of dendrograms. The utilization of dendrograms facilitates the examination and interpretation of clusters in a straightforward manner.

➢ *Two main Methods*

The bottom-up methods initially establish a similarity measure between two objects and subsequently expand it to encompass the similarity between clusters. Conversely, top- down methods directly establish the similarity between clusters.

➢ *Bottom-up Hierarchical Clustering*

The Bottom-up Hierarchical Clustering algorithms construct a cluster hierarchy by utilizing the dissimilarity between clusters, which is typically represented in the form of a dendrogram. The dendrogram is a graphical representation of the analyzed data using a bottom-up approach.

Fig 11 Dendrogram of Data using Hierarchical Clustering
Source: Dr. Soumen Atta, Ph.D , Hierarchical Clustering in Python: A Step-by-Step, Tutorial, Medium.com

ISSN No:-2456-2165

➢ *Algorithm (2): Pseudo-code for Agglomerative Hierarchical Clustering algorithm*

1 **Set** D be the index set of the initial clusters

$$D = \{1, \ldots, P\}$$

and let $C_p = \{x^p\}$, $n_p = 1, \quad p \in D$;
2 **Set** d_{rs} the distance between clusters C_r and C_s, $r, s \in D$, $r \neq s$;
3 **repeat**
4 **Find** the smallest values d_{rs}, $r, s \in D$, $r \neq s$;
5 **Merge** clusters C_r and C_s into a new cluster C_k with $n_k = n_r + n_s$;
6 **Remove** indices r and s from D;
7 **Add** index k to D;
8 **Compute** the distances of cluster C_k from all remaining clusters in D

$$d_{kl} = \alpha_r d_{rl} + \alpha_s d_{sl} + \beta d_{rs} + \gamma |d_{rl} - d_{sl}|$$

for all $l \in D$, $l \neq k$;
9 **until** $|D| = 1$;

The quantity drs can be interpreted as a measure of dissimilarity between clusters Cr and Cs. Various options for the parameters α and β in Algorithm 2 correspond to distinct Agglomerative Hierarchical Clustering algorithms. The steps 4-8 will be iterated P - 1 times.

➢ *Types of Hierchical Techniques:*

- *Single Linkage*

In this particular scenario, the distance separating two clusters is equivalent to the minimum distance observed between any element belonging to one cluster and any element belonging to the other cluster. In simpler terms, the distance between two clusters can be defined as the distance between their two nearest elements.

αr = αs = 0.5, β = 0, γ = −0.5.

- *Complete Linkage*

In this particular scenario, the distance separating two clusters is equivalent to the maximum distance observed between any element belonging to one cluster and any element belonging to the other cluster. Consequently, the distance between two clusters can be determined by measuring the distance between their two most distant elements. Here:

αr = αs = 0.5, β = 0, γ = 0.5.

- *Simple Average*

The distance between two clusters is defined as the average distance between each pair of elements (one for each cluster), weighted so that the two clusters have equal influence on the final result.

αr = αs = 0.5, β = 0, γ = 0.

- *Centers*

The distance between two clusters is defined as the distance between their corresponding centers.

$$a_r = \frac{n_r}{n_k}, \; a_s = \frac{n_s}{n_k}, \; \beta = -a_r a_s, \; \gamma = 0$$

ISSN No:-2456-2165

Source: Fatih Karabiber, Hierchical clustering on, LearnDataSci.com and Aurélien Géron, September 2019, Hands-On Machine Learning with Scikit- Learn, Keras, and TensorFlow.

- *Top down hierchical clustering*

Top-down clustering, also known as divisive hierarchical clustering, commences by considering the entire dataset as a single cluster and subsequently partitions it into smaller clusters based on a specific criterion, such as variance, entropy, or silhouette. An illustration of top-down hierarchical clustering involves the division of a cluster R into two subgroups, namely A and B.To perform the spitting, the following steps are necessarily executed. First A = R and B = ∅. Then the elements from cluster A are moved to cluster B. For each element r in A, compute the average dissimilarity to all the other objects of A:

$$\text{diss}(r, A \setminus \{r\}) = \frac{1}{|A|-1} \sum_{s \in A, s \neq r} d_{rs}$$

The element r′ that reaches the maximum value is moved from cluster A to cluster B:

$A \leftarrow A \cup \{r'\}$, $B \leftarrow B \cup \{r'\}$

Then, as long as A still contains more than one object, is computed:

$$\text{diss}(r, A \setminus \{r\}) - \text{diss}(r, B) = \frac{1}{|A|-1} \sum_{s \in A, s \neq r} d_{rs} - \frac{1}{|B|} \sum_{s \in B} d_{rs}$$

Let r′ be the element that is reaching the maximum value. If such value is a positive element, r′ is moved from cluster A to cluster B. Otherwise, if this value is negative the splitting is over.

Source: Fatih Karabiber, Hierchical clustering on, LearnDataSci.com

ISSN No:-2456-2165

CHAPTER THREE
EXERCISE

➢ *Presentation of the Algorithm Structure*

The k-means clustering problem can be better understood by expressing it as a Mixed- Integer Quadratically Constrained Problem in GAMS. Although it appears little at first, it actually needs some attention.

As previously demonstrated, the K-means algorithm may be applied to improve the quality, efficacy, and efficiency of advertisements by developing a large dataset that represents consumers or potential customers and selecting how to manage various ad formats based on different factors.

The k-means algorithm and approach used in gams are efficiently expressed by the following GAMS algorithm, which is followed by the results and closing remarks.

The general algebraic modeling system (GAMS) is a high-level modeling system for mathematical optimization. GAMS is designed for modeling and solving linear, nonlinear, and mixed-integer optimization problems. The system is tailored for complex, large-scale modeling applications and allows the user to build large maintainable models that can be adapted to new situations. The system is available for use on various computer platforms. Models are portable from one platform to another.

➢ *Algorithm*

```
$ontext

  K-means clustering heuristic

$offtext


option seed=101;

sets
   k  'clusters of clients' /k1*k4/
```

ISSN No:-2456-2165

```
    i   'data value' /i1*i100/
    ik(i,k) 'assignment of value to clusters of clients'
    xy 'coordinates of points that rapresents clients
evaluations depending on the type of advertisement' /x,y/
;
parameter
    m(k,xy) 'random clusters of potential clients to generate
data points'
    p(i,*)   'data points'
    numk     'number of clusters of potential clients'
    cluster(i)     'cluster of potential clients identification
number'
;

numk = card(k);
alias(ii,i);

*---------------------------------------------------------------------------
* generate a random potential clients data set
*---------------------------------------------------------------------------
m(k,xy) = uniform(0,4);
cluster(i) = uniformint(1,numk);
ik(i,k) = cluster(i)=ord(k);
p(i,xy) = sum(ik(i,k),m(k,xy)) + uniform(0,1);
display m,p;

sets
   ik(i,k) 'assignment of points to clusters'
   ik_prev(i,k) 'previous assignment of points to clusters'
   ik_diff
   trial    'number of trials' /trial1*trial15/
   iter     'max number of iterations' /iter1*iter20/
;
parameters
   n(k)         'number of points assigned to cluster'
   c(k,xy)      'centroids'
   notconverged  '0=converged,else not converged'
   d(i,k)       'squared distance'
   dclose(i)   'distance closest cluster'
   trace(trial,iter) 'reporting'
;

loop(trial,

*---------------------------------------------------------------------------
* Step 1
* Random assignment
*---------------------------------------------------------------------------

        cluster(i) = uniformint(1,numk);
        ik(i,k) = cluster(i)=ord(k);
```

```
      notConverged = 1;
      loop(iter$notConverged,

*-----------------------------------------------------------------------------
* Step 2
* Calculate centroids of each cluster of clients depending on
clients average characteristics and specifics that the
advertisement project want to enhance
*-----------------------------------------------------------------------------
         n(k) = sum(ik(i,k), 1);
         c(k,xy)$n(k) = sum(ik(i,k), p(i,xy))/n(k);

*-----------------------------------------------------------------------------
* Step 3
* Re-assign points depending on the results
*-----------------------------------------------------------------------------

         ik_prev(i,k) = ik(i,k);
         d(i,k) = sum(xy, sqr(p(i,xy)-c(k,xy)));
         dclose(i) = smin(k, d(i,k));
         ik(i,k) = yes$(dclose(i) = d(i,k));

*-----------------------------------------------------------------------------
* Step 4
* Check convergence
*-----------------------------------------------------------------------------

         ik_diff(i,k) = ik(i,k) xor ik_prev(i,k);
         notConverged = card(ik_diff);

         trace(trial,iter) = sum(ik(i,k),d(i,k));
      );
);

display trace;
```

Source: Tuesday, March 31, 2015, k-means clustering heuristic in GAMS and the XOR operator, Yet Another Math Programming Consultant.

ISSN No:-2456-2165

➢ *Results*

The algorithm's output in GAMS after it has run is as follows:

```
----    100 PARAMETER trace  reporting

              iter1       iter2       iter3       iter4       iter5
       iter6

trial1      366.178      28.135      14.566
trial2      271.458      26.741      14.566
trial3      316.975      23.912      14.566
trial4      299.522      24.511      14.566
trial5      346.148      77.747      14.566
trial6      313.978      64.730      14.865      14.566
trial7      310.735      27.412      14.566
trial8      330.829     213.308     184.504     102.191      76.611
      74.210
trial9      356.897      90.286      16.112      14.566
trial10     345.086      82.716      76.154      76.146
trial11     354.545     169.648      75.626      74.350      71.929
      67.385
trial12     310.257      52.175      18.800      14.566
trial13     289.293      23.505      14.566
trial14     337.024      20.783      14.566
trial15     336.969      28.747      14.566

      +       iter7       iter8       iter9      iter10

trial8       67.933      42.621      14.881      14.566
trial11      59.815      59.591
```

Based on the findings, it is crucial to utilize diverse initial configurations when attempting to optimize the sum of squared distances. The general agreement among researchers and experts in the field is that the optimal value for the sum of squared distances is 14.566.

Utilizing diverse initial configurations is essential because it allows for a more comprehensive exploration of the solution space. By starting with a variety of initial configurations, we can avoid getting trapped in local optima and increase the chances of finding the global optimum. This is particularly important in complex optimization problems where the solution space is vast and intricate.

The findings suggest that the optimal sum of squared distances, which amounts to 14.566, represents a significant achievement in the field. This value indicates the minimum possible sum of squared distances that can be achieved for the given problem.

Achieving this optimal value is crucial as it signifies the best possible solution and can have important implications in various domains.

Furthermore, the agreement among researchers regarding this optimal value adds credibility and reliability to the findings. When multiple experts independently arrive at the same conclusion, it strengthens the validity of the results and increases confidence in the optimal solution.

In conclusion, the utilization of diverse initial configurations is crucial when aiming to optimize the sum of squared distances. The general agreement among experts is that the optimal value for this metric is 14.566. This finding highlights the importance of exploring a wide range of initial configurations and provides a benchmark for evaluating the quality of solutions in this context.

Source: Tuesday, March 31, 2015, k-means clustering heuristic in GAMS and the XOR operator, Yet Another Math Programming Consultant.

➢ *Conclusion*

The GAMS software, which stands for General Algebraic Modeling System, is a powerful tool that can be used in conjunction with the k-means clustering algorithm to effectively manage and analyze a set of data representing clients and their corresponding variables in an advertisement campaign.

In an advertisement campaign, it is crucial to understand the characteristics and preferences of the target audience in order to create effective marketing strategies. The variables that can be considered in this analysis include the client's estimated interest in the product, geographical location, previous research and purchases made through partner companies, and client category based on age, salary, wealth, and general interests.

By utilizing the k-means clustering algorithm, which is a popular unsupervised machine learning technique, the GAMS software can group clients into distinct clusters based on their similarities and differences in these variables. This clustering process allows for the identification of different customer segments within the target audience.

Once the clusters are formed, the GAMS software can then be used to analyze the characteristics and preferences of each cluster. This analysis can provide valuable insights into the specific needs and preferences of different customer segments, allowing for the creation of tailored marketing strategies for each cluster.

By leveraging the power of the GAMS software and the k-means clustering algorithm, the effectiveness of the advertisement campaign can be significantly enhanced. Precise targeting of the campaign based on the analysis of client data can lead to increased profits and reduced costs.

For example, by identifying clusters of clients who have a high estimated interest in the product and have made previous purchases through partner companies, the campaign can be targeted towards these clusters to maximize the chances of conversion and increase sales. Similarly, by identifying clusters of clients who are located in specific geographical areas, the campaign can be localized to target these specific regions, reducing unnecessary costs associated with targeting a broader audience.

Overall, the utilization of the GAMS software with the k-means clustering algorithm allows for a data-driven approach to advertisement campaign management. By analyzing client data and tailoring marketing strategies based on the insights gained from this analysis, businesses can optimize their advertisement campaigns, resulting in increased profits and reduced costs.

Source: Tuesday, March 31, 2015, k-means clustering heuristic in GAMS and the XOR operator, Yet Another Math Programming Consultant.

REFERENCES

[1]. Aidan Wilson, December 7 2020 A brief introduction to Unsupervised Learning towardsdatascience.com
[2]. Pragati Baheti, October 1 2021, Supervised and Unsupervised learning [Differences & Examples], v7labs.com
[3]. Julianna Delua, March 12 2021, Supervised vs unsupervised learning, What's the difference?, IBM.com
[4]. Nixus, 2023, Unsupervised learning types, algorythms and applications, Nixus.com
[5]. Seldon, September 16, 2022, Supervised vs unsupervised learning compared, Seldon.com
[6]. January 2022, Ewen D. McApline, Pamela Michelow, Turgay Celik, The Utility of Unsupervised Machine Learning in Anatomic Pathology, PMID: 34302331, DOI: 10.1093/ajcp/aqab085, I am J Clin Pathol, PubMed website.
[7]. Cứ Máu Là Xong, May 19 2020, Advertising system architecture: How to build a production Ad platform, Khamd.com
[8]. Dr. Alhassan Ali Ahmed, Dr. Mohamed Abouzid, Prof. Elzbieta Kaczmarek, 2022, Deep Learning Approaches in Histopathology, Cancers 2022, 14(21), 5264, Fig Analysis and Computational Pathology in Cancer Diagnosis.
[9]. Cứ Máu Là Xong, May 20 2021, How to predict the success of your marketing campaign, Khamdb.com
[10]. Qortex, 2023, How AI Targets Audiences, Qortex.com
[11]. Forbes, December 2019, A commissioned study conducted by Forrest consulting on behalf of IBM.
[12]. Appen, March 10 2022, Machine learning in advertising – Predictive targeting and moderation, Appen.com
[13]. Louis Columbus, August 18 2023, 10 Ways AI and Machine Learning Are Improving Marketing In 2021, business2community.com
[14]. Chandrasekaran, Sriramakrishnan & Kumar, Abhishek. (2019). A Clustering Approach for Customer Billing Prediction in Mall: A Machine Learning Mechanism. Journal of Computer and Communications. 07. 55-66. 10.4236/jcc.2019.73006.
[15]. Dr. Soumen Atta, Ph.D, April 4 2023, Hierarchical Clustering in Python: A Step- by-Step, Tutorial, Medium.com
[16]. Sara Wendte, 2017, K-Means Clustering Applications in Digital Signal Processing, 14.2-Clustering-KMeansAlgorithm-Machine Learning - Professor Andrew Ng, Stanford CS 221,"K Means, Purdue ECE 438, "ECE438 - Laboratory 9: Speech Processing (Week 1)", October 6, 2010.
[17]. Dayanithi, March 13 2023, K-Means Clustering, Medium, levelup.gitconnected.com
[18]. Tuesday, March 31, 2015, k-means clustering heuristic in GAMS and the XOR operator, Yet Another Math Programming Consultant.
[19]. Renato De Leone, 3 February 17 2023, Machine Learning, Course notes - Version 4.1, Cap 5: Unsupervised Learning Techniques, pages (51 - 60), RDL.
[20]. Fatih Karabiber, Hierchical clustering on, LearnDataSci.com
[21]. Aurélien Géron, September 2019, Hands-On Machine Learning with Scikit- Learn, Keras, and TensorFlow, 2nd Edition, Publisher: O'Reilly Media Inc., ISBN: 9781492032649