INTERNATIONAL SCIENTIFIC AND VOCATIONAL JOURNAL (ISVOS JOURNAL)

Vol.: 5 Issue: 2 Date: 31.12.2021 Received: 27.11.2021 Accepted: 30.12.2021 Final Version: 31.12.2021

ISVOS Journal, 2021, 5(2): 241-253 - https://doi.org/10.47897/bilmes.1029139

An Optimization Focused Machine Learning Approach in Analysing Arts Participative Behavior with Fine Arts Education Considerations

Özerk Yavuz[a,1]

[a,1]Haliç Üniversity, İstanbul 34445, Turkey
ORCID ID: 0000-0002-1371-688X

Abstract

Fine arts education constitutes an important value for the development of nations, countries and geographies. In literature it is seen that societies and nations that give importance to fine arts, other art forms with respective education programs and agendas seem to be more developed in the fierce competitive environments. It is a known fact that art forms enhance the creativity of individuals as well as developing questioning skills and empathy. Therefore cultural plans, programs and agendas handled in the cultural ministry or similar levels should always be considered as a chance for investments made for the development of individuals and society at large. Therefore enough budgets for cultural programs should be maintained. Academic and pre-university degree art programs should be designed and formed in response to the changes and developments in the contemporary art world. Graduates should also be prepared in these programs to the career paths that they want to follow in these areas. Understanding the domain of art marketing as well as implementing marketing strategies, plans, programs in every level in this context can be helpful in enlarging the arts market which means creating more jobs and employment opportunities for more actors and actresses. Qualification of these talents should also be supported with academical or nonacademical artistic education forms. In this study theatres located in Istanbul Municipality and theatre participation behavior has been analyzed with recent machine learning trends and approaches available in data mining literature with the aim of enhancing exploratory and confirmatory understanding in such domains as theater, entertainment and art marketing with the aim of providing some insights and ideas to the arts community, leaders and society at large.

Keywords: "Fine Arts, Fine Arts Marketing, Fine Arts Education, Entertainment, Entertainment Marketing, Theater Marketing, Art Marketing, Clustering, Classification, Data Mining, Machine Learning, Quantitative Analysis, Supervised Learning, Unsupervised Learning"

1. Introduction

Fine arts education constitutes an important value for the development of nations, countries and geographies. In literature it is seen that societies and nations that give importance to fine arts, other art forms with respective education programs and agendas seem to be more developed in the fierce competitive environments. It is a known fact that art forms enhance the creativity of individuals as well as developing questioning skills and empathy. Therefore cultural plans, programs and agendas handled in the cultural ministry or similar levels should always be considered as a chance for investments made for the development of individuals and society at large. Therefore enough budgets for cultural programs should be maintained. Academic and pre-university degree art programs should be designed and formed in response to the changes and developments in the contemporary art world. Graduates should also be prepared in these programs to the career paths that they want to follow in these areas. Understanding the domain of art marketing as well as implementing marketing strategies, plans, programs in every level in this context can be helpful in enlarging the arts market which means creating more jobs and employment opportunities for more actors and actresses. Qualification of these talents should also be supported with academical or nonacademical artistic education forms. In this study theatres located in Istanbul Municipality and theatre participation behavior has been analyzed with recent machine learning trends and approaches available in data mining literature with the aim of enhancing exploratory and confirmatory understanding in such domains as theater, entertainment and art marketing with the aim of providing some insights and ideas to the arts community, leaders and society at large.

Marketing science has been in the focus of many businesses, industries and sectors for many decades. In this context, marketing scholars and practioners have long worked on building good relationships with the customers applying contemporary, state of the art customer centric marketing approaches. As a result marketing found place and gained popularity in many areas as in goods marketing, services marketing, political marketing, celebrity marketing, entertainment and art marketing. In this study some of the factors associated with the theaters which has known to be one of the most influential forms of arts in human history. In this context theatres located in Istanbul Municipality and theatre participation behavior has been analyzed with recent machine learning trends

[1] Sorumlu Yazar. Tel.: +90-212-924-2444.
E-posta adresi: ozerkyavuz@halic.edu.tr

and approaches available in data mining literature with the aim of enhancing exploratory and confirmatory understanding in such domains as theater, entertainment and art marketing [1,2,6,11,12,13,14,15,16, 19,20].

As many marketers would know or imagine, success of a business highly linked to the successful relationships that a business build with your customer. Therefore understanding the consumer, his/her needs, expectations, requirements and functionalities expected from the goods and services constitutes an important value. Therefore permanent qualitative and quantitative research approaches should be formulated and conducted to gain insights and knowledge about the targeted markets, market segments and consumers. As in many relational forms and relational marketing approaches in order to build a good relationship with the consumer a trust associated with the goods, services provided and the organizations which provides these. Then a relationship is built with the commitment of the two parties when the expectations match from such relationship. In this context analyzing and understanding the consumer and providing the goods and services expected with the right marketing mixes composed of good product or service functionalities with the expected quality, with an appropriate pricing strategy targeting the related market segments based on their purchasing power, effective use of logistical, sales, service channels and points with a good promotion strategy composed of successful advertisement and public relation approaches would provide the customer satisfaction which would lead to a long term profitable relationship with the customer. Organizations who successfully form these marketing mixes would make a positive impact in increasing market share, enhancing their customer portfolio, customer acquisition, customer retention and customer lifetime value. From this point forward, quantitative data mining techniques can be used to analyze the consumers, organizations, markets and several consumer behaviors. This way of an aristothelen approach of positivist research paradigm can be used to formulate more targeted segmentation, targeting and positioning in the market place. By conventional and new forms of communication approaches and tools right messages highlighting different facets of the brand equity and brand would be helpful in building a good image in the minds of the consumer. Therefore leveraging strong brand associations and a strong brand equity in the mind of the consumers would make a positive influence to be clustered in the consideration set in consumer decision making process which can lead to a purchase behavior of the consumer momentarily [45, 52 ,53].

As seen in many behavioral forms and several behavioral theorists, scholars indicate there are several elements and many antecedents that lead to a behavior to occur. Consumer decision making, theatre participation and art participation behaviors are not much different in this scope. As indicated in literature by several notable scholars in literature, individuals are in a tendency to engage in behaviors that are likely to provide rewards and positive outcomes whereas they do not engage in behaviors where it would lead to negative outcomes or consequences. There may be several influential factors which can lead to a behavior to be formed as, subjective norms, group norms, behavioral intention, past experiences, positive anticipated emotions, negative anticipated emotions, perceived behavioral control, group influence, environment, trust, social identity with several characteristics of the individuals in this context. In this study a quantitative research methodology has been followed by applying contemporary clustering and classification approaches available in literature. In this way it is aimed to gain an in depth understanding of theatre participation behavior with such phenomena in the city of Istanbul [45 ,52, 53].

When the history of theater and development of theater era investigated it is seen that early forms of theatres found place in Ancient Greeks which dates back to 6th century BC. Later different forms of theaters with other entertainment forms as cinema and art found place in several business settings and geographies. Greek Theatre, Roman Theatre, English Elizabethan theatre, Spanish Golden age theatre, French Classical theatre, American Theatre, Turkish Theatre, Neoclassical theatre, African theatre, Asian Theatre, 19th Century Theatre, 20th century theatre, 21st century theatre have been some of theatrical forms and ages found place in different geographies in history. If the history of the democracies and theatres are analyzed it seen that in some autocracies and undemocratic regimes theatre and artists were subjected to pressure, plays have been forbidden or censored whereas with the development of democracies, the emphasis on freedom of though and speeches artists and theatre found the place they deserved mostly in western, civilized democracies and many parts of the world in different geographies [6,8,9]. Many notable theatrical plays took place all over the world in different geographies and locations as Hamlet, Long Day's Journey Into Night, Who's Afraid of Virginia Woolf?, Death of a Salesman, Oedipus Rex, Look Back in Anger, Waiting for Godot, Uncle Vanya, Tartuffe, Candida, Twelfth Night, Playboy of the Western World, The Importance of Being Earnest. Shakespeare, William Shakespeare, Henrik Ibsen, Oscar Wilde, Sophocles, Tennessee Williams, Arthur Miller, Anton Chekhov, Molière, Euripides, Aristophanes, Samuel Beckett, Victor Hugo, Voltaire have been some of these notable playwrights whose plays have been played by countless valuable, theatres, artists and art lovers. Some have been transformed to distinguished and original works in cinema archives [1,2,6,11,12,13,14,15,16, 19,20].

When the scientific body of knowledge is investigated it is seen that several research methodologies, research techniques evolved which base on different research paradigms and designs over time. Among them data mining and applying machine learning techniques have gained momentum and interest in understanding different phenomena by applying several techniques. Data mining approach can be considered as a systematical, structured research process which focuses on situation analysis, data gathering, model formation and testing of the model. Later findings and knowledge discovered from these analyses can be used as a decision support point for leaders, science community and society at large. Machine learning technique which is a famous approach in data mining based quantitative research methodologies is a form of learning in machine forms. This learning process is usually triggered and activated by forward feeding approaches which is later followed with backpropagation processes which are stochastic in nature. With the help of mapping functions input layers in the model are mapped to the output layer considering the independent, dependent values. Functions and equations involved in this mapping are calculated. Later in many forms rules generated with the least error rate and which provides the most proximity to actual results are selected and presented as the

distinctive association rules. In the evaluation of this a stochastic backpropagation technique is used in many cases [46,47,48,49,50,51].

Supervised learning and unsupervised learning are two forms of machine learning in data mining. Supervised learning is a form of classification approach where input and output layer mappings are done with the transformation functions, with the aim of rule discovery and insights discovery. A stochastic backpropagation technique is used in this type of machine learning in general. In the mapping process independent multivariate variables are assigned to the respective class labels which are considered as the dependent variables in the output layer. Since the initial labeling of the dependent values in the form of nominal values technique, this type of machine learning is named supervised machine learning. On the other hand in unsupervised machine learning several attributes of different instances are assigned to respective clusters with respective values without the requirement of an initial class label declaration. In this form of machine learning several mathematical and statistical functions utilizing heuristics in many cases are applied. In most of the clustering analysis which is also known as unsupervised machine learning centroid values for each independent cluster is calculated and related attribute values for several instances in the data set are assigned to the cluster with the focus on similarity and convergence maximization in one cluster and divergence, difference maximization with other cluster members having other centroid values. Instances with the respective parameter values are assigned to the clusters based on the centroid value similarity. Following these assignments centroid values for each cluster is updated once new instances with respective attributes are assigned to the associated cluster. J48, JRip, Part, OneR Method, Multilayer Perceptron, Bayesian Networks, Hoeffding Tree, Random Tree, Kmeans have been some of the mostly cited supervised and unsupervised machine learning techniques which utilizes different classification and clustering approaches in literature [46,47,48,49,50,51].

In general in machine learning approaches, mainly the ones employing a tree/graph structure, an entropy value is calculated based on different probabilities and combinations of the variables then from the parent nodes to the child nodes paths with the low entropy values are followed. Paths with the minimum entropy value is chosen and respective boundary conditions are assigned to each node (decision nodes) in the tree/graph structures with the aim of reducing the chaos perceived from unpredictability. In this way the unpredictability, disorder randomness or uncertainty with selected paths are minimized. This way of assignment and tree/graph formations are managed in feedforwarding type of feeding and employing activation or mapping/transformation functions. Later with the backpropagation approach most optimum paths and functions are tried to be formed with the customization of the functions and function weights. This is known as a form of gradient descent learning rule or delta rule which aims to calculate most optimum weights for the input output mappings in the mapping/transformation functions. In this rule learning approach the weights which generates the most optimum solution are assigned as the slopes in the function [62, 63, 64 ,65 ,66 ,67 ,68]. A gradient in this context shows the slope which is the derivative of cost at the respective cost point. This slope can also be calculated by finding the tangent of that point which can be calculated by dividing edge looking to the angle(opposing edge) to adjacent edge in a triangular form. By employing a gradient descent rule learner as in stochastic processes as backpropagation the shortest and steepest route which indicates the local or global low with a slope zero. Weights at this point giving the minimum cost are used as the slopes of the function. In this way the total loss associated with the input layer variable weights are customized and updated. While the gradient descents, the total cost decreases and reaches to the minimum at t1 which is the optimum weight for the minimum cost. In this process gradient descents iteratively and approximates to the minimum cost with the respected weight of the independent variable. Derivative of cost find its minimum in the reverse parabol's minimum vertex (local minimum or global minimum) which gives the minimum cost and associated weight providing this minimum cost for the function. This point is reached in an iterative manner moving in the opposite direction of the gradient since it will approximate to the steepest and shortest point (descent). This is the point giving the most optimal weight when the slope is zero in local minimum or global minimum. Alternatively moving in the direction of the gradient would approximate the point to the local or global maximum which is the gradient ascent [62, 63, 64 ,65 ,66 ,67 ,68].

The rule with the lowest entropy value and lowest cost which is an indication of an optimal solution is presented and listed as the association rule which maps the input layers (spaces) to the output layers (spaces). In each of these mappings a cost or lost function is calculated which assigns an intuitive cost associated with the mapping function. Cost function is formulated as the average of loss functions whereas loss function is involved in finding the error rate for a single training example. It calculates the difference between actual and predicted values. Optimization approaches are mainly involved in minimizing this cost and try to find the best fitting parameters with techniques as gradient descent, delta rule learning or backpropagation. On the other hand an error function indicates the deviation of an actual value from its prediction for the instance

In this context employing the combination of supervised and unsupervised machine learning techniques can be used as an effective and efficient approach in understanding some of the macro and micro variables associated with theatre and art participation in different geographies. Understanding these factors would also be helpful in formulating influential marketing strategies, plans, programs which may have a positive impact on brand equity, brand image, customer sales, customer retention customer satisfaction, customer lifetime value and building long term profitable relationships with the customer. Also in the governmental level results of such studies can be used as a decision support point in making legislative actions and reforms which would provide a legislative framework formed based on the expectations and needs of the art and theatre lovers. With the help of effective and state of the art legislative reforms and actions freedoms of the artists, art lovers and art community can be preserved while preventing barriers for art lovers, art and theatre participation behavior [46,47,48,49,50,51].

2. Research Method

In this study a quantitative research paradigm has been employed. Quantitative research paradigms aim to explore the phenomena in applying deductive techniques. They have a positivistic and objectivist orientation of epistemology and ontology. Quantitative data is a type of structured knowledge which can be collected by several approaches in the form of primary data sources and secondary data sources. Conventional and unconventational forms of data gathering techniques can be applied as, sensor, user, service driven data collection or gathering, internet of things, paper administered surveys, online surveys etc. [46,47,48,49,50,51,61,69].

Data mining is defined as a methodological approach in quantitative data analysis as indicated in literature. Data mining process is composed of some set of structured steps that makes the data mining research process and methodology. Initially understanding and analysis of the situation and business problem is completed which is followed with the examination and pre-processing of data. Later a conceptual framework or model is devised following the literature review and analysis approaches. Testing of the model with supervised and unsupervised versions of machine learning approaches takes place. Finally predicted analysis results are evaluated and assessed [46,47,48,49,50,51,61,69].

Data mining approach can be considered as a systematical, structured research process which focuses on situation analysis, data gathering, model formation and testing of the model. Later findings and knowledge discovered from these analyses can be used as a decision support point for leaders, science community and society at large. Machine learning technique which is a famous approach in data mining based quantitative research methodologies is a form of learning in machine forms. This learning process is usually triggered and activated by forward feeding approaches which is later followed with backpropagation processes which are stochastic in nature. With the help of mapping functions input layers in the model are mapped to the output layer considering the independent, dependent values. Functions and equations involved in this mapping are calculated. Later in many forms rules generated with the least error rate and which provides the most proximity to actual results are selected and presented as the distinctive association rules. In the evaluation of this a stochastic backpropagation technique is used in many cases. Supervised learning and unsupervised learning are two forms of machine learning in data mining. Supervised learning is a form of classification approach where input and output layer mappings are done with the transformation functions, with the aim of rule discovery and insights discovery. A stochastic backpropagation technique is used in this type of machine learning in general. In the mapping process independent multivariate variables are assigned to the respective class labels which are considered as the dependent variables in the output layer. Since the initial labeling of the dependent values in the form of nominal values technique, this type of machine learning is named supervised machine learning. On the other hand in unsupervised machine learning several attributes of different instances are assigned to respective clusters with respective values without the requirement of an initial class label declaration. In this form of machine learning several mathematical and statistical functions utilizing heuristics in many cases are applied. In most of the clustering analysis which is also known as unsupervised machine learning centroid values for each independent cluster is calculated and related attribute values for several instances in the data set are assigned to the cluster with the focus on similarity and convergence maximization in one cluster and divergence, difference maximization with other cluster members having other centroid values. , J48, JRip, Part, OneR Method, Multilayer Perceptron, Bayesian Networks, Hoeffding Tree, Random Tree, Kmeans have been some of the mostly cited supervised and unsupervised machine learning techniques which utilizes different classification and clustering approaches in literature . Machine learning utilization of data mining can provide exploratory and confirmatory understanding in the phenomena in question and may provide insights and in-depth understanding with knowledge discovery, prediction or forecasting option it provides. In this context a data mining approach with the publicly supplied data set of Istanbul Municipality to understand art and theatre participation behavior has been employed. Such analysis may provide an in-depth understanding of the topic of interest for the individuals, researchers, leaders, managers and different stakeholder of the community and society at large. These type of approaches can be used and adapted to different projects, works and research designs for knowledge discovery, prediction or forecasting purposes for such domains [46,47,48,49,50,51,61,69].

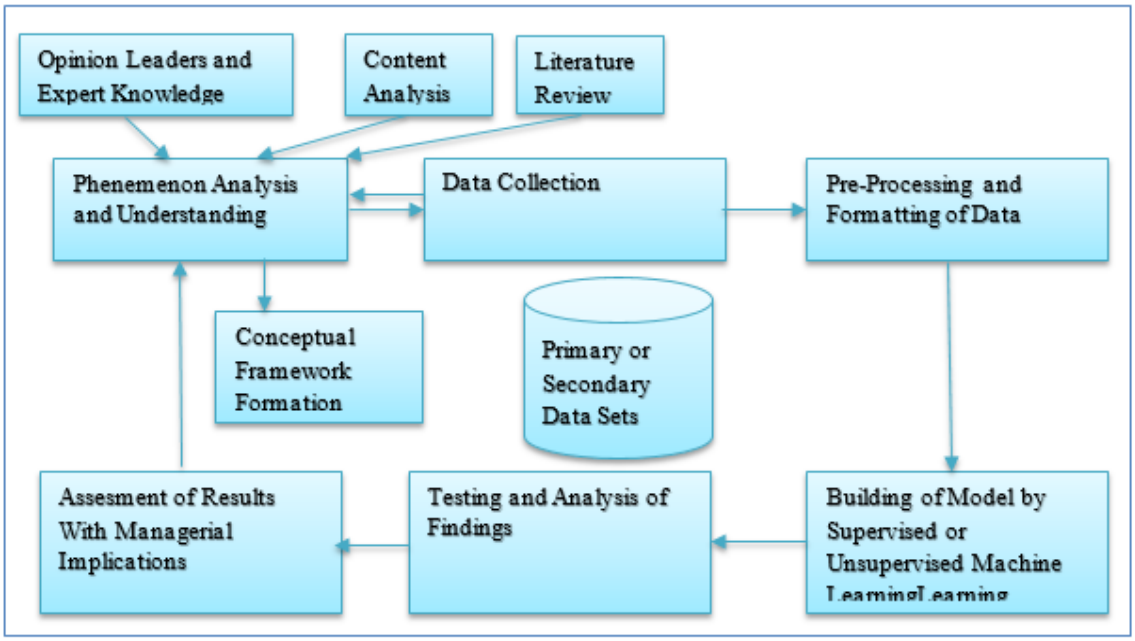

Figure 1. Data Mining Process (Prepared by the Researcher)

Data Mining methodology followed can be seen in Figure 1. In the analysis section, JRip, PART, OneR, Multilayer Perceptron, Bayesian Network, Kmeans have been used for the machine learning techniqes. Unsupervised machine learning algorithms here asseses the instance values and assigns these independent values to the respective segment clusters whereas supervised machine learning algorithms mainly focuses on mapping the multivariate variables in input layers to class labels in output layers with transformation and mapping functions, later applying stochastic backpropogation techniques in many cases. Additionaly class based metrics are evaluated. Performances of several machin learning approaches have been compared. Rules with rule functions have been generated in an reinforced fashion some applying forward feeding and backpropagation approaches based on the algorithmic designs and architectures they have [46,47,48,49,50,51,61,69]. Based on several factors as algorithmic design, algorithmic architecture, complexity of the algorithms these algorithms can generate different results for similar, same or distinct problem sets and domains [46,47,48,49,50,51,61,69]. By using the same data set with the same parameter values, performance indicators of the algorithms have been assessed and evaluated. The top scorer algorithm for this problem domain with respective data set and parameters has been discovered with the analysis conducted. Knowledge patterns and rules found out have been interpreted and listed. Unsupervised and supervised machine learning methodologies followed is seen in Figure 2 and Figure 3.

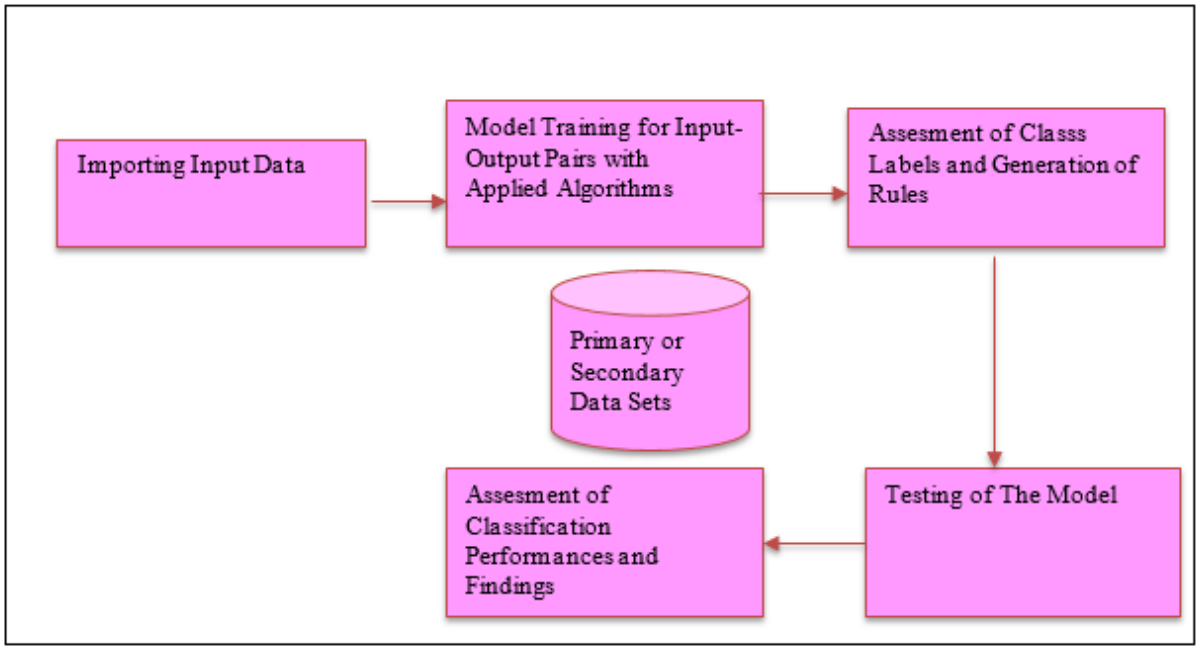

Figure 2. Supervised Machine Learning Process (Prepared by the Researcher)

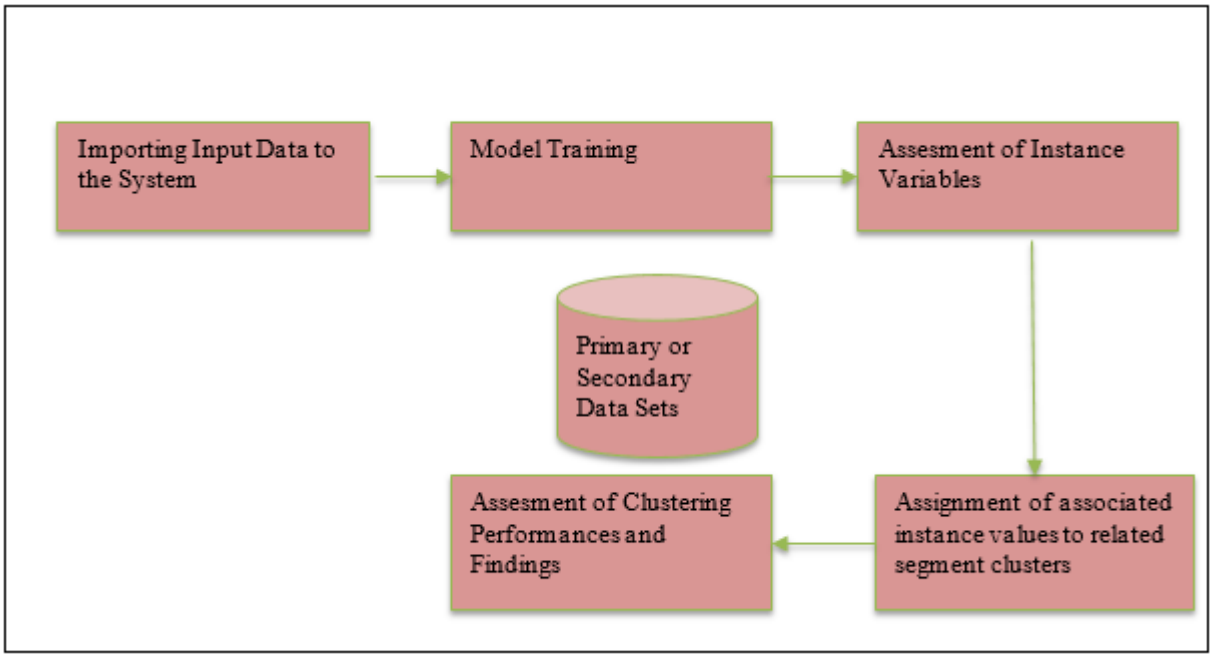

Figure 3. Unsupervised Machine Learning Process Composed of Model Building and Testing (Prepared by the Researcher)

In the data analysis which is composed of a supervised and unsupervised form of machine learning, pulic secondary data of Istanbul Municipality has been used. After the pre-processing of the data and the literature review, classification and clustering methodologises have been followed.

Table 1. List of Attributes

Play Date	Nominal
Play Category	Nominal
Guest Player Number	Numeric
Play Type	Nominal
Number of Audience	Numeric
Interest Status Indicator	Nominal
Theater Name	Nominal
Play Name	Nominal

In data mining analysis, association rules, knowledge and understandings are discovered with the help of classification and clustering algorithms for the relevant problem set and domain. In these approaches, input-output mapping functions are used to create association rules that map the outer layer to the inner layer. In some, feed forward and back propagation techniques have been applied. The relevant rules with the least error rate are presented as the main rules of the analysis [46,47,48,49,50,51,61,69].

As Özerk claims, many data mining processes today apply a technical approach in supervised learning where independent or multivariate indicators and variables are assigned to output class labels using functions of mapping. In unsupervised versions of data mining and machine learning, the core values of each cluster (centroids) are calculated, the corresponding sample and attribute values are assigned to the respective clusters so as to maximize convergence and minimize differences in the sam cluster, whereas a divergence is expected with the members of different clusters. In the process of supervised and unsupervised machine learning, rules are created to improve the exploratory and confirmatory understanding of the phenomenon [46,47,48,49,50,51,61,69]. In this context, an Aristotelian research design path can bring several advantages in understanding these phenomena and can be a good decision support tool for key business leaders, political leaders and society in general.

The same input load with the same parameters was tested using machine learning algorithms, J48, JRip, Part, OneR Method, Multilayer Perceptron, Bayesian Networks, Hoeffding Tree and Random Tree. The University of Waikato's Weka data mining package, which includes supervised and unsupervised machine learning applications, was used in the analysis. Then, the performance of classification and clustering was compared and evaluated. In the analysis, 69 percent of the data was used to train the model, which was then tested with the test data set for the same variables. Based on the performance indicators associated with the data mining analysis, a high performance algorithm was chosen and can be used for such areas and sets of problems to gain additional insight and insight. For this purpose, values of mean squared error, precision, correct classification rate and

misclassification rate were used [46,47,48,49,50,51,61,69]. The analysis revealed the performance indicator values and rules as in Tables 3 and 4.

Table 2. Performance Scores of Machine Learning Algorithms

Method Applied / Performance Indicator	Multilayer Perceptron	Bayesian Networks	J48	JRip	Part	OneR Method	Hoeffding Tree	Random Tree
RMSE	0.0737	0.0004	**0**	**0**	**0**	**0**	0.0097	0
Precision	N/A	1	**1**	**1**	**1**	**1**	1	1
Correctly Classified %	98.15	100	**100**	**100**	**100**	**100**	99.98	100
Incorrectly Classified %	1.85	0	**0**	**0**	**0**	**0**	0.02	0

Table 3. Association Rules Generated by Supervised and Unsupervised Machine Learning Algorithms

If number of audience is less than or equal to 99 then instances are classified under low interest label. If number of audience is smaller than or equal to 696 then it is medium interest group whereas it is higher than 696 then it indicates high interest
Total number of low interest group distinct play viewers are greater than total number of medium interest group distinct play viewers which is greater than total number of high interest group distinct play viewers
If number of audience is greater than 2819 then it indicates the play name "Hastalık Hastası" with the highest audience number whereas if the number of audience is less than or equal to 1 than it indicates the play "Hayal-i Temsil" with the lowest audience number.
If number of audience is greater than or equal to 3587.5 then it is adult category, If number of audience is greater than or equal to 3504 and less than 3587.5 then it is musical category whereas if If number of audience is greater than or equal to 123.5 and less than 124.5 then it is child category
The most interested category is the adult category which is later followed with musical and child categories
If number of audience is greater than or equal to 3320.5 then it is Harbiye Theatre whereas If number of audience is less than 1.5 then it is Kadiköy Theatre based on the data set used in analysis
The most crowded play took in Harbiye theatre whereas the least crowded audience took place in Kadiköy Theatre
If play category is child then number of audience is 62.85 If play category is adult then number of audience is 63.28 If play category is musical then number of audience is 100.96
In the cluster analysis Adult, Foreign play category, 21.34 audience number, low interest value, Sultangazi Theatre with "Uzlaşma" play are positioned in the first cluster, Musical, Foreign play category, 118.17 audience number, low interest value, Harbiye with "Bak Bizim Şarkımızı Çalıyorlar" play are positioned in the second cluster, Adult, Domestic play category, 62.91 audience number, low interest value, Üsküdar Theatre with "Kahvede Şenlik" play are positioned in the third cluster, Adult, Foreign play category, 154.69 audience number, low interest value, Ümraniye Theatre with "Nora Bir Bebek Evi" play are positioned in the fourth cluster.

In the cluster analysis Adult, Foreign play category, 21.34 audience number, low interest value, Sultangazi Theatre with "Uzlaşma" play are positioned in the first cluster, Musical, Foreign play category, 118.17 audience number, low interest value, Harbiye with "Bak Bizim Şarkımızı Çalıyorlar" play are positioned in the second cluster, Adult, Domestic play category, 62.91 audience number, low interest value, Üsküdar Theatre with "Kahvede Şenlik" play are positioned in the third cluster, Adult, Foreign play category, 154.69 audience number, low interest value, Ümraniye Theatre with "Nora Bir Bebek Evi" play are positioned in the fourth cluster.

Based on the analysis conducted it was found out that, if number of audience is less than or equal to 99 then instances are classified under low interest label. If number of audience is smaller than or equal to 696 then it is medium interest group whereas it is higher than 696 then it indicates high interest. Total number of low interest group distinct play viewers are greater than total number of medium interest group distinct play viewers which is greater than total number of high interest group distinct play viewers. If number of audience is greater than 2819 then it indicates the play name "Hastalık Hastası" with the highest audience number whereas if the number of audience is less than or equal to 1 than it indicates the play "Hayal-i Temsil" with the lowest audience number. If number of audience is greater than or equal to 3587.5 then it is adult category, If number of audience is greater than or equal to 3504 and less than 3587.5 then it is musical category whereas if If number of audience is greater than or equal to 123.5 and less than 124.5 then it is child category. The most interested category is the adult category which is later followed with musical and child categories. If number of audience is greater than or equal to 3320.5 then it is Harbiye Theatre whereas if number of audience is less than 1.5 then it is Kadiköy Theatre based on the data set used in analysis. The most crowded play took in Harbiye theatre whereas the least crowded audience took place in Kadiköy Theatre. If play category is child then number of audience is 62.85. If play category is adult then number of audience is 63.28. If play category is musical then number of audience is 100.96.

In the cluster analysis Adult, Foreign play category, 21.34 audience number, low interest value, Sultangazi Theatre with “Uzlaşma” play are positioned in the first cluster, Musical, Foreign play category, 118.17 audience number, low interest value, Harbiye with “Bak Bizim Şarkımızı Çalıyorlar” play are positioned in the second cluster, Adult, Domestic play category, 62.91 audience number, low interest value, Üsküdar Theatre with “Kahvede Şenlik” play are positioned in the third cluster, Adult, Foreign play category, 154.69 audience number, low interest value, Ümraniye Theatre with “Nora Bir Bebek Evi” play are positioned in the fourth cluster.

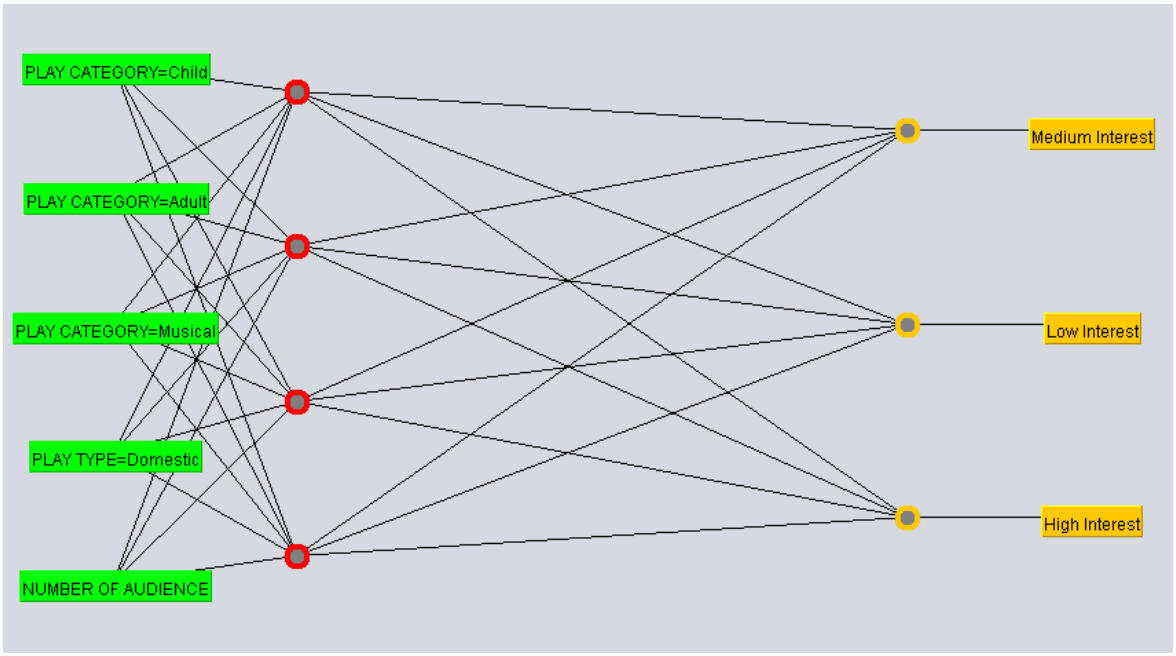

Figure 4. A Neural Network View of the Model Generated (Multi-Layer Perceptron- Interest Rate Indicator is the dependent variable)

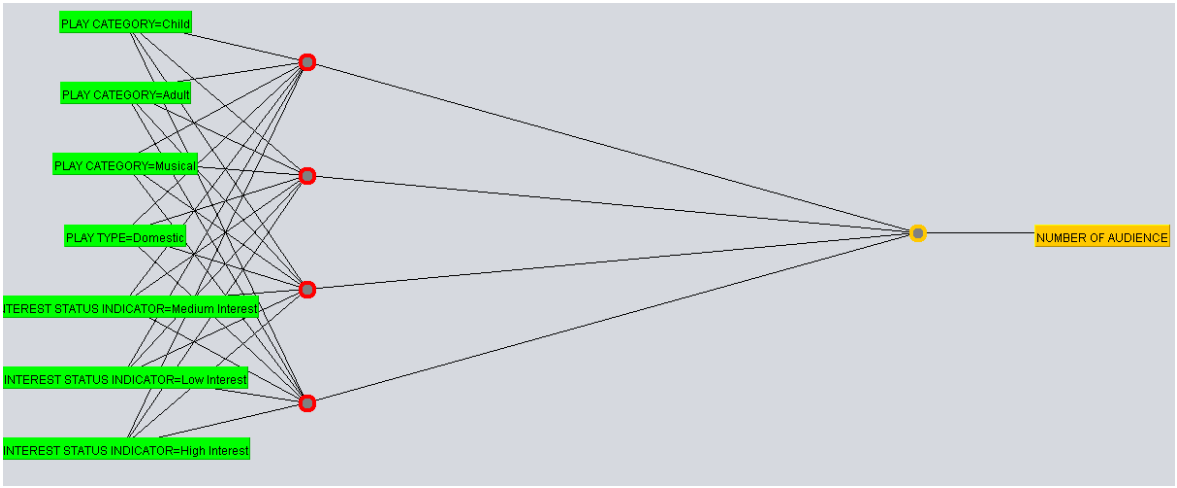

Figure 5. A Neural Network View of the Model Generated (Multi-Layer Perceptron- Number of Audience is the Dependent Variable)

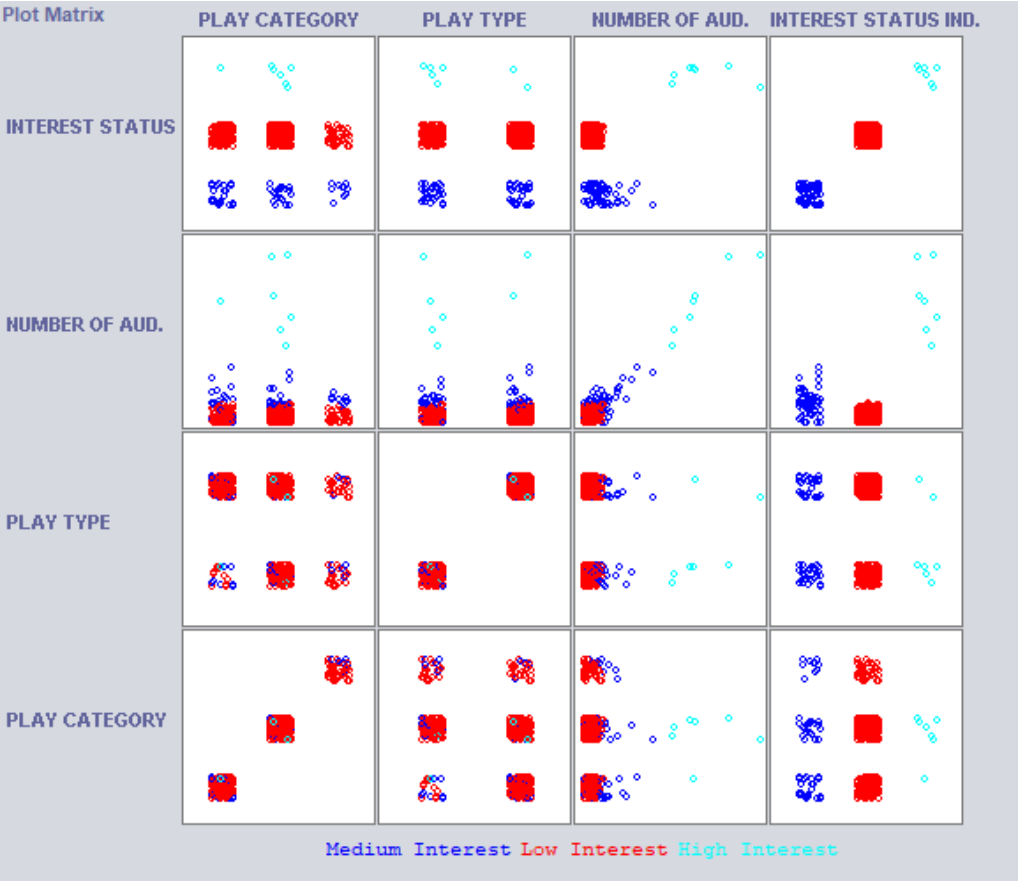

Figure 6. A Plot Matrix View of the Sampling

In the analysis J48, JRip, Part, OneR Method have been the top performing algorithms among other supervised machine learning approaches applied. All techniques generated %100 correct classification rate, 0 RMSE rate and a precision of 1 as the choosen performance metrics and indicators as suggested in literature. To sum up, supervised and unsupervised machine learning algorithms, which are also known as classification and clustering techniques in data mining literature can be used as an effective and efficient tool for knowledge discovery or confirmation in exploratory and confirmatory research designs. These insights may be considered by decision makers and society at large in such problem sets and domains. Based on the input loads, algorithmic design, architecture and performance of the algorithm which can be assessed with approximations, metrics as Big O or Big Ω which are used to assess the efficiency and the computational complexity [46,47,48,49,50,51, 70].

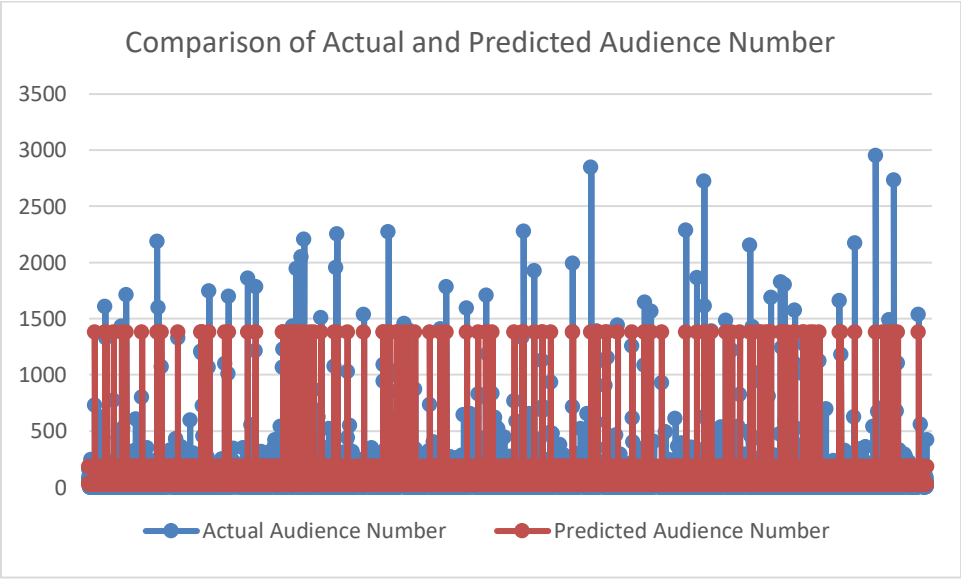

Figure 7. A Comparison of Actual and Predicted Audience Numbers for several plays

3. Discussion and Conclusion

Fine arts education constitutes an important value for the development of nations, countries and geographies. In literature it is seen that societies and nations that give importance to fine arts, other art forms with respective education programs and agendas seem to be more developed in the fierce competitive environments. It is a known fact that art forms enhance the creativity of individuals as well as developing questioning skills and empathy. Therefore cultural plans, programs and agendas handled in the cultural ministry or similar levels should always be considered as a chance for investments made for the development of individuals and society at large. Therefore enough budgets for cultural programs should be maintained. Academic and pre-university degree art programs should be designed and formed in response to the changes and developments in the contemporary art world. Graduates should also be prepared in these programs to the career paths that they want to follow in these areas. Understanding the domain of art marketing as well as implementing marketing strategies, plans, programs in every level in this context can be helpful in enlarging the arts market which means creating more jobs and employment opportunities for more actors and actresses. Qualification of these talents should also be supported with academical or nonacademical artistic education forms. In this study theatres located in Istanbul Municipality and theatre participation behavior has been analyzed with recent machine learning trends and approaches available in data mining literature with the aim of enhancing exploratory and confirmatory understanding in such domains as theater, entertainment and art marketing with the aim of providing some insights and ideas to the arts community, leaders and society at large.

Marketing science has been in the focus of many businesses, industries and sectors for many decades. In this context, marketing scholars and practioners have long worked on building good relationships with the customers applying contemporary, state of the art customer centric marketing approaches. As a result marketing found place and gained popularity in many areas as in goods marketing, services marketing, political marketing, celebrity marketing, entertainment and art marketing. In this study some of the factors associated with the theaters which has known to be one of the most influential forms of arts in human history. In this context theatres located in Istanbul Municipality and theatre participation behavior has been analyzed with recent machine learning trends and approaches available in data mining literature with the aim of enhancing exploratory and confirmatory understanding in such domains as theater, entertainment and art marketing.

Several machine learning techniques of data mining approach revealed that, if number of audience is less than or equal to 99 then instances are classified under low interest label. If number of audience is smaller than or equal to 696 then it is medium interest group whereas it is higher than 696 then it indicates high interest. Total number of low interest group distinct play viewers are greater than total number of medium interest group distinct play viewers which is greater than total number of high interest group distinct play viewers. If number of audience is greater than 2819 then it indicates the play name “Hastalık Hastası” with the highest audience number whereas if the number of audience is less than or equal to 1 than it indicates the play “Hayal-i Temsil” with the lowest audience number. If number of audience is greater than or equal to 3587.5 then it is adult category, If number of audience is greater than or equal to 3504 and less than 3587.5 then it is musical category whereas if If number of audience is greater than or equal to 123.5 and less than 124.5 then it is child category. The most interested category is the adult category which is later followed with musical and child categories. If number of audience is greater than or equal to 3320.5 then it is Harbiye Theatre whereas if number of audience is less than 1.5 then it is Kadiköy Theatre based on the data set used in analysis. The most crowded play took in Harbiye theatre whereas the least crowded audience took place in Kadiköy Theatre. If play category is child then number of audience is 62.85. If play category is adult then number of audience is 63.28. If play category is musical then number of audience is 100.96. In the cluster analysis Adult, Foreign play category, 21.34 audience number, low interest value, Sultangazi Theatre with “Uzlaşma” play are positioned in the first cluster, Musical, Foreign play category, 118.17 audience number, low interest value, Harbiye with “Bak Bizim Şarkımızı Çalıyorlar” play are positioned in the second cluster, Adult, Domestic play category, 62.91 audience number, low interest value, Üsküdar Theatre with “Kahvede Şenlik” play are positioned in the third cluster, Adult, Foreign play category, 154.69 audience number, low interest value, Ümraniye Theatre with “Nora Bir Bebek Evi” play are positioned in the fourth cluster.

In the analysis J48, JRip, Part, OneR Method have been the top performing algorithms among other supervised machine learning approaches applied. All techniques generated %100 correct classification rate, 0 RMSE rate and a precision of 1 as the choosen performance metrics and indicators as suggested in literature. To sum up, supervised and unsupervised machine learning algorithms, which are also known as classification and clustering techniques in data mining literature can be used as an effective and efficient tool for knowledge discovery or confirmation in exploratory and confirmatory research designs. These insights may be considered by decision makers and society at large in such problem sets and domains. Based on the input loads, algorithmic design, architecture and performance of the algorithm which can be assessed with approximations, metrics as Big O or Big Ω which are used to assess the efficiency and the computational complexity.

References

[1] https://library.harvard.edu/collections/harvard-theatre-collection

[2] https://plato.stanford.edu/entries/theater/

[3] https://www.gsb.stanford.edu/insights/why-does-movie-popcorn-cost-so-much

[4] https://arts.stanford.edu/directory/theater-performance/

[5] https://en.wikipedia.org/wiki/Theatres_Act_1968

[6] https://en.wikipedia.org/wiki/History_of_theatre

[7] https://www.timeout.com/newyork/theater/best-plays-of-all-time

[8] Goldstein R.J. (1989) Political Censorship of the Theatre. In: Political Censorship of the Arts and the Press in Nineteenth-Century Europe. Palgrave Macmillan, London. https://doi.org/10.1007/978-1-349-20128-0_4

[9] Tamarchenko A. Theatre Censorship. Index on Censorship. 1980;9(4):23-28.

[10] Moise, Daniel & Bogdan, Georgescu & Daniel, Zgură. (2012). The Use of Event Marketing Management Strategies. Procedia - Social and Behavioral Sciences. 46. 10.1016/j.sbspro.2012.06.448.

[11] Hackley, Chris & Hackley, Rungpaka Amy. (2006). Entertainment Marketing and Experiential Consumption. Journal of Marketing Communications. 12. 63-75. 10.1080/13527260500358608.

[12] Foutz, Natasha. (2017). Entertainment Marketing. Foundations and Trends® in Marketing. 10. 215-333. 10.1561/1700000049.

[13] Ulker-Demirel, Elif & Akyol, Ayşe & Şimşek, Gülhayat. (2018). Marketing and consumption of art products: The movie industry. Arts and the Market. 8. 00-00. 10.1108/AAM-06-2017-0011.

[14] Nasir, Suphan & Öcal, Figen. (2016). Film Marketing: The Impact of Publicity Activities on Demand Generation. 10.4018/978-1-5225-0143-5.ch019.

[15] Horvath, Adam & Gyenge, Balázs. (2020). Service Marketing Analysis Of Movies As Unique Service-Products. SocioEconomic Challenges. 4. 44-55. 10.21272/sec.4(2).44-55.2020.

[16] Kerrigan, Finola & Fraser, Peter & Ozbilgin, Mustafa. (2007). Arts Marketing. 10.4324/9780080472058.

[17] Lee, Jin Woo & Lee, Soo Hee. (2017). "Marketing from the Art World": A Critical Review of American Research in Arts Marketing. The Journal of Arts Management Law and Society. 47. 17-33. 10.1080/10632921.2016.1274698.

[18] https://hbr.org/2021/01/will-movie-theaters-survive-when-audiences-can-stream-new-releases

[19] https://www.independent.co.uk/arts-entertainment/films/features/40-greatest-best-plays-ever-written-theatre-drama-scripts-shakespeare-a8778031.html

[20] https://www.ranker.com/list/best-playwrights-ever/ranker-books

[21] https://www.hbs.edu/faculty/research/publications/Pages/default.aspx?q=Film%20Entertainment

[22] https://www.hbs.edu/faculty/research/publications/Pages/default.aspx?q=Hollywood

[23] https://hbr.org/2013/03/advertising-is-an-art-and-a-science

[24] https://www.hbs.edu/faculty/research/publications/Pages/default.aspx?q=Arts%20Marketing

[25] Blackmore, K., Bossomaier, T., 2002. Comparison of See5 and J48.PART algorithms for missing persons profiling. International Conference on Information Technology and Applications

[26] Frank E. and Witten I.H. (1998). Generating Accurate Rule Sets Without Global Optimization. In Shavlik, J., ed., Machine Learning: Proceedings of the Fifteenth International Conference, Morgan Kaufmann Publishers.

[27] Frank E. and Witten I.H. (2000). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann Publishers: San Francisco, CA.

[28] Lemeshow S., Hosmer D.W., Klar J. & Lwanga S.K., 1990. Adequacy of sample size in health studies. Chichester: John Wiley and Sons.

[29] https://pubs.niaaa.nih.gov/publications/aa72/aa72.htm

[30] Ramchoun, H. r., Idrissi, M. m., Ghanou, Y. y., & Ettaouil, M. m. (2017). New Modeling of Multilayer Perceptron Architecture Optimization with Regularization: An Application to Pattern Classification. IAENG International Journal of Computer Science, 44(3), 261-269.

[31] Rosenblatt, F., & Cornell Aeronautical Laboratory. (1958). The perceptron: A theory of statistical separability in cognitive systems (Project Para). Buffalo, N.Y: Cornell Aeronautical Laboratory.

[32] Shearer, C., 2000 The CRISP-DM model: the new blueprint for data mining. Journal of Data Warehousing, 5, 13-22.

[33] Simoudis, E. (1996). Reality Check for Data Mining. IEEE EXPERT, 11(5), pp.26-33

[34] Cohen, W. (1995). Fast effective rule induction. In A. Prieditis and S. Russell (eds.), Proceedings of the 12th International Conference on Machine Learning, Lake Tahoe, CA, pp.115-123.

[35] Saravanan, N., Gayathri V., 2018. Performance and classification evaluation of J48 algorithm and Kendall's based J48 algorithm (KNJ48). International Journal of Computer Trends and Technology

[36] Sasaki M., Kita K., 1998. Rule based text categorization using hierarchical categories, IEEE

[37] Edmondson, Amy C., and Tiona Zuzul. "Quantitative and Qualitative Methods in Organizational Research." In The Palgrave Encyclopedia of Strategic Management. Continuously updated edition, edited by Mie Augier and David J. Teece. Palgrave Macmillan, 2017. Electronic. (Pre-published, October 2013.)

[38] Taniguchi M., Haft M., Hollm´en J., and Tresp V. (1998). Fraud detection in communications networks using neural and probabilistic methods. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'98), Volume II, pp. 1241-1244.

[39] Venkatesan, E. V., 2015. Performance Analysis of Decision Tree Algorithms for Breast Cancer Classification. Indian Journal of Science and Technology.

[40] Ławrynowicz, A., Tresp, V., 2014. Introducing Machine Learning. Perspectives on Ontology Learning. AKA Heidelberg /IOS Press.

[41] Thomas, M., 2012. Root Mean Square Error Compared to, and Contrasted with, Standard Deviation. Surveying and Land Information Science, 72.

[42] Ławrynowicz, A., Tresp, V., 2014. Introducing Machine Learning. Perspectives on Ontology Learning. AKA Heidelberg /IOS Press.

[43] Thomas, M., 2012. Root Mean Square Error Compared to, and Contrasted with, Standard Deviation. Surveying and Land Information Science, 72.

[44] https://plato.stanford.edu/entries/aristotle-logic/

[45] Ajzen, I. & Fishbein, M., 1980. Understanding attitudes and predicting social behaviour. Englewood Cliffs, NJ: Prentice Hall

[46] Yavuz Ö., 2019, A data mining approach for desire and intention to participate in virtual communities. International Journal of Electrical and Computer Engineering, 9(5).

[47] Karahoca D., Karahoca A., Yavuz Ö., 2013. An early warning system approach for the identification of currency crises with data mining techniques. Neural Computing and Applications, 23(7-8)

[48] Yavuz, Ö., 2018. Marketing Implications Of Participative Behavior In Virtual Communities, Bahcesehir University Graduate School of Social Sciences, Management-Marketing Program, Istanbul

[49] Yavuz, Ö., 2009. An early warning system approach for the identification of currency crises, Bahcesehir University Graduate School of Sciences, Computer Engineering Graduate Program, Istanbul

[50] Yavuz, Ö. (2021). A Public Perceptions Analysis With Data Mining Algorithms, 2. International "Başkent" Congress On Physical, Social and Health Sciences, Ankara

[51] Yavuz, Ö. (2021). A Data Mining Analysis of Coronavirus Cases and Vaccinations in The City of London. Astana, Ankara.

[52] Kotler, Philip. Principles of Marketing. Englewood Cliffs, N.J. :Prentice Hall, 1991.

[53] Kotler, Philip. Marketing Management. Upper Saddle River, N.J. :Prentice Hall, 2000.

[54] Halldórsdóttir, S. (2000). The Vancouver School of doing Phenomenology. In: Fridlund, B. and Hildingh, C. (eds.) Qualitative research methods in the service of health. Lund: Studentlitteratur, pp. 47-84.

[55] Smith, J.D.. (2012). Single-Case Experimental Designs: A Systematic Review of Published Research and Current Standards. Psychological Methods. 10.1037/a0029312.

[56] Olsen, Wendy. (2004). Triangulation in social research: Qualitative and quantitative methods can really be mixed. Developments in sociology. 20. 103–118.

[57] Kuluchumila, Revocatus. (2018). Researching Education leadership in Tanzanian Secondary Schools.

[58] Rasmussen, C. E.; Williams, C. K. I. Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning); The MIT Press: 2005.

[59] http://old.opentox.org/dev/documentation/components/gaussianregressions

[60] https://en.wikipedia.org/wiki/Gaussian_process

[61] Malhotra, Naresh K. Marketing Research: An Applied Orientation. Upper Saddle River, NJ: Pearson/Prentice Hall, 2007.

[62] https://towardsdatascience.com/entropy-how-decision-trees-make-decisions-2946b9c18c8

[63] https://en.wikipedia.org/wiki/Feedforward_neural_network

[64] https://en.wikipedia.org/wiki/Backpropagation

[65] https://brilliant.org/wiki/backpropagation/

[66] https://en.wikipedia.org/wiki/Loss_function

[67] https://en.wikipedia.org/wiki/Error_function

[68] https://en.wikipedia.org/wiki/Gradient_descent

[69] Bryman, A. (2008). Social Research Methods. New York: Oxford University Press.

[70] https://cs50.harvard.edu/ap/2020/assets/pdfs/computational_complexity.pdf