The Analysis on the usage of the Video Conferencing Rooms using Classification

Arfa Hassan[1] Salma Aftab[2] Raheela Khan[3] Hira Asim[4]

Abstract

The ways and means of collaboration have been changed over the last few decades. They have extended beyond interaction within the same meeting room. Nowadays, all multinationals are installing video conference rooms in their offices globally in order to collaborate with their clients online to save travel cost and time. These video conference rooms are meant to capture the huge amount of data. Keeping in view the growth of the data in this situation, we performed an analysis on the usage of video conferencing rooms using data mining techniques. The data have been taken from a Norway based company named Cyviz . The data set is then further preprocessed and analyzed. Data reduction and data transformation have been done on the selected attributes to get better and appropriate results. A well-known data mining tool named WEKA[5] is used to perform the classification on the dataset taken into consideration. Classification algorithms named Naïve Bayes and Random Tree are applied to the dataset after preprocessing and their results are compared and analyzed. This study is an effort to analyze the usage of the video conference room so that appropriate usage of the resources can be ensured.

Keyword: Video Conference, Audio Conference, Naive Bayes, Random Tree, Classification, Data Mining, WEKA

1 Introduction

Cyviz is a technological organization which is based on research and development. They are having 120 employees who are operating globally. It has customers in 50 countries around the world. Cyviz started back in 1999 and with its 20 years of experience they now develop and produce softwares and hardwares that are used in collaboration systems and visualization. This comprises huge display walls with high resolution and collaboration rooms that enable consumers a proficient use of display walls. Previously, the users had ability to use display walls for presentations and high resolution videos while Cyviz introduced the flexible control system which is called CDC (Cyviz Display Controller) for the concept of the collaboration rooms. It enables consumers to utilize the same display wall for sharing presentations and video conferencing with other users. In advance, there was an only one point of control for display walls.

The CDC enables to configure the admittance of numerous consumers to control the system. The 20 years of focus and dedication on the technology of large displays has given

[1]*University of Management and Technology, Lahore | f2017114004@umt.edu.pk*
[2]*University of Management and Technology, Lahore | s2018114003@umt.edu.pk*
[3]*University of Management and Technology, Lahore | s2018179006@umt.edu.pk*
[4]*University of Management and Technology, Lahore | hira.asim@umt.edu.pk*

[5]*WEKA Website: https://www.cs.waikato.ac.nz/ml/weka/*

Cyviz a unique perception which enables them to understand the requirements of different businesses. In order to improve the portfolio of their product the company is determinedly spending in new technology [1].

Cyviz has provided meeting rooms to the Fortune 500 and the government customers. These meeting rooms have different features and they are installed according to the customer's requirement. These features are providing information about the usability of the conference rooms. The features of video conferencing rooms are like audio detected, room in use, the room booked through exchange, presence detected, picture in picture active, stereo on, audio or video conference active, audio call active, video call active and some others.

This study is based on the analysis of available features of the conference rooms and extracting meaningful information out of it.

2 Background Knowledge

Video conferencing is the means of communication using a combination of audio, video, text and graphics to support real time communication between distributed groups sharing same interests or working in the same domain like business meetings, playing games, learning and entertainment [4]. Our study is a statistical analysis of a system using video conferencing determining the availability and non-availability of rooms using the features provided by the organization. Weka is an application that runs almost on every system and is developed using JAVA. It provides an interface to multiple algorithm and also support pre and post processing of data to extract results from different data sets [4]. Random trees are used to predict results using multiple decision trees that grow in different subsets in the same domain, the idea was proposed by Leo Breiman in 2000 [5]. Naïve Bayes is a simple classifier that calculates the probabilities of frequency and uses different combination of values from a data set, the algorithm is helpful in supervised learning [6]. Depending upon our best knowledge and research study, we tried to analyze the data set received from Cyviz using above mentioned data mining techniques in order to find meaningful knowledge.

3 Methodology and Results

The machine learning view of Knowledge Data Discovery (KDD), shown in the Figure 1, is followed in the whole research process.

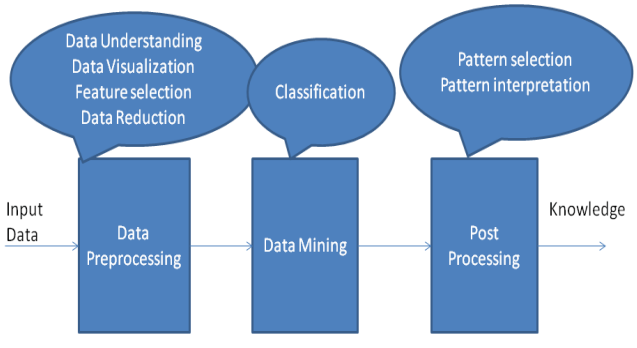

Figure 1: KDD process a machine learning view

A Data Preprocessing

The phase is handling the data manipulation Procedure of data preprocessing consists of data understanding, data visualization, feature selection and data reduction.

1) Data Understanding

The research started from getting familiarity and understanding of the available data. The initial data provided for analysis were having around 30 features of the video conference rooms. These features were capturing the usage time in number of seconds, more precisely the number of seconds each feature had been used in any particular room. For example, in one transaction for any particular date if Audio conference feature is having 3600 value, then the room used the feature of audio conference for an hour. The features were studied thoroughly to figure out their usage and any possible relation.

2) Data Visualization

In order to get more understanding of the data all available features were plotted in MATLAB. Some of the plots that are understandable are shown in Figure 2, Figure 3 and Figure 4.

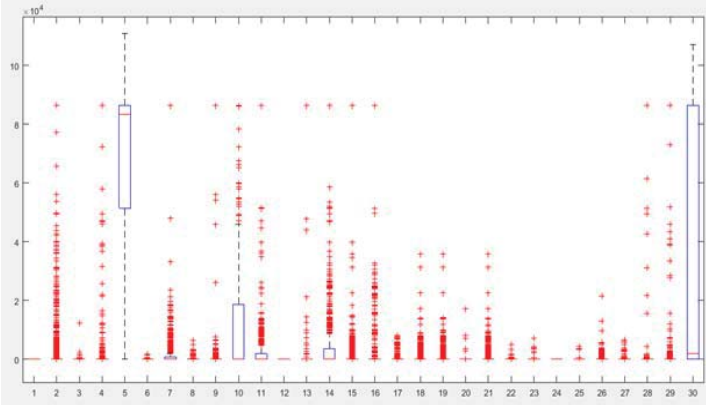

Figure 2: Boxplot showing all features of the video conference room

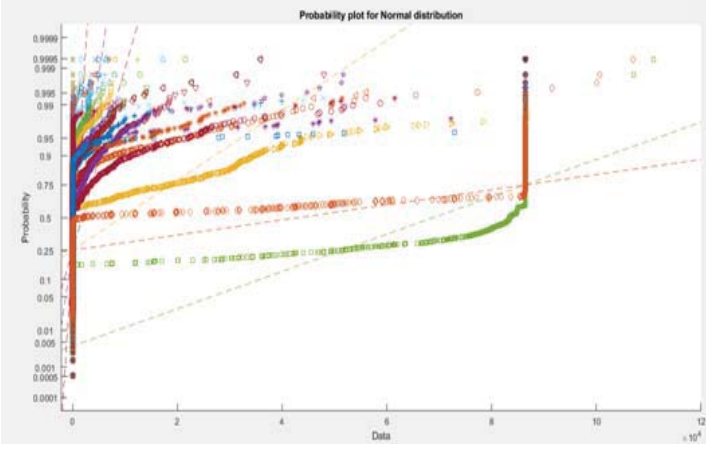

Figure 3: Probability plot for Normal distribution

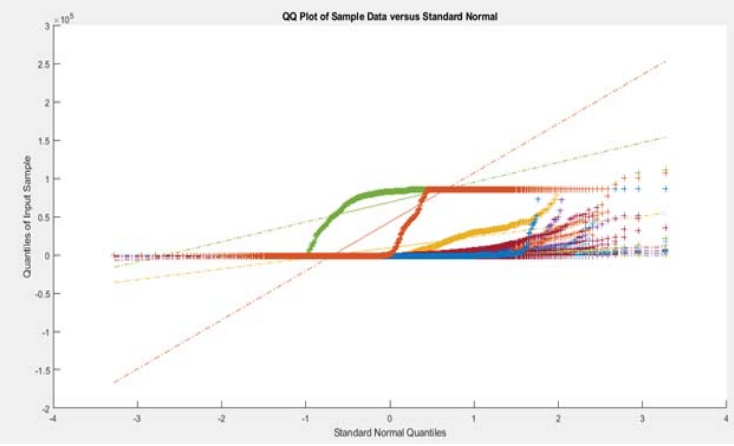

Figure 4: QQ Plot of sample data versus standard Normal

3) Data Reduction

From the visualization of data, it was figured out that in order to understand the patterns which were more informative, it was important to reduce the data size by selecting some maximum information giving features. The selected features of the study are shown in Table 1.

Table 1: The selected feature of video conference room

Feature	Description
Room Name	The name of the room. The name uses name space notation, showing the region, country and city of that room.
Day	The day of the week on which this information is recorded. The data is captured against particular date and day is derived from that date.
Is Room In Use	The number of seconds for which room was in use on a particular day
Is Pip Active	The number of seconds for which picture in picture was active on a particular day
Is Audio Video Conference Active	The number of seconds for which audio and video conference both were active. It has sum of Audio and video active time collectively on a particular day

4) Statistical Description

The statistical description of selected features is shown in the Table 2 below.

Table 2: The usage data in number of seconds

Description	Is Room In Use	Is Pip Active	Is Audio Video Conference Active
Mean	5114,80129	2627,081	1500,87895
Median	0	0	0
Mode	0	0	0
Standard Error	349,335813	229,8189	201,5143526

Standard Deviation	15094,3797	9930,199	8707,192431
Minimum	0	0	0
Maximum	89999	86400	86400

The average use of all rooms together is 85 minutes.

Mean: = 5114,80129/60 = 85 minutes

The average use of Picture in picture for all rooms is 44 minutes.

Mean: = 2627,081/60 = 44 minutes

The average use of audio and video conference feature for all rooms is 25 minutes

Mean: 1500,87895/60 = 25 minutes

5) Data Transformation

In order to apply data mining algorithms on all the selected features, they were transformed. The transformed data are shown in Table 3.

Table 3: The data before and after transformation

Feature	Actual Data	Transformed Data
Room Name	Number of seconds	Nominal
Day	Day against each date	Day
Room Used	Number of seconds	Boolean
Picture in picture	Number of seconds	Boolean
Audio or video conference	Number of seconds	Boolean

The main objective of data transformation was to generate meaningful knowledge after applying data mining techniques. The rules that were followed for transformation are defined in Table 4.

Table 4: The rules set for data transformation

Feature	Actual Data	Transformed Data
Room Used	= 0 > 0 &<= 120 mins >120 &<= 360 mins >360 mins	No Usage Low Usage Normal High
Picture in Picture	= 0 > 0	N Y
Audio or Video conference	= 0 > 0	N Y

Table 4 above is giving us the following insights about the features we took into consideration:

a) Room used

If there is no meeting at all then recorded data is 0 seconds, which is transformed into No usage of room. If the room is used for 1 minute to 2 hours (120 mins) then the usage of the room is Low. If the room is used for 2 hours to 6 hours (360 mins) then the usage is Normal. For more than 6 hours the usage is transformed into High.

b) Picture in Picture sharing

If the data of picture in picture sharing feature are 0 then it is transformed into N otherwise it is Y.

c) Audio and video conference

If the data of audio and video conference feature is 0 then it is transformed into N otherwise it is Y.

B Data Mining

1) Data Sample

There are 6 attributes and Room used is the class attribute. The total number of transactions is 2449. The sample data is shown in Table 5.

Table 5: The sample dataset after preprocessing

Room	Day	Room Used	PIP	Audio Video	Usability
Room 1	Tuesday	No Usage	N	N	no usability
Room 1	Wednesday	Low	Y	Y	full utilized
Room 1	Thursday	Low	Y	N	min-use
Room 1	Monday	Normal	Y	Y	full utilized

2) Weka Tool for Classification

Weka is providing different machine learning algorithms for the tasks of data mining [4]. The research classifiers are implemented in it.

3) Naïve Bayes Classification

The Naviebayes classifier is based on Baye's theorem of posterior probability. It is assumed that features or attributes are independent of each other. The value of one attribute is independent of the other attribute of a class [2]. The summary of accuracy generated by Naive Bayes Classifier is shown in Table 6.

Table 6: The summary of accuracy by Naive Bayes

Correctly Classified Instances	1838	75.051%
Incorrectly Classified Instances	611	24.949%

The implementation of Naive Bayes Classifier is shown in the Figure 5.

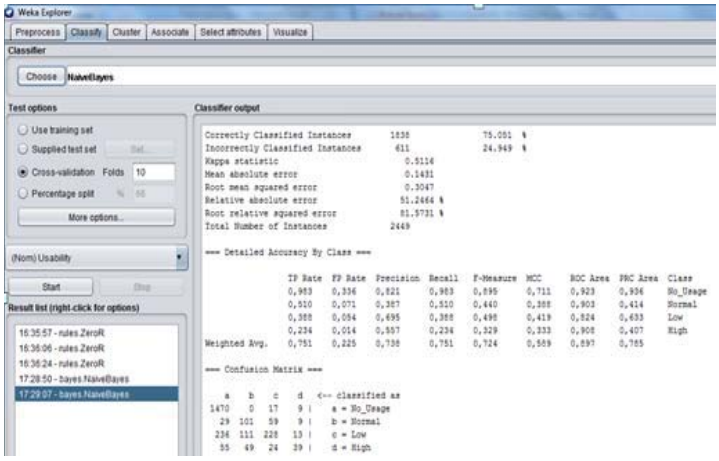

Figure 5: Implementation of Naive Bayes using WEKA

The confusion matrix using Naive Bayes classifier is shown in the Table 7.

Table 7: The confusion matrix using Naive Bayes classifier

A	B	C	D	Classified As
1470	0	17	9	A = No Usage
29	101	59	9	B = Normal
236	111	228	13	C = Low
55	49	24	39	D = High

4) Random Tree Classification

Random tree is an ensemble machine learning algorithm which generates numerous different learners. In order to produce a random set of data for constructing a decision tree it uses a bagging idea [4]. An alternative classifier is used in the research study to get better accuracy. The summary of accuracy generated by the random tree is shown in the Table 8.

Table 8: The summary of accuracy by random tree

Correctly Classified Instances	1878	76.684%
Incorrectly Classified Instances	571	23.316%

The implementation of Random tree classifier is shown in the Figure 6.

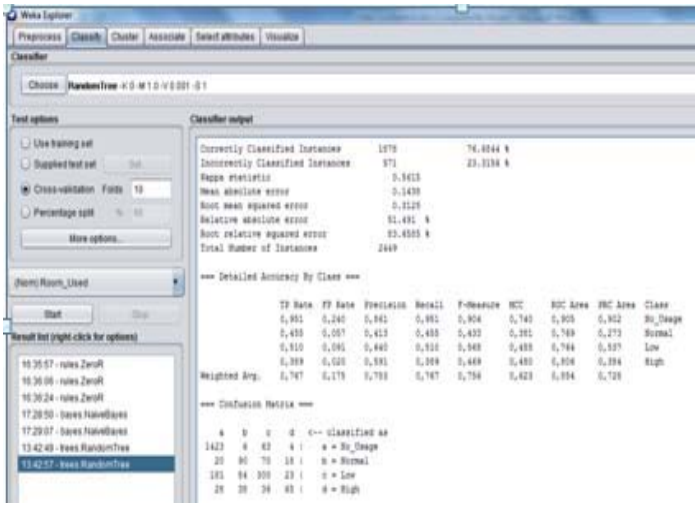

Figure 6: Implementation of Random Tree classifier using WEKA

The confusion matrix generated in WEKA using the random tree classifier is given below in Table 9.

Table 9: The confusion matrix using Random Tree classifier

A	B	C	D	Classified As
1423	6	63	4	A = No Usage
20	90	70	18	B = Normal
181	84	300	23	C = Low
28	38	36	65	D = High

C Post Processing

The proposed classifiers have divided the data into four categories No usage, low, medium and high room usage. They are giving two different accuracies and can predict the usage of the room on the basis of defined attributes.

4 Conclusion & Future Research Directions

The selected features of the video conference room are showing the level of utilization of that particular room. In future, work we are interested to figure out that on which day, week or a month meeting room is busy or it can predict about the availability of the room on any particular day or week. As the company is ready to provide more data for analysis so we have a room to expand our work in different directions. We also suggest the company to keep the record of the camera installed in the room in order to check the correct usage of a conference room. Through camera we can obtain image data and can relate this to the existing features.

Acknowledgment

We acknowledge the support of Cyviz for providing us their valuable data for analysis and the continuous guidance of Dr. Malik Tahir Hassan.

References

[1] Cyviz. (n.d.). Retrieved Febraury 01, 2019 from Cyviz: https://www.cyviz.com/about/

[2] Jiawei Han, M. K. (2012). Data Mining Concepts and Techniques. Waltham: Morgan Kaufmann.

[3] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.

[4] Kalmegh, S. R. (2015). Comparative analysis of weka data mining algorithm randomforest, randomtree and ladtree for classification of indigenous news data. International Journal of Emerging Technology and Advanced Engineering, 5(1), 507-517.

[5] Benedictus, F. (2011). Get more out of video conferencing (Master's thesis, University of Twente).

[6] Patil, T. R., &Sherekar, S. S. (2013). Performance analysis of Naive Bayes and J48 classification algorithm for data classification. International journal of computer science and applications, 6(2), 256-261.