opción

Revista de Antropología, Ciencias de la Comunicación y de la Información, Filosofía, Lingüística y Semiótica, Problemas del Desarrollo, la Ciencia y la Tecnología

Año 35, 2019, Especial N° 19

Revista de Ciencias Humanas y Sociales

ISSN 1012-1587/ ISSNe: 2477-9385

Depósito Legal pp 198402ZU45

Universidad del Zulia

Facultad Experimental de Ciencias

Departamento de Ciencias Humanas

Maracaibo - Venezuela

Opción, Año 35, Especial No.19 (2019): 1368-1393
ISSN 1012-1587/ISSNe: 2477-9385

Prediction of the Digital Game Rating Systems based on the ESRB

Khaled Mohammad Alomari[1],
[1]Faculty of Computer Sciences, Abu Dhabi University, Abu Dhabi, UAE
khaled.alomari@adu.ac.aeb

Ahmad Qasim AlHamad[1],
[1]Faculty of Computer Sciences, Abu Dhabi University, Abu Dhabi, UAE
aqd14@adu.ac.ae

Hisham o. Mbaidin[2]
[1]Faculty of Management Information Systems, Mutah University, Karak, Jordan
Hisham.mbaidin@adu.ac.ae

Said Salloum[3]
[3]Fujairah University, Fujairah, UAE
ssalloum@uof.ac.ae

Abstract

This study tries to find out the best model for prediction video game rate categories. A representation from four rating categories (everyone, everyone 10+, teen, mature) was used for the analysis. The paper follows CRISP-DM approach under Rapid Miner software to business and data understanding, Data preparation, model building and evaluation. The researchers compared prediction among six model and the results showed that the Generalized Linear Models (GLMs) achieved a best accuracy (0.9027), also results highlighted eight important content descriptions to have the highest influence on prediction.

Keywords: Video game, Digital Game Rating Systems, Machine Learning.

Predicción de los sistemas de clasificación de juegos digitales basados en el sistema de calificación ESRB

Resumen

Este estudio trata de encontrar el mejor modelo para las categorías de predicción de videojuegos. Para el análisis se utilizó una representación de

Recibido: 14-12-2019•Aceptado: 19-03-2019

cuatro categorías de calificación (todos, todos los mayores de 10 años, adolescentes, adultos). El documento sigue el enfoque CRISP-DM del software Rapid Miner para la comprensión de negocios y datos, preparación de datos, construcción de modelos y evaluación. Los investigadores compararon la predicción entre los seis modelos y los resultados mostraron que los Modelos lineales generalizados (GLM) lograron una mejor precisión (0.9027); los resultados también resaltaron ocho descripciones de contenido importantes para tener la mayor influencia en la predicción.

Palabras clave: Videojuego, Sistemas de Clasificación de Juegos Digitales, Aprendizaje Automático.

1. INTRODUCTION

In the wake emergence of the technological and communication revolution in the last quarter of the twentieth century and after the proliferation of “Globalization “, doubts and fears appeared among many peoples especially the developing communities on the impact of the modern technologies and media on cultures, traditions, heritage, and social structure (S.A. Salloum, AlHamad, Al-Emran, & Shaalan, 2018; Said A Salloum, Al-Emran, & Shaalan, 2017). Many studies have shown no communities are fully immune on the impacts of modern new technologies where it becomes an influential force on the global social structure. Because of its fast-rapid spread in our community, it becomes a source of information entertainment and cultural trait.

In the shadows of modern communities, information technology plays a huge impact on the members of the community. People find themselves automatically dealing with huge amount of information and managing their daily complex and varied problems through the usage of this technology (Said A Salloum, Al-Emran, & Shaalan, 2018). This technology provides them with the energy unprecedented thinking tool and the art of finding new modern and sophisticated solution (S. A. Salloum, Al-Emran, Monem, & Shaalan, 2017). In addition, technology gave the opportunity to the community to learn more and absorb new concepts to stay up-to-date with the current world, while children being sensitive and more adaptive to these changes where in many cases they become in competition between parents and educators in the socialization and education.

Perhaps the most important achievements of information technology are the advent of computers, Internet, mobile devices, games consoles, and tablet, which reshaped the child's life at home and school in unexpected and profound ways. Children of the ‘e-community’ are more susceptible to the pros and cons of this community. Computers push children to learn better, through the learning they found more efficient environment and while trying new technologies that makes them more prepared for the future. For children, learning information technology is essential for giving them tools of success at the future because in the modern world, technology plays a major role of knowledge provider without retreating to other fields.

2. BACKGROUND

2.1. Digital Game Effect

According to (Turner et al., 2012) as researchers, it is important to separate the hype from the reality. A fundamental issue for the study of problems related to digital gaming is that there is no agreement upon definition of types of problems digital games produce and also, there is no clear boundaries between normal play and excessive play. The digital games today are spur to players to become part of it. It plays a big role in physical and emotional effect on the player. It took them to a deeper level in engagement and interaction (Norcia, 2014). According to ,McGonigal, 2011, American children spend long hours in playing video games compared to the time spent on their studying. Where in many casesin other places, the time spent on video games vs. studying were equal. This would result in a fact that by the time this American child gets to the age of 21 he or she would have spent 10000 hours on playing video games. Other statistics (latest statistics) showed that 170 million American are video gamers. According to , Greitemeyer & Osswald, 2009, pro-social video game playing led to a short-term reduction in the tendency to see the world as hostile and an immediate reduction in anti-social thoughts. The pro-social game would lead to increased helpful and decreased hurtful behavior, relative to violent games, with neutral games yielding intermediate behaviors (Saleem, Anderson, & Gentile, 2012). Video games can also include opportunities for self-assessment and are often becoming important social learning environments that allow for additional learning from different perspectives (McLean & Griffiths, 2013). Some researchers have argued that video games are the “training wheels” for computer literacy (Gentile & Anderson, 2006); (McLean & Griffiths, 2013). Video game helps to improve perceptual skills and visual attention task regardless of previous video game experience. Other studies have documented relations between video game play and visual selective attention, mental rotation, spatial visualization, and reaction times (Gentile & Anderson, 2006). Video games can also provide opportunities for practice in different directions and in the use of fine motor skills (Gentile

& Anderson, 2006). In spite of the benefits that may be in some video games, however, drawbacks in the horizon arise, because most of the games used by children and adolescents with negative implications affecting them at all stages of growth. In addition, a large percentage of video games depends on the entertainment and enjoy killing others, destroying property and assaulting them unjustly. Consequently, children and adolescents methods of committing the crime and their arts, tricks, capabilities and skills of violence, aggression, and the outcome of crime developed in their minds. These capabilities acquired through the exercise of intimacy in those games (Weiss, 2010). Excess synapses during early adolescence in emotionally laden situations may lead to the increased aggression of early adolescence (more so for boys than girls) (Kirsh, 2002).

Researchers have disagreed in regard to whether digital games such as Mortal Kombat, Splatterhouse, and Grand Theft Auto have the capacity to ‘blur’ the social boundaries between good and evil and to invoke catastrophic effects through subsequent violent acts by gamers (Jaslow, 2013). According to ,Engelhardt, Bartholow, Kerr, & Bushman, 2011, it is confirmed that violent digital games with apocalyptic themes desensitize the gamer to the violent graphics over the short term; and that games such as Grand Theft Auto, Killzone and Hitman caused the participants to act more aggressively than the participants who played non-violent games. The study in which 70 participants played either violent or non-violent digital games for 25 minutes, and after which measurements were made on the participants’ brain activities which supported that the violent content and interaction, enhanced the neural desensitization to violence. According to (Engelhardt et al., 2011) the neural desensitization also served as a predictor to increased aggressive behaviors after repeated exposure to the game violence. However, Massachusetts Institute of Technology (MIT) Director, Henry Jenkins (Jenkins, 2006) argued that games are “merely games” and that the youth are well aware of the difference between game rules and rules that apply in the real world. Further, (Jenkins, 2006) supported that the youth who transfers the tragedy in digital game content to real-world tragedies are “severely emotionally disturbed”.

3. LITERATURE REVIEW

3.1.Prediction Models

Reliable outputs from predictive analyses of big data are of tremendous value to business analysts forecasting tools. Statistical models are typically used to build relationships and to test them in order to defend theories of causal relationships that are used as a basis for prediction. (Hee, 1966) outlined the importance of both qualitative and quantitative testing of the validity of forecasting model formulas and values. (Vucevic & Yaddow, 2012) supported that game design approaches that are unorganized and that are not clearly defined prevent the prediction and repetition of the quality of the product.

The need to test the validity of theoretical scientific proclamations has been well served by predictive modeling and testing methods which identify existing causal mechanisms and discover differentiations in construction operations (Shmueli, 2010). The predictive model may reveal potential improvements to existing models and bridge the gap between theoretical assumptions and application. Classical models for probability and prediction were developed prior to the information technology revolution and pre-programmed algorithms. Furthermore, the quantification of predictability within the model is commonly achieved by a statistical evaluation of distinct features through both explanatory, descriptive, and predictive approaches (Kumalasari, 2019).

According to (Shmueli, 2010) the predictive model is considered a tool for applications of data mining in order to make new and future predictions. The approach to making predictions may be Bayesian, frequentists, parametric, or statistical models. Statistical predictions within dynamic systems fit the Bayesian model; when time and relative system state are not referenced, the distribution selection is generated by equilibrium distribution. Model-to-data fit may be measured by a prior and posterior predictive checks or from a hybrid of checks designed for the hierarchical model (Gelman, Hwang, & Vehtari, 2014). The most common prediction model is the multiple linear prediction models which utilizes predictor and response variables.

The primary objective of logistic regression is the modeling of some probability. More specifically, predictive modeling is used to forecast unknown values based on other values or attributes that are known. The learning algorithm is used to generate a dependable rule for the prediction of probable outputs for future data (Friedman, 1997). In the case of non-stochastic prediction, the objective is the prediction of output (Y) for any new observations with input values (X), and observations to time (t) is the basis for forecasting future values at a time t + k, where k > 0 (Geisser, 1993).

4. METHODOLOGY

The methodology for this study is based on the assumptions and structures of the CRISP-DM to prediction models for digital game rating. The initial data is collected from large datasets of ESRB rating system; and is prepared, modeled and evaluated based upon the CRISP-DM general tasks. The video game data is extracted and segmented are created in order to create the game rating prediction model.

The first step of the CRISP-DM model is to identify and clarify the business objective. The business objective for this study is to gain an understanding of video game rating and prediction models performance through the extraction of data from a large dataset for analysis. This section will describe sample data and outline the method of data analysis for 2053 video game titles rating that ware extracted from a listing ESRB ratings system in order to explore different prediction models that will identify the most significant model to predict a rating for video game titles.

4.1 .Date

The research will use a total of 2053 video game titles which were selected from the listing of ESRB ratings. The method of selection was to extract all video game under the PlayStation four platform until June 2018. The final 2053 title sample is comprised of representations from four rating categories (Everyone, Everyone 10+, Teen, Mature). A total of 2053 video game titles were used for the analysis along with the 38 titles of Content Descriptors as shown Table 1 and Classification Rate generated for each game. The data was stored in Excel Sheet where content on (Game Title, Rate, 38 Content Descriptors). The Content Descriptors convert to binary data through using 1 if exist content descriptors and 0 if not exist according to the game data in the Content Descriptors based on ESRB data.

Table 1 Content Descriptors

Alcohol Reference	Animated Blood	Blood	Blood and Gore	Cartoon Violence
Comic Mischief	Crude Humor	Drug Reference	Fantasy Violence	Intense Violence
Language	Lyrics	Mature Humor	Mild Blood	Mild Cartoon Violence
Mild Fantasy Violence	Mild Language	Mild Lyrics	Mild Suggestive Themes	Mild Violence
Nudity	Partial Nudity	Real Gambling	Sexual Content	Sexual Themes
Sexual Violence	Simulated Gambling	Strong Language	Strong Lyrics	Strong Sexual Content
Suggestive	Tobacco	Use of	Use of Alcohol	Use of

Themes	Reference	Alcohol	and Tobacco	Drugs
Use of Drugs and Alcohol	Use of Tobacco	Violence		

4.2.Transformation

The data was scraped from the ESRB, cleaned and trained based upon the steps of the CRISP-DM model, using the RapidMiner tool. The data which was stored on excel files uploaded as sources which were used to create a dataset that could be selected for the different types of analytics. The dataset was split randomly to 80% train and 20% validation set as shown in Table 2. Where validation set outputs are then used to predict game rate classification for 411 game titles.

Table 2 Dataset split

Rate	**Everyone**	**Everyone 10+**	**Teen**	**Mature**	**Total**
Train (0.8)	409	366	543	324	1642
validation (0.2)	102	92	136	81	411
Total	511	458	679	405	2053

i. *Generalized Linear Models (GLMs)*

Executes GLM algorithm using H2O 3.8.2.6.

Generalized linear models (GLMs) are an extension of traditional linear models that allows the specification of models whose response variable follows different distributions. It covers widely used statistical models, such as linear regression for normally distributed responses, logistic models for binary data, log linear models for count data, complementary log-log models for interval-censored survival data, plus many other statistical models through its very general model formulation.

The operator starts a 1-node local H2O cluster and runs the algorithm on it.

Table 3 Generalized Linear Model result

<table>
<tr><td rowspan="10">Generalized Linear</td><td></td><td>true E</td><td>true T</td><td>true M</td><td>true E10</td><td>FN</td></tr>
<tr><td>pred. E</td><td>101</td><td>1</td><td>0</td><td>5</td><td>6</td></tr>
<tr><td>pred. T</td><td>0</td><td>124</td><td>14</td><td>8</td><td>22</td></tr>
<tr><td>pred. M</td><td>0</td><td>1</td><td>67</td><td>0</td><td>1</td></tr>
<tr><td>pred. E10</td><td>1</td><td>10</td><td>0</td><td>79</td><td>11</td></tr>
<tr><td>ground truth</td><td>102</td><td>136</td><td>81</td><td>92</td><td></td></tr>
<tr><td>Predict</td><td>107</td><td>146</td><td>68</td><td>90</td><td></td></tr>
<tr><td>FP</td><td>1</td><td>12</td><td>14</td><td>13</td><td></td></tr>
<tr><td>sum</td><td>411</td><td></td><td></td><td></td><td></td></tr>
<tr><td>TP & TN</td><td>371</td><td></td><td></td><td></td><td></td></tr>
</table>

Table 4 illustrate the Accuracy, Precision, Recall, and F-score of each class in the Generalized Linear model.

Table 4 Generalized Linear Model categories result

	class	**E**	**T**	**M**	**E10**
Generalized Linear	TP	101	124	67	79
	TN	270	247	304	292
	FP	1	12	14	13
	FN	6	22	1	11
	Accuracy	0.981	0.916	0.961	0.939
	Precision	0.99	0.912	0.827	0.859
	Recall	0.944	0.849	0.985	0.878
	F-score	0.967	0.879	0.899	0.868

ii. Decision Tree

Decision tree employs the divide and conquer method and recursively divides a training set until each division consists of examples from one class. A general algorithm for decision tree building will include creating a root node and assign all of the training data to it, selecting the best splitting attribute, adding a branch to the root node for each value of the split, splitting the data into mutually exclusive subsets along the lines of the specific split and at the end a repetition is made for steps 2 and 3 for each and every leaf node until the stopping criteria is reached. A prediction for the class label Attribute is determined depending on the majority of Examples which reached this leaf during generation, while an estimation for a numerical value is obtained by averaging the values in a leaf.

This Operator can process Example Sets containing both nominal and numerical Attributes. The label Attribute must be nominal for classification and numerical for regression.

Table 5 illustrate the prediction using the Decision Tree Model and collocation the False Negative (FN), False Positive (FP), True Negative (TN) and True Positive (TP) to evaluation model.

Table 5 Decision Tree Model result

		true E	true T	true M	true E10	FN
Decision Tree	pred. E	101	1	0	6	7
	pred. T	0	126	14	11	25
	pred. M	0	2	67	0	2
	pred. E10	1	7	0	75	8
	Ground Truth	102	136	81	92	
	Predict	106	149	67	89	
	FP	1	10	14	17	
	sum	411				
	TP & TN	369				

Table 6 illustrate the Accuracy, Precision, Recall, and F-score of each class in the Decision Tree model.

Table 6 Decision Tree Model categories result

	Class	**E**	**T**	**M**	**E10**
Decision Tree	TP	101	126	67	75
	TN	268	243	302	294
	FP	1	10	14	17
	FN	7	25	2	8
	Accuracy	0.97878	0.913366	0.958442	0.936548
	Precision	0.990196	0.926471	0.82716	0.815217
	Recall	0.935185	0.834437	0.971014	0.903614
	F-score	0.961905	0.878049	0.893333	0.857143

iii. Deep Learning

Deep Learning is a machine learning technique that imitate the human experience so that the computer will appear as normal human. In deep learning, a computer model learns to perform classification tasks directly from images, text, or sound. In deep learning, models are trained with stochastic gradient descent using back-propagation. The network can contain a large number of hidden labeled layers consisting of neurons with tech, rectifier, and max out activation functions. It also requires very large computing power.

Table 7 illustrate the prediction using the Deep Learning Model and collocation the False Negative (FN), False Positive (FP), True Negative (TN) and True Positive (TP) to evaluation model.

Table 7 Deep Learning Model result

		true E	true T	true M	true E10	FN
Deep Learning	pred. E	99	1	1	5	7
	pred. T	0	126	14	9	23
	pred. M	0	1	66	0	1
	pred. E10	3	8	0	78	11
	ground truth	102	136	81	92	
	Predict	108	151	69	83	
	FP	3	10	15	14	
	sum	411				
	TP & TN	369				

Table 8 illustrate the Accuracy, Precision, Recall, and F-score of each class in the Deep Learning model.

Table 8 Deep Learning Model categories result

	Class	**E**	**T**	**M**	**E10**
Deep Learning	TP	99	126	66	78
	TN	270	243	303	291
	FP	3	10	15	14
	FN	7	23	1	11
	Accuracy	0.973615	0.91791	0.958442	0.936548
	Precision	0.970588	0.926471	0.814815	0.847826
	Recall	0.933962	0.845638	0.985075	0.876404
	F-score	0.951923	0.884211	0.891892	0.861878

iv. Gradient Boosted

A gradient boosted model is an ensemble of either regression or classification tree models. Both are forward-learning ensemble methods that obtain predictive results through gradually improved estimations. Boosting is a flexible nonlinear regression procedure that helps to improve the accuracy of trees. By sequentially applying weak classification algorithms to the incrementally changed data, a series of decision trees are created that produce an ensemble of weak prediction models. While boosting trees increases their accuracy, it also decreases speed and human interpretability. The gradient boosting method generalizes tree boosting to minimize these issues.

The operator starts a 1-node local H2O cluster and runs the algorithm on it. Although it uses one node, the execution in parallel.

Table 9 illustrate the prediction using the Gradient Boosted Model and collocation the False Negative (FN), False Positive (FP), True Negative (TN) and True Positive (TP) to evaluation model.

Table 9 Gradient Boosted Model result

		true E	true T	true M	true E10	FN
Gradient Boosted	pred. E	100	2	1	5	8
	pred. T	0	119	13	7	20
	pred. M	0	0	67	0	0
	pred. E10	2	15	0	80	17
	ground truth	102	136	81	92	
	Predict	108	139	67	97	
	FP	2	17	14	12	
	sum	411				
	TP & TN	366				

Table 10 illustrate the Accuracy, Precision, Recall, and F-score of each class in the Gradient Boosted model.

Table 10 Gradient Boosted Model categories result

	Class	E	T	M	E10
Gradient Boosted	TP	100	119	67	80
	TN	266	247	299	286
	FP	2	17	14	12
	FN	8	20	0	17
	Accuracy	0.973404	0.908189	0.963158	0.926582
	Precision	0.980392	0.875	0.82716	0.869565
	Recall	0.925926	0.856115	1	0.824742
	F-score	0.952381	0.865455	0.905405	0.846561

v. Random Forest

A collection of several random trees that are defined by the amount of trees parameter form up a random forest. The creation/training of the trees takes place at the Example Set's bootstrapped sub-sets available at the Input Port. The splitting rule for a particular Attribute is signified for every node present in the tree. The splitting rule is chosen on the basis of a sub-set of Attributes, showed along with subset ratio. The values are separated in best possible way through this rule for the chosen parameter criterion. The values are separated according to the rule for various classes in classification, whereas, when regression is concerned, the errors due to the estimation are minimized by separating. Till the ending criteria are reached, new nodes are built continuously.

The model's complexity can be minimized by leverage of the pruning technique after replacing sub-trees, as its predictive power with leaves is quite limited. The details of parameter shall indicate the various kinds of pruning.

A somewhat identical technique to random forest is extremely randomized trees; the method for acquiring this includes assessing the split random parameter and disabling pruning. Tuning for this technique include the parameters that are minimal leaf size and split ratio, while disabling guess split ratio will undo it. The best picks for the minimal leaf size classification and regression problems include two and five respectively.

Table 11 illustrate the prediction using the Random Forest Model and collocation the False Negative (FN), False Positive (FP), True Negative (TN) and True Positive (TP) to evaluation model.

Table 11 Random Forest Model result

		true E	true T	true M	true E10	FN
Random Forest	pred. E	101	7	2	26	35
	pred. T	0	115	16	1	17
	pred. M	0	0	63	0	0
	pred. E10	1	14	0	65	15
	ground truth	102	136	81	92	
	Predict	136	132	63	80	
	FP	1	21	18	27	
	sum	411				
	TP & TN	344				

Table 12 illustrate the Accuracy, Precision, Recall, and F-score of each class in the Random Forest model.

Table 12 Random Forest Model categories result

	Class	E	T	M	E10
Random Forest	TP	101	115	63	65
	TN	243	229	281	279
	FP	1	21	18	27
	FN	35	17	0	15
	Accuracy	0.905263	0.900524	0.950276	0.891192
	Precision	0.990196	0.845588	0.777778	0.706522
	Recall	0.742647	0.871212	1	0.8125
	F-score	0.848739	0.858209	0.875	0.755814

vi. Naive Bayes

A small data set can be the basis for creating a reasonable model through Naive Bayes which is a high-bias, low-variance classifier. It’s quite user-friendly and computationally inexpensive. The major utilizations are for text categorization, including spam detection, sentiment analysis, and recommender systems. The Naive Bayes has a fundamental assumption on the theory that if the label (the class) is provided, the result from an Attribute is independent from the result of the other Attributes. The Naive Bayes classifier is quite useful, despite of the fact that these assumptions are rarely true (they are "naive"!). The mathematical work required for creating the Naive Bayes probability model is minimized with the independence assumption. Certain assumption regarding distributions of conditional probability should be carried out to give the final touch to the probability model, for specific Attributes according to the class. The

Gaussian probability densities are used by Operator for modeling the Attribute data.

Table 13 illustrate the prediction using the Naive Bayes Model and collocation the False Negative (FN), False Positive (FP), True Negative (TN) and True Positive (TP) to evaluation model.

Table 13 Naïve Bayes Model result

		true E	true T	true M	true E10	FN
Naive Bayes	pred. E	100	3	1	9	13
	pred. T	0	58	0	0	0
	pred. M	0	58	79	0	58
	pred. E10	2	17	1	83	20
	ground truth	102	136	81	92	
	Predict	113	58	137	103	
	FP	2	78	2	9	
	sum	411				
	TP & TN	320				

Table 14 illustrate the Accuracy, Precision, Recall, and F-score of each class in the Naive Bayes model.

Table 14 Naie Bayes Model categories result

	Class	**E**	**T**	**M**	**E10**
Naive Bayes	TP	100	58	79	83
	TN	220	262	241	237
	FP	2	78	2	9
	FN	13	0	58	20

	Accuracy	0.955224	0.80402	0.842105	0.916905
	Precision	0.980392	0.426471	0.975309	0.902174
	Recall	0.884956	1	0.576642	0.805825
	F-score	0.930233	0.597938	0.724771	0.851282

5. DISCUSSION

For model's analysis, the authors used two methods. First, through evaluating each class in the model used accuracy and F-Score. Second, evaluating the models used accuracy.

Table 15 compares each class in all models

	Class E		**Class T**		**Class M**		**Class E10**	
	Accuracy	**F-score**	**Accuracy**	**F-score**	**Accuracy**	**F-score**	**Accuracy**	**F-score**
Generalized Linear	0.981	0.967	0.916	0.879	0.961	0.899	0.939	0.868
Decision Tree	0.979	0.971	0.913	0.884	0.958	0.905	0.937	0.829
Deep Learning	0.974	0.943	0.918	0.878	0.958	0.88	0.937	0.891
Gradient Boosted	0.973	0.952	0.908	0.865	0.963	0.905	0.927	0.847
Random Forest	0.905	0.849	0.901	0.858	0.95	0.875	0.891	0.756
Naive Bayes	0.955	0.93	0.804	0.598	0.842	0.725	0.917	0.851

From table 15 above that illustrate the comparing Accuracy and F-score of class in each model and if we exclude the random forest model, it is obvious that the accuracy results in class E was very high. It is important to note that any game in principal is for everyone (means, class

E), the content description will change the class to be some other classes – please see table 17 for content descriptions. In general, the accuracy results were in somehow high in all categories, but not in the class T category, which implies that the prediction for this class was not achieved perfectly. The researchers believe that this is due to the conflict between class T content description with (class E10 and class M)' content descriptions.

Table 16 comparing the models evaluation

Model	**Overall Accuracy**
Generalized Linear	0.9027
Decision Tree	0.8978
Deep Learning	0.8978
Gradient Boosted	0.8905
Random Forest	0.837
Naive Bayes	0.7786

Table 16 shows that the generalized linear model achieved the best results in terms of accuracy (0.9027) compared to other models used in this research. Although, the decision tree and deep learning models achieved very well. The most important content description that has high influence on prediction were those who has weight more than 0.5 shown in table 17 and proved by the receiver operating characteristic curve (ROC) shown in figure 2 below.

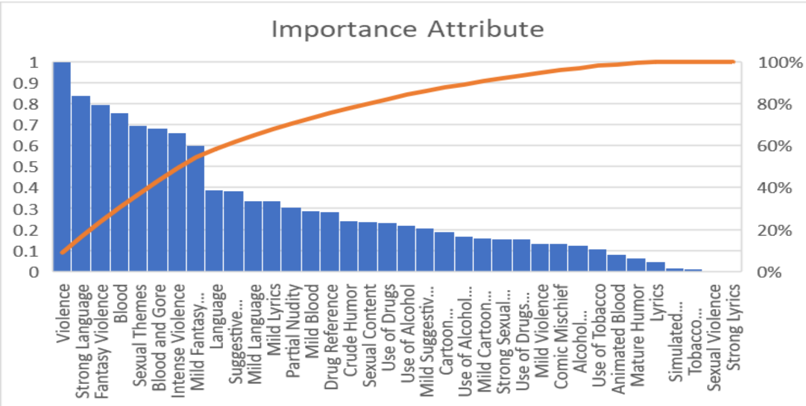

Figure1 content description weight

Table 17 content description weight result

Attribute	**weights**	**Attribute**	**weights**
Violence	1	Cartoon Violence	0.188934216
Strong Language	0.838057824	Use of Alcohol and Tobacco	0.168348965
Fantasy Violence	0.794854985	Mild Cartoon Violence	0.159409298
Blood	0.756097962	Strong Sexual Content	0.153716103
Sexual Themes	0.696178674	Use of Drugs and Alcohol	0.153716103
Blood and Gore	0.681380689	Mild Violence	0.133808617
Intense Violence	0.662858896	Comic Mischief	0.132589223
Mild Fantasy Violence	0.598541607	Alcohol Reference	0.124247704
Language	0.388589151	Use of Tobacco	0.10678485
Suggestive Themes	0.384235154	Animated Blood	0.082399214
Mild Language	0.33754685	Mature Humor	0.063136969
Drug Reference	0.284278128	Lyrics	0.048609269
Crude Humor	0.24096595	Simulated Gambling	0.017131237
Sexual Content	0.236182457	Tobacco Reference	0.013750678
Use of Drugs	0.230909405	Sexual Violence	0

Use of Alcohol	0.220620196	Strong Lyrics	0
Mild Suggestive Themes	0.204935285		

6. CONCLUSIONS

The researchers compared prediction among six model and the results showed that the Generalized Linear Models (GLMs) achieved a best accuracy (0.9027), also results highlighted eight important content descriptions to have the highest influence on prediction. The most important content description that has high influence on prediction were those who has weight more than 0.5 shown in table 17 which are: Violence, Strong Language, Fantasy Violence, Blood, Sexual Themes, Blood and Gore, Intense Violence, and Mild Fantasy Violence. These content description results are proved by the receiver operating characteristic curve (ROC).

The result is useful for researchers and digital rating system developer so that it formulates a base for parents to be advised before their children can use these games. Which could also be very beneficial in predicting their behaviors if these games are to be used, so the best is to have restriction in these games.

REFERENCES

Bartlmae, K. ,2000,. **An experience factory for knowledge discovery in databases: representing KDD-experience. In** Proceedings of the 2000 information resources management association international conference on Challenges of information technology management in the 21st century (pp. 1064–1065). IGI Global.

Diebold, F. X., & Mariano, R. S. ,1995,. **Comparing Predictive Accuracy.** Journal of Business & Economic Statistics, Vol.13,No.3,pp. 253–263.

Diebold, F. X., & Kilian, L. ,2001**,. Measuring predictability: theory and macroeconomic applications.** Journal of Applied Econometrics, Vol.6,No.6, pp. 657–669.

Engelhardt, C. R., Bartholow, B. D., Kerr, G. T., & Bushman, B. J. , 2011,. **This is your brain on violent video games: Neural desensitization to violence predicts increased aggression following violent video game exposure**. Journal of Experimental Social Psychology, Vol.47,No.5, pp. 1033–1036.

Felini, D. , 2015, **Beyond today's video game rating systems: A critical approach to PEGI and ESRB, and proposed improvements.** Games and Culture, Vol.10. No.1,pp. 106–122.

Geisser, S. ,1993, **Predictive inference, vol. 55. CRC Press, Boca Raton, FL. Gullett, W.(1997).** Environmental Protection and the Precautionary Principle: A Response to Scientific Uncertainty in Environmental Management. Environmental and Planning Law Journal, Vol.14,No.1, pp. 52–69.

Gelman, A., Hwang, J., & Vehtari, A. ,2014,. **Understanding predictive information criteria for Bayesian models.** Statistics and Computing, Vol.24,No.6, pp. 997–1016.

Gentile, D. A., & Anderson, C. A. ,2006**,. Video games.** Encyclopedia of Human Development, Vol.3, No.8, pp.1303–1307.

Greitemeyer, T., & Osswald, S, 2009, **Prosocial video games reduce aggressive cognitions.** Journal of Experimental Social Psychology, Vol.45,No.4,pp.896–900.

Gruber, M. , 2017**,. Decision Modeling and CRISP-DM for Modern Data Science Projects.** Decision Management Solutions. Retrieved from Http://Www.Decisionmanagementsolutions.Com/Crisp-Dm-and-Decision-Modeling-for-Modern-Data-Science-Projects/.

Hee, O. ,1966, **Tests for Predictability of Statistical Models.** Journal of Farm Economics, Vol.48, No.5, pp.1479–1484.

Jaslow, R. ,2013,. **Violent video games and mass violence: a complex**

link. CBS News, 18.

Jenkins, H. , 2006, **Reality bytes: Eight myths about video games debunked**. Impact of Gaming Essays. Retrieved Feb, 1, 2007.

Kirsh, S. J. ,2002, **The effects of violent video games on adolescents**. Aggression and Violent Behavior, Vol.7, pp.1–13.

Kumalasari, G. 2019. **Local government policy model of indonesia rattan handicraft creative industry in trangsan village**. Humanities & Social Sciences Reviews. Vol. 7, N° 3: 87-91. India.

Laczniak, R. N., Carlson, L., Walker, D., & Brocato, E. D., 2017, **Parental Restrictive Mediation and Children's Violent Video Game Play: The Effectiveness of the Entertainment Software Rating Board (ESRB) Rating System.** Journal of Public Policy & Marketing, Vol.36,No.1, pp.70–78.

McGonigal, J, 2011,. **Reality is broken: Why games make us better and how they can change the world.** Penguin.

McLean, L., & Griffiths, M. D. ,2013, **The psychological effects of video games on young people.** Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport, Vol.31,No.1.

Norcia, A. M. A., 2014**,. The Impact of Video Games**. Retrieved from Http://Www.Pamf.Org/Parenting-Teens/General/Media-Web/Videogames.Html.

Saleem, M., Anderson, C. A., & Gentile, D. A., 2012**,. Effects of prosocial, neutral, and violent video games on children's helpful and hurtful behaviors.** Aggressive Behavior, Vol.38,No.4,pp. 281–287.

Salloum, S. A., Al-Emran, M., Monem, A., & Shaalan, K,2017**, A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives.** Advances in Science, Technology and Engineering Systems Journal, Vol.2,No.1, pp. 127–133.

UNIVERSIDAD
DEL ZULIA

opción

Revista de Ciencias Humanas y Sociales

Año 35, Especial N° 19, 2019

Esta revista fue editada en formato digital por el personal de la Oficina de Pubñlicaciones Científicas de la Facultad Experimental de Ciencias, Universidad del Zulia.
Maracaibo - Venezuela

www.luz.edu.ve

www.serbi.luz.edu.ve

produccioncientifica.luz.edu.ve