Advances in Mathematical
Sciences and Applications
Vol. 27, No. 1 (2018), pp. 241–262

GAKKOTOSHO
TOKYO JAPAN

A COMPARISON OF CLASSIFIERS FOR MUSICAL GENRES CLASSIFICATION AND MUSIC EMOTION RECOGNITION

Beatriz C. Flamia de Azevedo *
Federal University of Technology of Paraná - Brazil, Mathematics Department
1640, Alberto Carazzai, 86300-000, Cornélio Procópio
(E-mail: beatrizazevedo@alunos.utfpr.edu.br)

and

Glaucia Maria Bressan †
Federal University of Technology of Paraná - Brazil, Mathematics Department
1640, Alberto Carazzai, 86300-000, Cornélio Procópio
(E-mail: glauciabressan@utfpr.edu.br)

Abstract. The automatic classification of music into genres is one of the most popular tasks in the music information retrieval research area and it can be framed as a pattern recognition problem. The goal of this paper is to investigate and to apply different classification methods in order to automatically classify Latin musical genres and predominant emotions associated to them. In order to do this automatic classification task, musical physical attributes (sound beats, timbre and frequency) are considered as input of the classification methods and the performances of classification algorithms are analyzed. The classifiers considered in this paper are based on data mining and statistical techniques: Decision Trees, Random Forest, k-Nearest Neighbor (kNN), Support Vector Machine (SVM) and Artificial Neural Network (ANN). Each Latin genre is associated with its predominant emotion, obtained from the literature. Thus, the proposed methodology can be easily used by music therapists in health treatments.

*Financial support Fundação Araucária
†corresponding author: glauciabressan@utfpr.edu.br
Communicated by Editors; Received May 5, 2018.
2010 AMS Subject Classification: 68T10; 68Q01; 62-07; 68R10.
Keywords: Latin Musical Genres, Music Emotions, Classification Methods, Pattern Recognition.

1 Introduction

Music is more than leisure activity; it is part of human development. Listening to music provides benefits for the body, mainly for the brain. People reaction when listening to music and the emotion power of music have been researched for different areas as medicine, psychology, mathematics, neuroscience. The scientists are interested in observe the music effects in the brain and body reaction to musical effects. The results of these researches are very positive and they demonstrate that music can induces changes in mind and consequently it improves humans life quality. This fact then has favored the use of music as a tool for the different diseases treatment, such as cardiovascular and respiratory [56], speech and memory [17, 23, 24], brain's disorder such as autism, schizophrenia, depression, anxiety [18, 23, 31, 34] and others.

Music is present in our life since early age. Nakata and Trehub (2004) report that singing is more effective than speech at holding an infant's attention [35], while Trappe (2012) affirms that music could be considered the first sensory impression for the child, since the music brings benefits during the pregnant and it has influence in the babies development [37, 56].

The study of music and music therapy needs to be multidisciplinary as well as theoretically and scientifically pluralistic [22]. To determine the effect of each music will provoke in a specific person is not trivial. Music emotions are influenced by many factors such as age, cultural and personality for example. These characteristics make different people listen to the same music and distinct emotions are stimulated. Thus, therapists had difficulty selecting the best music for specific patients. For example, to select a music for a depressed person, who needs to have happiness and positive feeling stimulated, or for a schizophrenic during a crisis, it is not a simple task, considering personal multi factors that involve musical emotions.

In this context, an automatic system that evaluates musical attributes in order to classify music, considering the predominant emotion of each music or genre, without human intervention, could be an important tool to help in music therapy. This process could be done analyzing the physical attributes of music.

Physical attributes of music have connection with emotions that are stimulated in the listener. Some studies that investigated the relation between emotion and music attributes concluded that emotions can be influenced by such attribute, as time, timbral, harmony and loudness [25] and the attributes related to rhythm and harmony are closer to human experience [12, 16]. Aucouturier and Pachet (2003) aim that music that belong to the same genre share similarities among themselves, which differentiate them from other genres [3]. While Lin et. al (2009) affirm that genre and emotion provide complementary descriptions of music content and often correlate with each other. Their study shows the existence of a correlation between genre and emotions. In addiction, they observed improvement in the music emotion classification results when the information about the genre are available [30].

In this paper, we investigate and apply different classification methods in order to automatically classify Latin musical genres and predominant emotions associated to them. Musical physical attributes (sound beats, timbre and frequency) are considered as input of the classification methods. Aiming to evaluate classifiers, the performances of classification algorithms are analyzed and compared. The classifiers considered in this paper are: (i) Decision Trees - using CART and C4.5 algorithms, (ii) Random Forest, (iii) k-Nearest Neighbor (kNN) - using IBK algorithm, (iv) Support Vector Machine (SVM) - using Sequential Minimal Optimization (SMO) and (v) Artificial Neural Network (ANN) - using Multilayer Perceptron. The database used is named *Latin Music Database* [49], from Silla-Jr et al.(2008), which present real numerical features of Latin music. Each Latin genre is associated with its predominant emotion obtained in the literature and mainly in the work [46], which investigate the emotion of the *Latin Music Database*.

This paper is organized as follow: Section 2 describes some researches about music therapy and music classification. Section 3 describes the classifiers used in this paper for musical genres classification and music emotion recognition. The evaluation of the classification task is presented in Section 4 and the database as well as the study involving Latin genre emotions are described in Section 5. In Section 6 the results obtained from classification methods considered in this work are presented. Finally, Section 7 comments on the conclusion obtained from the classification results.

2 Related Works

As described before, music is associated with emotion and it can induces reaction and feelings. Music can not cure a disease, but it can help to relax, to calm and to soft negative effects of various symptoms and diseases, making the healing process faster. Several publications indicate significant improvements in cure progression even in more advanced stages [17, 23, 54].

Many researchers have been studying the effect of music in humans' memories. The autobiographical memories count among the most poignant experiences associated with music [24]. The musical memory is considered as independent part from other memory systems [23]. Jacobsen et. al (2015) and Gagnon et. al (2009) studied the influence of music in Alzheimer's disease, and indicate that musical memory and the music emotion are preserved and significant improvements has been noticed with the music therapy.

Partanen et. al (2013), investigate the formation and retention of neural representations induced by exposure to melodies during the fetal period [37]. Tabarro et. al (2010) explore the use of the music during the labor and on the newborn when submitted to the same melodies heard by their own mothers during pregnancy [53]. In both studies, positive effects were noticed for both mothers and babies.

Thompson (2013) presents a study of music therapy involving women with breast cancer, which Decision Tree was used to provide a framework for determining applicable clinical interventions to meet the needs of participants [54]. Cabredo et. al (2012) use an electroencephalograph to record the reaction of people's brain while they listen music. The authors used an implementation of linear regression and the algorithm C4.5 to construct a Decision Tree to build the emotion models for each music emotions [9]. Zhang et. al (2016), propose an automatic emotion recognition system for the music by extracting different features from the music and machine learning method learning from common knowledge on emotional state of the trained data. The Random Forest classifier is chosen to ensemble all these features together and predict the emotion categories of the music [59].

In Chathuranga et. al (2014), the authors present an ensemble approach for the problem of automatic music genre classification using SVM and pattern recognition, considering audio signal [11].

Considering authors that investigate ANN, Mokhsin et. al (2014) detected emotion features in Malay popular music using ANN by extracting audio timbre features from both vocal and instrumental sound clips [33]; while Gerrero-Turrubiates (2014), proposed a music recognition system based on harmonic modification and ANN, this classification is performed by a feed-forward neural network or Multilayer Perceptron [19].

Others interesting approaches are described in Kothe et. al (2016) and Azarloo et. al (2012), in these papers the authors proposed machine learning models as kNN, ANN and SVM to detect and distinguish instruments considering different musical genres, based on music features [5, 27]. Chandwadkar and Sutaone (2013), in their work, classified four instruments considering piano, acoustic guitar, xylophone and violin. These instruments are identified using various music features and classifiers, Decision trees, k-Nearest Neighbour classifier, Multilayer Perceptron, Sequential Minimal Optimization Algorithm (SMO) and multi class classifier (meta classifier) [10].

In this paper, we propose an automatic classification of Latin music genres and their predominant emotion, applying data mining and statistical technique. In this sense, the main contribution of this paper is to propose a methodology capable of determining the correspondent music genre of each music, evaluating multiple music attributes relation to sound beats, timbre and frequency. Furthermore, the predominant emotion of each genre is considered in this paper and the emotion is associated to the genre (output class of the classifier). The proposed methodology can be easily used by music therapists helping to choose the best music for each patient or health treatment.

3 Classifiers

In this section, algorithms of the classification methods employed to classify the Latin Music Database are described. These algorithms are the more recent techniques presented in the literature and they have different construction criteria. All of them are available in free software WEKA*, developed by the University of Waikato, New Zealand. Thus, it is possible to compare the performance of different classification algorithms, since they are implemented using the same platform.

The use of different classifiers allow us to compare methodologies. The algorithms present different ways to extract information from database and to transform it into knowledge. For this reason, depending on the metric used by an algorithm, the results can be different.

In order to compare the performance of different classification methods, "Black Box" and "White Box" methods were chosen to classify Latin musical genres database. The Black Box methods can be viewed in terms of its inputs and outputs, without any knowledge of its internal workings. The opposite of them, is the White Box, which is a system where the inner components are available for inspection. White Box algorithms reveal the structure, allowing users to assemble algorithms from algorithm building blocks. Among

*(Waikato Environment of Knowledge Analysis - http://www.cs.waikato.ac.nz)

them, in this paper, *Artificial Neural Networks, k-Nearest Neighbor* and *Support Vector Machines* are Black Box methods, in the sense of users can not read the acquired knowledge in a comprehensible way, and *Decision Tree* and *Random Forest* are White Box methods. The latter ones generate understandable rules. So, users can understand each split, see the impact of that split and even compare it to alternative splits [29].

3.1 Decision Trees

Decision tree is a very useful data mining technique to extract information of a data set. This process is done by a TDIDT (Top-Down Induction Decision Tree) algorithm [45], which induces a tree structure by splitting data into subgroups on divide and conquer method [28]. A Decision Tree consists of a data set, partitioned into groups known as *nodes*. The top node is called *root node*, which is selected using some attribute selection measure or splitting criterion. Under the root node are the internal nodes, originating from the division of the data set; they constitute the tree branches. At the end of each branch is the terminal node, designed leaves, which represent the most appropriated class for the rule. One rule is composed by each terminal node, plus the internal nodes that belongs to one specific branch and the root node. The rules are describe by a IF-THEN model ("IF attribute w is $y1$ AND attribute x is $y2$ THEN the class is z"). These rules are used to classify unknown data at the inference process.

In order to construct the Decision Tree models, in this paper we employed the most commons algorithms to induce trees: CART (Classification and Regression Tree) [8] and C4.5 [41]. These algorithms construct the tree in two phases: growing and pruning. The growing is a recursive process that establishes the structure of the Decision Tree according to some splitting criterion, such as Information Gain, Gini Index, Twoing, and Gain Ratio. The mathematical formulation of C4.5 and CART can be seen in [45].

CART algorithm is characterized by the fact that it constructs binary trees, namely each internal node has exactly two outgoing edges [45]. The splitting criteria is named Twoing, that search for two classes that will make up together more than 50% of the data [55]. While C4.5 algorithm uses Gain Ratio as splitting criteria. The element with highest Gain Ratio is taken as the root node and data set is split based on the root element values [38].

The second phase, pruning, is responsible to reduce the complexity of the tree. It means reducing size of the tree that are too large and deep. The problem of noise and overfitting reduces the efficiency and accuracy of data. The overfitting happens when the tree lost the ability of generalizing to instances not present during the training process. By increasing the number of nodes, the training error usually decreases while at some point the generalization error becomes worse [45].

Pruning methods typically use statistical measures to remove the least reliable branches [39]. As a consequence it optimize the computational operation, eliminate overfitting, and improve the classification of unknown data. The pruning techniques varies according to the Decision Tree algorithm, Cost-Complexity Pruning, Error Based Pruning, Reduced Error Pruning and Minimum Error Pruning are some examples that can be used to prune trees. Their mathematical formulation can be seen in [45].

CART uses the Cost-Complexity to prune the tree, this method proceeds in two stages:

in the first stage, a sequence of trees $T_0, T_1, ..., T_k$ is built on the training data where T_0 is the original tree before pruning and T_k is the root tree and in the second stage, one of these trees is chosen as the pruned tree, based on its generalization error estimation [45]. The strategy used by C4.5 named Error Based Pruning it is an evolution of other method called Pessimistic Pruning. C4.5 equates the predicted error rate at a leaf with this upper limit, on the argument that the tree has been constructed to minimize the observed error rate [42]. More details about the prune methodology and the algorithm used to prune could be seen in [45, 39].

3.2 Random Forests

Random Forests or Random Decision Forests are an ensemble learning method for classification that construct a multitude of Decision Trees at training time and outputting the class that is the mode of the classes of the individual trees. Due to their power, versatility, and ease of use, Random Forest are quickly becoming one of the most popular machine learning methods [28]. Random Forest classifiers incorporate the principles of bagging (multiple drawing of samples with replacement) and random feature selection (finding the most discriminating features from a randomly selected subset of features) for learning the Decision Trees [26].

A Random Forest consists of multiple Decision Tree and predicts the class that is obtained by voting on the predictions made by the individual tree (ensemble classifier) [26]. An ensemble combines a series of n learned models (or base classifiers), $M_1, M_2, ..., M_n$ with the aim of creating an improved composite classification model, M_*. A given data set, D, is used to create n training sets, $D_1, D_2, ..., D_n$, where $D_i (1 \leq i \leq n-1)$ is used to generate classifier M_i. Given a new data tuple to classify, the base classifiers each vote by returning a class prediction. [20].

In order to grow ensembles, often random vector are generated that govern the growth of each tree in the ensemble. Bagging [6], Random split [14] and Random subspace [15] are some examples of split selectors. According to Breiman (2001) the common element in all of these procedures is that for the kth tree, a random vector Θ_k is generated, independent of the past random vectors $\Theta_1, ..., \Theta_{k-1}$, but with the same distribution; and a tree is grown using the training set and Θ_k, resulting in a classifier $h(\mathbf{x}, \Theta_k)$ where $\mathbf{x}$ is an input vector. After a large number of trees is generated, each tree casts a unit vote for the most popular class at input $\mathbf{x}$ [7].

The accuracy of a Random Forest depends on the strength of the individual classifiers and a measure of the dependence between them. The ideal is to maintain the strength of individual classifiers without increasing their correlation. Random Forest are insensitive to the number of attributes selected for consideration at each split [20]. The generalization error for a forest converges as long as the number of tree in the forest is large. Thus, Random Forest are less prone to overfitting [20, 28].

3.3 Support Vector Machines

A Support Vector Machine (SVM) is a binary classification method that divides the given data into two groups in the best possible way by using hyperplanes. This method is based on a structural risk minimization method to reduce the error rather than the empirical risk minimization method used in traditional statistical learning theory [29]. Without any knowledge of the mapping, the SVM finds the optimal hyperplane by using the dot product functions in feature space using kernel functions. The solution of the optimal hyperplane can be written as a combination of a few input points that are called support vectors [30].

Training a SVM requires the solution of a very large quadratic programming optimization problems. The Sequential Minimal Optimization (SMO) [40] is an analytical form to solve the problem of quadratic optimization programming.

The SMO decomposes the overall quadratic programming in quadratic sub-problems, using Osuna's Theorem [36] to ensure converge, without any extra matrix storage or using numerical quadratic programming optimization steps. SMO chooses to solve the smallest possible optimization problem at every step. For the standard SVM quadratic problem, the smallest possible optimization problem involves two Lagrange multipliers, because the Lagrange multipliers must obey a linear equality constraint. At every step, SMO chooses two Lagrange multipliers to jointly optimize, finds the optimal values for these multipliers, and updates the SVM to reflect the new optimal values [40].

SMO works fastest for linear SVMs and sparse data sets, as well as binary data and non-linear data [40]. This algorithm are able to work with missing values are replaced globally, nominal attributes are transformed into binary ones, and attributes are normalized by default [58]. The advantage of SMO lies in the fact that solving for two Lagrange multipliers can be done analytically. SMO requires no extra matrix storage at all. Thus, every large SVM training problems can fit inside of the memory of an ordinary personal computer or workstation, it is less susceptible to numerical precision problems [40]. The amount of memory required for SMO is linear, which allows SMO to handle very large training sets.

3.4 k-Nearest Neighbors

According to [20], the k-Nearest Neighboor (kNN) classifiers are based on learning by analogy, that is, by comparing a given test tuples with training tuples that are similar to it. The training tuples are described by n attributes. Each tuple represents a point in an n-dimensional space. In this way, all the training tuples are stored in an n-dimensional pattern space. When given an unknown tuple, a *k-Nearest Neighboor classifier* searches the pattern space for the k training tuples that are closest to the unknown tuple. These k training tuples are the k "nearest neighbors" of the unknown tuple. The closeness is defined in terms of a distance metric, such as Euclidean distance. The Euclidean distance between two points of tuples, which is commonly used in kNN algorithm could be described by: considering

$X_1 = (x_{11}, x_{12}, ..., x_{1n})$ and $X_2 = (x_{21}, x_{22}, ..., x_{2n})$, then

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2}.$$

In other words, for each numeric attribute, we take the difference between the corresponding values of that attribute in tuple X_1 and in tuple X_2, square this difference, and accumulate it. The square root is taken of the total accumulative distance count [20].

Specify the value o k, is not a trivial task. The value of k determines how well the model will generalize to future data [28]. So, choosing a large k reduces the impact or variance caused by noisy data, but can bias the learner such that it runs the risk of ignoring small, but important patterns. On the other hand, using a single nearest neighbor allows noisy data or outliers, to unduly influence the classification of examples. One common practice is to set k equal to the square root of the number of training examples, but normally k is set somewhere between 3 and 10 [28]. For this reason some authors suggest test k from 1 to 10 to find the best k.

The Based Instance Algorithm (IBK) is based on instances (instance based algorithm) [1]. This algorithm use the technique of "Nearest Neighbor" to classify each new instance. This methodology is known by lazy learner since training consist only of storing the given training examples alongside the class label [26].

In the IBK, the value of k can also be determined automatically using leave-one-out Cross-Validation, subject to an upper limit given by the specified value. Predictions from more than one neighbor can be weighted according to their distance from the test instance, and two different formulas are implemented for converting the distance into a weight. The number of training instances kept by the classifier can be restricted by setting the window size option. As new training instances are added, the oldest ones are removed to maintain the number of training instances at this size [58].

3.5 Artificial Neural Network

An Artificial Neural Network (ANN) models the relationship between a set of input signals and an output signal using a model derived from our understanding of how a biological brain responds to stimuli from sensory inputs. ANNs are best applied to problems where the input data and output data are well-understood or at least fairly simple, yet the process that relates the input to output is extremely complex [28].

A neural network performs pattern recognition by first undergoing a training session during which the network is repeatedly presented with a set of input patterns along with the category to which each particular pattern belongs. Later, the network is presented with a new pattern that has not been seen before, but which belongs to the same population of patterns used to train the network. The network is able to identify the class of that particular pattern because of the information it has extracted from the training data. Pattern recognition performed by a neural network is statistical in nature, with the patterns being represented by points in a multidimensional decision space. The decision space is divided into regions, each one of which is associated with a class. The decision boundaries are determined by the training process. The construction of these boundaries is made statistical by the inherent variability that exists within and between classes [21].

When the classes are linearly separable, we could use a simple perceptron. However when the problem is not linearly separable, the simple perceptron algorithm will fail to generate a separating hyperplane (in this two-dimensional instance space a hyperplane is just a straing line), in this case Multilayer Perceptron are used. Multilayer Perceptrons are usually trained by minimizing the squared error of the network's output, essentially treating it as an estimate of the class probability [58].

According to Haykin (2009), a popular method for the training of Multilayer Perceptrons is the *Backpropagation algorithm*, which includes the least-mean-square algorithm as a special case. The training proceeds in two phases: forward and backward, which are described below.

- In the forward phase, the synaptic weights of the network are fixed and the input signal is propagated through the network, layer by layer, until it reaches the output. Thus, in this phase, changes are confined to the activation potentials and outputs of the neurons in the network. A pattern is presented to the units in the input layer, and from this layer, the units calculate their response that is produced in the output layer.

- In the backward phase, an error signal is produced by comparing the output of the network with a desired response. The resulting error signal is propagated through the network, again layer by layer, but this time the propagation is performed in the backward direction. In this second phase, successive adjustments are made to the synaptic weights of the network. Calculation of the adjustments for the output layer is straightforward, but it is much more challenging for the hidden layers. The network learns the rules necessary to work with the knowledge, by adjusting the weights and to minimize the error between the actual data and the calculated by the network.

Since Backpropagation requires a known, desired output for each input value, it is a supervised learning method. It calculates the error contribution of each neuron after a batch of data is processed. This is used by an enveloping optimization algorithm to adjust the weight of each neuron, completing the learning process for that case [21]. It calculates the gradient of the *loss function*. This function maps values of one or more variables onto a real number intuitively representing some cost associated with those values. The loss function then calculates the difference between the network output and its expected output. It is used in the gradient descent optimization algorithm. The error is calculated at the output and distributed back through the network layers. After training, the network can be used as a tool to classify new data. For this, it should only be used in progressive mode (feed-forward),i.e., new entries are inserted, processed in intermediate layers and the results are presented in the output layer, but without the backpropagation of the error. More details about ANN theory and Backpropagation algorithm can be seen in [21].

4 Evaluation of Classification Methods

Evaluating the performance of a classification is a fundamental aspect of machine learning. In this work, Cross-Validation, Receiver Operating Characteristic and Confusion Matrix are applied as efficient indicators for assessing the quality of the analysis results.

In s-fold *Cross-Validation*, the data is randomly split into s mutually exclusive subsets of approximately equal size. An inducer is trained and tested s times; each time it is tested on one of the n folds and trained using the remaining $s-1$ folds. The advantage of this method is that it matters less how the data gets divided. Every data point gets to be in a Test Set exactly once, and gets to be in a training set $s-1$ times. The variance of the resulting estimate is reduced as s is increased. Therefore, the basic idea is that some of the data is removed before training begins. Then, when training is done, the data that was removed can be used to test the performance of the learned model on "new" data.

The *Confusion Matrix* is used as an indication of the properties of a classification rule. It contains the number of elements that have been correctly or incorrectly classified for each class. The main diagonal presents the number of observations that have been correctly classified for each class. The off-diagonal elements present the number of observations that have been incorrectly classified [45].

The Confusion Matrix $M(C_i, C_j)$, for $i, j = 1, ..., k$, classes provides an effective measure of the classification model performance, since it shows the number of elements correctly classified and predict, for each class. Let T be a set of examples and h the hypothesis. The Confusion Matrix of h is given by:

$$M(C_i, C_j) = \sum_{\forall (x,y) \in T : y = c_i} \parallel h(x) = C_j \parallel$$

where C_i is the true class and the C_j is the predict class. Table 1 illustrates the Confusion Matrix.

Table 1: Confusion Matrix

Class	Predict C_1	Predict C_2	$\ldots$	Predict C_k
True C_1	$M(C_1, C_1)$	$M(C_1, C_2)$	$\ldots$	$M(C_1, C_k)$
True C_2	$M(C_2, C_1)$	$M(C_2, C_2)$	$\ldots$	$M(C_2, C_k)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
True C_k	$M(C_k, C_1)$	$M(C_k, C_2)$	$\ldots$	$M(C_k, C_k)$

Receiver Operator Characteristic (ROC) curve, or ROC curve, is commonly used to present results for decision problems in machine learning. It examines the tradeoff between the detection of true positives, while avoiding the false positives. Curves are defined on a plot with the proportion of true positives on the vertical axis, and the proportion of false positives on the horizontal axis. These values are equivalent to *sensitivity* and $1-$*specificity*, respectively. The closer the curve is to the perfect classifier, the better it is at identifying positive values. This can be measured using a statistic known as the *area under the ROC curve*, which measures the total area under the ROC curve. The area ranges from 0.5 (for a classifier with no predictive value), to 1.0 (for a perfect classifier).

The points comprising ROC curves indicate the true positive rate at varying false positive thresholds. To create the curves, a classifier's predictions are sorted by the model's estimated probability of the positive class, with the largest values first. Beginning at the origin, each prediction's impact on the true positive rate and false positive rate will result in a curve tracing vertically (for a correct prediction), or horizontally (for an incorrect prediction), according to [28].

5 Database and Musical Emotions

In this section, the music database is described, as well as the procedure to collect the numerical input attributes. Besides, the musicology of the musical genres considered in this paper is presented in order to enable the association with the emotions prevalent in each one of them.

5.1 Latin Music Database

In this paper, we used a set of the available music in *Latin Music Database*, from [49], a 3000 musical recordings database from the ten different genres: Tango, Salsa, Forró, Axé, Bachata, Bolero, Merengue, Gaúcha, Sertanejo and Pagode. This database is organized in a matrix, where the row are called "instance" and contains the numerical information about the attributes and it corresponds to one genre; and column represents 30 input attributes, that represent the attributes extracted from the music; these attributes are better described in Table 2. The MARSYAS framework was employed for feature extraction [57]. The attributes employed in this paper comprise short-time Fourier transform, Mel frequency cepstral coefficients (MFCC), beat and pitch related features, inter-onset interval histogram coefficients, rhythm histograms and statistical spectrum descriptors. More details can be seen in [50]. The attributes can be split into three groups: the beat-related attributes (attributes 1 to 6) include the relative amplitudes and the beats per minute. Timbral texture attributes (attributes 7 to 25) account for the means and variance of the spectral centroid, rolloff, flux, the time zero domain crossings, the first five MFCCs and low energy. Pitch-related attributes (attributes 26 to 30) include the maximum periods and amplitudes of the pitch peaks in the pitch histograms. The attribute vector is outlined in Table 2.

Table 2: Description of the attribute vector

Attributes	Description
1	Relative amplitude of the first histogram peak
2	Relative amplitude of the second histogram peak
3	Ratio between the amplitudes of the second peak and the first peak
4	Period of the first peak in bpm
5	Period of the second peak in bpm
6	Overall histogram sum (beat strength)
7	Spectral centroid mean
8	Spectral rolloff mean
9	Spectral flow mean
10	Zero crossing rate mean
11	Standard deviation for spectral centroid
12	Standard deviation for spectral rolloff
13	Standard deviation for spectral flow
14	Standard deviation for zero crossing rate
15	Low energy
16	First MCFF mean
17	Second MCFF mean
18	Third MCFF mean
19	Fourth MCFF mean
20	Fifth MCFF mean
21	Standard deviation for first MFCC
22	Standard deviation for second MFCC
23	Standard deviation for third MFCC
24	Standard deviation for fourth MFCC
25	Standard deviation for fifth MFCC
26	The overall sum of the histogram (pitch strength)
27	Period of the maximum peak of the unfolded histogram
28	Range of the maximum peak of the folded histogram
29	Period of the maximum peak of the folded histogram
30	Pitch interval between the two most prominent peaks of the folded histogram

5.2 Musicology

In this subsection, a literature review that associates Latin musical genres with respective predominant emotion is presented. Specific books and publications about Latin music was consulted in order to find characteristics of each genre considered in this paper [2, 4, 13, 32, 43, 44, 46, 47, 48, 51, 52].

Tango: This musical genre has melancholic characteristics. The lyrics usually present an unmatched love as central theme, referring to the past, to the lost youth, then producing sadder emotions, combining drama, passion, sensuality and aggressiveness [51, 52].

Bachata: It is a genre that deals with love and suffering. Its origins refer to rural and poor communities in the Dominican Republic. Love, lack of love and disappointment are frequently sung in Bachata. José Manuel Calderón, considered the pioneer of Bachata music, affirms that this musical genre refers to “feeling”, capable of rising not only bitterness and disaffection, but also a range of emotions, which become a true message: “The Bachata sings love” [48].

Bolero: It is considered a romantic genre. It involves subjects like love among human beings, betrayal and unmatched love. Commonly the lyrics present a central theme that talk about a problem or an amorous intention [2, 52].

Merengue: It is considered the national dance of the Dominican Republic, Merengue rised in the Creole culture. It is a simple and engaging dance with sexual connotations, in which dancers interweave during the choreography. The lyrics usually talk about subjects like love and passion, in joyful and seductive way [4, 52].

Salsa: This genre instigates the idea of a music with “taste” [44]. It is dancing genre. Normally the lyrics talk about a seducer man, characterized by the pleasure and sexy expressions [52].

Forró: It started with the Brazilian musician Luiz Gonzaga, after his rhythm “Baião”. Gonzaga wanted “to sing the Brazilian Northeast”, talking about the man who left his land due to drought and the financial difficulties to live large urban centers [47].

Pagode: Pagode is a kind of samba, a popular Brazilian genre. The lyrics talk about daily activities. It is common in poor or traditional “samba communities” in Rio de Janeiro. Popularly, the Pagode is dancing and singing in joyful cluster of people [32].

Sertanejo: It is a common genre in countryside of Brazil, popular sung by duets or trios of musicians with strong connection of countryside. Its lyrics talk about arduous life of country people, as rural work, love suffered, or when the countryman is forced to move to a city, leaving his roots [43].

Gaúcha: This genre is popular in the South of Brazil with strong influence of Italians and Germans rhythms. It contemplates joy and theatricality, describing the country life [13].

Axé: This genre appeared in the state of Bahia, in Brazil, during the carnival party. It is considered a mix of African and Brazilian rhythms characterized by a cheerful, contagious, enthusiastic and very popular in Brazilian Carnival [32].

In Santos and Silla-Jr (2015), musical genres that compose the Latin Music Database was analyzed by professionally trained Brazilian musicians who indicate the predominant emotion perceived in each music [46]. In addiction, after a large literature review, the emotions associated to the musical genres are determined and the results are presented in Table 3.

6 Results

In order to classify the Latin musical genres, the performances of the classifiers described in Section 3 were investigated, by executing them for the same database.

The Latin Music Database, is compose by 30 input attributes, described in Table 2, and the 31th column contains the outputs, which are Latin genres. There are 3000

Table 3: Emotions associated to musical genres

Genre	**Emotion**	**Genre**	**Emotion**
Tango	Deception	Forró	Welcoming
Bachata	Love	Pagode	Happiness
Bolero	Romantic	Sertanejo	Sadness
Merengue	Passion	Gaúcha	Joyful
Salsa	Seduction	Axé	Enthusiastic

instances (rows of the matrix), being 10 instances of each one of the ten genres.

In order to compare the numerical and categorical data results, firstly the attributes were categorized according to tercile. Then, each input attribute is divided in 3 categories (low, medium and high), with 1000 values in each category.

The considered measures of models validation are: Test Set and Cross-Validation. For the Test Set, the database is partitioned in training set and Test Set. The training set is composed by 80% of full database (2400 instances) which is used to generated the models; and the Test Set, composed by 20% remaining data, is used to measure the ability of generalization of the generated models for unknown data. For the Cross-Validation, we consider n=30 as described in the Section 4. Cross-Validation uses the full data, however there is replacements of data subsets to test, during the testing process.

The results obtained from classifiers execution, considering numerical data, are presented in Table 4 and 5 present the results obtained from categorical data. The column named “accuracy”, in Tables 4 and 5, is defined as the number of instance correctly classified divided by the instances total. Accuracy values are provided by Confusion Matrices, since the main diagonal indicates the number of genres correctly classified. As described in the Section 4, the higher the values are concentrated in main diagonal, the better the classifier is.

The area under the ROC curve are also used as evaluation measure. As describe in Section 4, the closer the curve is to perfect classifier, the better it is at identifying positive values. Thus the closer to 1, the better the classification is.

Table 4: Numerical data results

Classifier	Test Set		Cross-Validation	
	Accuracy	ROC	Accuracy	ROC
C4.5 (pruning)	62.33	0.811	61.03	0.797
C4.5 (no pruning)	62.00	0.811	60.47	0.798
CART (pruning)	62.83	0.868	64.20	0.869
CART (no pruning)	60.17	0.804	63.10	0.812
Random Forest	78.00	0.972	80.33	0.972
Multilayer Perceptron	74.33	0.951	76.23	0.956
SMO	72.33	0.933	74.37	0.936
IBK (k=6)	60.33	0.882	63.17	0.889

Table 5: Categorical data results

Classifier	Test Set		Cross-Validation	
	Accuracy	ROC	Accuracy	ROC
C4.5 (pruning)	56.83	0.783	56.93	0.779
C4.5 (no pruning)	55.00	0.776	55.87	0.779
CART (pruning)	58.00	0.863	60.43	0.857
CART (no pruning)	57.50	0.773	57.60	0.785
Random Forest	73.83	0.955	73.37	0.972
Multilayer Perceptron	71.00	0.934	69.40	0.936
SMO	68.17	0.921	69.77	0.927
IBK (k=6)	64.50	0.921	68.90	0.932

As described in Section 3, Decision Trees models could be constructed using or not the pruning methods. The pruning process optimizes the computational operation, eliminates the overfitting and improves the classification of unknown data. In Table 6, measures of complexity of each Decision Tree model generated is presented. The complexity of the trees can be indicated by the number of leaves and the size of the tree.

Table 6: Measures of complexity of the Decision Trees algorithms

Algorithms	**Test Set**		**Cross-Validation**	
Numerical Data	**Leaves**	**Size**	**Leaves**	**Size**
CART(pruning)	58	115	67	133
CART(no pruning)	239	585	341	681
C4.5 (pruning)	301	601	364	727
C4.5 (no pruning)	310	619	379	757
Categorical Data	**Leaves**	**Size**	**Leaves**	**Size**
CART (pruning)	55	109	109	217
CART (no pruning)	385	769	485	969
C4.5 (pruning)	368	735	475	949
C4.5 (no pruning)	433	865	547	1093

Considering the models constructed using the kNN algorithm, different values for k were tested, varying k from 0 to 10. The best results are provided using Euclidean distance, with $k = 6$, which is nearly to the square root of the number of the attributes (30 attributes), as suggested in the literature.

As we can notice in Tables 4 and 5, the best accuracy and the best area under the ROC curve are presented by the models generated using the Random Forest. For this reason, Tables 7 to 10 present the confusion matrices for the Random Forest models. The Tables 7 and 9 correspond to Test Set, thus, the matrices must contain 600 data samples. And the Tables 8 and 10 consist of the results from Cross-Validation, which contains the 3000 the full dataset.

Table 7: Confusion Matrix using Test Set and numerical data

Genre		a	b	c	d	e	f	g	h	i	j
Tango	=a	60	0	0	0	0	0	0	0	0	0
Bachata	=b	0	58	0	0	2	0	0	0	0	0
Bolero	=c	2	1	43	0	3	3	2	3	3	0
Merengue	=d	0	2	0	53	1	0	1	0	2	1
Salsa	=e	0	3	0	3	44	2	2	1	2	3
Forró	=f	0	0	4	1	5	40	7	2	1	0
Pagode	=g	0	0	5	1	0	1	46	3	1	3
Sertanejo	=h	0	0	5	0	1	4	1	42	2	5
Gaúcha	=i	1	0	4	1	4	4	2	0	41	3
Axé	=j	0	2	0	2	3	2	4	1	5	41

Table 8: Confusion Matrix using Cross-Validation and numerical data

Genre		a	b	c	d	e	f	g	h	i	j
Tango	=a	290	0	10	0	0	0	0	0	0	0
Bachata	=b	0	283	2	4	7	1	0	0	3	0
Bolero	=c	8	1	239	0	9	5	11	23	4	0
Merengue	=d	0	7	0	277	2	3	3	1	4	3
Salsa	=e	0	2	10	6	243	11	5	4	17	2
Forró	=f	0	5	15	8	16	182	22	19	19	14
Pagode	=g	0	0	8	1	6	9	249	6	9	12
Sertanejo	=h	0	0	31	0	2	17	20	196	10	24
Gaúcha	=i	3	2	15	4	10	16	8	13	218	11
Axé	=j	0	5	5	10	3	9	7	8	20	233

Table 9: Confusion Matrix using Test Set and categorical data

Genre		a	b	c	d	e	f	g	h	i	j
Tango	=a	57	0	3	0	0	0	0	0	0	0
Bachata	=b	0	56	0	2	2	0	0	0	0	0
Bolero	=c	5	0	38	0	2	2	7	4	2	0
Merengue	=d	0	1	0	55	0	0	1	0	2	1
Salsa	=e	1	2	2	3	42	1	1	2	2	4
Forró	=f	1	0	4	1	2	33	12	4	1	2
Pagode	=g	0	1	5	1	0	1	42	5	2	3
Sertanejo	=h	1	0	6	0	2	1	4	35	0	11
Gaúcha	=i	2	0	2	1	2	2	1	2	45	3
Axé	=j	0	3	0	3	4	0	2	5	3	40

Table 10: Confusion Matrix using Cross-Validation and categorical data

Genre		a	b	c	d	e	f	g	h	i	j
Tango	=a	276	0	24	0	0	0	0	0	0	0
Bachata	=b	0	277	4	9	5	4	0	0	1	0
Bolero	=c	11	4	226	0	12	4	15	20	6	2
Merengue	=d	1	13	0	264	4	1	4	1	6	6
Salsa	=e	1	11	20	14	208	9	9	7	15	6
Forró	=f	2	7	23	11	18	142	36	24	20	17
Pagode	=g	0	2	12	4	8	10	224	11	15	14
Sertanejo	=h	1	1	34	1	8	18	21	177	11	28
Gaúcha	=i	12	6	15	6	14	9	8	12	201	17
Axé	=j	1	8	3	13	11	8	12	17	21	206

7 Conclusion

In this paper, we execute and discuss the results and the performance obtained from different classifiers for the musical genre classification task and music emotion recognition. From the obtained results, it is possible to conclude about the classifier that provided the best classification results and then it can be used by music therapists, who must select a determined musical genre for a specific patient, by considering the predominant emotion of each genre.

For the classification task, White Box and Black Box classifiers were considered, since they provide different classification models and results, which make possible to investigate their performances.

Analyzing the results presented in Tables 4 and 5, the higher accuracy and the higher area under the ROC curve are provided by Random Forests models. Due to this method deal with ensemble, it works better using large number of trees, consequently, large data sets. The Confusion Matrices can confirm this fact, since the higher values are concentrated in the main diagonal, indicating the excellent performance of the model.

The overffiting can be a problem for Decision Trees, but not for Random Forests classification method, since it consists of multiple Decision Trees and predicts the class that is obtained by voting on the predictions made by the individual trees. The most voted genre is elected.

The advantage of using White Box methods, such as Decision Trees and Random Forest, is the possibility to know the relationship between the input attributes and to obtain more information about the behavior of the data. Besides, the "IF-THEN" rules set, which is extracted from the classification model, is closer to the human understanding.

Neural networks that use Backpropagation can be seen as Black Box models, since the ways that network uses to obtain a specific result are unknown. A limitation of use refers to training time, which tends to be very slow. Thousands of cycles can be needed to reach acceptable levels of error, especially if serial computers are used in the simulation, since the processing unit must calculate the functions for each unit and its connections separately, which can be problematic for networks with a large amount of data.

There are no rules to define how many units must exist in intermediate layers, how many layers, or how the connections between these units should be. The processing is slow, however, for the data set used in this work, the neural network presents the second best accuracy, especially with numerical data, since it presents a larger amount of information to do the training process.

Tables 4 and 5 also show that the categorical data presented better results than numerical data when using kNN classifier, with IBK algorithm. This algorithm generated best results using Euclidean distance, with $k = 6$. The reason is because kNN classifier for categorical data uses only 3 classes to partitioning the categories, while for numerical data, it uses 10 classes (the 10 musical genres). Then, kNN presents higher accuracy for categorial data.

Table 6 presents the number of leaves and the size of each tree generated by Decision Tree models. When a prune process is applied, the complexity of the tree is reduced, consequently, the computational cost is also reduced. The prune process makes the trees easier to be interpreted without impact in the accuracy of the models. According to Tables 4 and 5, CART algorithm with pruning presents higher accuracy and higher area under ROC curve than other Decision Trees algorithms.

Despite of Decision Trees not present the best accuracy in relation to the other classifiers, the Decision Trees are the clearest models for to human understanding. These rules can provide important information for the musicotherapists, who will have an extra knowledge to guide them in the selection of the most appropriated music for the patient. An example of Decision Tree rule, from the CART model, extracted from Test Set validation, is "IF attribute 6 is greater than or equal to 0.067 AND attribute 13 is less than 0.464 AND attribute 8 is less than 0.276 AND attribute 11 is greater than or equal to 0.024 AND attribute 28 is less than 0.121 THEN the genre is Bolero". Thus, considering Table 3, the predominant emotion of Bolero genre is Romantic. For CART categorical model, an example of Test Set rule is "IF attribute 13 is low or medium AND attribute 16 is high or medium AND attribute 9 is medium or low AND attribute 6 is high AND attribute 21 is low AND attribute 18 is low or medium AND attribute 5 is low or high and attribute 28 is medium THEN the genre is Sertanejo". Thus, the predominant emotion is Sadness, according to Table 3. The number of the attributes can be checked in Table 2.

It is important to highlight that all classification models provided good results, accuracy higher than 60% and the area under ROC curve higher than 0.77. Then, all of them could be applied to classify new data and generate acceptable results. For this case study, Random Forest method presented the best results for the classification task, obtaining 0.972 for the area under ROC curve and 80.33% of accuracy for numerical data. Random Forest is a "White Box" method, as explained above, however, it does not generates a final tree from which a set of rules could be extracted.

References

[1] D. W. Aha, D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.

[2] J. P. Arzubiaga. Apuntes sobre el Bolero: desde la esclavitud africana hasta la globalización. *Revista Ciencias Sociales*, pages 95–117, 2007.

[3] J. J. Aucouturier and F. Pachet. Representing musical genre: a state of the art. *Jornal of New Music Research*, 32(1):83–93, 2003.

[4] P. Austerlitz. *Merengue: dominican music and dominican identity*. Temple University Press, 1997.

[5] A. Azarloo and F. Farokhi. Automatic musical instrument recognition using kNN and MLP neural networks. *IEEE-Fourth International Conference on Computational Intelligence, Communication Systems and Networks*, pages 289–294, 2012.

[6] L. Breiman. Bagging predictors. *Machine Learning*, 26(2):123–140, 1996.

[7] L. Breiman. Random Forest. *Machine Learning*, 45:5–32, 2001.

[8] L. Breiman, R. A. Olshen, and C. J. Stone. *Classification and regression trees*. Wadworth and Books/Cole Advanced Books and Software, 1984.

[9] R. Cabredo, R. Legaspi, P. S. Inventado, and M. Numao. An emotion model for music using brain waves. *13th International society for music Information Retrieval Conference*, pages 265–270, 2012.

[10] D. M. Chandwadkar and M. S. Sutaone. Selecting proper features and classifiers for accurate identification of musical instruments. *Insternational Journal of Machine Learning and Computing*, 3:172–175, 2013.

[11] D. Chathuranga and L. Jayaratne. Musical genre classification using ensemble of classifiers. *IEEE-Fourth International Conference on Computational Intelligence, Modeling and Simulation*, pages 237–242, 2014.

[12] D. C. Correa. *Artificial intelligence applied to musical genres analysis*. Doctoral Thesis. University of São Paulo. São Carlos Institute of Physics, 2012.

[13] F. Cougo-Jr. A historiografia da música gauchesca: Apontamentos para uma história. *Comtemporâneos Revista de Arte e Humanidades*, (10), 2012.

[14] T. Dietterich. An experimental comparison of three methods for constructing ensembles of Decision Trees: Bagging, Boosting and Randomization. *Machine Learning*, pages 1–22, 1998.

[15] T. Dietterich. The random subspace method for constructing Decision Forests. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[16] Z. Fu, G. Lu, K. M. Ting, and D. Zhang. A survey of audio-based music classification and annotation. *IEEE*, 13(2):303–319, 2011.

[17] L. Gagnon and I. Peretz. Musical structural determinats of emotional judgments in dementia of Alzheimer type. *Neuropsychology*, (1):90–97, 2009.

[18] M. Geretsegger, C. Elefant, K. A. Mssler, and C. Gold. Music therapy for people with Autism spectrum disorder. *Cochrane Database of Systematic Reviews*, 6, 2014.

[19] J. J. Guerrero-Turrubiates, S. E. Gonzalez-Reyna, and S. E. Ledesma-Orozco J. G. Avina-Cervantes. Pitch estimation for musical note recognition using artificial neural network. *IEEE*, pages 53–58, 2014.

[20] J. Han, M. Kamber, and J. Pei. *Data Mining: concepts and techniques.* Morgan Kaufmann, 2011.

[21] S. Haykin. *Neural networks and learning machines.* Pearson, 2009.

[22] T. Hillecke, A. Nickel, and H. V. Bolay. Scientific perspective on music therapy. *Annals New York academy of sciences*, pages 1–12, 2005.

[23] J. Jacobsen, J. Stelzer, T. H. Fritz, G. Chételat, R. La Joice, and R. Turner. Why musical memory can be preserved in advanced Alzheimer's disease. *Brain: A journal of neurology*, 138:2438–2450, 2015.

[24] P. Janata. The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, 19:2579–2594, 2009.

[25] Y. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull. Music emotion recognition: a state of the art review. *ISMIR*, pages 255–266, 2010.

[26] P. Knees and M. Schedl. *Music similarity and retrieval. An introduction to audio-and web-based strategies.* Springer, 2016.

[27] R. S. Kothe and D. G. Bhalke P. P. Gutal. Musical instrument recognition using k-nearest neighbour and support vector machine. *IEEE-International Conference on Advances in Eletronics, Communication and Computer Technology*, pages 308–313, 2016.

[28] B. Lantz. *Machine learning with R.* Packt Publishing, 2013.

[29] H. Lee and S. Kim. Black-box classifier interpretation using Decision Tree and Fuzzy Logic-based classifier implementation. *International Journal of Fuzzy Logic and Intelligent Systems*, 16(1):27–35, 2016.

[30] C. F. Lin and S. D. Wang. Fuzzy support vector machines. *IEEE Transactions on Neural Networks*, 13(1):464–471, 2002.

[31] A. S. Maratos, C. Gold, X. Wang, and M. J. Crawford. Music therapy for depression. *Cochrane Database of Systematic Reviews*, 1:1–22, 2008.

[32] C. McGowan and R. Pessanha. *The brazilian sound: samba, bossa nova and the popular music of Brazil.* Temple University Press, 1998.

[33] M. B. Mokhsin, N. B. Rosli, S. Zambri, N. D. Ahmad, and S. H. Hamidi. Automatic music emotion classification using Artificial Neural Network based on vocal and instrumental sound timbres. *Journal of Computer Science*, 12:2584–2592, 2014.

[34] K. Mssler, X. Chen, T. Heldal, and G. Christian. Musictherapy for people with schizophrenia and schizophrenia-like disorders. *Cochrane Database of Systematic Reviews*, 12, 2011.

[35] T. Nakata and S. E. Thehub. Infant's responsiveness to maternal speech and singing. *Infant behavior and development*, 27:455–464, 2004.

[36] E. Osuma, R. Freund, and F. Girosi. An improved training algorithm for support vector machines. *IEEE*, pages 276–285, 1997.

[37] E. Partanen, T. Kujala, M. Tervaniemi, and M. Huotilainen. Prenatal music exposure induces long-term neural. *Plos One*, 8:1–6, 2013.

[38] N. Patel and S. Upadhyay. Study of various tree pruning methods with their empirical comparison in Weka. *International Journal of Computer Applications*, 60:20–25, 2012.

[39] D. D. Patil, V. M. Wadhai, and J. A. Golhale. Evaluation of decision tree pruning algorithms for complexity and classification accuracy. *International Journal of Computer Applications*, pages 23–30, 2010.

[40] J. C. Platt. Sequential minimal optimization: a fast algorithm for training support vector machines. *Technical Report MSR-TR-98-14, Microsoft Research*, pages 1–20, 1998.

[41] J. R. Quinlan. Induction of Decision Trees. *Kluwer Academic Publishers*, pages 81–106, 1986.

[42] J. R. Quinlan. *C4.5: Programs for machine learning*. Morgan Kaufmann, 1993.

[43] J. H. Ribeiro. *Música caipira: as 270 maiores modas de todos os tempos*. Globo, 2006.

[44] A. G. Q. Rivera. *Salsa, sabor y control!: sociología de la música tropical*. Siglo Veintiuno, 1998.

[45] L. Rokach and O. Maimon. *Data mining with decision trees: theory and applications*, volume 69. World Scientific, 2008.

[46] C. L. Santos and C. N. Silla-Jr. The Latin music mood database. *Journal on Audio, Speech and Music Processing*, 2015.

[47] J. F. Santos. *A música como expressão do nordeste*. IBRASA, 2004.

[48] J. A. Sellers. *Bachata and dominican identity*. McFarland, 2014.

[49] C. N. Silla-Jr, A. L. Koerich, and C. A. A. Kaestner. The Latin music database. *Proc. International Society for Music Information Retrieval*, pages 451–456, 2008.

[50] C. N. Silla-Jr, A. L. Koerich, and C. A. A. Kaestner. A feature selection approach for automatic music genre classification. *International Journal of Semantic Computing*, 3(2):183–208, 2009.

[51] L. G. Soares. *É Necessário dois para bailar um tango*. Libertas, 2013.

[52] I. Stavans. *Latin music: musicians, genres and themes*. Greenwood, 2014.

[53] C. S. Tabarro, L. B. Campos, N. O. Galli, and N. F. Novo. Effect of the music in labor and newborn. *Rev Esc Enferm USP*, pages 441–448, 2010.

[54] S. Thompson. Decision making in music therapy the use of a Decision Tree. *The Australian journal of music therapy*, 24:48–64, 2013.

[55] Roman Timofeev. Classification and regression trees (cart) theory and applications - master thesis, 2004.

[56] H. Trappe. Music and medicine: The effects of music on the human being. *Applied cardiopulmonary pathophysiology*, 16:133–142, 2012.

[57] G. Tzanetakis and P. Cook. Marsyas: a framework for audio analysis. *Organized Sound*, 4:169–175, 1999.

[58] I. H. Witten, E. Frank, and M. A. Hall. *Data Mining: practical machine learning tools and techniques*. Morgan Kaufmann, 2011.

[59] F. Zhang, H. Meng, and M. Li. Emotion extraction and recognition from music. *Natural Computation, Fuzzy Systems and Knowledge Discovery*, 2016.