

**UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI**  
**Faculdade de Ciências Exatas**  
**Vinícius Ribeiro dos Santos**

**RECONHECIMENTO DE PADRÕES: UM ESTUDO SOBRE UMA BASE DE CIFRAS  
DE VIOLÃO.**

**Diamantina**  
**2019**



**Vinícius Ribeiro dos Santos**

**RECONHECIMENTO DE PADRÕES: UM ESTUDO SOBRE UMA BASE DE CIFRAS  
DE VIOLÃO.**

Monografia apresentada ao Curso de Sistemas de Informação do Departamento de Computação da Universidade Federal dos Vales do Jequitinhonha e Mucuri como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Eduardo Pelli

**Diamantina**

**2019**

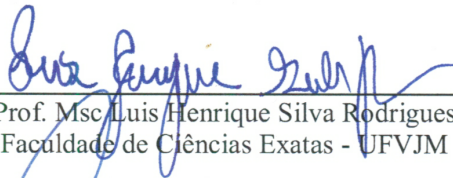
Vinicius Ribeiro dos Santos

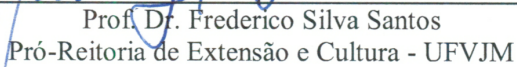
**RECONHECIMENTO DE PADRÕES: UM ESTUDO SOBRE UMA BASE DE  
CIFRAS DE VIOLÃO.**

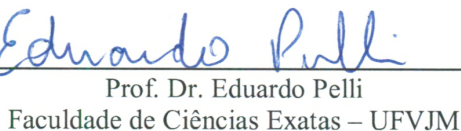
Monografia apresentada ao Curso de Sistemas de Informação do Departamento de Computação da Universidade Federal dos Vales do Jequitinhonha e Mucuri como requisito parcial para obtenção do título de Bacharel.

Orientador: Prof. Dr. Eduardo Pelli

Data de aprovação 19 / 07 / 2019.

  
\_\_\_\_\_  
Prof. Msc Luis Henrique Silva Rodrigues  
Faculdade de Ciências Exatas - UFVJM

  
\_\_\_\_\_  
Prof. Dr. Frederico Silva Santos  
Pró-Reitoria de Extensão e Cultura - UFVJM

  
\_\_\_\_\_  
Prof. Dr. Eduardo Pelli  
Faculdade de Ciências Exatas – UFVJM

**Diamantina**

**2019**

## AGRADECIMENTOS

Agradeço à meu orientador, Prof. Dr. Eduardo Pelli, pela paciência e dedicação durante a orientação deste trabalho. Agradeço também a todos os professores e colegas que, de diversas formas, me auxiliaram durante os anos de graduação em sistemas de informação.

Aos meus pais, pelos incontáveis esforços para quebrar o paradigma socioeconômico do qual viemos, me dando a oportunidade de ser, entre todas as gerações de nossa família, o primeiro a ingressar em um curso superior. Sempre dividi esse sonho com eles, e essa conquista só foi possível graças ao apoio incondicional que sempre recebi deles.

À Cíntia Sena, quem me deu muito incentivo e apoio em momentos de dificuldade e alegrias, especialmente durante a elaboração deste trabalho - e que também acabou se tornando minha principal referência acadêmica.

Aos professores do Conservatório Estadual de Música Lobo de Mesquita, pelos ensinamentos no curso de educação musical, que muito contribuíram para a decisão de envolver a música no meu trabalho de conclusão de curso. Em especial, Mayron e Luciana Vieira, por quem tenho admiração que vai além da técnica instrumental e didática.

A todos os músicos com quem já tive o prazer de dividir apresentações nos bares, auditórios, festas e repúblicas de Diamantina. Destaco nesse agradecimento todos os amigos que fiz nos projetos de extensão universitária: Coral Cênico UFVJM, Musicais Famosos e Coral Universitário - onde pude aprender muito sobre artes, de forma tão especial, em momentos de desconpressão dentro do próprio campus da universidade.

Por fim e não menos importante, agradeço aos meus amigos: Álvaro Vitor, Gleison Ferreira, Luciano Caldeira e minha irmã Laiane. Com eles com certeza dividi a maioria dos meus melhores momentos em Diamantina, fazendo com que eu sempre me sentisse em casa.



*“O sonho é a satisfação de que o desejo se realize.”  
(Sigmund Freud)*





## RESUMO

Com estudo de variáveis que influenciam no desestímulo do aprendizado de música, é possível inferir uma das causas frequentes do abandono do aprendizado: o material utilizado. E com a crescente ascensão da internet, muito material passou a ser compartilhado na rede mundial, deixando ao aluno a responsabilidade de filtrar todo o conteúdo em busca de material relevante para seu nível técnico. O objetivo deste trabalho é aplicar técnicas de descoberta de padrão, para estudar as características que definem o nível de dificuldade de uma cifra de violão, extraíndo e selecionando atributos em uma base de dados coletada, que possam ser relevantes para a tarefa de classificação apoiada por uma máquina de vetores de suporte. Os resultados da classificação considerando apenas dois atributos extraído e selecionados no experimento, sugerem o sucesso na aplicação da metodologia, apresentando acurácia de  $95,75\% \pm 1,86\%$  em um experimento com cifras de 240 músicas, estratificadas em relação à quatro níveis de dificuldade, e distribuídas entre treino (70%) e teste(30%).

Palavras-chave: Música. Reconhecimento de Padrões. Cifras. Aprendizado de Máquina. Classificação.



## **ABSTRACT**

By studying variables that influence the discouragement of music learning, it is possible to infer one of the frequent causes of learning abandonment: the material used. And with the growing rise of the Internet, much material is now being shared on the world wide web, leaving students with the responsibility to filter all content for material relevant to their technical level. The purpose of this paper is to apply pattern discovery techniques to study the characteristics that define the difficulty level of a guitar cipher by extracting and selecting attributes from a collected database that may be relevant to the classification task supported by a support vector machine. The results of the classification considering only two attributes extracted and selected in the experiment, suggest the successful application of the methodology, presenting accuracy of  $95,75\% \pm 1,86\%$  in an experiment with 240 stratified music codes in relation to four difficulty levels, and distributed between training (70 %) and test (30 %).

**Keywords:** Music. Pattern Recognition. Music Code. Machine Learning. Classification.



## LISTA DE ILUSTRAÇÕES

Figura 1 – Cifra da música Itamarandiba, de Milton Nascimento. . . . .	20
Figura 2 – Representação de funcionamento de uma SVM. . . . .	25
Figura 3 – Aplicação do kernel RBF em SVM. . . . .	26
Figura 4 – Representação da metodologia do experimento . . . . .	32
Figura 5 – Distribuição por classe da ocorrência dos acordes mais frequentes da base. . .	35
Figura 6 – Distribuição da quantidade de acordes por classe; escala natural. . . . .	36
Figura 7 – Distribuição da quantidade de acordes por classe; escala logarítmica. . . . .	36
Figura 8 – Distribuição estratificada das features <i>chordCount</i> e <i>chordFreq</i> . . . . .	37



## LISTA DE TABELAS

Tabela 1 – Sequência de notas da escala de <i>dó maior</i> . . . . .	21
Tabela 2 – Composição do acorde <i>dó maior</i> . . . . .	21
Tabela 3 – Representação de uma matriz de confusão. . . . .	26
Tabela 4 – Distribuição de registros por classe. . . . .	31
Tabela 5 – Descrição dos atributos da base de dados. . . . .	32
Tabela 6 – Análise descritiva dos dados de cifras. . . . .	33
Tabela 7 – Coeficientes de correlação de Pearson com a característica <i>hardness</i> . . . . .	37
Tabela 8 – Resultados de classificação para a classe 1 (Iniciante). . . . .	38
Tabela 9 – Resultados de classificação para a classe 2 (Fácil). . . . .	39
Tabela 10 – Resultados de classificação para a classe 3 (Intermediário). . . . .	40
Tabela 11 – Resultados de classificação para a classe 4 (Avançado). . . . .	41
Tabela 12 – Resultados estratificados de classificação. . . . .	42
Tabela 13 – Cifras consideradas no experimento. . . . .	50





## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>17</b>
<b>2 OBJETIVOS</b>	<b>19</b>
<b>3 FUNDAMENTAÇÃO TEÓRICA</b>	<b>20</b>
<b>3.1 Cifragem Musical</b>	<b>20</b>
<b>3.2 Reconhecimento de Padrões</b>	<b>21</b>
3.2.1 Tipos de Aprendizado	22
3.2.1.1 Aprendizado Supervisionado	22
3.2.1.2 Aprendizado Não Supervisionado	23
3.2.2 Pré-Processamento	23
3.2.2.1 Extração de Características	23
3.2.2.2 Seleção de Características	24
3.2.3 Classificação	25
3.2.3.1 Support Vector Machine	25
3.2.4 Avaliação	26
3.2.4.1 Acurácia	27
3.2.4.2 Precision	27
3.2.4.3 Recall	27
<b>4 MATERIAIS E MÉTODOS</b>	<b>28</b>
<b>4.1 Materiais</b>	<b>28</b>
4.1.1 Recurso Computacional	28
4.1.2 Anaconda Navigator	28
4.1.3 Jupyter Notebooks	28
4.1.4 Python	29
4.1.5 Scikit-Learn	29
<b>4.2 Métodos</b>	<b>30</b>
<b>5 RESULTADOS E DISCUSSÃO</b>	<b>35</b>
<b>6 CONCLUSÕES</b>	<b>43</b>
<b>6.1 Trabalhos Futuros</b>	<b>43</b>
<b>REFERÊNCIAS</b>	<b>43</b>
<b>APÊNDICE</b>	<b>50</b>
<b>Base de Dados</b>	<b>50</b>



## 1 INTRODUÇÃO

O estudo sobre as razões que motivam o ser humano a se submeter ao aprendizado de música já é abrangido em escala global, e tema de grandes trabalhos argumentativos.

Em McPherson (2012) os autores argumentam que o processo de aprendizado musical envolve uma tendência natural dos seres a demonstrar prazer em envolverem com o aprendizado de música, principalmente porque essa pode ser uma fonte de atenção da sociedade para uma habilidade das quais nem todos dispõem. Ainda de acordo com os autores, existe também um contraponto pra essa característica do aprendizado: o processo de aprendizado é constantemente interferido pelo temor de que fique evidente alguma eventual dificuldade durante o processo. Ambos os fenômenos são notados com maior frequência nas crianças, principalmente por causa de a infância ser considerada como a fase onde se desenvolvem os interesses pessoais, bem como o início da formação técnica e intelectual de qualquer cidadão.

Para AMATO (2008), um dos fatores que afastam o Brasil do ensino popular de música é o baixo status atribuído à essa disciplina no currículo de um cidadão em relação aos demais conteúdos considerados na matriz do ensino básico escolar. Ainda de acordo com o autor, isso se deve principalmente à atribuição de um título de área menos exigente em relação à cognição, uma vez que a música, assim como todas as artes, tem uma ênfase maior na própria liberdade de se expressar. Apesar disso, Vanzela, Oliveira e Carvalho (2017) apresenta uma contradição à esse argumento, afirmando que a riqueza técnica exigida para executar exigem mais do que conhecimentos espontâneos, e que o aprendizado se dá exclusivamente por meio de esforços nos estudos. Ainda de acordo com os autores, uma performance imprecisa por conter erros, ou faltar elementos técnicos ou afetivos em expressão, sim, pode ser considerada um exemplo de exercício de aprendizagens espontâneas, sendo claramente distinguidas dos formais.

Para Wigfield e Eccles (2000) o principal valor considerável para a construção do interesse de um determinado indivíduo em realizar uma tarefa de aprendizado é na verdade uma relação intrínseca da percepção criada pelo mesmo entre seus objetivos e suas memórias. O autor defende que um iniciante se mantém interessado em aprender música, basicamente, se for positiva a percepção sobre a sua competência assumida em relação a dificuldade atribuída para a realização de cada tarefa considerada no aprendizado. E, assim conclui que a dificuldade atribuída a cada tarefa dentro de um material ou método pode levar o aluno a se empenhar em evoluir ou desistir.

Apesar da complexidade envolvida no aprendizado musical, Green (2017) cita a importância de exercícios informais para relacionar o conhecimento absorvido em ambiente acadêmico, e manter o aluno em constante evolução. A música popular tem papel fundamental nessa forma de aprendizado, uma vez que o interesse, sobretudo de alunos jovens, é mais aguçado para a música tangente ao popular. Esse aprendizado é considerado bastante enriquecedor principalmente por desenvolver habilidades técnicas e fomentar o anseio pelo aprendizado.

(COUTO, 2009)

O método de cifragem é uma evolução da escrita musical tradicional que popularizou o violão, por ser acessível, e viabilizar o estudo informal de iniciantes em teoria musical, sobretudo com a ascensão da tecnologia e disponibilização gratuita de cifras na internet. Esse é um dos métodos mais populares de introdução ao violão, e frequentemente é utilizado como alternativa no aprendizado informal de música. (JUNIOR et al., 2017)

## 2 OBJETIVOS

O objetivo deste trabalho é apresentar uma abordagem de aplicação de técnicas de reconhecimento de padrões em um problema de classificação de níveis de dificuldade em cifras de violão. Extrair e selecionar características, treinar e testar um classificador e avaliar o desempenho na classificação são os principais processos que compõem a metodologia a ser seguida.

Espera-se ao final do experimento, ter conclusões a respeito da viabilidade da aplicação da abordagem mencionada, para o problema em questão.

### 3 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os conceitos básicos para a compreensão deste estudo. A primeira seção apresenta o método de escrita musical sobre o qual se desenvolve o experimento descrito no decorrer deste trabalho, enquanto a segunda seção descreve os fundamentos da tecnologia aplicada no desenvolvimento do mesmo.

#### 3.1 Cifragem Musical

A escrita musical é muito importante no processo de aprendizado de qualquer instrumento, por ser a forma com que os músicos comunicam-se e transmitem conhecimento. Diversas alternativas de escrita musicais surgiram com o objetivo de tornar mais precisa a representação das instruções para a execução de uma música através da escrita. (JUNIOR et al., 2017)

A cifragem<sup>1</sup>, popularmente conhecida apenas por ‘cifra’, é definida por Chediak (1986) como sendo o uso de símbolos criados para representar o acorde de uma maneira prática. A cifra é composta de letras, números e sinais. É o sistema predominantemente usado em música popular para qualquer instrumento.

Taborda (2011) complementa essa definição afirmando que a cifra comumente é associada ao conjunto de representação de acordes junto à letra da canção - conforme apresentado no exemplo da Figura 1, formando assim uma estrutura legível para interpretação da música através de um instrumento. Segundo a autora, esse é o padrão moderno mais utilizado em qualquer instrumento harmônico, devido principalmente ao fato de a cifragem trabalhar a representação de acordes.

Figura 1 – Cifra da música Itamarandiba, de Milton Nascimento.

**Itamarandiba**  
**Milton Nascimento**  
 Composição: Fernando Brant / Milton Nascimento  
 Tom: G

C G C G C G D7 G	C/D D7 G7+	Am G Am G Am G Em
No meio do meu caminho	Pedra miúda rolando sem vida	Sonharás com Pedra Azul
C G C G C G D7	Cm Bm	C D Em
Sempre haverá uma pedra	Como é miúda e quase sem brilho	Viverás em Diamantina
Am G Am G Am G Am Em	F#m C D Em	C G C G C G D7 G
Plantarei a minha casa	A vida do povo que mora no vale	No caminho dessa cidade
C D Em	C G C G C G D7 G	C G C G C G D7
Numa cidade de pedra	No caminho dessa cidade	As mulheres são morenas
A4/7 A7	C G C G C G D7	Am G Am G Am G Em
Itamarandiba, pedra corrida	Passarás por Turmalina	Os homens serão felizes
		C D Em
		Como se fossem meninos

Fonte: CifraClub (2019)

<sup>1</sup> Devido à concordância popular em referenciar a cifragem apenas por ‘cifra’, este trabalho adere esta referência em todo o texto, tratando toda a estrutura que representa uma música em escrita cifrada, apenas por cifra.

Um acorde, por sua vez, pode ser definido como sendo um conjunto de três ou mais notas que são tocadas simultaneamente. A principal nota do acorde é denominada fundamental (ou tônica), enquanto as demais recebem o nome de seu intervalo (distância dentro da escala musical) em relação à nota fundamental. (PASSAMAE, 2011)

A Tabela 1 apresenta a sequência de notas da escala de dó maior, enquanto a Tabela 2 mostra as notas da escala e seus intervalos em relação à nota dó, destacando as notas (grifadas) que compõem o acorde de dó maior.

**Tabela 1 – Sequência de notas da escala de dó maior.**

dó	ré	mi	fá	sol	lá	si	dó
----	----	----	----	-----	----	----	----

Fonte: Elaborada pelo autor.

**Tabela 2 – Composição do acorde dó maior.**

Nota	<b>dó</b>	ré	<b>mi</b>	fá	<b>sol</b>	lá	si	dó
Intervalo	<b>Fundamental</b>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	5 <sup>a</sup>	6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>

Fonte: Elaborada pelo autor.

Taborda (2011) afirma que a cifragem viabilizou a popularização do aprendizado de instrumentos musicais, sendo fortemente utilizada desde os anos setenta, quando ainda nem se contava com o advento da internet.

Com os avanços da tecnologia, e a massificação do acesso à internet o método ganhou ainda mais popularidade, e é inclusive citado ainda por Taborda (2011) como fator definitivo para o crescimento do número de interessados por aprender violão, tornando-o definitivamente o instrumento mais tocado entre todos os estilos musicais e classes sociais do Brasil.

### 3.2 Reconhecimento de Padrões

Bilhões de usuários geram, para as mais diversas áreas da ciência, uma abundância incalculável de dados através de diversos serviços de computação, sobretudo nos últimos anos, com o acesso massivo da população à rede mundial de computadores. Nestas circunstâncias, apresenta-se mais um novo desafio técnico para a computação: a criação de alguma ferramenta, técnica ou metodologia que possibilita-se retirar conhecimento relevante a partir de base de dados com bilhões de registros sendo acrescentados por todo tempo, e ainda tendo características de dinamismo em diversos casos. (JAGADISH et al., 2014)

O reconhecimento de padrões é um campo de estudo que trata de classificações ou agrupamento de registros em conjuntos de características rotuladas por classes. Estas características podem ser de diferentes tipos, a depender do tipo de aplicação a que se destina o

processamento. Problemas famosos envolvem a classificação de sinais de áudio, textos, e imagens.(THEODORIDIS et al., 2006)

Nos últimos anos as limitações computacionais começaram a ser vencidas, e o número de aplicações vem sugerindo um crescente interesse pela área, já que modelar e processar já não são tarefas custosas computacionalmente, mesmo em extensas bases de dados. (KASABOV, 1996)

Algumas aplicações relevantes que empregam de reconhecimento de padrões são levantadas por Kasabov (1996), entre elas:

- Análise de sentimentos em bases de textos;
- Reconhecimento de biometria;
- Apoio em laudos médicos;
- Predições de acidentes ambientais;
- Identificação de rostos humanos;
- Análises de sensoriamento remoto;

### 3.2.1 Tipos de Aprendizado

Dois paradigmas de aprendizado de máquina são os mais tradicionais e concentram a maioria dos trabalhos descritos em toda a bibliografia que referencia este assunto: o aprendizado supervisionado e o aprendizado não-supervisionado.

#### 3.2.1.1 Aprendizado Supervisionado

Nesta abordagem os dados são compostos pelas features (atributos) e uma dessas features é denominada *label* (ou simplesmente 'classe'), que é exatamente a característica de interesse. Os dados de treino são entregues ao algoritmo classificador na etapa de treinamento, ainda com a *label*. O objetivo é que o classificador calcule hipóteses sobre a *label*, e as valide com o acesso à variável. (MORAES, 2016)

No segundo momento, os dados de teste são entregues ao classificador, que se baseia nos padrões encontrados na base de treino, para que sejam rotuladas. O detalhe dessa segunda etapa é que os dados de teste não acompanham a *label*. O objetivo é criar uma representação tão consistente na primeira etapa, que o algoritmo possa responder, na segunda etapa, quais são as classes dos dados analisados. (MORAES, 2016)

Quando um problema trata dados com valor de classe representado por um número discreto, este é entendido como um problema de classificação. Se o dado de classe tiver representação por um número contínuo, então o problema é chamado de regressão. (SILVA; SANTOS, 2017)



### 3.2.1.2 Aprendizado Não Supervisionado

Em contraste ao aprendizado supervisionado, esse paradigma é totalmente independente de referências diretas para atribuir, à uma configuração de dados, determinada classe. Isso porque não uma etapa prévia de conhecimento das classes. Os dados são classificados (ou agrupados) com o apoio de medidas de similaridade entre as entradas submetidas. (LORENA; CARVALHO, 2007)

De acordo com Lorena e Carvalho (2007), esse tipo de abordagem é comumente aplicada em bases de onde se deseja tirar padrões que sejam capazes de apoiar uma representação esclarecida da base de dados.

### 3.2.2 Pré-Processamento

A etapa de pré-processamento da base de dados compreende processos elementares na construção da base a ser trabalhada. Obtenção e organização dos dados, identificação de padrões de interesse, remoção de ruído, normalização e trabalho em representações que sejam compactas, são alguns dos processos básicos da etapa de pré-processamento. (MORAES, 2016).

De acordo com Bishop (1996), um dos desafios frequentes em processamento de dados, sobretudo quando se tratam de grandes bases de dados, é o custo computacional para processar arquivos de entrada com alta dimensionalidade. Técnicas que combinam variáveis (features) dos dados de entrada, para gerar um único atributo com maior valor representativo, podem ajudar nesse problema nesse problema. Essas técnicas podem ser aplicadas de forma tradicional, baseadas no conhecimento da base de dados, e as peculiaridades, com o apoio de métodos metodológicos para tal. Esses métodos são conhecidos por extração e seleção de características.

#### 3.2.2.1 Extração de Características

Nessa etapa, o principal objetivo é reunir as principais características que descrevem as peculiaridades de cada uma das classes predefinidas no problema, em um conjunto menor de features que sejam bastante representativas. (MARQUES; REIS; CAMPOS, 2017)

Duda, Hart e Stork (2001) afirma que durante essa etapa, é muito importante utilizar análises de medidas que são associadas ao objeto que se deseja obter (em casos de classificação, esse objeto seria a variável que representa a classe), para ter sucesso na obtenção de features que valorizem a similaridade entre os objetos de mesma classe, e a discrepância entre os objetos de classes diferentes.

A etapa de extração de características é muito importante para alcançar resultados significativos em projetos de classificação. Isso porque grande parte da base de treinamento dos modelos de classificação são definidas nessa etapa. (SANTOS et al., 2007)

### 3.2.2.2 Seleção de Características

Félix (1998) afirma que extrair conhecimento diretamente de uma base de dados vem se tornando cada vez menos interessante, dado o aumento no fácil acesso à dados, que por sua vez gera acúmulo de dados que carecem de processamento, e a própria limitação na manipulação de dados por parte dos algoritmos de aprendizado de máquina. O resultado de grandes bases de dados para a entrada dos algoritmos é um espaço de busca demasiadamente amplo. Ademais, grandes volumes de registros, podem carregar muita redundância, forçando os algoritmos a reforçarem padrões irrelevantes para o retorno no processamento. O autor menciona ainda exemplos como “as pessoas que ficam grávidas são mulheres” e “se a cidade é São Carlos então o estado é São Paulo” para representar padrões que não acrescentam nos resultados.

Mannila (2012) aponta que a seleção de características é o processo que geralmente mais consome tempo dentro da descoberta de conhecimento, chegando a representar 80% do tempo total de um projeto. A autora justifica a importância desse processo como sendo o motivo para o custo do mesmo. Por sua vez, o processo se faz importante principalmente porque é onde são definidas as variáveis que melhor representam a base, baseando-se no conceito de máxima relevância e mínima redundância.

O princípio que define a relevância de uma feature (atributo) é a possibilidade nativa de a mesma evidenciar a classificação de dado registro, quer essa feature esteja ou não em conjunto com outras features. A redundância, por sua vez, em tese é o conceito do indicador que quantifica o nível de dependência observado entre duas ou mais features do conjunto de dados. (FERNANDES, 2017)

Segundo Félix (1998), seleções de características podem considerar aspectos como:

- Dimensionalidade da base de dados.
- Técnicas para obtenção de características (domínio do problema, índices estatísticos, etc.).
- Consistência das características.
- Comportamento dos dados em relação ao crescimento da base.

Um avaliador bastante utilizado na bibliografia (VINISKI; GUIMARÃES, 2017; MICHELON et al., 2018; OLIVEIRA et al., 2017) para mensurar a importância de cada feature, dentro da etapa de seleção de características, é o coeficiente de correlação de Pearson.

O coeficiente de correlação de Pearson é um índice que varia entre -1 e 1. O valor indica a magnitude da correlação enquanto o sinal corresponde ao sentido da correlação. Intuitivamente, quanto mais o valor de correlação que se aproxima de um, mais intensa é a correlação entre o par de variáveis. Em contrapartida, quanto mais o valor se aproxima de zero, menor é a associação. O caso em que o valor da correlação é igual a zero, indica que as variáveis não se correlacionam (são ortogonais). Duas variáveis com valor de correlação positivo alto, apontam

que quando uma cresce, a outra cresce também. Sendo assim, pode-se afirmar que os valores altos de uma variável está associado a valores altos de outra. (ALEXANDRE et al., 2014)

### 3.2.3 Classificação

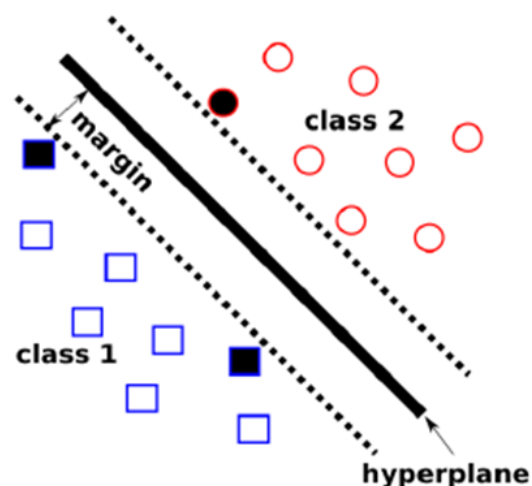
Uma das técnicas de mineração de dados mais conhecidas e utilizada é a classificação de dados. Essa técnica se fundamenta na ideia de encontrar um modelo matemático que seja capaz de descrever um subconjunto (classe) de dados, utilizando inteligência artificial. (HAN; PEI, 2012)

Algoritmos classificadores tem como finalidade a tarefa de rotular as instancias de uma base de dados, de forma sistemática, em classes, aplicando a modelagem que foi anteriormente aprendida. A construção do modelo pelo algoritmo parte dos atributos valorados, que estão instanciados nos dados de treinamento, para a aplicação dos classificadores. (LUCCA et al., 2013)

#### 3.2.3.1 Support Vector Machine

A base de uma *support vector machine* (SVM) é a busca por hiperplanos que permitam distanciar ao máximo os pontos mais próximos entre todas as classes. Os vetores de suporte são as estruturas que separam esses conjuntos, se limitando a uma única restrição: o ponto de início e fim precisa valorizar a maior distância possível de quaisquer pontos de todas as classes, conforme ilustrado na Figura 2. É desses vetores, e sua importância para o algoritmo, que o nome se deriva. (CESCHIN; BURDA, 2015)

Figura 2 – Representação de funcionamento de uma SVM.



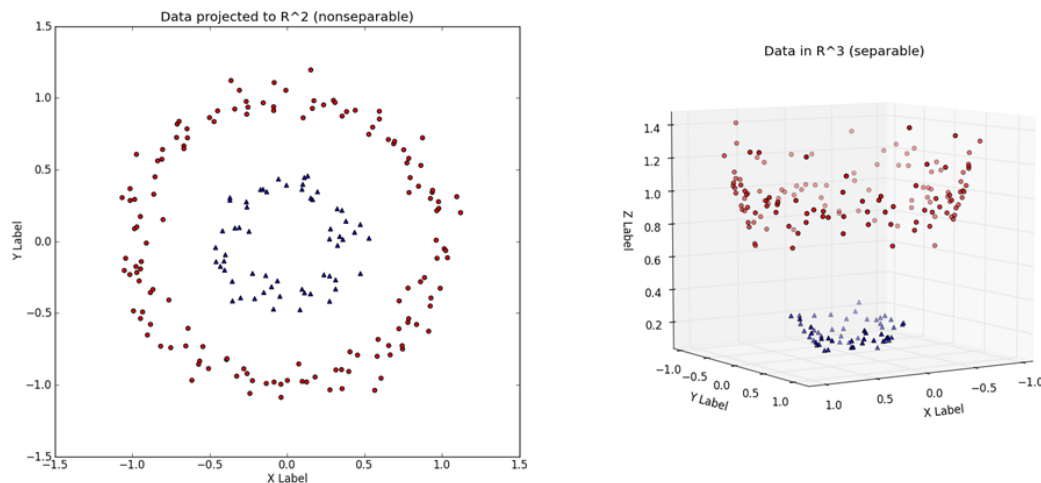
Fonte: (PEDREGOSA et al., 2011)

O SVM é naturalmente um classificador caracterizado por ser linear em sua natureza, mas ainda pode ser aplicável a dados cujo padrão não tenha separações lineares. (JÚNIOR, 2018)

Na maioria dos problemas, os dados não tem uma separação linear, então a função de *kernel* se faz necessária para projetar os dados em hiperplanos que viabilizem a separação por um vetor. Na implementação da biblioteca *scikit-learn*, essa função é habilitada através de um parâmetro passado no momento da declaração do classificador, esse parâmetro é chamado, não por coincidência, de *kernel*. (CESCHIN; BURDA, 2015)

A Figura 3 ilustra a aplicação de um kernel da implementação da *scikit-learn*: *radial basis function* (RBF). O exemplo representado considera um conjunto de dados cuja distribuição seria linearmente impossível de separar, apresentando como a técnica de kernel trick cria uma nova distribuição para os dados, criando uma simples possibilidade para agrupar os dados de diferentes classes.

**Figura 3 – Aplicação do kernel RBF em SVM.**



Fonte: (PEDREGOSA et al., 2011)

### 3.2.4 Avaliação

A matriz de confusão carrega resultados importantes a respeito da performance de um algoritmo, sendo as principais: a quantidade de instancias corretamente classificadas e a quantidade de classificações errôneas. Com a matriz ainda é possível entender a quantidade de vezes que o classificador julgou que uma instância seria pertencente a uma classe, quando na verdade pertencia a outra classe. (RABELO et al., 2017)

A matriz de confusão tem o formato no padrão representado na Tabela 3.

**Tabela 3 – Representação de uma matriz de confusão.**

	<b>Classe 1</b>	<b>Classe 2</b>
<b>Classe 1</b>	verdadeiroPositivos	falsoPositivos
<b>Classe 2</b>	falsoNegativos	verdadeiroNegativos

Fonte: Elaborada pelo autor.

### 3.2.4.1 Acurácia

Este é o índice de avaliação mais comum em trabalhos das mais diversas áreas, principalmente pela sua base teórica simplista. Representa a taxa de positivos e negativos que foram classificados de maneira correta. O cálculo da acurácia se dá pela equação 1:

$$Acurácia = \frac{\text{verdadeiroPositivos} + \text{verdadeiroNegativos}}{\text{(total)}} \quad (1)$$

### 3.2.4.2 Precision

Precision (ou *Precisão*) indica o total de verdadeiro positivos divididos pela soma de verdadeiro positivos e falso positivos, de acordo com a equação 2:

$$Precision = \frac{\text{verdadeiroPositivos}}{\text{(verdadeiroPositivos} + \text{falsoPositivos)}} \quad (2)$$

### 3.2.4.3 Recall

Recall (ou *Cobertura*) é uma métrica para representar a razão entre os verdadeiro positivo e a soma de verdadeiro positivo com falsos negativos. O cálculo que expressa o índice está representada pela equação 3:

$$Recall = \frac{\text{verdadeiroPositivos}}{\text{(verdadeiroPositivos} + \text{falsoNegativos)}} \quad (3)$$

## 4 MATERIAIS E MÉTODOS

Neste capítulo, é apresentado o material utilizado e as estratégias seguidas no decorrer do experimento. O material está descrito na primeira seção, enquanto os processos são descritos na segunda seção.

### 4.1 Materiais

Esta seção apresenta brevemente as principais ferramentas utilizadas para a realização do trabalho.

#### 4.1.1 Recurso Computacional

O experimento computacional descrito nesse trabalho foi inteiramente desenvolvido em uma máquina com processador quad-core @ 3,6 GHz, 8 gigabyte de memória RAM, e 120 gigabyte em solid-state drive. A máquina em questão é apoiada pelo sistema operacional Windows 10 Home.

#### 4.1.2 Anaconda Navigator

Anaconda Navigator é uma plataforma muito conhecida no meio científico, sendo utilizada por mais de 4,5 milhões de usuários pelo mundo. (NAVIGATOR, 2019)

Pode ser considerada um ambiente de facilitação de acesso a maioria das ferramentas relevantes no tratamento de dados através da linguagem python, inclusive a instalação da própria linguagem. O conda é o terminal por onde se pode instalar, remover ou atualizar quaisquer pacotes para a linguagem de programação python. O Anaconda é apenas uma interface gráfica que acrescenta muito ao projeto, tornando as ferramentas acessíveis a interessados que possam ter os mais diversos níveis de familiaridade com a programação. A instalação está disponível para download gratuito no site oficial <sup>2</sup> do projeto. (VIEIRA; VALADAO, 2017)

#### 4.1.3 Jupyter Notebooks

O Jupyter Notebook é uma ferramenta do projeto open-source *jupyter*<sup>3</sup> que funciona como um ambiente de interação amigável para a criação, edição, execução e até mesmo compartilhamento de código, textos com visualizações estilizadas, fórmulas matemáticas. Por ser construído com estrutura baseada em HTML5, o projeto é multiplataforma e é acessado através um web browser. O jupyter tem suporte a diversas linguagens de programação, além de ser referencia na criação de conteúdo interativo. (CASTELLUCCI, 2017)

---

<sup>2</sup> anaconda.org

<sup>3</sup> jupyter.org

Segundo José, Almeida e Paiva (2017) a palavra notebooks do nome, explica o formato de funcionamento do jupyter, onde cada arquivo realmente funciona como um caderno, onde temos várias possíveis páginas sendo facilmente acessíveis e editáveis. Essa organização pode ser muito proveitosa para a fácil execução de trechos isolados de código, que possibilita uma disposição intuitiva para códigos baseados em metodologias de desenvolvimento estruturadas em etapas. Essa forma de distribuição de código em células também torna os projetos mais legíveis e didáticos, sem a necessidade de preocupação em lidar com estruturas de importação e links de arquivos, tarefa que pode ser complexa para determinado níveis de experiência com a programação.

A estrutura de um projeto de descoberta de conhecimento segue exatamente o padrão descrito no parágrafo anterior, e principalmente por isso a ferramenta é utilizada neste trabalho para desenvolvimento de código da solução computacional proposta.

#### 4.1.4 Python

Python foi lançado por Guido Van Rossum, em 1991 e desde então é tida como uma tecnologia de código livre. É uma linguagem de programação popular, principalmente por ser ter uma sintaxe simples e fácil interpretação. É uma linguagem denominada multiplataforma que permite a programação através de paradigmas tanto conceitual como orientação a objetos. (DOWNEY, 2015)

Sua simplicidade muito se deriva de uma boa proposta de biblioteca nativa, que por ser robusta, facilita a programação. Isso faz com que códigos enormes em diversas linguagens sejam reescritos em poucas linhas de comando no ambiente do Python. (SILVA, 2014)

Marques et al. (2011) demonstrou a versatilidade da linguagem ao introduzir alunos do ensino médio à programação por meio da mesma. O autor concluiu que a possibilidade de trabalhar com jogos de uma maneira menos complexa foi um fator crucial para despertar e manter o interesse dos alunos, e o resultado do trabalho foram jogos codificados pelos próprios alunos, iniciantes na programação.

#### 4.1.5 Scikit-Learn

Scikit-Learn é uma biblioteca de código aberto, que reúne uma diversidade de algoritmos famosos para processamento de dados. As implementações são na linguagem python e podem ser utilizadas em qualquer sistema operacional, desde que previamente configurada. (SILVA, 2017)

Xavier e Souza (2018) destaca como características de relevância da biblioteca o fato de sua instalação ser simples e a documentação facilmente acessada através do site<sup>4</sup> oficial do projeto. Além de ter muita documentação oficial, a biblioteca é amplamente utilizada em problemas das mais distintas naturezas, o que torna a comunidade muito ativa e fonte de documentação extraoficial.

---

<sup>4</sup> [scikit-learn.org](http://scikit-learn.org)

O principal motivo de inúmeros trabalhos considerarem as implementações da biblioteca é a possibilidade de ter resultados de forma rápida, possibilitando maior dedicação na compreensão da base de dados e seus atributos. Além do fato de através da biblioteca ser possível usar tantos algoritmos quanto se queira, sem requerer habilidades aprofundadas em codificação, o que possibilita o levantamento de diversas possibilidades de algoritmos, e avaliação do impacto de cada abordagem sobre as peculiaridades da base submetida.

## 4.2 Métodos

Os dados estudados nesse trabalho foram coletados da plataforma cifraclub<sup>5</sup>, em maio de 2019. O conteúdo acessado e coletado é, em sua totalidade, publicado no site.

Para agilizar a coleta foi utilizado um *script* com um *crawler*, que recebe o link da página onde a cifra é publicada, e retorna um arquivo estruturado em formato *json* com os dados da cifra da música. O *crawler* foi escrito na linguagem python e utiliza métodos da biblioteca *urllib*<sup>6</sup> para abrir os endereços das cifras e coletar os dados publicados na página da mesma.

A base é composta por 240 registros de cifras de violão, filtradas entre as listas de maior acesso no contexto geral da plataforma. Além disso, para inserir maior diversidade no contexto do problema, também foram incluídas cifras mais acessadas de diferentes gêneros musicais. As listas mencionadas tem um formato de ranking temporal, e portanto são dinâmicas. Os registros considerados foram acessadas no decorrer deste trabalho, e podem não representar a realidade em outros momentos. A Tabela 6.1 descreve todas as entradas deste experimento.

A característica de interesse da base de dados estudada no experimento descrito neste trabalho, é a *tag* de dificuldade atribuída no registro. Como o experimento em questão descreve um trabalho de classificação, a variável de interesse pode ser referenciada apenas por ‘classe’ a partir deste ponto, indicando o atributo que agrupa os dados.

Uma preocupação em considerar cifras com uma *tag* de classe consistente com os dados em questão foi essencial pra que o trabalho pudesse ter continuidade. A *dificuldade* é responsável por agrupar os dados de diferentes classes do problema, e portanto precisa ser consistente. Para tal, as cifras foram submetidas à uma vista prévia, e validadas antes de ser coletadas.

A validação dessa consistência se baseia na definição das características de cada um dos quatro níveis de dificuldade considerados no problema: iniciante, fácil, intermediário, e difícil. Para definição dos critérios que compõem a dificuldade de cada um dos quatro níveis, foram considerados aspectos que envolvem a técnica (habilidades motoras) necessária no instrumento, bem como o conhecimento teórico-musical envolvido em cada nível de dificuldade atribuído. Assim sendo, as definições abrangem mais que a dificuldade envolvida em executar os acordes no violão, mas também a coesão entre conhecimentos esperados e correspondidos

---

<sup>5</sup> cifraclub.com.br

<sup>6</sup> docs.python.org/2/library/urllib2



pelo aprendiz em cada nível, fazendo com que o nível englobe apenas aquele que compreenda a complexidade na formação de cada estrutura musical embutida nas músicas executadas.

As definições dos níveis de dificuldade mencionados são:

- **Nível 1: Iniciante;** são cifras que contém apenas acordes simples de serem montados no braço do violão, sem nenhum tipo de variação. As sequências desses acordes são bastante repetidas, e consideradas bons exemplos de primeiras músicas a serem praticadas. (CABRAL et al., 2001)
- **Nível 2: Fácil;** algumas músicas tem muitos elementos em comum com o nível caracterizado anteriormente, mas com um elemento definitivo no aprendizado do violão: a pestana. Acordes que demandam dessa estrutura de formação, exigem muito estudo e postura de um iniciante, e é popularmente considerado um ponto de separação de entusiastas para estudantes do violão, conforme cita Brazil (2017).
- **Nível 3: Intermediário;** elementos dos níveis iniciante e fácil se repetem no nível intermediário, incluindo alguns acréscimos. Conceitos fundamentais da música aparecem também nas cifras deste grupo. Acordes com composição composta, ou seja, que consideram variações em sua construção (como sétimas, nonas, quintas), e também acidentes (sustenidos e bemóis) passam a compor as sequências. Essa inclusão expande muito o dicionário de acordes possíveis de aparecerem nas cifras, e como mostra DANTAS (2018), os conceitos aqui descritos já criam sequências de acordes que exigem um maior grau de experiência para a compreensão e execução.
- **Nível 4: Avançado;** considerados para estudantes aprofundados na teoria e prática do violão, os acordes apresentados nas cifras que compõe esse grupo são em sua maioria, acordes compostos por múltiplas variações. Outra característica desse grupo é o tamanho do dicionário de acordes presente em cada cifra. Executar muitos acordes tão elaborados tecnicamente, exige muito entendimento de teoria e prática em paralelo.

Com o objetivo de entender bem todos diferentes níveis de dificuldade, a base considera amostras de igual tamanho em todas as quatro diferentes classes. Isso caracteriza a base como balanceada. De acordo com Batista (2003), trabalhar com uma base de dados balanceada pode tornar mais simples as etapas de processamento e avaliação.

A Tabela 4 apresenta a distribuição de registros em relação à classe.

**Tabela 4 – Distribuição de registros por classe.**

	<b>Iniciante</b>	<b>Fácil</b>	<b>Intermediário</b>	<b>Avançado</b>	<b>Total</b>
<b>Registros</b>	60	60	60	60	240

Fonte: Elaborada pelo autor.

Os dados disponíveis inicialmente eram compostos pelo nome da música, nome do artista, gênero musical do artista, cifra (letra com acordes), tom da música e dificuldade.

Com base no conhecimento sobre o problema, os atributos disponíveis foram combinados na etapa de extração de características. Foram adicionados novos atributos à base de dados. A Tabela 5 apresenta todos os atributos da base após a extração de características.

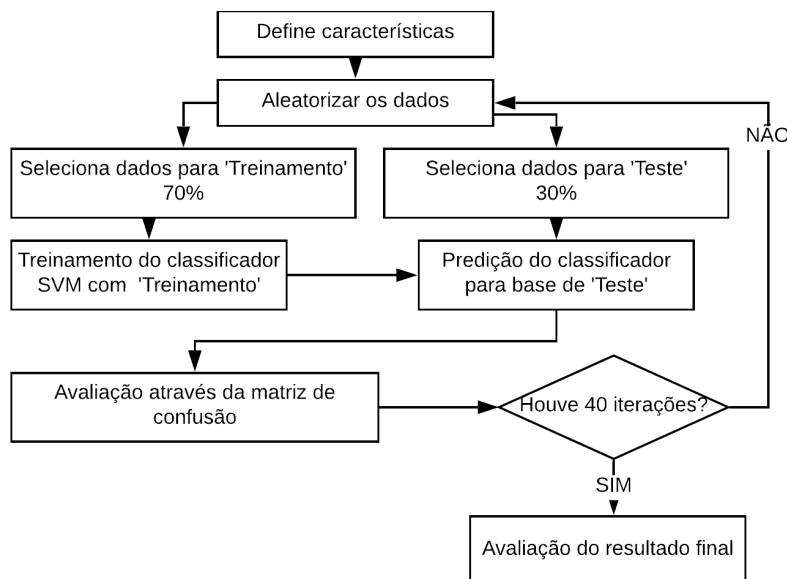
**Tabela 5 – Descrição dos atributos da base de dados.**

<b>Atributo</b>	<b>Descrição</b>	<b>Tipo</b>
hardness	Label de dificuldade.	<i>int64</i>
name	Nome da música.	<i>object</i>
artist	Nome do artista.	<i>object</i>
genre	Gênero musical do artista.	<i>object</i>
tone	Tom da música.	<i>object</i>
ratio	Razão entre a quantidade de acordes e o tamanho da letra.	<i>float64</i>
transitions	Número de diferentes transições entre acordes.	<i>int64</i>
accidents	Número de acordes com sustenido (#) ou bemol (b) na cifra.	<i>int64</i>
chords	Lista com os acordes que compõem a cifra.	<i>object</i>
topChord	Acorde com maior frequência na cifra.	<i>object</i>
chordPop	Frequência absoluta do topChord na base de dados.	<i>int64</i>
chordFreq	Frequência do topChord na classe da cifra.	<i>int64</i>
chordCount	Número de diferentes acordes na cifra.	<i>float64</i>

Fonte: Elaborada pelo autor.

Seguindo o modelo de Carvalho e Pelli (2017), a Figura 4 apresenta como aconteceu o fluxo de processos do experimento. São apresentadas todas as etapas, a partir da escolha de características até a avaliação dos resultados.

**Figura 4 – Representação da metodologia do experimento**



Fonte: Adaptação Carvalho e Pelli (2017)

O atributo *chordCount* foi convertido para escala logarítmica, com objetivo de aumentar a relevância do atributo.

A Tabela 6 apresenta uma análise dos principais indicadores descritivos das features numéricas em cada uma das classes, depois da extração e transformação de características.

**Tabela 6 – Análise descritiva dos dados de cifras.**

	Atributo	N	Min	Max	Média	Desvio Padrão
Iniciante (1)	ratio	60	0.12	0.76	0.34	0.13
	transitions	60	2.00	13.00	7.23	2.63
	accidents	60	0.00	0.00	0.00	0.00
	chordPop	60	4.00	40.00	26.00	10.52
	chordFreq	60	1.00	18.00	12.57	5.46
	chordCount	60	0.69	1.61	1.34	0.25
Fácil (2)	ratio	60	0.11	0.94	0.38	0.13
	transitions	60	4.00	17.00	9.60	3.14
	accidents	60	0.00	0.00	0.00	0.00
	chordPop	60	2.00	40.00	22.93	14.15
	chordFreq	60	1.00	20.00	10.10	7.23
	chordCount	60	1.10	2.20	1.58	0.23
Intermediário (3)	ratio	60	0.14	1.28	0.53	0.24
	transitions	60	6.00	58.00	24.88	12.02
	accidents	60	0.00	140.00	26.51	26.40
	chordPop	60	1.00	40.00	11.33	12.63
	chordFreq	60	1.00	7.00	3.26	2.17
	chordCount	60	1.61	3.37	2.40	0.41
Avançado (4)	ratio	60	0.22	1.44	0.65	0.22
	transitions	60	24.00	173.00	65.26	24.21
	accidents	60	6.00	173.00	43.90	25.06
	chordPop	60	1.00	27.00	2.50	3.80
	chordFreq	60	1.00	3.00	1.66	0.79
	chordCount	60	2.83	4.67	3.70	0.33

Fonte: Elaborada pelo autor.

A seleção de características é baseada no coeficiente de correlação de Pearson, calculado através do método ‘corr’ da biblioteca de modelagem de dados *pandas*<sup>7</sup>.

Para avaliar a assertividade global da classificação, são realizadas 40 repetições da etapa de classificação. A cada iteração a base de dados é randomizada, de maneira que a cada iteração o classificador possa trabalhar com dados diferentes tanto para o treino quanto para testes, além de ter sempre estes dados em nova disposição de ordem. Depois da classificação, ainda em cada iteração, são coletados os índices de acurácia, precisão e sensibilidade, bem como a matriz de confusão.

A quantidade de iterações foi definida após uma série de testes. O experimento foi executado 240 vezes, sempre com o incremento de uma iteração, e a coleta da media de acu-

<sup>7</sup> pandas.pydata.org

rácia e desvio padrão do indicador. Após a análise dos indicadores, foi percebido que após 40 iterações os mesmos pararam de sofrer variações significativas. Assim, o número de iterações foi definido como 40, por ser o menor número de iterações possíveis para a repetição do experimento com a obtenção de resultados esperados.

Ao final das iterações, a eficácia da classificação foi mensurada através da média e desvio padrão dos indicadores (acurácia, precisão e recall) coletados durante as 40 repetições predefinidas.

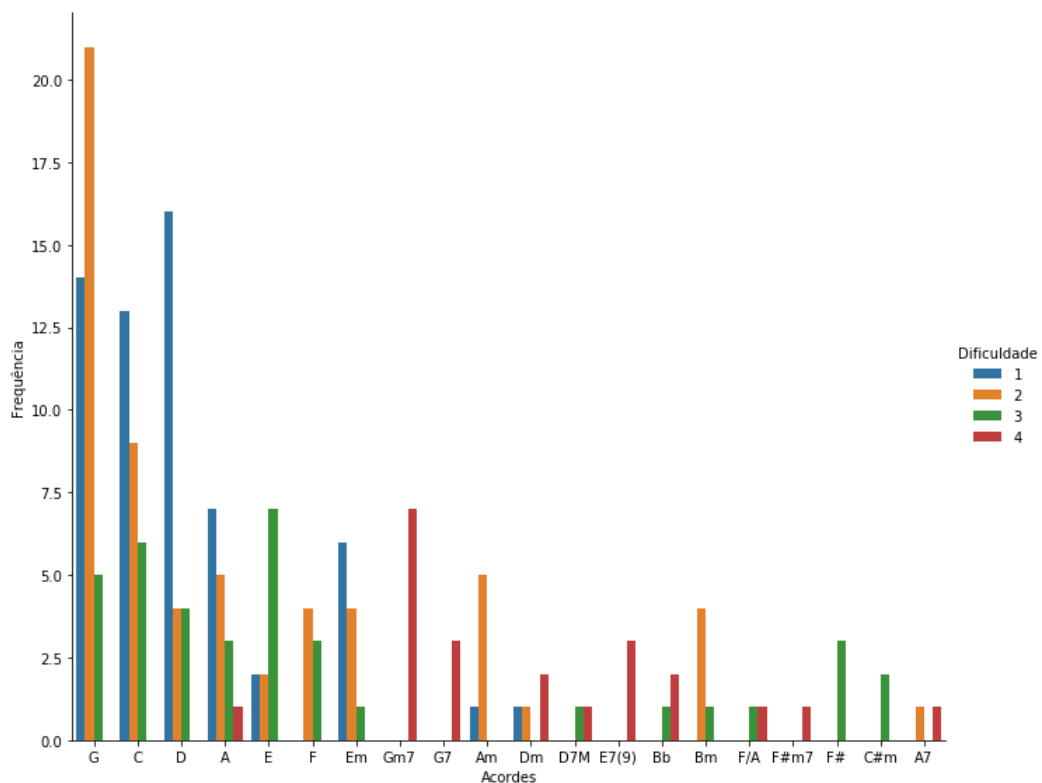
## 5 RESULTADOS E DISCUSSÃO

Os resultados estão apresentados na cronologia de decorrências da aplicação de cada etapa da metodologia do reconhecimento de padrões. As considerações são apresentadas no final, tratando do resultado geral.

O processo de extração de características resultou em sete novas features que tem por objetivo auxiliar na representação do conteúdo da base de dados. Todas estão listadas na descrição da Tabela 5.

O novo conjunto de features permitiram a obtenção de insights fundamentais a respeito do comportamento dos dados. A Figura 5 torna evidente quais os 20 acordes mais populares da base, e o quantitativo de vezes em que os mesmos aparecem como sendo o acorde mais frequente em alguma cifra. Esse resultado considera a separação de classes, o que permite visualizar na classe 1 uma concentração notável de acordes muito frequentes na base, enquanto que na classe 4 estes aparecem poucas vezes, e com frequência muito baixa.

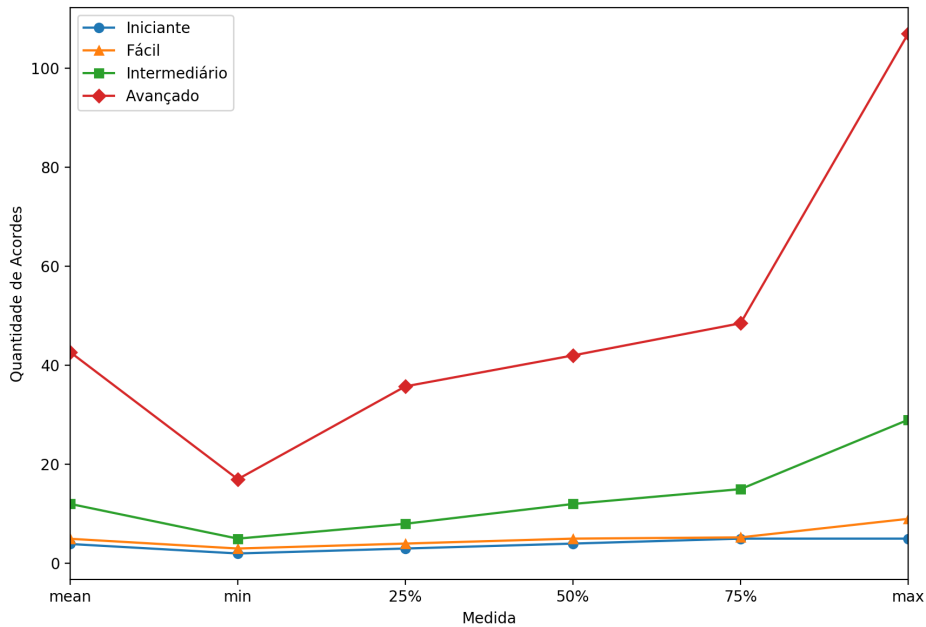
**Figura 5 – Distribuição por classe da ocorrência dos acordes mais frequentes da base.**



Fonte: Elaborada pelo autor.

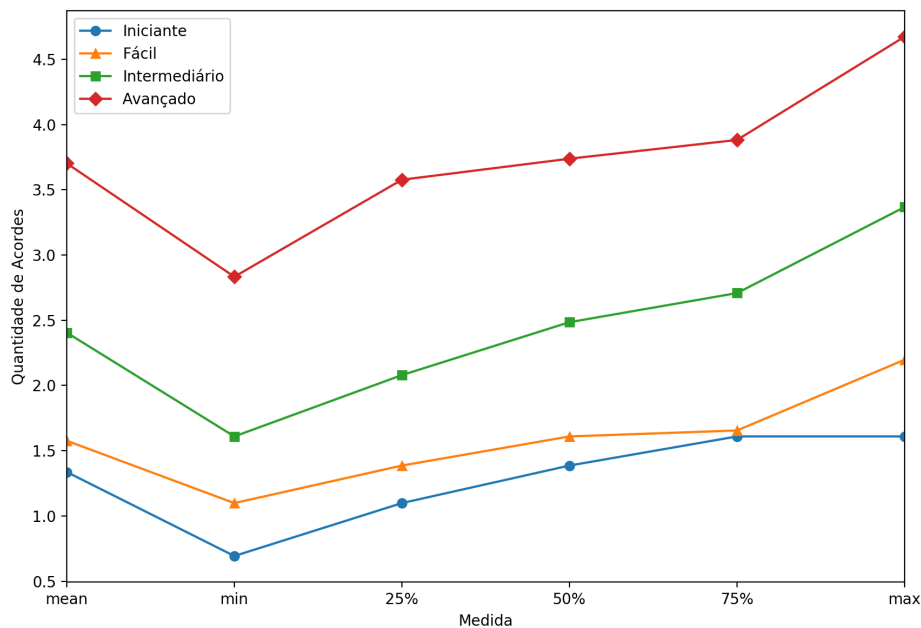
A transformação dos dados do atributo chordCount para escala logarítmica trouxe maior relevância para a feature, sendo possível perceber através das Figuras 6 e 7 - representação da distribuição dos dados antes e depois da transformação, respectivamente - que os dados estão contidos em intervalos visivelmente melhor separáveis, empregando a definição de maior relevância descrita por Félix (1998).

**Figura 6 – Distribuição da quantidade de acordes por classe; escala natural.**



Fonte: Elaborada pelo autor.

**Figura 7 – Distribuição da quantidade de acordes por classe; escala logarítmica.**



Fonte: Elaborada pelo autor.

A seleção de características foi baseada nos resultados de correlação entre as todas variáveis numéricas da base de dados e a característica de classe do problema, (*hardness*), que estão dispostos na Tabela 7.

É possível observar que a característica *chordCount* tem a maior relevância, com uma correlação positiva de 0,90. Este indicador aponta que à medida que a dificuldade das cifras

evolui (de 1 a 4), o número de acordes únicos aumenta. *chordFreq* é a característica de menor índice de correlação, o que aponta com uma menor intensidade que à medida que a dificuldade das cifras cresce, o número de vezes em que os acordes mais frequentes das mesmas aparece dentro da classe diminui.

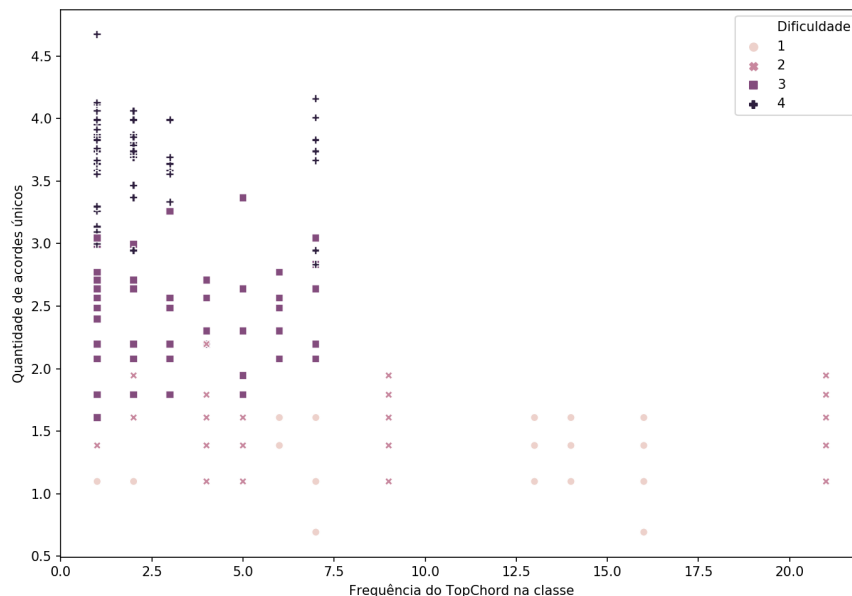
**Tabela 7 – Coeficientes de correlação de Pearson com a característica *hardness*.**

Nome	Pearson
chordCount	0.91
transitions	0.79
accidents	0.68
ratio	0.55
chordPop	-0.63
chordFreq	-0.64

Fonte: Elaborada pelo autor.

A Figura 8 apresenta um gráfico com a distribuição das características com os índices extremos de correlação com a classe, e é possível perceber a concentração dos grupos de instâncias das classes. Sugerindo que as características representam bem o conjunto de dados, em relação à classe.

**Figura 8 – Distribuição estratificada das features *chordCount* e *chordFreq*.**



Fonte: Elaborada pelo autor.

A classificação considerou 72 registros para teste. Os dados foram estratificados em relação à classe durante a divisão das bases de treino e teste. A partir das matrizes de confusão resultadas da SVM, foi possível calcular os índices de acurácia, precisão e recall de cada classe, e apresentar nas Tabelas 8, 9, 10, 11.

Tabela 8 – Resultados de classificação para a classe 1 (Iniciante).

Teste	VP	FN	FP	VN	Acurácia(%)	Precisão(%)	Recall(%)
1	18	0	0	49	100.0	100.0	100.0
2	18	0	0	48	100.0	100.0	100.0
3	18	0	0	46	100.0	100.0	100.0
4	18	0	0	51	100.0	100.0	100.0
5	17	1	0	47	98.46	100.0	94.44
6	18	0	0	47	100.0	100.0	100.0
7	18	0	0	48	100.0	100.0	100.0
8	18	0	0	47	100.0	100.0	100.0
9	18	0	0	48	100.0	100.0	100.0
10	16	2	0	50	97.06	100.0	88.89
11	18	0	0	46	100.0	100.0	100.0
12	18	0	0	49	100.0	100.0	100.0
13	18	0	0	49	100.0	100.0	100.0
14	18	0	0	50	100.0	100.0	100.0
15	18	0	0	45	100.0	100.0	100.0
16	18	0	0	43	100.0	100.0	100.0
17	18	0	0	48	100.0	100.0	100.0
18	18	0	0	47	100.0	100.0	100.0
19	18	0	0	50	100.0	100.0	100.0
20	18	0	0	48	100.0	100.0	100.0
21	18	0	0	47	100.0	100.0	100.0
22	18	0	0	52	100.0	100.0	100.0
23	18	0	0	47	100.0	100.0	100.0
24	18	0	0	48	100.0	100.0	100.0
25	17	1	0	47	98.46	100.0	94.44
26	18	0	0	51	100.0	100.0	100.0
27	18	0	0	46	100.0	100.0	100.0
28	16	2	0	49	97.01	100.0	88.89
29	18	0	0	52	100.0	100.0	100.0
30	18	0	0	47	100.0	100.0	100.0
31	18	0	0	50	100.0	100.0	100.0
32	18	0	0	49	100.0	100.0	100.0
33	18	0	0	49	100.0	100.0	100.0
34	18	0	0	49	100.0	100.0	100.0
35	16	2	0	48	96.97	100.0	88.89
36	18	0	0	48	100.0	100.0	100.0
37	18	0	0	45	100.0	100.0	100.0
38	18	0	0	50	100.0	100.0	100.0
39	18	0	0	49	100.0	100.0	100.0
40	18	0	0	50	100.0	100.0	100.0
<b>Média</b>	17.8	0.2	0	48.23	99.7	100.0	98.89
<b>Desvio Padrão</b>	0.56	0.56	0.0	1.91	0.85	0.0	3.13

Fonte: Elaborada pelo autor.



Tabela 9 – Resultados de classificação para a classe 2 (Fácil).

Teste	VP	FN	FP	VN	Acurácia(%)	Precisão(%)	Recall(%)
1	17	1	1	50	97.1	94.44	94.44
2	17	1	1	49	97.06	94.44	94.44
3	18	0	4	46	94.12	81.82	100.0
4	18	0	1	51	98.57	94.74	100.0
5	15	3	0	49	95.52	100.0	83.33
6	17	1	2	48	95.59	89.47	94.44
7	17	1	0	49	98.51	100.0	94.44
8	18	0	0	47	100.0	100.0	100.0
9	15	3	1	51	94.29	93.75	83.33
10	16	2	2	50	94.29	88.89	88.89
11	16	2	1	48	95.52	94.12	88.89
12	15	3	0	52	95.71	100.0	83.33
13	16	2	0	51	97.1	100.0	88.89
14	16	2	1	52	95.77	94.12	88.89
15	16	2	3	47	92.65	84.21	88.89
16	17	1	4	44	92.42	80.95	94.44
17	17	1	2	49	95.65	89.47	94.44
18	16	2	2	49	94.2	88.89	88.89
19	17	1	2	51	95.77	89.47	94.44
20	18	0	2	48	97.06	90.0	100.0
21	16	2	1	49	95.59	94.12	88.89
22	18	0	0	52	100.0	100.0	100.0
23	18	0	2	47	97.01	90.0	100.0
24	18	0	3	48	95.65	85.71	100.0
25	17	1	3	47	94.12	85.0	94.44
26	17	1	1	52	97.18	94.44	94.44
27	15	3	1	49	94.12	93.75	83.33
28	17	1	3	48	94.2	85.0	94.44
29	18	0	1	52	98.59	94.74	100.0
30	15	3	1	50	94.2	93.75	83.33
31	16	2	0	52	97.14	100.0	88.89
32	17	1	2	50	95.71	89.47	94.44
33	16	2	1	51	95.71	94.12	88.89
34	18	0	2	49	97.1	90.0	100.0
35	16	2	2	48	94.12	88.89	88.89
36	16	2	2	50	94.29	88.89	88.89
37	17	1	4	46	92.65	80.95	94.44
38	17	1	2	51	95.77	89.47	94.44
39	18	0	2	49	97.1	90.0	100.0
40	17	1	1	51	97.14	94.44	94.44
<b>Média</b>	16.73	1.27	1.57	49.3	95.86	91.79	92.92
<b>Desvio Padrão</b>	0.99	0.99	1.13	1.95	1.84	5.42	5.48

Fonte: Elaborada pelo autor.

Tabela 10 – Resultados de classificação para a classe 3 (Intermediário).

Teste	VP	FN	FP	VN	Acurácia(%)	Precisão(%)	Recall(%)
1	16	2	3	51	93.06	84.21	88.89
2	16	2	4	50	91.67	80.0	88.89
3	11	7	1	53	88.89	91.67	61.11
4	16	2	1	53	95.83	94.12	88.89
5	17	1	7	47	88.89	70.83	94.44
6	15	3	4	50	90.28	78.95	83.33
7	13	5	1	53	91.67	92.86	72.22
8	12	6	1	53	90.28	92.31	66.67
9	16	2	4	50	91.67	80.0	88.89
10	17	1	3	49	94.29	85.0	94.44
11	14	4	4	50	88.89	77.78	77.78
12	17	1	4	50	93.06	80.95	94.44
13	16	2	3	51	93.06	84.21	88.89
14	17	1	3	51	94.44	85.0	94.44
15	14	4	5	49	87.5	73.68	77.78
16	12	6	5	49	84.72	70.59	66.67
17	14	4	2	52	91.67	87.5	77.78
18	14	4	3	51	90.28	82.35	77.78
19	15	3	1	53	94.44	93.75	83.33
20	15	3	3	51	91.67	83.33	83.33
21	16	2	5	49	90.28	76.19	88.89
22	17	1	1	53	97.22	94.44	94.44
23	13	5	2	52	90.28	86.67	72.22
24	13	5	1	53	91.67	92.86	72.22
25	13	5	3	51	88.89	81.25	72.22
26	16	2	1	53	95.83	94.12	88.89
27	14	4	4	50	88.89	77.78	77.78
28	16	2	3	49	92.86	84.21	88.89
29	16	2	0	54	97.22	100.0	88.89
30	15	3	4	50	90.28	78.95	83.33
31	18	0	4	50	94.44	81.82	100.0
32	16	2	3	51	93.06	84.21	88.89
33	16	2	3	51	93.06	84.21	88.89
34	15	3	2	52	93.06	88.24	83.33
35	15	3	3	49	91.43	83.33	83.33
36	15	3	3	51	91.67	83.33	83.33
37	10	8	1	53	87.5	90.91	55.56
38	16	2	2	52	94.44	88.89	88.89
39	14	4	1	53	93.06	93.33	77.78
40	15	3	1	53	94.44	93.75	83.33
<b>Média</b>	14.9	3.1	2.73	51.12	91.89	85.19	82.78
<b>Desvio Padrão</b>	1.77	1.77	1.52	1.64	2.72	6.97	9.81

Fonte: Elaborada pelo autor.

Tabela 11 – Resultados de classificação para a classe 4 (Avançado).

Teste	VP	FN	FP	VN	Acurácia(%)	Precisão(%)	Recall(%)
1	16	2	1	51	95.71	94.12	88.89
2	15	3	1	51	94.29	93.75	83.33
3	17	1	3	47	94.12	85.0	94.44
4	17	1	1	52	97.18	94.44	94.44
5	15	3	1	49	94.12	93.75	83.33
6	15	3	1	50	94.2	93.75	83.33
7	18	0	5	48	92.96	78.26	100.0
8	17	1	6	48	90.28	73.91	94.44
9	17	1	1	49	97.06	94.44	94.44
10	17	1	1	49	97.06	94.44	94.44
11	16	2	3	48	92.75	84.21	88.89
12	17	1	1	50	97.1	94.44	94.44
13	17	1	2	50	95.71	89.47	94.44
14	17	1	0	51	98.55	100.0	94.44
15	15	3	1	48	94.03	93.75	83.33
16	14	4	2	47	91.04	87.5	77.78
17	17	1	2	49	95.65	89.47	94.44
18	17	1	2	48	95.59	89.47	94.44
19	18	0	1	50	98.55	94.74	100.0
20	15	3	1	51	94.29	93.75	83.33
21	15	3	1	50	94.2	93.75	83.33
22	17	1	1	53	97.22	94.44	94.44
23	16	2	3	49	92.86	84.21	88.89
24	17	1	2	49	95.65	89.47	94.44
25	17	1	2	47	95.52	89.47	94.44
26	18	0	1	51	98.57	94.74	100.0
27	17	1	3	47	94.12	85.0	94.44
28	16	2	1	49	95.59	94.12	88.89
29	18	0	1	52	98.59	94.74	100.0
30	17	1	2	48	95.59	89.47	94.44
31	16	2	0	52	97.14	100.0	88.89
32	16	2	0	51	97.1	100.0	88.89
33	17	1	1	50	97.1	94.44	94.44
34	16	2	1	51	95.71	94.12	88.89
35	17	1	3	47	94.12	85.0	94.44
36	17	1	1	49	97.06	94.44	94.44
37	18	0	4	45	94.03	81.82	100.0
38	17	1	0	51	98.55	100.0	94.44
39	17	1	2	50	95.71	89.47	94.44
40	18	0	2	50	97.14	90.0	100.0
<b>Média</b>	16.6	1.4	1.68	49.42	95.55	91.29	92.22
<b>Desvio Padrão</b>	1.01	1.01	1.29	1.74	2.02	5.75	5.6

Fonte: Elaborada pelo autor.

Até o presente momento, não foram identificados na literatura trabalhos que apliquem a metodologia de reconhecimento de padrões na identificação de características que sejam relevantes pra definir a dificuldade de cifras musicais. Diversos trabalhos vem implementando SVM para tratamento de problemas das mais variadas naturezas, e estes trabalhos são considerados na discussão dos parágrafos que seguem.

Através da Tabela 12 são apresentados os resultados de acurácia, precisão e recall médios da classificação, considerando uma avaliação global do experimento.

**Tabela 12 – Resultados estratificados de classificação.**

	Acurácia (%)		Precisão (%)		Recall (%)	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
Iniciante	99.7	0.85	100.0	0.0	98.89	3.13
Fácil	95.86	1.84	91.79	5.42	92.92	5.48
Intermediário	91.89	2.72	85.19	6.97	82.78	9.81
Avançado	95.55	2.02	91.29	5.75	92.22	5.6
<b>Global</b>	<b>95.75</b>	<b>1.86</b>	<b>92.07</b>	<b>4.53</b>	<b>91.70</b>	<b>6.0</b>

Fonte: Elaborada pelo autor.

É observada que a classe iniciante tem os melhores indicadores de avaliação, apresentando uma acurácia média de 99,7%, ou seja, o classificador identifica com acerto 99,7% das instancias da classe iniciante submetidas para teste.

Em contraste, a classe intermediário se mostrou a mais difícil em ser abstraída pelo classificador. A acurácia ficou em 91,89%±2,72%. Esse é um impacto direto da extração e seleção de características nos resultados. (BIANCHI, 2006)

Todavia, esse é considerado um percentual de acerto muito bom em trabalhos de classificação nas mais diversas áreas, como em (JESUS et al., 2018), (MEGETO et al., 2014), (NASCIMENTO et al., 2009), (PIMENTEL et al., 2008), (ANDRADE; FRANCISCO; ALMEIDA, 2013).

O percentual de acurácia global obtido foi de 95.75% , sob um desvio padrão de 1,86%. A precisão e recall se mantiveram aceitáveis, alcançando a média de 92,07%±4,53% e 91,70%±6,0% respectivamente.

## 6 CONCLUSÕES

Este trabalho apresenta um estudo sobre a aplicação da abordagem de reconhecimento de padrões em uma base de dados composta por cifras de violão, descrevendo a realização de um experimento aplicando as técnicas de coleta de dados, extração e seleção de características, classificação supervisionada e avaliação de resultados.

Como resultado do experimento foi obtida a acurácia média de cada classe, apontando o sucesso na abstração das características dos dados da classe iniciante, e a notável dificuldade em entender a classe intermediário, que resultou na menores indicadores de acurácia e precisão na etapa de predição. Apesar de tal, a taxa global de acurácia, de  $95,75\% \pm 1,86\%$  (média), foi considerada um ótimo resultado para a classificação.

Conclui-se neste trabalho que a abordagem de reconhecimento de padrões é válida, e eficiente, para a aplicação no problema de classificação de níveis de dificuldade de cifras de violão.

### 6.1 Trabalhos Futuros

A extração de características poderia ser revisada, afim de que novas features pudessem ser extraídas e avaliadas para validação e possível incremento na eficiência da abordagem aqui descrita.

Testes com aumento na quantidade de registros das bases de treino e teste se faz interessante para avaliar a escalabilidade da solução, bem como a viabilidade de aplicação em um cenário de produção.

Novas soluções aplicando técnicas e implementações baseadas em *deep learning* poderiam ser comparadas à abordagem tratada no presente trabalho, com o propósito de propor melhorias para os resultados.

Um trabalho mais aprofundado na área musical se mostra relevante para o aprimoramento das definições de níveis de dificuldades tratados neste trabalho, tornando-as próximas da realidade dos estudantes de violão.



## REFERÊNCIAS

- ALEXANDRE, J. et al. Desvendando os Mistérios do Coeficiente de Correlação. p. 66–95, 2014. ISSN 0104-7094. Citado na página 25.
- AMATO, R. d. C. F. A situação da música na educação básica sob a ótica dos educadores musicais. In: *XVII Congresso Nacional da Associação Brasileira de Educação Musical-Abem, Anais, São Paulo*. [S.l.: s.n.], 2008. Citado na página 17.
- ANDRADE, A. C.; FRANCISCO, C. N.; ALMEIDA, C. M. d. DESEMPENHO DE CLASSIFICADORES PARAMÉTRICO E NÃO PARAMÉTRICO NA CLASSIFICAÇÃO DA FISIONOMIA VEGETAL Evaluating the Performance of Parametric and Non-parametric Classifiers for Identifying Vegetal Physiognomies Universidade Federal Fluminense UFF INTROD. *Revista Brasileira de Cartografia*, v. 65, n. 2, p. 227–241, 2013. Citado na página 42.
- BATISTA, G. E. A. P. A. Pré-processamento de dados em aprendizado de máquina supervisionado. 2003. Citado na página 31.
- BIANCHI, M. F. d. *Extração de características de imagens de faces humanas através de wavelets, PCA e IMPCA*. Tese (Doutorado) — Universidade de São Paulo, São Carlos, 4 2006. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/18/18133/tde-10072006-002119/>>. Citado na página 42.
- BISHOP, C. *Neural networks for pattern recognition*. [S.l.: s.n.], 1996. ISBN 978-0198538646. Citado na página 23.
- BRAZIL, M. A. Leitura musical para iniciantes em aulas coletivas de violão: uma visão através da teoria da autoeficácia. p. 288, 2017. Citado na página 31.
- CABRAL, G. et al. Da Cifra Para o Braço: Estudo dos Problemas de Execução Musical em Violão e Guitarra. *Proceedings of VIII Brazilian Symposium on Computer Music. Fortaleza*, p. 2–8, 2001. Citado na página 31.
- CARVALHO, H. C. F. B.; PELLI, E. Técnicas de reconhecimento de padrões para identificação de ataque de DNS. *Revista Brasileira de Computação Aplicada*, v. 9, n. 2, p. 99–110, 2017. Citado na página 32.
- CASTELLUCCI, P. B. JULIA E JuMP: NOVAS FERRAMENTAS PARA PROGRAMAÇÃO MATEMÁTICA. *Pesquisa Operacional para o Desenvolvimento*, v. 9, n. 2, p. 48–61, 2017. Citado na página 28.
- CESCHIN, F. J. d. O.; BURDA, R. L. Classificação de Postagens de Grupos do Facebook. 2015. Disponível em: <<https://web.inf.ufpr.br/menotti/ci171-2015-2-1/files/final-FabricioRenan-artigo.pdf>>. Citado 2 vezes nas páginas 25 e 26.
- CHEDIAK, A. *Harmonia e Improvisação. 70 Músicas Harmonizadas e Analisadas*. [S.l.]: Lumiar Editora, 1986. 296 p. ISBN 8574072648. Citado na página 20.
- CifraClub. *O maior site de ensino de música do Brasil*. 2019. Disponível em: <<https://www.cifraclub.com.br>>. Citado na página 20.
- COUTO, A. C. Música popular e aprendizagem. p. 89–104, 2009. Citado na página 18.

DANTAS, L. M. T. D. S. PENSAR E AGIR: O PERCURSO CONSTRUTIVO DE UM MÉTODO PARA O ENSINO DE VIOLÃO POPULAR A PARTIR DA VIVÊNCIA DOCENTE. v. 9, p. 7–9, 2018. ISSN 09135685. Citado na página 31.

DOWNEY, A. *Think Python: How to Think Like a Computer Scientist*. [S.l.: s.n.], 2015. ISSN 01635948. ISBN 9780521898119. Citado na página 29.

DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern classification. *New York: John Wiley, Section*, 2001. Citado na página 23.

FÉLIX, L. C. M. Data Mining no Processo de Extração de Conhecimento de Bases de Dados. p. 109, 1998. Citado 2 vezes nas páginas 24 e 35.

FERNANDES, C. A. F. S. Algoritmo do Tipo Filter-Wrapper de Seleção de Features para Utilização na Seleção de Genes. *Dissertação de Mestrado Integrado em Engenharia Electrotécnica e de Computadores.*, p. 4–7, 2017. Citado na página 24.

GREEN, L. *How Popular Musicians Learn*. [S.l.: s.n.], 2017. Citado na página 17.

HAN, M. K. J.; PEI, J. Data Mining: Concepts and Techniques, Third Edition - Books24x7. *Morgan Kaufmann Publishers*, p. 745, 2012. ISSN 14337851. Citado na página 25.

JAGADISH, H. V. et al. Big data and its technical challenges. *Communications of the ACM*, 2014. ISSN 00010782. Citado na página 21.

JESUS, L. d. et al. Aplicação de técnicas de reconhecimento de padrões para estimar o comportamento de bovinos em função de dados de posicionamento GPS Leandro. *Anais 7º Simpósio de Geotecnologias no Pantanal, Jardim, MS, Embrapa Informática Agropecuária/INPE*, p. 41-50, p. 40–50, 2018. Citado na página 42.

JOSÉ, R.; ALMEIDA, D. A.; PAIVA, L. C. D. Investigação do processo de análises interativas em Ciência de Dados centradas na web com Jupyter Notebook. p. 53–54, 2017. Citado na página 29.

JÚNIOR, D. M. CLASSIFICAÇÃO DE EDITAIS LICITATÓRIOS EM ÁREAS DE ATUAÇÃO BASEADO EM APRENDIZADO SUPERVISIONADO. n. 4, 2018. ISSN 1413-8123. Citado na página 25.

JUNIOR, R. C. et al. A utilização de tecnologias da informação e comunicação para o ensino do violão. *Revista Intersaberes vol.12 nº25*, v. 12, p. 68–83, 2017. Citado 2 vezes nas páginas 18 e 20.

KASABOV, N. K. Learning fuzzy rules and approximate reasoning in fuzzy neural networks and hybrid systems. *Fuzzy Sets and Systems*, 1996. ISSN 01650114. Citado na página 22.

LORENA, A. C.; CARVALHO, A. C. P. L. F. d. Uma Introdução às Support Vector Machines. *Revista de Informática Teórica e Aplicada*, v. 14, n. 2, p. 43–67, 2007. Citado na página 23.

LUCCA, G. et al. Uma implementação do algoritmo Naïve Bayes para classificação de texto. *Centro de Ciências Computacionais - Universidade Federal do Rio Grande (FURG) Rio Grande - RS - Brasil*, p. 1–4, 2013. Citado na página 25.

MANNILA, H. Methods and problems in data mining. In: . [S.l.: s.n.], 2012. Citado na página 24.



MARQUES, D. L. et al. Atraindo Alunos do Ensino Médio para a Computação: Uma Experiência Prática de Introdução a Programação utilizando Jogos e Python. In: *Anais do XXII SBIE - XVII WIE*. [S.l.: s.n.], 2011. ISBN 2316-6541. ISSN 2316-6541. Citado na página 29.

MARQUES, L.; REIS, B.; CAMPOS, L. F. D. A. Detecção de Lesões de Câncer de Pele Utilizando Análise de Componentes Independentes e Análise Discriminante Linear. 2017. Citado na página 23.

MCPHERSON, G. E. *The Child as Musician: A Handbook of Musical Development*. [S.l.: s.n.], 2012. ISSN 1744-7607. ISBN 9780191689765. Citado na página 17.

MEGETO, G. A. S. et al. Modelagem Da Temperatura E Da Precipitação Em Ocorrências Da Ferrugem Asiática Da Soja Por Meio Da Técnica De Árvore De Decisão. *Engenharia Agrícola*, v. 34, n. 3, p. 590–599, 2014. Citado na página 42.

MICHELON, G. K. et al. Máquina De Vetores De Suporte Para Estimar a Produtividade Da Soja. *Revista Engenharia Na Agricultura - Reveng*, v. 25, n. 3, p. 240–248, 2018. ISSN 1414-3984. Citado na página 24.

MORAES, E. C. C. Método Não Supervisionado Baseado Em Curvas Principais Para Reconhecimento De Padrões. 2016. Citado 2 vezes nas páginas 22 e 23.

NASCIMENTO, R. F. F. et al. O algoritmo SVM: avaliação da separação ótima de classes em imagens CCD-CBERS-2. *XIV Simpósio Brasileiro de Sensoriamento Remoto*, p. 2079–2086, 2009. Citado na página 42.

NAVIGATOR, A. *Anaconda The Most Popular Python Data Science Platform*. 2019. Disponível em: <<https://docs.anaconda.com/>>. Citado na página 28.

OLIVEIRA, D. G. d. et al. CORRELAÇÃO ESPACIAL E LINEAR DE ATRIBUTOS FÍSICOS DO SOLO EM SISTEMA DE INTEGRAÇÃO LAVOURA PECUÁRIA. p. 69–77, 2017. Citado na página 24.

PASSAMAE, M. A. d. R. V. Radamés Gnattali, a Era Vargas, o Rádio e a Construção da Identidade Nacional. In: . Uberlândia: CONGRESSO DA ASSOCIAÇÃO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM MÚSICA, 2011. Citado na página 21.

PEDREGOSA, F. et al. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado 2 vezes nas páginas 25 e 26.

PIMENTEL, L. D. et al. Seleção precoce de maracujazeiro pelo uso da correlação entre dados de produção mensal e anual. *Pesquisa Agropecuaria Brasileira*, v. 43, n. 10, p. 1303–1309, 2008. ISSN 0100204X. Citado na página 42.

RABELO, H. et al. Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de EaD em ambientes virtuais de aprendizagem. *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*, v. 1, n. Cbie, p. 1527, 2017. Citado na página 26.

SANTOS, J. C. et al. Seleção de atributos usando algoritmos genéticos para classificação de regiões. *Xiii Sbsr*, n. June 2016, p. 6143–6150, 2007. Disponível em: <<http://marte.dpi.inpe.br/rep-/dpi.inpe.br/sbsr@80/2006/11.15.12.38>>. Citado na página 23.

- SILVA, J. P. S. Recursos Digitais de Apoio ao Ensino das Funções Trigonométricas. *Universidade de Aveiro*, 2014. Citado na página 29.
- SILVA, R. R. V. Visão computacional em python utilizando as bibliotecas scikit-image e scikit-learn. p. 407–428, 2017. Citado na página 29.
- SILVA, W. K. N. D.; SANTOS, A. D. M. Estratégias de Construções de Comitês de Classificadores Multirrótulos no Aprendizado Semissupervisionado Multidescrição. *Revista de Informática Teórica e Aplicada*, v. 24, n. 2, p. 71, 2017. ISSN 01034308. Citado na página 22.
- TABORDA, M. *Violão e Identidade Nacional*. [S.l.]: Editora José Olympio, 2011. Citado 2 vezes nas páginas 20 e 21.
- THEODORIDIS, S. et al. *Pattern Recognition, Second Edition*. [S.l.: s.n.], 2006. ISSN 10538801. ISBN 0123695317. Citado na página 22.
- VANZELA, A.; OLIVEIRA, L. C. d.; CARVALHO, M. A. d. A psicopedagogia e o ensino musical. *Revista Psicologia, Diversidade e Saúde*, v. 6, n. 2, p. 94, 2017. ISSN 2317-3394. Citado na página 17.
- VIEIRA, S. T.; VALADAO, E. Otimização do posicionamento de pontos de acesso wireless. 2017. Citado na página 28.
- VINISKI, A. D.; GUIMARÃES, A. M. Técnicas de seleção de atributos para mineração de dados de alta dimensionalidade gerados por espectroscopia no infravermelho próximo–NIR. *Anais SULCOMP*, v. 8, n. Md, 2017. Disponível em: <<http://periodicos.unesc.net/sulcomp/article/view/3142>>. Citado na página 24.
- WIGFIELD, A.; ECCLES, J. S. Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, v. 25, n. 1, p. 68–81, 2000. ISSN 0361476X. Citado na página 17.
- XAVIER, K.; SOUZA, S. D. Montagem de ambiente para classificação de solos usando ScikitLearn classification using ScikitLearn. *Resumos Expandidos XIV Mostra de Estagiários e Bolsistas*, p. 104–110, 2018. Citado na página 29.

## APÊNDICE

## APÊNDICE A - A Base de Dados

**Tabela 13 – Cifras consideradas no experimento.**

<b>Dificuldade</b>	<b>Artista</b>	<b>Título</b>
Intermediário	IKilo	Deixe-me Ir
Avançado	A-Ha	Start The Simulator
Avançado	Agnaldo Timóteo	Cadeira Vazia
Intermediário	Aline Barros	Cheio de Alegria
Intermediário	Aline Barros	Caminho de Milagres
Iniciante	Aline Barros	Arca de Noé
Intermediário	Aline Barros	Renova-me Senhor
Intermediário	Altamar Dutra	Não Se Esqueça de Mim
Iniciante	Ana Vilela	Bullet Train
Intermediário	Anderson Freire	Imensidão
Intermediário	Arctic Monkeys	When The Sun Goes Down
Iniciante	Armandinho	A Ilha
Avançado	Belchior	Chega de Saudade
Fácil	Beyoncé	Amor Gitano
Iniciante	blink-182	Sober
Intermediário	Bruninho & Davi	Dormi Na Praça / É Tarde Demais (pot-pourri)
Iniciante	Bruno e Marrone	Furacão
Iniciante	Bruno e Marrone	"Quem Tá Bebo. Quem Tá Bão"
Fácil	Bruno e Marrone	Foi Só Um Susto
Fácil	Bruno e Marrone	Aline
Iniciante	Bruno e Marrone	Coisa de Pele
Intermediário	Bruno e Marrone	Já Não Sei Mais Nada
Avançado	Caetano Veloso	Sampa
Avançado	Caetano Veloso	Noites do Norte
Avançado	Caetano Veloso	Chega de Saudade
Iniciante	Capital Inicial	Terceiro Mundo Digital
Iniciante	Capital Inicial	Construção Civil
Iniciante	Capital Inicial	O Passageiro
Iniciante	Capital Inicial	Água e Vinho
Intermediário	Cássia Eller	O Segundo Sol
Fácil	Cássia Eller	Partners
Fácil	Cássia Eller	"Um Branco. Um Xis. Um Zero"
Fácil	Católicas	"Aleluia. Glória Ao Senhor"
Fácil	Católicas	Até Quando
Fácil	Católicas	Anuncia-me
Fácil	Católicas	A Começar Em Mim
Fácil	Católicas	A Minha Paz Eu Vos Dou
Intermediário	Charlie Brown Jr.	Só Os Loucos Sabem
Iniciante	Chiclete Com Banana	Chiclete na Boca
Avançado	Chico Buarque	"Meninos. Eu Vi"
Avançado	Chico Buarque	Aquela Mulher
Avançado	Chico Buarque	Choro Bandido

Avançado	Chico Buarque	Bolero Blues
Avançado	Chico Buarque	A Noiva da Cidade
Avançado	Chico Buarque	A Mais Bonita
Avançado	Chico Buarque	João e Maria
Avançado	Chico Buarque	Chão de Esmeraldas
Avançado	Chico Buarque	Amor Barato
Intermediário	Chitãozinho & Xororó	Será Que Eu Sou
Fácil	Chitãozinho & Xororó	Aguenta Coração
Intermediário	Coldplay	The Scientist
Iniciante	Coldplay	I Hope
Intermediário	Corinhos Evangélicos	A Minha Alma Engrandece Ao Senhor
Intermediário	Corinhos Evangélicos	Cristo Tem Poder
Intermediário	Corinhos Evangélicos	Como É Precioso
Fácil	Corinhos Evangélicos	A Tua Paz
Intermediário	Corinhos Evangélicos	A Bela Cidade?
Iniciante	Daniela Araújo	Abril
Intermediário	Diante do Trono	Lugares Altos
Intermediário	Diante do Trono	À Sombra do Altíssimo
Intermediário	Diante do Trono	Espírito Santo
Intermediário	Diante do Trono	A Bíblia
Intermediário	Dilsinho	Péssimo Negócio
Avançado	Djavan	Orquídea
Fácil	Djavan	Vamos Fugir (part. Daúde)
Avançado	Djavan	A Noiva da Cidade
Fácil	Ed Sheeran	Afire Love
Avançado	Elis Regina	Corsário
Avançado	Elis Regina	Flora
Avançado	Elis Regina	Bolero de Satã (com Cauby Peixoto)
Avançado	Elis Regina	Pois é
Avançado	Elis Regina	Cadeira Vazia
Intermediário	Felipe Araújo	Atrasadinha (part. Ferrugem)
Intermediário	Fernandinho	Grande É o Senhor
Intermediário	Fernandinho	Nova História
Intermediário	Fernandinho	Seu Sangue
Iniciante	Fernandinho	Santa Euforia
Intermediário	Fernandinho	Ao Deus de Abraão
Iniciante	Fernandinho	Vem Cear
Iniciante	Fernandinho	Seu Nome É Jesus
Intermediário	Fernandinho	Nada Além do Sangue
Fácil	Fernando e Sorocaba	Calma Aí
Fácil	Fernando e Sorocaba	A Casa Caiu
Fácil	Fernando e Sorocaba	Bobeia Pra Ver
Fácil	Frank Aguiar	Pra Comer Batom
Avançado	Fundo de Quintal	Sutilmente... Sambando
Fácil	Gabriel Diniz	É Pra Tocar No Paredão
Fácil	Gabriela Rocha	Desperta (Wake)

Intermediário	Gabriela Rocha	Glória e Louvor
Intermediário	Gabriela Rocha	Aclame Ao Senhor
Avançado	Gal Costa	Se Todos Fossem Iguais a Você
Avançado	Gal Costa	Você Não Gosta de Mim
Avançado	Gal Costa	Chega de Saudade
Avançado	Gilberto Gil	Água de Meninos
Avançado	Guilherme Arantes	Trilhas
Avançado	Guinga	Bolero de Satã
Iniciante	Guns N' Roses	Dead Flowers
Iniciante	Gusttavo Lima	Até Amanhecer
Iniciante	Gusttavo Lima	Vai Não Vai
Fácil	Gusttavo Lima	A Cor da Minha Esperança
Fácil	Gusttavo Lima	Água de Bar
Intermediário	Gusttavo Lima	Na Hora de Amar
Iniciante	Gusttavo Lima	As Mina Pira
Intermediário	Harpa Cristã	Saudai Jesus
Fácil	Harpa Cristã	Achei Jesus
Intermediário	Harpa Cristã	Cristo Morreu Por Mim
Intermediário	Harpa Cristã	Caminheemos Na Luz
Iniciante	Henrique e Juliano	Jesus Apaga a Luz
Intermediário	Henrique e Juliano	Eu Quero Ser Seu Anjo (part. Gustavo Moura e Rafael)
Iniciante	Henrique e Juliano	A Fila Anda Depressa
Iniciante	Henrique e Juliano	Quem Ama Sempre Entende
Iniciante	Henrique e Juliano	Tão Distante de Tudo
Intermediário	Hinos Avulsos CCB	Não Mais Oprimidos Seremos
Intermediário	Hinos Avulsos CCB	É Jesus Meu Refúgio
Intermediário	Hinos Avulsos CCB	Guia Me Ó Senhor
Intermediário	Hmb	Não Me Deixes Partir
Intermediário	Hugo e Guilherme	Cara de Pau / Amor Sem Juízo / Essa Tal Liberdade (pot-pourri)
Iniciante	Iron Maiden	Strange World
Iniciante	Isadora Pompeo	Hey Pai
Avançado	Jorge Aragão	Nosso Romance
Avançado	Jorge Aragão	Peço Atenção
Intermediário	Jorge e Mateus	Amor Não É Jogo de Azar
Iniciante	Jorge e Mateus	Goiânia Me Espera
Intermediário	Jorge e Mateus	Tá Faltando Eu (part. Gusttavo Lima)
Fácil	Jorge e Mateus	Calma
Fácil	Jorge e Mateus	Armadilha
Iniciante	Jorge e Mateus	Amor Bandido
Iniciante	Jorge e Mateus	Amor Turista
Avançado	Jorge Vercillo	Monalisa
Avançado	Jorge Vercillo	Memória do Prazer
Avançado	Jorge Vercillo	Verdade Oculta
Avançado	Jorge Vercillo	O Que Eu Não Conheço
Avançado	Jorge Vercillo	Luzes Que Se Movem Pelo Céu
Avançado	Jorge Vercillo	Dindi

Avançado	José Augusto	Não Dá Pra Te Esquecer
Iniciante	Jota Quest	Tanta Mentira
Iniciante	Jota Quest	Waiting For You (Party On)
Iniciante	Jota Quest	Ela É
Intermediário	Lady Gaga	Shallow (feat. Bradley Cooper)
Intermediário	Legião Urbana	Pais e Filhos
Fácil	Los Hermanos	Alguem Pra Mim
Fácil	Los Hermanos	Vai Embora
Iniciante	Luan Santana	Virou Star
Iniciante	Luan Santana	Bailando (feat. Enrique Iglesias)
Iniciante	Maneva	Mil Promessas
Intermediário	Marília Mendonça	Como Faz Com Ela
Iniciante	Marília Mendonça	Ela Tá Virada
Avançado	Marisa Monte	Carinhoso
Intermediário	Matheus e Kauan	Fica (part. Ana Vitória)
Iniciante	Matheus e Kauan	O Amor Só Existe Um
Iniciante	Melim	Maju
Fácil	Melim	Não Demora
Fácil	Melim	Confusão
Avançado	Milton Nascimento	Beijo Partido
Avançado	Milton Nascimento	Corsário
Avançado	Milton Nascimento	Peixe Vivo
Iniciante	Molejo	Beijo Molhado
Iniciante	Músicas Infantis	A Casa
Fácil	Nando Reis	A Minha Gratidão É Uma Pessoa
Fácil	Nando Reis	Livre Como Um Deus
Fácil	Nando Reis	Driamante
Fácil	Nando Reis	Luz Antiga
Avançado	Nara Leão	Folhetim
Avançado	Nara Leão	Inclinações musicais
Fácil	Natiruts	Leve Com Você
Fácil	Natiruts	Bob Falou
Fácil	Natiruts	Aldeia
Fácil	Nx Zero	Estrada
Fácil	Nx Zero	O Destino
Fácil	Nx Zero	Fração de Segundo
Iniciante	Onze:20	Menina
Intermediário	Os Paralamas do Sucesso	Óculos
Intermediário	Padre Marcelo Rossi	Sonhos de Deus
Fácil	Paula Fernandes	Meu Eu Em Você
Fácil	Paula Fernandes	Don't
Fácil	Paula Fernandes	Canções do Vento Sul
Avançado	Paulinho da Viola	Aquela mulher
Intermediário	Pink Floyd	Wish You Were Here
Iniciante	Projota	Samurai
Intermediário	Queen	Keep Passing The Open Windows

Intermediário	Queen	Good Old Fashioned Lover Boy
Intermediário	Queen	Innuendo
Fácil	Queen	Small
Iniciante	Raimundos	Minha Cunhada
Iniciante	Raul Seixas	Mosca Na Sopa
Iniciante	Raul Seixas	Eu Não Tô Nem Aí
Iniciante	Raul Seixas	Best Seller
Fácil	Red Hot Chili Peppers	Bicycle Song
Fácil	Red Hot Chili Peppers	Animal Bar
Fácil	Red Hot Chili Peppers	A Millions Miles Of Water
Fácil	Roberto Carlos	A Garota do Baile
Fácil	Roberto Carlos	Com Muito Amor e Carinho
Fácil	Roberto Carlos	Detalles
Avançado	Roberto Carlos	Como É Grande o Meu Amor Por Você
Iniciante	Sandy & Junior	Vai Ter Que Rebolar
Intermediário	Skank	Ela Desapareceu
Iniciante	Skank	Cadê o Pênalti?
Iniciante	Skank	Los Pretos
Avançado	Só Pra Contrariar	Tudo acaba em perdão
Avançado	Soundgarden	Black Hole Sun
Intermediário	The Beatles	And I Love Her
Intermediário	The Beatles	Across The Universe
Iniciante	The Beatles	Why Don'T We Do It In The Road?
Iniciante	The Beatles	Yellow Submarine
Iniciante	The Beatles	Johnny B. Goode
Intermediário	The Beatles	Girl
Iniciante	The Beatles	Act Naturally
Fácil	The Beatles	Glad All Over
Iniciante	Tião Carreiro e Pardinho	Mineira de Diamantina
Avançado	Tom Jobim	Desafinado
Avançado	Tom Jobim	Chega de Saudade
Avançado	Tom Jobim	Gabriela
Avançado	Tom Jobim	Passarim
Avançado	Tom Jobim	Sabiá
Avançado	Tom Jobim	Ana Luiza
Avançado	Toquinho	Este seu Olhar / Corcovado / Se todos Fossem Iguais a Você
Avançado	Toquinho	O Caderno
Avançado	Toquinho	Cotidiano nº 2
Avançado	Toquinho	Maria Vai Com as Outras
Avançado	Toquinho	Na Boca da Noite
Avançado	Toquinho	Choro Chorado para Paulinho Nogueira
Iniciante	Victor e Leo	Senhorita
Iniciante	Victor e Leo	Nova York
Iniciante	Victor e Leo	Fuscão Preto
Fácil	Vitor Kley	Caminho do Hawaii
Intermediário	Vitor Kley	Flor



Intermediário	Zé Neto e Cristiano	Amigo Taxista
Fácil	Zé Neto e Cristiano	Bobo Fui Eu
Fácil	Zé Neto e Cristiano	Eu e Você
Fácil	Zé Neto e Cristiano	A Noite do Ano
Fácil	Zé Ramalho	Coração Bobo
Fácil	Zé Ramalho	Avôhai
Fácil	Zé Ramalho	Amanhã Eu Vou
Intermediário	Zé Ramalho	Chão de Giz
Fácil	Zé Ramalho	A Única Coisa Que Eu Quero
Iniciante	Zeca Pagodinho	Bagaço da Laranja
Intermediário	Zezé Di Camargo e Luciano	Cada Volta É Um Recomeço
Fácil	Zezé Di Camargo e Luciano	Curtindo Um Dia de Sol
Fácil	Zezé Di Camargo e Luciano	A Gente Fica Sem Se Amar
Iniciante	Zezé Di Camargo e Luciano	Passou da Conta

Fonte: Elaborada pelo autor.

