

**Automatic classification of music genres**

**Gustavo Costa Araújo Moreira**

**Dissertation to obtain Master’s Degree in Informatics Engineering, Specialization in Software Engineering**

**Supervisor: Fátima Rodrigues**

**Jury**:

Presidente:

[Nome do Presidente, Categoria, Escola]

Vogais:

[Nome do Vogal1, Categoria, Escola]

[Nome do Vogal2, Categoria, Escola] (até 4 vogais)

Porto, [Mês] [Ano] Calibri, 12pt

Página em branco [apagar este comentário]

Dedicatória (opcional)

….

Página em branco [apagar este comentário]

Resumo

**Palavras-chave**: Palavra-chave1, …, Palavra-chave6

Página em branco [apagar este comentário]

Abstract

O documento tese deve conter um resumo em português e outro em inglês que não excedam as 200 palavras ou 1 página A4. Quando a tese é escrita em português o abstract deve ser uma tradução em inglês do resumo.

Se a tese for escrita em inglês deve conter um resumo alargado em português que não exceda as 1000 palavras ou 2 páginas A4.

Após o resumo/abstract é obrigatório colocar as principais palavras-chave/keywords do tema em que se insere o trabalho desenvolvido, sendo permitido um máximo de 6 palavras-chave/keywords.

**Keywords**: Keyword1, …, Keyword6

Página em branco [apagar este comentário]

Agradecimentos <opcional>

Dirigidos ao(s) orientador(es) à família, colegas, Instituições, …

Página em branco [apagar este comentário]

Table of contents

1 Introduction 1

1.1 Context 1

1.2 Problem 2

1.3 Objectives 2

1.4 Value analysis 3

1.5 Approach 3

1.6 Document Structure 3

2 Value analysis 6

2.1 New Conecpt Development Model 6

2.2 Opportunity identification 8

2.3 Opportunity analysis 8

2.4 Idea Generation and Enrichment 8

2.5 Idea Selection 8

2.5.1 Analytic Hierarchy Process (AHP) 8

2.6 Concept Definition 8

2.6.1 Value proposition 8

3 Background 9

3.1 Deep Learning architecture 9

3.1.1 Benefits 9

3.1.2 Common use cases 9

3.1.3 Attention points 9

3.2 Other machine learning architectures 9

4 State of the art 10

4.1 Deep learning frameworks research 10

4.1.1 Existing frameworks 10

4.1.2 Comparison between frameworks 10

4.2 Existing applications of deep learning models for music genre classification 10

4.2.1 Showcase of existing application 10

4.2.2 Comparison between existing applications 10

4.3 Other possible approaches for music genre classification 10

4.3.1 Exploratory analysis to machine learning based models for music genre classification outside a deep learning approach 10

4.3.2 Exploratory analysis to music genre classification outside machine learning based models 10

5 Deep learning for music genre classification 11

5.1 Design 11

5.1.1 Requirements 11

5.1.2 Design alternatives 11

5.1.3 Design proposal for implementation 11

Table of Figures

[Figure 1 - The New Concept Development model as illustrated in the original paper (Koen et al., 2001) 7](#_Toc61712322)

[Figure 2 - The AHP hierarchy decision tree 12](#_Toc61712323)

**Inserir página em branco apenas se necessário de modo a que a próxima secção comece numa página à direita**

Lista de Tabelas

Tabela 1 – 1º Formato de tabela 4

Tabela 2 — 2º Formato de tabela 4

Tabela 3 — 3º Formato de tabela 4

**Inserir página em branco apenas se necessário de modo a que a próxima secção comece numa página à direita**

Acrónimos e Símbolos

**Lista de Acrónimos**

**IA** Inteligência Artificial

**Lista de Símbolos**

**β** Largura de banda

**Inserir página em branco apenas se necessário de modo a que a próxima secção comece numa página à direita**

# Introduction

This chapter has as exclusive purpose to introduce the work that is described by this document. It contains the scope definition, the purpose that the work tries to achieve, the intrinsic value that is provided to the society by it and an overall structure of this document.

## Context

Over the years that compose our humanity, it is commonly accepted that classifying a music genre by a human being is a fairly easy task Gjerdingen (2008). Classifying a music into genres has never been a problem for humans, as long as they previously understand the concept. It is a task that can be accomplished even by listening to only a small extract of music, even of five seconds or less. Subgenres are also easily identifiable by humans, assuming they are familiarized with the genre in question. For example, for someone who know trance music, it should also easily identify it’s subgenres, for instance, psy-trance, melodic-trance, hardstyle, amongst many others.

With the increasing number of songs reaching peak levels every day (Spotify, 2020), it becomes humanly impossible to classify each song individually.

The music streaming service industry is increasingly an average of 15% year over year (Statista, 2020), with growth revenue following the same percentage trends for the upcoming years.

As the consumption of music through music streaming services become more and more a standard, it also become more important to guarantee that a good quality service is provided.

That implies a state of the art music genre classification, with accurate subgenre labelling. As it becomes impossible for a human, or group of humans, to classify the entire catalogue of music available, the question that this document tries to answer is whether a machine algorithm is capable of doing the same classification that a human does, faster and more reliable.

Nowadays, the mainstream music streaming companies like Spotify, Apple, Amazon, and many other heavily really on deep learning, including the music genre classification (Tyagi, 2020).

Previous work on music genre classification based on machine learning techniques achieved a top accuracy of 74% (Landsdown, 2019). It is extremely important to keep researching on this topic, improving the accuracies of the classification models and expanding the genres that are classified.

## Problem

In 2019, on Spotify alone, nearly 4000 songs were uploaded to the platform every single day (Tim Hingham, 2019).

If the statement above is not enough to convince the reader that the amount of data that is being uploaded to the internet and made available to consumers is massive, the following mathematical analysis will help statistically understand the scale of the problem.

The average length of a song is three to five minutes. For the purpose of this section, let’s assume that the average length of a song is four minutes.

The regular work length during a single day is of eight hours, in other words, 480 minutes.

If a single person spends every single minute of it’s working hours listening to music with the single task of classifying it’s gender, it would be able to only classify 120 songs during a day.

Tracking back to the original statement made in this section, it would take 34 employees at Spotify to solely classify the songs musical genres manually, assuming the work is continuous and without interruptions, and without the need for repetitions in the process.

The point to retain here is that with the increase of available data every single day, it becomes humanly impossible to solely classify and provide good quality service to the customers by relying on humans to perform this task.

Machine learning is the main study topic, with the purpose of providing valuable assistant to humans in order to classify music genres.

## Objectives

It is the purpose of this document to present a deep learning based solution that improves the speed of music genre classification, without losing the accuracy that is obtained by humans.

This objective is wide in the spectrum regarding the meaning that can be associated to it. Therefore, defining sub objectives can help reduce the spectrum of the goal and provide valuable assistance to achieve the main objective. Having this in consideration, the sub objectives are the following:

* Study previous work developed with the single purpose of classify music genres automatically. Machine and deep learning based projects are preferrable and the biggest focus of research, but alternatives paradigms and implementation should not be discarded and should have a dedicated section to evaluate it’s performance and viability.
* Study the available deep learning frameworks that are widely available and embraced by the developer community and select one to perform the practical part of the dissertation. The usage of a widely available framework deeply reduces the risk of failure of a project. (Ben Yoss, 2019)
* Implement a machine learning based solution that classify a pre-selected number of music genres. This implementation should thrive to always use development best practices and reach the goal of performance better and faster than humans.
* Implement a basic application that connects to the machine learning based solution to classify songs without human assistance. This application should accept an audio file and output a music genre.

## Value analysis

The objective is empty in terms of value if there is no formal value proposal associated. The reader was already alerted to the fact that the amount of data in form of music genre available in the internet is massive and the trends is for this number to keep increasing in the next following years at a constant rate. With this statement in mind, with the end consumer in the centre of the focus, a solution should be formally specified.

To achieve the formal specification, there are three models that this document dedicates separate sections to further define a value analysis for the solution.

To help define the value, both New Concept Development and Analytic Hierarchy Process are used. From a business oriented point of view, the value analysis is formalized by the usage of the Canvas Business Model.

## Document Structure

This document is structured in 6(?) different chapters, each chapter containing sections, followed by all supporting references and the appendix for further details of the process.

Each chapter is summarized below:

1. Introductory chapter that contextualizes the reader about the topic of the document, the problem that is identified, the solution that is proposed and the approach that is used to achieve the proposed goal.
2. Presents a formal definition of the value analysis that this works intends to bring to the society, going through different models, such as New Concept Development an Analytic Hierarchy Process. From a business point of view, the Canvas Business Model is used.
3. A deep dive into the meaning of Machine Learning and Deep Learning, containing a dissection of the meaning of these two concepts, to help the reader easily identify terms and notations used in the following chapters.
4. A state of the art chapter dedicated to study relevant work previously done by other researchers on the topic, presenting the achieved results, highlighted concerns and identified future work that needs to be done to improve current proposals.
5. A chapter dedicated to the system requirements design for the solution to be implemented.
6. [TBD]

**Inserir página em branco apenas se necessário de modo a que o próximo capítulo comece numa página à direita**

# Value analysis

This chapter presents to the reader the value analysis made in the context of this work. Value analysis “is an organized creative approach which has as it’s purpose the efficient identification of unnecessary cost, i.e, cost which provides neither quality, nor use” (Lawrence Miles, 1946), therefore, the primary objective of this chapter is to assess how to increase the value of the work at the lowest possible cost without sacrificing quality.

In order to create value, this analysis reflects on the real problem that was introduced in the previous chapter and presents ideas that bring value to the customer. Automatic music classification is not a completely new topic, therefore, bringing value to the customer might be the increase of speed classifying one single song, increasing the accuracy of classification, or even reducing the cost of development and achieving similar results when comparing to existing solutions.

The following section of this chapter concretize the above based on the use of the New Concept Development (NCD) model of Peter Koen (Koen et al., 2001). The New Concept Development model ideas will further be complemented with an even more business oriented model, the business model canvas, to provide a clear vision of the value that this work brings to the society.

## New Concept Development Model

The New Concept Development model (NDC) is a framework developed with the single purpose of providing a concrete way of establishing new products based on an initial idea, increasing the act of creativity in the world. It was built based on the idea that there is a “widely-perceived lack of high-profit ideas entering the New Product Process Development (NPPD)” (Koen et al., 2001). To fight this issue, the New Concept Development model provides a common language (framework) of the key components that drives innovation further (Koen et al., 2001).

There are three key components that compose the New Concept Development model:

* The inner area, that defines the five key elements comprising the Front End of Innovation (FEI). These elements are:
  + Opportunity Analysis
  + Opportunity Identification
  + Concept & Technology Development
  + Idea Selection
  + Idea Genesis
* The engine, which drives the five front-end elements and is fueled by the leadership and culture of the organization
* The influencing factors, which drive product development and can be intrinsic or extrinsic, for example, organizational capabilities, outside world business strategies, among others.

The Figure1 provides a visual help to the bullet points defined above.

Imagem em preto e branco de um relógio

Descrição gerada automaticamente com confiança baixa

Figure 1 - The New Concept Development model as illustrated in the original paper (Koen et al., 2001)

Furthermore, several characteristics are worth noting.

“The inner parts of the NCD were specifically designated as elements rather than processes. Processes imply a structure that may not be applicable and could force a set of poorly designed NPPD controls to be used to manage front-end activities. The circular shape is meant to suggest that ideas are expected to flow, circulate and iterate between and among all the five elements”. (Koen et al., 2001)

## Opportunity identification

Opportunity identification is used to identify opportunities that are worth to be investigated and studied. “Business and technological opportunities are explicitly considered so that resources will eventually be allocated” to a determined area (Koen et al., 2001).

In broad terms, an opportunity “may be a near-term response to a competitive threat, a breakthrough possibility to capture competitive advantage, or a means to simplify/speed-up/reduce the cost of operations” (Koen et al., 2001).

The New Concept Development model understands that for opportunity identification to work well, sources and methods that are used to identify opportunities are an essential element.

Typical methods may be divided into two big groups:

* Formal – opportunity identification processes that are aligned with all the influencing factors, e.g: brainstorming, mind mapping and lateral thinking, trend analysis.
* Informal – more organic processes that allow the flow of opportunity identification ideas, eg: ad-hoc sessions, individual insights, or edicts from senior management.

This work was based on trend analysis, following the trend in the technological community regarding Artificial Intelligence, Deep Learning, Machine Learning and other technologies of the future.

**Automatic music classification is a hot topic**

In the Introduction chapter, it was discussed that there is an increase of data consumption regarding music content. Companies like Spotify, Apple and Tidal provide a streaming service that must be of extreme quality to it’s consumers.

In 2016, Spotify provides a total of 1387 music sub-genres in their service platform. (Nick Patch, 2016). Big corporations like Spotify are already on the edge of technology and have dedicated teams working on a sub area of machine learning called “machine listening”, and classify songs based on a set of factors, including tempo, acoustic-ness, energy, danceability, strength of the beat and emotional tone (Nick Patch, 2016). Spotify already builds it’s platform music genres classification using a still-evolving tool, but “the process is still imperfect. At one point, the computers confused the sound of the banjo with human singers” (Nick Patch, 2016).

In 2020 and with the increase of deep learning study materials, the number of studies published in recent times show that there is still a lot do understand and improve.

Leland Roberts developed a Convolutional Neural Network that classifies music genres with overall accuracy of 68%. (Leland Roberts, 2020).

The above information allows to identify an opportunity, which is to improve the accuracy of automatic music classification. Streaming services benefit from a more accurate algorithm, therefore this opportunity identification allows the generation of value for the customer, once a better accuracy is achieved.

## Opportunity analysis

The NDC model defines opportunity analysis as the validation of the previous identified opportunity. The first chronological moment of an opportunity is to identify it. The second is to study it’s viability and support the effort that might be done in the future based on facts, market needs and business alignments. (Koen et al., 2001). Taking this into consideration, this section decomposes the opportunity to validate the value that the work will bring to the society.

**Why is automatic music classification important**

From 2015 to 2020, the overall Artificial Intelligence revenues grew by 20%, and this number is expected to grow exponentially in the upcoming years (UBS, 2020).

Music genre classification is only a small portion of Artificial Intelligence, but it is currently explored by many students, researchers and big corporations that are already using Artificial Intelligence based algorithm to perform music genre classifications.

Currently, a deep investigation into the topic will not be able to find a common guideline to develop algorithms of the sort. The current documentation available is done on an individual level, within an academic context or researchers publishing investigation documents like this one, exposing a proposal of implementation and presenting the results.

No standard has been defined to this date.

In 2021, Spotify has 144 million paid subscribers, and this number more than doubled since 2017. (Spotify, 2021). The numbers tell us that streaming services are reaching a never reached number of new users before, and with the increment of users, the competition between corporations gets bigger, and therefore, the quality of the service should be better to improve customer retention.

Research on the topic that is currently done is targeted at the 11 most popular music genres, contrasting with the 1387 genres that Spotify currently provides.

There is a clear need in the market to standardize music genre classification based on genre, velocity of training, and general accuracy of the algorithm.

All of the three factors above, if improved, can bring a lot of value to the customers, and therefore, validating the value of the identified opportunity.

## Idea Generation and Enrichment

In the New Concept Development model, the Idea Generation and Enrichment, or Idea Genesis, is the process of transforming an opportunity into a concrete idea. Therefore, it “represents an evolutionary process in which ideas are built upon, torn down, combined, reshaped, modified and upgraded” (Koen et al., 2001). It is not expected that the ideas, once defined, are fixed and immutable. They go through a process of continuous improvement and may go through many iterations before is finalized. This process may include “a formal process like brainstorming sessions and idea banks so as to provoke the organization or individual into generating new ideas for the identified opportunity” (Koen et al., 2001).

For this work, both brainstorming and bank ideas were used when defining ideas for the identified opportunities. The bank of ideas came mostly from the author, while researching for chapters number 3 and 4, and also by the inclusion of members in informal discussions that have an interest in the topic (professora, aqui eu quero fazer a referência a algo que ficará em anexo, pois queria apresentar uma espécie de questionário que pretendo fazer para suportar este capítulo. Como devo fazer referência a isso?)

The following ideas are presented with the intention of identify possible future solutions to the identified opportunity.

1. **Continue the work of an individual researcher that has publicly made his work available to the community** – The objective of this idea is to not start from scratch and take advantage of the work that someone else has already done, and improve the solution to obtain better results.
2. **Build a technical solution from scratch, using a dataset that was previously defined** – The purpose of this idea is to start a technical solution from scratch. This solution might be based on a deep learning framework, but also other types of framework. The common point here is that is bases the classification on a dataset that is freely available to the community to use.
3. **Build a technical solution and a dataset from scratch** – The objective of this idea is to fully design and implement the solution as discussed in the idea number two, but using a dataset that is manually built for the context of this work.
4. **Use previous research to identify common points of implementation and propose a guideline of implementation that is accepted by the community** – The goal of this idea is to explore the work that was already done regarding music genre classification, identify common practices and guidelines that lead to the development of a reliable algorithm. Further implementation of this idea requires starting conversation with proper identities that can validate and certify the process. In other works, the end goal of this idea is to have a clear and define pattern of implementation for music genre algorithms, like Martin Fowler defined software design patterns previously.
5. **Build a music genre classification framework** – The objective of this idea is to develop a framework that is highly optimized to develop music genre classification algorithms, having as ultimate goal facilitate future work of the developer community by reducing speed of developing and facilitating the accuracy of the algorithm.

## Idea Selection

Idea Selection is the element of NDC that allows to select the idea that brings the most value. The “selection may be as simple as an individual’s choice among many self-generated options” (Koen et al., 2001). NDC does not identify a rigorous method to select the idea with the most value.

At this point, as presented in the previous section, there are 5 ideas identified, and selecting one is still left to the individual decision. To help identify the most viable solution for this dissertation, a second method is introduced, the Analytic Hierarchy Process.

### Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process (AHP) is one of the most used method in decision making environments. It was defined by Thomas L. Saaty, in 1980. This method uses both quality and quantity criteria in the evaluation process. The main idea is to divide the problem into decision tree levels, facilitating the comprehension and evaluation of the decision when selecting an idea.

The process processes the creation of a tree, where the first level represents the problem that is intended to be solved, and the second hierarchy level represents the factors of importance in order to achieve a problem solution. Finally, in the lowest level of hierarchy, are represented the ideas proposed to solve the problem.

The Figure 2 represents the decision tree of the AHP applied to the problem introduced in the first chapter of the document. The criteria to assess whether the idea is significant or not is the following:

* **Technical achievement** – The solution achieves a highly quality code structure and directly solves the proposed problem, by applying software engineering best practices.
* **Algorithm accuracy** – The solution achieves a good accuracy that is bigger than the current research standards.
* **Time restrictions** – If the solution is timeboxed to a specific time and needs to be delivered on a specific date
* **Community meaningfulness** – If the solution brings a more direct value to the community.

Diagrama

Descrição gerada automaticamente

Figure 2 - The AHP hierarchy decision tree

Having the AHP tree defined, it is necessary to proceed to the next stage. The criteria on the second hierarchy level of the AHP tree is directly compared to one another in terms of relative importance, accordingly to the model proposed by Saaty, in 1990. Table 1 provides guidance to the model proposed by Saaty.

Table 1 - Scale of criteria comparison (Saaty, 1990)

|  |  |  |
| --- | --- | --- |
| **Importance level** | **Definition** | **Explanation** |
| 1 | Same importance | Both activities equality contribute to the goal. |
| 3 | Week importance | The experience and judgment lightly favor one activity comparing to the other one. |
| 5 | Strong importance | The experience and judgment strongly favor one activity comparing to the other one. |
| 7 | Very strong importance | One activity is highly favored compared to the other one |
| 9 | Absolute importance | The evidence favors one activity with the highest certainty level possible. |
| 2, 4, 6, 8 | Intermediate values | When the criteria is in the middle of two definitions. |

Based on Table1, the comparison between criteria for the stated problem can be achieved. Table 2 provides a comparison between the different weights of each criteria, following the AHP scale. The information in the table assembles the conclusion that Algorithm Accuracy is the most important criteria, followed by Community Meaningfulness and Technical Achievement. Time restrictions on the other hand are considered to be the less importance criteria.

Table 2 - AHP Evaluation table for dissertation topic

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Evaluation Criteria** | **Technical Achievement** | **Algorithm Accuracy** | **Time Restrictions** | **Community Meaningfulness** |
| **Technical Achievement** | 1 | 0.5 | 3 | 0.75 |
| **Algorithm Accuracy** | 2 | 1 | 4 | 1,5 |
| **Time Restrictions** | 0.33 | 0.25 | 1 | 0.5 |
| **Community Meaningfulness** | 1.25 | 0.75 | 2 | 1 |
| **Sum** | 4.58 | 2.5 | 10 | 3.75 |

The next step of decision making presented in the AHP process is to normalize the table above, so that the sum of each criteria is equal to 1 and, enable the calculation of importance of each criteria in the next step. Table 3 presents the AHP normalized table of the content presented in Table 2.

Table 3 - AHP normalized table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Evaluation Criteria** | **Technical Achievement** | **Algorithm Accuracy** | **Time Restrictions** | **Community Meaningfulness** | **Importance** |
| **Technical Achievement** | 0.2183 | 0.2 | 0.3 | 0.2 | 23% |
| **Algorithm Accuracy** | 0.4367 | 0.4 | 0.4 | 0.4 | 41% |
| **Time Restrictions** | 0.0721 | 0.1 | 0.1 | 0.1333 | 10% |
| **Community Meaningfulness** | 0.2729 | 0.3 | 0.2 | 0,2667 | 26% |
| **Sum** | 1 | 1 | 1 | 1 | 1 |

Table 3 is importance since it provides guidance in the weights that should be considering when evaluating each idea individually. As stated before, Algorithm Accuracy is the most relevant criteria in idea selection for this work and should account for 41% of the decision. In the other hands, only 10% of the weight when assessing an idea should be related to time restrictions. Having this data, it is now possible to go through each idea and select the most valuable one.

* **Continue the work of an individual researcher that has publicly made his work available to the community –** This idea would be ideal in terms of time restrictions, improving speed of development and simultaneously provide a theoretical easy way to increase algorithm accuracy. However, there is no relevant data that suggest improved accuracy when building upon existing work. It is also not easy to determine if this idea is meaningful for the community since it does not bring a new perspective nor achieves technical details, since the implementation is already there.
* **Build a technical solution from scratch, using a dataset that was previously defined –** This idea starts from the premise that a good dataset exists and is able to classify more than 11 genres, and gives appropriate time to dedicate to build a more accurate and efficient classification algorithm. At the same time is contributing to the community by providing a new perspective of development and allows the visibility for technical achievement. It is not blocked from the beginning due to time restrictions.
* **Build a technical solution and a dataset from scratch –** This idea starts from the premise thar everything will be built from scratch. While building a dataset from scratch would definitely be helpful to the community and allowing for flexibility in terms of what can be classified, time consumed on this specific task would suck the entire time, not allowing enough time to improve algorithm accuracy, which is the most importance criteria. Therefore, this idea should not be considered.
* **Use previous research to identify common points of implementation and propose a guideline of implementation that is accepted by the community –** Although the idea would highly align with community meaningfulness, it would not provide direct response to the most important criteria, which is algorithm accuracy.
* **Build a music genre classification framework –** This idea aligns with the previous presented one. It emphases on community meaningfulness, but rejects or not direct aligns with algorithm accuracy.

Considering all the points above, the chosen idea for this dissertation in the idea “**Build a technical solution from scratch, using a dataset that was previously defined**”. In chapter 5 of this document, it is presented a design proposal to the solution based on this idea. All further topics in the document may focus on the concretization of this idea and aligned with it.

## Concept Definition

### Value proposition

## Canvas Business Model

## **Inserir página em branco apenas se necessário de modo a que o próximo capítulo comece numa página à direita**

# Background

## Deep Learning architecture

### Benefits

### Common use cases

### Attention points

## Other machine learning architectures

# State of the art

## Deep learning frameworks research

### Existing frameworks

### Comparison between frameworks

## Existing applications of deep learning models for music genre classification

### Showcase of existing application

### Comparison between existing applications

## Other possible approaches for music genre classification

### Exploratory analysis to machine learning based models for music genre classification outside a deep learning approach

### Exploratory analysis to music genre classification outside machine learning based models

# Deep learning for music genre classification

## Design

### Requirements

### Design alternatives

### Design proposal for implementation

Referências

|  |  |
| --- | --- |
| (DEEWR 2009) | Department of Education, Employment and Workplace Relations 2009, DEEWR, Canberra, viewed 21 November 2009, http://www.deewr.gov.au/ |