

**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 Document Structure 3

2 Value analysis 6

2.1 New Concept Development Model 6

2.2 Opportunity identification 8

2.3 Opportunity analysis 9

2.4 Idea Generation and Enrichment 10

2.5 Idea Selection 11

2.5.1 Analytic Hierarchy Process (AHP) 11

2.6 Concept Definition 15

2.6.1 Value proposition 16

Inserir página em branco apenas se necessário de modo a que o próximo capítulo comece numa página à direita 18

3 Background 19

3.1 Deep Learning architecture 19

3.1.1 Benefits 19

3.1.2 Common use cases 19

3.1.3 Attention points 19

3.2 Other machine learning architectures 19

4 State of the art 20

4.1 Deep learning frameworks research 20

4.1.1 Existing frameworks 20

4.1.2 Comparison between frameworks 20

4.2 Existing applications of deep learning models for music genre classification 20

4.2.1 Showcase of existing application 20

4.2.2 Comparison between existing applications 20

4.3 Other possible approaches for music genre classification 20

4.3.1 Exploratory analysis to machine learning based models for music genre classification outside a deep learning approach 20

4.3.2 Exploratory analysis to music genre classification outside machine learning based models 20

5 Deep learning for music genre classification 21

5.1 Design 21

5.1.1 Requirements 21

5.1.2 Design alternatives 21

5.1.3 Design proposal for implementation 21

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

**Nenhuma entrada de índice de ilustrações foi encontrada.**

**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, music has been part of the human nature. Rhythm and sounds are fairly easily categorized into music genres. 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, it becomes humanly impossible to classify each song individually.

The music streaming service industry is growing at a fast pace, and now there is more offer for consumers than ever before.

As the consumption of music through music streaming services become more and more a standard, it also becomes imperative 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 dissertation tries to answer is whether a machine learning system is capable of doing music genre classification faster and more reliable that a human does.

Nowadays, the mainstream music streaming companies like Spotify, Apple, Amazon, and many other heavily really on machine learning algorithms to automatically perform music genre classification.

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

Today the worldwide music catalog, thanks to globalization, increase of accessibility to the internet and quality of life improvement of the society in general, allows to the general population to consume music by paying a monthly fee to a streaming service, which on the other hand, should provide a service of quality an rigorous content.

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 duration of a song is three to five minutes. For the purpose of this study, 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 its gender, it would be able to only classify 120 songs during a day.

In chapter 3, the amount of data related to music upload is analyzed and explored in detail, where the reader can understand why relying only on humans to perform such task is currently considered hideous.

Machine learning is the main study topic, with the purpose of providing valuable assistance 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.
* 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 trend 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 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 concretizes 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 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 in data consumption regarding music content. Companies like Spotify, Apple, and Tidal provide a streaming service that must be of extreme quality to its consumers.

In 2016, Spotify provides a total of 1387 music sub-genres in its 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, the 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 shows that there is still a lot to understand and improve.

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

The above information allows identifying 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 previously identified opportunity. The first chronological moment of an opportunity is to identify it. The second is to study its 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 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 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 has 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, the 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 finished. 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.

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 it 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 methods 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 describes 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 Table 1, 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 important 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 important 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 previously 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

This dissertation has three main goals.

The first one is to identify what the community was already able to achieve regarding music genre classification. This classification might be achieved with a variety of different tools and techniques. Those are: machine learning, deep learning, machine listening, which all belong to a broader topic called Artificial Intelligence, or it might even be achieved by using techniques that are completely irrelevant for Artificial Intelligence. It is an objective of this work to identify all the possible techniques that are currently used. In parallel, a concrete report of what the community was already able to achieve, must be reported in this document. The first goal of this work should be able to answer to some questions: “What was the biggest accuracy already achieved?”, “What genres are currently being considered?”, “What is the cost and effort of creating a dataset capable of improve the accuracy?”. Those are question that should be clearly answered in the next chapters.

The second goal of this dissertation is to provide a technical solution that allow the automatic classification of music genres. This solution should follow the best practices of software development, and should provide value to users, in the sense that is capable to do what it is supposed to do: correctly classify a music genre and with better accuracy than previous works.

Finally, the third goal is to increase the overall knowledge of the community by providing additional information regarding music genre classification. The outcomes, improvements, blocking points and breakthrough achieved during this work must be shared among the tech community to help mature the technology.

To help clarify the goals above, and to turn this work into a viable business product, the next sections of this topic will clarify the goal by using the business model canvas.

### Value proposition

At this point the reader is familiar with music genre classification, the relation of this sub-genre as a small area of actuation of a broader topic of Artificial Intelligence, and understands that there is a deficit in accuracy in previous attempts to solve this problem.

The reader also understands that automatic music genre classification is more than just an academic research. Companies like Spotify and Apple already use machine listening to develop their algorithm, with different goals, like providing recommendations of playlist or classifying songs are uploaded to the platform (Paul Pandey, 2018).

This work intends to help solving the current issues that have been reported in previous work, understand the failures, and provide a solution for them, contributing to the main goal of this dissertation: improve the algorithm accuracy.

In section 2.5.1, it was identified that due to time restrictions, a dataset will not be developed from scratch, so it is important to clarify that the development will not face improvements on that specific topic.

Therefore, this works intends to provide to companies, individual researchers and academics additional information on how to correctly develop music genre classification algorithms, the steps to achieve it, and additionally, how can the algorithm be adapted to other use cases, if the study becomes relevant.

In order to transform the previous statements into a concrete business idea, the Canvas Business Model for the use case will be presented. The Canvas Business Model is a business model that describes how an organization creates value (Osterwalder, A. and Y. Pigneur, 2010).

Table 4 provides detailed information of the business plan to consider when implementing this project.

The Key Partners are identified and listed. The university in which this work is inserted is a natural key partner. Furthermore, specialized knowledge platforms like Udacity and Udemy are considered key partners, as they can provide additional knowledge for the algorithm development, which otherwise would not be of easy access. Key activities include an initial research of the literature and a later implementation of a new solution to the stated problem in the previous chapters. The open source community is identified as a customer, since one of the goals of this project is to directly provide knowledge to this community. It is simultaneously clear that previous work done is the back bone and the starting point of this project and will provide invaluable knowledge. Finally, it is considered that developing the work will have no associated cost, unless there is the need to setup a cloud provider to boost the development of the solution. An optional cost is the usage of platforms like Udacity or Udemy to have access to learning content related to the topic, and therefore, contributing to the overall goal of this work.

Table 4 - Business model canvas

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Key Partners** | **Key Activities** | **Value preposition** | **Customer Relationships** | **Customer Segments** |
| - ISEP, TOWARDSDATASCIENCE, DeepLearning.ai, Udacity, Udemy  - Researchers on the subject, other academics | - Intensive review of the existing literature, and identifying common issues and problems that make the accuracy lower  - Design and implement a technical solution for automatic music genre classification  - Contribute to the community by publishing an article with the findings. | By identifying the most commons issues, problems and restrictions when developing the algorithm, the literature provided by this dissertation must contribute to the addition and increase of knowledge in the community.  Furthermore, the develop code must also be publicly available at the end of this work. | Publish solution on Github. Find interested companies and researchers to keep working on the implemented solution. | - Open source community. By making the work available on Github, allow anyone interest proceeding with the work. |
| **Key Resources** | **Channels** |
| Public articles regarding automatic music genre classification.  Paid courses related to machine and deep learning in platforms like Udacity or Udemy. | Published articles  Tech blogs  Private company courses |
| **Cost structure** | | | **Revenue Streams** | |
| The results achieved by this work will be made available publicly with no costs.  To train the algorithm, the usage of a Cloud provider might be necessary and have some costs.  Additional knowledge by taking a course on Udemy or Udacity may be necessary. | | | Improved accuracy  Reduced training time | |

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

# Background

In this chapter, the reader is presented with the background that leads to the topic of this work. Key concepts will be presented and explained related to Artificial Intelligence, the broader topic that sits in the top of the chain (section 3.1), and subsequent topics of specialization will be further detailed, starting the Machine Learning (section 3.2), followed by Deep Learning (section 3.3), and finally explaining Machine Listening (section 3.4).

## Artificial Intelligence

The term “Artificial Intelligence” was first stated by John McCarthy in 1956 and it is the science and engineering of making intelligent machines, especially intelligent computer programs (McCarthy, 2004).

Despite being first stated and researched in 1956, Artificial Intelligence remained a topic that was deeply connected to science fiction amongst the general society. In recent years though, the term started to become relevant again to the tech community since it became possible to actuate on real world problems by using Artificial Intelligence (AI) techniques. This actuation became possible by the sudden availability of large amounts of data and the corresponding development and wide availability of computer systems that can process data in achievable human time (IBM, 2020).

Artificial Intelligence is a broad topic. When applied to a specific problem, AI is concretized into a specific domain.

Figure 3 introduces a simplified representation of current layers of AI. Artificial intelligence is the most external layer, and therefore, the most generic one.

The intermediate layer is related to machine learning, one of the many categories of AI. Machine learning includes every algorithm that given a pre-determined set of inputs, produces and output based. The term machine learning exists because the output is also based on past information that was collected by the algorithm.

In the inner layer lies deep learning, a subcategory of machine learning, where, as stated before, given a pre-determined set of inputs, an output is generated. However, deep learning uses what is usually called a neural network, imitating the human brain to perform a task (Andriy Burkov, 2019).

Diagrama

Descrição gerada automaticamente

Figure 3 - A representation of Artificial Intelligence scopes

AI is not only represented by the identified categories.

In order to provide a full context to the reader, Figure 4 represents a hierarchy tree of AI and the most common categories that are currently relevant to the tech industry (Jeff Giangiulio, 2017).

Artificial Intelligence splits into 2 main categories: Robotic Process Automation, and Cognitive Computing. The scope of this work relies only on Cognitive Computing.

Cognitive computing can also be further split into different categories:

* Natural language processing, used for automatic text translations, text completion, etc.
* Computer Vision, used for image recognition and machine vision. This category is currently used for self-driving car technologies.
* Speech, used for speech to text and text to speech algorithms.
* Machine learning

This work relies on the exploration of Machine Learning, specially with Deep Learning techniques. The next sections will further explore these two topics separately.

![Diagrama

Descrição gerada automaticamente]()

Figure 4 - Hierarchy tree of AI categories

## Machine Learning

Machine learning is a subfield of computer science that is concerned with building algorithms which, to be useful, rely on a collection of examples of some phenomenon. These examples can come from nature, be handcrafted by humans or generated by another algorithm (Andriy Burkov, 2019).

### Architecture

In Software Engineering, in order to fully understand a subfield, it is important to understand the relying architecture. Machine learning architecture is still volatile, but the community has accepted the following discussion points as concrete architectural cornerstones for this subfield.

**Supervised Learning**

In supervised learning, the training data is a mathematical model that consists of both inputs and respective desired outputs (Andriy Burkov, 2019). In this case, the system is able to find a relationship between the input and outputs obtained during the training data, and apply the same technique for a given input afterwards. This architecture is usually related to classification and regression analysis problems.

**Unsupervised Learning**

Unlike supervised learning, the input feature vector that corresponds to the training data of the algorithm does not contain output. Unsupervised learning identifies relations between various features of an input, such as trends, color schemas, between others, and generates output accordingly (Andriy Burkov, 2019).

**Semi-Supervised Learning**

In Semi-Supervised leaning, the training data simultaneously contain labeled and unlabeled data. Usually, the quantity of unlabeled data is higher than the labeled data (Andriy Burkov, 2019). The goal of this architecture is the same as supervised learning but, given a good dataset condition, is expected to generated a better model than the one generated by relying only on supervised learning data.

**Reinforcement Learning**

This is used in training the system to decide on a particular relevance context using various algorithms to determine the correct approach in the context of the present state. These are widely used in training gaming portals to work on user inputs accordingly (Andriy Burkov, 2019). This topic is not relevant for the context of this dissertation, so it will not be mentioned from now on.

**Components**

Machine learning system follow a component development structure.

In this work, in order to classify music genres, data collection and data generation are two major tasks related to this work. And end classification is the desirable end result.

From now on this document, the terms “data collection”, “data generation”, “dataset”, “training data” and “output” are used interchangeably. While chapter 4 explain in deeper details these concepts, here they are introduced for better engagement with the rest of the chapter. Figure 5 exposes the main components of a machine learning system architecture (Markus Schimtt, 2020).

1. Data Generation – Every machine learning project starts by data generation. This can be achieved by collecting the data from scratch, using previous generated data as a starting point, or by using new and previously obtained data.
2. Data Collection – After generating the data, this data should be easily accessible. This should be achieved through a well-formed database collection, cloud storage on any other kind of storage that is programmatically accessible without major difficulties evolved.
3. Feature Engineering Pipeline – The act of preparing the data for training. At his stage, raw data is selected, labeled, transformed and combined in order to become ready to be feed in the machine learning algorithm.
4. Training – Training is where the data prepared on step 3 is feed to the algorithm. A set of inputs correspond to a set of output that are understood by the algorithm for future classifications/predictions.
5. Evaluation/Validation – Happens after the training step. It is where the model obtained in step 4 is validated in terms of accuracy, correctness and prediction. A bad result achieved in this step should require rework of previous steps.
6. Task Orchestration – Non mandatory step of a machine learning algorithm. It facilitates the training, validation and data consumption of big datasets, by scheduling training times. When used, it is usually associated to a cloud provider, like AWS, Azure or Google Cloud Platform.
7. Prediction – Classify an input based on the previous training. In this work, prediction is used to classify any new music and obtain an output, the corresponding music genre.
8. Infrastructure – The underlying hardware required to run the solution. This might be achieved by using proprietary hardware or using cloud providers.

Diagrama

Descrição gerada automaticamente

Figure 5 - Basic Machine Learning System Architecture

### Benefits and common use cases

Machine learning gained attention in recent years because it brings benefits to past solutions to solve the same problems.

It is a fairly accepted statement that machine learning is capable of performing some tasks better than humans, and that is the single reason why there is an increased interest in building solutions from this subfield of computer science.

Some common use cases are spam detection, real time business decision making, creation of financial models, stock market prediction, amongst others (Nikhil Gupta, 2017).

In recent years, music genre classification has also been a major target of study, with some studies reporting 68% of accuracy (Leland Roberts, 2020).

Chapter 4 focus on providing more case studies related to the current status of music genre classification.

### Attention points

Despite growing usage of machine learning related technologies, one big attention point to have in consideration is related to Data Generation.

In order for a model to work correctly and achieve desirable results, the collection of good data to perform the task is an imperative step to achieve in first place.

Without a good training dataset the work is compromised from the beginning.

For this work, several datasets are considered. The GTZAN and A Million Song datasets are amongst the possibilities to be used as Data Generation and Data Collection steps of the machine learning system pipeline. Chapter 5 provides details about the design of the solution, and Chapter 7 focus on the solution.

## Deep Learning

Deep learning is a subtype of Machine Learning and is inspired in the structure of the human brain. Deep learning algorithms attempt to mimic similar conclusions as humans would by continually analyzing data with a given logical structure (Artem Oppermann, 2019).

The above behavior is achieved by using a multi-layered structure of algorithms typically called neural networks.

A neural network owes its name to the comparison of a human brain. Just as the brain identify patterns and classifies different types on information based on what is capable of process, neural networks can be taught to perform the same task given a pre-determined set of data (Artem Oppermann, 2019).  
Figure 6 represents a typical neural network used in deep learning algorithms. X vector represent the input, and Y vector represent the outputs.

W1, W2, W3 and W4 represent the weights associated within each layer, which are thereby represented by H1 Vector, H2 Vector and H3 Vector.

Diagrama

Descrição gerada automaticamente

Figure 6 - Typical architecture of an artificial neural network

**Feature Extraction**

A common machine learning algorithm requires manual feature extraction.

Feature extraction is a term used to represent a pre-processing step that is required to be done manually to raw data so that later a machine learning algorithm can perform classification or regression tasks on the pre-processed data.

Feature extraction is usually quite complex and requires detailed knowledge of the problem domain. This pre-processing layer should be adapted and refined several times before achieving optimal results (Pier Ippolito, 2019).

This is the big differentiation of deep learning and neural networks. With neural networks, the Feature Extraction step is not required. The neural networks layers implicitly learn how to process the raw data.

This statement can be easily explained with an example. A machine learning model that intends to classify if a given image represent a car or do not represent a car, feature extraction should be done manually. That is, the different features of a car, such as size, wheels, mirrors, colors should be extracted and feed into the algorithm as input data. On the other hand, with neural networks, the feature extraction step described becomes irrelevant, as the model recognizes unique features from the given input and makes correct predictions from there.

At this time of writing od this document, there are several identified neural network architectures that are accepted by the community as good enough to solve the majority of the problems by reusing and adapting the accepted neural network architectures.

Like any good software development practice, to develop a good deep learning system, a good architecture should be applied. The next section goes through the most common neural network architectures.

### Architecture

In this section, the most common used neural networks are described. After this section the reader should be familiar with deep learning architectures and identify the structure used in the project, described from Chapter 5 onwards.

**Perceptrons**

Perceptons are considered the first generation of neural networks.

This neural network uses feed-forward propagation, and the error is back propagated.

A limitation of perceptrons is the fact that all features are identified from the beginning, manually. A new requirement change in the feature set requires total rearrangement of the first neural network layer, which is time consuming and not practical. It is usually associated with machine learning only problems, not commonly associated with deep learning techniques.

**Convolutional Neural Networks**

Convolutional Neural Networks distinguishes from perceptrons since they use unsupervised learning to classify data. These networks are primarily used for image and audio processing.

They work by given an input, several convolutional layers are applied and perform feature extraction. Feature extraction is possible by transforming the initial image into different images based on similar aspects that the network identify as a feature. That is the definition of a convolution.

A Convolutional Neural Network is able to successfully capture the spatial and temporal dependencies in an image through the application of relevant filters. The architecture performs a better fitting to the image dataset due to the reduction in the number of parameters involved and reusability of weights. In other words, the network can be trained to understand the sophistication of the image better. (Yann LeCun, 1998).

**Recurrent Neural Networks**

A Recurrent Neural Network is a type of convolutional neural network which uses time series data. They are commonly used in natural language processing, speech recognition and image captioning problems (James Le, 2018). Similarly to Perceptrons and Convolutional Neural Networks, RNN use datasets to learn.

The distinguish point of RNN is that they introduce the concept of memory, as they take information from prior inputs to influence the next sequence of inputs and outputs. While traditional deep neural networks assume that inputs and outputs are independent of each other, the output of recurrent neural networks depend on the prior elements within the sequence (IBM, 2020).

**Long/Short Term Memory**

Long/Short Term Memory networks can be considered a specialized version of RNN’s, as they try to solve the memory vanishing problem of previous inputs. While RNN’s introduce the concept of memory, by influencing the next input sequence with weights from the previous analysis, with time and over iterations, that memory becomes irrelevant for newer inputs (Hochreiter & Schimhuber, 1997).

LSTM define memory cells, which store previous values and holds the data until the network explicity forgets that same data.

These memory cells are particularly import in speech translation, where an input word at the beginning of a sentence might be more or less relevant to predict the next word in the text.

**Gate Recurrent Unit**

Gate recurrent units are a slight variation of LSTM. The differentiation is that Gate recurrent units don’t need the cell layer to pass memory along the next iteration (James Le, 2018). The calculations within each iteration insure that the current values are being passed along either retain or discard a high amount of old information.

In general, GRU’s tend to be faster. A general rule of tomb to decide between Gate Recurrent Units or LSTM is related to expressiveness and accuracy. If speed is more important than accuracy, GRU’s are usually the chosen architecture.

**Hopfield Network**

**Boltzmann Machine**

**Deep Belief Network**

**Autoencoder**

**Generative Adversarial Network**

### Comparison

Table 5 - Comparison between different neural network architectures

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Architecture | Date of conception | Use Cases | Speed vs Accuracy | Uses Memory | Relevant for audio processing |
| Perceptrons |  |  |  |  |  |
| CNN’s |  |  |  |  |  |
| RNN’s |  |  |  |  |  |
| LSTM’s |  |  |  |  |  |
| GRU’s |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

## Machine Listening

### Architecture

### Benefits and common use cases

### Attention points

# 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 applications

### 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/ |