

Digital Business Innovation Lab

a.y. 2019/2020

SONIK

Powered by Sony



Bartesaghi Andrea	928656
Bersani Federico	926343
Gupta Pragati	938654
Khan Mohsin	943115
Palombini Elena	945573
Vigoni Silvia Francesca	921181
Vitali Camilla	920935

Table of CONTENTS

- 0 EXECUTIVE SUMMARY
- 1 INTRODUCTION
- 2 EXTERNAL ANALYSIS
- 3 INTERNAL ANALYSIS
- 4 LEAN CANVAS
 - 4.1 PROBLEM
 - 4.2 CUSTOMER SEGMENTS
 - 4.3 SOLUTION
 - 4.4 UNIQUE VALUE PROPOSITION
 - 4.5 UNFAIR ADVANTAGE
 - 4.6 CHANNELS
 - 4.7 KEY METRICS
 - 4.8 INVESTMENT ANALYSIS
- 5 VALIDATION
- 6 TEAM COMPOSITION

0 EXECUTIVE SUMMARY

The aim of the following report is to better understand the recent explosion of the TikTok's success and to help Sony Music to define a strategy for scouting emerging artists and promoting Sony's ones through this channel. All the analyses were made considering mainly the Italian market in order to have a modular project that can be applied also to other countries in case of scalability. The report was divided in three main phases, explained below.

In the **first phase**, the Vision, Mission and Purpose, that were initially defined to align the overall goals, are exposed. It follows an analysis of the environment, in which the company operates, focusing on the main factors that are affecting the music context and putting a particular attention on the TikTok phenomena.

The external analysis is accompanied by an internal one, highlighting the most important processes inside the company and, most of all, the key resources that Sony can leverage. Subsequently, a SWOT analysis provides a big picture of the threats to overcome in the industry, the internal weaknesses to avoid or to solve, the opportunities that could be exploited and the strengths that can support the company in defining the strategic alternatives. In conclusion, it is shown how the Strategic Posture and Strategy Palette respectively gave a further confirmation of the direction previously identified by the SWOT and a clear idea of the position of the company in terms of strategy and context-Indeed, the last tool suggested an adaptive approach, and so being fast and focusing on experimentations in a high dynamic and uncertain environment. In light of those analysis, the Lean Startup approach, with its iterative process "**build-measure-learn**", was embraced: first of all, the model was built, then validated and finally it was possible to learn from what has been observed.

Hence, the **second phase (build)** is mainly focused on the Lean Canvas and the creation of the solution, highlighting its main features and functionalities. The Lean Canvas has the advantage to shape quickly and easily the main elements needed to develop the business idea.

First, the analysis goes through the problems, that partially emerged in the first phase: low malleability and high unpredictability of the TikTok environment, high visibility by competitors and high profitability and riskiness linked to artists.

Only after, it was possible to come up with the idea of two automatized and structured processes, taking the perspective of an internal function of Sony. After defining the hierarchical structure of the customers, the solution and its grounding technologies are presented in detailed. The section goes through the definition of all the blocks of the Lean Canvas, finally shifting the attention to the economic perspective and trying to assess its attractiveness for Sony Music.

Lastly, the **third phase (measure-learn)** consists in the validation of some of the most important hypotheses over which the solution was built. For this operation the Business Model Canvas was created and divided in three different areas related to feasibility, desirability and viability of the idea. For each area the main assumptions were listed, but only three of them resulted to be testable during the project period. Then, using Test cards, the hypotheses were validated through varied methods (survey, collection of data from TikToK and music streaming platforms, training and testing of one algorithm) in order to demonstrate the resilience of the Business Model in the real world.

1 INTRODUCTION

Art influences us, whether it be painting, movie, sculpture or music. It has the power to unify diverse and distant people and to provide unique experiences, conveying the feeling of being part of something, a sense of belonging. Behind the deployment of such an emotional tool like music, there is a structured network of actors and processes that deserves attention.

The music landscape is formed by three main players:

Majors labels: under the control of multinational conglomerate organizations, the biggest in the industry.

Independent labels: not under the control of the majors.

Vanity labels: founded and managed by the artists themselves.

Sony Music Entertainment, founded in 1929 through the merger of several record companies, is one of the subsidiaries of the American multinational Sony. It is operating globally in the music industry and it is one of the three majors leading the sector.

During the years, the ways in which music has been distributed evolved. It was initially released through analogical products, then digital ones and finally through Cloud platforms, which allowed the listeners to be even closer to music.

Sony Music has always kept up with the technological influences that affected the industry, adapting its business model to the evolutions. Lately, they found out how the new lip-sync platform TikTok was taking over the music industry. Hence, the company decided to exploit this emerging trend, sensing its potential as a new mean to scout new artists and to foster the marketing activity of Sony's own artists.



Figure 1: Music Technology Evolution

According to that, the overall **vision**, the **mission** and the **purpose** were firstly defined. This was helpful to align all the team on the same pace, to assure an overall coherence within all the phases and to guide the process of the idea generation.

Vision

Exploit the power of digital technologies as a "geyser" able to push the company on a higher level and to get the best from the music landscape's evolutions.

Mission

Be the best and the first in discovering new viral artists. Engage with the artists suggesting always new and innovative ways to rapidly and broadly promote their songs, to transmit their personality and to put them closer to customers.

Purpose

Create a structured and technological process to enhance the performances of the A&R and Marketing functions, adapting to the evolutions that are affecting the market.

2 EXTERNAL ANALYSIS

PEST

To have a better understanding of the external environment and the exogenous factors that Sony has to deal with, a PEST analysis was carried out. This brought to the identification of the trends that are currently affecting the music industry from a Political, Economic, Social and Technological perspective.

A specific paragraph was dedicated to TikTok, the social trend of our interest that is rapidly and hugely impacting the sector.

Political

As a result of the digitalization process, even the legislative perspective had to be managed to regulate all the aspects related. Laws about copyright, aiming at protecting moral and economic rights of the author, started to be really threatened with the diffusion of piracy and illegal streaming. The Article 13 of the EU *Directive on copyright in the digital single market*, or *Directive 2019/790*, has the aim of harmonizing the regulatory framework of copyright with streaming platforms that are based on user generated content. According to the law, before a creation can be uploaded, the artist must give the license of use to the platform and so, indirectly, also to the users. The accessibility and use of customer data for marketing purposes, due to the spreading of digital music, are protected by the European *General Data Protection Regulation (GDPR)*, or *Regulation 2016/679*.

Economical

The diffusion of piracy in the late 90s was the biggest issue for the music industry. It resulted in a continuous erosion of the physical channels' global revenues, that decreased by 81% from 2001 to 2019. Simultaneously with the spreading of streaming platforms, that increased hugely since 2015, the music industry started to grow again. In 2019, the streaming growth was still huge (+24%) and this channel accounted for more than half of the global revenues (56%), mainly due to subscription audio streams. Italy showed the same global trends and statistics from Statista show that the forecasted revenues until 2024 will be characterized by a continuous growth of this channel.

Social

Thanks to social media platforms, musicians can get real-time feedback from their followers to create a true brand partnership connection. Now, musicians are able to present their music to a broad range of fans through Facebook, Instagram, Twitter and other social networks. In 2020, while the world is heavily impacted by Corona Virus, music artists are hugely using social media to entertain their fans.

Before the Internet, the only way for artists to sell their albums or other merchandise was in the form of physical interactions at music stores or shows. Social media's role in music is also significant to sponsor their merchandise.

Technological

Technology has opened the gates for independent artists that are now putting their fate in their own hands bypassing labels, uploading their projects to online streaming platforms and speaking to their fans directly through their social profiles. All the production, marketing and distribution efforts can be done locally just by using the available technology.

The second important trend is harnessing of the power of Artificial Intelligence. All the music platforms rely on algorithmic foundations: they use Machine Learning to work out music preferences and listening habits of their users and, starting from the consumer history, they can predict what is the next track the user would like to listen to. This is the basis of a Recommender System.



Analysis of TikTok

TikTok is a mobile application that allows users to create videos that last from 15 to 60 seconds with a music soundtrack. It is known as Douyin in its Chinese home market, and it was launched there in September 2016 by ByteDance and pushed globally the following year. In November 2017, the parent company acquired the famous rival app *Musical.ly* and merged it within TikTok in August 2018, maintaining the name of TikTok.

The most popular video contents made by the users include comedy skits, lip syncing and dancing. The aim and the power of the app is to enhance people to be creative and foster their imagination. There are many options available to express creativity: create your own video, react to a specific one, share it through different platforms and make duets with other users.

In the last few years, the app has been experiencing a **huge success**. This is also thanks to some of the world's most sophisticated AI, capable to analyse users' interests and preferences through the data provided by their navigation inside the platform and to display personalized contents in the "For You" page.

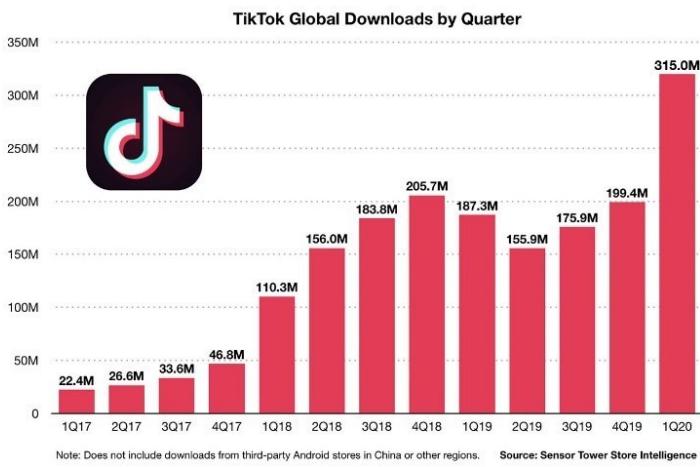


Figure 2: TikTok Global Downloads by Quarter, from SensorTower

This boom can be demonstrated by some data: in 2020, TikTok is one of the world's best-loved apps. The latest available and official statistics show that there are 500 million of active users around the world per month. This makes TikTok the **sixth-most used social app in the world**, only behind some giants like Facebook properties, YouTube and WeChat.

Another interesting data is the huge number of downloads until November 2019, accounting for 1.5 billion, that made the app become the one most downloaded in the world. By the end of the year, the year 2019 is responsible for 44% of the total downloads.

Between 2018 and 2019 the average session duration grew of 25%, that is an extremely high growth compared to other platforms, with an average value of 347 seconds. This has put TikTok in the first place as the app with the highest session's duration.

As regards demographic data, 60% of users are aged between 16-24 years old. The number of female users outnumbers male ones with almost a 2:1 ratio in US, but this ratio increases to 10:1 in some Asian countries.

Moreover, TikTok's great boom is hugely **impacting the music industry**: TikTok has signed agreements with the major labels, through which they provide the 15-60 seconds clips of tracks.

TikTok has the power to make a song trend even in one night, that is why major labels are struggling to use it as a mean for scouting or promotion: to be the first is fundamental in this highly **dynamic environment** that can rapidly **boost the virality** of a track.

Anyone can upload an audio to the app, and it can be searched and used by others. With the algorithm behind the app, the song can be displayed to more and more viewers in the For You section. However, it is the association between music and video that creates the trend, and this is the power and the distinctive advantage of this platform.

One of the most famous examples is *Old Town Road* by Lil Nas X. Before posting it on the platform, the song was unknown, but the associated hashtag #yeehaw, the dance and the cowboy outfit inspired a lot of people to reproduce it, which brought the song on to become a hit with global success. This is only an example, but there are plenty of similar cases, such as *The Git Up* by Blanco Brown and the Italian *Bando* by Anna.

In conclusion, the platform can be considered a "**pure digital**" **Big Bang disruption** for the following reasons: it is growing extremely fast; it is a multi-sided platform serving final users, music labels, advertisers and third parties acquiring data; it leverages existing assets and technologies; it creates a new value for the final users, that is disrupting the incumbents in the social media market.

3 INTERNAL ANALYSIS

The focus was then shifted to the internal environment of the Sony company. The Porter's Value Chain was used in order recognize which are the strengths to leverage on, going through the analysis of all Sony's activities.

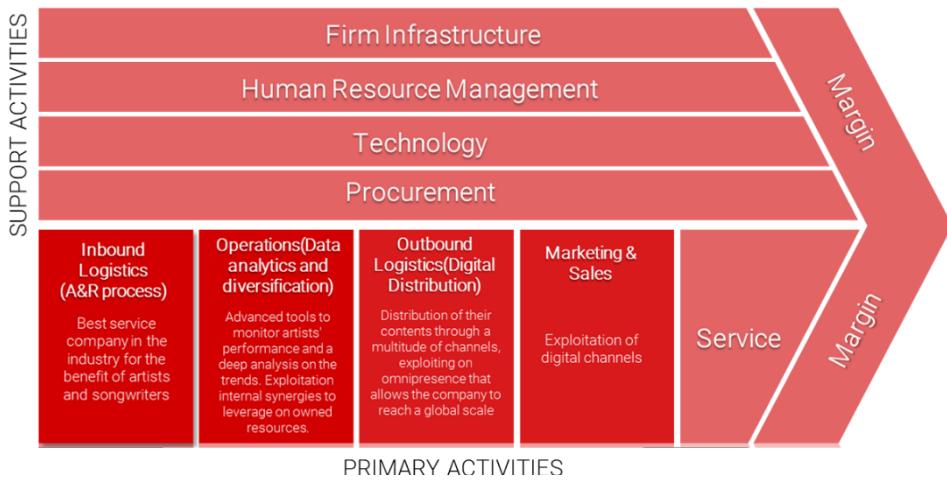


Figure 3: Porter's Value Chain

Inbound Logistics - A&R process (Artist & Repertoire) - Sony Music's mission is "Helping creators realize their dream" as stated in the annual report. With this aim, the Artist and Repertoire unit has a great importance and strives to support the creator in each step of her/his career. The main aim is to highlight the abilities and spirit of the creator to guarantee a long-term success that gives the artists their deserved recognition.

Operation – Diversification of Business - Thanks to the diversification of Sony business, which embeds the unit of Entertainment, Electronics and DTC services, the company can exploit internal synergies to leverage on owned resources. Being a source of differentiation, this helps in creating competitive advantage. Indeed, thanks to the diversity of Sony group, the company is able to support the artist through owned resources of high-level quality, exploiting in-house technologies that are under a constant update.

Operation - Data Analytics – Sony's employees use advanced tools to perform daily monitoring of the artists' performance and a deep analysis on trends, drawing from them a data-driven strategic action plan to maximize the impact and success of new releases. Furthermore, there is another purpose of data collection and analysis: to obtain insights on the audience and on the most effective channels to reach each artist's listeners, in order to optimize the track discovery. This allows to create a specific audience segmentation analysis to target listeners in a more focused way. To do all of this, third party data are also used from several streaming platforms.

Outbound Logistics: Sony can distribute its contents through a multitude of channels. The omnipresence of the company allows it to reach a global scale, and to make its music available to all the people in the world. The spread presence in many distribution channels enables Sony to improve its chances to make songs viral and make artists known all around the world. To do so, the company exploits teams of professional experts that focus on differentiated distribution strategies. What concretely allows Sony to reach people all around the globe and earn money from this, is the set of partnerships and licenses that the company has with the main providers of music platforms (Spotify, Apple Music, Amazon Music, YouTube, TikTok, ...). The main challenge of the company is to adapt its channels to the turbulence of the environment in which is operating.

Marketing & Sales: In order to promote music and artists, Sony makes use of its extensive expertise in marketing, driving its growth and enhancing its sales exploiting mainly social media.

Technologies: according to Sony's innovative culture, the company is in a constant update mood and it continuously looks for new solutions that they could be even able to develop in-house. In that sense, the

company is able, still thanks to the differentiation of the business, to support units like the A&R and Marketing, that are constantly looking for new smart solutions to support the artists in their success' enhancement. Great attention is paid to apps for the artist to monitor their success, Artificial Intelligence application fields, and content IP. Some other possibilities offered are the chance to record in the private Sony Music Studio and to be supported while working across different fields.

SWOT Analysis

The purpose of the SWOT analysis was to summarize the results of the previous examinations about the external market trends, that are affecting and transforming the music industry, and about the internal environment of Sony company.

This tool provided a general picture of the situation in which Sony music is operating and suggested some useful insights, that turned out to be crucial as a guide toward the solution.

<p>STRENGTHS</p>  <ul style="list-style-type: none"> • High-level quality of A&R department • Diversification of business and exploitation of internal synergies leveraging on owned resources such as proprietary technologies • Advanced tools to perform data analytics • Presence on a multitude of channels • Extensive expertise in marketing • Constant research in terms of new technologies and new partnerships in order to develop new solutions in-house • Partnership networks with streaming platforms and social networks impacting on music trends • Second major global record music company, but the real brand power is due to Sony's artists 	<p>WEAKNESSES</p>  <ul style="list-style-type: none"> • Lack of a structured process in order to discover and evaluate emergent artists in the new dynamic environments • Lack of a differential competitive advantage in the scouting of new talents and in the marketing strategies in the new dynamic environments • Low visibility of Sony brand in the market
<p>OPPORTUNITIES</p>  <ul style="list-style-type: none"> • Increasing trend of the market of streaming platforms • Real-time feedbacks from customers through social media • Broader potential fanbase thanks to social media • Power of AI and machine learning mainly used in the field for recommendation systems 	<p>THREATS</p>  <ul style="list-style-type: none"> • Continuous maintenance and control of solutions in order to protect the copyright and to safeguard customers' data • Independent artists have now the means and the channels to produce, distribute and promote their music thanks to new technologies • More and more dynamic and fast environment in terms of virality of music through social networks • High concentrated market: high visibility of other players' moves

Figure 4: SWOT Analysis

The starting points were the weakness of Sony's brand, in terms of low visibility by the customers on the market, and the threat represented by the high concentration of the music industry, characterised by a low number of big players that can easily notice competitors' moves and imitate each other.

A possible way to overcome these hostilities was to move in the direction of a back-end solution based on emerging technologies, i.e. Artificial Intelligence, Machine Learning and Data Analytics. This solution would enable Sony Music to be more agile than its competitors in discovering new artists, to build an effective marketing strategy and, therefore, to create a sustainable competitive advantage that is difficult to imitate.

Additionally, thanks to these technologies, other opportunities could be exploited in the market, such as the collection and analysis of customer feedback related to a broad fanbase.

This can be feasible only leveraging on Sony's strengths related to the high-quality level of the A&R department, the ownership of proprietary technologies, the innovative culture, the advancement in data analytics technologies and the broad network of famous artists and partners collaborating with the firm.

Strategic Posture

This model gave a further confirmation of the direction previously identified with the SWOT.

The strategic posture is used to assess how companies behave when they have to take decisions and make investments based on what other companies do. There are three dimensions that are fundamental: industry turbulence, concentration and demand growth.

The **turbulence** in the music industry is low: the entry rate is not so high since the industry is dominated by few big players and to compete with them, knowledge of the processes, capital (financial, human, physical and intellectual) and a strong IT organization are needed.

The **concentration** is high, few majors are operating and dominating the market with 66.2% of the market share. The remaining 33.8% is managed by many small independent labels. The concentration is still increasing due to mergers and acquisitions: the last one concerns Sony Music Entertainment's acquisition of EMI Music in 2018. Due to a high concentration, all the actions of the majors can be seen by competitors and easily imitated.

The **growth** of the **demand** is highly increasing, as emerged from the economic trends in the PEST analysis, mainly due to streaming channels. The market is based on new content creation and, since the product sold is music, the demand for it will hardly decrease. On the contrary, it will be always higher thanks to the global accessibility to music, enabled by streaming platforms and by the customers who are always "hungry" of new music.

Low turbulence, high concentration and high growth suggest to the companies inside the industry to apply a strategy that is coherent with what other players are doing, without differentiating and diverging from industry norms. Therefore, competitors will tend to imitate each other, and this is the confirmation that a key differentiating factor in this environment is something that cannot be visible to the other firms.

Strategy Palette

This tool provides a framework based on three main dimensions (predictability, malleability, harshness) that allowed to define the best strategic approach for Sony Music in the industry.

Since the interest is in finding a strategy to operate within the TikTok platform, it was better to consider as a reference environment only TikTok rather than the whole music industry.

Today the **malleability** is very low, the company finds difficulties in shaping and monitoring the environment; the main problem concerns the lack of data and the lack of available information that the company has regarding the app. In fact, the main barrier is the agreement between Sony and TikTok which is not so advantageous, since they must rely on fewer data that is not enough to understand how the environment is changing.

The lack of malleability is also explained by the fact that the company is operating in a new environment, which puts Sony in the condition of gaining awareness of which tools the company can rely on and which knowledge must be developed in order to maximize its responsiveness.

Furthermore, it is possible to state that Sony Music has little **predictability** on the ongoing trends in terms of music tastes toward a song or an artist, since they can burst overnight.

Sony is just a spectator that observes changes, but it is not able to shape them. In this kind of environment, the key is being faster than your competitors in finding, studying, following, scouting and promoting artists.

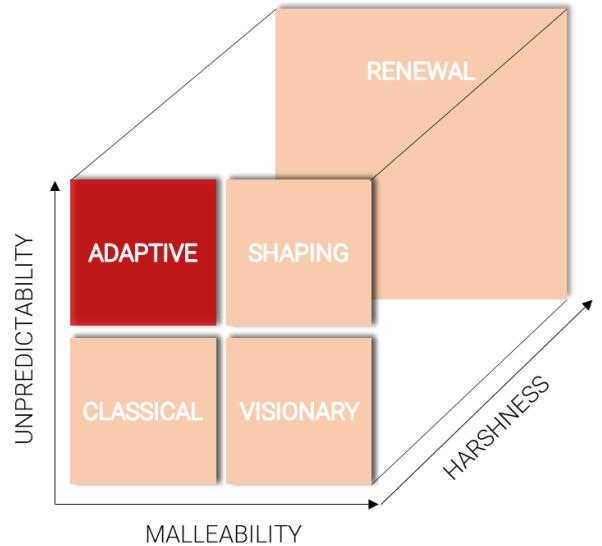


Figure 5: Strategy Palette

Despite the high level of unpredictability and the low level of malleability, being the industry highly concentrated and being Sony a major, the **harshness** is low because it is easy to survive and so, the competition is not so strong.

Hence, a low malleability, a high unpredictability and a low level of harshness locates Sony in the Adaptive quadrant. This means that Sony needs to be fast and the best way to be successful is through experimentation: it is important to try different directions generating new options to test, to select the most successful one and to implement it. Due to the fast-changing environment, it is also fundamental to continuously iterate the process in order to maintain their advantage.

4 LEAN CANVAS

The model is suitable for highly dynamic and uncertain environments, in which moving fast is a priority. The tool allows to shape in a quick and easier way all the necessary element to develop and test the business ideas as soon as possible.

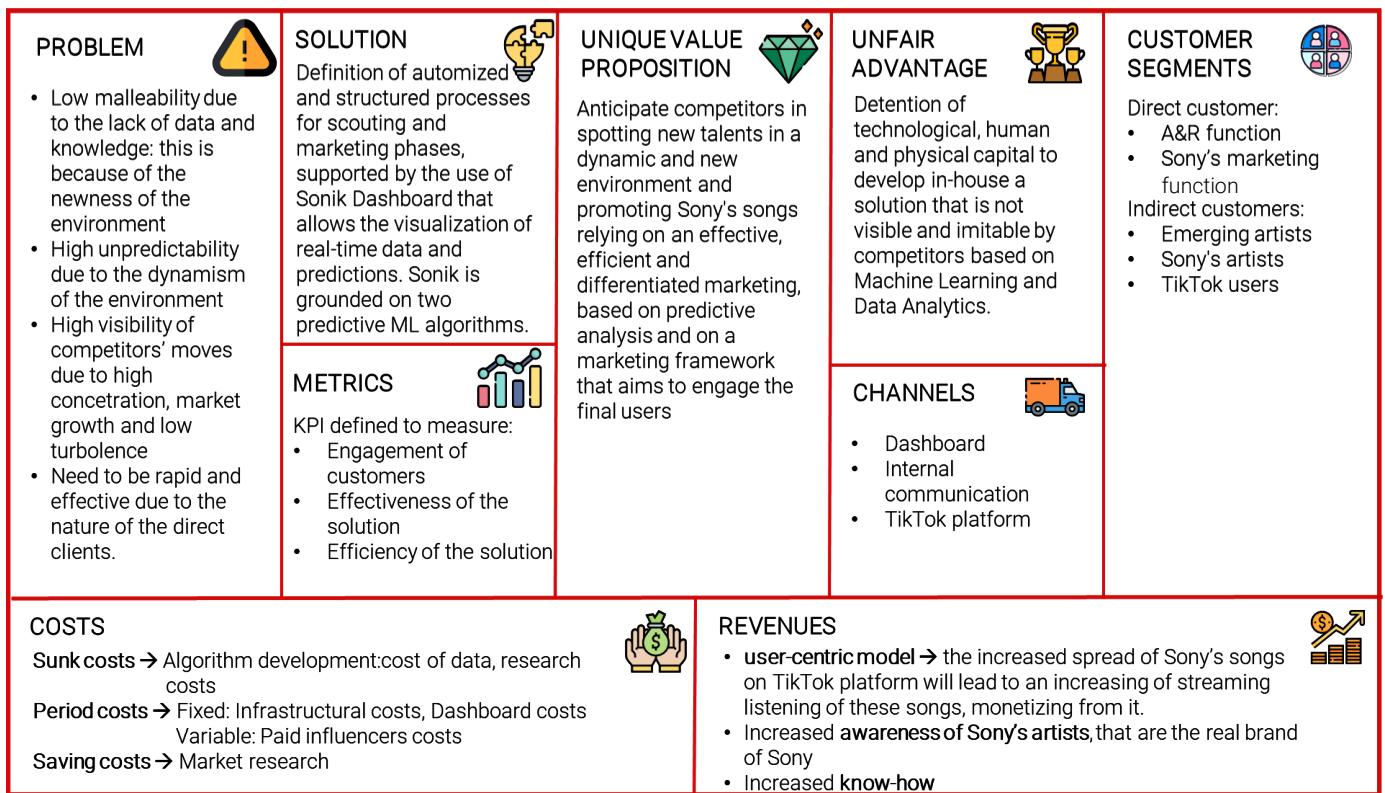


Figure 6: Lean Canvas

4.1 Problem

The direct customers of music labels are artists, who are highly profitable clients. Moreover, the risk and the variability associated to their success are high and they could be reduced through a diversification of the roaster of artists. Hence, it is fundamental not to lose them in the scouting phase and currently Sony's A&R is scouting artists on TikTok only once they are famous. The TikTok environment is highly dynamic and to be the first could be difficult: virality can be created in few days. Indeed, people are becoming more and more addicted to the platform, boosting views in short periods. Sony is interested in riding the wave of this success both for scouting and promotion purposes.

The problems, as emerged from the Strategic Posture and Strategy Palette, are that the industry is characterized by a high visibility of competitors moves; low malleability due to the lack of data, processes and knowledge of a new environment; low predictability due to the uncertainty and velocity of trends on TikTok.

Idea

The initial solution was to add a landing page to the website of Sony Music Italy where emerging artists could sign in with their works and their TikTok account in order to let Sony monitor their performances and understand which of them could be the next star in the music field. But Sony does not leverage on the website as a channel, since its brand is not known and to promote this kind of initiative would require a great marketing effort and expense. Therefore, a new idea has been developed focusing on the analysis carried out.

The new solution consists in the definition of a **structured process** both for the A&R function and for the Marketing one. Having some guidelines to follow and automatizing the processes will allow to the company to act fast.

The processes that were defined are enabled by the usage of the **Sonik Dashboard**, a tool that shows real-time results of a complex process of prediction. Indeed, a predictive analysis is performed by two different algorithms based on Machine Learning technologies.

As regards the purpose to promote Sony's artists, the marketing process is also enabled by a **marketing framework** that provides differentiated guidelines on how to exploit the TikTok channel, according to the different attitudes of the artists towards the platform.

4.2 Customer segments

According to the perspective taken, the proposed solution is addressing directly Sony Music functions. It was possible to distinguish a **hierarchical structure** to define the customer segments:

Direct customers: they will directly interact with the guidelines and with the Sonik Dashboard.

- **Sony A&R function**: it will benefit from the main advantages guaranteed by the solution for discovering faster new and talented artists.
- **Marketing function**: to which we are offering the chance to save time and costs for marketing researches in order to find the most "TikTok suitable" clip and a customizable solution for the Sony's roaster of artists for their promotion.

Indirect customer: the processes defined help Sony scout and support the artists, so it was fundamental that the idea could create value also for the direct clients of Sony's A&R and Marketing functions, who have been called the indirect clients of the solution:

- **Artists of Sony roster**: the complementary use of the algorithm and of the marketing framework helps them understand how to use the new channel of TikTok to increase their popularity and to strengthen their relationship with fans, increasing their retention.
- **Emerging artists**: they are the objective of the scouting process. Thanks to the solution ideated, they will be spotted in early stages, so there is the need to monitor them before signing a definitive contract. Hence, it is important to create a value that will keep them locked with Sony for the monitoring period.
- **TikTok users**: without the interest of TikTok users in the marketing strategy, no artist will be encouraged to embrace it. To develop a marketing framework for the campaign, it was necessary to consider what the final listener appreciates.

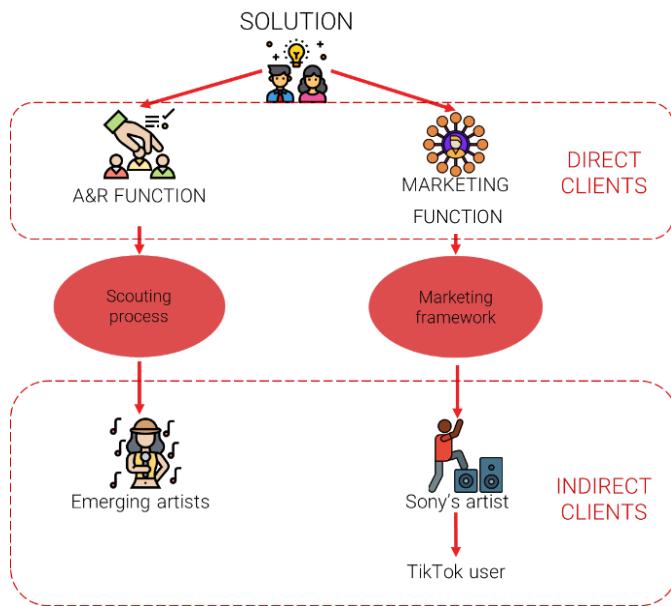


Figure 7: Customer Segments

4.3 Solution

Technologies

The idea is grounded on **Machine Learning technologies**. Nowadays, more and more companies are devising data-driven solutions to tackle a wide variety of issues. The reason is that these technologies are proving to be extremely effective to solve even the most complicated problems.

Machine Learning algorithms consist of mathematical models, which are trained on sample data, and then improve automatically through experience, eventually being able to make predictions or decisions.

The solution relies on **two algorithms** based on **supervised learning**, that have the aim to solve the difficult issue of understanding and predicting what enables a TikTok audio clip to stand out and become extremely popular.

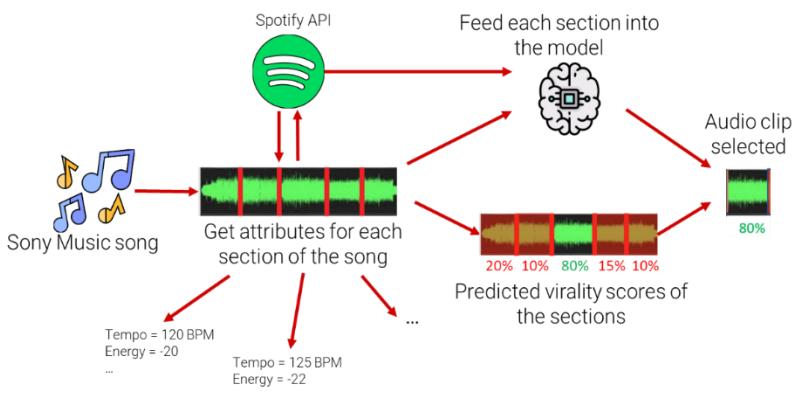


Figure 8: Selection of the best audio clip

with a sample of non-successful ones. The first step is the extraction of songs' viral features through the Spotify API. The algorithm will learn how to classify each clip of song. After that, there will be the testing phase (again using samples of successful and non-successful songs), in which the accuracy of the algorithm is assessed. When the model is created, the algorithm will work in this way: it receives Sony's songs from Sony's database and it returns a score associated to 30s clips of songs. It returns a 0 if the clip has low virality features or 1 in the opposite case.

On the other hand, the **second algorithm** will predict the future popularity of a given TikTok audio content. In this way, the A&R department can scout emerging artists from the platform before they go to the top of the charts. This algorithm will be called as the "**scouting algorithm**" since it supports the scouting process.

The scouting algorithm works on the same basis of the marketing one, but the main difference is that it should receive in input clips of new songs published on TikTok. It is possible to collect the list of songs just published on the platform by web scraping, but only if the new song uploaded is already present on Spotify, it will be possible to collect their main features through the Spotify API. At the moment, Sony does not have the access to data of songs only published on TikTok; two different scenarios were assumed:

- In the future Sony will redefine the agreement with TikTok in order to receive these data directly from the partner. the ideal situation is the one in which Sony can receive from TikTok real-time data about new published songs in order to be sure to be the first. However, it can be enough to receive, for instance, the top 200 songs two times per week, recently published on TikTok, and their main features. Sony Music is already receiving weekly data from the partner, so it is reasonable to think they can gather more if they can create an advantage even for the partner with the new solution. How this can be done will be explained next.
- Currently TikTok does not have an API suitable for our purposes, but probably in the future TikTok will create an API similarly to Spotify. In this way, it will be possible to get the main features from any recent TikTok clip even if it is still not present on Spotify.

Furthermore, this algorithm will return a score associated to the songs, that will then be visualized on the Dashboard in a more user-friendly way with a badging system that can help Sony's A&R to rapidly identify the different potential of different songs and artists.

Through the **first algorithm** it is possible to optimize a specific part of the process of releasing songs to TikTok: the choice of the 30 seconds snippet. This is essential because choosing a particular segment of a song instead of another, can determine its success or its failure in terms of likes, views and engagement of the users. In the following sessions, this algorithm will be called as the "**marketing algorithm**" since is related to marketing purposes. The marketing algorithm will be trained with a sample of songs that have been successful on TikTok and

In the following paragraphs, there is the detailed explanation of how the whole processes are working, highlighting all the touchpoints with the direct client. Due to the nature of the solution itself, which is internal, it is fundamental to show not only the steps through which the customer will interact with it, but even all the technologies that lay behind it. Finally, a service blueprint has been drafted for the two customers processes in order to better summarize the interconnections between technologies and interfaces.

A&R and scouting

Along the process, the manager interacts both with Sonik and the emerging artist.

In particular, she/he starts the journey by opening the Dashboard on her/his device and having access to the "scouting section".

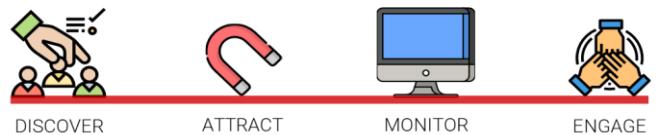


Figure 9: A&R Process

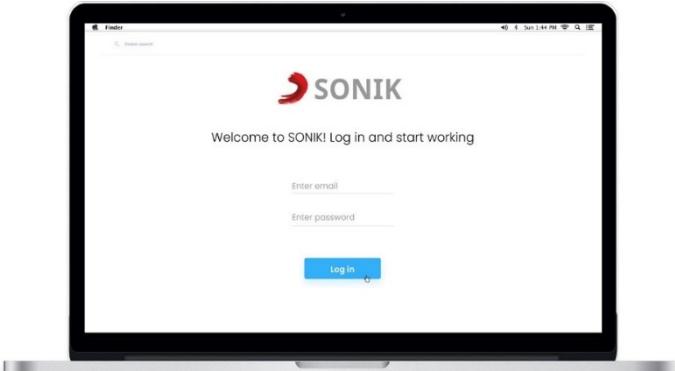


Figure 10: Welcome Page

Thanks to the connection with the scouting algorithm, the Dashboard shows the results, obtained from the predictive analysis, through a **badging system**; in this way the manager is able to check the trends and their potential virality in every moment.

What the user sees is a classification of songs considered as potentially viral in four main categories identified by a "badge":

- **Bronze** (less trending)
- **Silver**
- **Gold**
- **Platinum** (most trending)

The badge is a visual representation of performances and provides a quick way to analyze the trends and algorithm outputs. Its allocation takes into account different factors: it is initially assigned based on a "translation" of the score given by the algorithm, but then it may change according to the trend of the songs on TikTok in terms of views, re-uses in videos and the velocity of the growth of these numbers in time.



Figure 11: Scouting Page

The upper table in Fig. 11 represents the page with the long list of analyzed songs, but the user can filter it according to different criteria:

- Badges: if interested in the most performing.
- Views AND/OR videos: to connect the badges assigned to a quantitative representation of the trends.
- Publication times: it may happen that a song went viral in one night or in a week, so it can be useful to visualize only the latest songs to be sure to be the first in spotting them.
- Country: according to the location of the subsidiary, in this case Italy.

Thanks to those filters, it is possible to analyze a restricted amount of songs and to choose the ones with the highest potential to be successful. This phase is associated to human intelligence, since the evaluation of the performances of the artist cannot be only based on machines, but also to human emotions and experience. After the experts' selection, a reduced number of emerging artists are contacted to be offered with a **short-term contract**.

Short-term contract: it mainly consists in a "lock-in" period of time in which the emerging artist is asked to ensure her/his loyalty toward Sony without signing contracts with other labels. In turn, Sony gives her/him the chance to produce a new song through the advanced resources that a major like Sony is able to provide.

The short-term contract allows the A&R function to monitor the growth of the emerging artist giving time to verify the prediction of the algorithm; this contract should be also valuable from the emerging artist's point of view, otherwise she/he has no motivations in accepting it, therefore it was conceived as a double-exchange contract.

Once the emerging artist accepts the short-term contract, the A&R select her/his song and move it to the lower list of Sonik, the list of songs under short-term contract. Through this list, the manager can monitor the performances of these songs, and, if interested in more details, from the list it is possible to access to a page dedicated to the performances of a specific track or of a specific emerging artist.



Figure 12: Artist Page

The page dedicated to the artist (Fig. 12) helps to get overall information to understand if she/he is worth to be signed with a long-term contract. With this purpose, it is possible to monitor different parameters such as the average views and shares of all her/his songs uploaded on TikTok and other platforms, the average badge of her/his TikTok clips and the listening statistics on streaming platforms. These data are gathered through web scraping techniques. Once the short-term contract period ends, a conclusion on the performance of the artist is drafted as a consequence of the continuous monitoring, and just in case of confirmation of the prediction of the algorithm, she/he is offered with a long-term contract that allows her/him to be officially part of the Sony Music Roster.

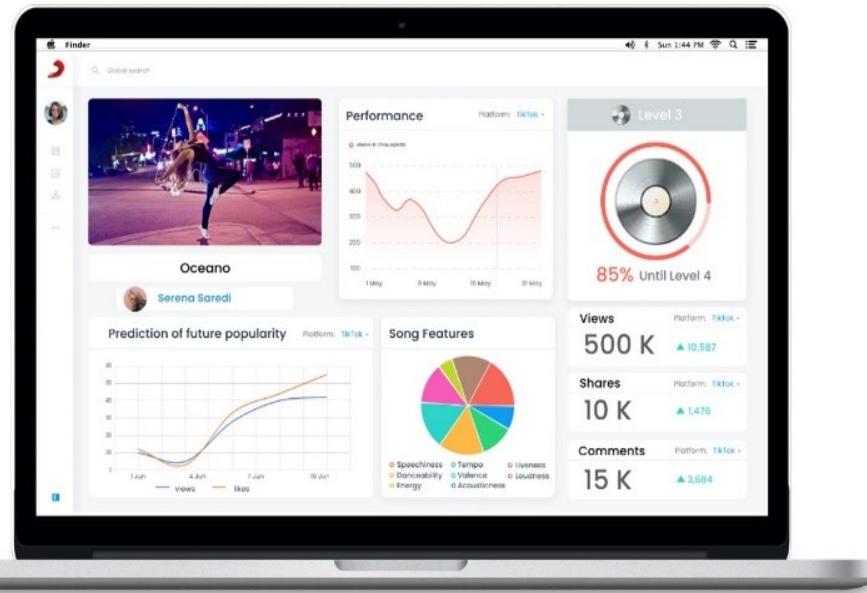


Figure 13: Song Page

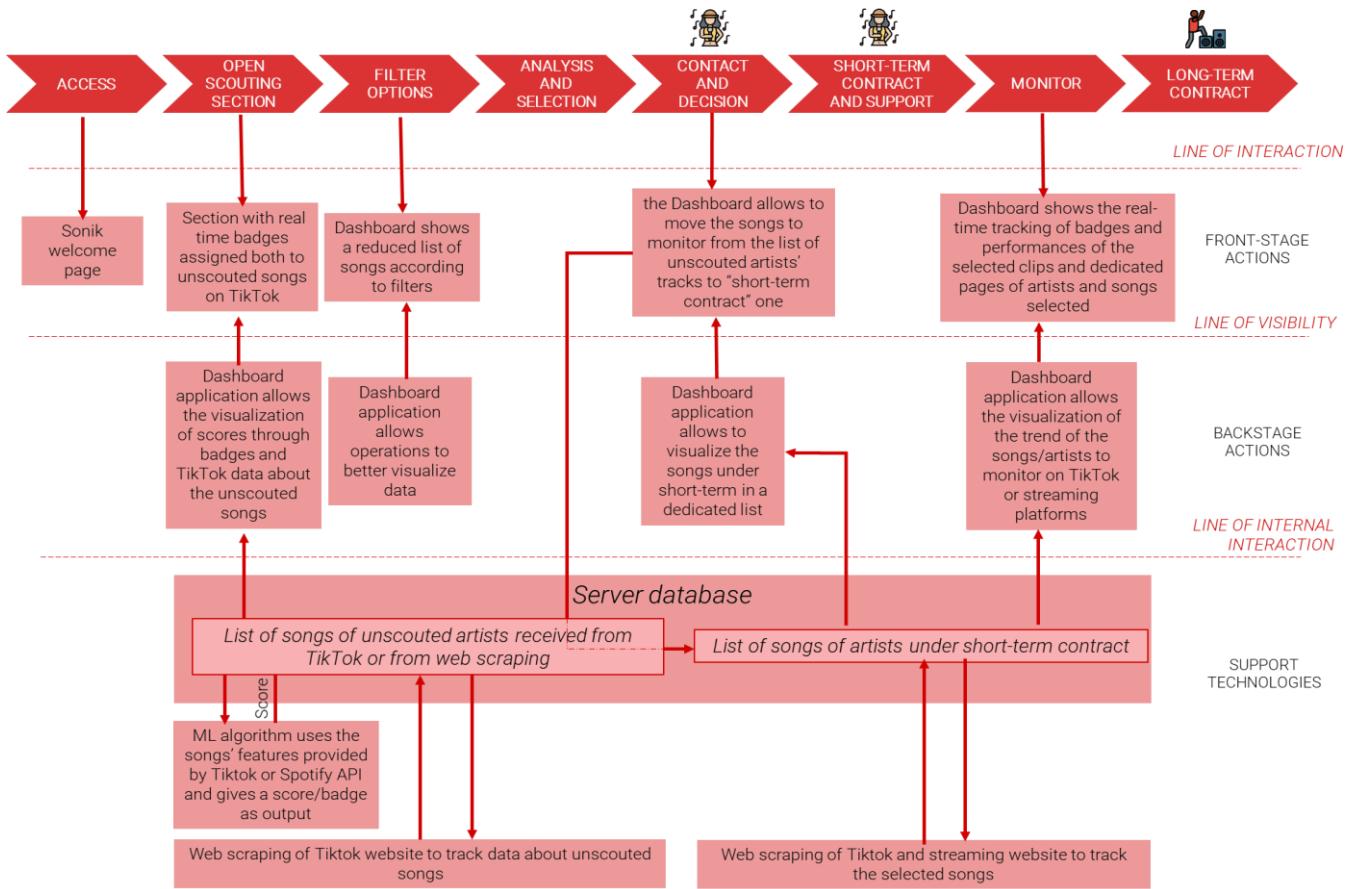


Figure 14: A&R Service Blueprint

Marketing process

First of all, it is necessary to make an important consideration. According to the observations made through the platform, three main elements have been identified as impacting on the success of content on TikTok: the **virality of the track**, the **video content** and the **personality of the creator**. The marketing algorithm aims at improving the first element, while for the second and third, a marketing framework is needed.

Sony's clips that are more likely to be viral according to the algorithm, are visualized by the Marketing function in a dedicated section of the Dashboard. There is also the possibility to filter, for instance by artist, by country or by track, in order to visualize a restricted number of options. From this list, the marketing experts choose the clips that can be object of the Marketing Framework and the artist is contacted in order to discuss together the strategy designed for her/him. This step is important since Sony believes that the contact with the artist is fundamental, indeed they are treated as partners rather than as clients. Moreover, in this way the personality of the artist can be preserved, as well as its genuineness, that is a key in the relationship with the audience. If the artist accepts to use the clip of song on TikTok, the Sony's employee selects the track from the upper list of Sonik and moves it to the lower one, that represents the list of clips selected for the marketing on TikTok.

After that, there is the marketing framework execution. In the case of "TikTok friendly" artists, it consists in the creation of a challenge, whose content and related hashtag are agreed with Sony's BU. In the case of TikTok distant artists, the Marketing unit supports the artist with the choice of the TikTok influencer, and the degree of creative freedom to leave to him.

At last, through the lower list of clips on Sonik, the Marketing unit is able to monitor the progress of the clip on TikTok in terms of views and shares, thanks to web scraping techniques. Moreover, it is possible to visualize more details, such as performances on streaming platforms, about a single clip in a dedicated page like the one of the scouting section (the only difference regards badges that here are absent).

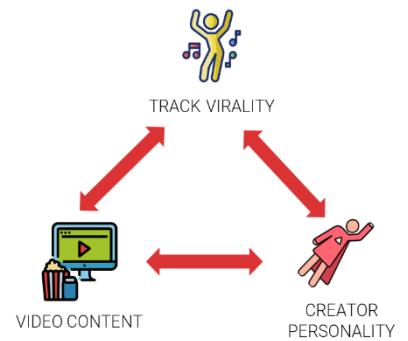


Figure 15: Virality Factors



Figure 16: Marketing Page

The algorithm can hugely reduce the marketing effort accelerating the analysis for the choice of the best segment of song to use on TikTok and increasing the amount of songs analysed in a short period of time. However, it must be considered that it cannot completely substitute human intervention. Indeed, there can be some facets in the evaluation of a clip that may not be taken into consideration by a machine, but that can be recognized only thanks to intuitions given by experience. Furthermore, the accuracy of the algorithm cannot

be 100%. Hence, probably the Marketing unit will still pick autonomously some tracks, that may not be considered viral by the algorithm, but in which they see potential. Sonik also shows the effectiveness of algorithm suggested clips comparing it to the one of autonomously picked.

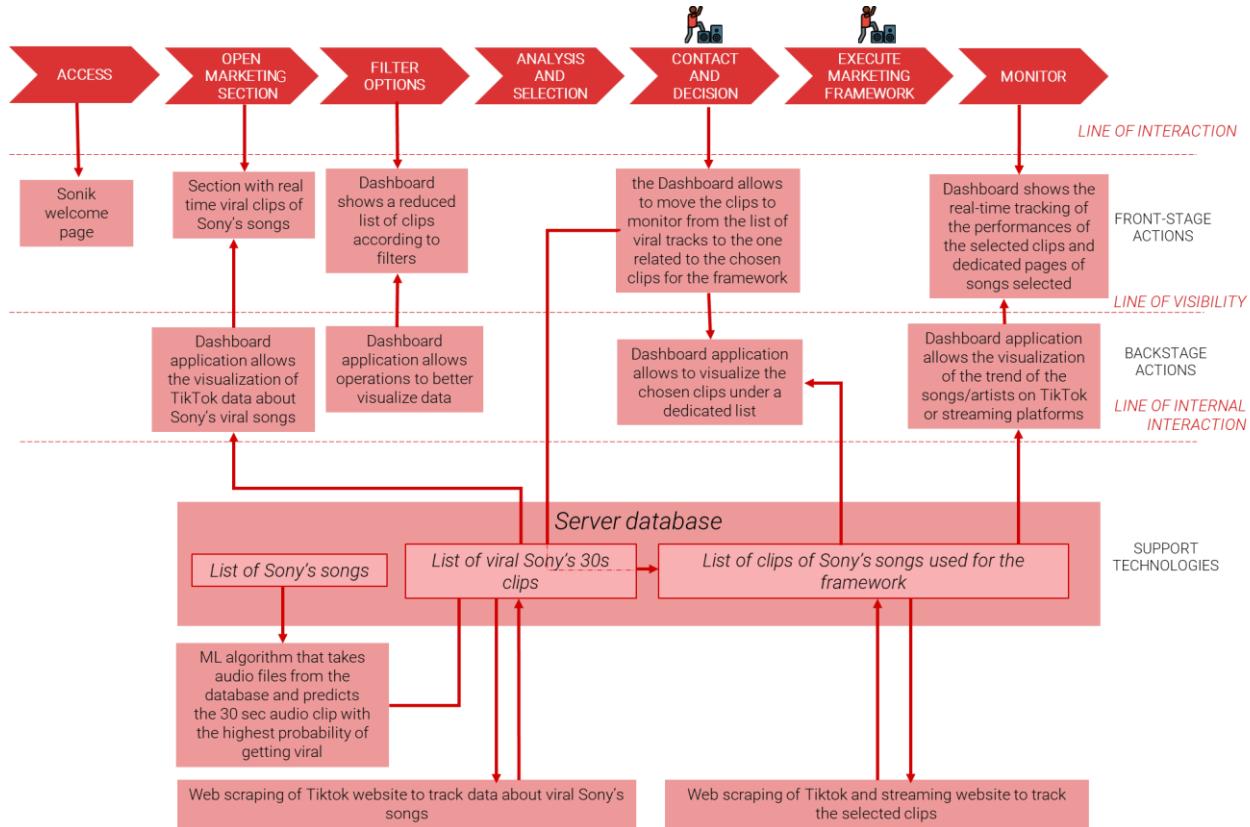


Figure 17: Marketing Service Blueprint

Marketing framework

Once the 30 seconds clips have been chosen, it is necessary to develop a meaningful strategy that can lead to the enhancement of the track's popularity, of the artists' fanbase and to the engagement of those that are already fans. To suggest the best marketing framework to Sony's artist, the roster of musicians was divided into two macro categories according to their attitude to TikTok. On one side, there are the more **TikTok friendly** ones; on the other side, there are those whose personality, age or artistic reputation are **far from the app** and/or its audience. Indeed, the belonging to a segment is influenced by the age of the artist, the genre, the artistic reputation and the image that she/he wants to transmit to her/his public. In order to make the distinction more understandable, two Personas has been drafted to embody concrete examples.

TIKTOK FRIENDLY ARTIST

As regards this first group of artists, it is important that they are present in first person inside the app. For them, this is the best way to acquire visibility and gain popularity in a social network where the average user is most likely part of her/his listening audience. However, the focus is to promote her/his musical repertoire reaching as many users as possible, even those that are not fans.



JULIET

Giulia Castelli
Female
19 years old
Milan (IT)
Pop, Urban
Activity from 2019 - present
1 Album
0 Musical Awards
Audience: teenagers, social addicted, trend follower and fashion-lovers

BELIEFS
«I need to directly feel engaged and connected with my fan, they are my lifeblood: they make me feel alive»

PERSONALITY
Social addicted
Extroverse
Playful
Unconventional
Friendly
Wrangler
Introspective

Most used social network



Figure 18: TikTok Friendly Artist

If not already done, the first step for the artist is to register and promote her/his new account on other social media platforms, where the artist is sure to benefit from the presence of its consolidated followers.

The overall aim is not limited to the promotion of new songs, but even to make old songs viral again. To do that, the artist creates a challenge using the 30 seconds clip with video content, that can be appreciated by the users and at the same time can trigger their creativity, pushing them to spread more and more the usage of the audio clip.

Hence, the content is a critical factor: it needs to be attractive and funny to encourage users to participate and so, Sony's Marketing function will help the artist with its creation and with the choice of the right hashtag. The challenge may consist in doing a sequence of dance moves, in doing a funny video or in making a TikTok duet with the artist.

Then, the artist will repost the most original and funny videos in order to get closer to the fans and to further stimulate the creation of video contents with her/his track.

The challenge and the users' contents will also be reposted on other social network in order to increase the number of reached people and so, the number of participants. This creates an advantage even for TikTok itself since artists are promoting the platform. Hence, it will be possible to leverage on this value creation in order to renew the agreements with the partner and gather more data for the scouting.

Users loves authenticity, so it is fundamental for the artists to feel free to express themselves on the platform. Being yourself is the best way to increase the followers inside this app and inspire them.

TIK TOK DISTANT ARTIST

For this kind of artists, it is better to leverage on other tools, even if they can open their account on this app, having fun and sharing contents they prefer the most.

The main strategy that can be applied requires the exploitation of influencers' marketing. Sony's marketing experts needs to find out the perfect influencers in the app in order to launch a challenge with the 30 sec. clip. They should be compliant with the personality and audience of the artist, and with the genre and meaning of the song. Once selected, due to the major knowledge of TikTok audience, the influencer will have the freedom to express her/his creativity to get the most from the 30 sec. music clips.

Of course, there could be a problem of costs: famous influencers have the power of asking big compensation to promote the artist. However, the fact that the app is basically new and influencers who are born there are newcomers in the influencer's marketing, their cost is lower respect to other social networks.

4.4 Unique value proposition

"Anticipate competitors in spotting new talents in a dynamic and new environment and in promoting Sony's songs relying on an effective, efficient and differentiated marketing, based on predictive analysis and on a marketing framework that aims to engage the final users"

- **"Anticipate competitors...in a dynamic environment"** The **predictive** logic that lays behind the scouting algorithm will give Sony the chance to "be the first" against competitors. It will avoid the fight for the artist and capture it before others. This is possible even thanks to the solution of the short-term contract that



JAKE LIT

Giacomo Amato

Male

38 years old

Rome (IT)

Rap, Hip-hop

Activity from 2000 - present

6 Album

Best Artist at MTV awards 2009,

4 platinum disks, 3 gold disks,

best Italian Song in 2007

Audience: people from 18 to 35,

against common trends, underground

style

ABOUT

Giacomo was born in the suburban area of Rome from a family of humble origins, he starts singing with some friends in the parks of the city as a way to express himself. Despite the enormous success and the age, his songs are still popular and meaningful also between young listeners because of the authenticity he transmits.

FRUSTRATION

He would like to reach fans on other channels, but his image as non-conformist are not consistent with the personal communication on some social networks.

BELIEFS

«I do not appreciate singers that use their music only as a commercial product, I believe in the pureness of music that comes from the street»

PERSONALITY

Social addicted

Extroverse

Playful

Unconventional

Friendly

Wrangler

Introspective

Most used social network



Figure 19: TikTok Distant Artist

creates a value also for the emerging artist. This will reduce the risk related to have the artist stolen and at the same time reducing the risk of scouting a talent which can reveal itself not to be that viral.

- **"Effective, efficient and differentiated marketing..."** the aim is to give to the artist **the perfect marketing solution**. It is **effective** since the virality of the track can be granted thanks to the marketing algorithm. But it is not just a matter of marketing, but also of her/his personality. Therefore, we offer to the Marketing unit the chance to provide a **tailored** solution to Sony's artists: the values of the artist are the foundation of her/his fame; hence they should be respected.

Moreover, it is possible to take advantage of the great number of users of TikTok and the mechanism that lays behind it to create relations with them and increase the artist's fan base. Thanks to myriad of possible connections among users, the 30 seconds video-song would spread in very little time starting from only one of the them, entering in users' mind due to the viral sound and bringing them on the streaming platform. The proposed solution is also **efficient** since it reduces marketing research's times and costs.

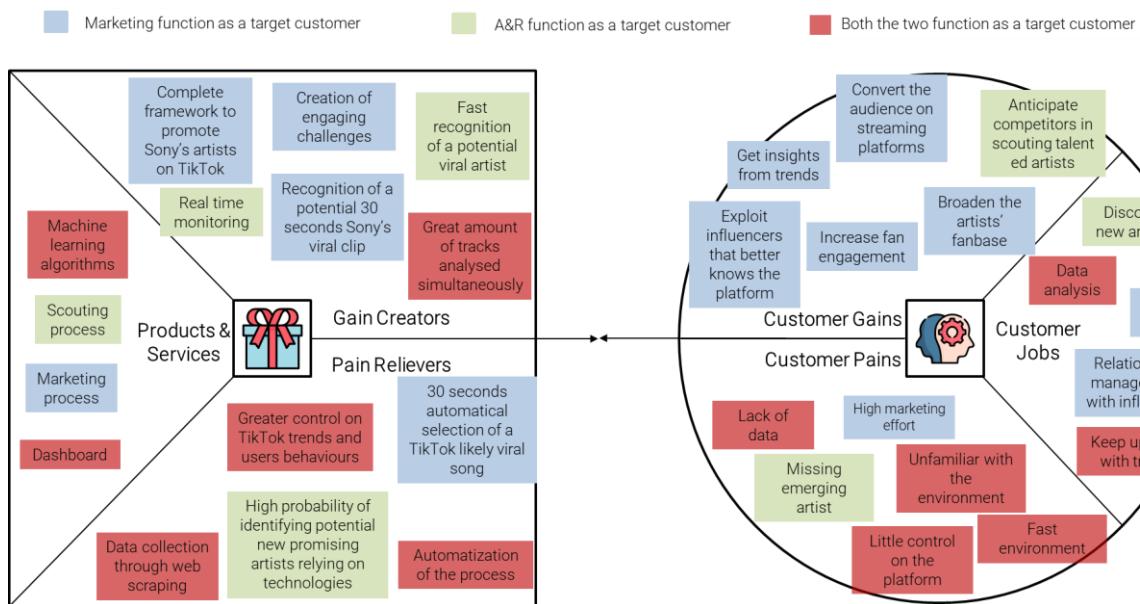


Figure 20: Value Proposition Canvas

4.5 Unfair advantage

Sony can leverage on some strategic advantages thanks to its unique resources. Therefore, the firm will play an "unfair" game against the competitors that do not have the means to respond in the same way and this will bring to the creation of an exclusive and sustainable competitive advantage.

Sony itself states that technology and employees are the two main areas that have to be enhanced in order to create a long-term value. Hence, its advantage is grounded mainly on intellectual and human resources, but also on the physical network of the company.

The first main advantage of Sony is the **intellectual** one. Indeed, the main differentiating factor between the firm and its competitors is the diversity of the business portfolio, that allows to exploit the synergies between different businesses and, in particular, to leverage the owned technologies coming from the electronics business area.

Moreover, Sony continues to invest incrementally for the extension of the content IP, for the development of Artificial Intelligence technologies and, in general, of proprietary technologies. Indeed, in-house development is a key differentiating advantage for the group.

However, the intellectual capital of the company is not only generated internally, but it is also represented by the network of collaborations with local universities and other institutions, for instance, the one with the Carnegie Mellon University (US) for the joint research in growth areas like robotics and AI.

The company can also rely on a solid **human** capital in terms of an extensive expertise in Data Analysis technologies, Marketing and A&R. Sony considers its employees the most important asset and they

are perceived as partners, accountable for the needs of the company and vice versa. The firm defines a strategy to "attract" and hire highly competent personnel, then to "develop" their skills assuring an exchange of knowledge between diverse expertise, and finally to "engage" them through the creation of strong partnerships.

Sony's opened culture towards innovation and the spotting of new technologies and talented employees is also represented by two initiatives: the Sony Startup Acceleration Program (SSAP) and the Sony Innovation Fund (SIF), that invests in startups active in growth areas such as robotics and AI.

Sony's technological advantage is made possible by its **physical** resources like the global network of R&D sites such as the R&D Center Europe Brussels Laboratory that mainly focuses on data analysis and security technologies.

4.6 Channels

Since the solution addresses different customers at different levels, the channels through which the company delivers the value to the final customers will be different for each segment.

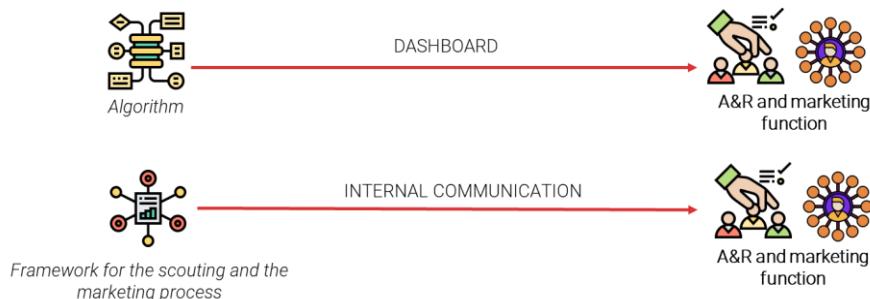


Figure 21: Main Channels

The main channels are the ones related to the direct customer.

The **Sonik Dashboard** allows the A&R and the Marketing unit to interact with the related algorithm.

Instead, the guidelines for the scouting of emerging artists and for the support in the marketing process for, are communicated respectively to the A&R and to the Marketing functions **internally**.

The other secondary channels deliver the value to the indirect customers.

The scouting process includes the signature of the short-term contract, that is proposed to the emerging artist by means of Sony's **TikTok** account.

According to the specific marketing framework assigned to Sony's artist, there are two different ways to reach the final client (the TikTok user): in the first case, the marketing framework and the content to publish on the platform are agreed between the artist and the Marketing function and according to it, the artist will spread the song on the app. In the second case, the marketing framework is in the hands of influencers, that will advertise the artist' song on the Tiktok platform.

Figure 22: Secondary Channels

Since the direct customers are internal functions of Sony, it is difficult to deeply monitor their "engagement" with Sonik Dashboard. Something resembling to the idea of engagement could be related to the effectiveness of the algorithm: if it works, of course the user will be satisfied and brought to an increased use of the "new way" of scouting and promoting on TikTok. However, it is important to consider also the indirect customers, that are fundamental as well and determine the effectiveness of the solution itself.

4.7 Key Metrics

The overall aim of metrics is to monitor the **engagement** of all the clients, the **effectiveness** and **efficiency** of the proposed solution. The KPI were defined only consequently to the identification of the general goals.

Since the direct customers are internal functions of Sony, it is difficult to deeply monitor their "engagement" with Sonik Dashboard. Something resembling to the idea of engagement could be related to the effectiveness of the algorithm: if it works, of course the user will be satisfied and brought to an increased use of the "new way" of scouting and promoting on TikTok. However, it is important to consider also the indirect customers, that are fundamental as well and determine the effectiveness of the solution itself.

Legend:

User engagement

Effectiveness of the solution

Efficiency of the solution

SONIK

KPI	GOAL	DESCRIPTION
#daily access to Dashboard	Dashboard interaction	How much the direct customers are engaged and rely on this tool, how many times they open and consult the Dashboard
#correct results in testing phase of the algorithms #total results in testing phase of the algorithms	Effectiveness of the algorithms	The number of correct recognitions of viral songs respect to the total results in the testing phase measures the effectiveness of the algorithms in terms of predictions accuracy.
#clips picked through Sonik that are inside the monthly list of Sony's trending songs #total clips picked through Sonik	Effectiveness of the scouting algorithm Effectiveness of the combination of the marketing algorithm and the marketing framework	Sony receives monthly a list of Sony's trending songs by TikTok. These data can be useful in order to measure the solution's performances.
#clips picked through Sonik that are inside the monthly list of Sony's trending songs #total clips picked through Sonik #clips autonomously picked that are inside the monthly list of Sony's trending songs #total clips autonomously picked through Sonik	Effectiveness of the combination of the marketing algorithm and the marketing framework compared to the one of traditional methods	Additionally to the absolute effectiveness, it is valuable also to compute differential performances respect to the ones obtained using traditional methods and so, picking a Sony's clip to use on TikTok without using the algorithm's suggestions.
#views, #likes, #shares on other platforms, #videos done, growth of followers on TikTok or other platforms	Effectiveness of the scouting algorithm Effectiveness of the combination of the marketing algorithm and the marketing framework Engagement of the TikTok user	Another confirmation of the goodness of predictions can come from the performance of a track or of an artist inside the TikTok application. For the marketing process these indicators are referred to the performance linked to the challenges launched, hence they measure the effectiveness of the combination of the marketing algorithm and of the marketing framework.
#long-term contracts #accepted short-term contracts	Effectiveness of the scouting algorithm	The more the ratio is equal to 1, it means that a high number of short-term contracts turned in long-term ones and confirmed the goodness of the prediction.
#short-term contracts accepted #short-term contracts proposed	Engagement of the emerging artist	This indicator shows the attractiveness of the short-term contract given to the emerging artist.
#number of accepted marketing proposal #total marketing proposal	Engagement of Sony's artist	It shows the attractiveness of the proposed marketing solution for the artists and the willingness to adopt it.
Conversion rate on streaming platforms	Effectiveness of the combination of the marketing algorithm and the marketing framework Engagement of the TikTok user	Increase of streaming listening after adopting the marketing solution.
#days to define a marketing strategy on TikTok with the solution - #days to define a marketing strategy on TikTok traditionally #human resources to define a marketing strategy on TikTok with the solution - #human resources to define a marketing strategy on TikTok traditionally	Efficiency of the marketing algorithm respect to traditional methods	Since the marketing process becomes more structured and automatized, the time needed to run marketing research is expected to be lower as well as the human resources that usually run the analysis

4.8 Investment Analysis

The economic analysis was made considering only the Italian market of streaming music, since the solution will be used by Sony Music Italy and, in case of scalability, the model will be applied to other countries. The period analysed is 4 years, since the environment is highly dynamic. The decision made was to consider always the worst case, since it is better to underestimate the benefits of the solution rather than the opposite. All the costs and revenues were expressed or converted in US dollars to standardize data gathered in different currencies.

Revenues

The main stream of revenue is not directly linked to the TikTok platform, but to streaming platforms. Therefore, the basic idea is that of using the app as a driver. The higher the number of future trending scouted artists, the higher Sony's listening statistics on streaming platforms. On the other hand, the more the marketing process will be effective, the more people will be reached and converted to Sony's songs.

The monetary benefits were not easily quantifiable since they depend on two highly volatile factors: the number of listens on streaming platforms and the number of people reached through the strategy. However, a possible scenario has been identified by making assumptions in order to give a general idea about the impacts. The numerical assumptions have been written in red.

It was necessary to compute the potential market that could be theoretically reached. At the beginning of 2020, in Italy **6.4 million** of people used the TikTok app with a growth of +202% in only three months. To forecast the growth of this number, the worst scenario, in which TikTok already reached its peak, was taken into account; the growth was considered as slowing down and then rapidly decreasing, resembling the **Big crunch phase** of the Big Bang disruption lifecycle.¹

$$\text{TikTok users} = \text{TikTok users}_{t-1} * (1 + \text{Assumed growth}_t)$$

According to the survey run during the project work, it emerged that within the sample of TikTok users, 63% have a paid subscription to a music streaming platform and, among them, 63% usually add songs discovered on TikTok to their playlists. These percentages were assumed to remain constant.

$$\text{Potential market} = \text{TikTok users} * 0,63 * 0,63$$

	2020	2021	2022	2023	2024
¹ Assumed TikTok users' growth		30%	20%	-20%	-50%
TikTok users	6.4	8.32	9.98	7.99	3.99
Potential market		3.3	3.96	3.17	1.59

The ideal situation should be the one in which all the potential consumers are reached and converted to streaming, but the reality is different. The hypothesis made is that the **percentage of users converted** will decelerate year after year according to imitating factors or to innovative solutions implemented by the competitors: the first year it was assessed at 50%, then 30% and finally at 20%.

The remuneration model for music labels from subscriptions to streaming platforms, is shifting from a pro-rata one to a user-centric: the subscription fee of the single user will be distributed, according to the overall number of listening to of the label artists songs. Hence, if the market share of Sony, in the paid streaming market, is 20% (based on the market share of Sony in the global market), this means that on average the **percentage of Sony's songs listened by paid streaming users** is around 20%. With our solution the user will experience an increase of this percentage, namely the number of Sony's songs listened will boost. Only for the portion of users converted, this percentage is considered to decelerate for the same reasons as before: 50%, 30%, 25%, according to the years. For the streaming users "not converted", this percentage it is assumed to remain 20%.

The **new market share** of Sony music Italy, calculated with respect to the paid streaming market, was computed according to a weighted average between the percentage of Sony's songs listened by the users reached and by the other individuals with a paid subscription:

$$\# \text{TikTok users with a paid subscription converted}^1 = \text{Potential market} * \% \text{users converted}^2$$



$$\text{new market share} = \frac{\# \text{TikTok users with a paid subscription converted}}{\# \text{italians with a paid subscription}} \times \% \text{Sony listening of users converted}^3 + \\ + \frac{(\# \text{italians with a paid subscription} - \# \text{TikTok users with a paid subscription converted})}{\# \text{italians with a paid subscription}} \times 20\%$$

Year	² %users converted	#converted users	#italians with a paid subscription*	³ % Sony listening of users converted	#other individuals with a subscription	New market share	Differential market share
2021	50%	1,651,104	11,100,000	50%	9,448,800	24.46%	4.46%
2022	30%	1,188,794	11,300,000	30%	10,111,200	21.05%	1.05%
2023	30%	951,036	11,400,000	30%	10,448,900	20.83%	0.83%
2024	20%	317,012	11,400,000	25%	11,082,900	20.14%	0.14%

*The number of Italian people with a subscription to a streaming music platform was taken from a forecast of Statista.

Finally, the **differential revenues** (in mln) were calculated as the product between the differential market share and the forecasted Italian paid streaming revenues:

	2017	2018	2019	2020	2021	2022	2023	2024
Global recorded music revenues	\$17,800	\$18,700	\$20,200	\$21,210	\$22,058	\$22,720	\$23,402	\$23,870
Growth	12.66%	5.06%	8.02%	5%	4%	3%	3%	2%
Italian recorded music revenues	\$251.64	\$258.06	\$278.93	\$292.88	\$304.59	\$313.73	\$323.14	\$329.60
Italian recorded music revenues Global recorded music revenues	1.41%	1.38%	1.38%	1.38%	1.38%	1.38%	1.38%	1.38%
Italian paid streaming revenues	\$42.87	\$66.64	\$86.63	\$103.96	\$114.354	\$125.79	\$132.079	\$138.68
Growth	2%	55%	30%	20%	10%	10%	5%	5%
Italian paid streaming revenues Italian recorded music revenues	17%	26%	31%	35%	38%	40%	41%	42%
Sony music Italy paid streaming revenues	\$8.57	\$13.33	\$17.33	\$20.79	\$22.87	\$25.16	\$26.42	\$27.74
Differential revenues					\$5.1	\$1.32	\$1.10	\$0.19

- **Global recorded music revenues:** data until 2019 were taken from the latest GMR (Global Music Report) of the International Federation of the Phonographic Industry. The following revenues were forecasted through a hypothetic growth percentage, assumed looking at the overall growing trend, but still considering the maturity phase of the market globally.
- **Italian recorded music revenues:** data until 2019 were taken from FIMI website. The following revenues were forecasted considering that the weight of Italian revenues on the global will remain constant.
- **Italian paid streaming revenues:** data until 2018 were taken from FIMI website. The following revenues were forecasted through a hypothetic growth percentage, assumed considering that the Italian market is still in a growing phase. However, putting in the worst case, the growth was assumed to slow down until the percentage of paid streaming revenues on total revenues will reach the 42%, that is a similar proportion looking at global numbers.
- **Sony music Italy paid streaming revenues:** computed as the product between Sony's assumed market share of 20% and the Italian paid streaming revenues.

Costs

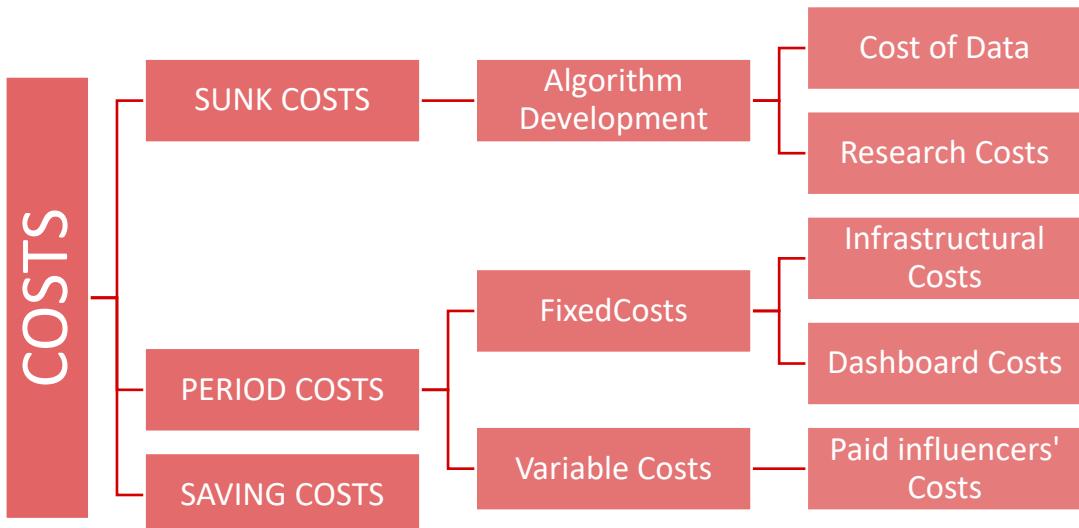


Figure 23: Costs Structure

Development costs

Disclaimer: In order to estimate some of the cost, a yearly compensation of \$60,000 has been assumed for every Machine Learning engineer: according to *Glassdoor* this is the average salary for ML engineers in Europe.

The cost of the Data

Data is fuel for any machine learning project. Using a service like [Amazon's Mechanical Turk](#) to crowd source the whole process, it would cost around **\$70,000** to generate a 100k samples dataset. This cost defines the quantity.

To compute the quality, most companies have one or two people in charge of validating and cleaning the data samples and the labelling. Having a small in-house validation team can add to the initial cost of outsourcing of the 100k data samples about **\$2,500** or more.

Hence, a solid dataset will cost around **\$72,500**, considering that the nature of data and the complexity of annotations is low.

Research Cost

The research is related to the initial feasibility study, the algorithm search and the experimentation phase. The Dimensional Research report states that most enterprise AI teams have 5 members on average. Out of these 5, 3 may be outsourced (either services or freelancers). The team, in this configuration, can probably work on 2 projects in parallel and they can conduct the research for a project in one to two months.

So, the cost for 2 employees ($2 \times \$5,000$) and 3 freelancers ($3 \times \$3,000$) results in **\$19,000 per month**. If the team can handle two projects at the same time and the length of research is 1.5 months, it means that the cost of this phase is about **\$28,500** for the two projects.

Infrastructural costs

Cloud VM		Flat File Storage	
CPU's	4 vCPU	Storage Capacity	30GBs
RAM	16 GBs	Transfer Inbound	1 TB
Storage	30 GBs	Transfer Outbound	1 TB

Price	244.63 \$/month	Price	92.07 \$/month
Database		Dashboard Infrastructure	
Engine	MySQL	Power Pro License	20
Storage	100 GBs	Cost Per Licence	9.99 \$/month
Price	171.44 \$/month	Price	200 \$/month
Total Cost		718.14 \$/month	

Dashboard costs

Microsoft's Power BI, part of its Office 365 cloud suite, ranges from \$10 to \$52 a month per user. Considering the maximum price and assuming that Sony Music Italy has 100 employees and only 30% of them use dashboard, the cost will be **\$1,560** per month.

Paid influencers costs

According to ONIM (Osservatorio Nazionale Influencer Marketing), the average remuneration of an influencer in Italy ranges from 300 to 600 €/post. Converting the cost into dollars, considering the maximum price and assuming to launch around 50 challenges per month, whose 20 are launched by influencers, the total cost is around **\$12,760**.

Saving costs

Thanks to the solution, is also possible to save costs related to marketing effort in terms of employees and expenses for marketing research. However, these costs were difficult to quantify and considering the worst case, these costs were not taken into account during the analysis.

NPV

	Now	2021	2022	2023	2024
Investment cost	\$ 101,000				
Differential revenues		\$ 5,100,000	\$ 1,320,000	\$ 1,100,000	\$ 190,000
Differential costs		\$ 180,457.68	\$ 180,457.68	\$ 180,457.68	\$ 180,457.68
Actualized cash flows		\$ 4,280,466	\$ 864,188	\$ 605,837	\$ 7,701
Cumulative NPV		\$ 4,179,466	\$ 5,043,653	\$ 5,649,490	\$ 5,656,562
		€ 3,694,669	€ 4,458,614	€ 4,994,117	€ 5,000,429

A discounted rate of 15% was considered, since the risk associated to the success of artists is high and, with the solution, it is even higher due to the uncertainty of predictions. Hence the rate was increased respect to the 10% of the industry.

Despite taking always in consideration the worst situation, the final NPV, converted to euros, is about **5 million** and due to the low investment costs, the payback time is equal to **1 year** because the initial costs are recovered during the first year of implementation. However, the benefits will be even higher taking into account the saving costs and the fact that probably TikTok success is going to last for a long time, becoming a new alternative to the already present social networks.

5 Validation

This is a fundamental part of the project in order to verify the main pillars of the solution. With this purpose, the business model was drafted to better recognize the different hypotheses to validate, considering three different areas related to the **feasibility**, **desirability** and **viability** of the idea.

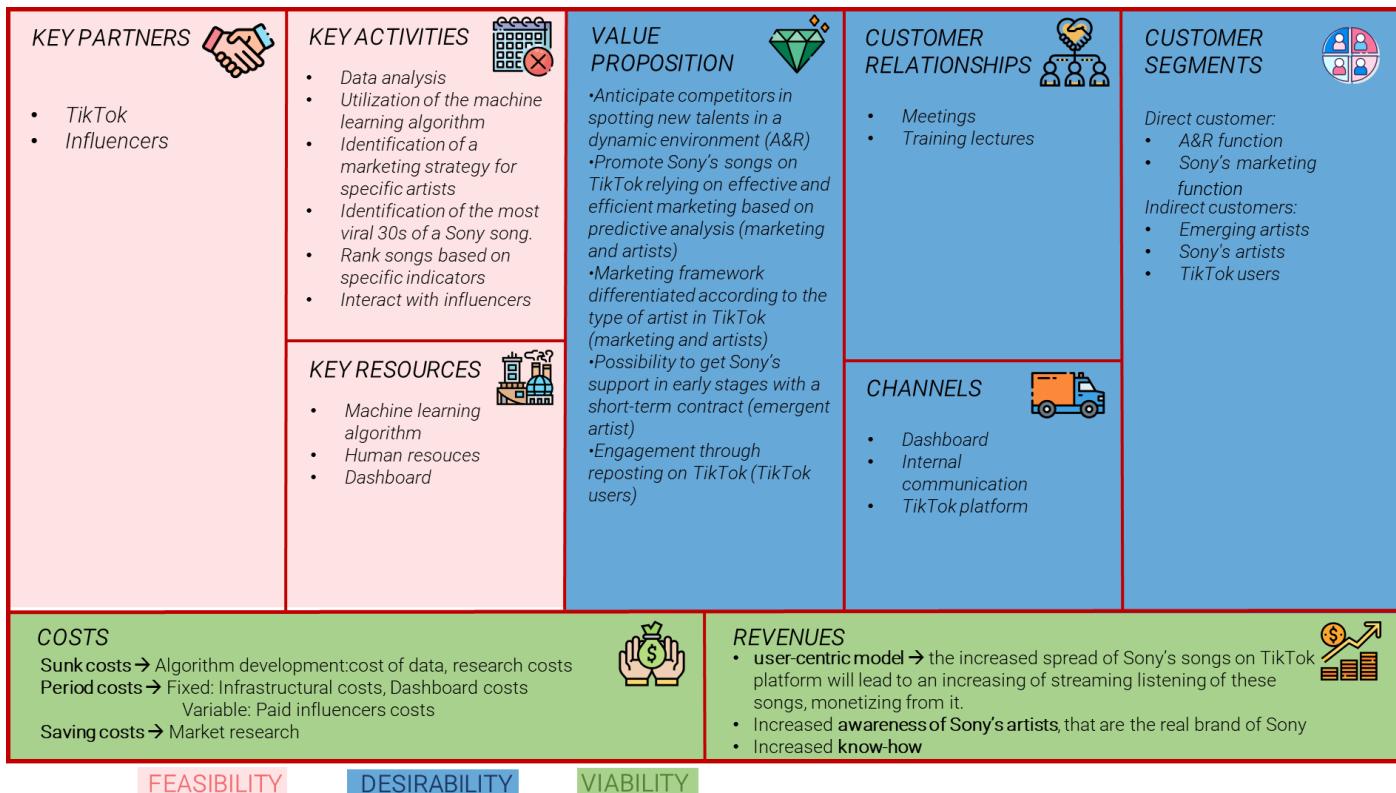


Figure 24: Business Model Canvas

After the definition of all the blocks, it was possible to clearly identify all the assumptions made through the solution. However, the only three assumptions that could be tested during the project were the 4th, the 6th and the 8th, respectively related both to the desirability, the viability and the feasibility areas. Even the last two areas were considered since they lead to the desirability of the direct customers.

Survey

In order to validate these hypotheses, a survey was run within a sample of 167 TikTok users and the Test Card tool was used to guide the process.

(Survey questions link: <https://www.shorturl.at/deCX2>; answers link: <https://www.shorturl.at/fitPW>.)

A survey has been drafted with the aim to analyze the behavior of users on the platform. So, using social-media-means like Whatsapp, Instagram and Facebook we delivered the survey asking to spread it in turn. The principle of "word of mouth" was used to collect as many answers as possible.

The survey has been sent with the clear objective of analyzing **TikTok users**, therefore only the answers of

1. Sony's A&R and marketing functions will be interested in implementing the solution.
2. Sony's artists will be willing to follow the marketing strategy.
3. Emergent artist will accept the short-term contract to be locked in with the company for a short period.
4. TikTok users will actively participate to challenges proposed by artists, spreading Sony's songs on TikTok.
5. Choosing the 30s track given by the algorithm will increase the view counts of the TikTok videos where the track has been used.
6. TikTok platform helps in spreading tracks and in boosting the listening on streaming platforms.
7. The scouting algorithm will predict the virality of a new song on TikTok with a good accuracy.
8. The marketing algorithm can extract the 30 most probable viral seconds from an existing song with a good accuracy.
9. The database and the web scraping will be able to support the collection of a great amount of data.

Figure 25 List of assumptions

respondents that have the app installed on their smartphone have been analysed. As it is shown by the graph, the respondents are mainly girls with an age that ranges between 13 and 23 years old. The majority of them is concentrated in a range going from 17 and 19 years old. So, the reference is to a very young population. Demographically, the sample is very near to the real segment of TikTok users.

Age	Respondent
8-12	1
13-17	65
18-22	81
23-25	19
25+	1

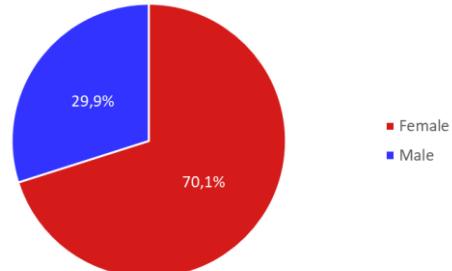


Figure 26: Gender Distribution

First Hypothesis: Desirability

The hypothesis has been considered **medium critical**, since it was hard to build an MVP able to reproduce a situation in which artists launch their own challenges. Hence, the testing consists in understanding customers' habits directly asking to them, but the answers could be partially biased because of the distance from a real situation.

For the same reason the **reliability of data** is **medium**. On the other hand, the **test costs** are **null**.

The **time required** is **medium**, considering the time to build the survey, to distribute it, to the reach a significant number of respondents and to analyse the answers.

A good percentage of people among TikTok users that will take part to Sony challenges was considered around **30%**.

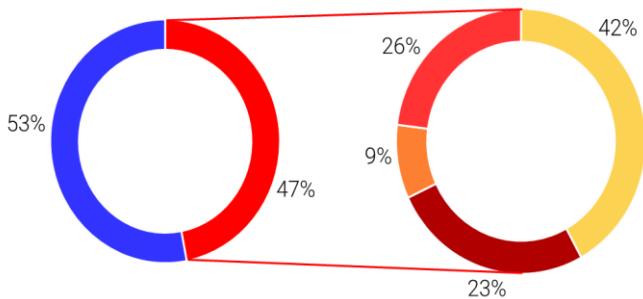


Figure 28: Survey's Results

Test card	
STEP 1: HYPOTHESIS <i>We believe that</i>	TikTok users will actively participate to challenges proposed by artists, spreading Sony songs on TikTok. Critical
STEP 2: TEST <i>To verify that, we will</i>	Run a survey in order to better understand the habits and the preferences of the users on the TikTok platform. Data Test Costs Reliability
STEP 3: METRIC <i>And measure</i>	The percentage of respondents that usually do challenges or create own videos on TikTok; the percentage of respondents that would take part to artist's challenges. Time Required
STEP 4: CRITERIA <i>We are right if</i>	The percentage of respondents that would take part to a challenge of an artist is higher than 30%.

Figure 27: First Test Card

What was observed is that **47%** of respondents would take part to artists' challenges, almost the half of the sample. The interesting thing is that among them, 23% is not used neither to do challenges nor to publish contents on TikTok.

This suggests that maybe they will start taking challenges or publishing contents only because of the initiative proposed by their favourite artists. We can also consider that the percentage observed may increase with the incentive of the reposting of the content by the artist.

Second Hypothesis: Viability

The level of **criticality** is **high**, since there is not the absolute certainty that TikTok drives part of the music trends and not the opposite, so that a track is present on TikTok because it is trending. However, during the testing, some interesting insights emerged.

The **reliability of data** is **high** because the survey question taken into consideration is related to a past behaviour of users, so it cannot be biased, and data were gathered directly by streaming platforms and TikTok playlists.

The **test costs** are **null** and the **time required** for the test is **medium**, considering times related to the survey and to the collection of data from streaming platforms and their analysis.

If the majority of TikTok users add TikTok songs to their playlists and the majority of TikTok trends are also trends on Spotify, the hypothesis is validated.

It was observed that the percentage of users that discover new songs on TikTok is the **83%**, while the percentage that add TikTok songs to their playlist is the **89%**.

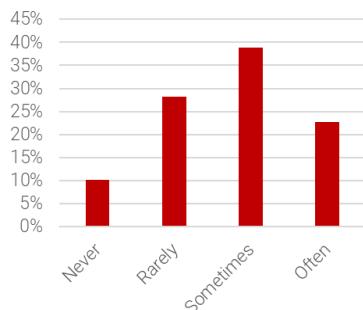


Figure 30: Percentage of people that add TikTok songs to their playlist with the relative frequencies

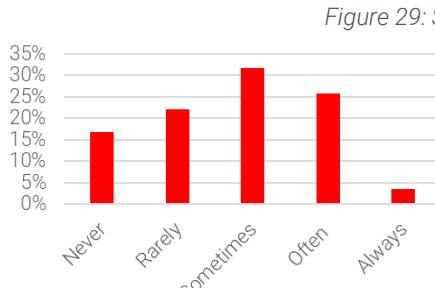


Figure 31: Percentage of people that discover new songs on TikTok with the relative frequencies

Furthermore, 97% of TikTok users use YouTube and 85% use Spotify, 54% with a payment subscription. Hence, being Spotify the streaming platform mostly used with a paid subscription, it has been considered for the next phase of the testing process.



Figure 33: Percentage of TikTok songs in Spotify Playlist

The percentage of songs that were present both in the playlists of Spotify 50 Top Global and 50 Viral Global and in the trends of TikTok was the **77%**; the percentage of songs that were present both in the playlists of Spotify 50 Top Italy and 50 Viral Italy and in the trends of TikTok was the **49%**. Hence, the hypothesis can be considered validated globally and even locally since it is really close to the majority.

Test card

STEP 1: HYPOTHESIS
We believe that
 TikTok platform helps in spreading tracks and in boosting the listening on streaming platforms.

Critical ⚠️ ⚠️ ⚠️

STEP 2: TEST
To verify that, we will
 Run a survey in order to better understand the link between the TikTok platform and streaming ones. Compare the chart of TikTok viral songs with the one of Spotify and YouTube, that emerged to be the most used streaming platforms.

Test Costs
Data
Reliability

coins graph checkmark

STEP 3: METRIC
And measure
 The percentage of people that usually discover new songs on TikTok and TikTok songs to their playlist; percentage trending songs on the most used streaming platforms that are also trending on TikTok.

Time Required
⌚ ⌚ ⌚

STEP 4: CRITERIA
We are right if
 The percentage of people that usually discover and/or add TikTok songs to their playlist is higher than 50%; the percentage of trending songs in the most used streaming platforms, that are also trending on TikTok, is higher than 50%.

Figure 29: Second Test Card

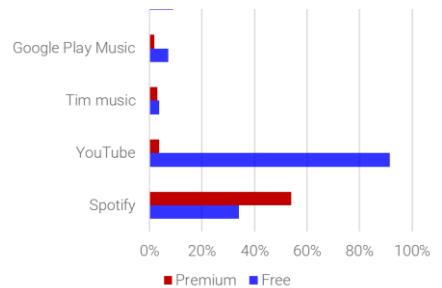


Figure 32: Most used streaming platforms

In order to overcome the criticalities mentioned first, some considerations have been made during the testing:

- The **results of the survey** suggest that TikTok is actually driving people to streaming platforms.
- **Correlated searches:** it emerged that searching on Google one of the songs trending both on TikTok and on Spotify, it suggested correlated search that “many users” made before us and the searches were about TikTok songs not present in Spotify trends. This may mean that many songs present in Spotify trends are mainly searched by TikTok users.
- **Comments on YouTube:** looking at comments on YouTube under songs trending both on TikTok and on Spotify, it is possible to notice that they are all about people that are “there thanks to TikTok”.

Third Hypothesis: Feasibility

The level of **criticality** and the **time required** are **high** due to the complexity of the testing procedure that consists in training and testing one of the two algorithms. The **reliability of data** is **high** and the **test costs** were **null**.

The validation process was composed by five steps:

1. Firstly, the “successful” TikTok hits to analyse were listed, taken from the list of the most popular Sony’s songs on TikTok provided by Sony.
2. Another list was prepared of “unsuccessful” Sony’s songs that did not make it to the top TikTok charts.
3. After obtaining these lists of songs an ETL job makes use of the Spotify API to fetch characteristics of all these songs as well as the 30 seconds sections of these songs and cleanses the data to extract correct Features.
4. Cleansed data is then split into two sets of Train data and Test data.
 - a. Machine Learning model is learned using the Training data.
 - b. After training Cross-Validation Technique is used to test the model using the Test data.
5. Having obtained the model, prediction score of the 30 seconds sections can be obtained which lead to the effectiveness of the decision making.

The code can be checked at the Github repository [here](#). The list of hit songs used has been provided to us by Sony and the relevant features have been fetched from [Spotify API](#). The project was built in Python using Jupyter Notebooks.

After collecting the two lists, the first step was to fetch characteristics through an ETL (Extract, Transform, Load) job.

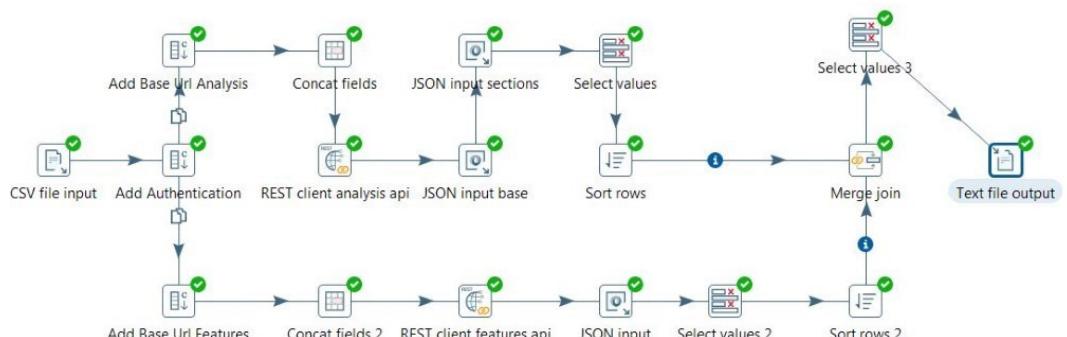


Figure 35: ETL Jobs for Data Preparation in Pentaho



Figure 34: Third Test Card

A total of 1,275 records were obtained: 1,000 records are potential hits (Boolean variable 1), while 275 are not (Boolean variable 0). Therefore, the data set is highly skewed, so there was the need to stratify the train test split. The test and train data set were created with train data set being 70% of the whole data set and test data set being 30%.

Tree based methods are supervised learning methods that empower predictive models with high accuracy, stability and ease of interpretation. In this step, Decision Tree was used as our Machine Learning model, in order to predict our classes.

Along with decision Tree, also Random Forest Classifier was applied. However, the prediction of Decision Tree was superior.

Decision Tree Accuracy on test data set with tree depth =3 (underfit)	81.2%
Decision Tree Accuracy on test data set with tree depth =4 (optimum)	83.5%
Decision Tree Accuracy on test data set with tree depth =5(overfit)	82.0%

Random Forest Classifier on test data set with tree depth =3	78.4%
Random Forest Classifier on test data set with tree depth =4	78.8%
Random Forest Classifier on test data set with tree depth =5	80.8%

It was observed that with current data and using Random forest ensemble methods, an accuracy between 80-85% can be achieved based on the tree depth. In the future, when the data available will be more, the accuracy can even improve using other ML techniques like Neural Networks.

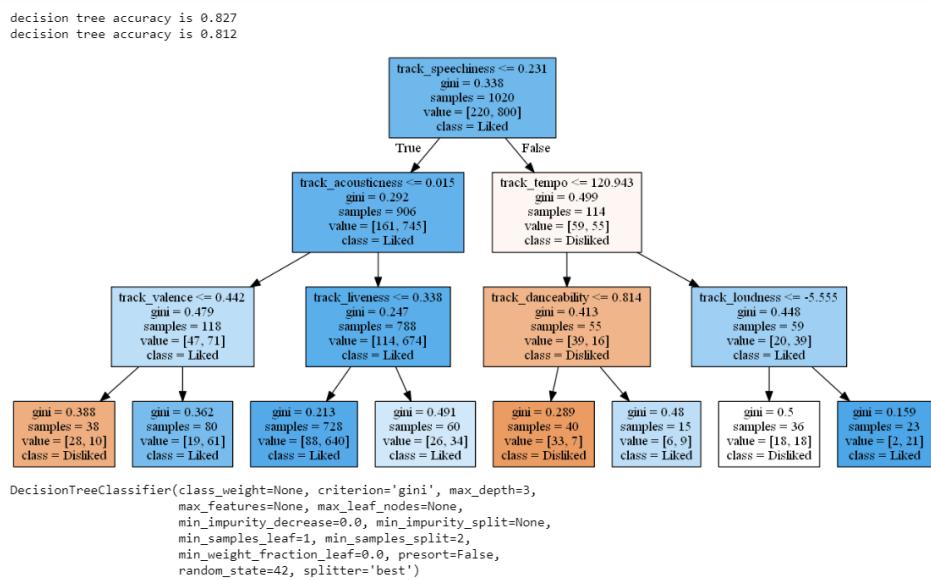


Figure 36: Decision Tree Model with Depth

Feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node probability can be calculated by the number of samples that reach the node, divided by the total number of samples. The higher the value the more important the feature. For each decision tree, Scikit-learn calculates a nodes importance using Gini Importance, assuming only two child nodes. Here using Decision Tree, we have predicted the relative importance's of all features present in the data.

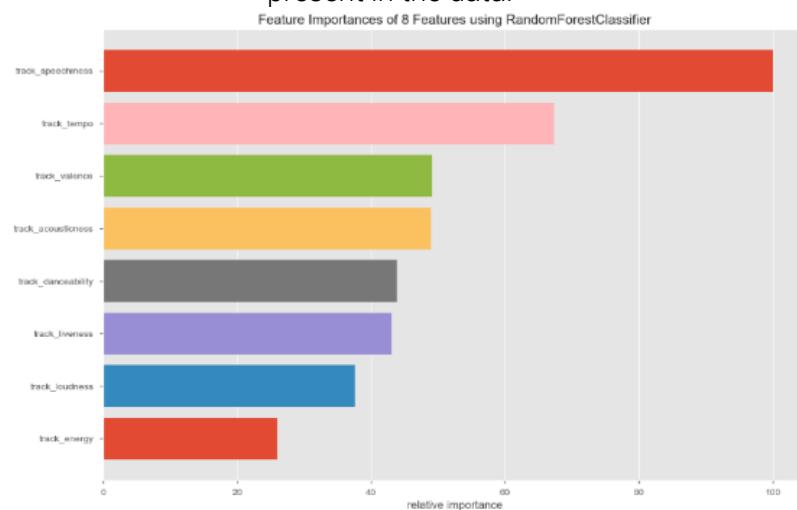


Figure 37: Feature Importance based on the Model

6 Team Composition



Bartesaghi Andrea
928656



Bersani Federico
926343



Gupta Pragati
938654



Khan Mohsin
943115



Palombini Elena
945573



Vigoni Silvia Francesca
921181



Vitali Camilla
920935

Bibliography

- <https://eur-lex.europa.eu/eli/dir/2019/790/oj>
- <https://eur-lex.europa.eu/eli/reg/2016/679/oj>
- <https://www.ifpi.org/recording-industry-in-numbers.php>
- <https://www.statista.com/statistics/457535/digital-music-revenue-in-italy-forecast/>
- <https://www.fim-musicians.org/streaming-pro-rata-vs-user-centric-distribution-models/>
- <https://www.businessofapps.com/data/tik-tok-statistics/>
- <https://sensortower.com/blog/tiktok-downloads-2-billion>
- <https://www.statista.com/statistics/422926/record-companies-market-share-worldwide-physical-digital-revenues/>
- <https://www.fimi.it/mercato-musicale/dati-di-mercato/>
- https://www.sony.net/SonyInfo/IR/library/corporatereport/CorporateReport2019_E.pdf
- <https://musically.com/2019/05/09/user-centric-streaming-payouts-apple/>
- <https://www.statista.com/statistics/457538/digital-music-users-in-italy-forecast/>
- <https://it.semrush.com/blog/report-influencer-marketing-italia-2019/#:~:text=Valutando%20i%20costi%20canale%20per,300%E2%82%AC%20e%20i%20600%E2%82%AC.>
- <https://www.tiktok.com/trending?lang=it>

ANNEX

Index

1. Second hypothesis validation on Spotify	32
2. Extended version of the third hypothesis validation	34
3. Implementation idea of the scouting algorithm	41



1. Second hypothesis validation on Spotify

TikTok trending songs present in Spotify top or viral playlist.

	<i>50 viral global</i>	<i>50 top global</i>	<i>50 viral Italy</i>	<i>50 top Italy</i>	<i>Publication</i>
Laxed (Siren Beat) - Jawsh 685	✓		✓		2020
Banana (feat. Shaggy) - Conkarah	✓		✓		2019
Il bacio di Klimt – Emanuele Aloia	✓		✓	✓	2020
GOOBA – 6ix9ine		✓		✓	2020
Fiori di Chernobyl – Mr.Rain			✓	✓	2020
Il Passo – MamboLosco			✓		2020
Auto Blu – Shiva & Eiffel 65			✓	✓	2020
Problemi Con Tutti (Giuda) – Fedez				✓	2020
Missili – Frah Quintale & Giorgio ...			✓	✓	2018
Bando – Anna				✓	2020
Nena – Boro Boro, Geolier			✓	✓	2020
Auto Blu – Sofia Del Baldo			✓	✓	2020
Party Girl – StaySolidRocky	✓		✓		2020
I like him – Princess Nokia	✓		✓		2020
Supalonely – BENEE		✓			2019
Play Date – Melanie Martinez	✓	✓	✓		2015
Surrender – Natalie Taylor	✓				2015
Stelle – Ras&Calle			✓		2020
Toosie Slide – Drake		✓		✓	2020
Amor de colegio – Luxian			✓		2019
Roses – Imanbek Remix		✓		✓	2016
Blinding Lights – The Weeknd		✓		✓	2019
Vuoto – Liner, Biondo			✓		2020
Blun7 a swishland – the Supreme				✓	2019
Goosebumps – Travis Scott		✓		✓	2016
MAMACITA – Black Eyed Peas, Ozuna, J. Rey Souls		✓	✓	✓	2020
Chega – Gaia				✓	2020
Illy – Surf Mesa, Emilee	✓	✓	✓		2019
Per sentirmi vivo – Fasma, GG				✓	2019
Testa fra le nuvole – Alfa				✓	2019
Cacao (feat. Pyrex) – Ghali			✓	✓	2020
Sexy – JoeVille	✓				2020
Boss Bitch – Doja Cat		✓		✓	2020

	<i>50 viral global</i>	<i>50 top global</i>	<i>50 viral Italy</i>	<i>50 top Italy</i>	<i>Publication</i>
ROCKSTAR – DaBaby, Roddy Ricch	✓	✓	✓	✓	2020
Lemons – Brye	✓				2020
Death bed (coffee for your head) – Powfu, beabadoobee	✓	✓	✓	✓	2020
THE SCOTTS - THE SCOTTS, Travis Scott, Kid Cudi	✓	✓	✓	✓	2020
Stuck with U – Ariana Grande, Justin Bieber		✓			2020
Don't Start now – Dua Lipa		✓			2020
Dance monkey – Tones and I		✓		✓	2019
Blueberry Faygo – Lil Mosey		✓			2019
The box – Roddy Ricch		✓			2019
Break my heart – Dua Lipa		✓			2020
Intentions – Justin Bieber, Quavo		✓			2020
Sunday Best – Surfaces		✓			2019
Falling – Trevor Daniel		✓			2018
Say So – Doja Cat	✓	✓			2019
Safaera – Bad Bunny, Jowell & Randy, Nengo Flow		✓			2020
Savage Remix – Megan Thee Stallion, Beyoncé	✓	✓			2020
Life is Good – Future, Drake		✓			2020
After Party – Don Toliver	✓	✓			2020
ROXANNE – Arizona Zervas		✓			2019
Be Kind – Marshmello, Halsey	✓	✓			2020
Tusa – KAROL G, Nicki Minaj		✓		✓	2019
SICKO MODE – Travis Scott		✓		✓	2018
Adore You – Harry Styles		✓			2019
WHATS POPPIN – Jack Harlow		✓			2020
In Your Eyes – The Weeknd	✓	✓			2020
HIGHEST IN THE ROOM – Travis Scott		✓			2019
Before you go – Lewis Capaldi		✓			2019
Memories – Maroon 5		✓			2019
Rojo – J Balvin		✓			2020
Breaking me – Topic, A7S		✓	✓		2019
Bad guy – Billie Eilish		✓			2019
Senorita – Shawn Mendes, Camila Cabello		✓			2019
If the world was ending – JP Saxe, Julia Michaels		✓			2019



2. Extended version of the third hypothesis validation

Implementation Overview:

1. Firstly, the “successful” TikTok hits to analyse were listed, taken from the list of the most popular Sony’s songs on TikTok provided by Sony.
2. Another list is prepared of other Sony’s songs that didn’t make it to the top charts. This is the list of the unsuccessful songs on TikTok.
3. After obtaining this list of songs an **ETL job** makes use of the Spotify API’s to fetch characteristics of all these songs as well as the sections of these songs and cleanses the data to extract correct **Features**.
4. Cleansed data is then split into two sets of Train data and Test data.
 - a. Machine Learning model is learned using the Training data.
 - b. After training Cross-Validation Technique is used to test the model using the Test data.
5. Having obtained the model, Prediction score of section can be obtained which lead to the effectiveness of the decision making.

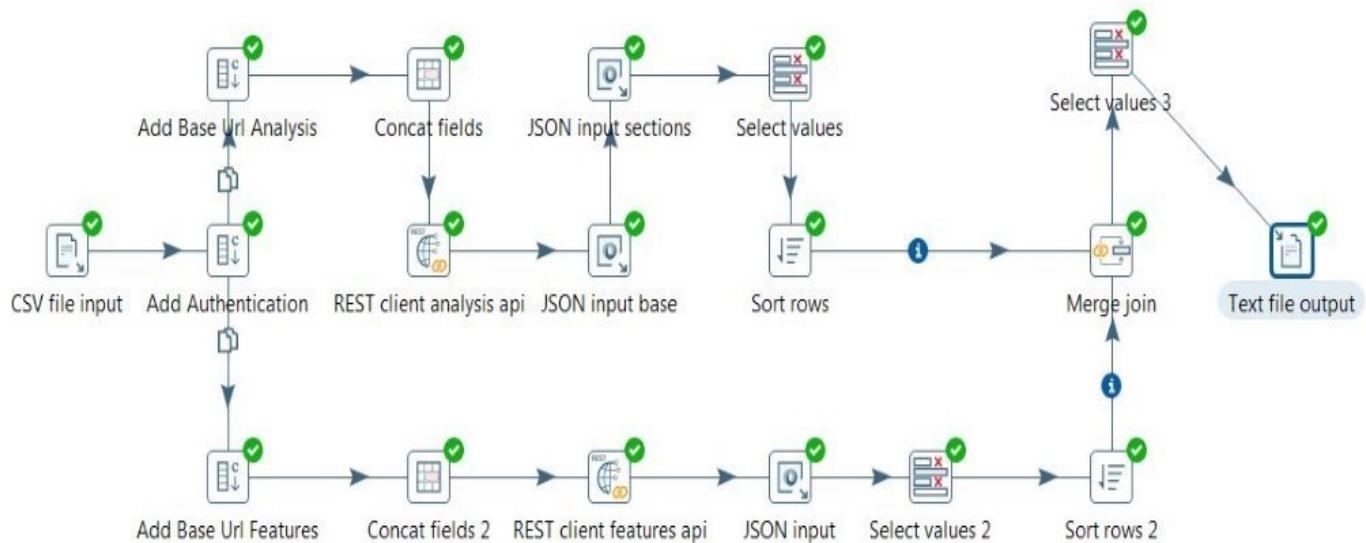
ETL Job:

ETL is referred to as Extract, Transform and Load, where the Extract part involves getting data from varies heterogenous sources; Transform entails the modifying of data in a form that could be fed to further processes; finally Loading this data to a consolidated space and format.

In this project context, process pipeline uses the Pentaho Data Integration tool to allow the ETL of data from Spotify which is not only extracted from multiple API’s but also is in JSON format. So ETL job extracts data from two different API’s and then parses and flattens the data, which allows the smooth functioning of the algorithm.

Data Features:

Like any other Machine Learning algorithm, the algorithm under discussion also makes use of the following features extracted from Spotify API’s.



Features	Explanation
Track Id	Uniquely represents each track
Track Duration	Total duration of the entire track
Section Start Time	Start time of particular section
Section Duration	Total duration of section
Section Tempo	Tempo of a particular section
Section Loudness	Loudness of a particular section
Track Danceability	How much a track is danceable
Track Energy	How much a track is energetic
Track Loudness	How much a track is loud
Track Speechiness	How much lyrics a track contains
Track Acousticness	How much is a track hearable?
Track Liveness	Detects the presence of an audience in the recording.
Track Valence	Describes the musical positiveness conveyed by a track.
Track Tempo	It is the speed or pace of a given piece and derives directly from the average beat duration.

The empirical efforts done is explained below. The code is deployed in the Github repository and can be checked [here](#). The project is built in Python using Jupyter Notebooks.

- **Task A: Import Libraries**

In the project, the following Python libraries are exploited:

S.No	Library Name	Purpose
1	Numpy	For simplification of complex computations
2	Pandas	For easy data analysis
3	Matplotlib	This is use for plotting of data
4	Seaborn	High level data visualization library
5	Pandas_profiling	For Profile Reports
6	YellowBrick	Suite of visual analysis and diagnostic tools
7	Sklearn	For all ML models
8	ipywidgets	For interactive User Interface of Notebooks
9	graphviz	For plotting complex graphs

- Task B: Exploratory Data Analysis

The data is read into Pandas df from the exported CSV file from the Spoon tool. These are the first five rows of the data.

track_id	track_danceability	track_energy	track_loudness	track_speechiness	track_acousticness	track_liveness	track_valence	track_tempo	class
162G8KlwWHSPAcNzS	0.800	0.497	-8.203	0.0362	0.00438	0.0740	0.329	110.996	0
0BILiPbbV3eiPTlIE98BJ	0.685	0.645	-6.030	0.2290	0.49100	0.0973	0.557	97.311	1
bkrW4m1Tac5xySEJ4M	0.754	0.725	-6.058	0.0661	0.01040	0.1920	0.271	120.002	1
5Rc5cEq9417o5qDOdo	0.414	0.359	-10.129	0.0300	0.79300	0.3200	0.695	112.299	1
ncqXDgegVEncnEosjCs	0.684	0.669	-5.950	0.0335	0.14600	0.1120	0.616	123.049	1

A total of 1275 records are used in the model right now. The following image displays the overall brief idea about the data.

Dataset info

Number of variables	10
Number of observations	1275
Total Missing (%)	0.0%
Total size in memory	99.7 KiB
Average record size in memory	80.1 B

Variables types

Numeric	8
Categorical	1
Boolean	1
Date	0
Text (Unique)	0
Rejected	0
Unsupported	0

Now, every feature of the data will be visualised.

1. Feature#1 Track_Danceability

track_danceability	Distinct count	500	Mean	0.64211
Numeric	Unique (%)	39.2%	Minimum	0.157
	Missing (%)	0.0%	Maximum	0.975
	Missing (n)	0	Zeros (%)	0.0%
	Infinite (%)	0.0%		
	Infinite (n)	0		



2. Feature#2 Track_Energy

track_energy	Distinct count	480	Mean	0.68005
Numeric	Unique (%)	37.6%	Minimum	0.128
	Missing (%)	0.0%	Maximum	0.996
	Missing (n)	0	Zeros (%)	0.0%
	Infinite (%)	0.0%		
	Infinite (n)	0		



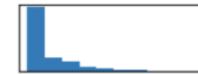
3. Feature#3 Track_Loudness

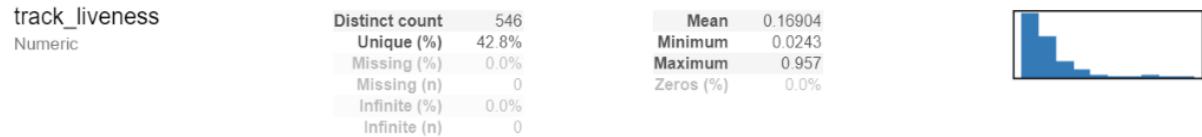
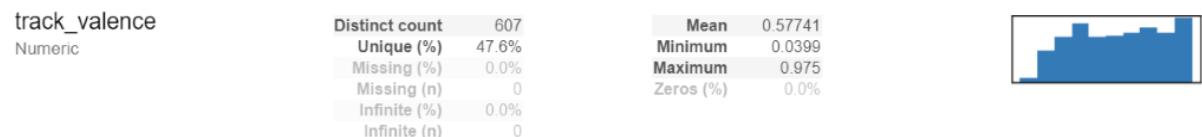
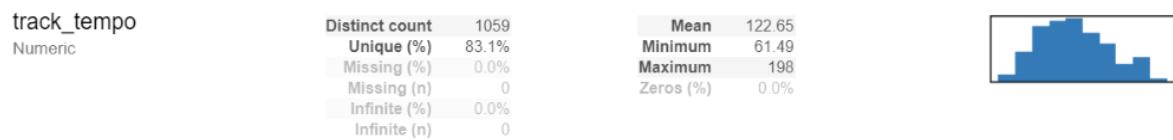
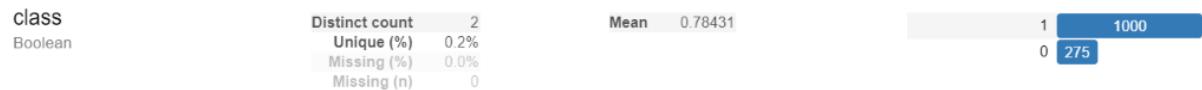
track_loudness	Distinct count	1011	Mean	-6.5682
Numeric	Unique (%)	79.3%	Minimum	-18.899
	Missing (%)	0.0%	Maximum	0.175
	Missing (n)	0	Zeros (%)	0.0%
	Infinite (%)	0.0%		
	Infinite (n)	0		



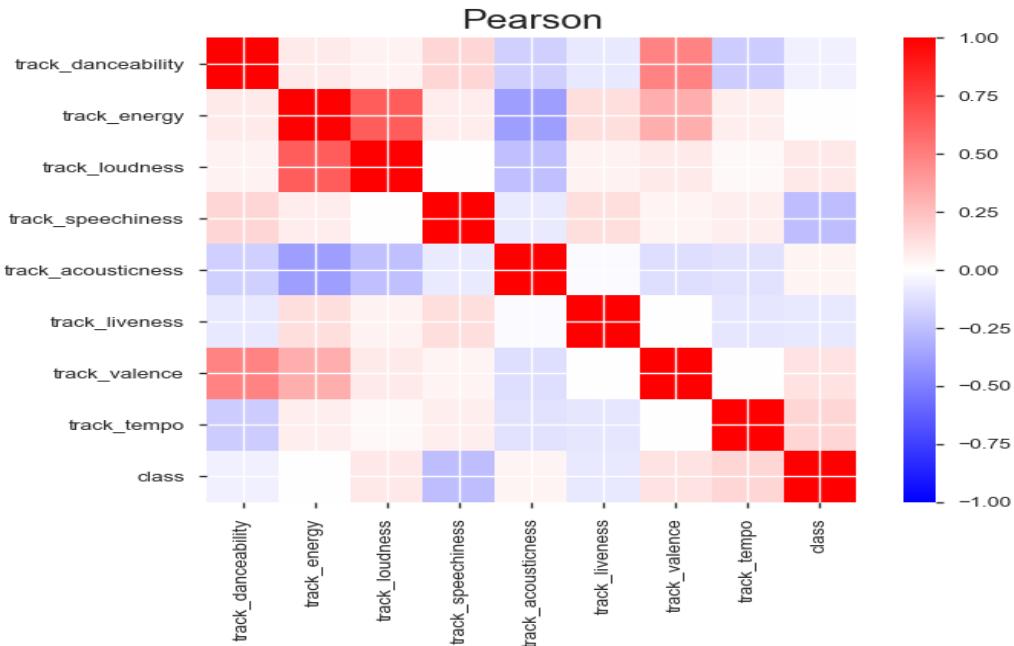
4. Feature#4 Track_Speechiness

track_speechiness	Distinct count	617	Mean	0.10415
Numeric	Unique (%)	48.4%	Minimum	0.0234
	Missing (%)	0.0%	Maximum	0.781
	Missing (n)	0	Zeros (%)	0.0%
	Infinite (%)	0.0%		
	Infinite (n)	0		



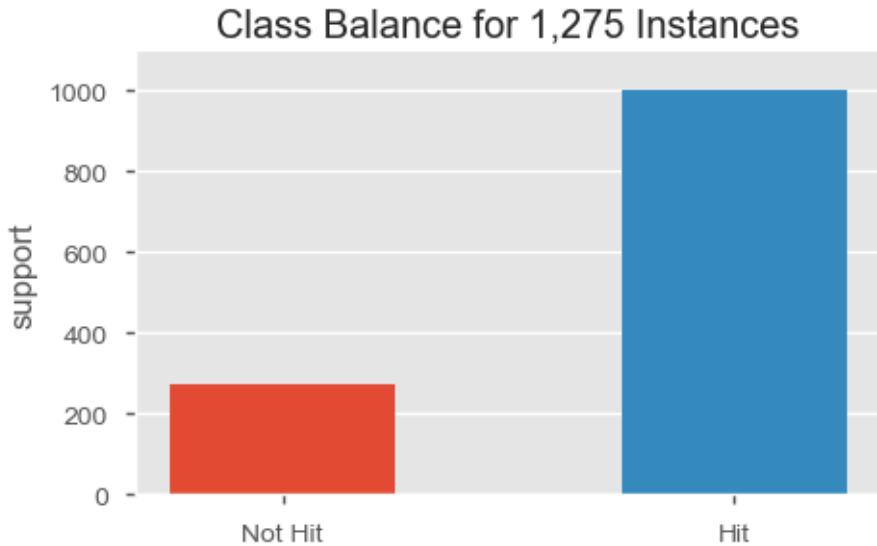
5. Feature#5 Track_Acousticness6. Feature#6 Track_Liveness7. Feature#7 Track_Valence8. Feature#8 Track_Tempo9. Feature#9 Class

Pearson's Correlation between the variables



- Task C: Visualise class imbalance

Now in this task, the skewedness of the data set is displayed, i.e positive to negative class ratio. This is done to ensure the level of stratification required in the train test split.



This visualisation confirms that our data set is highly skewed, so we need to stratify the train test split.

- Task D: Creating training and test set data

The test and train data set are created with train data set being 70 % of the whole data set, and test data set being 30%.

- Task E: Build an Interactive Decision Tree Classifier

In this step, Decision Tree algorithms are used as Machine Learning model, in order to predict our classes.

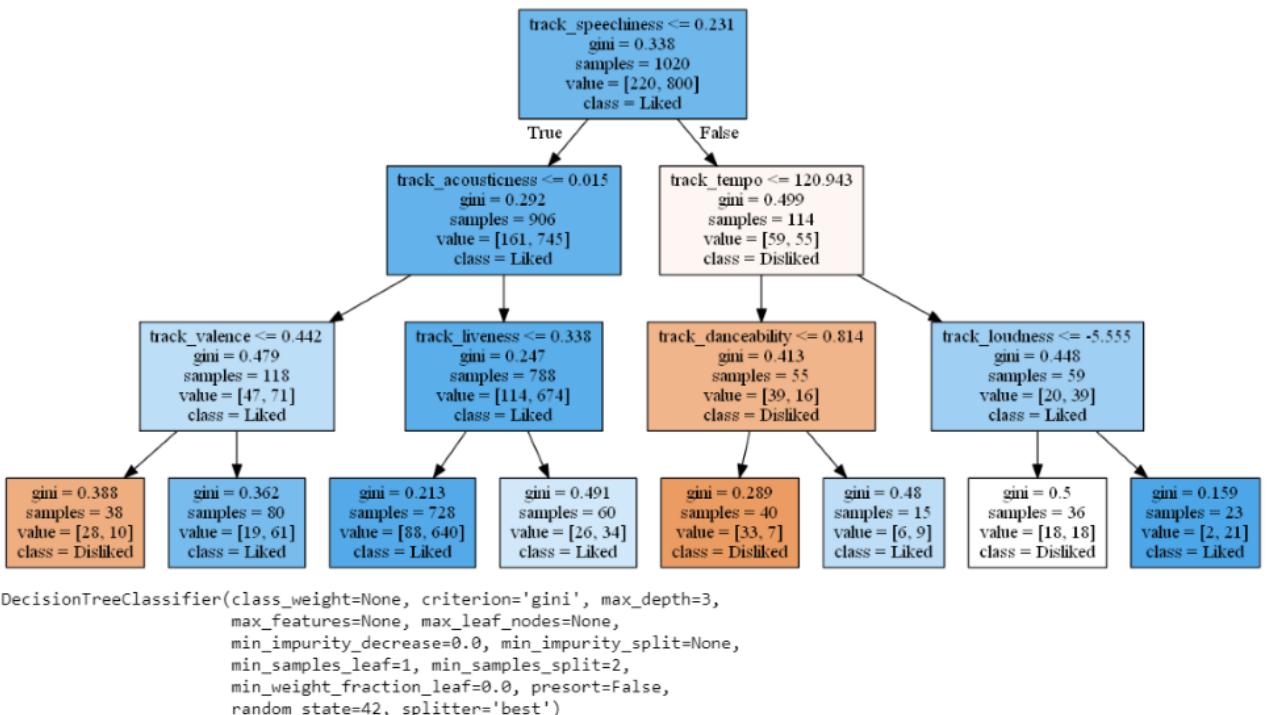
A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. Tree based learning algorithms

are considered to be one of the best and mostly used supervised learning methods. Tree based methods empower predictive models with high accuracy, stability and ease of interpretation.

- Decision trees are non-parametric models which can model arbitrarily complex relations between inputs and outputs, without any a priori assumption
- They implement feature selection, making them robust to noisy features (to an extent)
- Robust to outliers or errors in labels
- Easily interpretable by even non-ML practitioners.

After applying the Decision Tree on the Algorithm, the following results can be seen:

```
decision tree accuracy is 0.827
decision tree accuracy is 0.812
```



Decision Tree Accuracy on test data set with tree depth =3 (underfit)	81.2%
Decision Tree Accuracy on test data set with tree depth =4 (optimum)	83.5%
Decision Tree Accuracy on test data set with tree depth =5(overfit)	82.0%

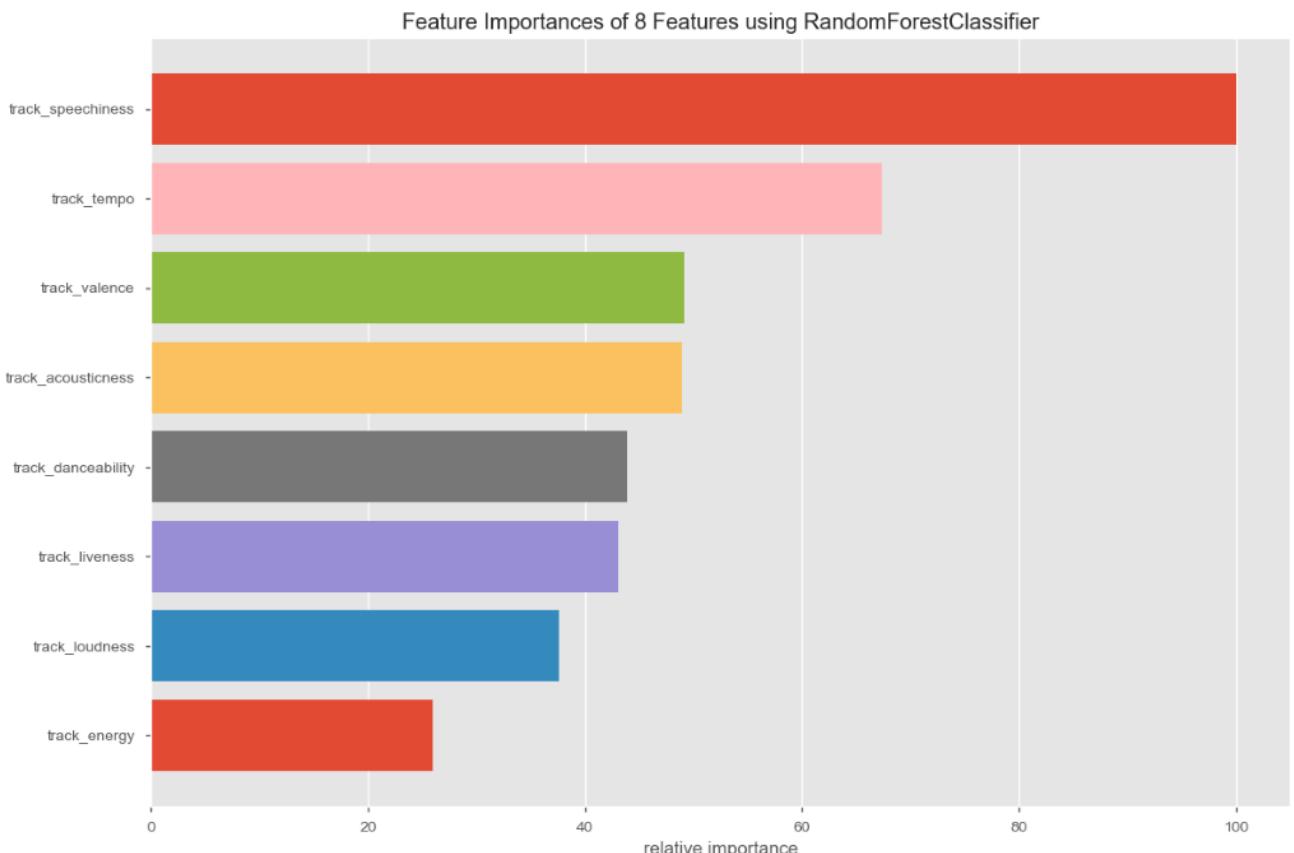
- Task F: Build an Interactive Random Forest Classifier

Along with decision Tree, Random Forest Classifier was also applied. However, the prediction of Decision Tree was superior.

Random Forest Classifier on test data set with tree depth =3	78.4%
Random Forest Classifier on test data set with tree depth =4	78.8%
Random Forest Classifier on test data set with tree depth =5	80.8%

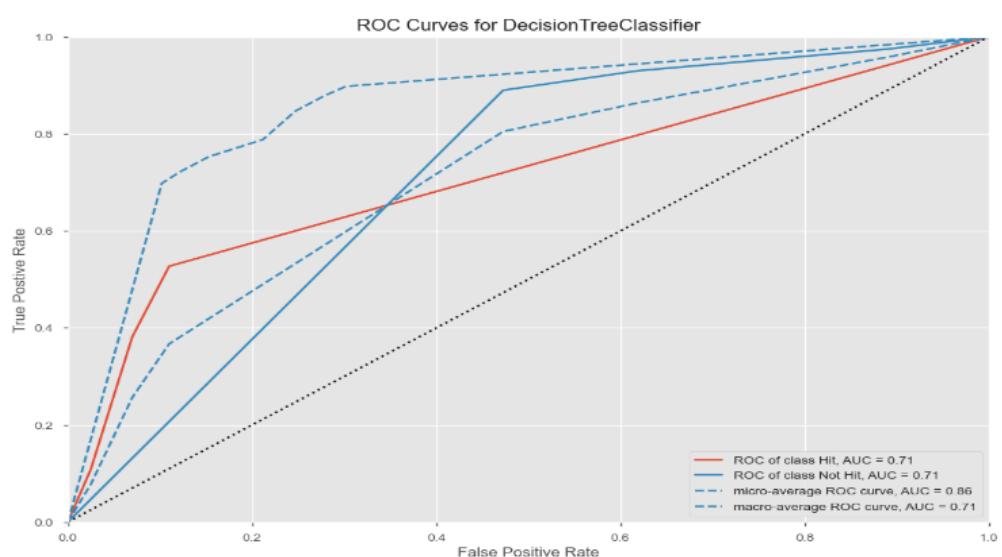
- **Task G: Feature Importance and Evaluation Metrics**

Feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node probability can be calculated by the number of samples that reach the node, divided by the total number of samples. The higher the value, the more important the feature. For each decision tree, Scikit-learn calculates a nodes importance using Gini Importance, assuming only two child nodes. Here using Decision Tree, the relative importance's of all features present in the data has been predicted.



- **Task H: Accuracy Graph (ROC)**

Finally, the receiver operating characteristic curve (ROC) is plotted.

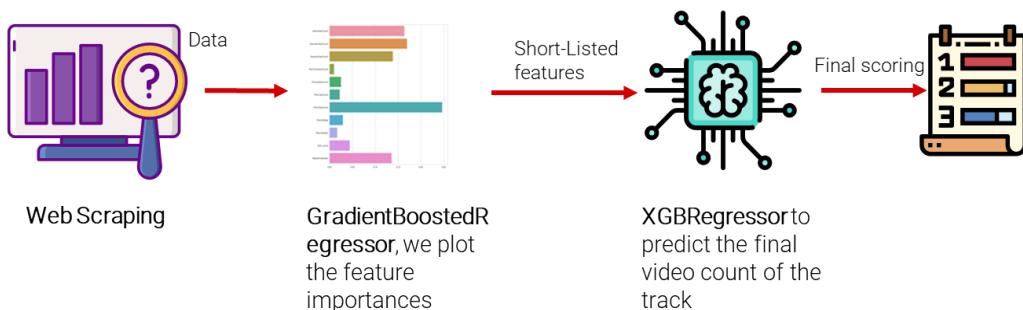


3. Implementation idea of the scouting algorithm

IDEA

The algorithm will be using Ensemble methods, which are a machine learning techniques that combine several base models in order to produce one optimal predictive model. Among Ensemble methods, boosting techniques were selected, that consider homogeneous weak learners, learn them sequentially in a very adaptive way (i.e. a base model depends on the previous ones) and combine them following a deterministic strategy.

Some steps have to be executed in order to predict the final virality. In the following paragraph, each step is discussed one by one.



STEP 1 - THE DATA EXPLORATION AND DATA CLEANING

It was not possible to find an appropriate dataset, so only the methodology that can be applied will be discussed.

1. Data Features

The following features of a composed music would be obtained for each track. These all features are measured with respect to TikTok.

Feature Name	Representation
No. of videos where the track is being used	This feature represents the number of videos, where the artist's track is being used in the background
Total no. of likes of these videos	This feature represents the cumulative sum of all the likes on all the videos where the track is being used.
Total number of shares	This means the cumulative sum of all the shares of all the videos where the track is used.
Total number of comments and total number of shares	This indicated the cumulative sum of all the comments of all the videos where the track is used.
Title of the original video	Title of the original videos where the track featured for the first time
Date published	Data when the track was published for the first time on TikTok
Follower count of the artist's account	The total number of the followers of the artist of the original track
Number of videos posted by the artist	Total number of videos/Tracks posted by the artist till now on his account

Comment count of the previous video posted by the artist	Sum of all the comments on all the previous videos.
View count of the previous video posted by the artist	Sum of all the views on all the previous videos of the artist
Age of the artist's account	Age in days of the artist's account

2. Data Cleaning

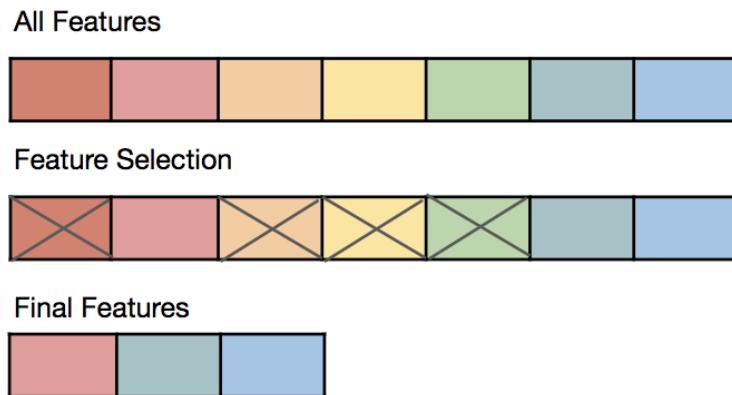
Data cleaning is the process of ensuring that the data is correct, consistent and useable by identifying corruptions in the data, correcting or deleting them, manually processing them as needed to prevent the error from happening again. In order to do that, the following steps are performed:

- First, some outliers are dropped: artists that have already gone “viral,” which were defined as artists crossing a fixed milestone of the counts.
- Then standardisation the data is done, i.e. all the numeric fields into the range 0-1 were normalised.
- For all the object type of data, Label Encoding is used and one hot encoding in order to make them into numerical fields.
- For all the missing values, the values are substituted by 0.

STEP 2 - FEATURE SELECTION

Feature Selection is one of the core concepts in machine learning which can hugely impacts the performance of our model. The data features that are used to train your machine learning models have a huge influence on the performance that can be achieved.

Feature Selection is one of the core concepts in machine learning which can hugely impacts the performance of our model. The data features that you use to train your machine learning models have a huge influence on the performance you can achieve. Irrelevant or partially relevant features can negatively impact model performance.



STEP 3 - FINAL MODEL

Gradient Boosting Techniques will be used for feature scaling as well as for the model. Then GradientBoostedRegressor will be used and then evaluate feature importances. Finally, a XGBRegressor to predict the future virality score of the track.



OUTPUT

Based on the training data set, the model is finally ready. Now the unknown data set, for which the outcome is unknown, is fed onto this model. The model will predict the scores out of 100. The score out of 100 will evaluate how much is the probability of the artist getting viral in the future.

