Geographical Classification of Traditional and Ethnic Genre Music using ML Approach

Claude Widmer
Department of Geography
University of Zurich
Zurich, Switzerland
claude.widmer@uzh.ch

ABSTRACT

This study investigates the interplay between music and geography, utilizing tree-based machine learning, specifically the CatBoost algorithm, to categorize traditional, ethnic, or 'world' genre-based music. Employing the MARSYAS framework for audio feature extraction, the research integrates geographical clustering methods, including manual clustering, DBScan, and KMeans. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE) and Principal Component Analysis (PCA) address data imbalances and reduce dimensionality.

The methodology involves data preparation, PCA/SMOTE application, exploratory data analysis (EDA), and clustering techniques (manual clustering, DBScan, and KMeans). The study concludes with training a CatBoost classifier on six datasets (each clustering method with PCA and PCA&SMOTE).

Key findings underscore the substantial impact of SMOTE, with performance enhancements ranging from 0.04 to 0.1 across metrics. Manual clustering yields suboptimal scores (0.4 to 0.5), while DBScan performs better, achieving an F1 score of 0.6. KMeans emerges as the top-performing method with an impressive F1 score of 0.91. In the K-Means algorithm, applied to all attributes rather than only location, no distinct geographic pattern is discernible. This deviation arises from the pervasive influence of music that transcends geographic boundaries, allowing similar musical styles to emerge in entirely disparate regions. The K-means method strategically emphasizes this phenomenon by attributively clustering the data, spotlighting commonalities in musical attributes over geographic proximity. This approach acknowledges and accommodates the idea that musical genres possess the ability to transcend traditional geographic boundaries.

In the conclusion, the study addresses various classification performance issues, including imbalanced and insufficient data, inappropriate model selection, and potential shortcomings in features and geographical representation. Furthermore, it points to inadequacies in the MARSYAS program as potential contributors to suboptimal music analysis.

1 Introduction & Motivation

Music, a universal language, reflects the cultural and geographical contexts of its origins [1]. This paper seeks to use advanced data analysis techniques to uncover the geographic information embedded within music. With the foregoing field of data science and the increasing integration of artificial intelligence (AI) across disciplines [2], the combination of music and geography through advanced machine learning (ML) techniques is becoming more used [3]. This new approach of viewing and categorizing data opens up more applications.

The integration of geographic data into music analysis is a known concept in Music Information Retrieval (MIR) [4]. This step involves the integration of geographic data into music analysis. Further, with the geographic data acquired, GIS principles such as clustering methods, provides a structured approach to unveil patterns and relationships within the landscape of musical genres. This combination of geographic expertise and data science is an interesting combination.

At its core, this research seeks to unravel the relationship between music and geography by using GIS techniques and data-driven methodologies. The main goal is to develop a framework that utilizes Machine Learning algorithms, such as CatBoost [5], and the MARSYAS framework [6] to accurately classify traditional, ethnic, or 'world' genre-based music. By doing so, this paper aims to contribute not only to the field of music classification but also to the broader discourse on how or if geographic nuances shape and define musical genres.

The significance of this study lies in its potential to provide an understanding of the geographical influences on music and of the methods used. Beyond the conventional categorization of genres, the research aims to capture the essence of geographic diversity within musical compositions, even inside the same area. As a geography student, this work represents a bridge between traditional

geographical analysis and the evolving landscape of Machine Learning, using a multidisciplinary approach to explore the cultural and spatial dimensions of music.

2. Literature Review

2.1 MIR (Music Information Retrieval)

Music Information Retrieval (MIR) represents a specialized field within information retrieval that focuses on the extraction, analysis, and organization of musical information from various sources. MIR encompasses a wide range of tasks, including music classification, genre identification, chord recognition, and recommendation systems. The main goal is to develop computational techniques that enable machines to understand, organize, and retrieve music-related data efficiently [7]. MIR has three different key tasks: **Information retrieval**, such as cover-sond identification, similarity measures or music recommendations, **Classification & estimation**, such as identification of performer, genre, or tempo. Lastly, **Sequence labeling**, which uses optical music recognition (OMR), Melody extraction and Chord estimation [7]. It is seen that MIR is connected to geographic data in the tasks of information retrieval and classification & estimation. Geographic Information Retrieval (GIR) is also linked with the classification of music in a geographical context.

2.2 Geographic Information Retrieval (GIR) in Music

Geographic Information Retrieval (GIR) is a specialized field within information retrieval that focuses on extracting and organizing spatially relevant information from large datasets. In the context of GIR, the emphasis is on retrieving data based on geographic parameters, such as coordinates, place names, or spatial relationships. GIR uses processes such as ranking, georeferencing, indexing or evaluation [8] in order to get the geographic information needed. GIR is often coupled with unstructured text-based information such as books, news articles and databases. In many cases, music is not considered as text-based information. But if we look closely at the structure of music, we can evaluate similarities between music and text-based information.

Exploring the parallels between textual and musical information reveals similarities in how individuals perceive and process both forms of expression. Both text and music are powerful vehicles of emotion [9], functioning as complex channels through which human experiences are received and interpreted. While textual narratives have the power to trigger empathy, convey stories, and evoke feelings [10], music achieves a similar impact through its melodic and rhythmic elements. Both serve as conduits for emotional expression, tapping into the realms of human sentiment and experience.
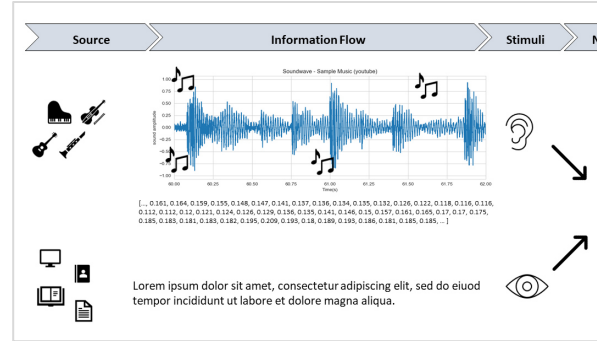
Figure 1: Information flow of music compared to text-based information (Part 1)

The ordered arrangement of information is another similarity between text and music. Textual information is structured through language, syntax, and semantics, creating a coherent narrative [11]. Similarly, music organizes sound patterns, melodies, and harmonies into a structured composition, guiding listeners through a sonic journey [12]. In essence, both text and music rely on the meticulous organization of elements to give meaning and cause specific responses.

Furthering the parallels, the processing of both text and music engages our sensory organs and activates similar neural pathways. Textual information is primarily accessed through the visual stimuli of reading, engaging the eyes and the corresponding brain regions [11]. Musical information is processed similarly. The soundwaves are processed by our ears and transported through the nervous system to the brain. According to the University of Central Florida [13], both music and written language is processed in similar regions of our brain. The experience of music with hearing loss is similar to the experience of reading with vision loss. In a comparable manner, music, primarily an auditory experience, is processed through the ears, activating overlapping regions in the brain associated with auditory perception and emotional response [14].
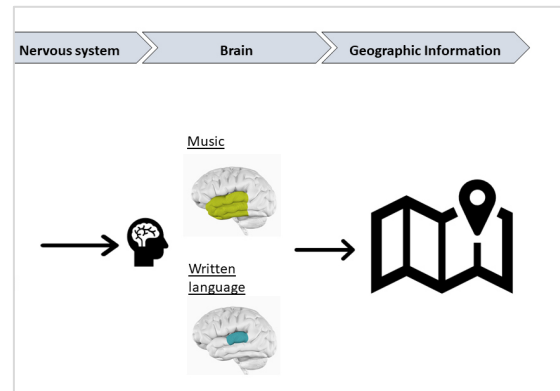
Figure 2: Information flow of music compared to text-based information (Part 2) – based on [13]

Studying the connections between text and music can help us find geographic details in music. The way emotions, information organization, and our brain processes both text and music are similar. There exist frameworks that try to capture this information of music into text form with different attributes. One of the frameworks is called MARSYAS.

2.2 MARSYAS and other feature extraction frameworks

The MARSYAS (MusicAl Research SYstem for Analysis and Synthesis) framework, a robust tool in music information retrieval (MIR), was developed by the Music Technology Group at Pompeu Fabra University. It is optimized for performance and user-friendliness because it has a user-interface. Its creation is due to the need for a flexible and comprehensive platform capable of extracting complex audio features from music signals, enabling a deeper understanding and analysis of musical compositions [6].

The primary objective behind MARSYAS' development was to address the multifaceted challenges associated with music analysis. Traditional methods often fell short in capturing the richness of audio data, especially when dealing with complex musical genres and diverse instrumentation [6]. MARSYAS was designed as a solution to these limitations, aiming to provide researchers, musicians, and data scientists with a tool capable of not only processing audio signals but also uncovering nuanced patterns and features within them. The main advantages of MARSYAS are:

- Extensive Feature Extraction: MARSYAS offers a comprehensive set of tools for extracting audio features. Its strength lies in its ability to capture a wide array of features, including spectral, temporal, and rhythmic components. This ability allows for a more nuanced representation of musical content, making it particularly suitable for diverse genres and complex compositions [15].

- Modularity and Customization: One of the distinguishing features of MARSYAS is its modular design, enabling users to customize and adapt the framework to their specific analytical needs. This modularity facilitates the creation of tailored workflows, allowing researchers to focus on particular aspects of music analysis relevant to their study objectives [6].

- Open-Source Nature: MARSYAS is an open-source framework that encourages collaboration and community-driven development. This characteristic encourages a continuous refinement of the tool, ensuring that it remains at the forefront of advancements in music analysis. Its open-source nature also makes it accessible to a broader audience, promoting transparency and reproducibility in research [6].

Various feature extraction toolboxes are available to analyze audio signals, each offering unique capabilities. MARSYAS is just one among several options. Aubio, written in C++, stand out for its high-level feature extraction, including onset detection, beat tracking, and melody [16]. Essentia serves as a comprehensive workflow environment with both high and low-level features, facilitating audio input, preprocessing, and statistical analysis in C++ with Python bindings [17]. Librosa provides a Python API for feature extraction [18], while jAudio offers a Java-based standalone application with a GUI and CLI for batch processing [19]. Other tools also contribute to this type of application, each designed with specific focus and functionality. According to [15], MARSYAS performs the worse in terms of feature range, but has its advantages in the easy-to-use interface and easy implementation.

In addition to these toolboxes, Spotify has also focused on music feature extraction using its own API. Leveraging the capabilities of its API, Spotify has developed methods for extracting various musical features that help analyze and categorize audio content within its platform. This initiative by Spotify further expands the range of tools available and demonstrates how industry leaders are actively engaged in advancing feature extraction techniques to improve music understanding and user experience [20].

2.4 CatBoost Classification

In the project the Machine Learning is performed with the CatBoost Classifier, developed by Yandex. It specializes in classification tasks and stands out in handling categorical features [5]. The algorithm constructs decision trees iteratively during the training process, emphasizing the correction of errors from previous iterations – a characteristic of gradient boosting [21]. This gradient boosting approach will improve the predictive power of this classifier.

A distinctive advantage of CatBoost lies in its seamless integration of categorical variables into the training workflow. Unlike many other algorithms, CatBoost eliminates the need for extensive preprocessing steps, simplifying the implementation process. This is particularly beneficial for practitioners seeking an efficient solution for real-world applications where datasets often comprise a mix of numerical and categorical features.

CatBoosts implementation is quite straightforward, offering its implementation in programming languages like Python and R. After installation, the CatBoost library aligns seamlessly with the broader ecosystem of data science libraries [5].

There are a few considerations to keep in mind. CatBoost's computational intensity, especially for larger datasets, can impact performance, even though a GPU training is possible [5]. Fine-tuning of hyperparameters with manual calibration or Cross Validation may be necessary to achieve optimal results, although the algorithm often shows competitive performance with default settings.

2.4 Clustering methods

Clustering plays an important role in the following classification tasks. Clustering groups similar items or data points together to identify patterns or relationships based on shared characteristics. This can be manual (manual calibration) or with specialized algorithms such as DBScan (automatic calibration). Clustering can be done with geographic data, such as defining the data into spatial categories such as regions or continents according to the UNN [22].

Advantages of human-made regional clusters lie in interpretability and alignment with established geographic boundaries, though they may oversimplify [23], particularly when organizing data into distinct human-made regional clusters, such as continents or regions. Other than manually clustering, there is an option to automatically cluster the data. Automatic clustering is a process where a computer algorithm organizes data into distinct groups or clusters based on inherent similarities, without explicit instructions or predefined categories. An example is the DBScan (Density-Based Spatial Clustering of Applications with Noise), which is quite well-known and often used in geographically based algorithms.

DBScan identifies dense regions of data points in a space, separated by sparser areas. It works by defining clusters as continuous regions of high data point density, allowing for the automatic discovery of clusters with even irregular shapes. DBScan categorizes points as core, border, or noise, making it robust in handling clusters of varying shapes and sizes while being resistant to outliers. This algorithm is particularly useful in situations where clusters have varying densities and when traditional methods may struggle to capture such complexities [24]. It is important that DBScan is generally only used for spatial clustering tasks.

As already mentioned, clustering will group similar items or data points together. This can be done with non-spatial data aswell. There are many different types of automatic clustering methods that take into account different characteristics and structures of the data. Examples include k-means algorithms, which is a type of partitioning clustering method. K-means works by dividing a dataset into a predetermined number of clusters, with each data point assigned to the cluster that has the nearest mean. This algorithm is known for its simplicity, understandability and efficiency in handling large datasets [21].

2.5 SMOTE

In scenarios where one class (or in this case cluster) is significantly underrepresented, traditional machine learning algorithms may struggle to accurately learn patterns and make predictions for the minority class. SMOTE (Synthetic Minority Over-sampling Technique) is a resampling method used in machine learning to address class imbalance. It operates by generating synthetic examples of the minority class to balance the dataset, particularly beneficial when the minority class is underrepresented. SMOTE selects a data point from the minority class and creates synthetic instances along the line segments connecting it to its nearest neighbors. This technique helps prevent biased models that favor the majority class and enhances the performance of classifiers by providing a more balanced representation of the classes [25].

2.6 PCA

Another technique, Principal Component Analysis (PCA), is used for dimensionality reduction, helping in computational efficiency and pattern recognition. While SMOTE (see chapter 2.5) focuses on class balance, PCA complements it by simplifying the data structure. Principal Component Analysis (PCA) serves as a key

dimensionality reduction technique in machine learning and data analysis. Its core function involves transforming complex datasets into a lower-dimensional space while retaining crucial information. This is achieved by identifying principal components, representing directions where variance in the data is maximized [26]. The more principal components, the more variance is explained. If less principal components are chosen, the less variance is explained. Therefore, selecting the optimal number of principal components is critical to maintaining a satisfactory level of variance explanation.

3. Research

Exploring the intersection of music and geography this study investigates the capability of tree-based machine learning, specifically utilizing the CatBoost algorithm, to accurately classify traditional, ethnic, or 'world' genre-based music. Western music is excluded from the data since western music influence is global. The study utilizes the MARSYAS framework for audio feature extraction, incorporating geographical clustering methods, such as manual clustering and DBScan, to reveal patterns within the data. Additionally, techniques like SMOTE and PCA are applied for data balancing and dimensionality reduction.

The goal is to contribute to the evolving landscape of music analysis by combining machine learning, audio feature extraction, and geographic clustering. The exploration of different methods aims to explore the complexity of music and to show that music is also within the domain of Geographic Information Retrieval (GIR).
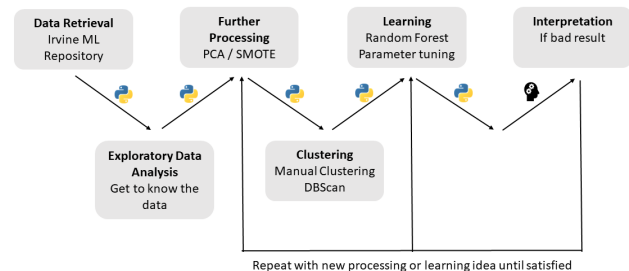
4.   Methods

Figure 3: Overview of methods

The project unfolds through a series of systematic steps implemented in Python, aiming to explain the complex relationship between music and geography. Figure 3 provides an overview of the methods used in temporal order. The first phase involves data retrieval from the Irvine Machine Learning Repository [27], providing the dataset for analysis.

The Dataset itself is about 1059 tracks long and covering 33 countries. Geographical information, specifically the origin of each track, is manually gathered from CD sleeve notes. In instances where the provided information is insufficient, additional sources are consulted. The precision of location data is constrained to the country of origin. Determination of the country of origin relies on the primary residence of the artist or artists, excluding tracks with

ambiguous origin. The geographical point of origin is established as the latitude and longitude of each country's capital city (or the province of the area). Audio features are extracted from the wave files using the MARSYAS program (see Chapter 2.2) [28].
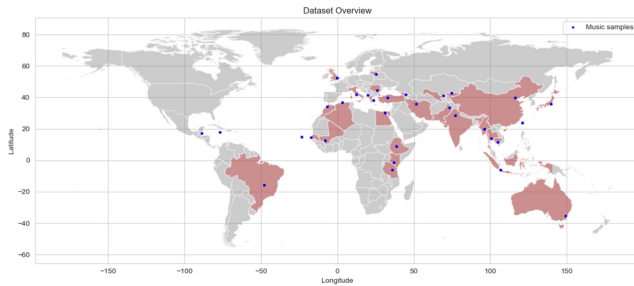
Figure 4: Geographical Overview of the Dataset [python]

An exploratory data analysis (EDA) is conducted, to gain insights into the characteristics and distribution of the dataset. This analytical phase aims to improve understanding by systematically examining the dataset's patterns and key features. Building on the insights gained from EDA, further processing is undertaken to refine the data. This includes adjustments to the data structure, using Principal Component Analysis (PCA) for dimensionality reduction, and employing Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance.

The clustering phase includes both manual and automatic clustering. Manual clustering and DBscan aligns with predefined geographic regions, while K-means operates on the attributes and internal data structures to reveal latent patterns within the dataset. These three clustering methods are all used for later processing. The machine learning phase focuses on training the CatBoost algorithm on these 6 datasets: Three clustering methods each with PCA and SMOTE. The number of features is different for the PCA dataset and the PCA&SMOTE dataset. This is because SMOTE adds more instances to the data to overcome the cluster imbalance.

Parameter tuning is performed on the CatBoost classifier to optimize the performance of the model, and cross-validation (CV) is used to reduce its robustness. The results obtained from the machine learning model establish the foundation for an in-depth interpretation using various metrics such as accuracy, precision, recall and F1 score.

This steps, illustrated in Figure 3, were performed. If the model did not perform well, the methods were adjusted in order to improve the predictive power of the model.
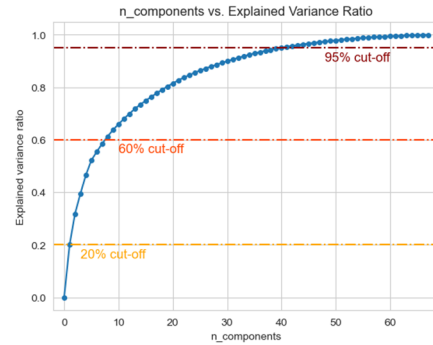
5. Results

In the following chapter, key findings for the PCA, both clustering methods and models are presented.

5.1 PCA

As described in Chapter 4, once the data had been structured, a PCA was carried out. Figure 5 illustrates this well. It shows the relationship between the number of principal components

(n_components) and the percentage of variance explained in a principal component analysis (PCA).
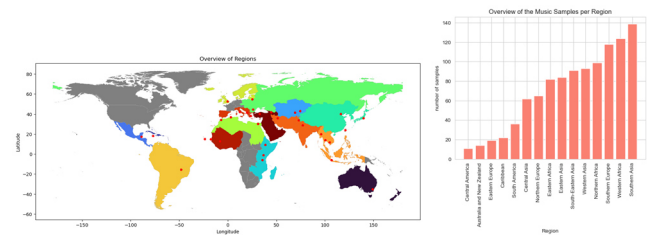

Figure 5: Overview Manual Clustering [python]

Dashed lines at 95%, 60%, and 20% indicate reference points for the percentage of variance retained. The 95% cutoff was chosen for further analysis, resulting in a reduction of the data set by 28 features. With 41 components, 95% of the variance is captured, helping to determine the optimal balance between dimensionality reduction and retained information. The resulting 41 features extracted out of the PCA were used for further analysis. In the next step, the clustering was performed.
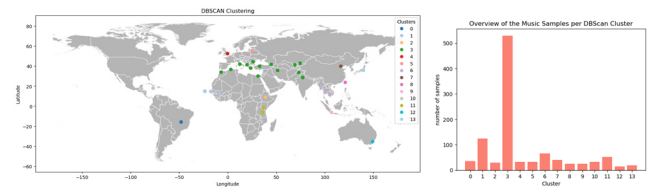
5.2 Clustering

In manual clustering (refer to Fig. 6), an imbalance in class distribution is observed. Southern Asia, with approximately 140 datapoints, has the highest amount of data for further training, whereas Central America has the fewest.

Figure 6: Overview Manual Clustering [python]

The DBScan method, denoted as method number 2, primarily focuses on the geographic pattern of datapoints. The algorithm has clustered more than half of the datapoints into the same cluster number 3, as illustrated in the histogram in Figure 7. In the DBScan method, Cluster 3 encompasses regions from Europe to the Middle East and India.

Figure 7: Overview DBScan Clustering [python]

In the K-Means algorithm, conducted not on geographic location but on the internal data structure of the music samples, no obvious geographic pattern is visible. The points also overlap, as many are in close proximity. This geographic imbalance will be further analyzed in the discussion. The 9 classes exhibit imbalance, ranging from approximately 270 music samples in cluster no. 1 to around 20 music samples in cluster no. 7.
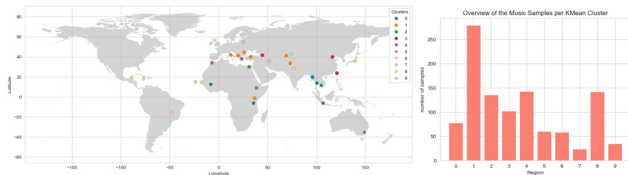

Figure 8: Overview K-Means Clustering [python]

5.3 SMOTE

Due to the observed imbalance especially in Cluster 3 of the DBScan clustering, the application of the SMOTE (Chapter 2.5) algorithm proves to be particularly useful. In Figure 9, it is evident that the SMOTE algorithm has synthetically added points to address the underrepresented classes, effectively mitigating the imbalance in the dataset.
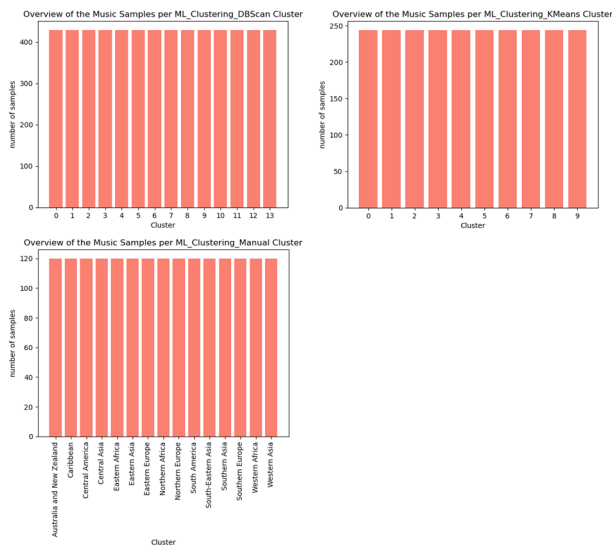

Figure 9: Overview SMOTE [python]

Having addressed the clustering aspect of the analysis, the focus now shifts towards the core objective of this paper: the classification and prediction of the resulting six datasets.

5.4 Classification & Prediction

Figure 10 presents the key findings with the corresponding 4 scores: Accuracy, Precision, Recall and the F1 Score. The key findings, outlined below, serve as a reference for further discussion.

- **Key finding 1**: Using SMOTE improves performance across all metrics by 0.04 to 0.1.
- **Key finding 2**: Manual clustering yields poor scores ranging from 0.4 to 0.53.
- **Key finding 3**: DBScan provides better results than manual clustering.
- **Key finding 4**: DBScan achieves an F1 score of 0.6.
- **Key finding 5**: Kmeans emerges as the best performing method overall.
- **Key finding 6**: Kmeans has a good F1 score of 0.91.
- **Key finding 7**: The F1 score is the lowest of the four metrics. Precision und Recall have different values.
- **Key finding 8**: There is a trade-off between precision and recall especially noticeable in Manual PCA and DBScan PCA.
- **Key finding 9**: The datasets which were processed with SMOTE have reduced imbalance between precision and recall.
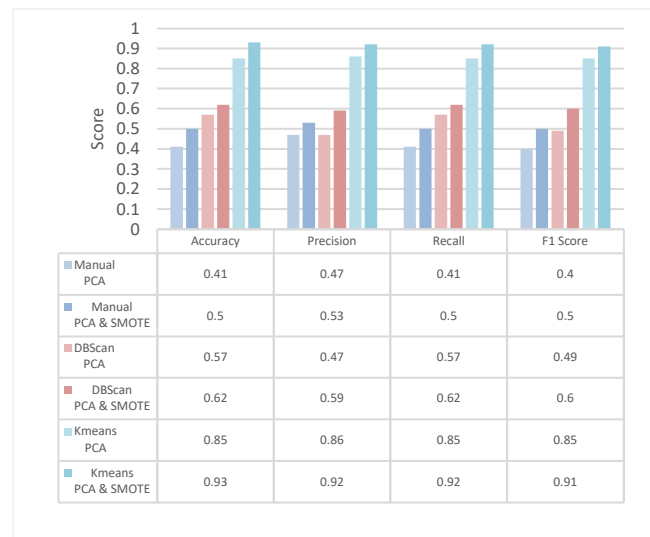

	Accuracy	Precision	Recall	F1 Score
Manual PCA	0.41	0.47	0.41	0.4
Manual PCA & SMOTE	0.5	0.53	0.5	0.5
DBScan PCA	0.57	0.47	0.57	0.49
DBScan PCA & SMOTE	0.62	0.59	0.62	0.6
Kmeans PCA	0.85	0.86	0.85	0.85
Kmeans PCA & SMOTE	0.93	0.92	0.92	0.91

Figure 10: Overview Results

6. Discussion

In this discussion, we first critically analyze the cluster methods, followed by an examination of classification and other relevant topics.

6.1 Cluster Analysis Insights

Manual clustering proves challenging, given that human-defined boundaries are constantly shifting, and traditional music does not adhere to these delineations. Because human-made boundaries are always shifting, and traditional music is not following necessarily those boundaries, the strong and clear-cut separation between groups may not be advantageous. However, this approach is necessary as it pertains to the classification of a single cluster, not a percentage breakdown of music origins. For instance, imagine conducting a manual clustering analysis on a diverse dataset of world music. Rather than seeking to determine the percentage distribution of musical origins, much as DNA testing identifies percentual genetic heritage, the focus is on classifying music into distinct

clusters. The challenge lies in grouping musical compositions that share similarities across geographic boundaries and cultural nuances.

The DBSCAN clusters in this analysis tend to align more closely with geographical patterns, examining the true density of data points. It is not influenced by human-made boundaries, because this algorithm focuses on the distribution of the points. However, DBScan forms a single large cluster (Cluster 3), which could introduce bias during classification, as discussed in Chapters 6.3/6.4. This is likely to result in a higher score compared to manual clustering (see Key Finding 3 and Key Finding 4). The formation of a single large cluster may result in a higher classification score for several reasons. First, the inherent emphasis on geographic proximity may lead to the inclusion of different musical styles within the same cluster, contributing to a more inclusive representation. While this inclusivity may capture a wide range of musical diversity, it does not capture the geographic information behind the music.

The K-means method produces the best results, as summarized in Key Finding 5. This result is expected because K-means focuses on the internal structure of the data rather than on geographic proximity. It is not influenced by human-made boundaries. When examining the generated classes, it becomes apparent that these clusters are not always geographically nearby. This shift occurs because the influence of music transcends geographic boundaries, with similar musical styles appearing in completely different regions. The K-means method aims to emphasize this by attributively clustering the data, emphasizing commonalities in musical attributes rather than geographic proximity. This approach recognizes that musical genres can transcend geographic boundaries. It also contributes to a more nuanced understanding of the dataset by revealing the diverse influences that shape musical landscapes.

6.2 Classification Performance

The classification performance ranges from unsatisfactory to very good, depending on the dataset. Manual clustering, particularly, yields the lowest scores.

Key finding 8 suggests that there is a trade-off between precision and recall especially noticeable in Manual PCA and DBScan PCA. In the case of Manual PCA, the higher precision than recall suggests that the model is cautious in predicting positive instances, aiming for accuracy but potentially missing some actual positives. On the other hand, in DBScan PCA, where precision is higher than recall, it indicates a trade-off where the model is focused on being precise in positive predictions, even if it means sacrificing the ability to capture all actual positive instances. This suggests that in Manual PCA, the cautious approach results in a higher certainty of correct positive predictions at the expense of potentially missing some positives, while in DBScan PCA, precision is prioritized even if it means sacrificing a bit of recall, possibly to avoid false positives. This difference between precision and recall could be due to class imbalance. After applying SMOTE, the difference between precision and recall is reduced (key finding 9). This difference is seen in the F1 score, which will take this difference between

precision and recall into account. The F1 scores highlight various issues within the data [29]. First problem that could explain the low F1 scores is the **imbalanced data**. When dealing with an imbalanced dataset, where one class is significantly more prevalent than the other, the model may face challenges in knowing the minority class, resulting in a performance loss and a resulting low score. This imbalance was addressed using SMOTE to minimize its impact, leading to improvement (Key finding 1). However, its effectiveness is limited, especially when there is a scarcity of data available for SMOTE in a particular underrepresented class. In such cases, despite attempts to reduce bias, SMOTE may still generate biased data. While the performance has been enhanced, as indicated by Key Finding 1, a residual bias will persist in the dataset.

The second explanation is the **insufficient data.** Limited dataset size or an inadequate number of representative examples for each class can hinder the model's capacity to acquire a robust representation. For this classification task, the dataset with only 1059 different music samples may not be enough to classify the data [28]. This number could be increased by adding more data to the classification. An investigation revealed that adjusting the train and test split from 20%/80% to 30%/70% resulted in a better fit, suggesting that a larger dataset could improve the predictive power of the models.

Additionally, a low score may be attributed to **inappropriate model selection**, where the chosen model is unsuitable for the specific task or lacks proper tuning. Despite the models being appropriately tuned, alternative approaches, such as neural networks or manual classification, might have the potential to outperform them in the classification task.

Moreover, **inadequate features** can contribute to low scores. If the selected features fail to capture relevant information for the task, the model may struggle to identify meaningful patterns in the data. The geographical imbalance in the data and potential shortcomings in the MAYRAS program may have resulted in improper music analysis. In comparison to other audio feature extraction toolboxes, MAYRAS is less developed [15]. Essentia [17], for example, can cover up to 4 times more audio features than the MAYRAS framework [15].

6.3 Training time of the model

Implementing the models proved to be time consuming, even though the CatBoost algorithm is resource friendly. Training times, especially with cross-validation, were notably long, especially on my laptop. The training time for a single model among the six reached about 5 minutes, which increased cumulatively while experimenting with different preprocessing approaches and test settings. In contrast, training times were significantly reduced to about 40 seconds per model when performed on my stationary desktop PC. This disparity underscores the importance of a robust CPU when working with ML models.

While the training process required time and computational resources, it's worth noting that ML models can be efficiently stored after training using the dedicated CatBoost library. While this

storage capability proves valuable for model analysis, it doesn't directly reduce the time constraints during the initial training phase.

7.   Conclusion

This project illustrates that the geographical classification of music is challenging and depends on several factors, such as the uncertainties in the feature extraction software, the nature of the music itself, and the amount of training data. The use of about 1000 samples may have been insufficient; the consideration of a larger dataset such as the "A Million Songs Dataset" [30] with extracted geographic data could provide a solution. Music is ambiguous and transcends traditional boundaries, making it difficult to understand and categorize. This project highlights this complexity. While the methods used in the project, in particular Kmeans clustering, produced good results, the difficulty was in establishing a meaningful geographic association for the music.

Understanding human creation is not straightforward, and music, with its complexity and its omnipresence, is a reflection of this complexity. Due to the inherent nature of music, it is challenging to classify its geographic origin with 100% certainty. While certain genres of music may be associated with certain regions, accurate geolocation proves to be quite difficult.

8.   Suggestions for future work

While working on this project, several ideas for future work have come to mind. The online implementation of the MARYAS framework could be explored to create a map that automatically classifies music to specific geographic points, enhancing the user experience. Additionally, a more traditional analysis of music structure, with a focus on elements such as chord progressions and rhythms, could be considered. Furthermore, experimentation with various machine learning models, including ensemble methods and neural networks, can be undertaken to contribute to a comprehensive evaluation of classification performance. Lastly, the development of a percentile classification model could offer a better understanding by providing a percentage breakdown of the music's origin, moving beyond a singular cluster approach.

9.   GitHub

The code and project data can be found on GitHub: <https://github.com/widmerc/GEO871>

REFERENCES

[1]   G. Carney, 'Music Geography', *J. Cult. Geogr.*, vol. 18, no. 1, pp. 1–10, Sep. 1998, doi: 10.1080/08873639809478309.

[2]   W. U. Yuchen, 'The Data Mining Application Based on WEKA: Geographical Original of Music', *Manag. Sci. Eng.*, vol. 10, no. 4, pp. 36–46, 2016, doi: http://dx.doi.org/10.3968/8997.

[3]   E. Gó mez, M. Haro, and P. Herrera, 'Music and geography: Content description of musical audio from different parts of theworld', in *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*, 2009, pp. 753–758.

[4]   K. R. Page, B. Fields, B. J. Nagel, G. O'Neill, D. C. De Roure, and T. Crawford, 'Semantics for Music Analysis through Linked Data: How Country is My Country?', in *2010 IEEE Sixth International Conference on e-Science*, Dec. 2010, pp. 41–48, doi: 10.1109/eScience.2010.49.

[5]   A. V. Dorogush, V. Ershov, and A. G. Yandex, 'CatBoost: gradient boosting with categorical features support'.

[6]   G. Tzanetakis and P. Cook, 'MARSYAS: a framework for audio analysis',

[7]   J. S. Downie, 'Music Information Retrieval (Chapter 7)', *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. Chapter 7, pp. 295–340, 2003.

[8]   R. S. Purves, P. Clough, C. B. Jones, M. H. Hall, and V. Murdock, 'Geographic Information Retrieval: Progress and Challenges in Spatial Search of Text', *Found. Trends Inf. Retr.*, vol. 12, no. 2–3, pp. 164–318, 2018, doi: 10.1561/1500000034.

[9]   D. Västfjäll, 'Emotion induction through music: A review of the musical mood induction procedure', *Music. Sci.*, vol. 5, no. 1_suppl, pp. 173–211, Sep. 2001, doi: 10.1177/10298649020050S107.

[10]   S. Aman and S. Szpakowicz, 'Identifying Expressions of Emotion in Text', in *Text, Speech and Dialogue*, vol. 4629 LNAI, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 196–205.

[11]   S. J. DeRose, D. G. Durand, E. Mylonas, and A. H. Renear, 'What is text, really?', *ACM SIGDOC Asterisk J. Comput. Doc.*, vol. 21, no. 3, pp. 1–24, Aug. 1997, doi: 10.1145/264842.264843.

[12]   R. Typke, F. Wiering, and R. C. Veltkamp, 'A survey of music information retrieval systems', in *ISMIR 2005 - 6th International Conference on Music Information Retrieval*, 2005, pp. 153–160.

[13]   University of Central Florida, 'Your brain on music', 2024. https://www.ucf.edu/pegasus/your-brain-on-music/.

[14]   Pegasus Magazine, 'Music and the Brain: What Happens When You're Listening to Music', *Pegasus Magazine- University of Central Florida*, 2022. https://www.ucf.edu/pegasus/your-brain-on-music/ (accessed Jan. 13, 2024).

[15]   D. Moffat, D. Ronan, and J. D. Reiss, 'An evaluation of audio feature extraction toolboxes', in *DAFx 2015 - Proceedings of the 18th International Conference on Digital Audio Effects*, 2015, pp. 1–7.

[16]   P. M. Brossier, 'The aubio library at mirex', *Synthesis (Stuttg).*, pp. 1–3, 2006.

[17]   D. Bogdanov *et al.*, 'Essentia: An audio analysis library for music information retrieval', in *Proceedings of the 14th International Society for Music Information Retrieval Conference, ISMIR 2013*, 2013, pp. 493–498.

[18]   B. McFee *et al.*, 'librosa: Audio and Music Signal Analysis in Python', in *Proceedings of the 14th Python in Science Conference*, 2015, no. Scipy, pp. 18–24, doi: 10.25080/Majora-7b98e3ed-003.

[19]   D. Mcennis, C. Mckay, and P. Depalle, 'JAUDIO : A feature extraction library', in *Proceedings of the international conference on music information retrieval*, 2005, pp. 600–603.

[20]   D. Duman, P. Neto, A. Mavrolampados, P. Toiviainen, and G. Luck, 'Music we move to: Spotify audio features and reasons for listening.', *PLoS One*, vol. 17, no. 9, p. e0275228, Sep. 2022, doi: 10.1371/journal.pone.0275228.

[21]   H. Drucker, C. Cortes, L. D. Jackel, Y. LeCun, and V. Vapnik, 'Boosting and Other Machine Learning Algorithms', *Proc. 11th Int. Conf. Mach. Learn. ICML 1994*, pp. 53–61, Jan. 1994, doi: 10.1016/B978-1-55860-335-6.50015-5.

[22]   U. Nations, 'Continents', 2024. https://www.un.org/geospatial/document-type/continent (accessed Jan. 02, 2024).

[23]   M. Alexandrov, A. Gelbukh, and P. Rosso, 'An Approach to Clustering Abstracts', in *Lecture Notes in Computer Science*, vol. 3513, 2005, pp. 275–285.

[24]   S. U. Rehman, S. Asghar, S. Fong, and S. Sarasvady, 'DBSCAN: Past, present and future', in *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*, Feb. 2014, pp. 232–238, doi: 10.1109/ICADIWT.2014.6814687.

[25]   J. Wang, M. Xu, H. Wang, and J. Zhang, 'Classification of Imbalanced Data by Using the SMOTE Algorithm and Locally Linear Embedding', in *2006 8th international Conference on Signal Processing*, 2006, vol. 3, pp. 6–9, doi: 10.1109/ICOSP.2006.345752.

[26]   A. Maćkiewicz and W. Ratajczak, 'Principal components analysis (PCA)', *Comput. Geosci.*, vol. 19, no. 3, pp. 303–342, Mar. 1993, doi: 10.1016/0098-3004(93)90090-R.

[27]   M. Kelly, R. Longjohn, and K. Nottingham, 'The UCI Machine Learning Repository', 2024. https://archive.ics.uci.edu (accessed Jan. 13, 2024).

[28]   F. Zhou, 'Geographical Original of Music', *UCI Machine Learning Repository*. 2014, doi: https://doi.org/10.24432/C5VK5D.

[29]   L. Derczynski, 'Complementarity, F-score, and NLP evaluation', in *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, 2016, pp. 261–266.

[30]   T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, 'The Million Song Dataset', 2011.

Organised Sound, vol. 4, no. 3, pp. 169–175, Dec. 2000, doi: 10.1017/S1355771800003071.