1

Project Artificial Intelligence, Supervised Learning.

Diana Lorena Balanta Solano
dianalorenabalantasolano@gmail.com

Carlos Javier Bolaños Riascos
carjabo12@gmail.com

Danna Alexandra Espinosa Arenas
alexandraarenas1705@gmail.com

Universidad Icesi
Department of Computing and Intelligent Systems

I. ABSTRACT

In this document we will explain and analyze the process and results obtained in the development of the Artificial Intelligence I course project. The project to be developed deals with the "Classification of musical genres", specifically Rock and Hip-Hop. The objective of this project is to demonstrate whether it is possible to predict the musical genre of a song given its metrics and characteristics.

To approach this project, three deliveries were proposed, the first one covered the research process and understanding of the problem, the search for databases and sources necessary for our development. The second part of the project involved data processing and training of the supervised learning models. Finally, the last part of the project consisted of the refinement of the trained models, the implementation of the product (deployment) and the conclusions.

II. INTRODUCTION

The classification of music genres is important for music platforms and user recommendations, allowing them to offer products that are associated with the tastes of users according to their visualizations or requests. In the context of the project and artificial intelligence, it is relevant to understand how the models learn and process data captured in real life.

During the research process different metrics and characteristics that differentiate the genres were discovered, so its application and development is a great academic exercise that helps to understand the logic and implementation of recommendation to users, according to their searches and tastes. It is interesting to see how from a complex reality, abstractions and mathematical models are applied to help us explain phenomena, and thus provide approximate solutions to the great challenges that society faces today.

III. THEORY

In the research and development of the project, several concepts about artificial intelligence, metrics and audio processing were applied. In the latter we found some metrics, mathematical characteristics of the audios, in order to train our supervised learning models. The metrics used were:

A. Acousticness: measures the probability that a song is acoustic. This metric ranges from 0 to 1, where 1 is more likely that the song is fully acoustic.
B. Danceability: Evaluates how suitable a song is for dancing based on a combination of musical elements such as tempo, rhythm stability, beat strength, and overall regularity. Values also range from 0 to 1, where 1 means that a song is more danceable.
C. Energy: Measures the perceived intensity and activity of a song, considering dynamics such as speed, loudness, and generality of sound. The range is from 0 to 1, where higher values indicate a more energetic song.
D. Instrumentalness: This metric predicts whether a track contains no vocals. Values close to 1.0 indicate tracks with a high probability of being instrumental.
E. Liveness: Detects the presence of an audience in the recording. A higher value (near 1.0) suggests that the song was performed live.
F. Speechiness: Identifies the presence of spoken words in a track. A higher value means that the track contains more spoken parts, as in a podcast or spoken book, and a value near 0.0 indicates non-verbal music.
G. Tempo: Measures the speed or tempo of a track in beats per minute (BPM). It is a quantitative measure of speed.
H. Valence: Measures the positive musicality conveyed by a track. Tracks with high valence sound more positive (happy, euphoric), while tracks with low valence sound more negative (sad, depressed).

These characteristics are calculated by applying learning models and audio signal analysis. The exact formulas and models behind these metrics are generally kept private by the companies that develop them, such as Spotify, and are not publicly available.

IV. METHODOLOGY

A. The CRISP-DM methodology, which is used in the development of Data Science projects, was applied in the solution of the project. The stages included in the project were:
B. Business understanding. Research was done on audio

1

processing, in addition to databases appropriate to the musical genres to be classified: Rock and Hip-Hop.

C. Data understanding. After doing the respective research and choosing the databases to use, we went on to make an exploration of the data, review their nature and what they indicate in the scales of their context, which in this case is music - audio processing. For this, two databases were used, one with basic data (description) about the songs, the other with the metadata, that is, with the characteristics described in Theory.

D. Data preparation. Data validation was performed, checking for the existence of null or repeated values. At the end of this, it was decided to apply techniques such as Standar Scaler and PCA to reduce the complexity and dimensionality of the data.

E. Modeling. With the processed data ready, we proceeded to the training of the supervised learning models seen in the course: Logistic Regression and Decision Tree.

F. Evaluation. At the end of the training of the models, the results and predictions were reviewed. It could be seen that the models made the predictions but tended to a specific class (Rock), i.e., that the model had biases, so it was decided to balance the database and retrain the models. By applying this, the results and predictions improved.

G. Deployment. In order to show how the solution works to the client, a Python script/application was made, where he could interact and see how the system classifies the genre of the song he wants.

V. Results

When applying the exploration to the databases, it was observed:

1) 'Rock' genre has a higher frequency than Hip-Hop, so this could influence the models, i.e., the possibility of bias by majority class.

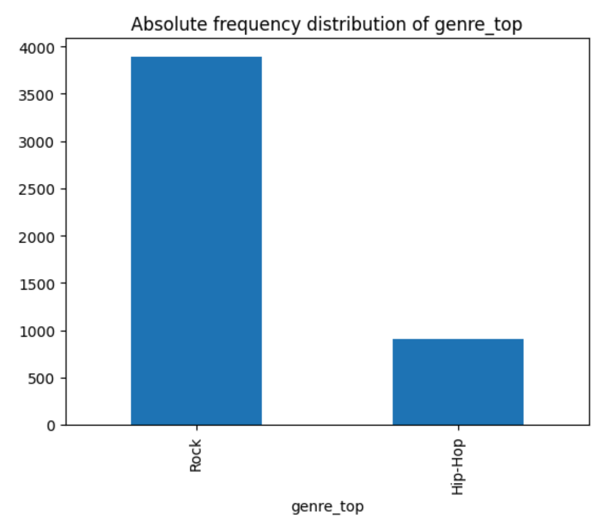

Fig 1. Absolute frequency of genera in the databases.

2) Distribution of audio metrics. When applying the exploratory analysis to the data.

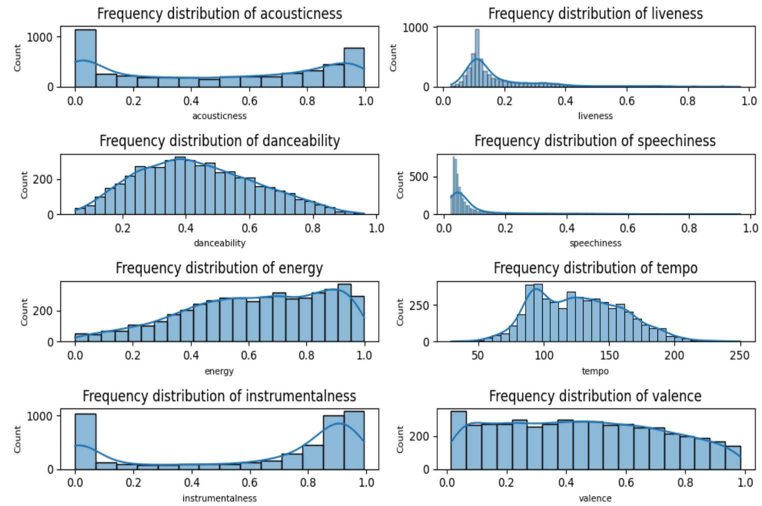

Fig 2. Distribution of audio metrics.

With this, we can say that:

A. Acoustics: ranges from practically 0 (9.491e-07) to almost 1 (0.9957965), with a median (0.5156888) suggesting that half of the tracks have a moderate level of acoustic elements.

B. Danceability: Varies widely (min. 0.051307, max. 0.961871) with a median around 0.436, indicating a moderate tendency for the tracks to be suitable for dancing.

C. Energy: Also shows a wide range (min. 0.000279, max. 0.999768), with a higher mean (0.625), suggesting that the tracks generally have a moderate to high energy level.

D. Instrumentality: Most tracks have a high degree of instrumental content (median 0.808752), which might suggest a prevalence of tracks without vocal content.

E. Vividness: Typically low across all tracks (median 0.188), indicating few live recordings.

F. Expressiveness: Generally low (mean 0.104877), suggesting that spoken words are not a dominant component in most tracks.

G. Pace: Ranges from 29.093 to 250.059 beats per minute, with a mean close to the typical pace of popular music (124.625 bpm).

H. Valence: shows a wide range (min. 0.014392, max. 0.983649) with a median of 0.446240, indicating varied emotional positivity across tracks.

3) Correlation matrix. When doing the correlation matrix analysis between the variables, we can see that

there is no strong relationship between them, indicating that none of these characteristics completely determines the other.

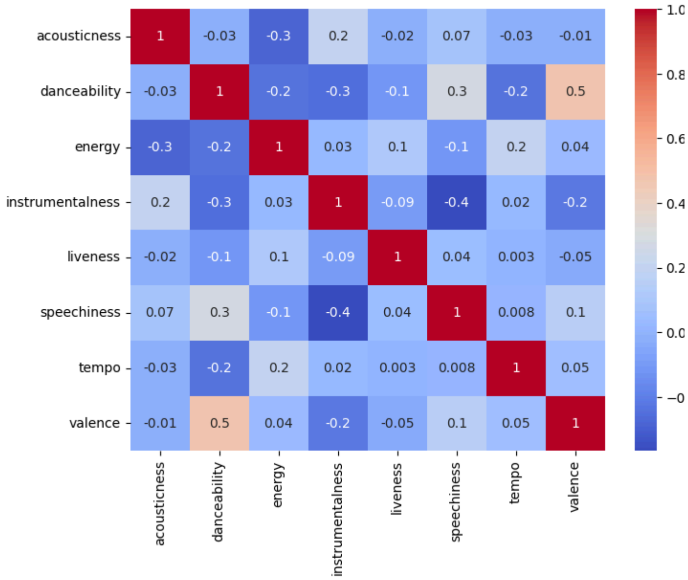

Fig 3.Correlation matrix between variables (audio metrics).

VI. Results analysis

When the models were trained, the following results were obtained.

Decision Tree:

	precision	recall	f1-score	support
Hip-Hop	0.63	0.63	0.63	233
Rock	0.91	0.91	0.91	968
accuracy			0.86	1201
macro avg	0.77	0.77	0.77	1201
weighted avg	0.86	0.86	0.86	1201

Logistic Regression:

	precision	recall	f1-score	support
Hip-Hop	0.76	0.46	0.57	233
Rock	0.88	0.96	0.92	968
accuracy			0.87	1201
macro avg	0.82	0.71	0.75	1201
weighted avg	0.86	0.87	0.85	1201

Fig 4.First training of supervised learning models (Decision Tree and Logistic Regression).

With the results of this first training, we can say that:

A. Models perform better overall in the Rock class compared to Hip-Hop. This could be due to an imbalance in the number of samples (larger support for Rock), which often influences the performance of the models.
B. Logistic Regression seems to be slightly better in terms of accuracy and F1-score for Rock, but has a lower recall performance for Hip-Hop compared to the Decision Tree.
C. The lower performance in Hip-Hop could indicate a bias in the model or in the data, possibly due to the lower support for this class in the dataset.

It is evident that the frequency imbalance between classes (Rock and Hip-Hop) makes the models biased. For this reason, a second training of the models with frequency balanced data will be applied to see how it behaves.

Decision Tree:

	precision	recall	f1-score	support
Hip-Hop	0.77	0.76	0.76	226
Rock	0.76	0.78	0.77	229
accuracy			0.77	455
macro avg	0.77	0.77	0.77	455
weighted avg	0.77	0.77	0.77	455

Logistic Regression:

	precision	recall	f1-score	support
Hip-Hop	0.81	0.84	0.82	226
Rock	0.83	0.81	0.82	229
accuracy			0.82	455
macro avg	0.82	0.82	0.82	455
weighted avg	0.82	0.82	0.82	455

Fig 5.Second training of supervised learning models (Decision Tree and Logistic Regression). Balanced data.

New training results show improvements, balanced data contributed to:

A. Giving balance between classes. Compared to the previous data, the performance between Hip-Hop and Rock, these results show a more balanced and uniform performance.
B. Improved accuracy and recall for Hip-Hop in Logistic Regression. In the previous results the Logistic Regression showed significantly lower recall for Hip-Hop, in these data, the recall and precision for Hip-Hop are noticeably better.
C. Generalization of the models. Precision and recall in both models are more consistent across classes in these results, suggesting better generalization and ability to handle equally represented classes.

VII. deployment.

In order to allow users to interact with the results of the project, a music genre classification system (currently only available for Rock and Hip-Hop) was implemented with the results of the trained models. In this software solution the user will be able to type the name of the song he/she wants to classify. Underneath, the system connects through a Spotify web API, with which it can extract the metrics and thus use them as input to make the predictions in the models.

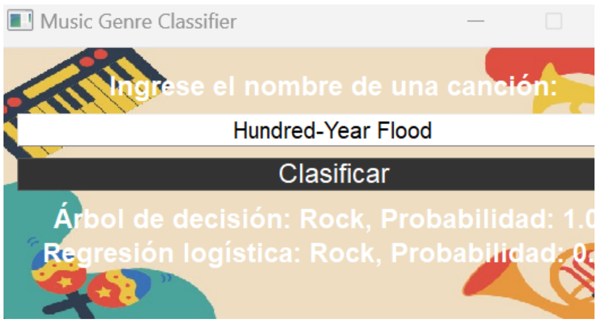

Fig 6. Deployment of the software solution to classify Rock and Hip-Hop music genres.

To use the application, you must have a python programming environment with the PyQt5, spotipy, pandas and pickle libraries.

VIII. CONCLUSIONS

Through the implementation of different models such as Decision Tree and Logistic Regression, we have been able to explore not only the theory behind these techniques, but also their ability to accurately handle and predict real data.

1) Model Performance: Evaluation of the models revealed that Logistic Regression, in particular, showed superior overall performance compared to Decision Tree, especially in terms of accuracy and recall for both genders. This underlines the importance of choosing an appropriate model based on the specific characteristics of the dataset and the objectives of the analysis.

2) Importance of Data Quality: One of the key lessons from this project has been the importance of good data preparation and management. The balance between classes, selection of relevant features, and normalization were crucial to improve the accuracy and generalizability of the models.

3) Practical Applicability: The results obtained have significant implications for music streaming platforms. The ability to accurately classify music genres can improve user experience by personalizing music recommendations according to their preferences, which is vital to retain and satisfy users in a competitive market.

4) Challenges and Future Directions: Despite successes, the project also faced challenges, such as handling class imbalances and variability in the quality of genre labels. Future research could explore more advanced class balancing techniques, and the use of deep learning methods that could more accurately capture the complexities and subtleties of musical features.

IX. RESOURCES

Kumar, "Classify Song Genres from Audio Data," Kaggle, 2021. [Online]. Available: https://www.kaggle.com/code/ashishkumarak/classify-song-genres-from-audio-data.

D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor, G. Roma, J. Salamon, J. Zapata, and X. Serra, "ESSENTIA: An Audio Analysis Library for Music Information Retrieval," in Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013), Curitiba, Brazil, November 4-8, 2013, pp. 493-498.

G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals," IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293-302, July 2002.