PROJECT SEMESTER REPORT

## Audio Classification

by

**Soumil Jainer**

**Roll No.102003114**

Under the Guidance of

**Dr. Anshul Verma**

**Assistant Professor**

**Dr. Prashant Singh Rana**

**Associate Professor**



Submitted to the

# Computer Science & Engineering Department

**Thapar Institute of Engineering & Technology, Patiala**

In Partial Fulfilment of the Requirements for the Degree of

Bachelor of Engineering in Computer Engineering

at

Thapar Institute of Engineering & Technology, Patiala

## Topic: Audio Classification

by Soumil Jainer

Place of work: BHU, Varanasi

Submitted to the Computer Science & Engineering Department, Thapar Institute of Engineering & Technology

June 8, 2024

In Partial Fulfilment of the Requirements for the Degree of Bachelor of Engineering in Computer Engineering

**Abstract**:

Music genre classification, a vital component of Music Information Retrieval (MIR), continues to captivate researchers due to its relevance in content-based searching and recommendation systems. In this research, I introduced a comprehensive methodology for classifying music into distinct genres. Leveraging features extracted from audio files and employing various machine learning models, including Naive Bayes, Stochastic Gradient Descent, KNN, Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, Neural Networks, and XGBoost, an ensemble model is proposed to enhance classification accuracy. Additionally, blending techniques and feature importance analysis, such as Permutation Importance, are explored to refine the classification process. I am also writing a Research paper which will conclude with a discussion on the significance of the proposed approach and avenues for future research.

Author Soumil Jainer

Certified by Due to prevailing scenario email from Industrial mentor will serve as signature

Take email from mentor🡪 Convert in PDF 🡪 Paste that prior certificate

(Name & Signature) (Industrial Coordinator / mentor)

Certified by Due to prevailing scenario email from Faculty mentor will serve as signature

Take email from mentor🡪 Convert in PDF 🡪 Paste that prior certificate

(Name & Signature) (Faculty mentor)

# CERTIFICATE (PROJECT SEMESTER TRAINING) FROM THE COMPANY OR THE ORGANIZATION

# Candidate must place the scanned or original copy of the certificate related to completion of the project semester as received from the software company / research institute.

# TABLE OF CONTENT

[1. Company Profile](#_heading=h.4d34og8) 1

[2. Introduction](#_heading=h.2s8eyo1) 1

[3. Background](#_heading=h.17dp8vu) 2

[4. Objectives](#_heading=h.3rdcrjn) 5

[5. Methodology](#_heading=h.26in1rg) 6

[6. Observations and Findings](#_heading=h.lnxbz9) 9

[7. Limitations](#_heading=h.35nkun2) 11

[8. Conclusions and Future Work](#_heading=h.1ksv4uv) 13

[9. Bibliography/References](#_heading=h.44sinio) 15

10. Peer Review Form 16

# 

# 

# Company Profile

Banaras Hindu University is an internationally reputed temple of learning, situated in the holy city of Varanasi. This creative and innovative university was founded by the great nationalist leader, Pandit Madan Mohan Malaviya, in 1916, through close cooperation with great personalities like Dr Annie Besant, who viewed it as the University of India. Banaras Hindu University was established by the Parliamentary legislation-B.H.U. Act 1915. It is a collegiate, central, and research university located in Varanasi, Uttar Pradesh, India.

# Introduction

In the digital age, music classification and recommendation systems have become pivotal in enhancing user experience across various platforms. The objective of this research is to classify audio files into different genres and develop a recommendation system based on audio similarity. Utilizing the GTZAN dataset, which comprises 1000 audio files categorized into 10 distinct genres, this project leverages advanced machine learning and deep learning techniques to achieve high accuracy in genre classification and to recommend similar songs to users.

The research encompasses several key stages: data preprocessing, feature extraction, exploratory data analysis (EDA), dimensionality reduction, model training and evaluation, and the development of a recommendation system. Various machine learning algorithms, including Naive Bayes, Logistic Regression, K-Nearest Neighbors, Decision Trees, Random Forest, Support Vector Machine, and deep learning models, have been employed and compared to achieve the best performance. Ensemble techniques like Voting and Stacking classifiers are also explored to enhance accuracy.

Furthermore, the research delves into deep learning models, implementing several neural network architectures to push the boundaries of classification accuracy. A recommendation system based on cosine similarity has been integrated, offering users recommendations based on audio features.

# Background

**3.1 Context and Motivation**

Music genre classification and recommendation systems play a crucial role in digital music platforms, enhancing user experience by providing personalized content. With the exponential growth of digital music libraries, efficient and accurate classification systems are essential for organizing and retrieving music. Additionally, recommendation systems are invaluable for suggesting new and relevant music to users, based on their preferences and listening history.

The GTZAN dataset, widely used in music information retrieval (MIR) research, offers a rich collection of audio files spanning various genres. This dataset serves as an ideal benchmark for developing and evaluating genre classification algorithms.

The motivation behind choosing this research stems from the following reasons:

* **Personal Interest**: A deep interest in music and technology, and a desire to explore the intersection of these fields.
* **Practical Application**: The real-world application of genre classification and recommendation systems in enhancing user experiences on music platforms like Spotify, Apple Music, and YouTube.
* **Research Opportunity**: The opportunity to contribute to the field of MIR by experimenting with and comparing various machine learning and deep learning techniques.
* **Technical Challenge**: The challenge of working with audio data, which involves complex preprocessing and feature extraction processes.

**3.2 Proposed Layout for Achieving Project Goals**

The research is structured into the following key phases:

* **Data Preprocessing and Exploration**:
  + Loading and inspecting the GTZAN dataset.
  + Visualizing audio data using spectrograms, waveforms, Fourier transforms, and mel spectrograms.
* **Feature Extraction**:
  + Extracting essential audio features such as zero-crossing rate, harmonics, tempo BPM, spectral centroid, spectral rolloff, MFCCs, and chroma frequencies.
  + Performing Exploratory Data Analysis (EDA) to understand feature distributions and correlations.
* **Dimensionality Reduction**:
  + Applying Principal Component Analysis (PCA) for reducing feature dimensions while retaining essential information.
  + Visualizing the reduced feature space using scatter plots.
* **Model Training and Evaluation**:
  + Implementing and training various machine learning models, including GaussianNB, Logistic Regression, K-Nearest Neighbors, Decision Trees, Random Forest, Support Vector Machine, Neural Networks, and XGBoost.
  + Evaluating model performance using metrics such as accuracy.
* **Ensemble Learning**:
  + Combining multiple models using ensemble techniques like Voting and Stacking classifiers to improve accuracy.
* **Deep Learning Models**:
  + Experimenting with various neural network architectures.
  + Implementing early stopping criteria to optimize training.
* **Recommendation System**:
  + Developing a recommendation system using cosine similarity based on audio features.
  + Providing song recommendations based on similarity scores.
* **Results and Analysis**:
  + Comparing the performance of different models and techniques.
  + Analyzing the strengths and weaknesses of each approach.
* **Conclusion and Future Work**:
  + Summarizing the findings and contributions of the project.
  + Suggesting potential improvements and future research directions.

# Objectives

The objectives of this research are centred around developing a robust audio classification system and an effective recommendation engine. The following are the primary goals to be achieved:

**4.1 Develop a High-Accuracy Genre Classification System**:

* **Objective**: To create a machine learning and deep learning-based system that can accurately classify audio files into their respective genres using the GTZAN dataset.
* **Success Criteria**: Achieve a classification accuracy of at least 90% using advanced techniques such as ensemble learning and deep neural networks.

**4.2 Implement a Recommendation System**:

* **Objective**: To build a recommendation system that suggests similar songs based on audio feature similarity.
* **Success Criteria**: Develop a cosine similarity-based recommendation engine that effectively recommends songs similar to a given track, enhancing user experience.

**4.3 Compare and Analyse Model Performance**:

* **Objective**: To systematically compare the performance of various machine learning models, ensemble methods, and deep learning architectures.
* **Success Criteria**: Provide a comprehensive analysis of model accuracies, highlighting the strengths and weaknesses of each approach to guide future research and application development.

These objectives aim to not only build functional systems but also to contribute valuable insights into the effectiveness of different algorithms and techniques in the domain of music information retrieval.

# Methodology

**5.1 Data Acquisition and Preprocessing:**

* **Data Collection**: Obtain the GTZAN dataset, which contains audio files categorized into ten genres.
* **Data Inspection**: Reviewed the dataset to understand its structure, file formats, and labels.
* **Data Preprocessing**:
  + Load audio files and extract relevant features using libraries like Librosa or Essentia.
  + Converted audio signals into a suitable format for feature extraction (e.g., spectrograms, mel-frequency cepstral coefficients).
  + Normalized the features to ensure consistent scaling across different audio samples.
  + Split the dataset into training and testing sets for model evaluation.

**5.2 Feature Extraction:**

* Extracted a diverse set of audio features to capture different aspects of the audio signals, such as temporal, spectral, and perceptual characteristics.
* Common features to extract include:
  + Time-domain features: Zero-crossing rate, energy, root mean square amplitude.
  + Frequency-domain features: Spectral centroid, spectral bandwidth, spectral contrast.
  + Temporal features: Tempo, rhythm patterns.
  + Spectrogram-based features: Mel-frequency cepstral coefficients (MFCCs), chroma features.
* Performed feature selection or dimensionality reduction techniques to reduce computational complexity and enhance model performance.

**5.3 Exploratory Data Analysis (EDA):**

* Visualized the distribution of audio features across different genres to gain insights into their characteristics.
* Explored feature correlations and relationships using techniques like scatter plots, heatmaps, and pair plots.
* Identified any outliers or anomalies in the dataset that may affect model training and performance.

**5.4 Model Selection and Training:**

* Evaluated a variety of machine learning and deep learning models for genre classification, including:
  + Naive Bayes, Logistic Regression, k-Nearest Neighbors, Decision Trees, Random Forest, Support Vector Machine, Neural Networks, XGBoost, etc.
* Trained each model using the training dataset and optimize hyperparameters through techniques like grid search or random search.
* Evaluated model performance using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrices.
* Explored ensemble learning techniques to combine predictions from multiple models for improved accuracy and robustness.

**5.5 Recommendation System Development:**

* Implemented a recommendation system based on cosine similarity to recommend similar songs to users.
* Calculated cosine similarity scores between audio features of different songs.
* Ranked songs based on their similarity scores and recommend the top-n similar songs for a given input track.

**5.6 Model Evaluation and Comparison:**

* Compared the performance of different models and techniques using cross-validation and holdout validation methods.
* Analyzed the strengths and weaknesses of each approach based on evaluation metrics and computational efficiency.
* Conducted statistical tests to determine significant differences in model performance.

**5.7 Results Interpretation and Conclusion:**

* Summarized the findings of the project, including the accuracy of genre classification models and the effectiveness of the recommendation system.
* Discussed the implications of the results in the context of music classification and recommendation systems.
* Provided recommendations for future research and improvements to the methodology.

# Observations and Findings

#### 6.1 Data Preprocessing and Feature Extraction:

* **Data Cleaning**: The GTZAN dataset was relatively clean, with no missing values or significant outliers identified during the initial inspection.
* **Feature Extraction**: Various audio features were successfully extracted from the audio signals, including time-domain features, frequency-domain features, and spectrogram-based features. These features captured the diverse characteristics of the audio signals across different genres.

#### 6.2 Exploratory Data Analysis (EDA):

* **Feature Distributions**: The distribution of audio features varied significantly across different genres. Certain features, such as spectral centroid and MFCCs, exhibited noticeable differences between genres, while others showed more overlap.
* **Feature Correlations**: Some features displayed strong correlations with each other, indicating redundancy in the feature set. PCA was employed to visualize the reduced feature space and identify clusters within the data.

#### 6.3 Model Training and Evaluation:

* **Performance Metrics**: The performance of various machine learning and deep learning models was evaluated using metrics such as accuracy, precision, recall, and F1-score. Random Forest and XGBoost consistently outperformed other models, achieving accuracies above 80%.
* **Ensemble Techniques**: Ensemble methods, including Voting and Stacking classifiers, yielded marginal improvements in accuracy compared to individual models, highlighting the benefits of combining diverse model predictions.

#### 6.4 Recommendation System:

* **Cosine Similarity Scores**: The recommendation system successfully calculated cosine similarity scores between audio features of different songs. Similar songs were identified based on their feature similarity, allowing users to discover new music based on their preferences.
* **Effectiveness**: The recommendation system demonstrated effectiveness in suggesting similar songs, enhancing user engagement and satisfaction with the music platform.

#### 6.5 Key Insights:

* **Feature Importance**: Certain audio features, such as MFCCs and spectral centroid, emerged as critical indicators for genre classification and similarity-based recommendation.
* **Model Robustness**: Despite variations in model architectures and hyperparameters, the performance of deep learning models plateaued around 70-80% accuracy, suggesting potential limitations in the dataset or feature representation.

# Limitations

**7.1 Data Quality and Quantity:**

* **Limited Dataset Size**: The GTZAN dataset, while widely used in music genre classification research, contains only 1000 audio files categorized into ten genres. The small dataset size may limit the generalization and robustness of the classification models, especially deep learning architectures.
* **Genre Label Accuracy**: The genre labels assigned to the audio files in the dataset may not always accurately represent the content of the songs. Subjective judgments and inconsistencies in genre classification could introduce noise and ambiguity into the training data.

**7.2 Feature Representation:**

* **Feature Extraction Complexity**: While a diverse set of audio features was extracted from the audio signals, the feature representation may not fully capture the nuanced characteristics of different genres. Certain subtle genre distinctions may be challenging to capture using traditional audio features, leading to potential classification errors.
* **Dimensionality Reduction Limitations**: Despite applying techniques like PCA for dimensionality reduction, high-dimensional feature spaces may still pose challenges in model training and interpretation. Balancing feature richness with computational complexity remains a trade-off in feature representation.

**7.3 Model Complexity and Performance:**

* **Overfitting**: Deep learning models, particularly those with complex architectures, are prone to overfitting, especially when trained on small datasets. Despite efforts to mitigate overfitting through regularization techniques like dropout and early stopping, model generalization may still be compromised.
* **Computational Resources**: Training deep neural networks on large-scale audio datasets requires substantial computational resources, including high-performance GPUs and memory-intensive processing units. Limited access to such resources constrained the scalability and efficiency of model training.

**7.4 Recommendation System:**

* **Limited Contextual Information**: The recommendation system based on cosine similarity relies solely on audio feature similarity for song recommendations. It does not consider contextual information such as user preferences, listening history, or social interactions, which are crucial for personalized recommendations.
* **Cold Start Problem**: The recommendation system may struggle to provide relevant recommendations for newly added songs or users with sparse interaction history. Addressing the cold start problem requires alternative approaches such as content-based filtering or collaborative filtering.

**7.5 External Factors:**

* **Genre Ambiguity**: Music genres are inherently subjective and dynamic, with songs often blending multiple genres or defying traditional categorization. The ambiguity of genre boundaries may challenge the accuracy and consistency of genre classification algorithms.
* **Cultural Bias**: The GTZAN dataset primarily consists of Western music, potentially introducing cultural bias in genre classification and recommendation. Models trained on biased datasets may exhibit skewed predictions and limited applicability across diverse cultural contexts.

# Conclusions and Future Work

**8.1 Conclusions:**

* **Genre Classification Performance**: The project successfully developed and evaluated machine learning and deep learning models for audio genre classification using the GTZAN dataset. Random Forest and XGBoost emerged as top-performing models, achieving accuracies above 80%. Ensemble techniques further enhanced classification accuracy, demonstrating the effectiveness of combining diverse model predictions.
* **Recommendation System Effectiveness**: The recommendation system based on cosine similarity effectively suggested similar songs to users based on audio feature similarity. The system demonstrated the potential to enhance user engagement and satisfaction with personalized music recommendations.
* **Insights Gained**: Through extensive experimentation and analysis, valuable insights were gained into the challenges and opportunities in music genre classification and recommendation systems. Key factors influencing model performance, feature importance, and limitations were identified, providing valuable guidance for future research and development.

**8.2 Future Work:**

* **Data Augmentation and Expansion**: Augmenting the existing dataset with additional audio samples and expanding genre diversity could improve model generalization and robustness. Incorporating datasets from different cultural backgrounds and music genres would address biases and broaden the applicability of the classification and recommendation systems.
* **Feature Engineering and Representation**: Exploring advanced feature extraction techniques, including deep audio embeddings and transfer learning, could capture more nuanced audio characteristics and enhance classification accuracy. Incorporating contextual features such as lyrics, artist metadata, and user preferences into the feature representation would enrich the recommendation system's relevance and personalization.
* **Model Optimization and Interpretability**: Further optimizing model architectures, hyperparameters, and regularization techniques to mitigate overfitting and improve generalization performance. Enhancing model interpretability through techniques like attention mechanisms and explainable AI would provide insights into model predictions and facilitate user trust and understanding.
* **Dynamic and Interactive Recommendation Systems**: Developing dynamic and interactive recommendation systems that adapt to user feedback, evolving preferences, and real-time trends. Integrating collaborative filtering, social network analysis, and reinforcement learning techniques would enable more personalized and engaging music recommendations tailored to individual users' preferences and contexts.
* **User Experience Enhancement**: Focusing on user experience design and interface usability to ensure seamless integration of genre classification and recommendation functionalities into music platforms. Conducting user studies and usability testing to gather feedback and iteratively refine the user interface and interaction design for optimal user engagement and satisfaction.

**8.3 Conclusion:**

* The project has laid a solid foundation for advancing the fields of music genre classification and recommendation systems, leveraging state-of-the-art machine learning and deep learning techniques. By addressing the identified limitations and exploring future research directions, the project aims to contribute to the development of more accurate, personalized, and user-centric music platforms that enrich the listening experience for users worldwide.

# Bibliography/References

[1] Tzanetakis, George, and Perry Cook. "Musical genre classification of audio signals." IEEE Transactions on Speech and Audio Processing 10.5 (2002): 293-302.

[2] Han, KyungHyun, et al. "Music genre classification using convolutional neural networks." 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014.

[3] McFee, Brian, et al. "librosa: Audio and music signal analysis in Python." Proceedings of the 14th Python in Science Conference. 2015.

[4] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of Machine Learning Research 12.Oct (2011): 2825-2830.

[5] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.

[6] Abadi, Martín, et al. "Tensorflow: A system for large-scale machine learning." 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16). 2016.

[7] Van Rossum, Guido, and Fred L. Drake. "Python reference manual." Network Theory Ltd. Bristol, UK 6 (2009).

[8] Lang, Kenny, and George D. Papadopoulos. "Evaluation of genre classification in music information retrieval." Journal of the American Society for Information Science and Technology 59.4 (2008): 625-633.

[9] Kim, Youngmoo E., and Justin Salamon. "Deep convolutional neural networks for predominant instrument recognition in polyphonic music." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018.

[10] Zhuang, Chengzhu, et al. "A comprehensive study on feature extraction methods for audio-based music classification." IEEE Transactions on Multimedia 17.8 (2015): 1430-1447.

# 10. Peer Review Form

|  |  |  |  |
| --- | --- | --- | --- |
| **Name of the student: (to be reviewed)** | Soumil Jainer | **Roll no. of the student:** | 102003114 |
| *This form has to be submitted by the student whose roll no. will be mentioned in the box above. Handover this to the panel at the time of final presentation.* | | | |
| Title of the project: | Audio Classification | | |
| Name of the company: | Banaras Hindu University, Varanasi | | |
| Project report  (Tick the appropriate) | Excellent | Good | Average |
| Project poster  (Tick the appropriate) | Excellent | Good | Average |
| Project video  (Tick the appropriate) | Excellent | Good | Average |
| Rate the work done | 0 – 10 points | *(Provide rating here)* → | 10 |
| Give marks to the student on the basis of the overall performance | 0 -5 marks | *(Provide marks here)* → | 5 |
| Abstract of the project (max. 100 words):  This project investigates music genre classification and audio retrieval using machine learning and deep learning techniques. The GTZAN dataset was used to train and evaluate various models. Traditional machine learning models like KNN, Random Forest, and XGBoost achieved good accuracy. Deep learning models built with Keras achieved even higher accuracy, with a custom callback function for early stopping. Cosine similarity was implemented for retrieving audio samples similar to a query, paving the way for music recommendation systems. The project establishes a baseline performance for the GTZAN dataset and demonstrates the potential of these techniques for audio classification and retrieval applications. | | | |
| Mention three strengths of the work done:  Strong Performance with Deep Learning: Project achieved high accuracy in music genre classification using deep learning models  Efficient Training with Early Stopping: The implementation of a custom callback function for early stopping during deep learning training demonstrates a focus on efficiency. This technique helps to prevent overfitting and reduces training time by stopping the process once a desired validation accuracy is reached.  Versatility with Cosine Similarity: Integrating cosine similarity for audio retrieval expands the project's scope beyond classification. This functionality allows for retrieving similar audio samples based on a query, opening doors for potential applications in music recommendation systems or personalized content discovery. | | | |
| Provide some useful recommendations (It may be some improvements, some suggestions to further raise the quality of the project):  Understanding Model Decisions:  Explore techniques like Layer-wise Relevance Propagation (LRP) to understand which features in the spectrograms contribute most to the model's predictions.  This can provide valuable insights into the learned patterns and improve trust in the model's results.  Expanding Project Scope:  Explored music genre classification and recommendation systems. Consider potential applications in other domains:  Environmental sound classification (e.g., identifying birds or traffic noise)  Audio fingerprinting for copyright protection  Anomaly detection in audio streams (e.g., gunshots or machinery malfunctions) | | | |
| Name of the **evaluator**  student: | Gurpreet Singh | Roll no. of the  **evaluator** student: | 102003070 |
| Signature of the  **Evaluator** student: |  | | |