Received: Added at production | Revised: Added at production | Accepted: Added at production
DOI: xxx/xxxx

ORIGINAL ARTICLE

Hindi Podcast Genre Prediction using Support Vector Classifier

Mudeet Jain[1] | Mehul Mahrishi[2,1] | Girish Sharma[3] | Samira Hosseini[2]

[1] Swami Keshvanand Institute of Technology,Management & Gramothan, Rajasthan, India
[2] Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Monterrey, NL 64849, Mexico
[3] Manipal University Jaipur, Rajasthan, India

Correspondence
*Mehul Mahrishi, Email: mehul@skit.ac.in

Abstract

India experienced a 23% rise in podcast listening after the Covid-19 pandemic. The pandemic and screen fatigue led people to seek their favorite old audio podcasts. Podcast genre classification allows listeners to compile a playlist of their favorite tracks; it also helps podcast streaming services provide recommendations to users based on the genre of the podcasts they enjoy. Since the COVID-19 pandemic, the need for educational content in all forms, including podcasts, has skyrocketed, making it even more crucial to anticipate the genre of educational podcasts. Educational podcasts are a sub-genre of the broader education genre and typically involve audio recordings of discussions, lectures, or interviews on educational topics. Education podcast genre prediction is required to efficiently classify and arrange educational content and make it simpler for listeners to access and absorb pertinent information. This study focuses on Podcast Genre Prediction, specifically for the Hindi language. In this study, our developed PodGen dataset was used, which consists of 550 five-minute podcasts with 26,867 sentences, where every podcast was manually annotated into one of the four genre categories (Horror, Motivational, Crime, and Romance). The performance comparison of state-of-the-art machine learning techniques on the PodGen dataset was used to demonstrate accuracy. The best performance on testing data was observed in the Support Vector Classifier model with balanced accuracy:82.42%, precision (weighted):83.09%, recall (weighted):82.42%, and F1 score(weighted):82.39%.

KEYWORDS:
Educational Innovation, Genre Classification, Hindi Language, Machine Learning, Podcast, Support Vector Classifier, Text Analysis

1 | INTRODUCTION

A podcast is a collection of recorded audio files released in episodes and available for download online [1,2]. They are the new fuel propelling the content business, offering a simple way to consume all information, from news to entertainment. This simplicity may have aided its meteoric ascent worldwide, with more and more people tuning in. With the help of podcasts, people and organizations may now engage in innovative activities and offer listeners new kinds of audio content. It is now used to support medical education, and libraries are gradually adopting it to create free downloadable audio content and reach out to users, even when they are far from the library building and occupied with other activities. Podcasts can be an intimate bridging medium, a way of communicating, and can support learning. Not only are the characteristics of podcast media utilized to convey narrative but they are also employed to influence and define the storytelling technique. Tourism organizations may create numerous hours of audio content affordably; before their trip, tourists can download customized information from websites and bring the audio devices storing the material to their travel destinations [3,4,5,6].

---

[0] **Abbreviations:** SVC, Support Vector Classifier; KNN, K-Nearest Neighbour; CNN, Convolutional Neural Networks; TF-IDF, Term Frequency- Inverse Document Frequency

. Despite Hindi being one of the most widespread languages worldwide and the most spoken in India, fewer podcasts are available in Hindi than in English. Because of the multilingual talent provided by websites, Hindi has significantly increased in value in various industries, including information technology, during the past ten years. Researchers have generally attempted to concentrate on English text processing, despite Hindi text processing's relevance in the information technology and digital spheres. Researchers have not worked much on processing Hindi text, possibly because insufficient resources are available. The task of stemming and preparing Hindi text is challenging[7]. Thirty-three consonants, ten numbers, and 13 vowels make up the Devanagari-written language of Hindi. It is distinctive and different from others due to the numerous conjunctions and combinations, which also require high concentration and critical thinking. It contains a vocabulary and a set of grammatical rules that are defined and organized[8].

. Recognizing emotions in Hindi text can help create systems serving many users. Hindi text emotion recognition can aid in developing powerful natural language processing tools that can work with a wide range of languages and dialects. The knowledge gained from researching Hindi text emotion recognition can be used to improve cross-lingual emotion identification systems in other languages. The identification of emotions in Hindi literature may assist in creating aids for language acquisition. For instance, the tool can give feedback and enhance grammar and language skills by identifying the emotions portrayed in a sentence. Businesses may increase consumer engagement and offer a more individualized experience by comprehending the emotions in Hindi text. In mental health, emotion recognition can be used to spot people who need support and intervention. It aids in understanding the feelings and sentiments of people during protests, natural calamities, and other occurrences. Authorities can act quickly and make wiser decisions by examining the emotional state of social media users during these events. Based on commonalities in the story components or viewers' emotional reactions, the "genre" categorizes a film[9]. As many researchers explain, subject experts create a film, and the genre is a trustworthy content indicator for the film 10. Movie genres are crucial aspects of films since they can be used as tags for targeted searches on multimedia platforms and to create individualized movie recommendations[10]. The categories to describe music genres are used to personalize the music. The feeling of rhythm and texture could be considered a quality of the specific musical structure for categorizing music into a genre[11]. Genre is an abstract yet characteristic feature of music[12]. The genre of a film is an essential indicator of its central themes and serves many functions. This factor heavily impacts moviegoers' preferences[13]. Accurate movie genre classification has been added to the recommended systems[14]. Labels for musical genres help classify songs, records, and artists into larger groups that share musical traits. Music classification has frequently used musical genres, from physical music shops to streaming platforms[15]. The music genre and the mood are closely related. People often listen to specific genres to regulate their moods[16].

Since the genre is such important, the classification process is needed to classify and predict different classes, which can be helpful for the users. Genre classification is the method of grouping related elements from the music's management data, such as aesthetics, common themes, or goals. It is used along with the classification of topics[11]. Music Genre Classification or grouping of music into various categories or genres is a concept that helps the masses to differentiate between two genres based on their composition or the beats they withhold. As more and more musical subgenres appear worldwide, classifying music by genre has recently gained much popularity. The idea of categorizing music genres using machine learning has recently gained attention. Although music genres have been known for a long time, machines can now categorize them in the contemporary world where everyone listens to music. Genre prediction in podcasts is relatively new; therefore, there are scarce datasets.

The popularity of educational podcasts among academicians has increased recently[17]. "educational genre" refers to written or visual content created to impart knowledge or skills in a specific field. Textbooks, training videos, online learning, and educational games are examples of academic genres. It is possible to determine the specific genre of an educational resource based on elements like format, target audience, and content focus. A sub-genre of the giant education genre, educational podcasts frequently feature audio recordings of talks, lectures, or interviews about learning areas.

Here are some popular podcast categories for education: These podcasts are primarily on teaching and learning methodologies, industry standards, and advice for teachers and students. Higher education: These podcasts cover subjects like admissions, financial aid, and student life relevant to schools and universities. K–12 education: These podcasts include curriculum, policy, teacher preparation, current concerns, and trends in elementary and secondary education. Professional development: These provide guidance and ideas on starting and developing a career, including networking and job-searching techniques. Individual improvement: These podcasts emphasize self-improvement and personal development, covering subjects such as productivity, mindfulness, and goal-setting.

Although various studies have been done on the subject, significant research gaps still need to be filled. Podcast genre prediction is a relatively young area of study. Some focus on the scarcity of datasets available for predicting the genre of podcasts; most are brief and don't cover all of them, which makes it difficult to train and test models accurately. Most studies on predicting the genre of podcasts rely on audio elements like mel frequency cepstral coefficients (MFCCs), but these variables don't capture the podcast's semantics or content. Further study on feature engineering is required, particularly regarding using machine learning algorithms to extract useful features from podcast transcripts. Cross-lingual genre prediction research is necessary for algorithms to predict the genre of podcasts in several languages. These study gaps demonstrate that predicting the genre of podcasts is a difficult and developing topic that needs further research. Filling in these gaps contributes to creating more precise and efficient algorithms for predicting podcast genres.

To train several machine learning models to accurately predict the genre of Hindi podcasts, we created our dataset called PodGen. The major phases of this study are finding and downloading podcasts, converting them into Hindi text files, creating a dataset, cleaning the data, training machine learning models, and finally, determining the podcast genre. The current study is novel because it uses ML models for performance evaluation and podcast genre prediction on a corpus of Hindi podcasts.

The remainder of the paper is outlined thus: The relevant work in genre classification is covered in Section 4, our dataset and its creation are explained in Section 5, and Section 6 focuses on our suggested framework. Section 8 presents the findings from our experiments, and Section 9 concludes the paper.

2 | MOTIVATION

The current boom in podcast popularity motivated this research. A podcast is a vast collection of recordings that one can browse, download, and listen to conveniently. It can pertain to a person's favorite blogs, shows, and topics and be enjoyed anywhere: in the vehicle, at work, at home, or while exercising. As a result, a sizable amount of podcasts feature user opinions that are rich in content. There is a rising demand for curated and customized content in Indian languages, including Hindi, as India's podcast listeners continue to increase.

Since most research is done only in English, creating a framework for predicting the genre of Hindi podcasts is necessary. Thus, genre prediction in non-English languages, primarily regional and local, poses an important task and opportunity. This framework could open the door to further improvements in natural language processing methods and machine learning in Hindi. It provides a foundation for future research on podcast genre prediction in Indian languages. Therefore, the research potential to enhance content discovery, recommendations, market research, and audience analytics in natural language processing for Indian languages motivates the researchers to study Hindi podcast genre prediction.

3 | RESEARCH CONTRIBUTIONS & OBJECTIVES

The contributions of this research resulted from achieving the following objectives:

- • The PodGen dataset we created includes 550 podcasts in four genre classes: Horror, Motivational, Romance, and Crime.
- • Hindi Text Classification using a Support Vector Classifier with Word Embedding Techniques.
- • Our novel approach for predicting genres of Hindi Podcasts using a Support Vector Classifier makes it simple to determine the genre of podcasts.
- • The study opens research dimensions and provides a comparative benchmark for further research to produce a significant impact.

4 | LITERATURE REVIEW

Over time, several techniques have been put forward to create Genre Classification models. There is no one-size-fits-all technique because many data types and use cases are available. However, most suggested strategies can be categorized as identifying film, music, website, book, and video genres. The majority of the work has been done on datasets based on the English language, with only a tiny portion of the classification work done on Hindi language datasets. Research done on genre classification using English databases is discussed in this section.

4.1 | Hindi Text Classification

Bafna et al.[7] proposed utilizing Concept Learning Algorithms, a Hindi Verse Class Predictor. They contrasted lazy machine learning techniques for categorizing the corpus of 565 Hindi poetry into four categories. The lazy machine learning methods are k-nearest neighbors (KNN) and linear regression. In their work, the authors collected and prepared the corpus, consisting of 565 Hindi verses. The model was then trained and evaluated. KNN outperformed linear regression. KNN is applied to corpora of various sizes. A training-to-testing data ratio of 8:2 was used, with 452 poems provided as the training data. It was noted that for k=8, 0.97 was the maximum accuracy achieved.

Sindhu et al.[18] focused on sentiment classification in Hindi text. The proposed research activity tried to test a strategy independent of the availability of dictionaries for different languages. To achieve this, they generated corresponding numerical data to the text., The model also combined recurrent neural and convolutional neural network models with a 1 Dimensional Convolutional Neural Network (1D CNN) to extract the subjective information from the provided dataset of movie reviews, which contained roughly 3000 positive and 3000 negative Tweets. The auxiliary tags and mentions were initially deleted from the data, which helped standardize the data and prevented unnecessary input that could reduce the effectiveness of the neural network. Ambiguous text was turned into numerical data to eliminate it. These patterns are independent of word length or word structure. Adding a vector model raised the model's accuracy by a range of +6.72 to +10.78.

Bafna et al.[19] worked on Hindi poetry classification using eager supervised machine learning Algorithms. On a corpus of 450 Hindi poems that are roughly classed as "Bal geet," "Updesh geet," and "bhajans," two enthusiastic machine learning techniques were deployed. A misclassification error was used to assess the classifiers. The initial stage of the methodology was cleaning and preprocessing using several techniques, followed by TF-Idf, which was used to extract the essential terms from the processed corpus. Select tokens were subjected to naive byes and random forest, and classifier models were created. Random Forest has been found to perform better than Naive Byes. Without any other method that makes predictions on the Hindi corpus, the misclassification error comparison between the random forest and naive byes classifiers proved the methodology's superiority.

Named entity recognition (NER) for Hindi and Marathi, two low-resource Indian languages model, was introduced by Litake et al.[20]. For NER tasks, transformer-based models are frequently employed. Based on publicly accessible Hindi and Marathi NER datasets, each having nearly 12000 sentences, the authors evaluated various BERT variants, including baseBERT, RoBERTa, and AlBERT. The monolingual MahaRoBERTa model performed well for Marathi NER, whereas the multilingual XLM RoBERTa model outperformed Hindi NER. They also demonstrated that monolingual instruction does not always guarantee better performance. Hindi monolingual models did not perform along with Marathi models. In addition, Hindi NER datasets generalized the mahaBERT models successfully.

Puri et al.[8] worked on a Hindi Text Classification model named HTC–SVM. It takes a set of known 4 Hindi documents of two classes—documents with simple Hindi statements that have been gathered from various Hindi blogs and official websites, preprocesses them at the document, sentence, and word levels, extracts features, and trains an SVM classifier, which then further categorizes a set of unknown Hindi documents. Hindi's extensive range of potential conjuncts, letter combinations, sentence structure, and multisense words make this text classification more difficult. The authors used a set of Hindi documents accurately identified using SVM for the experiments.

A model proposed by[21] classified musical moods based on Latent Dirichlet Allocation lyrical analysis of Hindi songs. The significant elements used for analysis were term frequency and unigrams to determine the listener's mood from song lyrics in Hindi. The songs were trimmed down; the topic-modeling Latent Dirichlet Allocation model extracted the mood from each song in the 1,894 ghazals, bhajans, and Bollywood songs collection. The findings suggested that 530 songs in the annotated dataset fell into Class S, which was the sad class, and that 1530 songs were positive and 370 portrayed negative feelings.

4.2 | Music Genre Classification

Ozakar et al.[22] introduced seven innovative high-level characteristics derived from song structures tested using CNN and a voting classifier to see how well they performed. Four different song structures were generated, containing 2,786 full-length songs and preprocessed. CNN achieved 44% accuracy, while the Voting classifier attained 49%.

. An RNN-based model for genre categorization that is trained on brief clips of only three seconds in length was proposed by Bisharad et al.[12]. The dataset was the benchmark GTZAN dataset, which consisted of 1000 music clips. For the top-1, top-2, and top-3 genre class predictions, the model had 82%, 91%, and 94.5% accuracy rates, respectively.

Leartpantulak et al.[11] conducted investigations using the stacking ensemble method to enhance prediction. KNN, DT, Random Forest, SVM, and NB classifiers comprised the base classifier. It was the best classification model (wrapper approach) with an accuracy value of 88.33%, compared to SVM's 85%, DT's accuracy value of 84.1%, and Random forest's accuracy value of 80.83%.

Pothina et al.[23] successfully predicted the movie and music genre identification using a heterogeneous ensemble technique and feature selection algorithms based on viewers' feelings when watching media clips. Deaf and Decaf datasets contained the magnetoencephalographic (MEG) data of 30 individuals and the electroencephalographic (EEG) data as they viewed movie and music clips.

Oramas et al.[15] developed a method for learning and combining multimodal data representations to classify music genres. In a single-label experiment, it could be seen how visual features outperformed audio features in some classes. Saari et al.[16] explored whether considering the genre is advantageous for automatic music mood annotation.

4.3 | Educational Podcasts and Genre Classification

The three educational podcast genres explored by Drew et al.[17] have been used to illustrate and forecast how podcast genres might support and improve student learning in e-learning contexts. For learners outside of formal institutions, short "Quick Burst" podcasts, in-depth "Narrative" podcasts, and interactive "Chat Show" podcasts have all established themselves as popular educational materials for learning and interacting with communities of like-minded learners. The discussion has revealed that the authors made predictions based on the genres' "moves" and "steps." Every podcast genre has a particular move and step that can identify its genre.

The vast volume of published podcasts and the wide range in quality and reputation can challenge listeners seeking to find excellent podcasts. Independent of the information they include, some podcasts are enjoyable to listen to, while others are uninteresting. A paradigm for analyzing listener appeal was developed and applied to automatically anticipate users' listening preferences by Tsagkias et al.[24] for PodCred. They created the framework to assist automated prediction of listener preference for a particular podcast. It comprised four indications relating to the technical execution of the podcast as well as the podcaster, context, and content. A rudimentary categorization system was implemented by converting these signs into quickly extractable surface attributes. The evaluation trials showed that basic surface features generated from the PodCred framework effectively categorized podcasts using popularity levels in iTunes as the ground truth.

To conduct a scoping review, author[25] searched the English-language databases of PubMed and Embase in June and July 2020 for studies of audio-only medical education podcasts in undergraduate, graduate, and continuing medical education. Data on descriptive outcomes such as podcast use, subject matter, structure, and educational results categorized using Kirkpatrick's four levels of evaluation were taken from the included studies by the authors. According to an analysis of this research, podcast usage has grown over time. They are a top resource for resident education and are now being introduced into the official medical curriculum. According to the 29 investigations that evaluated learners' attitudes and reactions to podcasts, listeners valued them for their mobility, effectiveness, and worth in entertainment and education. The ten studies that evaluated knowledge retention found that podcasts were comparable to conventional teaching strategies. The 11 studies that evaluated behavior change found that medical students with better documentation skills and residents and practicing doctors with self-reported practices changed after listening to podcasts.

4.4 | Movies and Video Genre Classification

Roy et al.[10] proposed that a new movie genre should address this issue using hybrid filtering. GAHF and GSHF exhibit lower MAEs than item-item collaborative filtering by 18% and 23%, respectively.

Kundalia et al.,[14] suggested a unique technique for employing neural networks with knowledge transfer learning to predict the multi-label movie genre from the movie's poster. They created a new dataset with over 30,000 images from IMDb, including 12 movie genres.

Radhika et al.[26] proposed a framework using Resnet-152. The In-depth visual characteristics are extracted from video frames using an LSTM model trained on the ImageNet dataset to capture the temporal dynamics of the visual information.

Yadav et al.[27] proposed a unique, deep affect-based modeling approach for movie trailers. ILDNet was trained using the EmoGDB dataset, which consists of 100 Bollywood movie trailers annotated with six different types of emotions and considered major movie genres.

Yu et al.[28] proposed ASTS, an attention-based, spatiotemporal, sequential approach to categorize movie trailers' genres. 14,415 annotated movie trailers were added to the MovieLens dataset. There are 1,128 different movie genres in total.

Wi et al.[13] explored using a Gram layer in a CNN and extracted the ideal details and traits from movie posters to help classify films into genres. 2,664 movie film posters with 12 multi-genres comprised the dataset in the experiment.

Bouyahi et al.[29] presented a new approach to video segmentation in scenes based on genre prediction. They finally achieved successful performances in videos with various genres using RAI and BBC datasets.

5 | THE PODGEN DATASET

The critical issue for our research was the lack of a Hindi podcast corpus. As a result, we formed the "PodGen dataset" containing information like filename, content (including transcripts of Hindi podcasts), and 550 Hindi podcasts' genre classifications. Each episode has a duration of five minutes. The dataset features 550 rows with data from 550 podcasts. It is organized using the 3 Filename, Content, and Annotation parameters in 4 Genre classes: Horror, Motivational, Crime, and Romance, as shown in figure 1. The dataset aims to build a brand-new database of Hindi podcasts, analyze it using machine learning techniques, and predict the genre based on the data. There are 107 podcasts in the crime genre, 149 in the horror genre, 129 in the motivational genre, and 165 in the romance genre. A .csv file is saved to disk for every genre, containing all transcripts, title names, and label markings. Labels are annotated manually, corresponding to their genre class. The numbers 1, 2, 3, and 4 correspond to horror, motivational, crime, and romance genres.

Figure 2 displays the dataset. More than three lakhs Hindi words were in our dataset, then randomly shuffled. After this, the data was cleaned using tokenization, punctuation mark removal, pronoun removal, and stop word removal. The number of irrelevant words was reduced by the data-cleaning method. The number of Hindi sentences was 26,867 across all genre classes. The dataset was in the ratio of 70:30, split into test and train datasets for the training and testing of various classifiers.

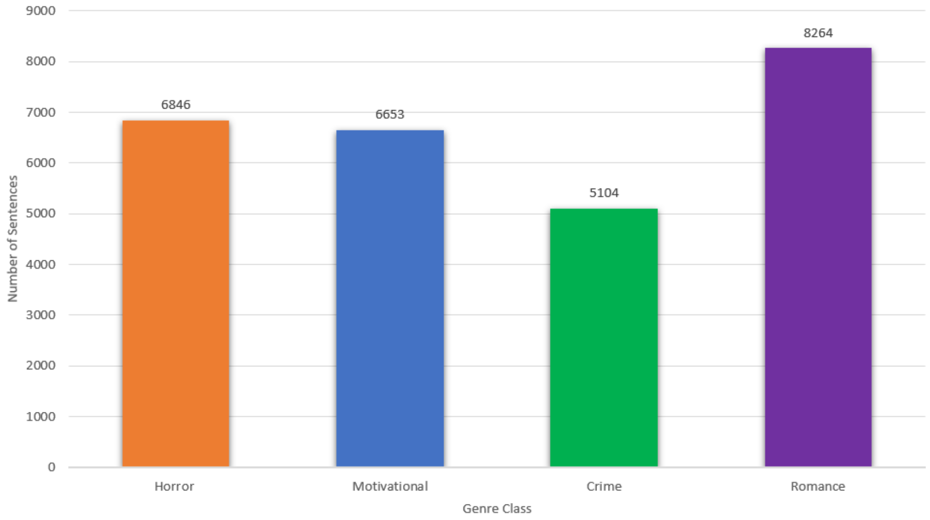
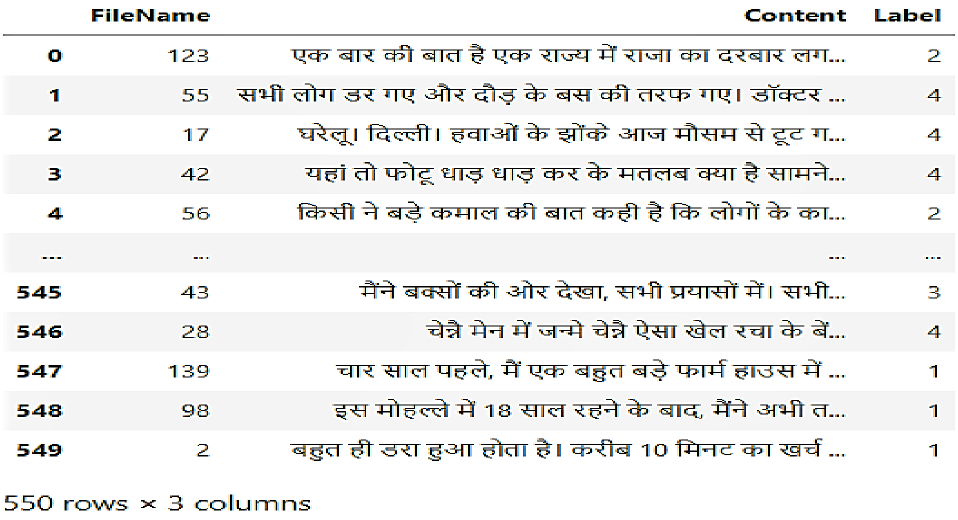


Figure 1 Number of Sentences in Each Genre Class



	FileName	Content	Label
0	123	एक बार की बात है एक राज्य में राजा का दरबार लग...	2
1	55	सभी लोग डर गए और दौड़ के बस की तरफ गए। डॉक्टर ...	4
2	17	घरेलू। दिल्ली। हवाओं के झोंके आज मौसम से टूट ग...	4
3	42	यहां तो फोटू धाड़ धाड़ कर के मतलब क्या है सामने...	4
4	56	किसी ने बड़े कमाल की बात कही है कि लोगों के का...	2
...	...	...	...
545	43	मैंने बक्सों की ओर देखा, सभी प्रयासों में। सभी...	3
546	28	चेन्नै मेन में जन्मे चेन्नै ऐसा खेल रचा के बें...	4
547	139	चार साल पहले, मैं एक बहुत बड़े फार्म हाउस में ...	1
548	98	इस मोहल्ले में 18 साल रहने के बाद, मैंने अभी त...	1
549	2	बहुत ही डरा हुआ होता है। करीब 10 मिनट का खर्च ...	1

550 rows × 3 columns

Figure 2 Podgen Dataset

6 | GENRE PREDICTION FRAMEWORK

The technique of identifying a text, piece of music, or other content's genre or category is known as genre prediction. Several methods can be used, including machine learning, text analysis, and audio analysis. Genre prediction aims to classify information into pre-defined categories automatically, which promotes easy search and indexing. Numerous applications, including music streaming services, TV program and movie recommendation systems, and bookstores, can benefit from this. This study focuses on podcasts and their prediction, particularly for Hindi. The two media aspects considered in the study were speech and text. The iNLTK library that supports Indian languages was used for natural language processing. Figure 3 depicts the critical stages for genre prediction.

- **Transcript Extraction:** The first step is downloading Hindi podcasts from various websites such as Spotify, Gaana, Saavan, and YouTube. Hindi Podcasts are very few compared to podcasts in English. Four different genres of podcasts are chosen for the model. The four genre categories are Horror, Motivational, Crime, and Romance. The downloaded podcasts are manually classified into four folders corresponding to their genres. Since the podcasts are very long, they are shortened to 5 minutes each to improve the performance of the API and algorithms.

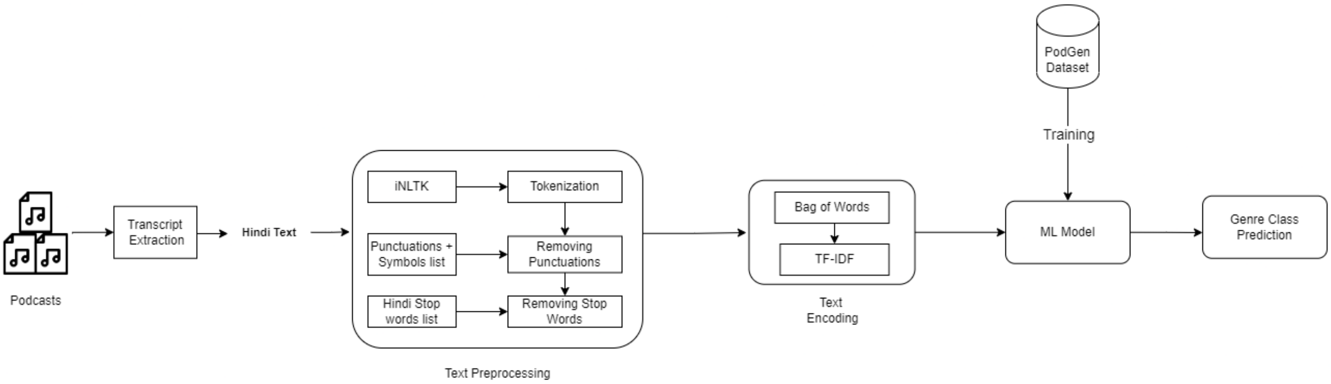


Figure 3 Genre Prediction Framework

The next step after gathering podcasts is to use an API to convert audio files to text files. Compared to other APIs like Google's API, which has a transcription rate of approximately 70%, this API is about 90% acceptable for the Hindi language.

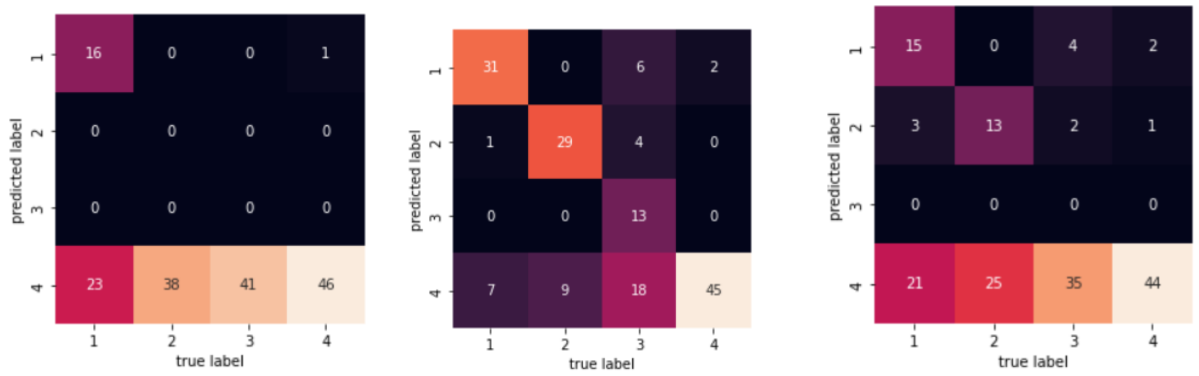
- **text Preprocessing:** Preprocessing the data after collection is essential, including text cleaning, preparation, annotation, and adding helpful computer-readable tags. For text preprocessing, the iNLTK library was used. The essential stages are tokenization, punctuation mark removal, pronoun removal, and stop word removal. The following steps were used for the preprocessing:
  - **Tokenization:** Tokenization is a typical activity in natural language processing(NLP). Advanced deep learning-based systems like Transformers and traditional NLP techniques like the Count Vectorizer utilize it. Tokenization breaks apart an extracted sentence from a text into tokens. In this context, tokens can be letters, subwords, or words. Thus, there are three categories of tokenization: character, n-gram characters (subword), and word. The token is the most popular method of processing raw text because it is the fundamental building block of Natural Language. Additionally, transformer-based approaches tokenize raw data. The same holds for the most popular deep learning techniques, including LSTM, RNN, and GRU.
  - **Removing punctuations:** When working with text in Natural Language Processing, text cleaning, often called text preprocessing, is crucial. Before providing this type of noisy text input to the machine learning model, one must clean it because it comprises a range of words with poor spelling, short words, special symbols, emojis, and other noise. Several techniques must be applied to clean the data. In this instance, we dealt with special punctuation or characters in the text data:,! | $ () * % @. While working with NLP, we frequently employ word embedding programs like GloVe, fastText, and others. Punctuation is handled in various ways throughout the entire word-embedding matrix.
  - **Removing stop words:** Stop words are frequently used in languages. Hindi stop words include "ka," "se," "ki," and others. In text mining and NLP, stop words are typically used to remove keywords so commonly used that they do not provide much valuable information. Stop words frequently replace words with little information but can also add context. Additionally, while using a published list of stop phrases is relatively simple, they are often insufficient. A stop word list that was publicly available for the Hindi language was used. Popular pronouns and nouns were also eliminated because they were not required for the emotion analysis. The list was extracted from the corpora of the Hindi Wordnet developed and managed by IIT Bombay[30].
- **NLP Techniques used for Data Preprocessing:** A combination of iNLTK (the Indian Natural Language Processing Toolkit) and Stanford NLP (the Natural Language Processing Toolkit) was used to preprocess Hindi data. We used the inLTK collection of tools. Indian languages have their version of Python's popular NLTK package, the iNLTK library. Any features a programmer might want to implement in an NLP app can be found in this library. Most of the qualities needed for modern NLP activities, such as creating a vector encoding for the input sentence, sentence similarity, and tokenization, are supplied by iNLTK in a straightforward and intuitive API interface.
- **Text Encoding:** Text encoding converts meaningful text into numerical or vector form while maintaining the relationship between words and sentences. This allows a machine to comprehend the pattern associated with any text and to recognize the context of sentences. This procedure uses Term Frequency-Inverse Document Frequency (TF-IDF Encoding) and Bag of Words techniques.

- **Machine Learning Model:** The research uses the PodGen dataset to train and evaluate a variety of ML models, which are Naive Bayes, Logistic Regression, Support Vector Classification(SVC), Decision Trees, K - nearest neighbors (KNN), Random Forest, Boosted Trees, and the Zero-Shot classifier.
- **Genre Prediction:** The genre prediction job signifies podcasts' genre classification. The classifiers' effectiveness was assessed using various parameters, including accuracy, precision, recall, and F1. When comparing different classification strategies, accuracy is often taken into account. How often predictions turn out to be accurate is called "precision." In statistics, recall measures how many true positives are recognized. One measure of classification accuracy is a recall or the percentage of class instances correctly identified. Essentially, it is the midpoint between recall and accuracy. Both false positives and false negatives are considered.

7 | PERFORMANCE OF ML CLASSIFIERS FOR GENRE PREDICTION

The experimental setup included a central processing unit (CPU) with a 4 GHz clock speed, a graphics processing unit (GPU) with 16 GB of memory, and 16 GB of random access memory (RAM). The research's code was written in Python, and the platforms used were Google Colab and Jupyter Notebooks. Whereas Jupyter Notebooks is a desktop tool that enables users to execute Python code on their local workstations, Google Colab is an online platform that allows users to run Python code and ML algorithms on Google servers with GPUs for quick processing of the algorithms. The PodGen dataset is a novel dataset created to compare genre prediction in Hindi. Therefore, the results and discussion are limited to this domain only. All the results depicted are in terms of the improvement of various machine learning algorithms applied to the PodGen dataset. We considered eight Machine Learning algorithms: Naive Bayes Classifier, Logistic Regression Classifier, Decision Trees Classifier, Boosted trees Classifiers, Random Forest Classifiers, KNN Classifiers, Zero-Shot Classifier, and Support Vector Classifier.

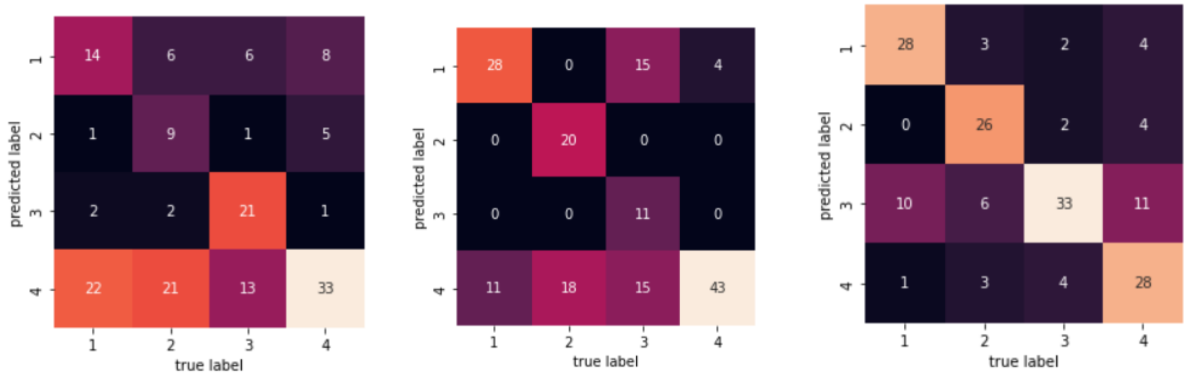
1. **Naive Bayes Classifier:** A classifier is a machine learning model that distinguishes different items using a set of attributes. A probabilistic machine learning model called a Naive Bayes classifier is used to solve classification problems. The classifier's foundation is the Bayes theorem. Figure 4a shows the confusion matrix of this model on the test data. We noted that the model correctly predicts only a few classes of the romance genre and can only forecast the horror genre with the greatest degree of accuracy. Accuracy:37.57 %, Precision(weighted):31.09 %, Recall(weighted):37.57 %, and F1 score(weighted):26.94 %.
2. **Logistic Regression Classifier:** Calculating the likelihood of a particular class or occurrence can be done using the supervised machine learning technique known as logistic regression. This method is applied when the data can be linearly separated, and the outcome is binary or dichotomous. By default, logical regression is only applicable to two-class classification problems. Figure 4b shows the confusion matrix of this model on the test data. We observed that all the podcasts in the Crime genre were predicted correctly, and many podcasts in the Romance genre were not predicted correctly. Accuracy:71.51%, Precision(weighted):79.50%, Recall(weighted):71.51%, and F1 score(weighted):69.65%.
3. **Decision Trees Classifier:** A supervised machine learning technique called a decision tree employs rules to make decisions as people do. A machine learning classification algorithm does have some judgment-making capabilities. Figure 4c shows the confusion matrix of this model on the test data, indicating it performed worst for the Crime genre as no podcast genre was predicted, and relatively good performance for the Horror and Motivational genres. Accuracy:43.63%, Precision (weighted):42.66%, Recall (weighted):43.63%, and F1 score (weighted):36.89%.
4. **Boosted Trees Classifier:** Boosting is a method for transforming many weak classifiers (trees) into strong ones. It employs boosted decision trees, where each tree is constructed iteratively. The weight assigned to the tree's output is then determined by its accuracy. According to the confusion matrix figure 4d, the Crime genre predicted the most podcasts, while the remaining three genre classes were not well predicted. Accuracy:46.66%. Precision (weighted):53.31%, Recall (weighted):46.66%, and F1 score (weighted):46.14%.
5. **Random Forest Classifier:** The random forest model makes predictions by additively merging the results of several base models. In this instance, each base classifier is a straightforward decision tree. Figure 4e shows the confusion matrix of this model on the test data, revealing that the Crime genre was predicted with great accuracy as it predicted 21 classes correctly, and five podcasts were not predicted correctly. Accuracy:61.81%, Precision (weighted):76.03%, Recall (weighted):61.81%, and F1 score (weighted):60.06%.
6. **KNN Classifier:** The nearest neighbors classifier predicts a data point's class as the most prevalent class among that point's neighbors. Careful thought and subject-matter knowledge are needed when defining the neighborhood criterion for a prediction data point. The size of " neighborhoods" must be defined concerning a distance function measuring the distance between data points. The best-predicted



(a) Naive Bayes Classifier

(b) Logistic Regression Classifier

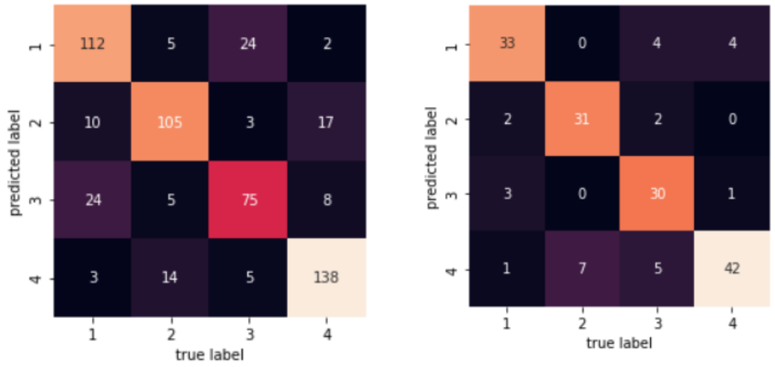
(c) Decision Trees Classifier



(d) Boosted Trees Classifier

(e) Random Forest Classifier

(f) KNN Classifier



(g) Zero-Shot Classifier

(h) Support Vector Classifier

Figure 4 Confusion matrix of state-of-the-art Machine Learning Classifiers for Genre Prediction

classes were the Motivational and the Romance, according to Figure 4f, showing the confusion matrix of this model on the test data. Accuracy:69.69%, Precision(weighted): 72.42%, Recall(weighted): 69.69%, and F1 score(weighted): 69.98%.

7. **Zero-Shot Classifier:** Zero-shot learning (ZSL) has historically been used to describe a rather specific task: training a classifier on one set of labels and then evaluating it on a different set of labels that it has never seen before. ZSL involves a machine learning problem scenario in which a learner observes samples from classes not observed during model training and predicts the category to which they belong. Figure 4g shows the confusion matrix of this model on the test data. Accuracy: 78.18%, Precision (weighted): 78.36%, Recall (weighted): 78.18%, and F1 score (weighted): 78.24%.

8. **Support Vector Classifier:** SVC has a reasonably high accuracy compared to other classifiers like logistic regression and decision trees. It is well known for its kernel approach to handling nonlinear input spaces. SVC assists in categorization by determining the ideal hyperplane for new data points. Figure 4h shows the confusion matrix of this model on the test data. The confusion matrix indicates that all the classes were predicted correctly. Accuracy:82.42%, Precision (weighted):83.09%, Recall (weighted):82.42%, and F1 score (weighted):82.39%. According

to the Heat map in Figure 4, the confusion matrix shows that 33 podcasts in the Horror genre were predicted correctly, and four podcasts were expected in the crime and romance genres each. Concerning the motivational genre, 31 podcasts were correctly predicted, and four were wrong. In the prediction of the crime genre, 30 were correctly predicted while four were not; lastly, the romance genre indicated 42 correct podcasts, and 13 were expected wrong.

8 | RESULTS & DISCUSSIONS

The two classifiers with the worst performance among the models trained and evaluated on our PodGen dataset were Decision Tree and Naive Bayes, as per the experimental results shown in Table 1. There are many possible reasons for this poor prediction, but one may be the number of words in a single paragraph; they may function effectively for a sentence line with only a few words. Logistic regression is helpful for this as it uses a sigmoid function. Because of factors including data sparsity, continuous word representations like word embeddings, and a lack of feature selection, irrelevant features are included, or if the feature set has become too large for the dataset, the Naive Bayes algorithm does not perform well. Logistic regression uses maximum likelihood estimation (MLE) to obtain the model coefficients that relate predictors to the target. Decision trees suffer badly in such high-dimensional feature spaces. Decision trees directly partition the sample space at each node. As the sample space increases, the distances between data points increase, making finding a ”good” split much harder. The Random Forest (RF) classifiers are suitable

Table 1 Comparison Table for Various Models

S.No.	Model Name	Accuracy	Precision(W.)	Recall(W.)	F1(W.)
1	Naive Bayes	37.57%	31.09%	37.57%	26.94%
2	Logistic Regression	71.51%	79.50%	71.51%	69.65%
3	Decision Trees	43.63%	42.66%	43.63%	36.89%
4	Boosted Trees	46.66%	53.31%	46.66%	46.14%
5	Random Forest	61.81%	53.31%	61.81%	60.06%
6	KNN Classifier	69.69%	72.42%	69.69%	69.98%
7	Zero-Shot	78.18%	78.36%	78.18%	78.24%
8	**SVC**	**82.42%**	**83.09%**	**82.42%**	**82.39%**

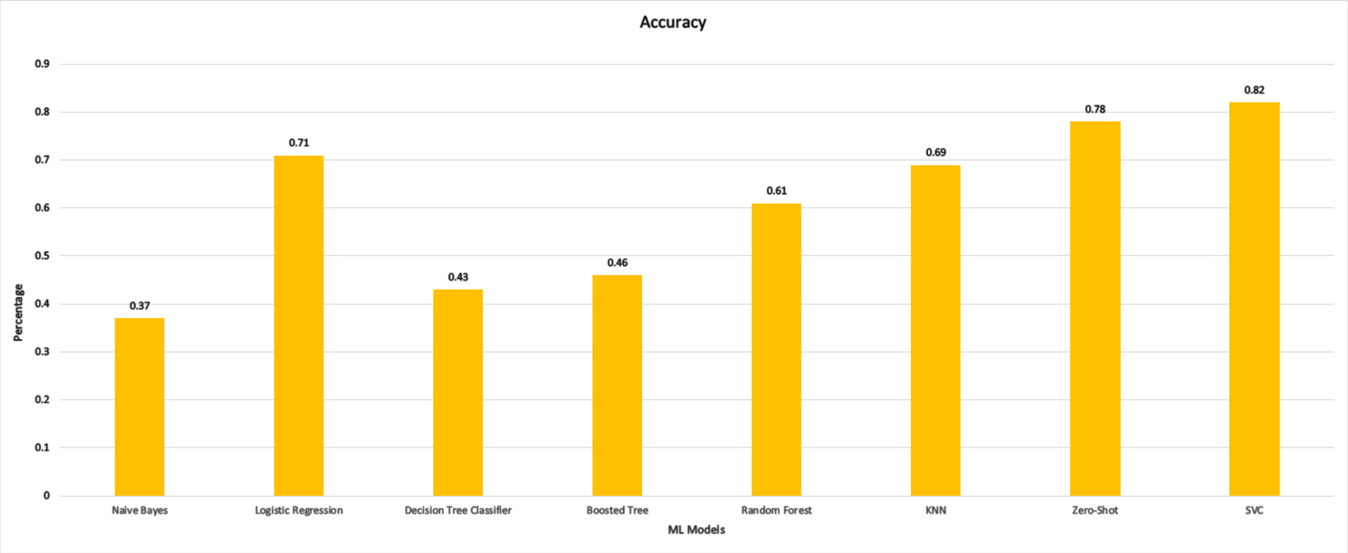


Figure 5 Accuracy comparison of Support Vector Classifier with other state-of-the-art algorithms

for dealing with high dimensional noisy data in text classification. An RF model comprises a set of decision trees, each trained using random feature subsets. Zero-Shot Classifier classifies the text documents without using labeled data or seeing any labeled text. Support Vector Classifier has comparatively high accuracy. It is well known for its kernel approach to handling nonlinear input spaces. The classifier divides data points into groups using the hyperplane with the largest margin, and SVC assists in categorization by determining the ideal hyperplane for new data points. We observed that the Support Vector Classifier outperformed all others regarding its parameters. Accuracy:82.42%, Weighted Precision:83.09%, Weighted Recall:82.42%, and Weighted F1 score :82.39%. One hundred thirty-six podcast genres were correctly predicted, including 33 in horror, 31 in motivational, 30 in criminal, and 42 in romantic. The result parameters like accuracy, weighted precision, weighted recall, and weighted f1 score of the used models are shown in Figures 5, 6, 7, and 8.

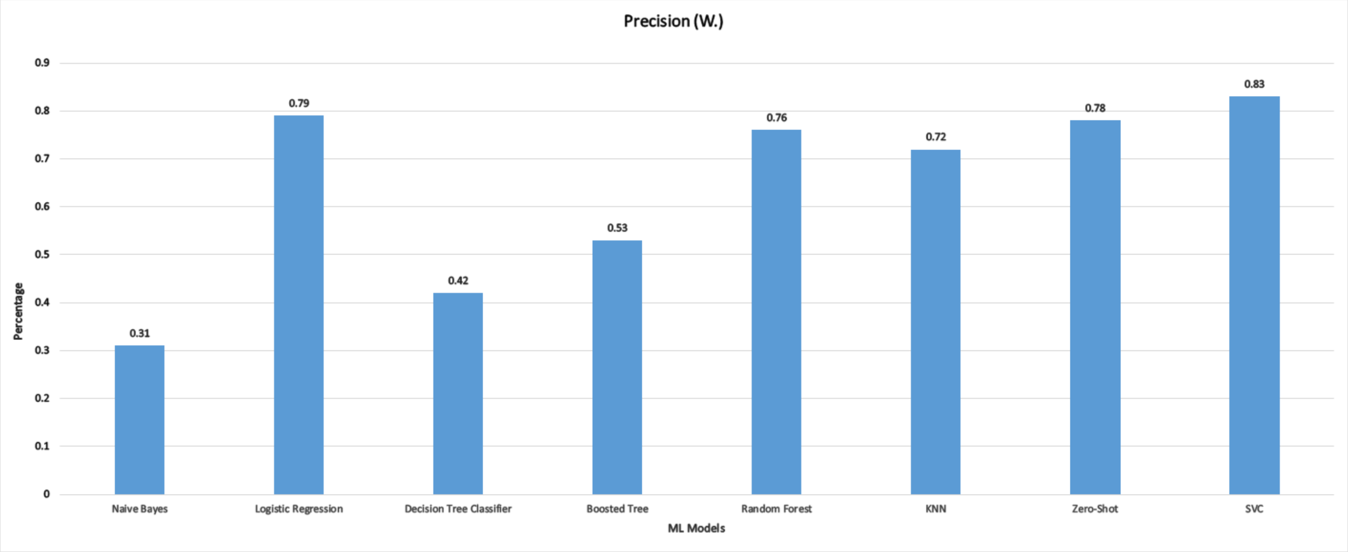


Figure 6 Weighted Precision Comparison of Support Vector Classifier with other state-of-the-art algorithms

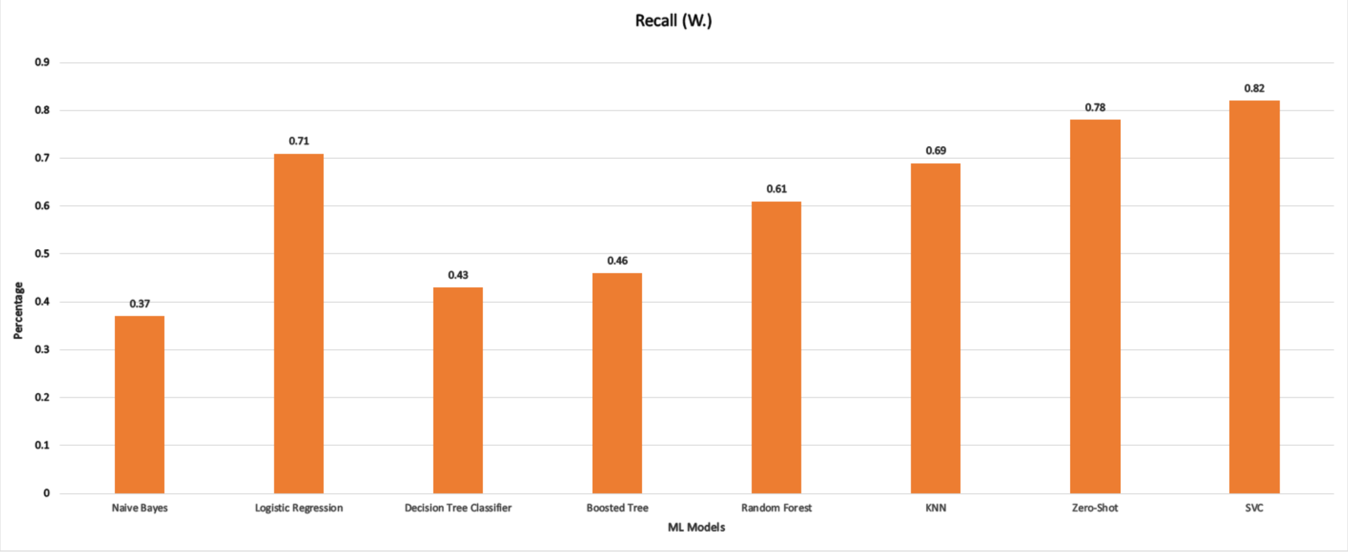


Figure 7 Weighted Recall Comparison of Support Vector Classifier with other state-of-the-art algorithms

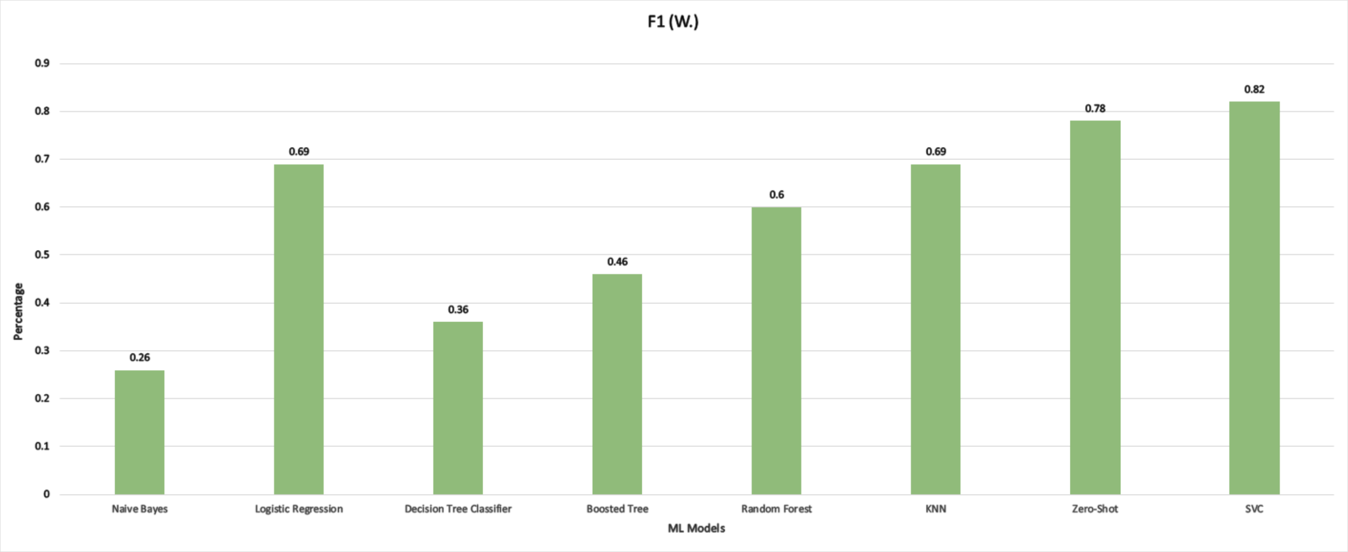


Figure 8 Weighted F1 score Comparison of Support Vector Classifier with other state-of-the-art algorithms

9 | CONCLUSION & FUTURE WORK

Using machine learning techniques to identify patterns and trends in a vast dataset has proven helpful. The ability to categorize a podcast's genre enables users of podcast streaming services to receive suggestions depending on the genre of the podcasts they prefer. It also allows listeners to create playlists of their favorite episodes. Education podcast genre prediction is required to efficiently classify and arrange educational content, making it more straightforward for listeners to access and absorb pertinent information. Listeners may find podcasts that suit their interests and requirements immediately by rapidly determining the genre of an education podcast, and educators can better focus their content on the right audience. Additionally, genre prediction can help producers of education-related podcasts better understand their listeners' format and content preferences and position their show more effectively in the market. Accurate genre prediction of educational genres can be helpful in this situation, like improving access to relevant content for remote learning for students and instructors by accurately categorizing educational podcasts and enhancing discovery as more and more educational podcasts are produced. Genre prediction can assist listeners in finding fresh, pertinent material while preventing information overload. Increased accessibility of users to locate content suited to their needs quickly and accurate genre prediction can help make educational content more accessible to a more extensive range of users, including those with unique learning needs or disabilities. This study applied multiple ML models to our PodGen dataset to predict the genres of Hindi podcasts. The outcomes showed that the Support Vector Classifier performed better than other state-of-the-art techniques for predicting genre using the same dataset, with a balanced accuracy of 82.42%.

The study presented here can be expanded in various ways, one is by extending the dataset and adding new genres to the model to improve prediction accuracy. Speech-based genre prediction is another scope for this study. As more and more educational podcasts are created, the genres in this particular field increase, helping to quickly classify and predict the podcasts by topic or level of expertise. This is advantageous for the podcasting industry as it allows people looking for instructional podcasts find precisely what they need with minimal effort.

PRODUCTION NOTES

Conflict of Interest Statement

All authors declare that they have no conflicts of interest.

Dataset Availability

Data will be made available on request.

Acknowledgement

The authors would like to acknowledge the technical support of Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Mexico, in the production of this work.

References

1. De Sarkar T. Introducing podcast in library service: an analytical study. *Vine* 2012.
2. Nwosu AC, Monnery D, Reid VL, Chapman L. Use of podcast technology to facilitate education, communication and dissemination in palliative care: the development of the AmiPal podcast. *BMJ supportive & palliative care* 2017; 7(2): 212–217.
3. Kang M, Gretzel U. Effects of podcast tours on tourist experiences in a national park. *Tourism Management* 2012; 33(2): 440–455.
4. Hancock D, McMurtry L. ‘I Know What a Podcast Is’: Post-Serial Fiction and Podcast Media Identity. In: Springer. 2018 (pp. 81–105).
5. Swiatek L. The podcast as an intimate bridging medium. In: Springer. 2018 (pp. 173–187).
6. Indahsari D. Using podcast for EFL students in language learning. *JEES (Journal of English Educators Society)* 2020; 5(2): 103–108.
7. Bafna PB, Saini JR. Hindi Verse Class Predictor using Concept Learning Algorithms. In: IEEE. ; 2020: 318–322.
8. Puri S, Singh SP. An efficient Hindi text classification model using SVM. In: Springer. 2019 (pp. 227–237).
9. Pal A, Barigidad A, Mustafi A. Identifying movie genre compositions using neural networks and introducing GenRec-a recommender system based on audience genre perception. In: IEEE. ; 2020: 1–7.
10. Roy A, Ludwig SA. Genre based hybrid filtering for movie recommendation engine. *Journal of Intelligent Information Systems* 2021; 56(3): 485–507.
11. Leartpantulak K, Kitjaidure Y. Music genre classification of audio signals using particle swarm optimization and stacking ensemble. In: IEEE. ; 2019: 1–4.
12. Bisharad D, Laskar RH. Music Genre Recognition Using Residual Neural Networks. In: IEEE. ; 2019: 2063–2068.
13. Wi JA, Jang S, Kim Y. Poster-based multiple movie genre classification using inter-channel features. *IEEE Access* 2020; 8: 66615–66624.
14. Kundalia K, Patel Y, Shah M. Multi-label movie genre detection from a movie poster using knowledge transfer learning. *Augmented Human Research* 2020; 5(1): 1–9.
15. Oramas S, Barbieri F, Nieto Caballero O, Serra X. Multimodal deep learning for music genre classification. *Transactions of the International Society for Music Information Retrieval. 2018; 1 (1): 4-21.* 2018.
16. Saari P, Fazekas G, Eerola T, Barthet M, Lartillot O, Sandler M. Genre-adaptive semantic computing and audio-based modelling for music mood annotation. *IEEE Transactions on Affective Computing* 2015; 7(2): 122–135.
17. Drew C. Educational podcasts: A genre analysis. *E-Learning and Digital Media* 2017; 14(4): 201–211.
18. Sindhu C, Adak S, Tigga SC. Opinionated text classification for hindi tweets using deep learning. In: IEEE. ; 2021: 1217–1222.
19. Bafna P, Saini JR. Hindi poetry classification using eager supervised machine learning algorithms. In: IEEE. ; 2020: 175–178.
20. Litake O, Sabane M, Patil P, Ranade A, Joshi R. Mono vs multilingual BERT: A case study in hindi and marathi named entity recognition. *arXiv preprint arXiv:2203.12907* 2022.
21. Chauhan S, Chauhan P. Music mood classification based on lyrical analysis of Hindi songs using Latent Dirichlet Allocation. In: IEEE. ; 2016: 72–76.
22. Ozakar R, Gedikli E. Music Genre Classificatio Using Novel Song Structure Derived Features. In: IEEE. ; 2020: 117–120.

23. Pothina H, Gopakumar G, others . Multimedia Genre Prediction by Analysing Neurophysiological Signals. In: IEEE. ; 2020: 364–368.

24. Tsagkias M, Larson M, De Rijke M. Predicting podcast preference: An analysis framework and its application. *Journal of the American Society for information Science and Technology* 2010; 61(2): 374–391.

25. Kelly JM, Perseghin A, Dow AW, Trivedi SP, Rodman A, Berk J. Learning through listening: A scoping review of podcast use in medical education. *Academic Medicine* 2022; 97(7): 1079–1085.

26. Radhika V, Swaraj K. Movie Genre Prediction and Recommendation Using Deep Visual Features from Movie Trailers. In: IEEE. ; 2020: 1–6.

27. Yadav A, Vishwakarma DK. A unified framework of deep networks for genre classification using movie trailer. *Applied Soft Computing* 2020; 96: 106624.

28. Yu Y, Lu Z, Li Y, Liu D. ASTS: attention based spatio-temporal sequential framework for movie trailer genre classification. *Multimedia Tools and Applications* 2021; 80(7): 9749–9764.

29. Bouyahi M, Ayed YB. Video scenes segmentation based on multimodal genre prediction. *Procedia Computer Science* 2020; 176: 10–21.

30. Bhingardive S, Redkar H, Sappadla P, Singh D, Bhattacharyya P. IndoWordNet:: Similarity-Computing Semantic Similarity and Relatedness using IndoWordNet. In: IIT, Bombay. ; 2016: 39–43.

☐