Genre Sentiment Classification Based on Movie Reviews

Abdallah Adel, Mohamed Kamal
Team 13

June 2, 2023

Abstract

Your abstract.

1 Introduction

A fascinating and difficult subject in the realm of natural language processing is the classification of sentiment into genres based on movie reviews. The work entails categorising the sentiment of a movie review text according to its genre. Sentiment analysis is a branch of natural language processing (NLP) that deals with finding and extracting subjective information from text data. It has a wide range of real-world uses, including marketing, studying consumer feedback, and examining movie reviews, among others.

Genre Sentiment Classification is a difficult assignment since it calls for the model to comprehend the context and subtleties of movie reviews, such as the language used, the tone, and the genre of the film being reviewed. Due to the subjectivity of sentiment analysis, the model must be able to reliably classify the review's sentiment.

Researchers have created a number of NLP techniques, including tokenization, word embedding, and neural network designs, to overcome these difficulties and achieve high accuracy in genre sentiment classification. Tokenization is the act of separating individual words from a sentence or paragraph so that they can be used as input for a model. Utilizing a method called word embedding, the relationship between words can be represented by the model by mapping words into a high-dimensional vector space. In numerous studies, neural network designs like Long Short-Term Memory (LSTM) have been utilized to accurately classify the sentiment in movie reviews.

Even with the advancements made in this area, more can be done. Dealing with sarcasm, irony, and other figurative language is one of the difficulties that still has to be overcome. Another is enhancing the model's performance on less well-known genres. Taking care of these issues will be essential for the future success of genre sentiment classification as the demand for automated sentiment analysis increases.

Genre sentiment classification using movie reviews is a difficult task. The main goal is to identify the tone of the review and categorize it into one of the five moods: hilarious, scary, horrible, good, or dramatic. Reviews can be lengthy, complex, and filled with a wide variety of language and sophisticated sentence structures, making this task difficult. Additionally, sentiment classification is a very subjective procedure because various people may have different interpretations of the same remark depending on their own backgrounds and viewpoints.

Another difficulty with this work is that it can be challenging to discern sarcasm or irony in movie reviews, which can occur frequently. It's possible for a reviewer to mention something that seems positive but is actually negative, or the other way around. This may result in incorrect sentiment categorization and deceive the reader.

The number of reviews that need to be analyzed can often be overwhelming. Each of the tens of thousands of films that are released each year may receive hundreds or even thousands of reviews. Each review would need to be read and categorized manually, which would take a lot of time and money. As a result, there is a critical demand for automated sentiment classification systems.

To address this difficulty, a variety of methods and resources are available. To preprocess the text input, the typical method uses Natural Language Processing (NLP) techniques such tokenization,

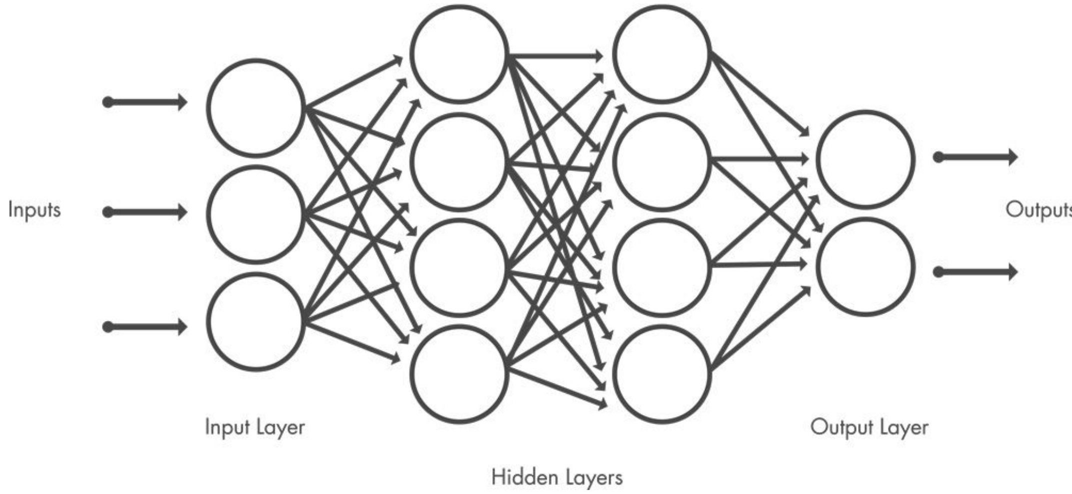

Figure 1: Deep neural network.

padding, and one-hot encoding. Once the emotion of the movie review has been classified, a deep learning model—specifically, an LSTM neural network—is used. However, there are a number of different machine learning methods and algorithms that can be applied to this problem, including Support Vector Machines, Random Forest, and Naive Bayes. To achieve accurate sentiment categorization, the right technique must be chosen because every algorithm has strengths and disadvantages.

In conclusion, classifying genre sentiment based on movie reviews is a difficult and complex undertaking. To accurately classify sentiment, large amounts of data must be processed, cutting-edge NLP methods must be used, and the right machine learning algorithm must be used. However, it is expected that sentiment classification systems will grow more accurate as deep learning and NLP techniques advance, making them invaluable tools for moviemakers, critics, and viewers alike.

2 Literature Review

2.1 Deep Neural Networks

There are many research trends on the topic of NLP using deep learning and machine learning techniques. We will first discuss more about Deep Neural Networks.

A DNN has appeared to cope with the high scale of data and powerful computation that outperforms other simple neural networks. It consists of three main layers: input layer, hidden layers, and output layer. Each layer receives the preceding layer's neuron activation as input and performs a basic computation. The network's neurons work together to create a complex nonlinear mapping from input to output. The weights of each neuron are adjusted using a technique called error back propagation to learn this mapping from the data. This architecture is mainly used in deep learning as DNN could digest a large amount of data.

2.1.1 Activation

The weighted inputs are summed and forwarded via an activation function, commonly termed a transfer function. An activation function is a straightforward mapping of summed weighted input to the output of the neuron. It is termed an activation function because it regulates the threshold at which the neuron is triggered and the intensity of the output signal. Historically simple step activation functions were utilized where if the cumulative input was over a threshold, for example 0.5, then the neuron would output a value of 1.0, else it would produce a 0.0. Traditionally, nonlinear activation functions are utilized. This permits the network to mix the inputs in more sophisticated ways and in turn gives a greater capacity in the functions they can model. Nonlinear functions like the logistic function also called the sigmoid function were utilized that output a value between 0 and 1 with an s-shaped distribution, and the hyperbolic tangent function also called Tanh that outputs the same distribution throughout the range -1 to +1. More recently, the rectifier activation function has been found to produce superior outcomes.

2.2 Natural Language Processing

A branch of computer science and artificial intelligence called "natural language processing" (NLP) aims to give computers the ability to comprehend, analyze, and produce human language. Using methods from linguistics, computer science, mathematics, and statistics, NLP is an interdisciplinary field. NLP aims to develop computer systems that can communicate with people using natural language.

Numerous industries, including healthcare, banking, e-commerce, and customer service, use NLP in diverse ways. With NLP, meaningful information can be gleaned from medical records, clinical notes, and other medical papers in the field of healthcare. NLP can be used in finance to analyze news and financial reports to forecast stock values. To enhance product suggestions and the consumer experience in e-commerce, NLP can be used to comprehend customer evaluations and feedback. NLP can be used in customer service to create chatbots and virtual assistants that can comprehend consumer requests and offer useful responses.

Tokenization, part-of-speech tagging, named entity identification, parsing, and sentiment analysis are only a few of the essential techniques used in the creation of NLP systems. Tokenization is the process of dissecting a text into its component words and phrases. Identifying the grammatical category of each word in a sentence is known as part-of-speech tagging. Identifying and classifying entities in a text, such as persons, businesses, and locations, is known as named entity recognition. Analyzing a sentence's grammatical structure is part of the parsing process. Sentiment analysis entails determining the text's emotional tone.

The number, quality, and correctness of the training data, together with the available computational resources, all affect how well NLP systems perform. The performance of NLP systems has considerably increased as a result of recent developments in deep learning, particularly the creation of neural network topologies like recurrent neural networks and transformers.

Overall, NLP is a rapidly developing field with a wide range of real-world uses. The necessity for improved NLP approaches will only grow as the volume of digital data continues to rise.

2.3 Dataset

In the discipline of natural language processing (NLP), the IMDB dataset, made available by Keras, is a frequently used dataset. There are 50,000 reviews of films from the Internet Movie Database (IMDB) in it, all of which are classified as either good or negative. A training set of 25,000 reviews and a test set of 25,000 reviews were used to partition the dataset into two sets.

Each review in the dataset is represented as a series of integers, each of which stands for a different word. Each word in the dataset has a distinct integer index due to preprocessing and indexing. The get_word_index() method offered by Keras can be used to discover the index for each word in the dataset.

Each review in the dataset is preprocessed to be a fixed-length sequence of integers, with reviews with fewer words being padded with zeros to make them the same length as reviews with more words. The longest review in the dataset, at its maximum length, determines the length of the sequence. For the purpose of training neural networks, which demand inputs of a specified length, a preprocessing step is necessary.

The task of categorizing the sentiment of a piece of text as positive or negative uses the IMDB dataset, which is frequently used in sentiment analysis. The IMDB dataset is frequently used by researchers to develop and test sentiment analysis NLP algorithms, such as neural networks. Accuracy, precision, recall, and F1 score are common measures used to assess these models' performance.

2.3.1 Analysis

In this dataset, after applying our manual genre classification based on specified keywords, we found out that the classification of IMDb reviews is as in 2. We noticed that funny is dominant while sci-fi and crime were recessive, so we removed both in order to balance the classification. The distribution after removing both can be seen in 3.

Another distribution that is seen in 4 is used for one of the LSTM models discussed below.

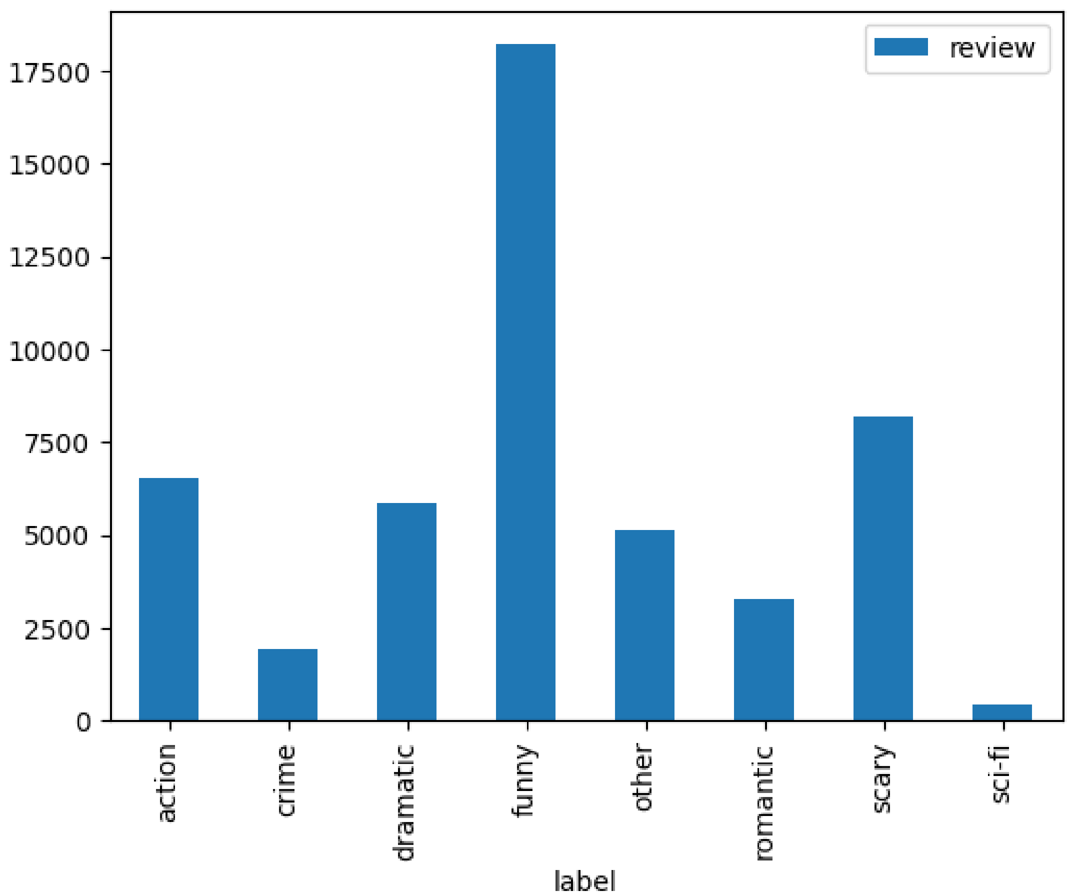

Figure 2: Distribution of classified genre before removal of sci-fi and crime

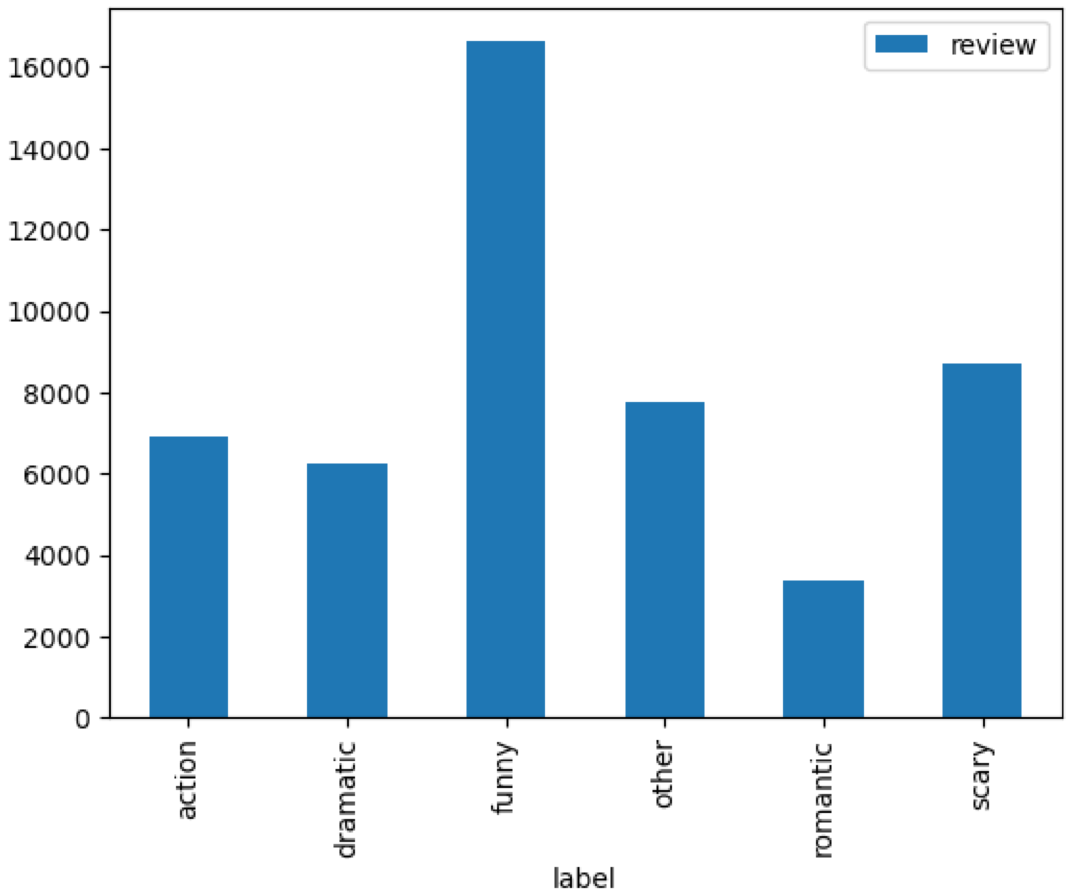

Figure 3: Distribution of classified genre after removal of sci-fi and crime

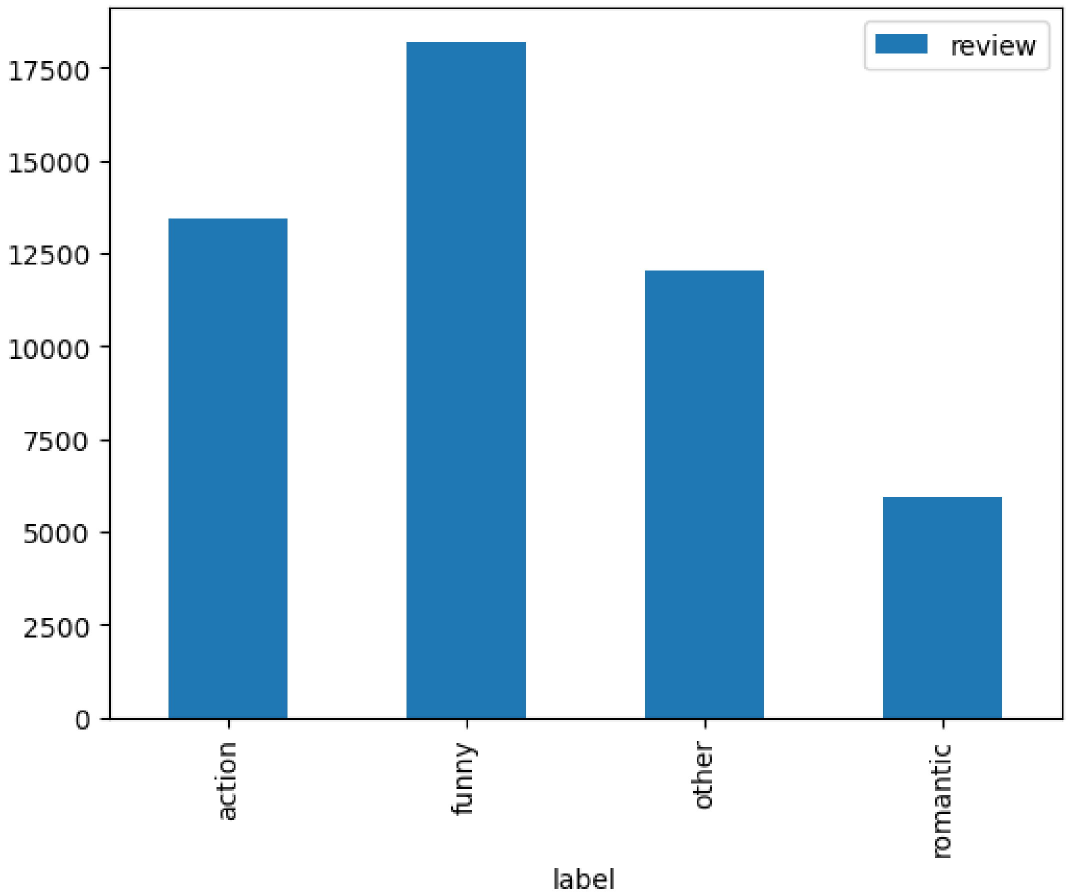

Figure 4: Distribution of classified genre after removal of dramatic, scary, sci-fi, and crime.

3 Methodology

In this section, we will discuss the methodology behind our approaches for genre classification based on movie reviews.

3.1 Intuition and Approaches

Classification of movie genre based on a review is an entirely different task (with different challenges) from sentiment classification, yet they may have some resemblance between each other. Sentiment classification usually requires acquiring a dataset (preferably a labeled one with positive or negative), preprocessing the review text, then acquiring or training a model for sentiment classification on the data gathered. This is a fairly transparent workflow that is not very prone to errors and problems. However, as genre classification is pretty similar to sentiment classification in terms of acquiring the data, and preprocessing the review text, one main difference lies in labeling the data. Reviews for any movie, or even any product, infer either a positive or a negative meaning behind them, this makes the problem of handling an anonymous review easier since it doesn't really rely on the movie itself rather than the actual meaning. However, in the case of genre classification, the problem here is to get the correct genre of a movie based on the review, and this is usually extremely difficult because most of the reviews have a straight forward meaning: positive or negative, without giving much info about the genre of the movie itself, i.e: you critique the movie as a whole not its ability to represent its proclaimed genre. So for this project we had several approaches tested. One was a KNN approach, another was a K-Means clustering approach, and lastly, manually labeling the data and training a regular LSTM model on it.

3.1.1 Manually labeling the data

In both the KNN model and LSTM models, we applied a manual genre labeling method to the reviews based on some keywords so that we can train our models based on these values. We also used these manually labeled reviews to use them as test samples for the K-means model. More info on this can be found in the Analysis section 2.3.1 under Dataset 2.3.

3.1.2 K-Means

We tried a K-means approach with k = 6 for the 6 genres, we believed such approach would show promising results and we be extremely meaningful as we thought that guessing the genre based on the review text would make a good use case for unsupervised clustering, and that reviews tend to be similar for the same genre.

3.1.3 KNN

We tried a KNN approach with k = 5 for the same reason as K-Means approach but this time with labeled data at train time, hence a supervised approach instead of an unsupervised one.

3.1.4 LSTM

Lastly, we tried the typical approach of an LSTM model, we trained two models. One using 6 genres: Action, Dramatic, Funny, Romantic, Scary, and Other. The other model was trained on same genres but we discarded Scary and Dramatic by random elimination.

4 Results

In this section, we will discuss the results obtained from each of our approaches.

4.1 K-means

For the k-means model we got a silhouette score of -0.01, which means that there is an overlapping in the clusters. This can be seen in figure 5. The silhouette score for each sample can be seen in figure 8.

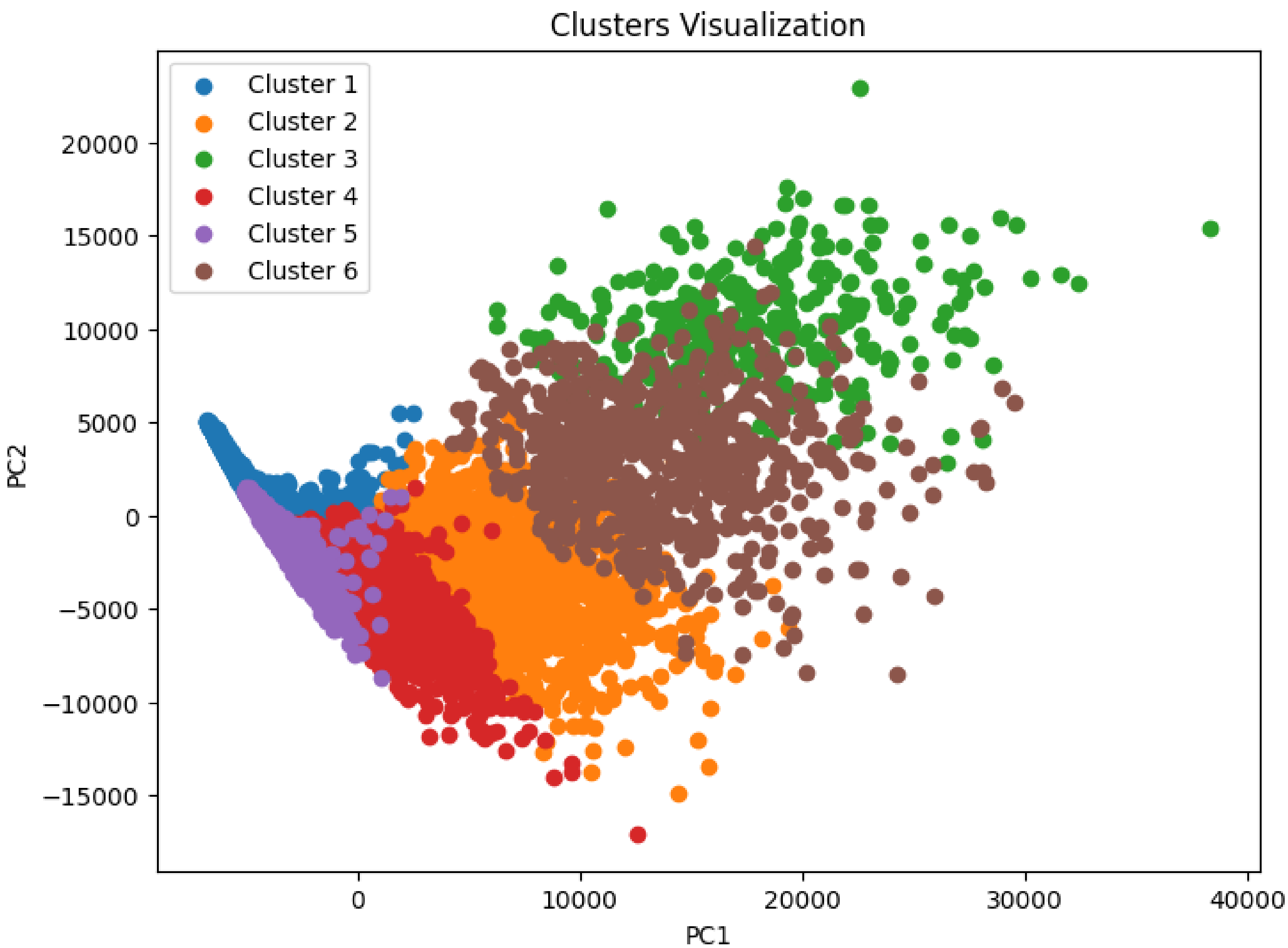

Figure 5: Unlabeled clusters of the trained K-Means model.

4.2 KNN

We received an initial accuracy score of 10% for the KNN model so we quickly discarded it and moved on to the LSTM models.

4.3 LSTM

For the LSTM approach, we tried two models. One with 6 genres: Action, Dramatic, Funny, Other, Romantic, Scary, we call it model_1. Another with 4 genres: Action, Funny, Other, Romantic, we call it model_2.

For model 1 we got an overall training accuracy of 69.47%, and test accuracy of 66.57%. We also had a weighted average of 63% for precision, 67% for recall, and 51% for recall.

For model 2 we got an overall training accuracy of 72.98%, and test accuracy of 71.89%. We also had a weighted average of 71% for precision, 72% for recall, and 69% for recall.

You can see the ROC curves for both model 1 and model 2 on

5 Future Work, Limitations, and Conclusion

For future work we can find a dataset that already features genre labels instead of manually assigning them, we also still need to find a way label the clusters generated from the K-means model in order

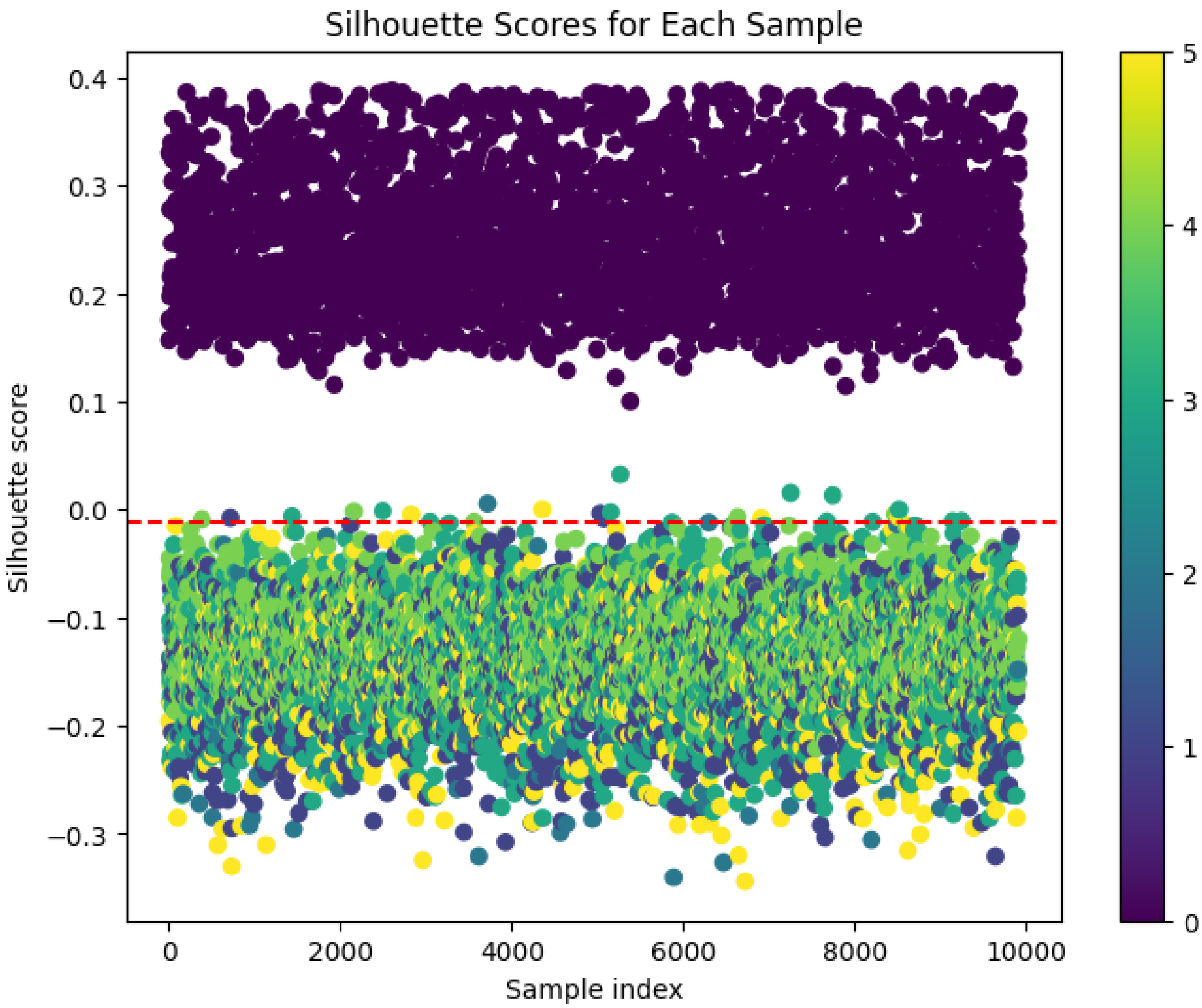

Figure 6: Silhouette score for each sample.

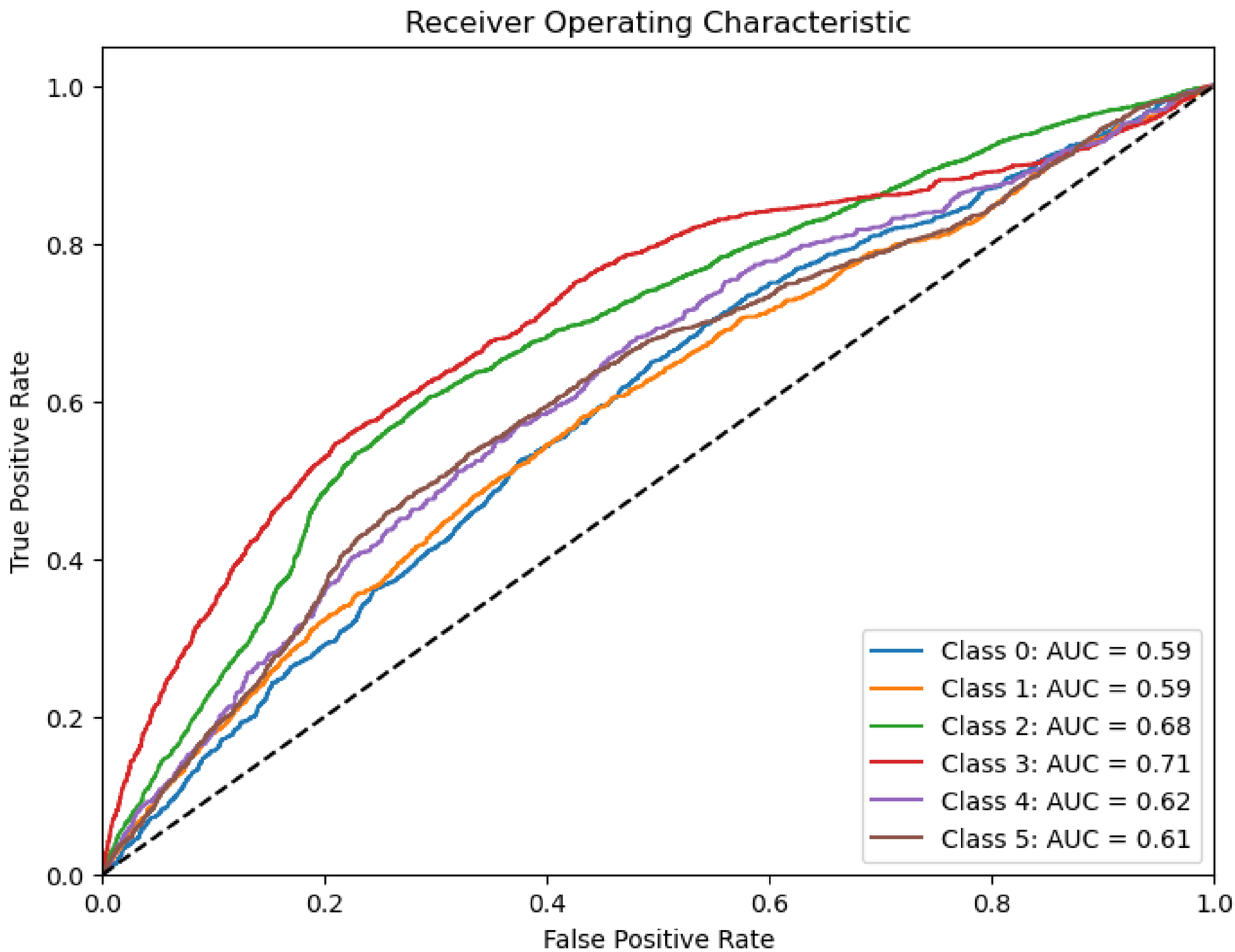

Figure 7: ROC curve for LSTM model 1.

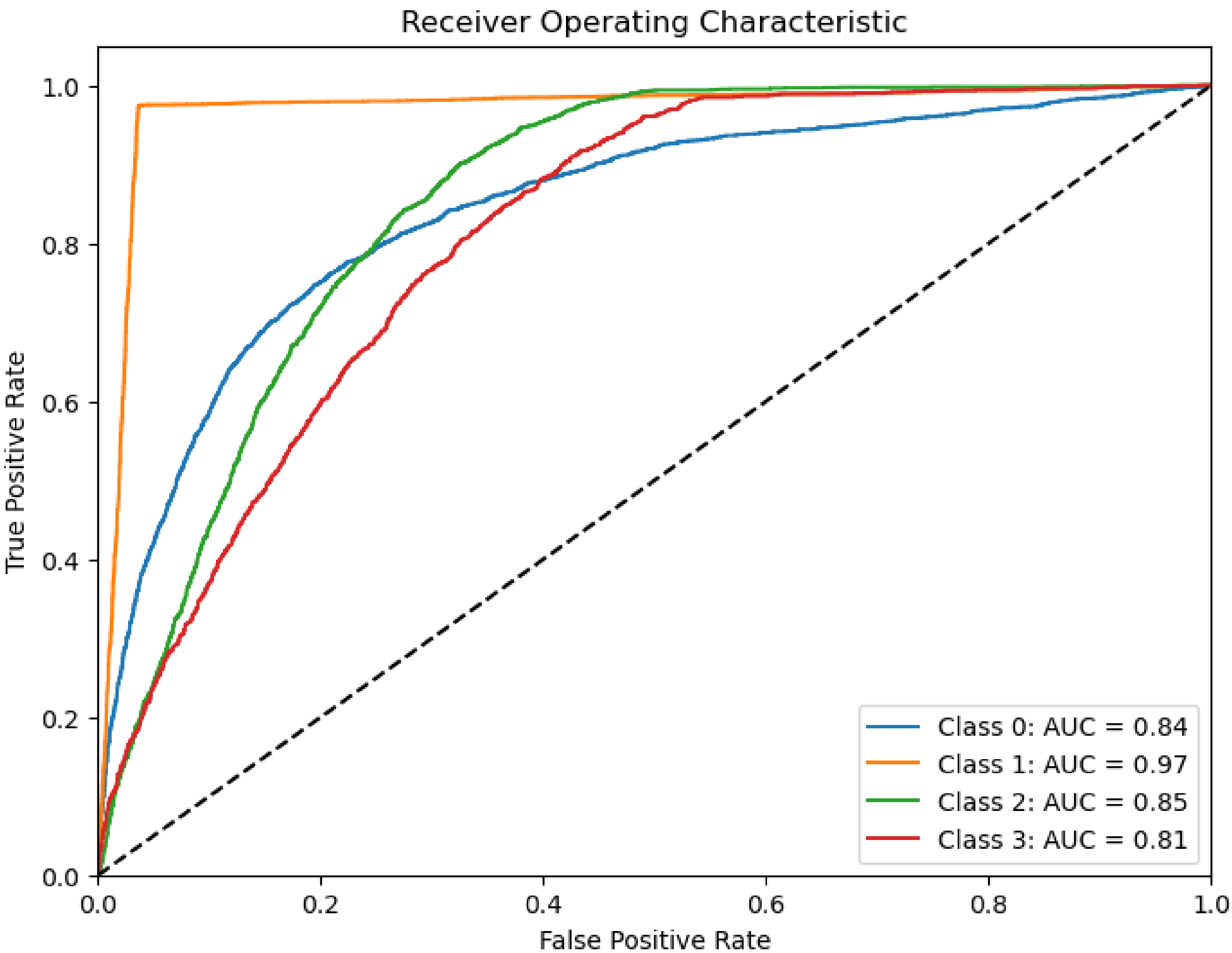

Figure 8: ROC curve for LSTM model 2 (reduced genres).

to see how good the model is.

Limitations include, but are not limited to: Limitations on training hardware, and availability of datasets.

In conclusion genre classification based on review can surely show whether the movie remains true to its proclaimed genres or not and can help viewers decide whether they want to watch the movie or not by adding more value to the sentiment classification. It is definitely not an easy task but it does add more value to review classification models, and tools.