A comparative study of LSTM, RNN and CNN at predicting sale figures of sings based on text abstract and musical features

Stephen Loftus

M.Sc. Data Analytics

College of Engineering & Informatics

National University of Ireland, Galway

Supervised by:

Dr. Mathieu D'Aquin

September 2021

Declaration

I, Stephen Loftus do hereby declare that this thesis entitled A comparative study of LSTM, RNN and CNN at predicting sale figures of sings based on text abstract and musical features is a bonafide record of research work done by me for the award of MSc in Computer Science Data Analytics from National University of Ireland, Galway. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

_____

Stephen Loftus

September 2021

Acknowledgements

I'd like to thank my supervisor Dr. Mathieu D'Aquin for his guidance and help during this project. His attention to detail and expertise greatly benefitted me in the completion of this project.

I'd also like to thank my friends and family who helped supported me throughout completion of this project and my Masters.

Abstract

This project is a comparative study of three different deep learning models Long Short-Term Memory(LSTM) , Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) and the success each model has in predicting sales figures of songs based on various types of data. The data is a combination of cultural data comprised of data such as the artist and release date which is collected from Wikipedia, and musical data which is related to the musical components of the song such as its tempo or key which is gathered from Spotify. Overall, nine models are developed , three models for each of the different types of deep learning model LSTM RNN and CNN, for each type of deep learning model one model is trained on the cultural data, one is trained on the musical data and finally one is trained on the combination of both musical and cultural data. For each of the nine models the percentage of predictions within a ±5% range of the actual sales value is calculated along with Mean Absolute Error(MAE) and Root Mean Square Error(RMSE) values. The models are then compared using these results to identify which of the three model types preform the best and also to identify which type of data gives the best results. T-tests are used to check if the performance difference between models is significant. Overall LSTM models were found to have the best performance with the greatest number of predictions within a 5% range of the actual and the lowest MAE and RMSE values across all of the 3 different types of training data. It was also found that the training the models using the combination of cultural and musical data led to a statistically significant improvement in performance as opposed to when models were trained using only one type of data.

Keywords :

Long Short-Term Memory , Recurrent Neural Network , Convolutional Neural Network , Spotify , Music Sales , Prediction , Cultural Data , Musical Data, Sales Prediction , Music

List of Figures

List of Tables

List of Abbreviations

IFPI	International Federation of the Phonographic Industry
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
MAE	Mean Absolute Error
RMSE	Root Mean Square Error

Contents

Chapter 1

Introduction

1.1 Outline

The goal of this project proposal is to summarise other work which will be completed with regards to the prediction of music performance measured using sales figures and the proposed process for tackling the problem of predicting music performance using a variety of Deep Learning methods trained on a combination of data relating to different aspects of a song within the music industry. The performance of the different models will be evaluated and compared based on their accuracy and performance at predicting sales figures.

1.2 Motivation

The Music Industry is a billion-dollar industry with a revenue of $21.6 Billion as reported by the International Federation of the Phonographic Industry (IFPI) in their global music report [1] , music sales make up $5.4 Billion or around 25% of total revenue, while revenue from streaming services equated to $13.4 Billion or just over 60% [1] , together revenue from sales and streams make up over 85% of the total revenue within the music industry. While this figure is for the music industry as a whole it gives a good representation of the revenue of individual artists. The Billboard Top 100 charts rank the top 100 songs within different criteria such as genre based off sales , streams, and radio airplay. Hence, one can conclude that being accurately able to predict sales figures for individual songs would be highly beneficial to artists and their labels as combined they can give us insight into both the potential revenue to be generated from a song and the success it will have within the charts.

It has been shown by that stream and sale figures are highly corelated therefore for the purpose of this project the goal is to focus on predicting sales figures, sales figures were chosen over streams as despite streams making up a higher percentage of revenue as more accurate and reliable sales figures are available then streaming figures.

Music is a multi-faceted industry with a large variety of factors that can impact on the overall sales of a song such as radio plays, social media following of the artist and musical features such as tempo to name just a few. The data related to these factors can be comprised of an assortment of data types such as text data , numeric data, or audio data. In order to make accurate predictions all of the varying factors should be considered, therefore models that can handle the varying data types should be implemented. The majority of work done utilise regular machine learning

algorithms such as Support Vector Machines when making predictions of song performance and have focused on only one of the factors that impact performance. Newer more advanced Deep Learning Neural Network models such as LSTMs (Long Short-Term Memory) which are better suited for use with textual data and can provide better predictions with regards to song specifics such as different audio features being more impact for different genres or artists due to their higher learning capabilities would provide more accurate predictions. By combining multiple factors and implementing these newer deep learning models' better accuracy and performance for predicting sales figures is expected.

1.3 Data Used

Spotify provides a developer platform which provides access to pre extracted audio features. A table of the extracted Musical Features and their description can be found on the last page of this section.

DBpedia is an online database of information created from Wikipedia a free online encyclopaedia which can be easily queried by users. Wikipedia contains a wide range of cultural data related to a song which can be easily obtained via Dbpedia, it also has an abstract for each song , since Wikipedia entries can be altered by anyone these provide an all-round representation of song as information about performance, awards a song received , the artist as well as their background information can all be contained within the abstract.

If streaming figures were to be predicted they would have been taken from Spotify. Spotify is the largest music streaming service second only to YouTube according to a report carried out on behalf of the IFPI in 2016 with 40 million paid users [3] and as of the first quarter of 2021 this figure has risen to 158 million paid users [4]. Thus, predicting Spotify streaming figures for a particular song would provide an accurate representation of streaming figures as a whole for a song , however it was decided to focus on sales figures during this project.

1.4 Research Question

The goal of this project will be to answer the following questions:

1. Can deep learning models namely Convolutional Neural Networks, Recurrent Neural Networks and Long Short-Term Memory Networks be used to accurately predict sales and streaming figures for Songs based on Cultural Data and Musical features ?

2. Which of these approaches has achieves the best overall performance ?

3. Which of the data types are most important to accurately predict sales and streaming figures?

The following table provides a description of the Musical Features extracted from Spotify; all descriptions are taken directly from the Spotify Developer Documentation [5]:

Musical Feature	Description
Danceability	**Danceability** is calculated by Spotify and describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.
Energy	**Energy** represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy
Key	The **Key** the track is in. Integers map to pitches using standard Pitch Class notation
Tempo	The overall estimated **tempo** of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
Loudness	The overall **loudness** of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks
Acousticness	**Acousticness** is a confidence measure of whether the track is acoustic.
Valence	**Valence** is a measure describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive, while tracks with low valence sound more negative

Table 1.3.1 Table of Spotify variables

Chapter 2

Background Research and Literature Review

2.1 Musical Performance

Musical data is a combination of different data sources, firstly Musical Feature Data which is made on from data directly related to the piece of music such as the Energy , Key, or Tempo, and secondly Cultural Data that is not a direct part of the piece of music such as the artist, genre, or online reviews.

2.1.1 - Music Performance from Cultural Data

A larger amount of research has gone into predicting the performance of music based off its Cultural Data related to that piece of music. For example, Abel, Diaz-Aviles, Henze, Krause and Siehndel [6] analysed how blogs post related to the success of a music album. Using machine learning algorithms with features they extracted from blogposts , they identified that the change in the number of blog posts about an album could be successfully predicted with an accuracy of around 60% , They found the number of blogs that mentioned the album title and artist in the blog to be most impactful variables when making predictions. Furthermore, they used the same algorithm to predict the sales position of an album within the Amazon Sales Rank with a precision of ~50%. In a similar study Dhar and Chang [7] also used blog data when predicting performance of albums, but they predicted actual sales figures of the albums. They also found that album performance is highly corelated to the number of blog posts, with release label and reviews in mainstream music sources also having an impact on sales figures.

Other work has been completed on trying to directly predict album sales, for example Lee, Boatwright, and Kamakura [8] implemented a Bayesian model with sales data from previously released albums to predict sales prior to launch for new albums. The models developed are able to estimate generalised sales volume for new album prior to release and could further improve upon forecasted sales once the album had been released and more data was available. Social media is a huge aspect of musical cultural data as it allows artists to communicate with and grow their fan base as well as being a primary source of information gathering for a lot of users. Knowing this Kim, sun, and Lee [2] investigated how Twitter users impacted on the position and duration of a song's placement within the Billboard Top 100 charts. The information extracted from tweets

that contained music related hashtags were for to be highly corelated with Billboard Top 100 rank, regression models built using the extracted data were also successfully able to classify a sing as being a hit (placing in the range 1 – 50 of the charts) or not a hit (placing in the range 51 -100) with an accuracy, precision and recall all above 80%.

How fans consume music has greatly changed with the technological advancements of the 21st century, as streaming services such as Spotify have become the primary way people consume music. As of 2020 IFPI report streaming accounts for 62% of the global music revenue [1]. M.Leea, Choia, Choa, and H.Leea [9] did research to see the impact the increase in online streams had on song sales, they found that music streams positively impact on song sales. Despite streaming being the primary source of consumption songs with a higher number of streams also have higher sales numbers.

Overall, from these works one can see that the Cultural Data related to an album or song has a great impact of its performance. Using features extracted from cultural data one can build models that accurately predict performance . Cultural data can be comprised of a wide range of different data types, such as text , numeric or date, and can come for a range of place such as traditional media or social media.

2.1.2 - Music Performance from Musical Features

Just like with Cultural Data a variety of work has been completed with regards to predicting the performance of a piece of music based on its Musical Features. One such example is the research carried out by Junghyuk Lee and Jong-Seok Lee [10] , who utilised Support Vector Machines with three main musical components, Chroma , Rhythm and Timbre. They aimed to predict performance based on positioning within the Billboard Hot 100 chart. It was found that the SVMs built using the extracted musical features performed significantly better than base line measures.

More recent work carried out by Rutger Nijkamp [11], focused on predicting performance measured by Spotify Streams using Musical Features extracted from the Spotify Database. A Linear Regression model was built using the extracted data and each of the features, Correlation Analysis was run on the model to identify the relationship between each of the Musical Features and their impact on Streaming Counts. The model was also tested to find the impact of combining the musical features on streaming count, it was found that around 20% of the variation in streaming count could be explained by the Musical Features. While this seems like a low percentage the approach taken doesn't account for discrepancies within genres, as mentioned by Nijkamp [14] different features have greater impact in different genres and the model used could not account for this.

A good example of this is the work carried out by Herremans, Martens and Sörensen [12] who looked into predicting if a dance song will be a hit. Five machine learning models ranging from Decision trees to Naïve Bayes to Logistic Regression were implemented and compared , all

models achieved an accuracy above 60% with some models achieving an accuracy of 85%. These outcomes are much greater than the results achieved by Nijkamp [11] and show that different musical features have greater importance for different genres and thus implementing a model that can learn these differences will yield greater results.

These papers are a curated sample of all work carried out in the domain of music performance based on musical features, they show not only that it is possible to predict performance from musical features but also that the impact of a particular musical feature on performances changes depending on the specific genre, while no work has been carried out on this question as if yet it could also be hypothesised that different musical features might have greater impact depending on artist also. Therefore, selecting models that have the capabilities to learn such underlying features such as deep learning models will outperform those that cannot.

2.1.3 - Music Performance from Combined Musical Data

So far, I have highlighted a variety of research based on predicting performance of music using only either Cultural Data or Musical Features, however some research has been done to investigate how the two types of data available can be combined.

Early research into the benefits of combining Cultural Data and Musical Features was carried out by Whitman and Smaragdis [13] who were trying to improve automatic genre recognition by combining the 2 types of data. The model developed from the combined data achieved significant performance improvements correctly classifying all test samples and overcame many of the problems that occurred when using only one of Cultural Data or Musical Features. This research can be backed up by McKay and Fujinaga [14] who also works on automatic genre classification for music by combining the available data types. Just like with Whitman and Smaragdis [13] , McKay and Fujinaga found that classification results significantly improved when features from the varying data sources were combined than when features from a single source was used.

Other research has been carried out by Dhanaraj and Logan [15] to predict performance of songs by combining Cultural Data and Audio Features. Similar to Junghyuk Lee and Jong-Seok Lee [10] performance was measured by using Support vector machines to predict if a song was a hit or not. They found that while Cultural Data had a greater impact on performance than Musical Features , by combining the two data types, models achieved better accuracy and performance than when only a single data type was utilised.

While it has been discussed how both Cultural Data and Musical features can be effectively used when predicting the performance of music, combining both aspects gives a fuller scope on all aspects that effect a song or albums performance. Overall, this leads to models that outperform models that are trained on only one of Cultural Data or Musical features.

2.2 Deep Learning

Deep Learning is a subset of machine learning, that utilises neural networks with at least 3 hidden layers . These Neural Networks allow Deep Learning algorithms to automatically extracted important features from data meaning they can be used with unstructured data such as text data which cannot be done with regular machine learning algorithms like Linear Regression.

As can be seen in figure 1 [14] on the next page deep neural networks are comprised of number of connected node layers , an input layer which takes in the data , a number of hidden layers which learn the data features and an output layer which is gives the output for example a prediction.

Each node takes in an input from the previous later which is multiplied by associated weight which determines the importance of each input. All inputs are summed, and a bias added to create an initial output, which is then passed to an activation function to find the final output . If the final output is greater than predefined threshold the output is passed to the next layer. This process allows the Network to learn the features of the data.

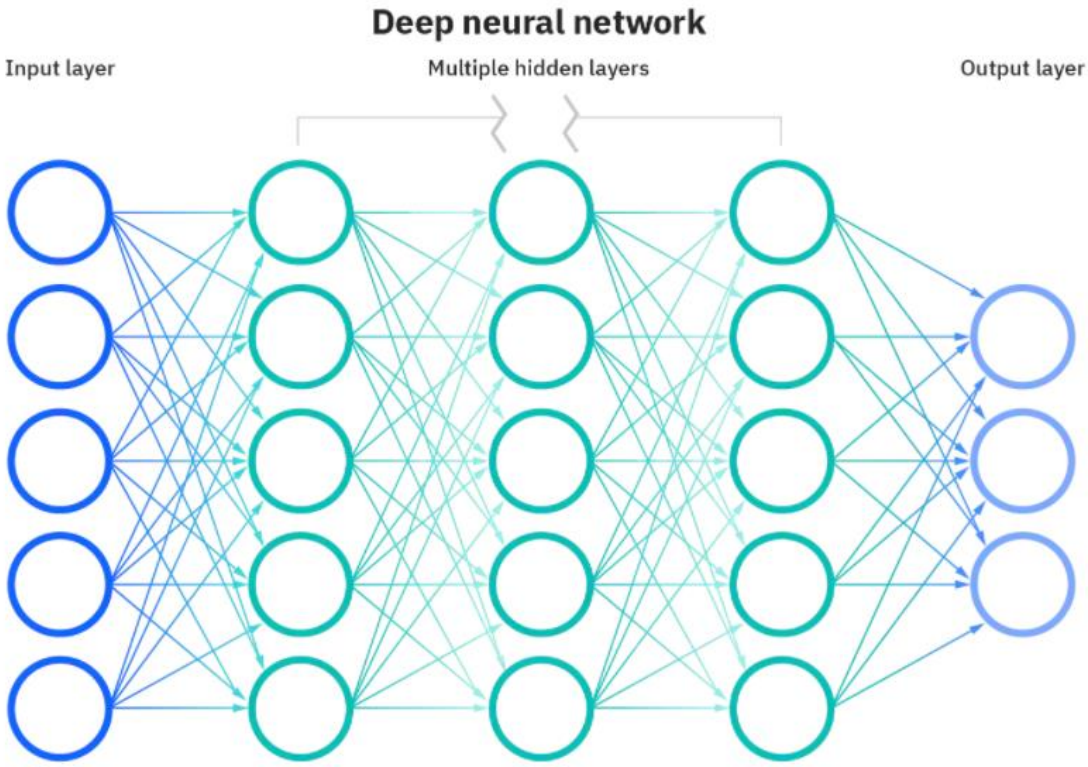

Figure 2.2.1 Deep Learning Neural Network [16]

2.2.1 - Convolutional Neural Networks(CNN)

Convolutional Neural Networks(CNNs) are feed forward neural networks and are made up of three main layers , a Convolutional layer , a Pooling layer, and a Fully connected layer. The convolution layer applies filtering to the input data , the pooling layer reduces dimensionality by decreasing the number of parameters in the input and finally the fully connected layer take the output from the series of convolutional and pooling layers and makes predictions which are then sent to the output layer.

CNNs are typically use for classification or computer vision task however they have been shown to outperform RNNs and LSTMs on stock price prediction by Sreelekshmy , Vinayakumar , Gopalakrishnan , Vijay, Soman [18]. While the task being carried out in this research is not the same there is sufficient reason to compare CNNs to RNNs and LSTMs for the task

2.2.2 - Recurrent Neural Networks(RNN)

Recurrent Neural Networks(RNNs) are slightly different from other Neural Network models due to their “memory” , RNNs consider the entire sequence of inputs to determine the output not just the input from the previous stage figure 2 [19]. This allows RNNs to learn information and context from previous stages of the learning process which can lead to better performance and makes them particularly useful for language and text processing tasks as often knowing the prior words in a sentence can give context to the current word.

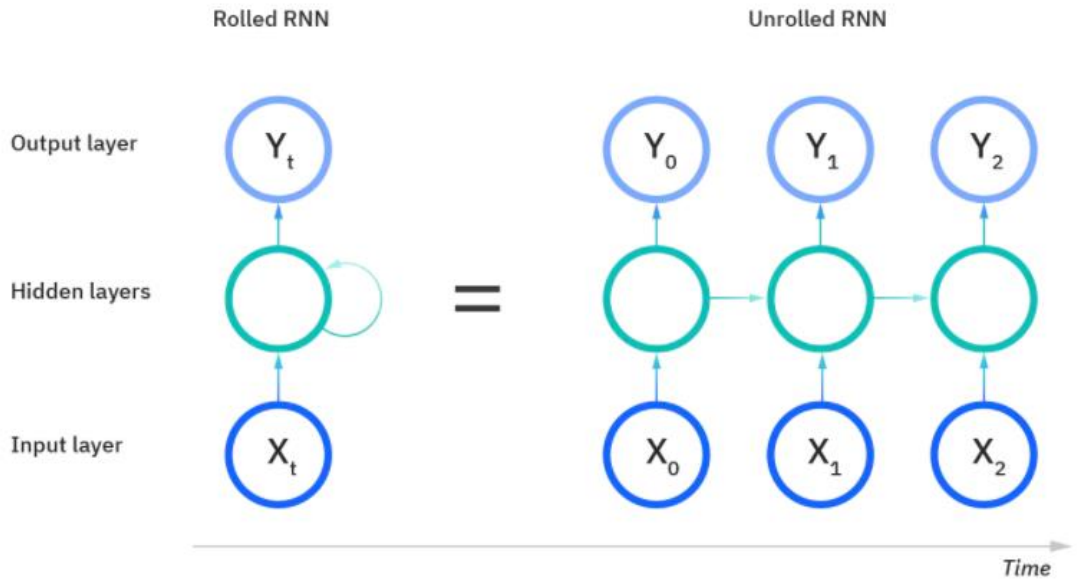

Figure 2.2.2 Recurrent Neural Network [19]

2.2.3 - Long Short-Term Memory (LSTM)

Long short-term memory(LSTM) is a type of RNN which was developed by Schmidhuber and Hochreiter to overcome the vanishing gradient problem that can arise with regular RNNs, this is where is where very small gradients are repeatedly multiplied causing them to go to 0 potentially losing important information. Regular RNNs have trouble learning long term dependencies where information they require is too far back in the sequence due to vanishing gradients. To overcome this LSTMs, use "cells" which store the required information in the hidden layers of the network. Gates which are comprised of a sigmoid layer and a multiplication pointwise operator and are used to regulate the information stored within the cells[20].

2.3 Short comings of research

From the literature review, it can be inferred that a variety of methods and approaches have been taken for forecasting and predicting music performance using both Cultural Data and Musical Features. Despite the huge advancements in deep learning techniques in recent years, no work seems to have been carried out to use these approaches within the music field for tasks such as genre classification or performance prediction instead many of the papers have opted to utilise machine learning methods. Herremans (2014) [12] successfully predicted if a dance song will be a hit showing the importance of genre specification in the research. Machine Learning techniques don't have the capabilities to learn and extract features from text data , this means that any research so far carried out on cultural data has been unable to look at the actual words and text within a review or blog for example .

I propose to implement and compare a variety of Deep Learning approaches using a combination of both cultural data and musical features. Deep Learning approaches have the ability to learn discrepancies between genres or artist which should lead to improved performance as suggested by Herremans[12]. It will also allow for the actual text data to be analysed to investigate the correlation between it and sales.

Chapter 3

Methodology:

3.1 Data Collection and Preparation

The data to be used in this project will be manual gathered and curated. The Song Name, Abstract , Sales Figure, Release Date , Duration ,Artist and Genre will be taken from Dbpedia and combined to make up the Cultural Data. Dbpedia has an inbuilt SPARQL querying system that allows users to run queries on Wikipedia data. The extracted data is then stored in a variety of formats, for this project the data is stored in a csv file.

Then for each song in this dataset the accompanying Musical Features made up from Danceability, Energy , Key, Tempo, Loudness, Acousticness and Valence will be extracted from Spotify for the Musical Features. A python script will be used in conjunction with the Spotipy package to scrape the required data. This data is then stored in a excel file.

Once both the cultural and musical data has been gathered it will then be pre-processed for use with the varying models. Pre-processing will be done on the text data contained in the Cultural Dataset to ensure that the models are learning from the actual text rather than discrepancies between texts, all data will be normalized and converted into the required shape for each of the models. Other pre-processing such as converting the release date into a new variable, days since release is also carried out to create the required final dataset.

The two datasets are then combined into the final dataset, which is split into 80% training data and 20% testing data.

3.2 Model Building

3.2.1 - Read in data and pre-process

The data is read, and pre-processing done in order to be used with the deep learning models the text data is converted to a standard format and is tokenized for use within the deep learning models. The data is normalized and finally converted into the required shape. Finally, the data is split into training and test.

3.2.2 - Design and Train models

The model parameters including number of layers, epochs , batch size and activation function are decided upon before building the models, for this project model parameters remain consistent across all models in order to ensure any change in performance is caused by the changes in model type or training data used. Sequential models are used for each model created as part of this project; sequential models build the model layers in a linear stack. The LSTM and RNN models are made up of one LSTM or RNN layer with 256 cells with Relu used as the activation function. The CNN models are slightly different from the other models with 1 convolutional layer containing 256 cells using Relu as the activation function like in the previous models and contains an additional Global 1-Dimensional Pooling layer. For all models , a dropout layer with dropout set to 0.1 is used to avoid over fitting and final a dense layer with 1 cell is added. The model is then trained for 200 epochs with batch size set to 16 in order to learn the patterns present in the data.

3.2.3 – Model Predictions and Evaluation Metric Calculation

The trained models are then used to make predict the sales values for the tests data. Once predictions have been made the three-evaluation metrics are calculated for each model. The first evaluation metrics is Accuracy measured as the percentage of predictions made by a model that are within ±5% of the actual sales figure . The second is Mean Absolute Error (MAE) which is the average or mean difference between the predicted and actual sales values . The third is Root Mean Square Error (RMSE) which is the square root of the squared mean difference between the predicted and actual sales values and is used to represent the sample standard deviation.

$$Accuracy = \frac{\sum_{i=1}^{n}(\ |PredictedSales_i - ActualSales_i| < 0.05 \ )}{n}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|PredictedSales_i - \ ActualSales_i|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(PredictedSales_i - ActualSales_i)^2}$$

Where:

n - total number of test cases

PredictedSales - The predicted sales figure

ActualSales – Actual Sales figure

The Accuracy evaluation metric is a custom metric implemented to make provided an understandable and simple metric to compare the various models. MAE and RMSE were chosen based off the paper done by Swalin [21].

3.3 Models Used

Model 1 – is an LSTM model designed using the parameters described in section 3.2.2, this model is trained using only the Cultural Data and is the used to predict the sales values of the test data using cultural data

Model 2 – is just like the first except it is an RNN model as opposed to an LSTM, as discussed in section 3.2.2 the LSTM and RNN models have the same overall setup. Once again, the model is trained using only Cultural Data and the predictions are made on the test data.

Model 3 – is a CNN model and is the final model to be trained using only the Cultural Data , sales values are predicted in order to calculate prediction performance just like with Models 1 and 2.

Model 4 – is an LSTM model and is identical to Model 1 except this time instead of the model being trained on the cultural data the model is instead trained using only the musical data. As usual the sales values for the test data are predicted and compared with the actual values to create the evaluation metrics.

Model 5 – is an RNN model trained only on the musical data and the other aspects are identical to the previous RNN model , Model 2.

Model 6 – is the final model to be trained solely using musical data , it is a CNN model, and all other aspect are the same as previous models.

Model 7 – is the last LSTM model as is trained using the combination of both musical and cultural data apart from the different training data Model 7 is identical to the previous LSTM models Model 1 and Model 5.

Model 8 – is the penultimate model and the concluding RNN model, this model is trained using the combined data and like all previous models is used to predict the sales figures of the data in the test set for evaluation purposes.

Model 9 – is the final model once again it is trained on the combined cultural and musical data and the trained model is used to predict the sales values of the test data before being evaluated.

3.4 Evaluation of Models:

Predicted sales values will be is compared with the actual amounts to estimate the model's accuracy and performance. The models are not expected to predict the exact sales figures instead any prediction within a 5% range above or below the actual sales amount will be considered as being correctly predicted. The ±5% range was chosen over a set value as it considers the discrepancies between sales and still requires predictions to be within close proximity to the actual sales figures. Additionally, for each of the models the Mean Absolute Error (MAE) and Root Mean Square Error(RMSE) will also be calculated so further comparison between the models can be carried out.

Comparison of performance is carried out on each of the three models to identify which model preforms best at predicting sales based for each type of training data. Results measured in terms of accuracy of sale prediction, MAE and RMSE are interpreted and compared to identify the best preforming model. R is used to plot the values to allow for visual interpretation of the results. A t-test is preformed to identify if there is statistically significant difference in performance between the model types, the plots of the various model performances alongside the results from the t-test are presented and discussed in the next section of this report.

Then, analysis will be carried out to identify which type of data is most useful when building deep learning models for music sales predictions. Once again results measured in terms of accuracy of sale predictions, MAE and RMSE will be compared to identify which type of data gives the most accurate predictions. Like with the model comparison plots of the performance metrics are presented using R and once again a t-test is run to identify if any difference in performance due to the change in data type is statistically significant.

3.5 Tools and Packages Used:

In this section a brief overview of the main programming languages and packages used is provided.

3.5.1 - Python

Python was chosen as the main programming language for this project for a variety of reasons,

1. Firstly, python contains a number of highly useful packages that are designed for data science and are particularly useful when designing and implementing Deep Learning Models Namely Pandas ,TensorFlow and Keras. These packages are easily available ,free and a large number of tutorials are available for them. They help to greatly reduce the complexity of the code and the time it takes to design and build deep learning models.

2. Python is a very popular programming language this meaning that a wide variety of tutorials and forum posts on sites such as stack overflow are available to help if required throughout the project if difficulties are to arise.

3.5.2 – Spotipy

Spotipy is Python library that is used to collect data from the Spotify Developer API, in order to use Spotipy one must first register for a Spotify developer account but once users have access to this, they can use Spotipy to seamlessly access the Web API. In this project Spotipy is used in order to collect the musical data related to each song in the dataset

3.5.3 – Pandas

Pandas is a Python library that provided a range of useful functions with regards to working with data structures and is a fundamental part of data manipulation and analytics. It is used throughout this project in order to store data as well as write data to excel or csv files and to read in data from excel or csv files as required.

3.5.4 - NumPy

NumPy is a Python library that provides a wide range of mathematical and scientific functions that are typically used with arrays

3.5.5 – TensorFlow & Keras

Is an end-to-end machine learning platform that allows for easy development and testing of machine learning models. Keras is high-level neural network APIs written in Python that is run on top of TensorFlow library. TensorFlow and Keras are utilised in combination to develop, built and test each of the nine models developed in this project.

3.5.6 - R

R is a statistical programming language typically used for data analytics and visualisation. R is used as the second programming language in this project to analyse and visualise the performance results of the deep learning models/

3.5.7 - Tidyverse

Tidyverse is a collection of R packages that are used for data analytics and visualisation. Tidyverse is used as part of this project to plot the evaluation metrics for each model to allow for comparison.

Chapter 4

Results:

4.1 Model Performance Results

4.1.1 – Long Short-Term Memory (LSTM) Performance Results

The LSTM models are made up of one LSTM layer with 256 cells and Relu used as the activation function , a dropout layer with dropout set to 0.1 is used to avoid over fitting and final a dense layer with 1 cell is added. The model is then trained for 200 epochs with batch size set to 16.

The LSTM models are fitted on the training data and then is used to predict the sales values for the test data. The predicted sales values are compared with the actual sales values of the test values to measure the performance of each of the LSTM models. Performance is measured based on 3 metrics, MAE (Mean Absolute Error) , RMSE (Root Mean Square Error) and Accuracy measured as the number of predicted sales that are within ±5% of the actual sales figure.

Data	MAE	RMSE	Accuracy
Cultural Data	269993.45	690409.05	52.19
Musical Data	218634.79	571987.97	41.85
Combined Data	108254.16	223797.28	68.49

Table 4.1.1 Table of performance results for LSTM model

4.1.2 – Recurrent Neural Network (RNN) Performance Results

The RNN models are made just like LSTM models with 1 RNN layer containing 256 cells and Relu used as the activation function , a dropout layer with dropout set to 0.1 is used to avoid over fitting and final a dense layer with 1 cell is added. The RNN model is trained for 200 epochs with batch size set to 16.

The trained RNN models are used to predict the sales just like with the LSTM models and performance is measured using the same 3 metrics.

Data	MAE	RMSE	Accuracy
Cultural Data	375102.34	861329.01	40.60
Musical Data	319851.14	767046.20	31.82
Combined Data	183142.04	450388.15	60.51

Table 4.1.2 Table of performance results for RNN model

4.1.3 – Convolutional Neural Network (CNN) Performance Results

The CNN models are slightly different from the previous models with 1 convolutional layer containing 256 cells using Relu as the activation function like in the previous models and contains an additional Global 1-Dimensional Pooling layer . Like with the LSTM and RNN models a dropout layer with dropout set to 0.1 is used to avoid over fitting and final a dense layer with 1 cell is added. The CNN model is trained for 200 epochs with batch size set to 16.

The trained CNN models are used to predict the sales just like with the LSTM and RNN models and performance is measured using the same 3 metrics.

Data	MAE	RMSE	Accuracy
Cultural Data	392190.49	849070.21	44.51
Musical Data	262718.69	623311.05	34.64
Combined Data	167084.29	489984.64	58.62

Table 4.1.3 Table of performance results for CNN model

4.2 Model Comparison

The evaluation results for each model in section 4.1 are compared in the graphs below, in each comparison the training data remains consistent for all models to ensure difference in performance is due to the variation in model type.

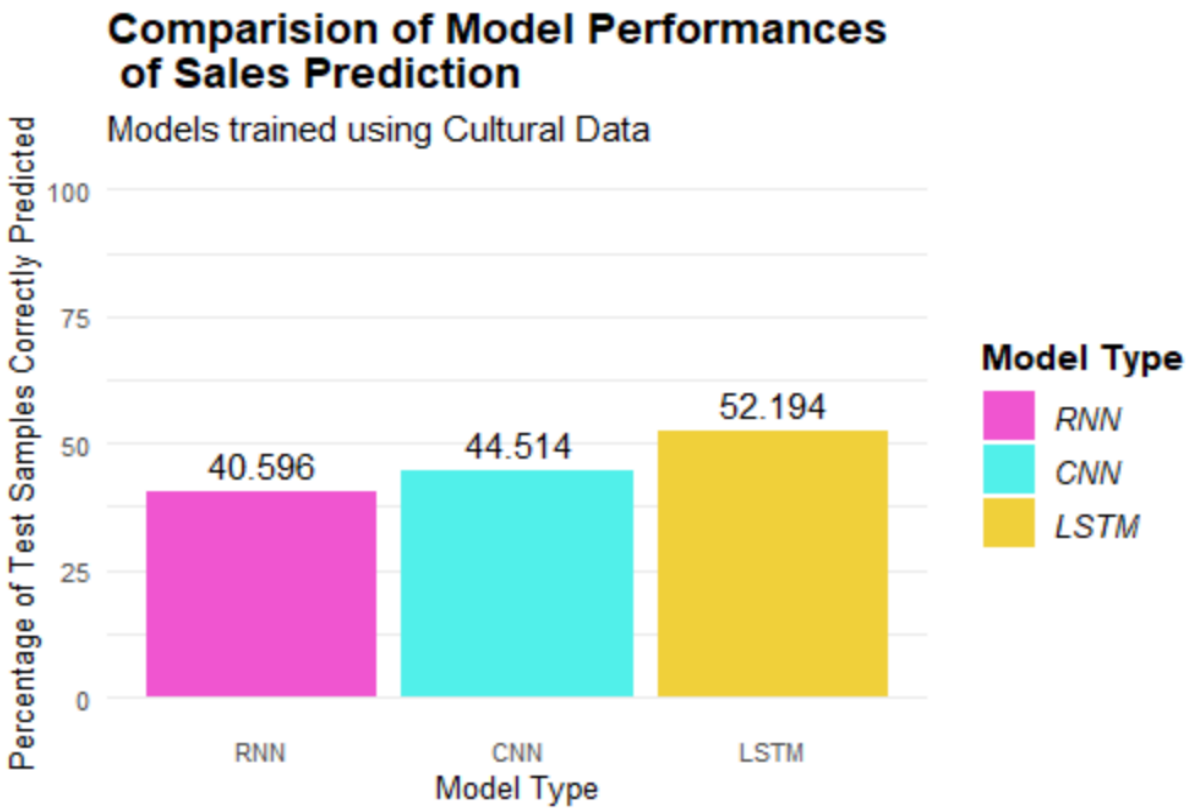

Figure 4.2.1 Plot of sales prediction Accuracy for each model type trained using cultural Data

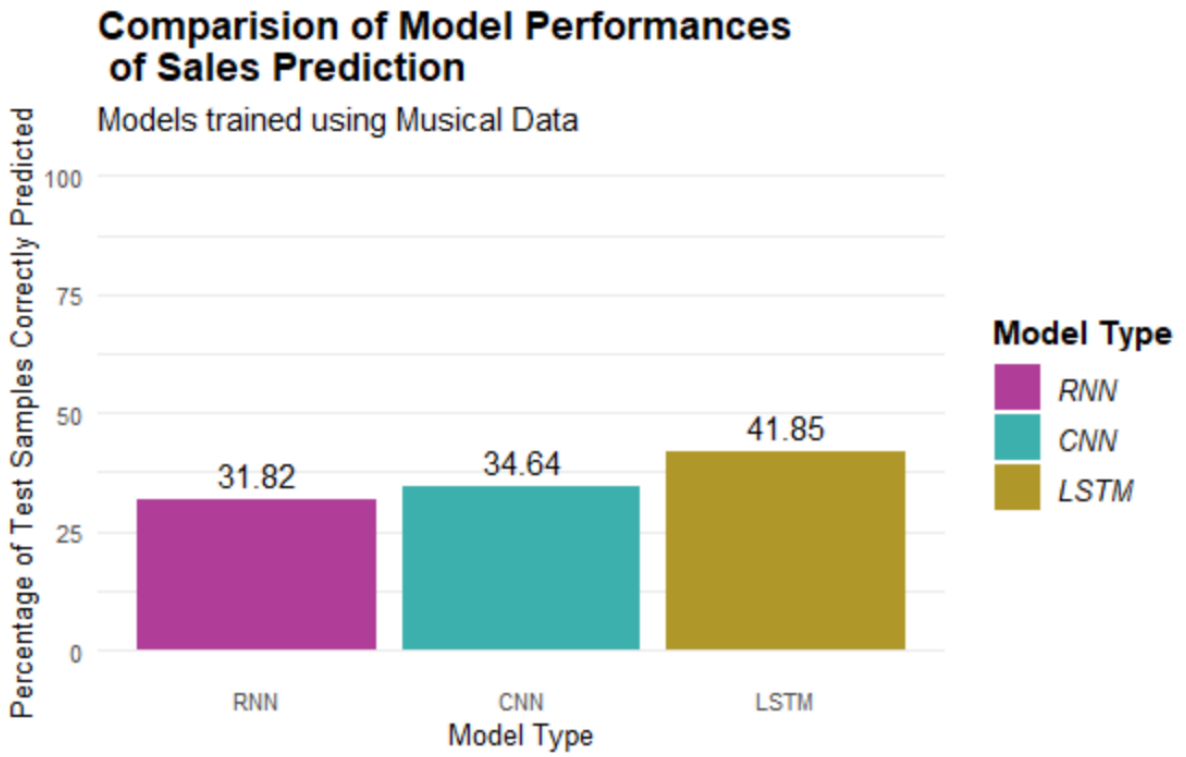

Figure 4.2.2 Plot of sales prediction Accuracy for each model type trained using musical Data

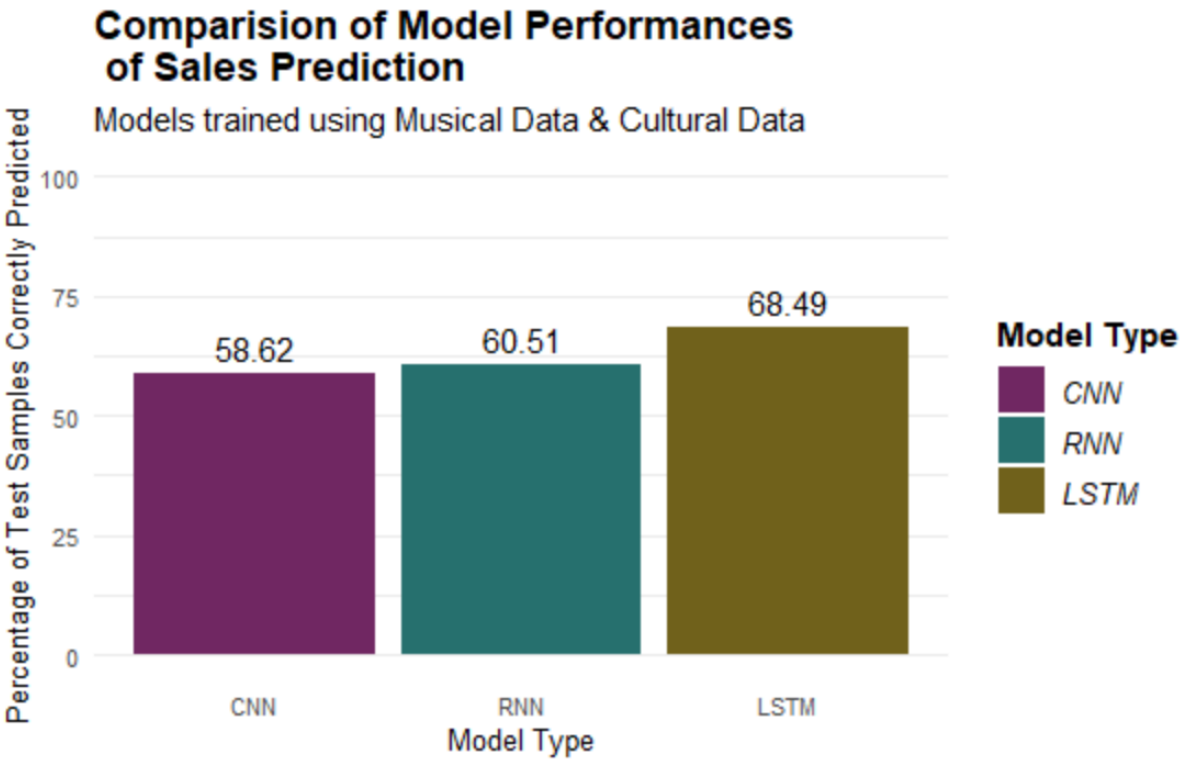

Figure 4.2.3 Plot of sales prediction Accuracy for each model type trained using combined Data

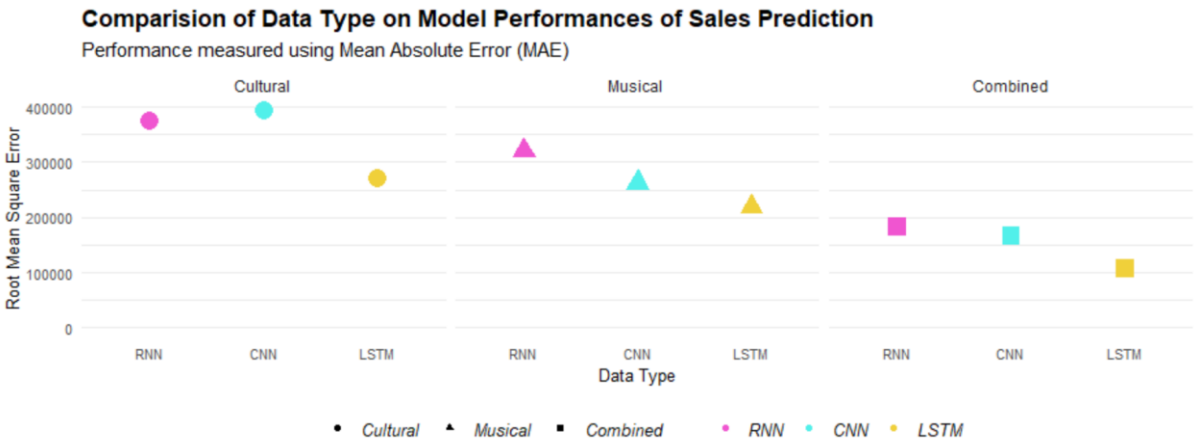

Figure 4.2.4 Plot of MAE values for each model type based on trained Data used

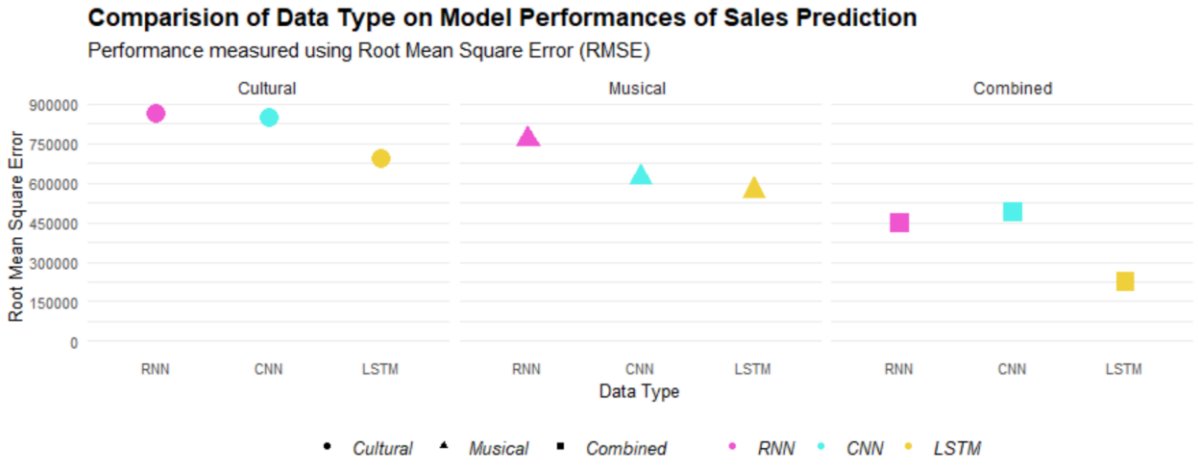

Figure 4.2.5 Plot of RMSE values for each model type based on trained Data used

As can be seen from each of the comparison plots above LSTM models achieve a higher percentage of accurate predictions (within the ±5% range) than both RNN and CNN models, for all 3 types of training data . LSTM models have on average 7% higher accuracy than CNN models and an average accuracy almost 10% higher than RNN models. MAE and RMSE values are also smaller for LSTM models when compared with the other model types further emphasising the increased prediction performance gained with LSTM models due to their smaller mean error values for both accurate (within ±5% range) and inaccurate (outside ±5% range) predictions, once again this improved performance is gained across all data types.

The results for the CNN models are almost identical to the results for the RNN models for predicting sales across all 3 data types suggesting no significant difference in performance between the 2 models , there is a maximum difference of around 4% between the 2 models while on average they only differentiate by approximately 2.5%. A similar situation is seen when comparing MAE and RMSE values where very similar values are achieved by both model types when trained on cultural data and the combination of cultural and musical data, however when trained on just musical data the CNN model seems to have a more significant performance than the RNN model when analysing MAE and RMSE values

For each of the 3 types of training data the model prediction performance is compared using a t-test to identify if there is a statistically significant difference in performance. The below table shows results from a t-test preformed to identify if the variation in performance between models is significant:

Models Compared	**Data**	**P-value**
LSTM ~ RNN	Cultural	0.393
LSTM ~ CNN	Cultural	0.00677
RNN ~ CNN	Cultural	0.0678
LSTM ~ RNN	Musical	0.00802
LSTM ~ CNN	Musical	0.0511
RNN ~ CNN	Musical	0.591
LSTM ~ RNN	Combined	0.393
LSTM ~ CNN	Combined	0.499
RNN ~ CNN	Combined	0.396

Table 4.2.1 Table of t-test results to identify if differences in performance caused by model type is statistically significant

When models are trained using cultural data the t-test shows that there is no statistically significantly improvement in performance was achieved by the LSTM model over the RNN model , however the LSTM model did preform significantly better than the CNN model. When the RNN model is compared to the CNN model the p-value of 0.0678 is slightly greater than the significance threshold of 0.05 , this suggests that there is a statistically significant difference between the models, and one could expect a better overall performance from the RNN model than the CNN model. These results show that despite the LSTM model having a better performance when comparing MAE, RMSE and accuracy (predictions within ±5% of actual) than the other model type it only significantly out preforms the CNN model. While CNN models have similar MAE , RMSE and Accuracy values to RNN models there is a significant difference in performance.

Comparison of the t-test values for models trained on musical data that this time the LSTM model does significantly outperforms both the RNN and CNN model , (the p-value for LSTM ~ CNN comparison is close enough to the threshold to be accepted in these circumstances). However, unlike when trained using cultural data there is no significant difference between the RNN and CNN models. This aligns with what we would expect to see from the analysis of the models using MAE, RMSE and prediction accuracy.

Finally, when models are trained on the combined data containing musical and cultural data despite differences in MAE , RMSE and accuracy value being observed particularly from the LSTM model the t-test suggests that there is no statistically significant change in prediction performance between the models.

4.3 Data Comparison

The evaluation results for each model in section 4.1 are compared in the graphs below, in each comparison the model remains consistent while training data is changed to ensure difference in performance is due to the variation in training data.

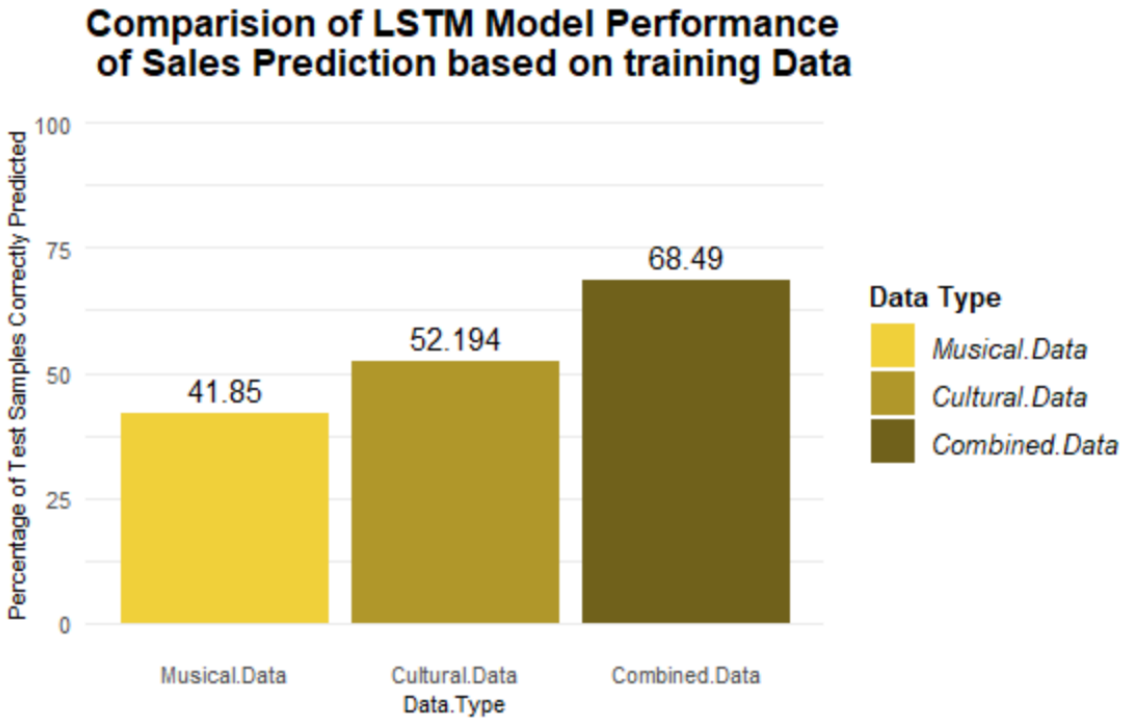

Figure 4.3.1 Plot of sales prediction Accuracy for LSTM models based on trained data used

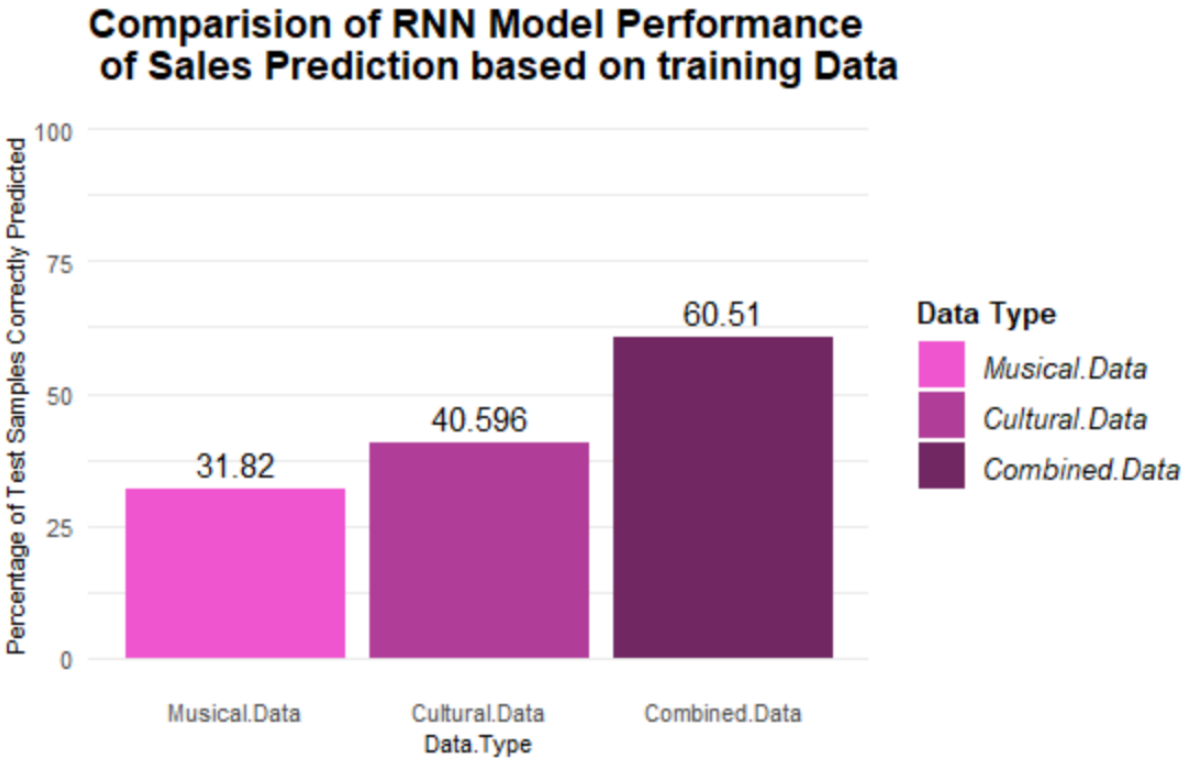

Figure 4.3.2 Plot of sales prediction Accuracy for RNN models based on trained data used

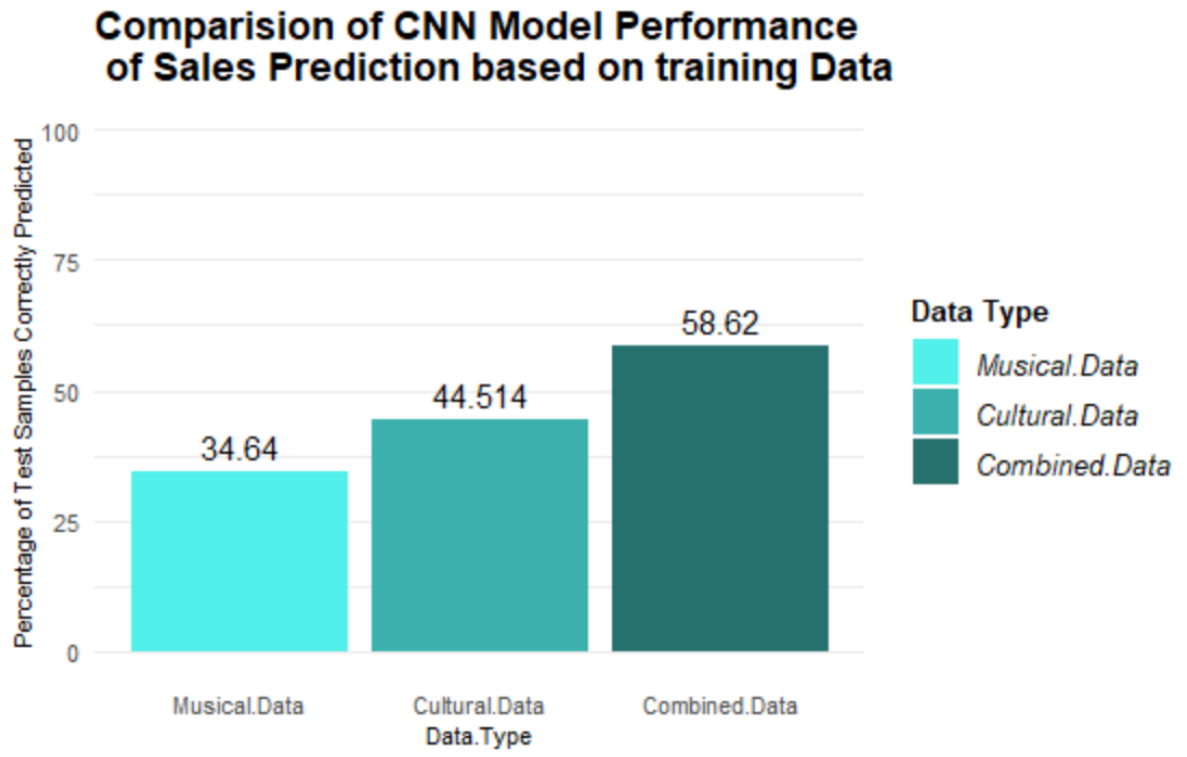

Figure 4.3.3 Plot of sales prediction Accuracy for CNN models based on trained data used

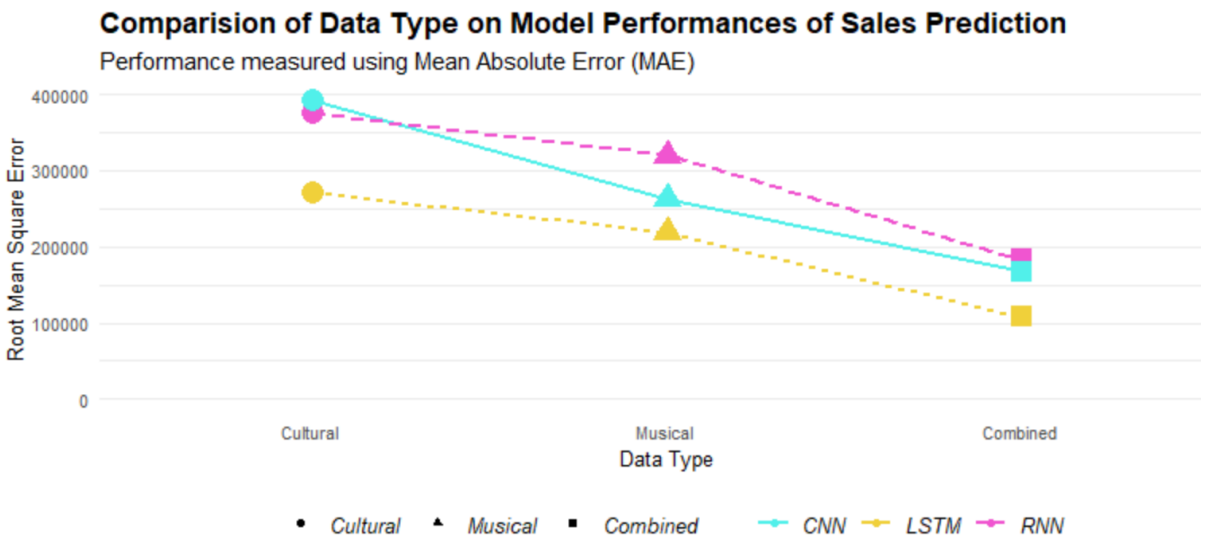

Figure 4.3.4 Plot of MAE values for each type of training data based on model

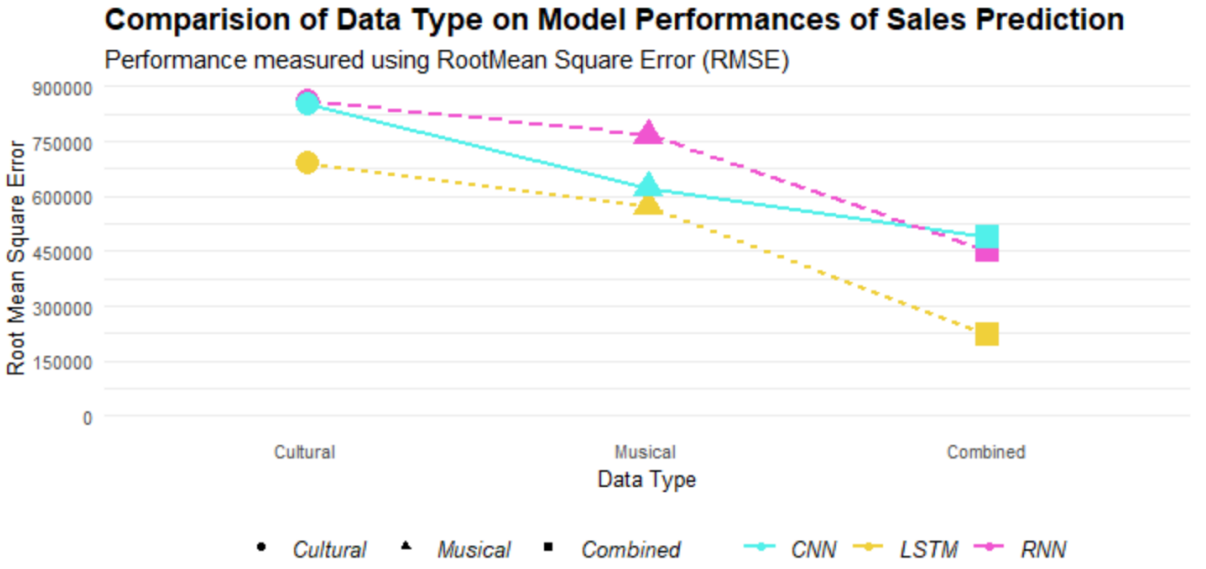

Figure 4.3.5 Plot of RMSE values for each type of training data based on model

From the plots above models trained using musical data have the worst accuracy scores. However, models trained on musical data have very comparable MAE and RMSE values to those trained using cultural data suggesting there might not be much difference in overall performance. Models trained using combined data achieves the best results having the highest accuracy and lowest MAE and RMSE values of the 3 data types.

Unlike with models the type of data appears to have a more significant effect of performance , in order to confirm this assumption a t-test is performed on the evaluation results:

Models Compared	**Model**	**P-value**
Cultural ~ Musical	LSTM	0.179
Cultural ~ Combined	LSTM	0.00427
Musical ~ Combined	LSTM	0.0974
Cultural ~ Musical	RNN	0.437
Cultural ~ Combined	RNN	0.0000168
Musical ~ Combined	RNN	0.00000603
Cultural ~ Musical	CNN	0.148
Cultural ~ Combined	CNN	0.000000502
Musical ~ Combined	CNN	0.00315

Table 4.3.1 Table of t-test results to identify if differences in performance caused by training data is statistically significant

The results from the t-tests run to compare the impact of the type of training data of model performance are presented in the table above 4.3.1. For all model types there is no statistically significant difference is observed between model trained using cultural or musical data. On the other hand, for all models using combined data achieves statistically superior performance than when cultural data is used , similarly for both RNN and CNN models the combined data significantly outperforms the musical data and for the LSTM model while not below the 0.05 threshold the low p-value of 0.09 suggests that a better performance is once again achieved by all models.

Chapter 5

Conclusion and Future work:

5.1 Conclusion

In this research a comparative study is done to between LSTM RNN and CNN models for song sale figure predictions , and the effect of the type of training data utilised Cultural , Musical or a combination of both. A comprehensive review of previous work carried out in the domain of performance prediction for songs was carried out and a brief overview of these works is provided. This previous work is the foundation for the decisions made during the completion of this work. The data used for training and testing the models was curated from various sources which were identified as being good representations of the wider areas of cultural data and musical data and combined to create the final dataset. Then models were created in order to identify which model type and which combination of training data provided the best results. LSTM models outperformed their counterparts in the RNN and CNN models , achieving more precise predictions as can be seen from the lower mean average error and root mean square error values. Also, the LSTM models obtain higher accuracy scores by predicting a higher percentage of sales within a 5% range of the actual figure. Despite the LSTM models enhanced performance , t-tests show that it doesn't achieve statistically better results than the other models notably when models are trained on the combined data.

Comparably models trained using the combined data containing both cultural and musical data outclass the models trained on only a single type of data attaining lower mean average error and root mean square error values and higher accuracy scores across all models. T-tests show that the increased performance achieved by models trained using the combined data is statistically significant. This suggests that cultural and musical data are linked and that combining them allows models to learn potential difference between aspects of musical data such as tempo and cultural data such as artist or genre leading to improved performance.

Overall, the LSTM model trained using the combined musical and cultural data achieves the best results , MAE - 108254.16 , RMSE - 223797.28 and Accuracy – 83.7%. Despite not achieving a statistically significant improvement over the other models trained on using the same data, the LSTM model achieves the lowest MAE and RMSE values and the highest Accuracy values. When the models are trained on the other data types the LSTM model does statistically outperform the other models

Overall, this work has managed to answer the research questions it outlined at the start of the project. Firstly, it has shown that Deep learning models can be used to accurately predict sales figures using cultural and musical data. Secondly LSTM models as previously discussed tend to outperform the other models tested and finally it found that the combination of both cultural and musical data leads to the overall best performance.

5.2 Future Work

If someone was going to further the research carried on as part of this project the following are improvements that they could consider to potentially improve upon this project :

- For comparison purposes model parameter were kept consistent throughout the training process , this means the model parameters were not adjust in order to find the best parameters for each model as this could have led to discrepancies caused by model parameters, work could be conducted in order to identify which specific model parameters produce the best results for each of the model and data types

- Analysis of the data could be performed to identify which parts of the cultural and musical data are most statistically significant and if this varies across genres and artists.

- As discussed at the start of the document it was decided to focus on predicting sales data for this project however future work could focus on streaming figures and see if the designed models are able to accurately predict streaming figures and compare performance of models predicting sales to those predicting streams

- Finally, no aspect of social media data was included in the cultural data for this project, work carried out by Kim, Sun, and Lee [2] has shown the impact Twitter has on the position and longevity of a song within the billboard top 100 charts, future work could include information from social media source such as twitter followers or tweets within the cultural data in order to identify the impact this would have on the models developed.

Code

The code along with the data used as part of this project can be found in the following GitHub repository :

https://github.com/stephen000000/DataAnalytics_Thesis.git

The GitHub repository contains:

1. The training data compromised of the DBpedia data and Spotify data.

2. The Results Data which contains the sales predictions made by each model along with the calculated accuracy, MAE , RMSE values

3. T-test values

4. Python Script for scraping Spotify

5. Python Script for creating the final Dataset

6. Python Script for building and testing the Models

7. A number or R scripts for plotting the performance results

References

1. Gmr2021.ifpi.org. 2021. *IFPI GLOBAL MUSIC REPORT 2021*. Available at: <https://gmr2021.ifpi.org/>

2. Kim, Y., Suh, B. and Lee, K., 2014. # nowplaying the future Billboard: mining music listening behaviors of Twitter users for hit song prediction. In *Proceedings of the first international workshop on Social media retrieval and analysis* (pp. 51-56).

3. Statista Infographics. 2021. *Infographic: The World's Largest Music Streaming Service?*. Available at: <https://www.statista.com/chart/5866/online-music-listening-platforms/>

4. Statista Infographics. 2021. *Spotify users - subscribers in 2020 | Statista*. Available at: <https://www.statista.com/statistics/244995/number-of-paying-spotify-subscribers/>

5. Developer.spotify.com. 2021. *Web API Reference | Spotify for Developers*. Available at: <https://developer.spotify.com/documentation/web-api/reference/#category-tracks>

6. Abel, F., Diaz-Aviles, E., Henze, N., Krause, D. and Siehndel, P., 2010, August. Analyzing the blogosphere for predicting the success of music and movie products. In *2010 International Conference on Advances in Social Networks Analysis and Mining* (pp. 276-280). IEEE.

7. Dhar, V. and Chang, E.A., 2009. Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, *23*(4), pp.300-307.

8. Lee, J., Boatwright, P. and Kamakura, W.A., 2003. A Bayesian model for prelaunch sales forecasting of recorded music. *Management Science*, *49*(2), pp.179-196.

9. Lee, M., Choi, H., Cho, D. and Lee, H., 2016. Cannibalizing or complementing?. The impact of online streaming services on music record sales. *Procedia Computer Science*, *91*, pp.662-671.

10. J. Lee and J. Lee, Nov. 2018. Music Popularity: Metrics, Characteristics, and Audio-Based Prediction, in *IEEE Transactions on Multimedia*, vol. 20, no. 11, pp. 3173-3182,

11. Nijkamp, R., 2018. *Prediction of product success: explaining song popularity by audio features from Spotify data*.

12. Herremans, D., Martens, D. and Sörensen, K., 2014. Dance hit song prediction. *Journal of New Music Research*, *43*(3), pp.291-302.

13. Whitman, B. and Smaragdis, P., 2002, October. Combining Musical and Cultural Features for Intelligent Style Detection. In *ISMIR*.

14. McKay, C. and Fujinaga, I., 2008, September. Combining Features Extracted from Audio, Symbolic and Cultural Sources. In *ISMIR* (pp. 597-602).

15. Dhanaraj, R. and Logan, B., 2005, September. Automatic Prediction of Hit Songs. In *ISMIR* (pp. 488-491).

16. Education, IBM., 2021. *What are Neural Networks?*. Available at: <https://www.ibm.com/cloud/learn/neural-networks>

17. Education, IBM., 2021. *What are Convolutional Neural Networks?*. Available at: <https://www.ibm.com/cloud/learn/convolutional-neural-networks>

18. S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, *2017*. Stock price prediction using LSTM, RNN and CNN-sliding window model, *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 1643-1647

19. Education, IBM., 2021. *What are Recurrent Neural Networks?*. Available at: <https://www.ibm.com/cloud/learn/recurrent-neural-networks>

20. Stanford. 2021. *Understanding LSTM Networks* Available at: <https://web.stanford.edu/class/cs379c/archive/2018/class_messages_listing/content/Artificial_Neural_Network_Technology_Tutorials/OlahLSTM-NEURAL-NETWORK-TUTORIAL-15.pdf>

21. Swalin, A., 2021. *Choosing the Right Metric for Evaluating Machine Learning Models — Part 1*. [online] Medium. Available at: <https://medium.com/usf-msds/choosing-the-right-metric-for-machine-learning-models-part-1-a99d7d7414e4>

22. TechRepublic. 2021. *Why Python is so popular with developers: 3 reasons the language has exploded*. [online] Available at: <https://www.techrepublic.com/index.php/price/freetotry/index.php/article/why-python-is-so-popular-with-developers-3-reasons-the-language-has-exploded/>