SALES FORECASTING SYSTEM

A PROJECT REPORT

Submitted by

SANKETH M [CB.EN.U4CSE19242]
A ADITYA SAI SANDEEP REDDY [CB.EN.U4CSE19402]
N KARTHIK KUMAR REDDY [CB.EN.U4CSE19444]
MARRI HARANATH REDDY [CB.EN.U4CSE19465]

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

AMRITA SCHOOL OF COMPUTING

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112

JUNE 2023

AMRITA VISHWA VIDYAPEETHAM

AMRITA SCHOOL OF COMPUTING, COIMBATORE – 641 112

BONAFIDE CERTIFICATE

This is to certify that the project report entitled **SALES FORECASTING SYSTEM** submitted by Sanketh M(CB.EN.U4CSE19242),A Aditya Sai Sandeep(CB.EN.U4CSE19402),N Karthik Kumar,(CB.EN.U4CSE19444) Marri Haranath Reddy(CB.EN.U4CSE19465) in partial fulfillment of the requirements for the award of Degree **Bachelor of Technology** in Computer Science and Engineering is a bonafide record of the work carried out under our guidance and supervision at the Department of Computer Science and Engineering, Amrita School of Computing, Coimbatore.

Ms. Sathiya R. R.
(Associate Professor)
Department of CSE

Dr. Vidhya Balasubramanian
Chairperson
Department of CSE

Evaluated on:

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

We, the undersigned solemnly declare that the project report **SALES FORECASTING SYSTEM** is based on our own work carried out during the course of our study under the supervision of Ms. Sathiya R. R., (Associate Professor), Computer Science and Engineering, and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgement has been made wherever the findings of others has been cited.

Sanketh M[CB.EN.U4CSE19242]- signature
A Aditya Sai Sandeep Reddy [CB.EN.U4CSE19402]- signature
N Karthik Kumar Reddy[CB.EN.U4CSE19444]- signature
Marri Haranath Reddy[CB.EN.U4CSE19465]- signature

ABSTRACT

This study examines the development of a machine learning-based sales forecasting system with a focus on the system's ability to assess historical data and identify significant sales performance indicators. A universal approach is not workable due to the distinctive characteristics of each firm. We recommend a customised approach, emphasising the requirement for an adaptable and customizable forecasting module. Solutions that are specific to the industry are needed to improve forecast accuracy. Traditional methods of model replication across various industries do not work because of their distinctive characteristics.Our customised approach aims to give businesses precise sales projections and insight into crucial success factors, helping them create informed sales and resource allocation strategies that will eventually spur growth and enhance sales performance.

We further investigate the methodology of the proposed machine learning-based sales forecasting system. By examining historical data, we can detect patterns and trends that serve as excellent indicators of how sales will perform in the future. A thorough grasp of the dynamism and complexity of sales performance, as well as the factors influencing it, is part of the research. It involves more than simply the statistics. .

In order to improve sales forecasting accuracy, the current situation necessitates sector-specific solutions. We make it apparent that since every industry has different characteristics and variables, just copying models from one industry to another won't work. Instead, we advocate for a customised approach that takes into account the specifics of each industry. .

ACKNOWLEDGEMENT

We would like to express our deep gratitude to our beloved Satguru **Sri Mata Amritanandamayi Devi** for providing the bright academic climate at this university, which has made this entire task appreciable. This acknowledgment is intended to thank all those people involved directly or indirectly with our project. We would like to thank our Pro Chancellor **Swami Abhayamritananda Puri**, Vice Chancellor **Dr.Venkat Rangan.P** and **Dr.Bharat Jayaraman**, Dean, Faculty of AI & Computing(Computing), Amrita Vishwa Vidyapeetham for providing us with the necessary infrastructure required for the completion of the project. We express our thanks to **Dr.Vidhya Balasubramanian**, Chairperson, Department of Computer Science Engineering and Principal, School of Computing, Amrita Vishwa Vidyapeetham, **Dr.C.Shunmuga Velayutham and Dr.N.Lalithamani,** Vice Chairpersons of the Department of Computer Science and Engineering for their valuable help and support during our study. We express our gratitude to our guide, **Ms. Sathiya R.R.** for her guidance, support and supervision. We feel extremely grateful to **Dr. Rajathilagam B , Ms.Prathilothamai M , Mr.R. Krishankumar, Ms.Shanmuga Priya S.** for their feedback and encouragement which helped us to complete the project. We would also like to thank the entire fraternity of the Department of Computer Science and Engineering. We would like to extend our sincere thanks to our family and friends for helping and motivating us during the course of the project. Finally, we would like to thank all those who have helped, guided and encouraged us directly or indirectly during the project work. Last but not the least, we thank God for his blessings which made our project a success.

Sanketh M[CB.EN.U4CSE19242]
A Aditya Sai Sandeep Reddy[CB.EN.U4CSE19402]
N Karthik Kumar Reddy[CB.EN.U4CSE19444]
Marri Haranath Reddy[CB.EN.U4CSE19465]

TABLE OF CONTENTS

ABSTRACT **iv**

ACKNOWLEDGEMENT **v**

List of Tables **viii**

List of Figures **x**

1 Introduction **xi**

2 Literature Survey **xii**
2.1 A. Overview of Sales Forecasting Research xii

3 Methodology **xiv**
3.1 Data Collection and Description xiv
3.1.1 Video Game Sales Dataset xv
3.1.2 E-Commerce Sales Dataset xviii
3.1.3 Walmart Sales Dataset xx
3.1.4 BigMart Sales Dataset xxi
3.2 Overview of Machine Learning Models xxiii

4 Implementation and Testing **xxiv**
4.1 Model Application and Evaluation xxiv
4.1.1 Model 1: LSTM Performance and Results xxv
4.1.2 Model 2: CNN Performance and Results xxix
4.1.3 Model 3: RNN Performance and Results xxxiii
4.1.4 Model 4: GRU Performance and Results xxxvii
4.2 Model 5: Hybrid Models Performance and Results xli
4.2.1 CNN-RNN xli
4.2.2 LSTM-FCN Model xlii
4.2.3 1D-CNN-LSTM Model xliii
4.2.4 1D-CNN-FNN xliv
4.3 Ensemble Models Performance and Results xlv
4.3.1 Ensemble Models of CNN, RNN and LSTM xlv
4.3.2 LSTM with Attention Mechanism xlvi
4.3.3 Variational Autoencoder (VAE) xlvii

5 Results and Discussion **xlviii**
5.1 Comparison of Model Performance xlviii
5.2 Customized Approach for Different Datasets l

6 Conclusion **li**

7 Future Enhancement **lii**

LIST OF TABLES

3.1 Video Game Sales Data Feature Table xv
3.2 E Commerce Sales Data Feature Table xviii

4.1 Performance of LSTM Model on Video Game Sales Dataset xxv
4.2 Performance of LSTM Model on BigMart Sales Dataset xxvi
4.3 Performance of LSTM Model on Walmart Game Sales Dataset . . . xxvii
4.4 Performance of LSTM Model on E-Commerce Dataset xxviii
4.5 Performance of CNN Model on Video Game Sales Dataset xxix
4.6 Performance of CNN Model on Bigmart Sales Dataset xxx
4.7 Performance of CNN Model on Walmart Game Sales Dataset xxxi
4.8 Performance of CONV 1D on E-Commerce Dataset xxxii
4.9 Performance of RNN Model on Video Game Sales Dataset xxxiii
4.10 Performance of RNN Model on BigMart Sales Dataset xxxiv
4.11 Performance of RNN Model on Walmart Game Sales Dataset xxxv
4.12 Performance of RNN Model on E-Commerce Dataset xxxvi
4.13 Performance of GRU Model on Video Game Sales Dataset xxxvii
4.14 Performance of GRU Model on BigMart Sales Dataset xxxviii
4.15 Performance of GRU Model on Walmart Game Sales Dataset xxxix
4.16 Performance of GRU Model on E-Commerce Dataset xl
4.17 Performance of CNN-RNN Hybrid Model on Video Game Sales Dataset xli
4.18 Performance of LSTM-FCN Model on E-Commerce Dataset xlii
4.19 Performance of 1D-CNN-LSTM Model on BigMart Sales Dataset . xliii
4.20 Performance of 1D-CNN-FNN Model on BigMart Sales Dataset . . xliv
4.21 Performance of Ensemble Models of CNN, RNN and LSTM on Video Game Sales Dataset xlv
4.22 Performance of Ensemble Models of CNN, RNN and LSTM on Video Game Sales Dataset xlvi
4.23 Performance of CNN Model on Walmart Game Sales Dataset xlvii

5.1 RMSE - Part 1 xlviii
5.2 RMSE - Part 2 xlviii
5.3 MAE - Part 1 xlviii
5.4 MAE - Part 2 xlix

LIST OF FIGURES

3.1 Yearly Global Sales Trend xvi
3.2 Genre Global Sales Trend xvi
3.3 Sales Genre Correlation Heat Map xvii
3.4 Platform Global Sales Trend xvii
3.5 Publisher Global Sales Trend xvii
3.6 Feature Correlation Map xvii
3.7 Money Spent vs Country xix
3.8 Density Plot xix
3.9 Quantity vs Density xix
3.10 Invoice Date vs Sales xix
3.11 Weekly Average Sales vs Holidays xx
3.12 Monthly Sales vs Years xx
3.13 Item outlet sales vs Outlet size xxi
3.14 Item outlet sales vs Outlet location type xxii
3.15 Item outlet sales vs Outlet type xxii
3.16 Outlet Identifier based on sales xxii
3.17 Item type based on sales xxii

4.1 Video Game data LSTM Plot() xxv
4.2 Video Game data Loss plot xxv
4.3 BigMart data training and validation LSTM plot xxvi
4.4 BigMart data Loss Plot xxvi
4.5 Walmart LSTM Training and Validation RMSE Plot xxvii
4.6 Walmart LSTM Training and Validation Loss Plot xxvii
4.7 Walmart LSTM Training and Validation MAE Error Plot xxvii
4.8 Ecommerce LSTM Training and Validation Accuracy Plot xxviii
4.9 Ecommerce LSTM Training and Validation Loss Plot xxviii
4.10 VideoGames CNN Training Plot xxix
4.11 VideoGames CNN Loss Plot xxix
4.12 BigMart CNN Training,Validation loss xxx
4.13 BigMart CNN Training,Validation Plot xxx
4.14 Walmart CNN RMSE Plot xxxi
4.15 Walmart CNN Loss Plot xxxi
4.16 Walmart CNN MAE Plot xxxi
4.17 Ecommerce CNN Training and Validation Accuracy Plot xxxii
4.18 Ecommerce CNN Training and Validation Loss Plot xxxii
4.19 VideoGame RNN Accuracy Plot xxxiii
4.20 VideoGame RNN Loss Plot xxxiii
4.21 BigMart RNN Training and Validation MAE Plot xxxiv
4.22 BigMart RNN Training and Validation RMSE Plot xxxiv
4.23 Walmart RNN RMSE Metric Plot xxxv
4.24 Training and Validation Loss Plot xxxv
4.25 Walmart dataset RNN MAE Plot xxxv

4.26 Ecommerce RNN Training and Validation Plot xxxvi
4.27 Ecommerce RNN Loss Plot xxxvi
4.28 VideoGames GRU Accuracy Plot xxxvii
4.29 VideoGames GRU Loss Plot xxxvii
4.30 BigMart GRU Training and Validation Metrics Plot xxxviii
4.31 BigMart GRU Training and Validation Loss Plot xxxviii
4.32 Walmart GRU Training and Validation Loss Plot xxxix
4.33 Ecommerce data GRU Accuracy Plot xl
4.34 Ecommerce data GRU Loss Plot xl
4.35 VideoGames CNN-RNN RMSE Plot xli
4.36 VideoGames CNN-RNN MAE Plot xli
4.37 Ecommerce LSTM-FCN Accuracy Plot xlii
4.38 Ecommerce LSTM-FCN Loss Plot xlii
4.39 BigMart CNN-LSTM Training and Validation Metrics Plot xliii
4.40 BigMart CNN-FNN Training and Validation Metrics Plot xliv
4.41 Training and Validation RMSE Plot xlv
4.42 Training and Validation MAE Plot xlv
4.43 VideoGames LSTM attention Loss Plot xlvi
4.44 Walmart VAE Training and Validation Loss Plot xlvii

Chapter 1

INTRODUCTION

Forecasting sales is a critical component of any successful firm. It acts as a road map, assisting businesses in making educated choices about resource allocation, inventory management, budgeting, and strategic direction. However, projecting sales is a difficult task that depends on a wide range of factors, making it difficult to navigate and synthesise manually. Therefore, it has become increasingly important in the modern corporate landscape to integrate technology, especially machine learning.

This research focuses on the creation of a machine learning-based sales forecasting module that uses previous sales data to determine important performance indicators, allowing firms to estimate future sales patterns. The system offers insights that act as the cornerstone for making strategic business decisions by examining past performance and sales patterns.The fact that companies operate in a variety of markets with distinct characteristics, market dynamics, and consumer behaviour must be acknowledged. Due to this variability, a flexible forecasting system that meets the specific requirements of each industry is required. .

Due to their inability to adapt to varied business situations, one-size-fits-all models frequently fail to give correct results. In order to overcome the problems caused by traditional forecasting techniques, we investigate the creation and application of particular, machine learning-based forecasting models. We provide a thorough analysis of how solutions tailored to a particular industry can raise prediction accuracy noticeably, enabling companies to plan properly, spur growth, and improve sales performance.

Chapter 2

LITERATURE SURVEY

2.1 A. Overview of Sales Forecasting Research

The accuracy of the prediction is a critical component of sales forecasting. As a result, several efforts have been undertaken to improve this method' accuracy. After comparing the actual sales with the projected outcomes, publications in this thesis use error measuring techniques like RMSE and MAPE to determine accuracy. Several pertinent publications are reviewed in this section.In the field of time-series forecasting, there is no single technique that can be considered the best for solving all problems. Each forecasting problem may require a different approach. The Recurrent Neural Network (RNN) has gained significant popularity for time-series prediction in recent times. It effectively captures temporal dependencies by incorporating loops within the network, enabling the retention of information from past events Gamboa (2017). Among the various RNN variations, the Long Short-Term Memory (LSTM) has garnered substantial attention as it addresses the limitations of short-term memory. Malhotra et al. (2016) proposed an LSTM-based model for detecting anomalies in time-series data, showcasing the effectiveness of LSTM in capturing complex patterns. In the field of forecasting, Zhao et al. (2017)[3] demonstrated the efficacy of LSTM in accurate prediction tasks.Moreover, LSTM has found application in studying the impact of external events on time-series data. Ghosh and Sanyal (2021) recently employed LSTM to predict market fear and identify hidden patterns of influence during the COVID-19 timeline, showcasing its potential in analyzing and forecasting dynamic scenarios.These studies highlight the versatility and effectiveness of LSTM in time-series forecasting and anomaly detection tasks, showcasing its potential to capture long-term dependencies and intricate patterns in the data.In recent years, Convolutional Neural Networks (CNNs) have gained attention as a powerful deep learning technique for time-series forecasting. Zhao and Wang (2017) proposed an innovative approach that leveraged CNNs to automatically extract effective features from data, employing this method for sales forecasting.

Their study demonstrated the effectiveness of CNNs in capturing patterns and making accurate predictions in the context of sales forecasting.The application of CNNs extends beyond sales forecasting. Selvin et al. (2017) proposed a deep learning-oriented model for stock price forecasting, utilizing CNNs to detect changes in trends. Their research highlighted the ability of CNNs to capture complex relationships and identify significant variations in stock market data.

Furthermore, CNNs have been applied to short-term forecasting tasks in the domain of renewable energy. Koprinska et al. (2018) explored the use of CNNs for short-term forecasting of solar power and electricity. Their findings indicated that CNNs, along with multilayer perceptron neural networks, outperformed other traditional forecasting approaches, emphasizing the superiority of CNNs in capturing temporal patterns and producing accurate forecasts in renewable energy domains.To enhance the robustness of time-series forecasting, Momeny et al. (2021) developed a CNN-based method that considered the inherent uncertainties and variability of the problem. Their approach demonstrated improved forecasting performance by incorporating robustness considerations into the CNN framework.

Some other advanced models such as the Long Short Term Memory Fully Convolutional Network (LSTM-FCN) and the Attention LSTM-FCN, have been introduced by Karim et al. (2019) and further developed by Karim et al. (2018) and . These deep learning models pair a Fully Convolutional Network (FCN) with a Long Term Short Term Recurrent Neural Network (LSTM) to perform time series dataset classification. Ensemble techniques also have been recently utilized for product demand forecasting. For instance, Seyedan et al. (2022) advanced an ensemble learning method to forecast product demand using cluster-based Bayesian Model averaging. This strategy involves customer segmentation based on their buying behaviors like recency, frequency and monetary attributes. In a different study, Sharma and Shafiq (2020) introduced an ensemble learning model consisting of Random CNN,Random Forest, XGBoost. Their approach surpassed the existing solutions in performance when evaluated.

Chapter 3

METHODOLOGY

3.1 Data Collection and Description

We collected and utilized four distinct datasets from reputable sources, each representing a different industry. The datasets chosen for analysis include the:

- Video Game Sales Dataset
- Bigmart Sales Dataset
- Ecommerce Sales Dataset
- Walmart Sales Dataset

These datasets were sourced from Kaggle, a well-known platform for accessing and sharing datasets, ensuring transparency and credibility in the data collection process.

The selection of these datasets was purposeful, aiming to include a range of industries and business sectors. By working on datasets from different industries, we aimed to test the effectiveness of our sales forecasting module across diverse contexts. These datasets serve as valuable resources for understanding the key drivers and metrics that significantly impact sales performance within each industry. We explore the specific attributes, variables, and factors contained within these datasets, shedding light on the unique aspects of each industry.

3.1.1 Video Game Sales Dataset

The Video Game Sales Dataset provides a comprehensive overview of the sales performance of various video games across different platforms, genres, and regions. This dataset includes a wide range of information, including rankings, game titles, release platforms, release years, genres, publishers, and sales figures in North America, Europe, Japan, and the rest of the world. Additionally, it includes the total global sales for each game.

This serves as a valuable resource for understanding the dynamics of the video game industry, as it captures the popularity and commercial success of games across different regions and platforms. Researchers and industry professionals can gain insights into the factors that contribute to the sales performance of video games.

The dataset also provides an indication of the relative success of each game compared to others. The platform information highlights the diverse range of platforms on which games are released, such as PC, PS4, and others. By examining the release years, users can identify trends and patterns in game sales over time.

Table 3.1: Video Game Sales Data Feature Table

Feature	**Interpretation**
Rank	Ranking of overall sales, integer
Name	The game's name
Platform	Platform of the game's release
Year	Year of the game's release
Genre	Genre of the game
Publisher	Publisher of the game
NA_Sales	Sales in North America (in millions)
EU_Sales	Sales in Europe (in millions)
JP_Sales	Sales in Japan (in millions)
Other_Sales	Sales in the rest of the world (in millions)
Global_Sales	Total worldwide sales

Making decisions based on data greatly benefits from business intelligence (BI), which provides useful information into decision making of a business operation. Businesses can use BI to find patterns, trends, and spikes in their data by utilising the power of data analytics. This allows them to make informed and better decisions. By looking at visualisation charts and making inferences from them to use on their businesses, we can do that.

We now concentrate on the use of business intelligence methods to examine a particular dataset pertaining to video game sales. We seek to derive significant insights to understand the fundamental structure of the video game business via the use of various visualisation plots and statistical observations.

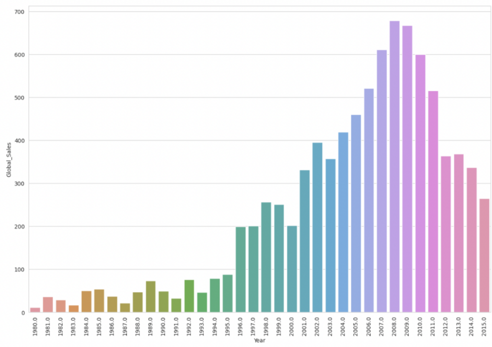

Figure 3.1: Yearly Global Sales Trend

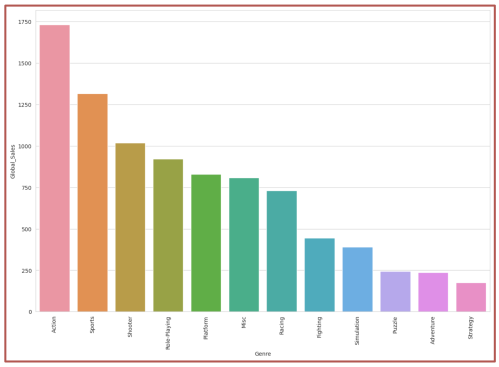

Figure 3.2: Genre Global Sales Trend

Figure 3.1: We can see that the sales around the year 2008 were the highest. This suggests that there may be some trend during those years that we are currently unaware of.

Figure 3.2: Overall, sales of games in the action genre were high. However, this information alone cannot predict if action genre games will sell well in the coming year. While it may not be a reliable indicator for future sales, it is still important information to consider.

Figure 3.3: Some genres may be more suitable for specific markets or regions. By analyzing sales data, it is possible to identify genres that perform poorly in certain places and consider removing them from the product portfolio. Conversely, it may be wise to invest in genres that are selling well in those markets to maximize profits.

Figure 3.3: Sales Genre Correlation Heat Map

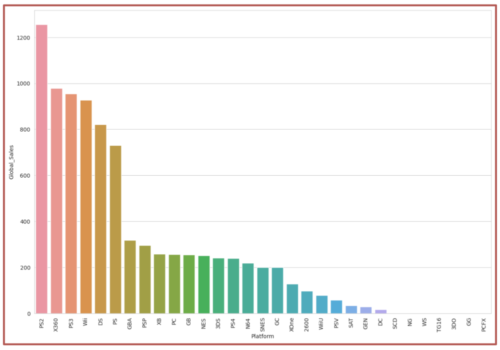

Figure 3.4: Platform Global Sales Trend

Figure 3.4: Considering all the years, investing in PS2 console games would be a wise idea since they have consistently performed well. This is different from game genre, as genre alone may not be a reliable indicator of sales. Developing games for a console would be a better option when considering all market conditions.

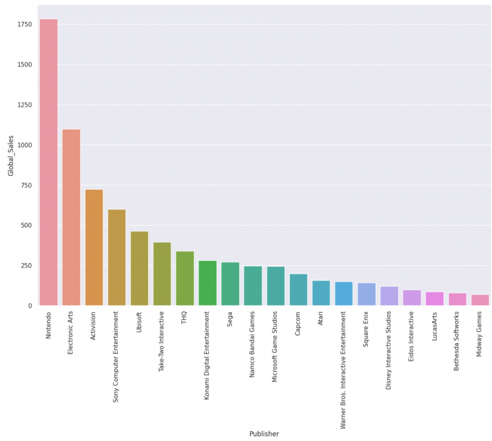

Figure 3.5: Publisher Global Sales Trend

Figure 3.6: Feature Correlation Map

Figure 3.5: Brand reputation is a critical factor when predicting sales since it has a lasting impact on consumer behavior. Established brands often have a loyal fan base, resulting in increased word-of-mouth marketing and stronger sales performance. Nintendo is a prime example of a brand with a strong reputation, as evidenced by their success in global sales.

Figure 3.6: It is very clear from the sales data that North America Sales and EU Sales have a significant impact on Global Sales, as the North American and European markets are among the largest in the video game industry

3.1.2 E-Commerce Sales Dataset

The study is based on a E-commerce dataset that includes transactions made by a retailer from the UK between December 1, 2010, and December 9, 2011. The business, which serves to both retail and wholesale consumers, focuses in one of a kind gifts for all occasions.

Understanding about consumer behaviour, sales trends, and e-commerce business trends are among the goals. The study aims to provide valuable findings and recommendations to improve business strategies in the online retail landscape through preprocessing, exploratory analysis, customer segmentation, sales and revenue analysis, geographic pattern exploration, looking into wholesalers, and using forecasting techniques.

The dataset features columns such as InvoiceNo, Stock Code, Description, Quantity, Invoice Date, Unit Price, CustomerID, and Country. Initial EDA involves addressing missing values, dropping the 'CustomerID' feature, imputing missing values in the 'description' column, and identifying patterns related to valid and non-valid items. Outliers are removed through scatter plots, box plots, and IQR techniques, resulting in a balanced dataset. Insights are gained regarding price range, customers' buying patterns, sales distribution per order, and holiday periods with zero sales.

Table 3.2: E Commerce Sales Data Feature Table

Feature	**Interpretation**
CustomerID	ID of the customer
InvoiceNo	Document Number of products
InvoiceDate	Date of the invoice generation
Quantity	Quantity of the product purchased
UnitPrice	Price per unit of the product
Country	Country of the customer
Stockcode	The stock code of product
Description	The details of the product

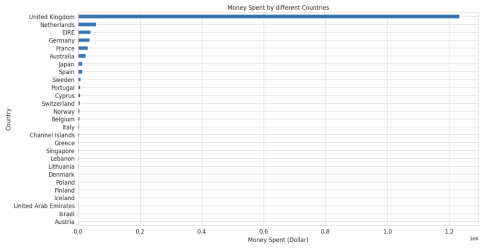

Figure 3.7: Money Spent vs Country

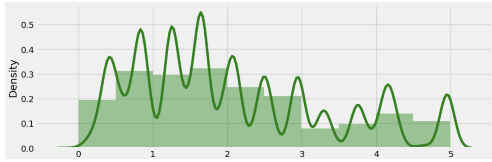

Figure 3.8: Density Plot

Figure 3.7: Since the company is situated in the UK, it makes sense that most of the products sold would be in UK. It is important to remove UK transactions from the information in order to study sales patterns of other nations.

Figure 3.8: The histograms reveal that the store primarily sells items priced between 0 and 3 pounds, stressing a focus on affordable products. This underscores the importance of providing to budget conscious customers.

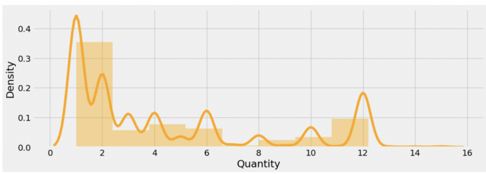

Figure 3.9: Quantity vs Density

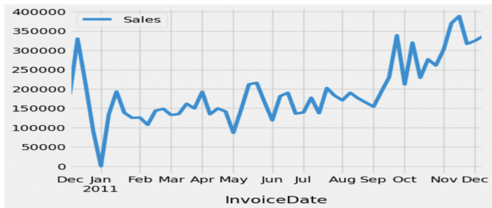

Figure 3.10: Invoice Date vs Sales

Figure 3.9: The histograms suggest that customers tend to purchase items in quantities of 1-5 or 10-12. This pattern indicates the potential presence of promotional offers or bundled sets, encouraging customers to buy multiple items together.

Figure 3.10: The week with zero sales in January coincides with the New Year holiday period, suggesting the store was closed during that time. The significant increase in sales from January to February indicates a positive trend in customer purchasing activity during that period.

3.1.3 Walmart Sales Dataset

The dataset includes columns such as 'Store', 'Dept', 'Date', 'Weekly Sales', and 'IsHoliday' to provide information about the store and department identifiers, sales records, date and time of sales, and whether the week includes a holiday. Other columns such as 'Type', 'Size', 'Temperature', 'Fuel Price', 'MarkDown1' to 'MarkDown5', 'CPI', 'Unemployment', and time components like 'Year', 'Month', 'Week', and 'DayOfTheMonth' offer additional details for analyzing sales patterns and factors influencing them.

The dataset is considered sequential high-frequency data as it contains a large volume of data collected at regular intervals, such as daily or hourly, enabling the study of patterns and trends over time. The content also outlines a data preprocessing pipeline for handling missing values and encoding categorical variables. Missing values in the 'CPI' and 'Unemployment' columns are filled with their respective means, while missing values in the 'MarkDown1' to 'MarkDown5' columns are filled with zeros. Dummy variables are created for the 'Type' column using one-hot encoding, and these encoded variables are appended to the original dataset. Finally, the 'Type' column is dropped from the dataset.

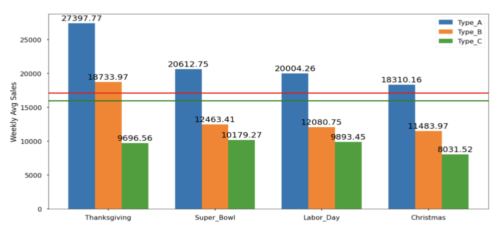

Figure 3.11: Weekly Average Sales vs Holidays

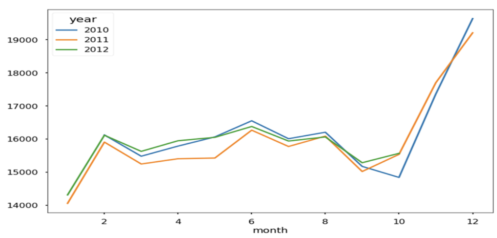

Figure 3.12: Monthly Sales vs Years

Figure 3.11: Based on the graph, it is evident that Thanksgiving week experiences the highest average sales compared to other holiday weeks. Additionally, during all holidays, Type A stores consistently achieve the highest sales average. The sales average during non-holiday periods is indicated by the green line on the graph, while the red line represents the average sales during holidays.

Figure 3.12: The graph indicates that, in general, the sales in 2011 were lower compared to 2010. However, when we consider the average sales, it is evident that 2010 had higher values. It is worth noting that the graph does not provide information about the sales in November and December for 2012, which are typically higher. Despite this missing data, the average sales for 2012 are still relatively close to those of 2010. It is highly likely that if we obtain the results for November and December in 2012 and include them, 2012 would take the first place in terms of sales.

3.1.4 BigMart Sales Dataset

The BigMart organization has gathered sales data for 1559 products across 10 stores in various cities. Their team of data scientists aim to develop a predictive model that can determine the sales of each product at specific stores. By utilizing this model, BigMart intends to gain insights into the characteristics of products and stores that have a significant impact on increasing sales.

The dataset used in this study comprises various attributes related to product and store information. Each product is identified by a unique ItemIdentifier. Additional details include Item Weight, indicating the weight of the product, and ItemFatContent, which specifies whether the product is classified as low fat or not. The attribute ItemVisibility represents the percentage of total display area in a store allocated to a specific product. The ItemType attribute categorizes the products into different categories. The maximum retail price of the product, known as ItemMRP, is also included and these are the factors influencing product performance.

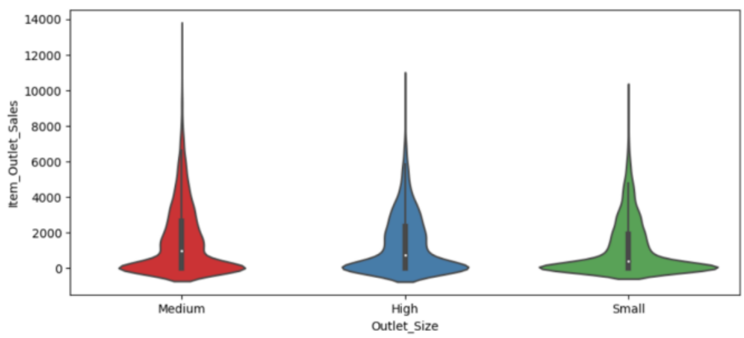

Figure 3.13: Item outlet sales vs Outlet size

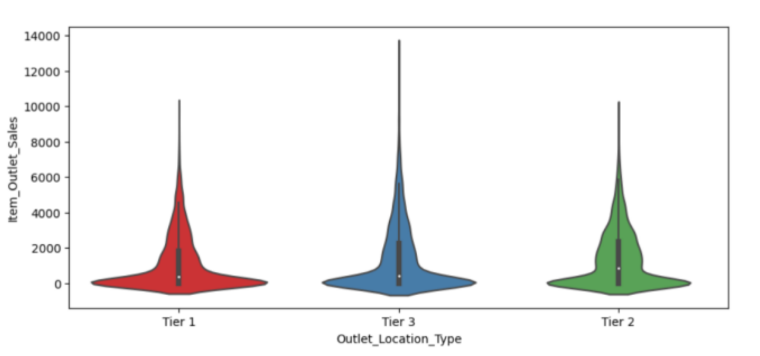

Figure 3.14: Item outlet sales vs Outlet location type

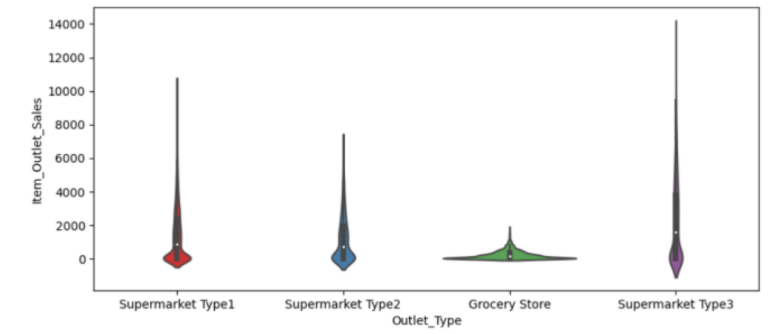

Figure 3.15: Item outlet sales vs Outlet type

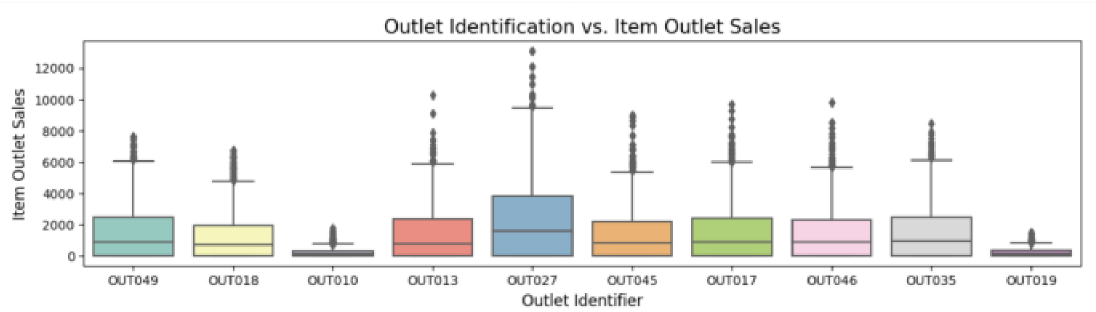

Figure 3.16: Outlet Identifier based on sales

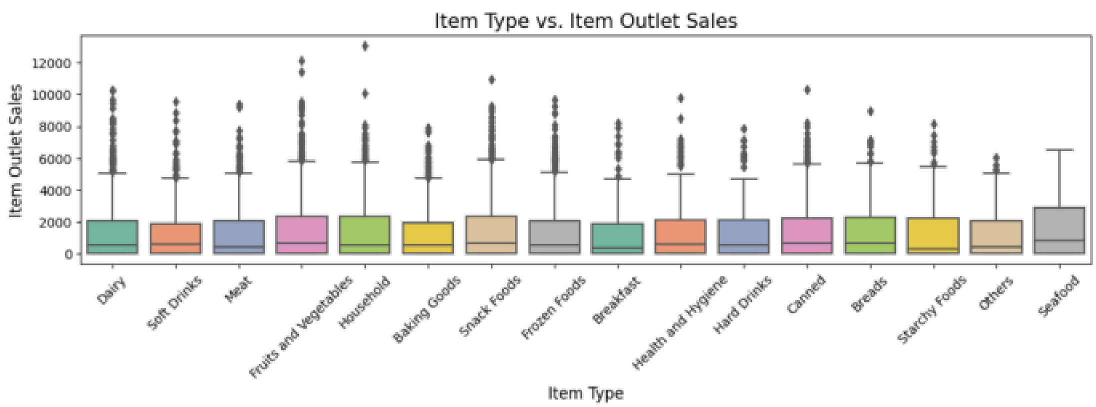

Figure 3.17: Item type based on sales

3.2 Overview of Machine Learning Models

To predict sales performance in our investigation, we used a wide variety of machine learning models. Our strategy uses a variety of models, including time series models, deep learning models, ensemble models, and volatility models, all of which were chosen for the task at hand.

We have used deep learning architectures such as LSTM , CNN , RNN , and GRU in addition to time series models. These models have shown good performance in identifying intricate connections and patterns within sequential data, which qualifies them for tasks requiring sales forecasting.

We have used ensemble models like Random Forest to further increase the forecasting potential of our models. To produce more precise and reliable projections, ensemble models aggregate the results of various independent models.

We compared how well each of these models performed across all of our datasets in our initial analysis. Based on the outcomes, we discovered that when applied to certain datasets, LSTM, CNN, RNN, and GRU consistently outperformed other models. In order to produce more precise and dependable projections, these models demonstrated improved capacity in capturing the complex dynamics and nonlinear interactions inside the sales data.

We will go into greater detail about the particular models we have selected to focus on in the following sections of this study. We will also discuss any new models that have been applied in light of the distinct traits and features of each dataset. We want to maximise the precision and efficiency of our sales forecasting system by customising the models we use to the particular information at hand.

Chapter 4

IMPLEMENTATION AND TESTING

4.1 Model Application and Evaluation

In this section, we focus on the application and evaluation of four main models: LSTM, CNN, RNN, and GRU. These models have been applied to four distinct datasets: the Video Game Sales Dataset, Big Mart Sales Dataset, E-commerce Sales Dataset, and Walmart Sales Dataset. Our objective is to assess and compare the performance of each model on these diverse datasets, providing valuable insights into their effectiveness for different industries.Each dataset represents a unique industry, encompassing specific features, characteristics, and behavioral patterns. The key drivers contributing to sales performance vary significantly across these industries. By analyzing the performance of different models on these datasets, we can gain a comprehensive understanding of their strengths and limitations in capturing the complexities of various industries' sales dynamics.

To facilitate the comparison, we present tables that highlight the performance of the same model across different datasets. However, it is important to note that the scales of measurement employed in these datasets may differ significantly. Some datasets might contain values in millions, while others may use normal numbers. Consequently, direct comparisons between these values may not be fair. To address this challenge, we employ a normalization technique to equalize the values, enabling a more meaningful and accurate comparison of the models' performance.Through this rigorous evaluation and comparison, we aim to provide a comprehensive analysis of how LSTM, CNN, RNN, and GRU models perform on distinct datasets from various industries. This knowledge will contribute to a deeper understanding of the applicability and effectiveness of these models in capturing the complexities of real-world sales scenarios.

4.1.1 Model 1: LSTM Performance and Results

Video Game Sales Dataset

When applying LSTM (Long Short-Term Memory) on the Video Game Sales Dataset, we aimed to explore the effectiveness of this model in analyzing low-frequency sequential data. The model is built by stacking multiple LSTM layers with 138 number of units and 0.2 dropout rate. It is compiled with the loss function, optimizer, and evaluation metrics. The function returns the constructed model, which is then trained using the provided training data, validation data, 100 epochs, 128 batch size, and an early stopping callback(es_patience = 10) The low RMSE and MAE values indicate that the

Error Metric	Value
RMSE	0.4417
MAE	0.1042

Table 4.1: Performance of LSTM Model on Video Game Sales Dataset

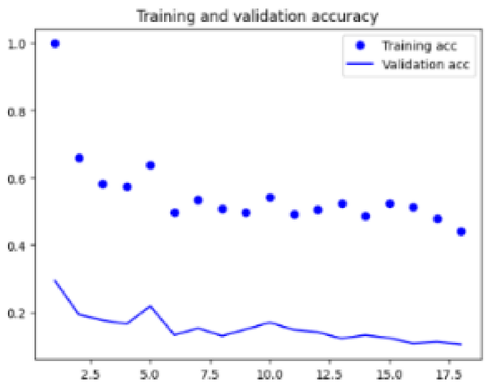

Figure 4.1: Video Game data LSTM Plot()

Figure 4.2: Video Game data Loss plot

LSTM model performs well in capturing the underlying patterns and dynamics of the dataset. It demonstrates its ability to effectively analyze the sequential nature of the data and generate accurate predictions for video game sales.

The plots revealed promising results. The dots, representing the accuracy and loss values, were consistently close to the lines, indicating a good fit to the dataset. This suggests that the LSTM model effectively captured the underlying patterns and relationships within the data.

BigMart Sales Dataset

The BigMart sales dataset, a non-sequential dataset, We aimed to determine whether LSTM can effectively model and predict sales in this type of dataset. During the implementation, we incorporated regularization, dropout, and early stopping techniques to improve the model's performance and prevent overfitting.

Our model achieved an RMSE of 1727.8 and an MAE of 1322.03, indicating the average prediction error. Notably, we observed instances where the validation RMSE

Error Metric	Value
RMSE	1727.8
MAE	1322.03

Table 4.2: Performance of LSTM Model on BigMart Sales Dataset

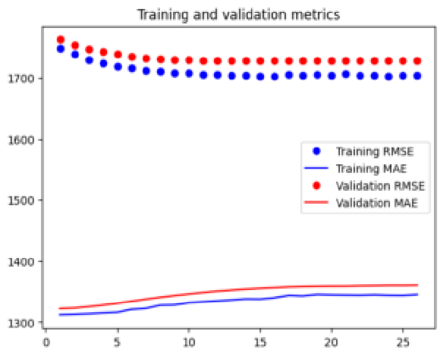

Figure 4.3: BigMart data training and validation LSTM plot

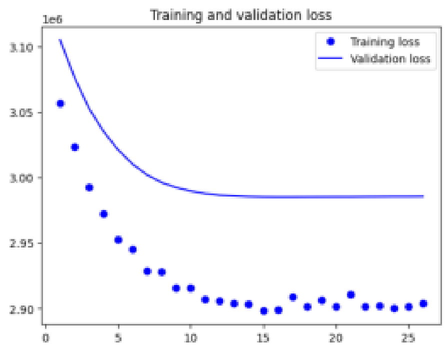

Figure 4.4: BigMart data Loss Plot

was higher than the training RMSE, and the validation loss was larger than the training loss. These findings suggest that our model was prone to overfitting, as it may have excessively focused on the training data and failed to generalize well to unseen data.
To address overfitting and improve the model's generalization, we employed dropout, a technique where randomly selected neurons are ignored during training, reducing the model's reliance on specific features or patterns. Additionally, we applied early stopping, which monitored the validation loss and halted training when the loss no longer improved significantly.

Walmart Sales Dataset

LSTM (Long Short-Term Memory) networks for training and testing. It addresses the issue of vanishing or exploding gradients in RNNs and highlights LSTM as a solution due to its ability to maintain long-term memory. The specific parameters used for the LSTM model are 80% training and 20% testing proportions, mean squared error loss, Adam optimizer, 20% dropout rate, batch size of 128, and 100 epochs with early stopping after 10 consecutive epochs without new learning. Various metrics were calculated for each epoch, including root mean squared error, mean absolute error, loss, validation loss, validation root mean squared error, and validation root mean absolute error. The lowest mean absolute error observed was 8992, and the lowest mean squared error was 16868.

Error Metric	**Value**
RMSE	16868.9
MAE	8982.6

Table 4.3: Performance of LSTM Model on Walmart Game Sales Dataset

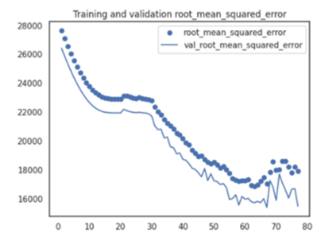

Figure 4.5: Walmart LSTM Training and Validation RMSE Plot

Figure 4.6: Walmart LSTM Training and Validation Loss Plot

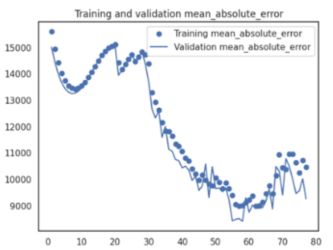

Figure 4.7: Walmart LSTM Training and Validation MAE Error Plot

The training with validation RMSE, validation loss, and validation MAE graphs exhibit a strong similarity, indicating that the model is a good fit.

E-Commerce Dataset

The LSTM model for analyzing e-commerce datasets utilizes the Adam optimizer with a mean squared error loss function. It incorporates a dropout rate of 0.3 to mitigate overfitting. The model consists of LSTM layers with 64 units each. Training is performed over 152 epochs with a batch size of 32. Early stopping is implemented with a patience of 15 epochs to prevent overtraining.

Thus,the best (lowest) validation MAE and RMSE values are: The lowest validation RMSE: 51.6373. The lowest validation MAE: 8.3668. The model shows progressive

Error Metric	Value
RMSE	51.6373
MAE	8.3668

Table 4.4: Performance of LSTM Model on E-Commerce Dataset

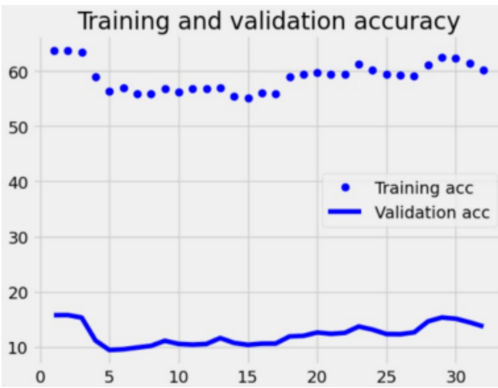

Figure 4.8: Ecommerce LSTM Training and Validation Accuracy Plot

Figure 4.9: Ecommerce LSTM Training and Validation Loss Plot

learning with decreasing loss metrics for both training and validation sets. It demonstrates improved performance on the validation set at epoch 4, balancing memorization and generalization. The model handles problem complexity well, maintaining stability in validation metrics while reducing training loss.

4.1.2 Model 2: CNN Performance and Results

Video Game Sales Dataset

CNNs are commonly associated with image processing tasks due to their ability to capture spatial features. However, we investigated their suitability for sequential data like the Video Game Sales Dataset, which poses unique challenges.The model is built by stacking several layers, including “2” convolutional layers, each with 32 filters and a ReLU activation function, followed by “2” LSTM layers(each 32 units) and a dense layer(1 unit) with linear activation. The model is compiled with the specified optimizer (Adam with a learning rate of 0.003), and evaluation metrics along with using 100 epochs, a batch size of 128, and an early stopping callback. These values of metrics

Error Metric	Value
RMSE	1.7281
MAE	0.6755

Table 4.5: Performance of CNN Model on Video Game Sales Dataset

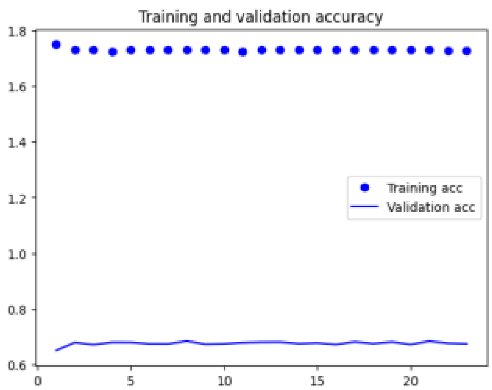

Figure 4.10: VideoGames CNN Training Plot

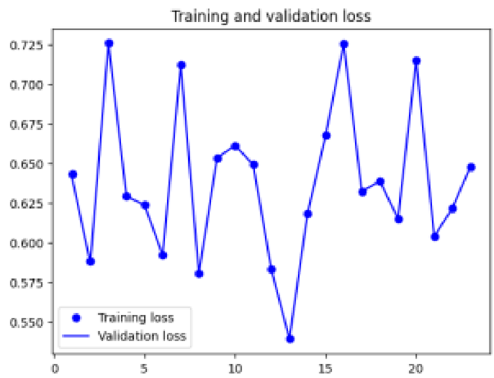

Figure 4.11: VideoGames CNN Loss Plot

suggest that the model did not provide a good fit for the Video Game Sales Dataset. The relatively poor performance of the CNN model on the Video Game Sales Dataset highlights the challenges of applying CNNs to sequential or time-series data. These findings imply that CNNs may not be well-suited for analyzing low-frequency sequential data, such as the Video Game Sales Dataset. Overfitting observed in plots suggests that the model might have learned the noise or specific patterns within the training data but failed to generalize to new, unseen instances.

BigMart Sales Dataset

The BigMart sales dataset was further analyzed using the 1D-CNN (Convolutional Neural Network) model. We aimed to explore the effectiveness of this architecture in predicting sales, considering its ability to capture spatial patterns within the data. The evaluation of the 1D-CNN model resulted in an RMSE of 1088.9 and an MAE of 749.8.

Error Metric	Value
RMSE	1088.9
MAE	749.8

Table 4.6: Performance of CNN Model on Bigmart Sales Dataset

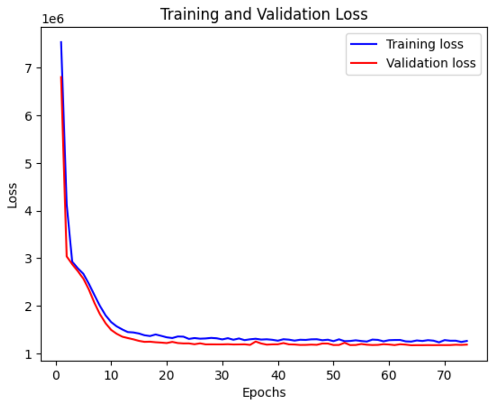

Figure 4.12: BigMart CNN Training,Validation loss

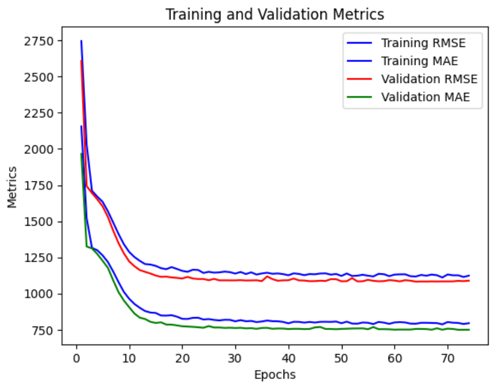

Figure 4.13: BigMart CNN Training,Validation Plot

In addition to the error metrics, we examined the training and validation curves to gain insights into the model's performance. Notably, the observed trends in the plots were encouraging. The validation loss was significantly lower than the training loss, indicating that the model was able to generalize well to unseen data.

These results indicate that the 1D-CNN model effectively captured the underlying patterns and relationships within the data, leading to accurate sales predictions. The comparatively low RMSE and MAE values signify that the model's predictions were close to the actual sales values.

Walmart Sales Dataset

A Convolutional Neural Network (CNN) is utilized in this research. CNN is a kind of deep neural network which consists of one or more convolutional layers, and it is mainly used for image processing and text classification. However, because of its ability for identifying patterns, it can be utilized in forecasting as well. The results from the previous CNN are fed into the next CNN layer. There is the max-pooling layer that takes the maximum number in the sliding window, and prevents the model from overfitting . Between this layer and the dense layer, a flatten layer is used. Fig. 5 shows a typical CNN architecture using a 1D convolutional layer . In this study, a 1D CNN model is proposed which consists of 32 filters with kernel size 3. Similar parameters have been taken for cnn as for that of LSTM the observed lowest root mean squared error and lowest mean absolute error observed are 22912 and 13421.

Error Metric	Value
RMSE	22912.6
MAE	15152.5

Table 4.7: Performance of CNN Model on Walmart Game Sales Dataset

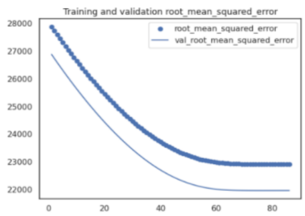

Figure 4.14: Walmart CNN RMSE Plot

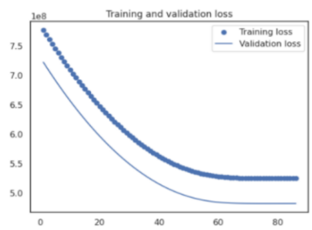

Figure 4.15: Walmart CNN Loss Plot

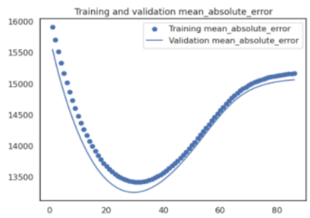

Figure 4.16: Walmart CNN MAE Plot

The training with validation RMSE, validation loss, and validation MAE graphs exhibit a strong similarity, indicating that the model is a good fit.

E-Commerce Dataset

CNNs are a type of deep neural network that have one or more convolutional layers and are mostly used for text classification and picture processing. But it can also be used in predicting because of its capacity for pattern recognition. The following CNN layer receives the output from the previous CNN layer. The sliding window's maximum number is used by the max-pooling layer, which keeps the model from overfitting. A flattened layer is employed in between this layer and the dense layer. A basic CNN architecture utilizing a 1D convolutional layer is shown in Fig. 5. A proposed 1D CNN model in this work consists of 32 filters with kernel size 3.

The lowest measured root mean squared error and lowest mean absolute error observed for cnn are 22912 and 13421, respectively. Similar parameters were used for LSTM.

Error Metric	Value
RMSE	63.7080
MAE	15.7860

Table 4.8: Performance of CONV 1D on E-Commerce Dataset

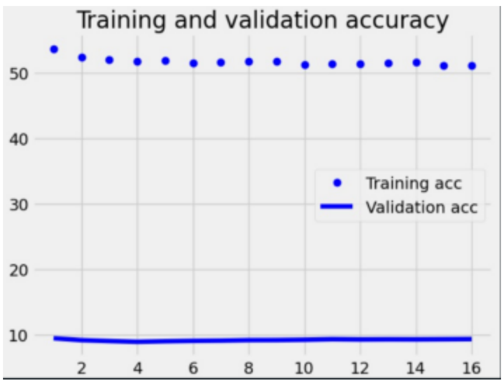

Figure 4.17: Ecommerce CNN Training and Validation Accuracy Plot

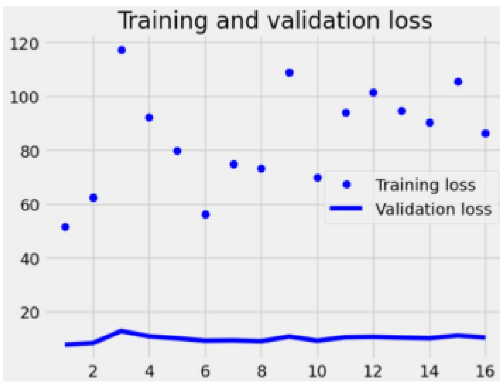

Figure 4.18: Ecommerce CNN Training and Validation Loss Plot

The training loss decreases over time, which means the model is learning and improving its performance on the training dataset.

4.1.3 Model 3: RNN Performance and Results

Video Game Sales Dataset

We explored the application of the RNN (Recurrent Neural Network) model to predict video game sales. RNNs are designed to capture temporal dependencies in sequential data, making them potentially suitable for analyzing the sequential nature of the dataset. The model is built with 138 units in each SimpleRNN layer, a dropout rate of 0.2, and a linear activation function in the final dense layer. The model is compiled with a mean squared error loss, Adam optimizer with a learning rate of 0.003, and uses root mean squared error and mean absolute error as evaluation metrics. The training is performed with 100 epochs, a batch size of 128, and early stopping patience of 10. The differences

Error Metric	Value
RMSE	0.7696
MAE	0.4988

Table 4.9: Performance of RNN Model on Video Game Sales Dataset

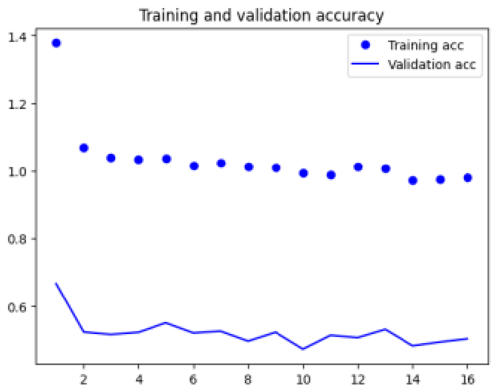

Figure 4.19: VideoGame RNN Accuracy Plot

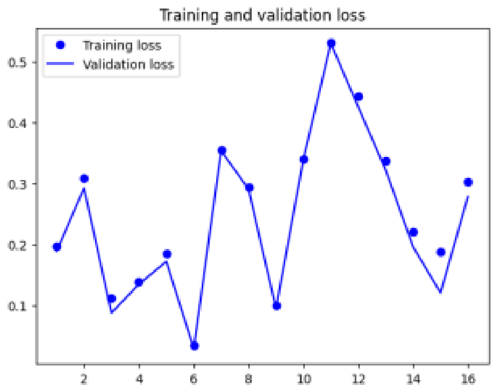

Figure 4.20: VideoGame RNN Loss Plot

between the predicted and actual sales values were comparatively smaller, indicating a better alignment with the underlying patterns and dynamics present in the dataset. The plots demonstrated the accuracy of the model's predictions during training and validation. A closer alignment between the dots (accuracy values) and the lines indicates a better fit and generalization of the model. This indicates that the RNN model effectively captured the temporal dependencies and patterns present in the sequential nature of the Video Game Sales Dataset.

BigMart Sales Dataset

We also applied the RNN (Recurrent Neural Network) model for sales prediction on the BigMart sales dataset. The model yielded an RMSE (Root Mean Squared Error) of 2575.8 and an MAE (Mean Absolute Error) of 1925.7. These metrics provide insights into the average prediction error of the RNN model compared to the actual sales values in the BigMart dataset.

Error Metric	Value
RMSE	2575.8
MAE	1925.7

Table 4.10: Performance of RNN Model on BigMart Sales Dataset

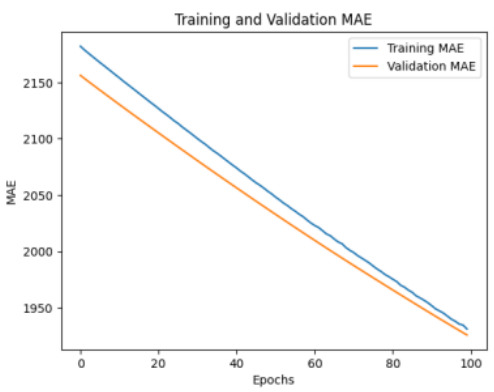

Figure 4.21: BigMart RNN Training and Validation MAE Plot

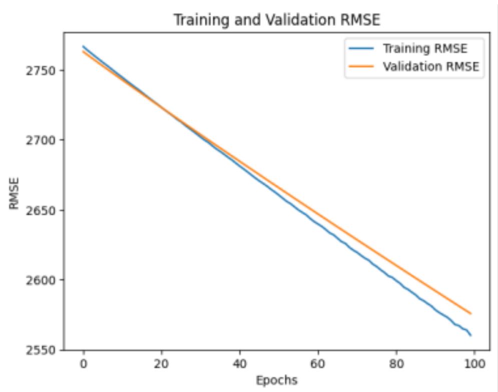

Figure 4.22: BigMart RNN Training and Validation RMSE Plot

Upon analyzing the training and validation plot specifically for the BigMart sales dataset, we made some key observations. The validation RMSE was found to be higher than the training RMSE, indicating a potential overfitting issue. Additionally, the validation loss was greater than the training loss, further reinforcing the notion of overfitting in the model when applied to the BigMart sales dataset.

Walmart Sales Dataset

Recurrent Neural Network (RNN) model using the SimpleRNN layer in Keras. The model architecture consists of stacked SimpleRNN layers with dropout layers to prevent overfitting. The final layer is a dense layer with a linear activation function. The model is compiled with mean squared error loss, Adam optimizer, and metrics such as RMSE and MAE. A dictionary of parameters is defined, including loss function, optimizer, dropout rate, LSTM units, number of epochs, batch size, and early stopping patience. The get_model function uses these parameters to construct the RNN model and return it. The model is trained using the fit method, providing training and validation data, parameters, and callbacks. The EarlyStopping callback monitors validation RMSE and stops training if no improvement is observed within a specified number of epochs. After training, the model's results are obtained, with RMSE and MAE values of 17779.1 and 9394.06, respectively.

Error Metric	Value
RMSE	17779.1
MAE	9394.06

Table 4.11: Performance of RNN Model on Walmart Game Sales Dataset

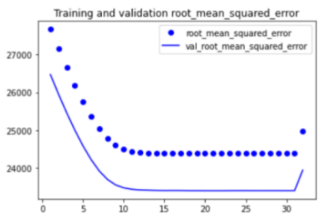

Figure 4.23: Walmart RNN RMSE Metric Plot

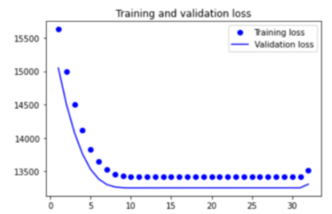

Figure 4.24: Training and Validation Loss Plot

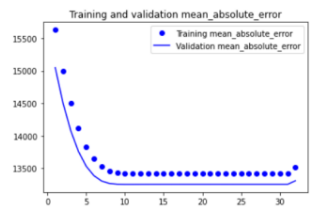

Figure 4.25: Walmart dataset RNN MAE Plot

The difference between the root mean square error (RMSE) values of the training and validation sets is significant, indicating a considerable gap between their performance. However, when we compare the training loss and validation mean absolute error (MAE) graphs, we observe some similarity, suggesting a relatively good fit between the two.

E-Commerce Dataset

Recurrent Neural Networks (RNNs) are particularly suited for analyzing e-commerce datasets due to their capacity to handle sequential data and capture temporal dependencies. This makes them ideal for modeling customer behaviors and sales patterns that inherently possess sequential and time-dependent characteristics. Its flexibility in managing varying input dimensions allows for a comprehensive analysis of different features, such as customer details, product attributes, and transaction history. In summary, the use of RNNs can provide businesses with valuable insights into customer preferences and trends, and contribute to the optimization of marketing and sales strategies in the e-commerce sector. Based on the results in the last epoch, the root mean squared error (RMSE) on the validation set is approximately 62.2050, and the mean absolute error (MAE) is about 17.0190.

Error Metric	Value
RMSE	62.2050
MAE	17.0190

Table 4.12: Performance of RNN Model on E-Commerce Dataset

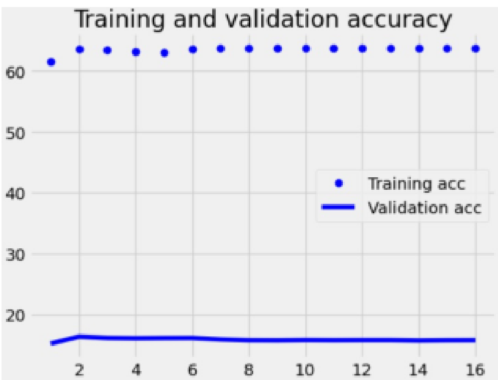

Figure 4.26: Ecommerce RNN Training and Validation Plot

Figure 4.27: Ecommerce RNN Loss Plot

The plot seems like the model's performance is fairly consistent throughout training. The RMSE and MAE values do not fluctuate drastically from epoch to epoch. This stability could potentially be a good sign, indicating that the model is not overly sensitive to the specific subset of the data used for training.

4.1.4 Model 4: GRU Performance and Results

Video Game Sales Dataset

We applied GRU (Gated Recurrent Unit) model to forecast video game sales. GRU models, a variant of the popular Recurrent Neural Networks (RNN), possess the ability to capture and retain crucial information over extended sequences, making them an appealing choice for dissecting sequential data. The model has 138 units in each GRU layer, a dropout rate of 0.2, and a linear activation function in the final dense layer. It is compiled with a mean squared error loss, Adam optimizer with a learning rate of 0.003, and uses root mean squared error and mean absolute error as evaluation metrics. The training is performed with 100 epochs, a batch size of 128, and early stopping patience of 10. These relatively low values signify a good enough level of accuracy

Error Metric	Value
RMSE	0.5095
MAE	0.1733

Table 4.13: Performance of GRU Model on Video Game Sales Dataset

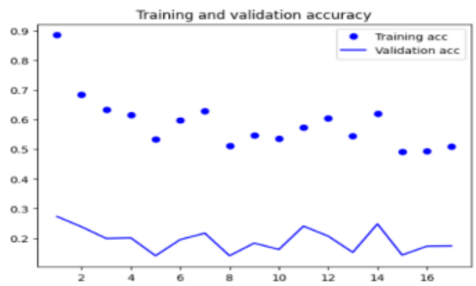

Figure 4.28: VideoGames GRU Accuracy Plot

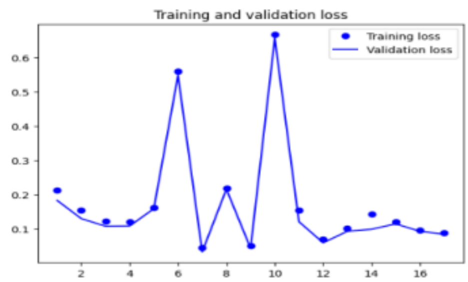

Figure 4.29: VideoGames GRU Loss Plot

when compared against the dataset. Consequently, we can assert that the RNN model aptly captured the underlying patterns and dynamics of the Video Game Sales Dataset. The "Training and Validation Accuracy" plot showcased a good alignment between the dots (representing accuracy values) and the lines, thereby affirming the RNN model's successful fit to the dataset. This alignment is indicative of the model's proficiency in capturing the sequential dependencies and patterns intrinsic to the Video Game Sales Dataset. These findings affirm the suitability of RNNs for analyzing low-frequency sequential data, such as video game sales.

BigMart Sales Dataset

We also applied the GRU model with dropout and early stopping for sales prediction on the BigMart sales dataset. The model yielded an RMSE of 1768.7 and an MAE of 1321.9 . These metrics provide insights into the average prediction error of the GRU model, considering the dropout and early stopping techniques, compared to the actual sales values in the BigMart dataset.

Error Metric	Value
RMSE	1768.7
MAE	1321.9

Table 4.14: Performance of GRU Model on BigMart Sales Dataset

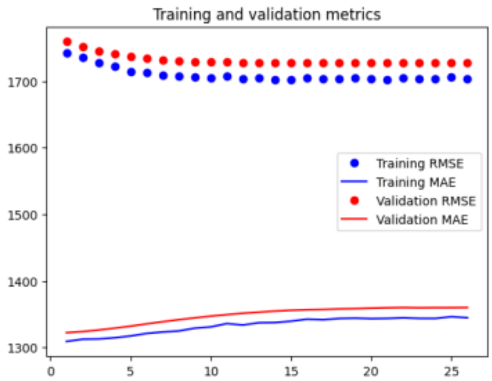

Figure 4.30: BigMart GRU Training and Validation Metrics Plot

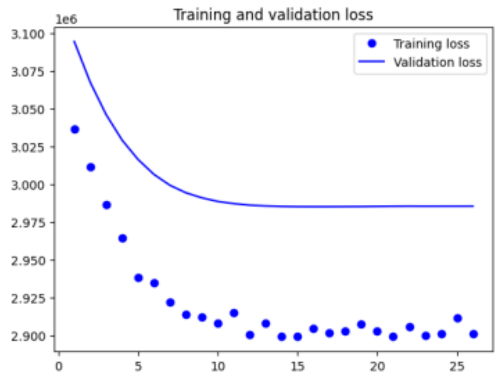

Figure 4.31: BigMart GRU Training and Validation Loss Plot

Analyzing the training and validation plot specifically for the BigMart sales dataset, we made some key observations. Notably, the validation RMSE remained higher than the training RMSE, suggesting the possibility of overfitting even with the application of dropout and early stopping. Additionally, the validation loss was greater than the training loss, further indicating potential overfitting challenges when applying the GRU model with dropout and early stopping to the BigMart sales dataset.

Walmart Sales Dataset

A GRU (Gated Recurrent Unit) model was implemented on Walmart sales data. The model architecture consists of multiple GRU layers with dropout regularization applied after each layer to prevent overfitting. The last GRU layer is followed by a dense layer with a linear activation function, which generates the predicted output.The model is compiled using the mean squared error (MSE) loss function and the Adam optimizer with a learning rate of 0.003. Two evaluation metrics, root mean squared error (RMSE) and mean absolute error (MAE), are used to assess the model's performance. The training progresses for a specified number of epochs with a given batch size. Progress information is displayed during training. To prevent overfitting and monitor the validation RMSE, an early stopping callback (es_callback) is included. The training process will stop if there is no improvement in the validation RMSE for a certain number of epochs.

Error Metric	Value
RMSE	18297.8
MAE	10821.6

Table 4.15: Performance of GRU Model on Walmart Game Sales Dataset

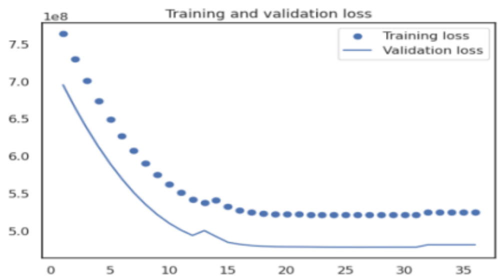

Figure 4.32: Walmart GRU Training and Validation Loss Plot

After training, the reported results include the RMSE and MAE values obtained on the test set, which are 18297.8 and 10821.6, respectively. These metrics indicate the performance of the model in terms of the average prediction error and the root mean squared error. The training and validation loss graphs, as well as the validation RMSE (Root Mean Square Error) graphs of the GRU model, show some dissimilarity. This indicates that the model may not be a good fit for the data since the validation results deviate from the training results.

E-Commerce Dataset

Due to its capacity to avoid the vanishing gradient problem and successfully capture long-term dependencies, the Gated Recurrent Unit (GRU) model is well-suited for analyzing sequential data, including e-commerce datasets. GRUs outperform LSTM models in terms of training and prediction times, making them ideal for large e-commerce datasets. They can produce good results even with limited training data, which is common in e-commerce. Because of their ease of use and interpretability, GRUs are ideal for studying client behavior and predicting. The 4th epoch yielded the best validation loss of 2728.27, an RMSE of 52.23, and an MAE of 9.07. By the 19th epoch, however, the validation loss had climbed to 3147.59, the RMSE had increased to 56.10, and the MAE had increased to 12.53.

Error Metric	Value
RMSE	52.2328
MAE	8.8274

Table 4.16: Performance of GRU Model on E-Commerce Dataset

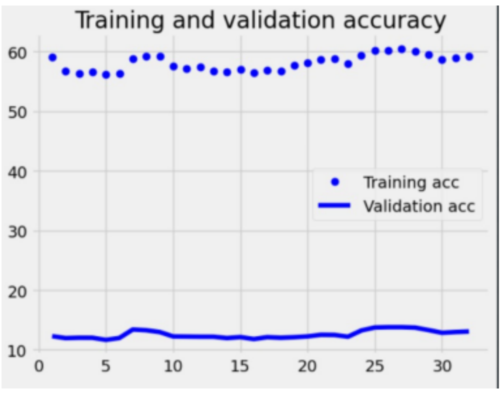

Figure 4.33: Ecommerce data GRU Accuracy Plot

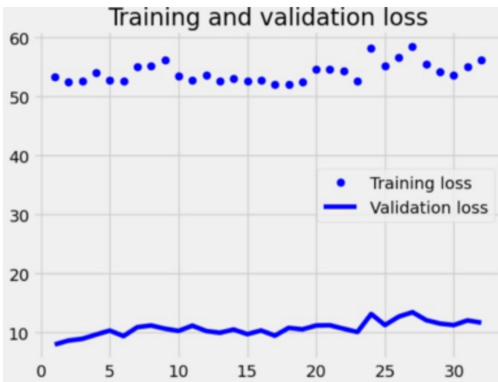

Figure 4.34: Ecommerce data GRU Loss Plot

From the data provided, it's clear that the model's performance during the initial training epochs was positive. There is a steady decline in the root mean squared error (RMSE) and mean absolute error (MAE) metrics from the first to the fourth epoch, indicating that the model was learning and improving its predictions during these early stages.

4.2 Model 5: Hybrid Models Performance and Results

4.2.1 CNN-RNN

Video Game Sales Dataset

We recognized the potential benefits of combining the strengths of both CNN (Convolutional Neural Network) and RNN (Recurrent Neural Network) models to create a hybrid approach. This hybrid model aimed to leverage the advantages of each architecture to achieve enhanced performance in capturing the intricate patterns and dependencies within the dataset. The model starts with a CNN layer with 64 filters, a kernel size of 2, and ReLU activation. A max-pooling layer is added with a pool size of 2(reduces spatial features). A dropout layer with a dropout rate of 0.2 is then included. Next, a SimpleRNN layer is added with 50 units. Finally, a dense layer with 1 unit and a linear activation function is included to produce the final output. When solely employ-

Error Metric	Value
RMSE	0.6439
MAE	0.1385

Table 4.17: Performance of CNN-RNN Hybrid Model on Video Game Sales Dataset

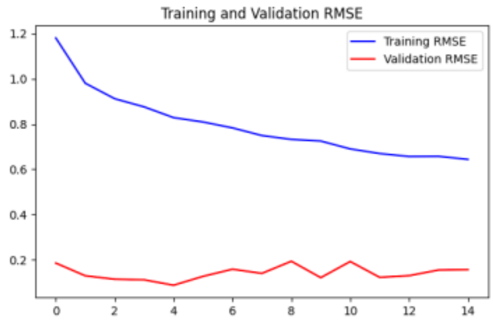

Figure 4.35: VideoGames CNN-RNN RMSE Plot

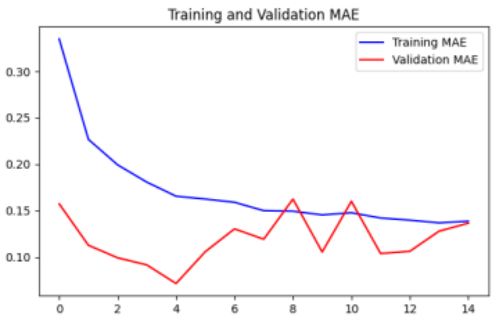

Figure 4.36: VideoGames CNN-RNN MAE Plot

ing the CNN architecture, we observed higher error metrics, suggesting limitations in accurately predicting video game sales. The same with RNN alone too. However, by integrating the CNN and RNN architectures in a hybrid model, we achieved a remarkable improvement in performance.

4.2.2 LSTM-FCN Model

E-Commerce Dataset

The LSTM component of the model is ideal for sequential data, capturing temporal dependencies and retaining memory. The FCN component excels at capturing spatial patterns and features through convolutional layers. By combining both LSTM and FCN, the model can leverage the strengths of each architecture to achieve a comprehensive understanding of the ecommerce dataset. This hybrid approach enables the model to capture both the temporal dynamics and spatial patterns in the data, making it suitable for tasks such as sales forecasting, anomaly detection, and customer segmentation in the ecommerce domain. By the 17th epoch, the model achieved a training loss of 2695.54, a root mean squared error (RMSE) of 51.92, and a mean absolute error (MAE) of 8.76. On the validation set, the loss was 4360.34, the RMSE was 66.03, and the MAE was 8.77.

Error metric	**Value**
RMSE	49.43
MAE	7.87

Table 4.18: Performance of LSTM-FCN Model on E-Commerce Dataset

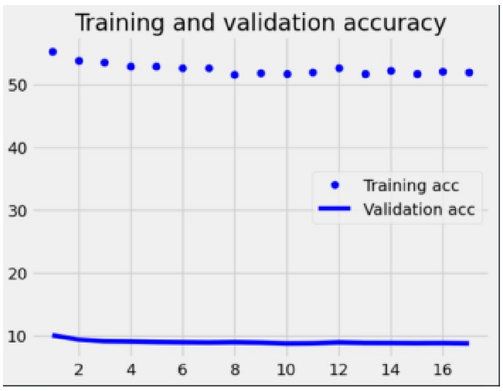

Figure 4.37: Ecommerce LSTM-FCN Accuracy Plot

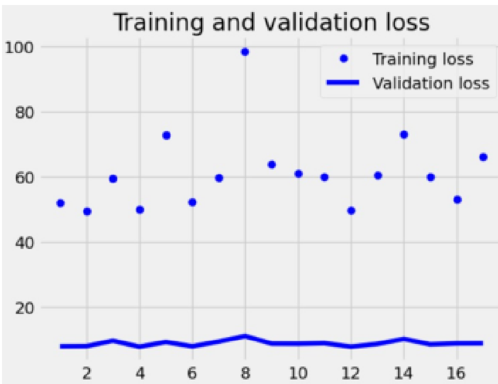

Figure 4.38: Ecommerce LSTM-FCN Loss Plot

The model is learning, as the loss is decreasing on the training set. There are epochs (like epoch 2 and 4) where both the training and validation losses decrease. This means that the model is capable of generalizing well to unseen data in these instances.

4.2.3 1D-CNN-LSTM Model

BigMart Sales Dataset

We also applied the 1D-CNN-LSTM (1-Dimensional Convolutional Neural Network with Long Short-Term Memory) model for sales prediction on the BigMart sales dataset. This model combines convolutional and recurrent layers to capture both spatial and temporal patterns in the data. The model yielded an RMSE (Root Mean Squared Error) of 1143.9 and an MAE (Mean Absolute Error) of 816.8, showcasing the average prediction error of the 1D-CNN-LSTM model compared to the actual sales values in the BigMart dataset.

Error metric	**Value**
RMSE	49.43
MAE	7.87

Table 4.19: Performance of 1D-CNN-LSTM Model on BigMart Sales Dataset

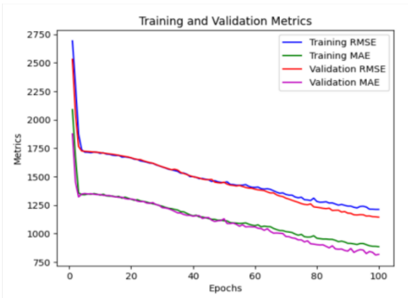

Figure 4.39: BigMart CNN-LSTM Training and Validation Metrics Plot

Analyzing the training and validation plot specifically for the BigMart sales dataset, we made some intriguing observations. Notably, the validation RMSE was lower than the training RMSE, indicating that the model performed even better on the validation set. Similarly, the validation MAE was lower than the training MAE, suggesting that the model's predictions on the validation set were closer to the actual sales values compared to the training set. These results indicate that the 1D-CNN-LSTM model exhibited excellent generalization capabilities and effectively captured the underlying patterns in the BigMart dataset.

4.2.4 1D-CNN-FNN

BigMart Sales Dataset

A hybrid model that combined 1D-CNN and FNN (Feedforward Neural Network) models for sales prediction on the BigMart sales dataset. This model leverages the strengths of both architectures, using the 1D-CNN to capture spatial patterns and the FNN to process the extracted features. Notably, the combined model yielded promising results, with the validation RMSE and MAE being smaller than the training RMSE and MAE, respectively. These results highlight the model's ability to generalize well and make accurate sales predictions on unseen data.

Error metric	Value
RMSE	49.43
MAE	7.87

Table 4.20: Performance of 1D-CNN-FNN Model on BigMart Sales Dataset

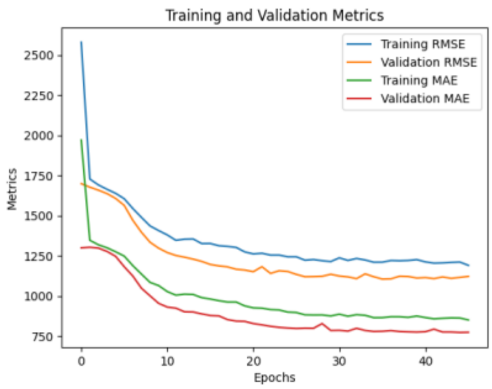

Figure 4.40: BigMart CNN-FNN Training and Validation Metrics Plot

Analyzing the training and validation plots specifically for the BigMart sales dataset, we made exciting observations. The validation RMSE and MAE consistently remained lower than their training counterparts, demonstrating the model's capacity to generalize effectively.

4.3 Ensemble Models Performance and Results

4.3.1 Ensemble Models of CNN, RNN and LSTM

Video Game Sales Dataset

We explored the potential of ensemble models that combine the strengths of CNN , RNN, and LSTM architectures. The ensemble prediction is obtained by averaging or using weighted averages of the individual model predictions. These remarkably low values indicate a high level of accuracy in predicting video game sales when compared to the average values within the dataset.

Error metric	Value
RMSE	49.43
MAE	7.87

Table 4.21: Performance of Ensemble Models of CNN, RNN and LSTM on Video Game Sales Dataset

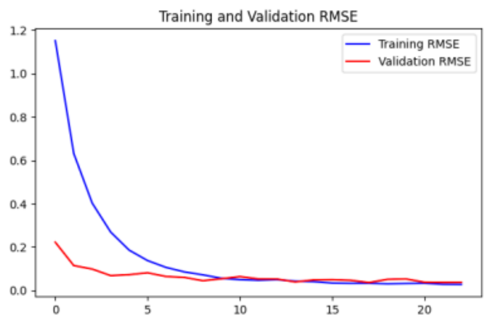

Figure 4.41: Training and Validation RMSE Plot

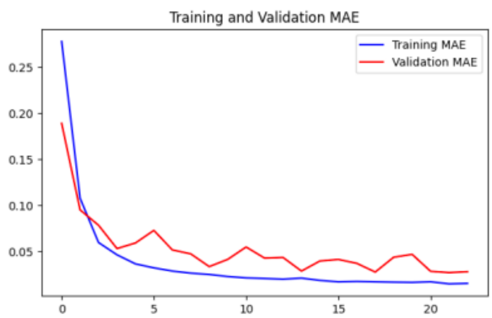

Figure 4.42: Training and Validation MAE Plot

The above plots revealed an outstanding alignment between the accuracy line and the validation line. These characteristics signify a robust fit and effective generalization of the ensemble models to both the training and validation data. The decision to adopt ensemble models that combine CNN, RNN, and LSTM architectures arises from their distinct advantages. CNNs excel in capturing spatial features, making them well-suited for extracting patterns from individual data points, such as genre, platform, and publisher. RNNs and LSTM models, on the other hand,are adept at capturing temporal dependencies and capturing sequential patterns, ensuring the integration of time-related information, such as the order of game releases.

4.3.2 LSTM with Attention Mechanism

Video Game Sales Dataset

In our exploration of models for predicting video game sales within the Video Game Sales Dataset, we investigated the efficacy of LSTM (Long Short-Term Memory) with an Attention Mechanism. The addition of an Attention Mechanism to the LSTM architecture allows the model to focus its attention on specific parts of the data, aiding in capturing relevant and important information. This attention mechanism helps overcome the challenges faced by basic LSTM models.

Error metric	Value
RMSE	49.43
MAE	7.87

Table 4.22: Performance of Ensemble Models of CNN, RNN and LSTM on Video Game Sales Dataset

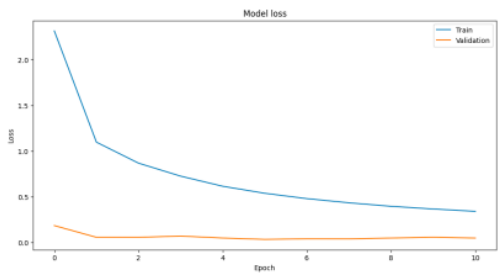

Figure 4.43: VideoGames LSTM attention Loss Plot

Upon examining the plot, it becomes evident that the LSTM model with Attention Mechanism displays favorable results. The loss curve demonstrates a consistent downward trend, indicating a gradual reduction in the discrepancy between predicted and actual values as the model undergoes training iterations. This signifies that the model is effectively learning from the data and improving its predictive capabilities. The LSTM model with an Attention Mechanism demonstrates promising performance on the Video Game Sales Dataset, as evidenced by the low loss value obtained. These findings highlight the significance of incorporating attention mechanisms in LSTM models when dealing with sequential data, such as the Video Game Sales Dataset.

4.3.3 Variational Autoencoder (VAE)

Walmart Game Sales Dataset

The VAE model architecture was defined with an input dimension of the scaled data and a latent dimension of 2. The encoder part of the model consisted of a dense layer with a ReLU activation function, which was then connected to two separate dense layers representing the mean and logarithm of the variance of the latent space. A sampling function was defined to generate latent space samples based on the mean and variance. The decoder part of the model included a dense layer with a ReLU activation function, followed by a dense layer outputting the reconstructed data. The VAE model was compiled with the Adam optimizer. The loss function was defined as a combination of the mean squared error (MSE) for reconstruction loss and the Kullback-Leibler (KL) divergence loss. The VAE model was trained with early stopping based on the validation loss.

Error metric	Value
RMSE	49.43
MAE	7.87

Table 4.23: Performance of CNN Model on Walmart Game Sales Dataset

Figure 4.44: Walmart VAE Training and Validation Loss Plot

The VAE model achieved a training loss of 1.0001 and a validation loss of 1.2667 and RMSE pf 1.9784 and MAE of 1.4547. These metrics indicate the reconstruction accuracy of the VAE model. Lower values of the loss indicate better performance and closer resemblance between the original and reconstructed data.

Chapter 5

RESULTS AND DISCUSSION

5.1 Comparison of Model Performance

Table 5.1: RMSE - Part 1

	A	B	C	D	E
1	0.4417	1.7281	0.7696	0.5095	0.6439
2	1727.8	1088.9	2575.8	1768.7	-
3	51.6373	63.7080	62.2050	52.2328	-
4	16868.9	22912.6	17779.1	18297.8	-

Table 5.2: RMSE - Part 2

	F	G	H	I	J	K
1	0.0279	0.2528	-	-	-	-
2	-	-	1143.9	1121.5	-	-
3	-	-	-	-	49.43	-
4	-	-	-	-	-	1.9784

Table 5.3: MAE - Part 1

	A	B	C	D
1	0.1042	0.6755	0.4988	0.1733
2	1322.03	749.8	1925.7	1321.9
3	8.3668	15.7860	17.0190	8.8273
4	8982.6	15152.5	9394.06	10821.6

- Column descriptions:
  - A: LSTM.
  - B: CNN.
  - C: RNN.
  - D: GRU.
  - E: CNN-RNN.
  - F: Ensemble of CNN,RNN,LSTM.
  - G: LSTM with attention mechanism.

  - H: CNN-LSTM.
  - I: CNN-FNN.
  - J: LSTM-FCN.
  - K: VAE.

- Row descriptions:
  - 1: Videogames Dataset.
  - 2: BigMart Dataset.
  - 3: E-commerce Dataset.
  - 4: Walmart Sales Dataset.

Table 5.4: MAE - Part 2

	E	F	G	H	I	J	K
1	0.1385	0.0279	0.2276	-	-	-	-
2	-	-	-	816.8	773.6	-	-
3	-	-	-	-	-	49.43	-
4	-	-	-	-	-	-	1.4547

5.2 Customized Approach for Different Datasets

In our study, we found that the four datasets we looked at each have unique qualities and fit into different categories. The Big Mart Sales Data was labelled as "Non-Sequential Data," the Video Game Sales Dataset as "Sequential Low Frequency Data," and the Walmart Sales Data and Ecommerce Sales Data as "Sequential High Frequency Data." These variations compelled us to take a unique strategy, adjusting our modelling methods to each dataset's unique properties.We examined different models that are particularly suited to the peculiarities of the data and industry as we dug deeper into each dataset to hone our method. We identified several models that performed better when used with particular datasets by assessing and testing these specialised models.

For instance, we discovered that the sequential low frequency data from the Video Game Sales Dataset performed remarkably well with the LSTM with Attention Mechanism model. The Bigmart Sales Dataset, Ecommerce Sales Dataset, and Walmart Sales Dataset, which are classified as sequential and non-sequential high frequency data, respectively, also showed greater performance on these models.

In order to address the specific characteristics of various datasets and sectors, it has been found to be efficient to customise the application of the same models while simultaneously discovering and utilising dataset-specific models.

We have learned a lot about the models that work best for each category of dataset by customising our modelling techniques to fit the unique features and behaviours of each dataset. Through the use of this strategy, firms may better utilise industry-specific knowledge and improve decision-making procedures while also increasing the accuracy of sales estimates.

This customised approach can be improved by continued investigation and experimentation with other datasets and sectors.

Chapter 6

CONCLUSION

Through our in-depth analysis using various models and datasets, We observed that certain models exhibit better performance when applied to specific features and datasets within a given industry. However, it is challenging to arrive at a definitive conclusion regarding specific models for specific industries, as the effectiveness of a model depends on the characteristics and nature of the data. Nonetheless, we can conclude that specific models can be successfully applied to a specific set of features and dataset types.Among the models we examined, the LSTM model demonstrated strong predictive capabilities for sequential data across various frequencies, including both high and low frequencies. On the other hand, the CNN model proved effective for non-sequential data, as it excelled in capturing spatial information. However, our exploration extended beyond single models, as we explored the benefits of hybrid models such as CNN-RNN and LSTM-FCN. The results were remarkable, showcasing improved performance compared to using single models alone. Similar success was observed with ensemble models that combined CNN, RNN, and LSTM.Furthermore, we ventured into employing recently developed models, such as LSTM with Attention Mechanism, Variational Autoencoder, and CONV 1D. Our analysis also highlighted the importance of recognizing the uniqueness of each dataset. Models that perform well on one dataset may not yield the same results when applied to another due to inherent differences in dataset features.Therefore, a customized approach that incorporates specific models tailored to the dataset characteristics is crucial. Building hybrid or ensemble models that leverage the strengths of individual models can enhance performance. Additionally, exploring the application of cutting-edge models is valuable, even though they may not always yield optimal results.In conclusion, our findings underscore the need for a nuanced approach in selecting and applying models. By considering the specific requirements of the dataset, embracing hybrid or ensemble models, and exploring recent advancements, we can unlock the full potential of models and maximize their predictive capabilities in various business intelligence applications.

Chapter 7

FUTURE ENHANCEMENT

Going forward, we have several exciting opportunities to fine-tune our custom sales forecasting across different datasets and industries. By diving deeper into research and experimentation, we can polish our models, paving the way for precise, industry-focused forecasting methods.

1) Broadening Dataset Analysis: Our study centered around four distinct datasets, but there's room to extend our focus to include more diverse sectors. This broader analysis can reveal novel patterns and trends, potentially demanding specialized modeling techniques.

2) Leveraging Modern Techniques: With the continuous evolution of technology and methodologies, it's crucial for us to keep pace with the newest advances in sales forecasting. This exploration can yield valuable insights, enhancing the precision of our forecasts.

3) Considering External Influences: Sales forecasts don't exist in a vacuum; they're shaped by a myriad of external elements - market trends, economic shifts, and customer behaviors. By infusing our models with these factors, we can fine-tune our forecasting accuracy.

4) Building Industry-Specific Models: Each sector has its unique nuances, demanding a tailored approach. By crafting models designed for specific industries, we can boost both the accuracy and relevance of our predictions.

By embracing these future strategies, we can further elevate our tailored sales forecasting approach. Through a process of ongoing exploration, adjustment, and teamwork, we will continually refine our models, leading to sharper predictions.

REFERENCES

1. Gamboa, J. C. B. (2017). "Deep learning for time-series analysis." *arXiv preprint arXiv:1701.01887*.

2. Ghosh, I. and Sanyal, M. K. (2021). "Introspecting predictability of market fear in indian context during covid-19 pandemic: An integrated approach of applied predictive modelling and explainable ai." *International Journal of Information Management Data Insights*, 1(2), 100039.

3. Karim, F., Majumdar, S., and Darabi, H. (2019). "Insights into lstm fully convolutional networks for time series classification." *IEEE Access*, PP, 1–1.

4. Karim, F., Majumdar, S., Darabi, H., and Chen, S. (2018). "Lstm fully convolutional networks for time series classification." *IEEE Access*, 6, 1662–1669.

5. Koprinska, I., Wu, D., and Wang, Z. (2018). "Convolutional neural networks for energy time series forecasting." IEEE, 1–8.

6. Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., and Shroff, G. (2016). "Lstm-based encoder-decoder for multi-sensor anomaly detection." *arXiv preprint arXiv:1607.00148*.

7. Momeny, M., Latif, A. M., Sarram, M. A., Sheikhpour, R., and Zhang, Y. D. (2021). "A noise robust convolutional neural network for image classification." *Results in Engineering*, 10, 100225.

8. Selvin, S., Vinayakumar, R., Gopalakrishnan, E., Menon, V. K., and Soman, K. (2017). "Stock price prediction using lstm, rnn and cnn-sliding window model." IEEE, 1643–1647.

9. Seyedan, M., Mafakheri, F., and Wang, C. (2022). "Cluster-based demand forecasting using bayesian model averaging: An ensemble learning approach." *Decision Analytics Journal*, 3, 100033.

10. Sharma, A. and Shafiq, M. O. (2020). "Predicting purchase probability of retail items using an ensemble learning approach and historical data." IEEE, 723–728.

11. Zhao, K. and Wang, C. (2017). "Sales forecast in e-commerce using convolutional neural network." *arXiv preprint arXiv:1708.07946*.

12. Zhao, Z., Chen, W., Wu, X., Chen, P. C., and Liu, J. (2017). "Lstm network: a deep learning approach for short-term traffic forecast." *IET Intelligent Transport Systems*, 11(2), 68–75.