

# Entity Embedding in Artificial Neural Networks: A Novel Approach to Sales Data Analysis and Forecasting

Christina Jayakumaran<sup>1</sup>, Sengole Merlin<sup>2</sup>, Vaishali R Kulkarni<sup>3,\*</sup>, Thompson Stephan<sup>4</sup> and Punitha S<sup>5</sup>

<sup>1</sup>Department of Computer Science and Engineering, Loyola-ICAM College of Engineering and Technology, Chennai, Tamil Nadu, India - 600034

<sup>2</sup>Strategic Education Inc., Texas, USA.

<sup>3</sup>Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India - 248002

<sup>4</sup>Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India - 248002

<sup>5</sup>Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India - 248002

## Abstract

This study delves into the realm of sales forecasting, a critical component for strategic business planning, encompassing staff scheduling, inventory management, and supply chain optimization. At its core, the research investigates the efficacy of advanced predictive analytics in sales forecasting, leveraging historical and current data to unearth patterns that guide business decision-making. The focus is primarily on the application and comparative analysis of two sophisticated algorithms: eXtreme gradient boost (XGBoost) and deep neural networks (DNNs). These methods are explored for their potential to enhance forecasting accuracy using sales data. A notable aspect of this research is exploring entity embedding within the artificial neural network (ANN), highlighting its relevance and application in the context of sales data analysis. This comprehensive approach aims to offer insights into the most effective predictive models for sales forecasting, contributing to the broader field of predictive analytics in business.

## Keywords

ANN, Entity Embedding, Sales Forecasting, Time Series Analysis, XGBoost

## 1. Introduction

In the rapidly evolving business landscape, the ability to accurately forecast sales has become a cornerstone for strategic planning and operational efficiency. Sales forecasting, a critical component in the vast domain of supply chain management, significantly impacts areas like inventory management, staff scheduling, and supply chain optimization. The accuracy of sales forecasts directly influences a company's ability to make informed decisions, manage resources effectively, and maintain a competitive edge in the market [1]. Despite its critical importance, traditional sales forecasting methods often fall short in today's dynamic and complex market environments. These methods, typically grounded in statistical analysis of historical data, struggle to adapt to the nonlinear and evolving patterns of consumer behavior and market trends. The limitation of traditional approaches in handling large and varied datasets underscores the need for more advanced and sophisticated forecasting techniques. In response to these challenges, this paper introduces machine learning as a transformative solution for enhancing the accuracy and dynamism of sales forecasts. Machine learning, with its capability to process and learn from vast amounts of data, presents an opportunity to develop more robust and adaptable forecasting models. Among the various machine learning algorithms, XGBoost and DNNs are notable for their efficacy in predictive modeling tasks. This research aims to explore and evaluate

*Symposium on Computing & Intelligent Systems (SCI), May 10, 2024, New Delhi, INDIA*

\*Corresponding author.

✉ christina@licet.ac.in (C. Jayakumaran); sengolemerlin@gmail.com (S. Merlin); vaishali@ieeee.org (V. R. Kulkarni);

thompsoncse@gmail.com (T. Stephan); punitharesearch@gmail.com (P. S)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

the application of these two algorithms in the context of sales data analysis, offering a comparative analysis of their performance and suitability.

A novel aspect of this study is the incorporation of entity embedding within ANN for sales data analysis. Entity embedding, a technique for transforming categorical variables into numerical forms, is particularly relevant in sales forecasting, where data often comprises a mix of categorical and numerical information. By integrating entity embedding, the research aims to enhance the predictive capability of ANNs, enabling them to handle the intricacies of sales data more effectively.

The primary objective of this paper is to present a comprehensive analysis of XGBoost and DNNs in sales forecasting, highlighting the benefits and limitations of each approach. Additionally, the paper explores the innovative application of entity embedding within ANNs, aiming to contribute to the broader field of predictive analytics in business. The structure of the paper includes a detailed examination of existing sales forecasting techniques, the methodology for implementing the machine learning models, and a comparative analysis of the performance of XGBoost and DNNs, culminating in a discussion of the implications and potential applications of entity embedding in sales data analysis [2]. Through this research, we aim to provide valuable insights and methodologies for businesses seeking to improve their sales forecasting capabilities in the face of rapidly changing market dynamics.

## 2. Literature Review

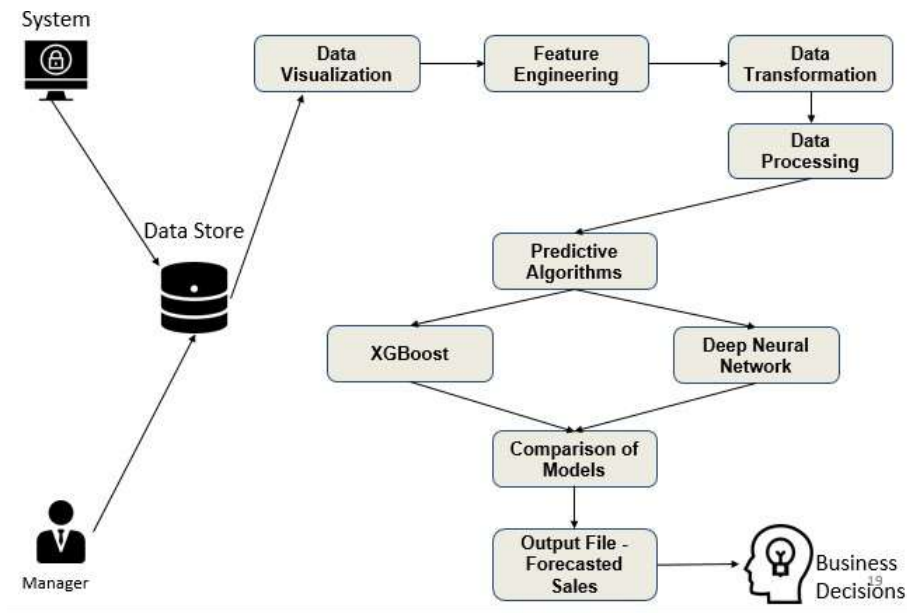
The use of machine learning (ML) techniques in predictive analytics for businesses has been widely studied. For instance, a Machine Learning Model and Rule Engine can be used to predict sales forecasts on historical data, which can be fed to a front-end application for processing [3]. This application will show the predicted data for the following week. The XGBoost Sales prediction model created for the Walmart dataset was used as a reference point for the model presented in this chapter [4]. This paper outlines the advantage of using an XGBoost model as the corresponding error scores are 16.3% and 15.4% lower than Linear Regression and Ridge regression models respectively. In [5] the application of the XG Boosting algorithm is implemented. The insights on feature engineering through the use of XGBoost feature importance ranking, have been implemented in the proposed model.

XGBoost is frequently cited as a highly effective ML model that can be applied to many problems. However, the key to developing an efficient model is to choose the right features beforehand, by performing data analysis and feature engineering techniques [6]. In recent years, XGBoost has been widely used in forecasting as it performs comparatively better than classic regression models. In other words, XGBoost outperforms the existing models by performing better in a shorter period [7]. In comparison to traditional models such as Linear regression models and Support Vector Machines, ANN also perform well. On applying ANN to Walmart Sales data ANN were found to have the lowest RMSE scores when compared to the previously mentioned traditional models [8]. In fast-paced industries, quick and reliable sales forecasting models are an invaluable resource. One such example is the fashion or retail industry. Sales predictions using NN models in this industry have proven to perform well, with low root mean square percent error (RMSPE) and mean square error (MSE) scores [9]. Another essential practice in the retail industry is promotional sales, which can be cold-start forecasted using gradient boosting algorithms [10].

## 3. Existing Systems

There are two types of sales forecasting used by most companies: quantitative and qualitative Sales Forecasting. Quantitative sales forecasting requires numerical data, so some commonly used data includes consumer spending and economic trends. In the simplest form, quantitative sales forecasting uses linear equations to calculate predicted values as given in (1).

$$Y = a_0 + a_1 + a_2^2 + \dots + a \quad (1)$$



**Figure 1:** Proposed Architecture

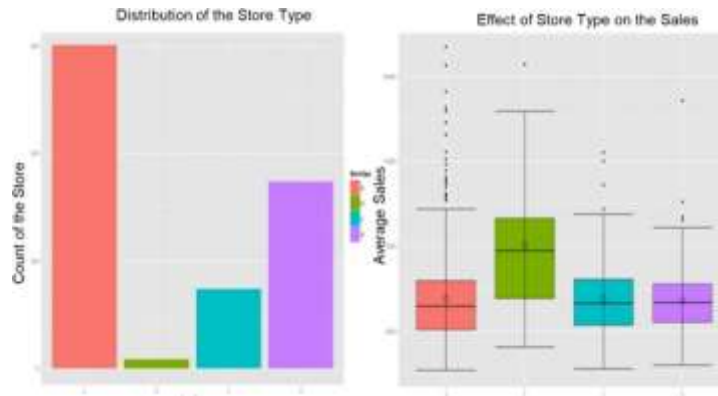
**Table 1**  
Distribution of Stores over Assortments

Store Type/Assortment Level	Level 'a'	Level 'b'	Level 'c'	Total Stores
Store Type A	381	0	221	602
Store Type B	7	9	1	17
Store Type C	77	0	71	148
Store Type D	128	0	220	348
Total Stores	593	9	513	1115

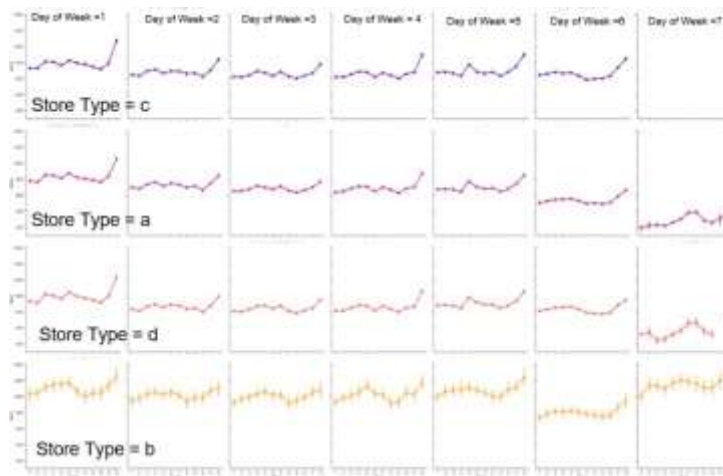
As more data is collected, forecasting models become more complex, so generic linear models will not predict reliable forecast values. Quantitative models are also not useful for businesses with no historical data, or with limited historical data. These models perform much better with data from established businesses. Qualitative Sales forecasting is used in multiple applications with a bigger scope in comparison with quantitative sales forecasting. Qualitative sales forecasting is classified as the jury of executive opinion, delphi technique, sales force composite, and surveyor of buyer intentions.

## 4. Proposed Architecture

For the proposed system, sales data was collected from 1115 Rossmann stores in Germany. This data was wrangled and analyzed. This system will predict the sales of each store in the company for the next 6 weeks. The different activities in promoting, distributing, and consumption of various products and services are covered under economic features. The forecasting of sales is based on various economic features, and it covers several visual and display options called temporal features. The other parameters include promotion policies and steps in promoting the product, the season of the sale that comprises school holidays, locations and other competing stores, the different times of the year, etc. The overall proposed architecture for the sales data forecasting and analysis is depicted in Figure 1. Commonsense reasoning and analytic knowledge methods are used for building the model. These two methods are used during the analysis of the data and are helpful in the development of a definitive conclusion. Two machine learning models are developed, one named XGBoost specific to the Rossmann company, and the other a neural network generic to all companies for their sales forecasting. The details of distribution of stores over assortments is as given in Table 1.



**Figure 2:** Distribution of Store Models and Average Sales



**Figure 3:** Day of the Week Versus Sales in each store

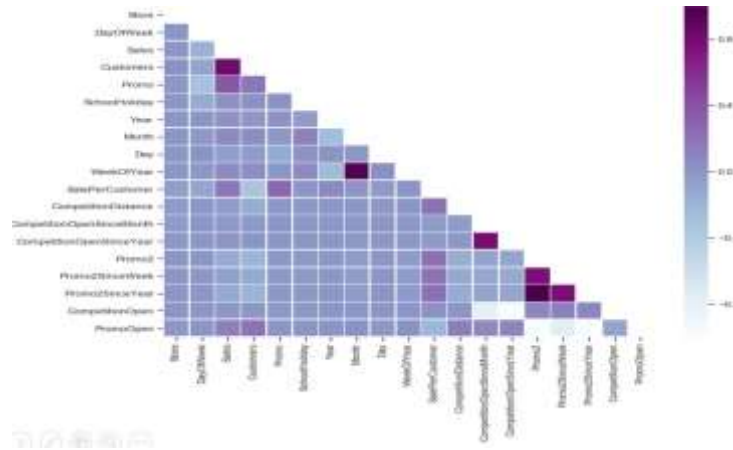
## 5. Implementation

### 5.1. Data

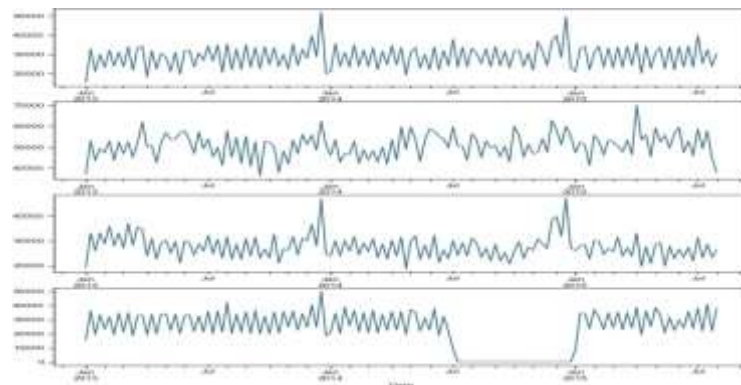
The data used here has been collected from Kaggle, a popular website for Data Science related projects. The dataset of Rossmann Drugstore that was published for a competition for predicting sales is used. The Google Trends data, weather data, and state (location of the store) data of each store of Rossmann on each day were used as external factors to predict the sales. The training data contains 1, 017, 210 records and the testing data contains 41089 records. On performing Exploratory Data Analysis (EDA), there are some significant correlations. Figure 2 denotes the distribution of the store types and its effects on the sales value. Figure 3 depicts the day of the week and its correlation with the sales value for each of the store type. The number of sales and the number of customers have a strong positive correlation, which means that more customers lead to better sales. This, however, was an expected trend. If the store offers a promotion (Promo 1), then the number of customers increases, which leads to more sales. Adversely, if the store offers a consecutive promotion (Promo 2), then the number of customers remains the same or even decreases slightly. The Pearson's correlation matrix of the attributes is depicted in Figure 4. The weekly sales status trend and yearly sales trend is depicted in Figure 5 and Figure 6 respectively. The following points give a general overview of the data as a whole:

- The most selling and crowded Store Type is A.
- The best Sale per Customer (Store Type D) shows the higher Buyer Cart. We could also assume that the stores of this type are in rural areas, so customers prefer buying more but less often.

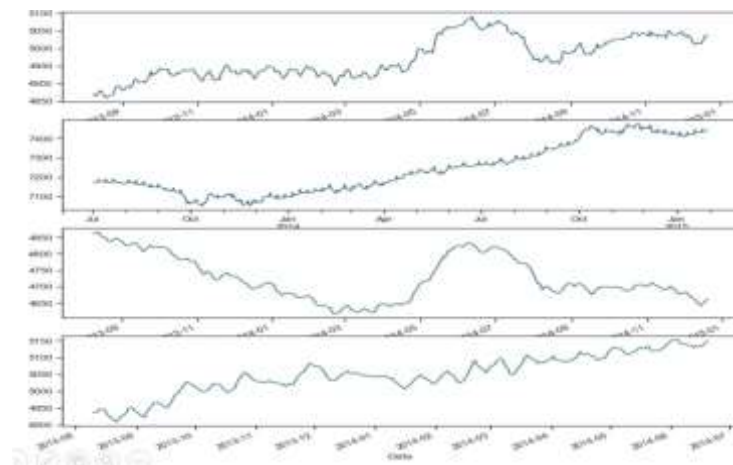
- Low Sale\_Per\_Customer amount for Store Type B shows the possibility that people shop there essentially for small things. This can also indicate the label of this store type - “urban” - as it is more accessible to the public, and customers don’t mind shopping there from time to time during the week.
- Customers tend to buy more on Mondays when there is one promotion running (Promo) and on



**Figure 4:** Pearson's Correlation Matrix

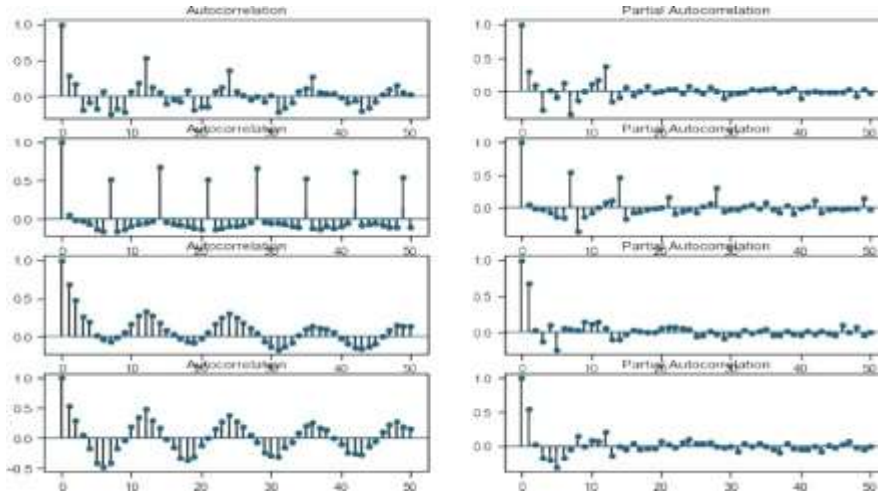


**Figure 5:** Weekly Status



**Figure 6:** Yearly Status





**Figure 7:** Autocorrelation and Partial Correlation of Each Store

Sundays when there is no promotion at all.

- Promo2 alone does not seem to be correlated to any significant change in the Sales amount.

## 5.2. Time Series Analysis of Data

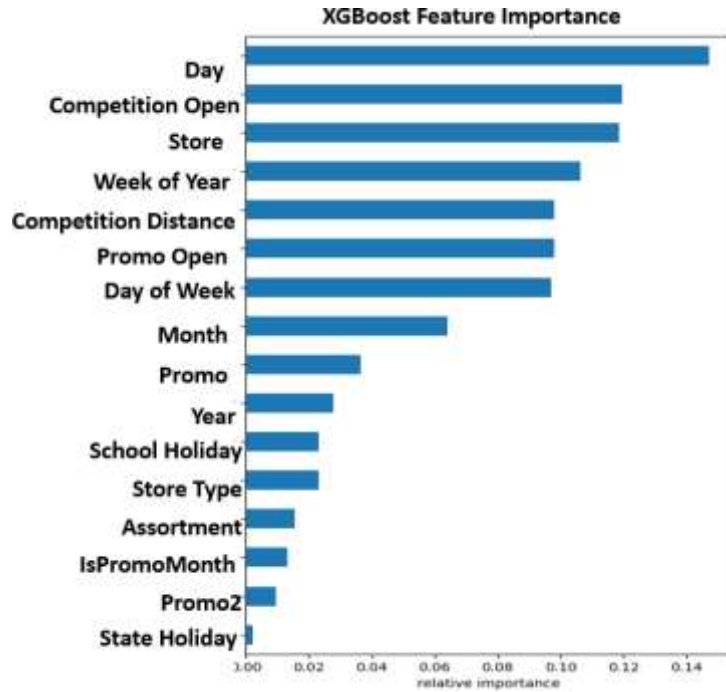
A time series analysis is done using the different store types but not by refereeing the individual store. Overall sales seem to increase, but not for Store Type C (a third from the top). Even Though Store Type A is the most selling store type in the dataset, it can follow the same decreasing trajectory as Store Type C did. The non-randomness of the time series and high lag-1 are common things for each pair of plots. The probability is that these two entities may probably need a higher order of differencing  $d/D$ . Type A and type B: Both types show seasonality at certain lags. For type A, it is each 12th observation with positive spikes at the 12 (s) and 24(2s) lags and so on. For type B it's a weekly trend with positive spikes at the 7(s), 14(2s), 21(3s) and 28(4s) lags. Type C and type D: Plots of these two types are more complex. The auto correlation and partial correlation of each store is shown in Figure 7. It seems like each observation is correlated to its adjacent observations.

## 5.3. XGBoost

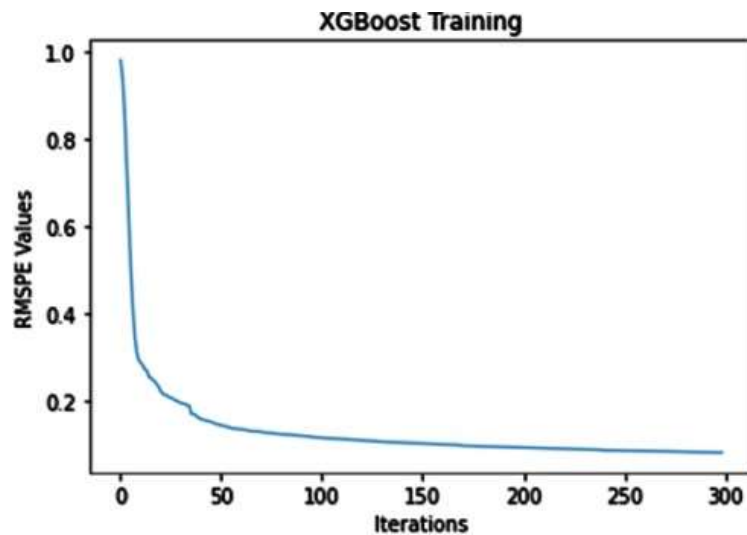
Boosting is considered to be an optimization method. Gradient boosting is a machine learning algorithm that is frequently used in regression and classification. It produces prediction models, mostly in the form of decision trees. Weak learners can be combined to create one strong learner. XGBoost, in particular, is used for supervised learning tasks. The XGBoost makes use of the term regularization to manage the complexity of the model. With the help of regularization, the problem of overfitting can be prevented. A popular tree ensemble model is an XGBoost and it consists of a set of classification and regression trees (CART). The tree includes the family members as distinct leaves and each leaf is allotted a score value. The CART includes a real score along with the decision value. The use of real scores helps in interpreting value in a better way than just the classification. XGBoost is optimized for boosting tree algorithms. Figure 8 indicates the order of importance of each feature and the RMPSE values against the iterations are denoted. The XGBoost provides a better framework than gradient boosting and it is faster than existing gradient boosting algorithms. XGBoost is based on the linear solver model, and it includes various objective functions. The functions make use of regression, classification, and ranking methods. The objective function used is given in (2).

$$O(\theta) = L(\theta) + \lambda(\theta) \quad (2)$$

where  $L(\theta)$  is a training loss and  $\omega(\theta)$  refers to regularization.



(a) Feature Importance based on XGBoost



(b) RMSPE - During the Training Phase of the XGBoost

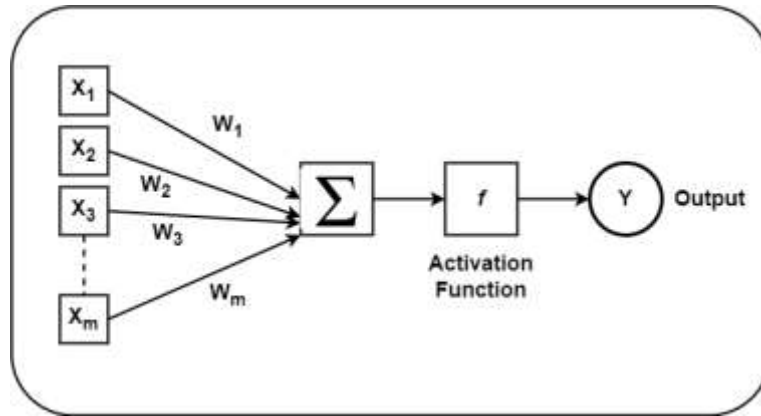
**Figure 8:** XGBoost Details

## 5.4. Artificial Neural Network

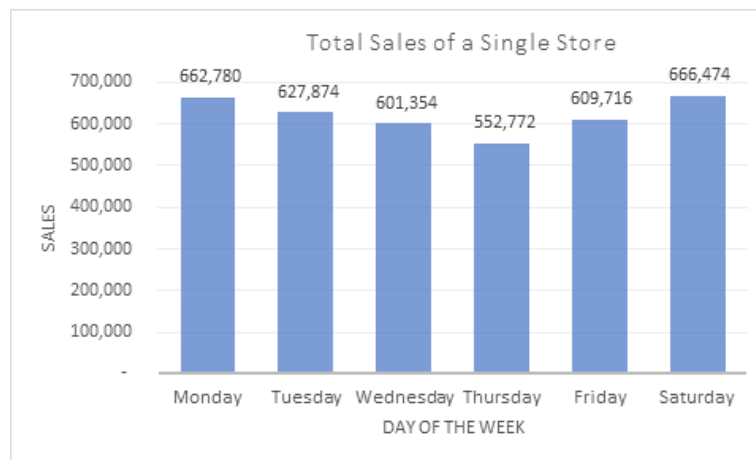
In a neural network, the inputs are multiplied by their corresponding weights and then summed together. This is depicted in 9. The sum is passed through an activation function which selects which neurons to activate within the network. The corresponding output is a single value. The output function  $\mathcal{Y}$  is represented in (3).

$$\mathcal{Y} = \mathcal{F}(o + i_1 + 2_2 + 3_3 + \dots + ) \quad (3)$$

The activation functions considered are the sigmoid and ReLU function. The sigmoid function maps values between 0 to 1. The sigmoid function is frequently used in machine learning models that work on probabilities. This function is both monotonic and differentiable. Rectified Linear Unit, or ReLU, is a widely used activation function. In ReLU, the gradient is positive for any positive input values, but there is no gradient (or gradient = 0) for negative values. ReLU ranges from 0 to infinity. Both the



**Figure 9:** ANN



**Figure 10:** Total Sales for Store ID = 1

function and its derivative are monotonic.

## 5.5. Entity Embedding

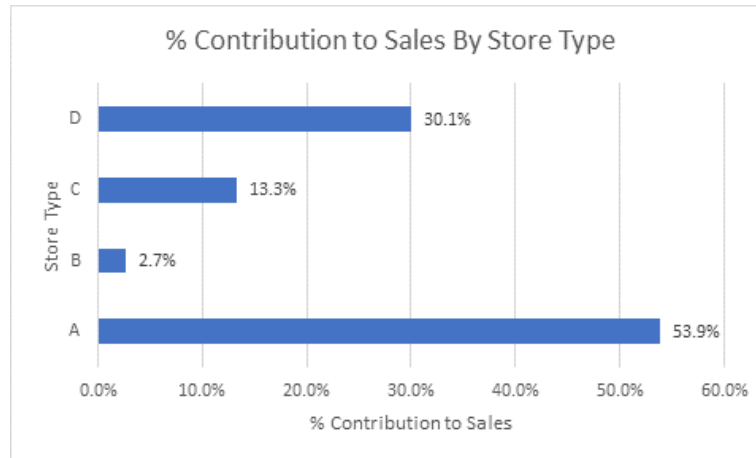
Entity embedding involves the mapping of categorical variables into Euclidean spaces. The neural network learns this mapping during the training phase. Embedding reduces memory usage and increases the speed of neural networks. Similar values are mapped in the embedding space for disclosing the intrinsic properties of the categorical variables. When the data sets have large high cardinality features, the other methods are overfit and cannot be used. In this study, the data sparsity problem is overcome by representing the discrete category features in a continuous space. The category similarity is reflected using the distance between category points. The idea is to utilize the data points close by, which is used for approximating missing data points.

In Figure 10 and Figure 11, the focus is on the overall sales performance of a single store (Store ID = 1). It can be seen that Saturdays are the most profitable days of the week. This figure highlights sales

**Table 2**  
Performance Evaluation

Algorithm	RMSPE Metric	Time (seconds)
XGBoost	0.094663	541.53 s (9 mins)
Deep Neural Networks	0.1015	1839.65 s (30 mins)





**Figure 11:** Contribution to Overall Sales By Store Type

optimization techniques at the granular level (in a single store, on a particular day). Also the focus is given to how each store type performs. Store Type A has the highest percentage of overall sales with an overwhelming 53.9%. The performance measurement based on the RMSPE metric and time is expressed in Table 2.

## 6. Conclusion

Two models based on machine learning models namely XGBoost and entity embedding neural network are used for sales forecasting in stores. The task involved predicting the sales on any given day at any store. The studied previous work is adapted in the domain including time series algorithms in machine learning. The patterns and outliers are identified using analysis of the data. The analysis has boosted the prediction algorithm. XGBoost has performed best at prediction and slightly better than neural networks. However, a neural network is suggested to be used for forecasting sales of those companies whose sales trend deviates from that of Rossmann's sales used in this project as an experimental dataset. The major parameter used for fitting is the measurement of the overall prediction error rather than the specific decomposition of error into bias and variance. The uncorrelated sales responses in various data stores are presented using RMSPE.

## Acknowledgement

Authors acknowledge the support received from Graphic Era Deemed to be University, Dehradun, India.

## References

- [1] A. Ahlemeyer-Stubbe, S. Coleman, A Practical Guide to Data Mining for Business and Industry, 1st ed., Wiley Publishing, 2014.
- [2] M. Seyedan, F. Mafakheri, Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities, *Journal of Big Data* (2020).
- [3] M. A. Khan, S. Saqib, T. Alyas, A. Ur Rehman, Y. Saeed, A. Zeb, M. Zareei, E. M. Mohamed, Effective demand forecasting model using business intelligence empowered with machine learning, *IEEE Access* 8 (2020) 116013–116023.
- [4] X. dairu, Z. Shilong, Machine learning model for sales forecasting by using xgboost, in: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2021, pp. 480–483.

- [5] Y. Niu, Walmart sales forecasting using xgboost algorithm and feature engineering, 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE) (2020) 458–461.
- [6] S. Ghosh, C. Banerjee, A predictive analysis model of customer purchase behavior using modified random forest algorithm in cloud environment, in: 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE), 2020, pp. 239–244.
- [7] R. P, S. M, Predictive analysis for big mart sales using machine learning algorithms, in: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1416–1421.
- [8] J. Chen, W. Koju, S. Xu, Z. Liu, Sales forecasting using deep neural network and shap techniques, 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) (2021) 135–138.
- [9] C. Giri, Y. Chen, Deep learning for demand forecasting in the fashion and apparel retail industry, *Forecasting* 4 (2022) 565–581.
- [10] C. Aguilar-Palacios, S. Muñoz-Romero, J. L. Rojo-álvarez, Cold-start promotional sales forecasting through gradient boosted-based contrastive explanations, *IEEE Access* 8 (2020) 137574–137586.