Retail Sales Forecasting using Deep Learning

Vishnu Varakala

U00864767

Independent Studies

**Contents:**

1. Project Background & Overview 3

2. Problem Statement 3

3. Research Questions being Answered 3

4. Literature Review 4

5. Machine Learning Techniques used for Current Work 4

6. Methodology 5

6.1 Overview 5

6.2 Data Source 6

6.3 Data Cleaning and Transformation 7

7.Feature Selection and Engineering 11

8. Model Selection and Evaluation 11

9. Model Results and Discussion 16

Model Evaluation 19

10. Discussion on Key Findings 22

10. Conclusion 23

11. Reference 25

# **Project Background & Overview**

Demand forecasting is essential for the growth and sustenance of any business. **S**upply chain and Inventory management is very vital any large- or small-scale business. For any business to grow and sustain a long run in a market maintaining balanced flow is not an option but a must. Having an intuition on upcoming increase or decrease in demand helps businesses in making better strategic decisions related to inventory, which in turn not only improves cashflow and customer experiences (product availability) but also reduces retail shrinkage (due to wastage).

# **Problem Statement**

*Complex problems needs innovative solutions”.* Due to frequent changing & complex consumer purchase patterns traditional approaches like Moving Average, ARIMA and ARIMAX etc., might not provide expected demand forecast accuracy levels. With appropriate data treatment advanced Deep Learning techniques like MLP, LSTMs etc., can provided satisfactory results. The current project scope is to predict future retail store sales using MLP, CNN and LSTM deep learning techniques.

Each technique is evaluated on test and train datasets and final model is selected post comparative study.

# **Research Questions being Answered.**

By analyzing sales data using deep learning techniques such as artificial neural networks and convolutional neural networks, we can gain insights into the patterns and trends that affect retail sales. These insights can be used to develop more accurate forecasting models and help retailers make better decisions about product development, pricing, marketing, and inventory management.

1. Which products are likely to sell the most during a particular time period?
2. What factors influence the sales of a particular product in a specific location?
3. How accurate can deep learning models predict retail sales compared to traditional forecasting methods?

# **Literature Review**

Traditionally approaches like ARIMA (Jamal Fattah, Latifa Ezzine. 2018) etc., were being used for forecasting demand across various business domains including retail. With the increase in usage as well as acceptance of deep learning models in solving business problems, application of deep learning has also increased in the field of demand forecasting. Deep learning algorithms like CNN (Getu Tadele Taye, Han-Jeong Hwang & Ki Moo Lim, 2020) and LSTM (Rathipriya, R., Abdul Rahman, 2022) have been successful applied in the past for forecasting problem in Finance, Pharma etc., domains. In the current work we have carried out a comparative study on the performance of each of these models on retail data.

# **Machine Learning Techniques used for Current Work**

In the current work Three different deep learning methods have been tried out in the current project for predicting store sales. Which includes Multilayer Perceptrons, Convolutional Neural Networks and Long Short- Term Memory

**5.1 Multilayer perceptron** **models** are one of the simplest forms of neural networks. As the name suggests MLP model consists of multiple layers of stacked perceptions units (simple neural network units). Due to the basic nature of MLP, they don’t have any concept of attention/memory which is essential for prediction tasks like forecasting were outcome from previous period plays a key role for predicting future.

Selection Intuition:

MLP approach has been selected as baseline model for current prediction task. Though MLP models don’t have the concept of looking back in time from algorithmic perspective, it can perform well if we induce it manually by creating engineering features based on time lag.

**5.2 Convolutional Neural Networks (CNN)**

Convolutional Neural Networks or CNN are special forms of neural network originally developed for computer vision related tasks. CNNs are designed to extract useful patterns/features from a given dataset using key concepts of convolutions & filters.

Though CNNs are majorly used for computer vision related tasks, recently there has been an increase in application of CNNs to time series data.

Selection Intuition:

CNNs in computer vision achieves prediction tasks through key feature/pattern extraction by individual filters. Time series forecasting can be viewed similar to that of computer vision where key patterns in peaks/valleys can be identified by filters. Due to this reason CNN is selected experimentation.

**5.3 Long Short-Term Memory (LSTM)**

Long Short-Term Memory or LSTM models are advanced forms of neural networks (especially RNN) designed to overcome the problem of vanishing gradient and achieving long term memory dependencies. LSTM is majorly applied in NLP field as they can generate predictions using data sequencies.

Similar to the problem of predicting the next word based on previous words in context, in NLP area, forecasting involves predicting future values based on past trends. Due to the capability of LSTMs model the trend behavior using memory, they can be applied to current forecasting problem with minimum data preparation.

Models on the three selected algorithms, listed above, are developed, and tested on preprocessed dataset. Were

* Model performance on hold out data (test data) is used for final model selection.
* Model development and evaluation is performed at store-month level data.
* MAPE and MAE are used as the primary metric for model validation and selection.

# **Methodology**

**6.1 Data Collection and Pre-processing**

**Data Collection:**

Data used for the project consists of daily sales data at transaction level (Store – Item). Originally data belongs to on the Russia’s largest software firm – 1C Company. Which then was made available in [Kaggle.](https://www.kaggle.com/competitions/demand-forecasting-kernels-only/data)( <https://www.kaggle.com/competitions/demand-forecasting-kernels-only/data>) Overall, 4 different tables pertaining to daily sales, product info, product category info and store info are available as part of data procured. Final master dataset (flat table format) is created by appropriately joining individual data sources. **Error! Reference source not found.** provides a view of key data fields and quick EDA comments.

**6.2 Data Cleaning and Transformation**

**Data cleaning:**

Post the data has been collected checked for missing values there are no missing in the data and checked if there is any issue with datatypes of features in the dataset there is no need of such. We then performed the exploratory data analysis.

**Preprocessing:**

Overall data considered for analysis in the current project has 2.9M records where each record is at a granularity of store-item-date. Data has 6 numeric, 3 categorical and 1 datetime field. And then worked to perform exploratory data analysis to find insights or pattern in the data.

* Below shows that 30/60 (~50%) of the store are contributing to 80% of the sales. Nailing forecast for these 50% of the stores will generate quantifiable impact.

Calendar

Description automatically generated

* From the Items which are sold during model period, only 8% of items contributed 80% sales.

Calendar

Description automatically generated

* We have noticed a spike every December, this might be due to festive period. But 2015 observed a decline which is counter intuitive, showing a possibility of data capture issue.

Chart, line chart

Description automatically generated

* Low sales at the beginning of every month and spike in the mid of month

Chart, line chart

Description automatically generated

* Low sales captured in weekends.

Chart, line chart

Description automatically generated

# **Feature Selection and Engineering**

* Initial data consists of multiple files items.csv, category.csv, sales train, shop.csv all these files are merged to form a single file.
* We have main feature date has been transformed to develop multiple features like day, month, year, weekday, year month these variables feature engineered which helped in find the sales trend.
* Since my motive of the project is forecast the sales of the store dropped few features like shop name, item name, item category name, item id

1. **Model Selection and Evaluation**

Multiple experiments were performed on proposed three models by changing Network structure (Hidden units, Hidden layers etcs.,) and other model parameters like optimizers, activation functions etc.,

Graphical user interface

Description automatically generated

Figure 3: High Level Model Architecture

## **Multilayer Perceptrons (MLP) Experimentation**

* + - Multiple experimentation/iterations performed on MLP models; Best model exhibited 45% test MAPE on dataset with 29% train MAPE
    - Iterations with high train error exhibited high test error
    - MLP model showed consistent performance on both Test and Validation dataset suggesting consistency of model prediction.

Refer to [Table](#_bookmark4) below for sample MLP model experiment results log

Text

Description automatically generated

Table 3: MLP Model Experimentation Log (sample)

**8.2 Convolutional Neural Networks (CNN) Experimentation**

* Multiple experimentation/iterations performed on MLP models; Best model exhibited 52% test MAPE on dataset with 36% train MAPE
* Iterations with high train error exhibited.
* CNN model **showed** moderate performance variation from in Train, Test and Validation datasets.

Text

Description automatically generated

Refer to [Table](#_bookmark5) below for sample CNN model experiment results log

**8.3 Long Short-Term Memory (LSTM) Experimentation**

* + - Multiple experimentation/iterations performed on LSTM models; Best model exhibited 68% test MAPE on dataset with 70% train MAPE
    - Iterations with high train error exhibited high test error
    - LSTM model showed high variation on Train vs Validation dataset. This can be a sign of moderate overfitting.

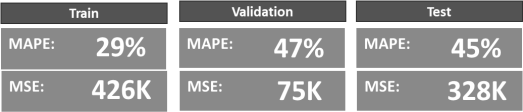
Text

Description automatically generated with medium confidence

LSTM Model Experimentation Log (sample)

1. **Results and Discussion.**

The best test MAPE of 45% is observed in one of the MLP iteration/experiment. MLP model exhibited better as well as consistent performance amount other to methods selected. Error! Reference source not found. and 8 shows MAPE across all the three.

MLP Model Best Iteration Results (Model stats reported at Store-Month level

Graphical user interface, text, application, chat or text message

Description automatically generated

CNN Model Best Iteration Results (Model stats reported at Store-Month level)

Graphical user interface, text, application, chat or text message

Description automatically generated modeling approaches (for best experiment).

LSTM Model Best Iteration Results (Model stats reported at Store-Month level)

MLP models exhibited consistent model accuracy performance across test and validation dataset in comparison with CNN & LSTM. Refer to [Figure](#_bookmark7) [4: Train, Validation and Test Accuracy for various](#_bookmark7) [iterations of MLP, CNN and LSTM model](#_bookmark7) [respectively](#_bookmark7)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Fig 4: Train, Validation and Test Accuracy for various iterations of MLP, CNN and LSTM model respectively

1. **Discussion of Key Findings**

Traditionally approaches like ARIMA (Jamal Fattah, Latifa Ezzine. 2018) etc., were being used for forecasting demand across various business domains including retail. In my approach I have applied three model MLP, CNN, LSTM out of all these model MLP performed well compared to CNN and LSTM.

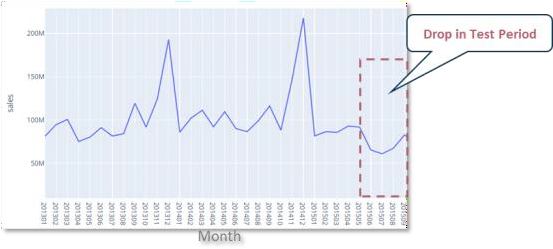
* + - MLP model showed consistent performance on both Test and Validation dataset suggesting consistency of model prediction. While CNN model **showed** moderate performance variation from in Train, Test and Validation datasets and LSTM model showed high variation on Train vs Validation dataset. This can be a sign of moderate overfitting.

1. **Conclusion**

Of the three model architectures explored in current work Multilayer Perceptron model exhibited best test error (MAPE) in comparison with Convolutional Neural Networks and Long- Short-Term-Memory models. Best experiment/iteration of MLP model produced 45% Test MAPE with is ~7% more than the best experiment of CNN and 23% more than the LSTM’s best experiment.

Model performance on train and validation datasets is relatively consistent in MLP experiments than that of CNN and LSTM. Though Convolutional showed good train MAPE of 36%, but it failed to perform on test dataset, while LSTM models failed to

perform during both training and testing. One of the major reasons for all models exhibiting high test error is due to abnormal sales dip during the test period which was never observed in the past. Refer to Figure 5 below on the dip observed.

Exploring the reason for abnormal dip can help in treating the data for improved model performance and validation.

(Monthly sales trend with abnormal drop in the during test period)

1. **References**

[1] Jamal Fattah, Latifa Ezzine. 2018. Forecasting of demand using ARIMA model. International Journal of Engineering Business Management

[2] Getu Tadele Taye, Han-Jeong Hwang & Ki Moo Lim. 2020. Application of a convolutional neural network for predicting the occurrence of ventricular tachyarrhythmia using heart rate variability features. Sci Rep10,676. <https://doi.org/10.1038/s41598-020-63566-8>

[3] Rathipriya, R., Abdul Rahman, A.A., Dhamodharavadhani, S. et al. 2022. Demand forecasting model for time-series pharmaceutical data using shallow and deep neural network model. Neural Comput & Applic. <https://doi.org/10.1007/s00521-022-07889-9>

[4] R. Carbonneau et al.. 2008. Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research

[5] Linda Eglite and Ilze Birzniece. 2022. Retail Sales Forecasting Using Deep Learning: Systematic Literature Review. Published online by RTU Press, [https://csimq-journals.rtu.lv.](https://csimq-journals.rtu.lv/)

[6] Purvika Bajaj, Renesa Ray. 2020. SALES PREDICTION USING MACHINE LEARNING

ALGORITHMS. International Research Journal of Engineering and Technology (IRJET)

[7] Md. Shiblee, P. K. Kalra & B. Chandra. 2019. Time Series Prediction with Multilayer Perceptron (MLP): A New Generalized Error Based Approach. Part of the Lecture Notes in Computer Science book series (LNTCS,volume 5507)

[8] H.D.Nguyen, K.P.Tran, S.Thomassey, M.Hamad. 2019. Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with applications in supply chain management. International Journal of Information Management, 102282

[9] Ackermann, N.(2018). Introduction to 1D convolutional neural networks in Keras for time sequences. [online] Medium. Available at: [https://blog.goodaudience.com/introduction-to-1d-](https://blog.goodaudience.com/introduction-to-1d-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf) [convolutional-neural-networks-in-keras-for-time-](https://blog.goodaudience.com/introduction-to-1d-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf) [sequences-3a7ff801a2cf](https://blog.goodaudience.com/introduction-to-1d-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf)

[10] I. Alon, M. Qi, R.J. Sadowski (2001). Forecasting aggregate retail sales: A comparison of Artificial Neural Networks and traditional methods. Journal of Retailing and Consumer Services, 8 (3) (2001),pp. 147-156

[11] A. Gensler, J. Henze, B. Sick, N. Raabe (2016). Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM neural networks. Proceeding of the IEEE international conference on systems, man, and cybernetics (SMC), IEEE, pp. 002858-002865

[12] S. Selvin, R. Vinayakumar, E.A. Gopalakrishnan, V.K. Menon, K.P. Soman. Stock price prediction using LSTM, RNN and CNN-sliding window model. Proceedings of the international conference on advances in computing, communications and informatics (ICACCI), IEEE (2017), pp. 1643-1647 International Journal of Information Management, 57 (2021), Article 102282