Retail Sales Forecasting using Deep Learning

Vishnu Varakala and Sukesh Donthineni

Abstract

Demand forecasting is very essential for growth and sustenance of any business. 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.

1 Introduction

Supply chain and Inventory management is very vital any large- or small-scale business. For any business to grow and sustain 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).

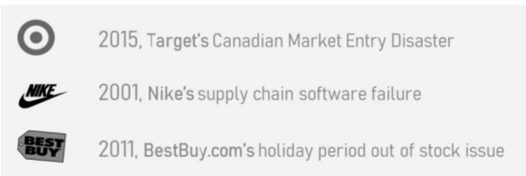

Figure 1: Historical Supply Chain Disasters in Retail

Be it 2001, Nike's supply chain failure or 2011, BestBuy.com holiday period out of stock issue the importance of Demanding forecasting is time tested.

2 Related Work

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.

3 Data

3.1 Data Description

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. 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.

3.2 Key Summary Stats

Overall data consider for analyzing in the current article 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. Refer to Table 1 and Table 2 showing summary stats for each numerical field

Refer to Figure 2: Key Data Trends showing key data trends discussed above

Metric	Description	Analysis Comment
shop_id	Unique identifier of a shop	Total 80 unique shops
item_id	Unique identifier of a product	~20K Items
item_category_id	Unique identifier of item category	64 Categories
item_cnt_day	Number of products sold. You are predicting a monthly amount of this measure	-
item_price	Current price of an item	-
date	Date in format dd/mm/yyyy	Jan' 2013 to Dec' 2015
date_block_num	A consecutive month number, used for convenience.	-
item_name	Name of item	~20K Items
shop_name	Name of shop	Total 80 unique shops
item_category_name	Name of item category	64 Categories

Table 2: Data Dictionary of Key Field along with Analysis Comments

```
n [58]: train.describe()
ut[58]:
```

	date_block_num	shop_id	item_id	item_price	item_cnt_day	item_category_id	sales	year	month	yearmo
count	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06	2.935849e+06
mean	1.456991e+01	3.300173e+01	1.019723e+04	8.908532e+02	1.242641e+00	4.000138e+01	1.157732e+03	2.013777e+03	6.432552e+00	2.013841e+05
std	9.422988e+00	1.622697e+01	6.324297e+03	1.729800e+03	2.618834e+00	1.710076e+01	5.683604e+03	7.684790e-01	3.504695e+00	7.662597e+01
min	0.000000e+00	0.000000e+00	0.000000e+00	-1.000000e+00	-2.200000e+01	0.000000e+00	-6.897000e+04	2.013000e+03	1.000000e+00	2.013010e+05
25%	7.000000e+00	2.200000e+01	4.476000e+03	2.490000e+02	1.000000e+00	2.800000e+01	2.490000e+02	2.013000e+03	3.000000e+00	2.013080e+05
50%	1.400000e+01	3.100000e+01	9.343000e+03	3.990000e+02	1.000000e+00	4.000000e+01	4.490000e+02	2.014000e+03	6.000000e+00	2.014030e+05
75%	2.300000e+01	4.700000e+01	1.568400e+04	9.990000e+02	1.000000e+00	5.500000e+01	1.078200e+03	2.014000e+03	9.000000e+00	2.014120e+05
max	3.300000e+01	5.900000e+01	2.216900e+04	3.079800e+05	2.169000e+03	8.300000e+01	1.829990e+06	2.015000e+03	1.200000e+01	2.015120e+05

Table 1: Key Summary Stats

3.3 Data Key EDA (Exploratory Data Analysis) Insights

- 30/64 (~50%) of the store are contributing to 80% of the sales. Nailing forecast for these 50% of the stores will generate quantifiable impact
- From the Items which are sold during model period, only 8% of items contributed 80% sales
- 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
- Low sales at the beginning of every month and spike in the mid of month
- Low sales captured in weekends

4 Method

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

4.1 Multilayer Perceptrons (MLP)

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 doesn'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.

4.2 Convolutional Neural Networks (CNN)

Convolutional Neural Networks or CNN are special form of neural network originally developed for computer vision related task. 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 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 this reason CNN is selected experimentation.

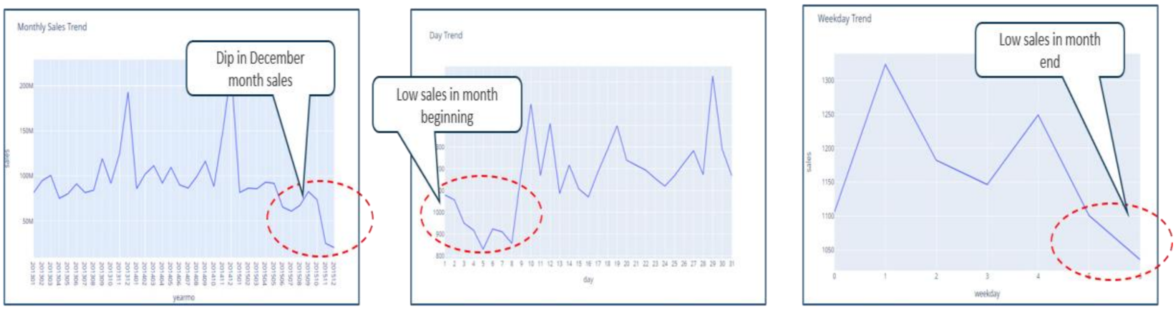

Figure 2: Key Data Trends

(Monthly Sales [Left], Day of Months Sales [Middle], Weekday Sales [Right])

4.3 Long Short-Term Memory (LSTM)

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

Similar to the problem of predicting 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 to model the trend behavior using memory, they can be applied to current forecasting problem with minimum data preparation.

Model 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

Refer to Figure 3 showing end to end model architecture.

5 Experiments

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

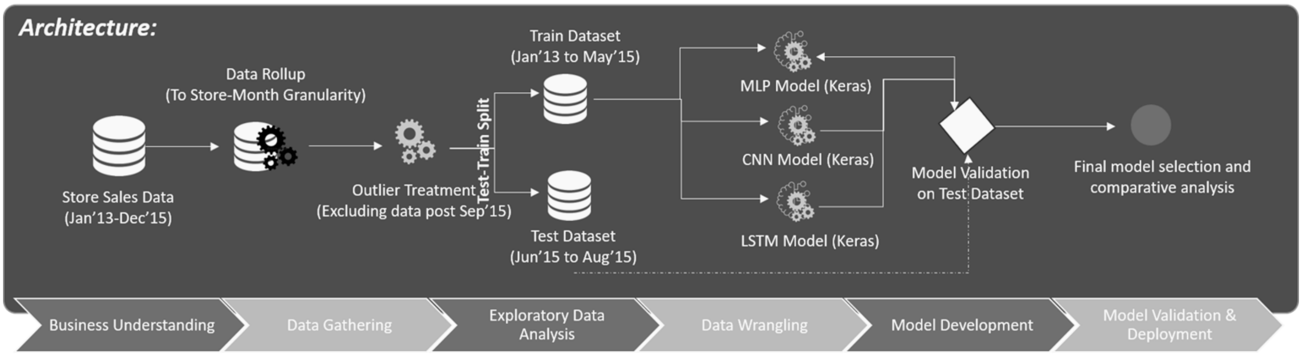

Figure 3: High Level Model Architecture

5.1 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 below for sample MLP model experiment results log

trainMAPE	trainMSE	testMAPE	testMSE	validMAPE	validMSE
26%	3,36,124	50%	3,23,323	51%	74,617
28%	3,97,434	48%	3,36,085	46%	74,006
28%	3,01,614	53%	3,23,323	56%	74,157
29%	4,26,577	45%	3,28,963	47%	75,733
30%	3,86,639	49%	3,21,858	53%	76,264
30%	4,10,014	45%	3,23,328	50%	75,806
31%	4,46,275	48%	3,37,963	48%	79,297
31%	3,53,668	50%	3,26,949	50%	73,248
31%	3,98,437	50%	3,40,225	53%	78,306
44%	4,61,647	60%	3,76,600	67%	83,362
46%	4,75,513	63%	4,08,059	71%	87,207
47%	5,22,063	56%	3,98,559	65%	86,903
63%	12,41,367	60%	5,93,638	61%	1,46,770

Table 3: MLP Model Experimentation Log (sample)

5.2 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 high test error
- CNN model showed moderate performance variation from in Train, Test and Validation datasets

Refer to Table below for sample CNN model experiment results log

trainMAPE	trainMSE	testMAPE	testMSE	validMAPE	validMSE
42%	4,15,195	54%	3,46,991	67%	83,665
66%	6,92,733	73%	4,02,233	89%	1,06,204
64%	7,55,990	62%	3,95,794	86%	1,09,565
59%	6,20,929	74%	4,19,152	85%	1,02,650
36%	4,24,550	52%	3,51,178	67%	86,388
30%	3,80,023	50%	3,50,350	64%	83,848
37%	3,36,140	64%	3,56,468	81%	95,092
39%	3,75,052	54%	3,54,781	71%	83,234

Table 4: CNN Model Experimentation Log (sample)

5.3 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
- CNN model showed high variation on Train vs Validation dataset. This can be a sign of moderate overfitting

Refer to Table 5 below for sample LSTM model experiment results log

trainMAPE	trainMSE	testMAPE	testMSE	validMAPE	validMSE
66%	8,63,664	86%	4,44,669	85%	1,16,532
68%	8,56,714	78%	4,31,523	89%	1,15,730
70%	8,68,453	74%	6,15,724	89%	1,16,945
70%	7,95,221	68%	6,20,039	86%	1,11,336
71%	8,86,312	88%	4,57,277	92%	1,19,576
72%	7,92,023	72%	6,36,024	90%	1,10,751
73%	8,02,793	75%	6,36,585	92%	1,13,814
76%	8,84,260	88%	4,52,108	96%	1,20,060
79%	9,09,531	82%	6,30,926	99%	1,21,989
81%	9,11,266	85%	6,54,229	101%	1,23,043
99%	18,20,039	99%	########	99%	2,15,190

Table 3: LSTM Model Experimentation Log (sample)

6 Results

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.**, **Error! Reference source not found.** and 8 shows MAPE across all the three

	Train	Validation	Test
MAPE:	29%	47%	45%
MSE:	426K	75K	328K

Table 4: MLP Model Best Iteration Results (Model stats reported at Store-Month level)

	Train	Validation	Test
MAPE:	36%	67%	52%
MSE:	424K	86K	351K

Table 5: CNN Model Best Iteration Results (Model stats reported at Store-Month level)

modeling approaches (for best experiment).

	Train	Validation	Test
MAPE:	70%	86%	68%
MSE:	795K	111K	620K

Table 6: 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 4: Train, Validation and Test Accuracy for various iterations of MLP, CNN and LSTM model respectively

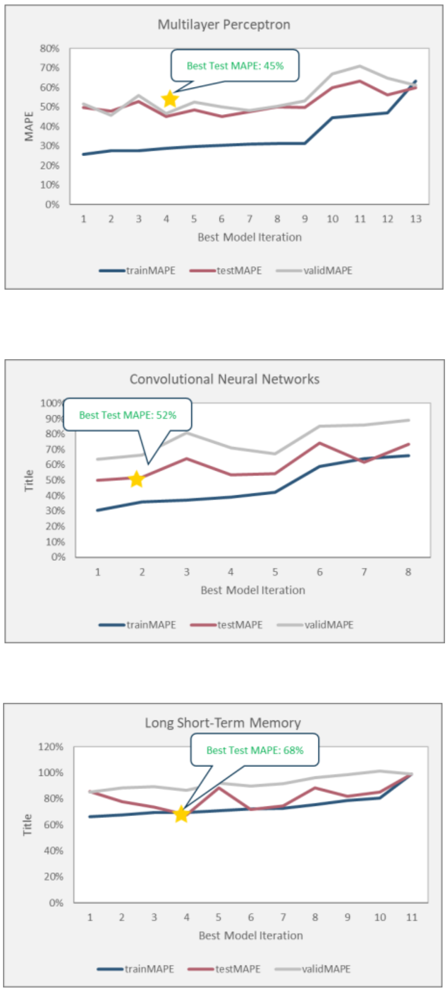

Figure 4: Train, Validation and Test Accuracy for various iterations of MLP, CNN and LSTM model respectively

7 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 very consistent in MLP experiments than that of CNN and LSTM. Though Convolutional showed good train MAPE of 36% but failed to perform on test dataset, while LSTM models failed to perform during both training and testing. One of the major reasons for all model 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

Pavlyshenko, B. M. 2019. *Machine-learning models for sales time series forecasting.Data,4(1), 15.*

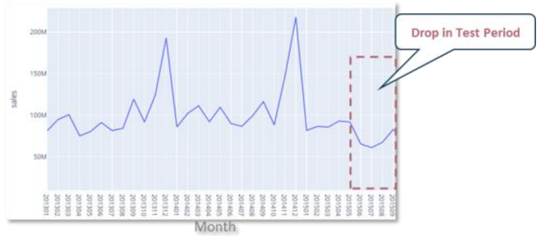

Figure 5: Monthly sales trend with abnormal drop in the during test period

8 Team Members Contribution

Vishnu Varakala: Responsible for leading the project and implementation of three algorithms MLP, LSTM, CNN

Sukesh Donthineni: Responsible for gathering the data for the project, preprocessing the data and reporting the insights from the data required for the project

9 References

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

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 Rep 10, 676. https://doi.org/10.1038/s41598-020-63566-8

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

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

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.

Purvika Bajaj, Renesa Ray. 2020. *SALES PREDICTION USING MACHINE LEARNING ALGORITHMS*. International Research Journal of Engineering and Technology (IRJET)

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)

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

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-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf

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

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

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