Original Article

A Novel Deep Belief Network Based Approach for Retail Store Sales Prediction During Peak Demand Seasons and its Performance Comparison over K-Nearest Neighbour Technique

B. Ruchitha[1], N. Deepa [2]

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602 105.
[2]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.
Email: [1]bruchitha18@saveetha.com, [2]ndeepa.sse@saveetha.com

Abstract

Aim: The research is about the Novel Deep Belief Network (NDBN) approach for Retail Store Sales Prediction during peak demand seasons and its performance comparison over K-Nearest Neighbour Technique (KNN). **Materials and Methods**: Deep Belief Network (N=10) and K-Nearest Neighbour algorithm (N=10) samples were considered based on the clinc calc online sample size calculator for predicting the accidents that happened in terms of accuracy. Two sample groups are taken into consideration and tested, G-power is the calculation that contains two different groups, alpha (0.05), power (80%), and environment ratio. **Results**: The Novel Deep Belief Network algorithm achieved 84.53% accuracy and K-Nearest Neighbour has 74.24%. This NDBN appears to have significance of p equal to 0.02 for the K-Nearest Neighbour, that is p less than 0.05 using independent sample T-test analysis. From the result, it proves that the Deep Belief Network approaches predict the retail sales store prediction.

Keywords: Machine Learning, Prediction, Retail Store, Sales, Novel Deep Belief Network, K-Nearest Neighbour.

DOI: 10.47750/pnr.2022.13.S04.218

INTRODUCTION

The sales trends of various businesses depend on the knowledge about their customers and demands of products according to the environmental conditions (Berg 2014). So, there is a purpose for predicting the people's needs after conducting various field work investigations, and to find a reliable deal pattern forecasting component, which is done through data mining. strategies to accomplish the most ideal income. This present business handles a tremendous vault of information (Cheriyan et al. 2018). The volume of information is relied upon to fill further in a remarkable way. The E-business industry is gravely in need of new information mining procedures and a shrewd expectation model of deals patterns with the most elevated conceivable degree of exactness and dependability (Maingi 2015). This study concerns the retail stores for sales forecasting and to produce outcomes that are reliable and easy to analyze. Deals define the workforce, income, and assets. (Cheriyan et al. 2018). It is a significant essential for big business arranging and dynamic. It permits organizations to design their business systems adequately.

Related to this study 53 papers were published in IEEE Xplore and 68 articles were published in Science Direct (Zhu et al. 2021). They Studied and trained the model with 4 different classifiers and chose the best model for execution (Gradient Boost, K-Nearest Neighbour) for predicting the retail sales store prediction search. For static systems, the K-Nearest Neighbour algorithm is best for sales prediction. According to my study, most researchers take accuracy into consideration. In much of the time, high accuracy value gives the best (Gentry, n.d.). In this paper, we predicted detailed forecast sales predictions (Lam, Vandenbosch, and Pearce 1998). Based on an exhibition assessment, a most appropriate prescient model is recommended for the business pattern figure. Stock forecasting has a significant influence on a company's success and profitability. Organizations face a few difficulties in regard to precise figures (Berg 2014). To be adequately capable and to create higher income,

business associations are continually upgrading basic information (Saylı, Ozturk, and Ustunel 2016). The business industry faces extreme difficulties to recognize an exact information mining procedure and powerful prediction methodology (Maingi 2015) because of the outstanding development of the colossal volume of information utilized in web-based business exchanges. (Kaneko and Yada 2016).

Our institution is passionate about high quality evidence based research and has excelled in various fields (Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020). The existing system faces a parcel of issues and major parts of deals capacities are recognizable proof of item property, value obsession, net deals acknowledgement, dispatch of new items and accuracy is not more than 75%. To overcome these issues, the proposed approach aims to use the different expectation techniques to deal with anticipating system and Expectation-Maximization (EM) calculations in this prediction system.

Materials And Methods

The experiment was set up in the Open Source lab, in the Computer Science Engineering department in Saveetha School of Engineering affiliated with Saveetha Institute of Medical and Technical Sciences, Chennai. Basically, it is considered that two groups of classifiers are used, namely Deep Belief Network and K-Nearest Neighbour algorithms, which are used to recommend the diet selection for diabetic patients. Group 1 is the Deep Belief Network algorithm with a sample size of 10 and the K-Nearest Neighbour algorithm is group 2 with a sample size of 10 for the comparison of performance. The sample size was evaluated using a clinical online calculator by keeping G power of 80% and threshold of 0.05%, confidence interval of 95%. The nutrition dataset was used in the study. This dataset was taken from the kaggle open-source website. The Deep Belief Network algorithm was chosen for implementation in this study, and it was compared to the K-Nearest Neighbour algorithm (Richards 2013).

Novel Deep Belief Network

The proposed Novel Deep Belief Network is trained using the analyzed visible feature units with the support of Contrastive Divergence (CD) algorithm. In the next stage, the learned features are mapped with the balanced activation weightage identified during the training using visible units. When the learning for the last hidden layer is completed, the Novel Deep Belief Network as a whole is trained.

Algorithm

1. Bring the Dataset
2. Explore the data and analyze the dataset how it looks
3. Preprocess of dataset
4. Organize the information into characteristics and labels.
5. Separate the data into two groups: training and testing.
6. Training the Novel Deep Belief Network
7. Make some recommendations

K-Nearest Neighbour

The K-Nearest Neighbour algorithm for calculating the distance between two points. Machine Learning's regression is an unsupervised learning approach. Based on the similarities with other recent cases, it forecasts the target. The distance metric is used to calculate the accuracy, with Distance measure being the most popular method. Predictions are made by searching the entire dataset for the K most comparable cases, i.e., the testing point's neighbours. Forecasting sales is a regression problem rather than a time series problem. As opposed to time series methods, practice demonstrates that using regression approaches can often produce better results. Patterns in time series can be discovered using machine learning methods.(Davies 1973). Using supervised machine learning approaches, identify complex patterns in sales dynamics. Tree-based machine learning algorithms are among the most used.

Algorithm

1. Importing the datasets to the algorithm
2. Explore the data and analyze the dataset how it looks
3. Pre-process the information
4. Organize the information into attributes and labels.
5. The data into training and testing should be divided
6. K-Nearest Neighbour datasets ore trained
7. Make some recommendations

The findings of the group statistics on all variables use the ensemble approach for recommending diet selection. The Novel Deep Belief Network algorithm acquires the best accuracy and standard deviation compared to K-Nearest Neighbour. As a result of these transformations, it obtains the best boundary between the viable outcomes of Machine Learning. Since the relevance of equality of variance, the probability value states that the results in the research effort are significant and correlated with each other, the table demonstrates the difference in the accuracy of both Deep Belief Network and K-Nearest Neighbour (Çoker et al. 2022).

Statistical Analysis

In this research, the obtained result values are recorded in Google spreadsheets, the collected data were fed into IBM SPSS V22 for performing statistical analysis.The independent variables are discount amount, discount rate, review volume, average ratings, number of pages, sentiment, and interaction effects and the dependent variables are sales price. The independent sample T test was performed between the values obtained from the iterations of the Independent Sample T-test was performed the Novel Deep Belief Network and K-Nearest Neighbour (Çoker et al. 2022).

Result

The accuracy comparison of the novel Deep Belief Network and K-Nearest Neighbour algorithm is taken for sample size N=10 per group. Because of its efficient classification feature based on the Deep Belief Network, the algorithm outperforms the K-Nearest Neighbour.

Table 1 shows the comparison of groups namely NDBN and K-Nearest Neighbour for the sample 10, it achieved 84.53% and 74.24 accuracies respectively.

Table shows the result of an independent sample T-test statistical comparison. The frequency of 3.32 at the time 1 the significance received as 0.023 for the alpha 0.05 and the confidence interval 95%.

Figure 1 represents the performance comparison of NDBN and K-Nearest Neighbour, where the graph was plotted accuracy in x axis and algorithms were plotted in y axis for the error bar mapped for the CI as 95 % and SD +/- 1.

Discussion

From the obtained independent sample t-test result, the significance was achieved was 0.023, where $p<0.05$. This evidence shows the existence of significance among the groups for this experimental analysis. The group's analysis derives the mean accuracy of 84.53% for NDBN and 74.24% for K-Nearest Neighbour, from the analysis Novel Deep Belief Network performs better than K-Nearest Neighbour.

Recognized paintings withinside the area of sale forecasting were executed through (Castillo et al. 2017), they did income predicted on newly posted articles in a piece of writing about business control surroundings through making use of computational methodologies. The business personal income forecast also predicted using Artificial Neural Networks (ANN) (Chang and Wang 2006). The prediction performance was enhanced using Fuzzy Neural Networks (FNN), additionally optimization was performed using Radial Basis Function Neural Network (RBFN) (Majhi, Panda, and Sahoo 2009) (Arunraj and Ahrens 2015). The literature on this subject indicates that now no longer a lot of paintings have been executed for effective prediction models using swarm intelligence approach (Majhi et al. 2008). The schooling of ANN models was deliberated using the ML algorithms such as Genetic Algorithm (GA), which leverage the capabilities of the algorithm. There are loads of works on this subject which has helped the agency to expect the destiny income that they could make via means of making an investment in the right vicinity at the proper time (Majhi et al. 2009). The researchers have determined the prediction framework for sales to tackle the need for business associations for the deals with large volumes of information. The business decision taking demands are influenced by the accuracy of prediction in short duration (Andersen-Rodgers et al. 2021).

The proposed system outreached with its best performance, even though it has some limitations when it comes to inflation in the customer's country the prediction is not accurate in such cases due to fluctuation in market pricing. In future works, this can be incorporated with the price prediction along with economic money supply and interest rate data in the dataset.

Conclusion

The experimentation using Novel Deep Belief Network and K-Nearest Neighbour was performed for retail store sales prediction. The results are compared using various statistical analyses. Finally, it is concluded that NDBN (84.53%) accuracy is more than K-Nearest Neighbour (74.24%) and performs better.

DECLARATION

Conflicts of Interest

No conflict of interest in this manuscript.

Authors' Contributions

Author BR was involved in data collection, data analysis, and manuscript writing. Author JK was involved in the conceptualization, guidance, and critical review of the manuscript.

Acknowledgments

The authors would like to express their gratitude to Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Reliance Industries Ltd.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

References

1. Andersen-Rodgers, Elizabeth, Xueying Zhang, Tam D. Vuong, Liz Hendrix, Cheryl Edora, Rebecca J. Williams, Lauren Groves, et al. 2021. "Are California's Local Flavored Tobacco Sales Restrictions Effective in Reducing the Retail Availability of Flavored Tobacco Products? A Multicomponent Evaluation." *Evaluation Review* 45 (3-4): 134–65.
2. Arunraj, Nari Sivanandam, and Diane Ahrens. 2015. "A Hybrid Seasonal Autoregressive Integrated Moving Average and Quantile Regression for Daily Food Sales Forecasting." *International Journal of Production Economics*. https://doi.org/10.1016/j.ijpe.2015.09.039.
3. Berg, Bettina. 2014. "Study 2: Retail Branding and Local Competition: The Importance of Retail Brand Equity and Store Accessibility for Store Loyalty in Local Competition." *Retail Branding and Store Loyalty*. https://doi.org/10.1007/978-3-658-01596-1_3.
4. Castillo, Pedro A., Antonio M. Mora, Hossam Faris, J. J. Merelo, Pablo García-Sánchez, Antonio J. Fernández-Ares, Paloma De las Cuevas, and María I. García-Arenas. 2017. "Applying Computational Intelligence Methods for Predicting the Sales of Newly Published Books in a Real Editorial Business Management Environment." *Knowledge-Based Systems*. https://doi.org/10.1016/j.knosys.2016.10.019.
5. Chang, P., and Y. Wang. 2006. "Fuzzy Delphi and Back-Propagation Model for Sales Forecasting in PCB Industry." *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2005.07.031.
6. Cheriyan, Sunitha, Shaniba Ibrahim, Saju Mohanan, and Susan Treesa. 2018. "Intelligent Sales Prediction Using Machine Learning Techniques." *2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE)*. https://doi.org/10.1109/iccecome.2018.8659115.
7. Çoker, Elif Naz, Rachel Pechey, Kerstin Frie, Susan A. Jebb, Cristina Stewart, Suzanne Higgs, and Brian Cook. 2022. "A Dynamic Social Norm Messaging Intervention to Reduce Meat Consumption: A Randomized Cross-over Trial in Retail Store Restaurants." *Appetite* 169 (February): 105824.
8. Davies, R. L. 1973. "Evaluation of Retail Store Attributes and Sales Performance." *European Journal of Marketing*. https://doi.org/10.1108/eum0000000005104.
9. Devarajan, Yuvarajan, Beemkumar Nagappan, Gautam Choubey, Suresh Vellaiyan, and Kulmani Mehar. 2021. "Renewable Pathway and Twin Fueling Approach on Ignition Analysis of a Dual-Fuelled Compression Ignition Engine." *Energy & Fuels: An American Chemical Society Journal* 35 (12): 9930–36.
10. Dhanraj, Ganapathy, and Shanmugam Rajeshkumar. 2021. "Anticariogenic Effect of Selenium Nanoparticles Synthesized Using Brassica Oleracea." *Journal of Nanomaterials* 2021 (July). https://doi.org/10.1155/2021/8115585.
11. Gentry, Matthew L. n.d. "Displays, Sales, and In-Store Search in Retail Markets." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2050620.
12. Kamath, S. Manjunath, K. Sridhar, D. Jaison, V. Gopinath, B. K. Mohamed Ibrahim, Nilkantha Gupta, A. Sundaram, P. Sivaperumal, S. Padmapriya, and S. Shantanu Patil. 2020. "Fabrication of Tri-Layered Electrospun Polycaprolactone Mats with Improved Sustained Drug Release Profile." *Scientific Reports* 10 (1): 18179.
13. Kaneko, Yuta, and Katsutoshi Yada. 2016. "A Deep Learning Approach for the Prediction of Retail Store Sales." *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. https://doi.org/10.1109/icdmw.2016.0082.
14. Lam, Shunyin, Mark Vandenbosch, and Michael Pearce. 1998. "Retail Sales Force Scheduling Based on Store Traffic Forecasting." *Journal of Retailing*. https://doi.org/10.1016/s0022-4359(99)80088-8.
15. Maingi, Mathew Ngwae. 2015. "A Survey on the Clustering Algorithms in Sales Data Mining." *International Journal of Computer Applications Technology and Research*. https://doi.org/10.7753/ijcatr0402.1009.
16. Majhi, Ritanjali, Ganapati Panda, Gadadhar Sahoo, and Abhishek Panda. 2008. "On the Development of Improved Adaptive Models for Efficient Prediction of Stock Indices Using Clonal-PSO (CPSO) and PSO Techniques." *International Journal of Business*

Forecasting and Marketing Intelligence. https://doi.org/10.1504/ijbfmi.2008.020814.
17. Majhi, Ritanjali, G. Panda, Babita Majhi, and G. Sahoo. 2009. "Efficient Prediction of Stock Market Indices Using Adaptive Bacterial Foraging Optimization (ABFO) and BFO Based Techniques." *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2009.01.012.
18. Majhi, Ritanjali, G. Panda, and G. Sahoo. 2009. "Development and Performance Evaluation of FLANN Based Model for Forecasting of Stock Markets." *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2008.08.008.
19. Nandhini, Joseph T., Devaraj Ezhilarasan, and Shanmugam Rajeshkumar. 2020. "An Ecofriendly Synthesized Gold Nanoparticles Induces Cytotoxicity via Apoptosis in HepG2 Cells." *Environmental Toxicology*, August. https://doi.org/10.1002/tox.23007.
20. Parakh, Mayank K., Shriraam Ulaganambi, Nisha Ashifa, Reshma Premkumar, and Amit L. Jain. 2020. "Oral Potentially Malignant Disorders: Clinical Diagnosis and Current Screening Aids: A Narrative Review." *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation* 29 (1): 65–72.
21. Perumal, Karthikeyan, Joseph Antony, and Subagunasekar Muthuramalingam. 2021. "Heavy Metal Pollutants and Their Spatial Distribution in Surface Sediments from Thondi Coast, Palk Bay, South India." *Environmental Sciences Europe* 33 (1). https://doi.org/10.1186/s12302-021-00501-2.
22. Pham, Quoc Hoa, Supat Chupradit, Gunawan Widjaja, Muataz S. Alhassan, Rustem Magizov, Yasser Fakri Mustafa, Aravindhan Surendar, Amirzhan Kassenov, Zeinab Arzehgar, and Wanich Suksatan. 2021. "The Effects of Ni or Nb Additions on the Relaxation Behavior of Zr55Cu35Al10 Metallic Glass." *Materials Today Communications* 29 (December): 102909.
23. Richards, Romeo. 2013. *How to Increase Retail Sales with Store Design and Visual Merchandise Display*. CreateSpace.
24. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). https://doi.org/10.1002/ep.13696.
25. Saylı, Ayla, Isil Ozturk, and Merve Ustunel. 2016. "Brand Loyalty Analysis System Using K-Means Algorithm." *Journal of Engineering Technology and Applied Sciences*. https://doi.org/10.30931/jetas.287786.
26. Tesfaye Jule, Leta, Krishnaraj Ramaswamy, Nagaraj Nagaprasad, Vigneshwaran Shanmugam, and Venkataraman Vignesh. 2021. "Design and Analysis of Serial Drilled Hole in Composite Material." *Materials Today: Proceedings* 45 (January): 5759–63.
27. Uganya, G., Radhika, and N. Vijayaraj. 2021. "A Survey on Internet of Things: Applications, Recent Issues, Attacks, and Security Mechanisms." *Journal of Circuits Systems and Computers* 30 (05): 2130006.
28. Zhu, Siyao, Cassandra Mitsinikos, Lisa Poirier, Takeru Igusa, and Joel Gittelsohn. 2021. "Development of a System Dynamics Model to Guide Retail Food Store Policies in Baltimore City." *Nutrients* 13 (9). https://doi.org/10.3390/nu13093055.

TABLES & FIGURES

Table 1. Statistical analysis of NDBN and novel K-Nearest Neighbour. Mean data transmission speed value, Standard deviation and Standard Error Mean for NDBN and novel K-Nearest Neighbour algorithms. It is observed that the K-Nearest Neighbour algorithm has a better data transmission speed than the NDBN algorithm.

	Algorithm	N	Mean	Std.Deviation	Std.Error Mean
Accuracy	NDBN	10	84.5310	.75250	.23796
	K-Nearest Neighbour	10	74.2490	.36638	.11586

Table 2. Independent sample T-test for significance and standard error determination. P-value is less than 0.05 considered to be statistically significant and 95% confidence intervals were calculated.

Leven's Test for Equality of Variances				**t-test for Equality of Mean**					**95% Confidence Interval of the Difference**	
		F	Sig	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper

Accuracy	Equal Variance assumed	3.323	.023	18	38	000	10.28200	.26467	9.72596	10.83804
	Equal Variance not assumed			18	37.437	000	10.28200	.26467	9.71040	10.85360

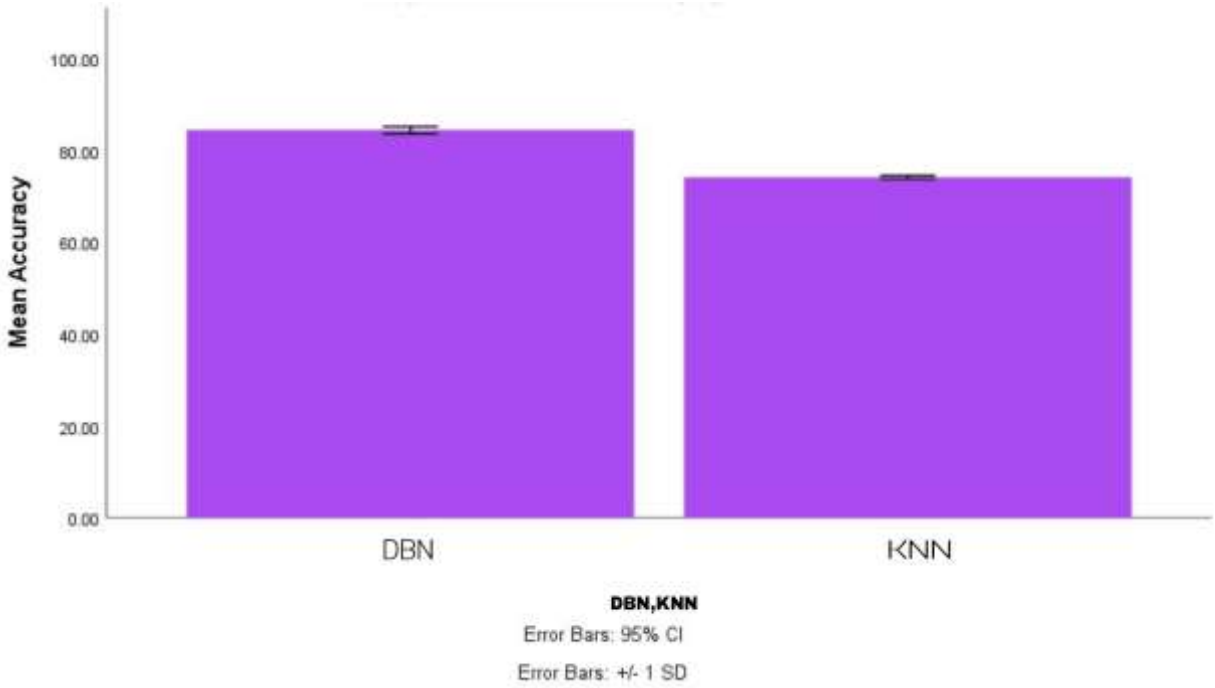

Fig. 1. Comparison of K-Nearest Neighbour algorithm and NDBN in terms of mean data transmission speed. The mean accuracy given in X-axis for NDBN Vs KNN and Y-axis is error bars +/- 1 SD.