A Cognitive Analytics Framework for Improving Sales Prediction Accuracy

Hanan Wagih Elgezery
School of MOT
Nile University
Cairo, Egypt
h.wagih@nu.edu.eg

Mohamed Mamdouh Awny
School of MOT
Nile University
Cairo, Egypt
mmawny@nu.edu.eg

2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES) | 979-8-3503-8103-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/NILES59815.2023.10296687

***Abstract*— accurate prediction of sales data is an important and challenging problem in the retail industry. However, due to the diversity of data types, it is difficult for traditional time series prediction models to generate good prediction results for different types of data. This research develops a cognitive analytic framework to help transform retailers from the reactive environment to automated continuously learning environment utilizing the power of AI technology in predicting future sales with high accuracy. It proposes a multi-stage prediction framework based on the integration of Statistical Forecasting, Machine Learning Forecasting (ML), and Deep Learning Forecasting (DL) algorithms. The results of the statistical forecasting stage are taken as a baseline to the others consecutive two stages. This approach uses different forecast models based on different forecast mechanisms. It takes the error of each model as an input feature to the following model. The research shows that this process of creating forecasts using a proposed multi-stage framework that combines the error of each forecast stage as an input feature to the next stage produces better accuracy than the output of any single stage alone as well as when not using the error. This research work presents an advanced approach for more accurate prediction of the sales' volume during a defined future period. Accuracy reaches as high as 90% with 82% as an average accuracy._The generalization of the proposed framework outperforms the results in both the category and the product levels of forecasting. This is much preferred in E-commerce platforms where there is a high need to generate product-level operational demand forecasts and when there is no historical data for a long time period to capture the seasonality of a target variable.**

Keywords: AI; Machine Learning; Deep Learning; Cognitive Analytics; Predictive Analytics.

I. INTRODUCTION

Worldwide, there is currently more information generated in a single day than anyone could possibly absorb in an entire lifetime. These information help decision makers becoming more adept at predicting the future. The retail business is a data-intensive industry which would gain high benefit from the advanced data analytics technology. The automated data collections systems, using sensors' Radio Frequency Identification (RFID), have significantly increased the quantity and quality of retail data. Coupled with consumer generated content, retail industry now can have more data than ever [1]. This research identifies key emerging technologies that are essential to get benefit from this huge and complex data in transforming the retail industry. These technologies include Artificial Intelligence (AI), ML, and DL technologies. According to Bill Shorten [2] "AI is the engine that is poised to drive the future of retail to all-new destinations." It is expected that AI will contribute up to 45 percent of all economic gains by 2030 [3].

AI is an aggregate term for computer systems that can detect some current environment, think, learn, and make a move considering what has been detected and according to a set of targets. Big Data is presently the most frequently discussed and used technology in relation to AI technologies. It deals with a collection of data with complexity, diversity, heterogeneity, and high potential value that are difficult to process and analyze with traditional tools in reasonable time [4]. AI enables the extraction of meanings and trends from Big Data to solve problems, improve products and increase productivity [3]. Data analytics is the practice of extracting information from existing data sets in order to determine patterns and predict future outcomes and trends. The role of analytics has changed significantly over the past 30 years. Data scientists are experiencing a progression in analytics maturity levels ranging from descriptive analytics which does not predict, to diagnostic analytics, to predictive analytics, to prescriptive analytics, up to machine learning and cognitive computing. Data analytics process includes a variety of statistical techniques, data mining, machine learning, and deep learning algorithms [5]. Statistics is the science of analyzing ~~the~~ data and the statistical modeling is focused primarily on making inferences and understanding the characteristics of the variables. In a statistical model, a hypothesis is a testable way to confirm the validity of the specific algorithm. Indeed, machine learning models leverage statistical algorithms and apply them to activate analytics [6]. Data mining is intended to discover patterns that can be used by humans and not intended to make predictions to back up hypotheses. In contrast, machine learning automates the process of identifying patterns that are used to make predictions [6]. In fact, ML is a continuation of the concepts around predictive analytics with one key difference which is its ability to make assumptions, test, and learn autonomously. The advances in computer technology such as central processing units (CPUs), graphics processing units (GPUs), and the very large in-memory capacity combined with the automated data collections systems, make it possible to process very large and complex data in real time. Yet, ~~the~~ statistical forecasting methods cannot face the challenges of dealing with large and complex data.

Thus, ML is becoming an essential part for creating analytics and predictive models. It uses a variety of algorithms that iteratively learn from data to improve models, describe data, discover patterns, and predict outcomes. ML algorithms continuously refine the models by continuously processing new data in near real-time and train the system to adapt to changing patterns and relationships in data.

In the absence of AI technologies, the sales forecasting models may ignore important factors such as promotional waves, market cannibalization, and seasonal changes which are too complex for traditional forecasting tools to recognize. Egyptian market is a price sensitive market, and the Egyptian retail environment is a highly dynamic because of the high competition. The existing demand forecasting methods are traditional techniques that treat every transaction as an isolated event and relies on the guesswork of human sales experts. They are not able to handle efficiently the retail data characteristic, which is highly non-stationary with irregular sales patterns, sparse sales data, and highly intermittent sales. Non-stationary data are unpredictable and cannot be modeled or forecasted by using traditional forecasting techniques. As a result, demand forecasting in Egypt carries various challenges and may give non-accurate results utilizing the presently used techniques. Therefore, the use of the state-of-the-art of artificial intelligence and real-time predictive techniques are highly recommended to help retailers moving to a more effective demand-driven supply chain and respond quickly to changing demand.

This research focuses on the influence of two emerging technologies in the retail industry, one is machine learning that can test and learn autonomously and the other is deep learning that can get the most accurate sales demand forecasting. It develops a multi-stage forecasting framework. The main features of the proposed framework are the integration of a two-step approach (where statistical models are used in conjunction with machine learning models) with the deep learning networks that produce forecast results with the least error. The model overcomes the obstacles of the complex demand sales which includes critical factors such as promotional waves, advertise cannibalization and occasional changes. In this paper, the experimental setup is demonstrated the results obtained by applying the proposed framework to a dataset collected from one of the biggest stores in Egypt. The research work proves how this approach achieves competitive results on demand sales' category and product levels datasets by using the good opportunity to utilize the correlated and similar sales patterns available in a product portfolio of any category of products.

This research does not go into the details of the AI and ML infrastructure from a communications and networking perspectives but focuses on the use and management side of the subject. Finally, this research opened some more research points which can be addressed in the future.

II. PROPOSED COGNITIVE ANALYTICS FRAMEWORK

The proposed framework presents a short to medium-term demand forecasting (that is carried out for period of 3 to 24 months) with different and integrated predictive analytics techniques. This proposed framework is designed considering the sales demand to predict on category level then generalized the model to predict on the product level. It introduces new functionalities to understand what drives sales and leverage patterns in sales data to generate useful features for better demand forecasting. It can combine static and dynamic (weather) features to extract high accuracy sales prediction. Where sales of products can be greatly influenced by many factors such as holidays, events, campaigns, competitors, and weather. The main target of this framework is to achieve the highest accuracy of sales' demands prediction by using the multi-stage forecasting approach which is integrating the deep learning techniques with the conventional two-step forecasting approach.

The proposed framework has the unique advantage of reaching the least error in the prediction process quickly, since the deep learning technique utilizes the Artificial Neural Networks to extract higher-level features from datasets in a progressive mode. The methodology for applying this proposed cognitive analytics framework is as follows:

- In the "multi-stage" approach, the prediction of sales demands will be carried out in three consecutive integrated stages (figure 1). In each stage a different forecast algorithms type is applied.

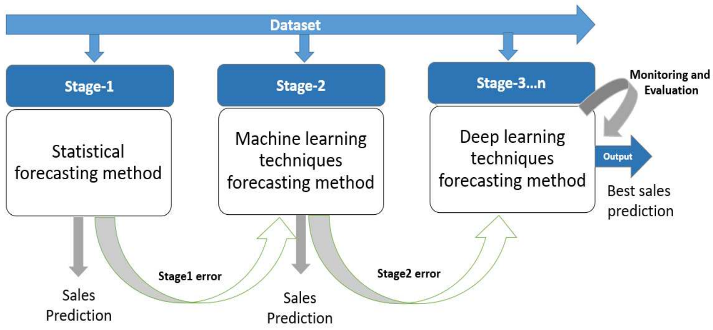

Figure 1: The Proposed Multi-Stage Cognitive Analytics Approach

- The third stage will result in the highest sales prediction accuracy. Hence this stage will be used for the prediction of sales demand for the required time.
- The error value of each stage is used as feedback and input with the dataset to the next forecasting stage. This method of preserving a seasonal pattern that is present in the errors' history while projecting the effect of both trend and seasonality for a product sales' prediction into the next period, minimized the final forecast error.
- The sales historical transactions (datasets) to train and test the proposed framework should be of two historical years minimum. The first year of sales data will be taken as a training dataset and the sales data of the other year(s) will be used as a test dataset. Models will be developed using the training dataset and will make predictions that will be compared with the test dataset.
- The forecast algorithms type to be used in the different stages are the statistical algorithms, supervised machine learning algorithms, and supervised deep learning algorithms.
- Six forecasting techniques and algorithms are applied in the three stages. A comparison is to be made between each two forecasting algorithms in every stage to select the algorithm that produces the least error (high accuracy).
- The selection of the above forecasting techniques depends on many factors such as the problem of the forecast, the relevance and availability of historical data, the degree of accuracy desired, and the period to be forecast.
- Accuracy, Root Mean Squared Error (RMSE), Mean Absolute percentage Error (MAPE), Mean Absolute Error (MAE), and Bias are the performance evaluation metrics used in this research to evaluate the different forecast models and do a comparison between each forecast technique and the baseline technique.

A. The Design Of The Framework

The research case is a sales demand forecast problem where data is ~~a~~ linear regression dataset, the input and output datasets are known, and the target is to obtain an accurate forecasting for the volume of sales with no overfitting or under-fitting. The proposed cognitive analytic framework for the retail sector consists of five main components: data collection, statistical forecasting, data preparation for advanced analytics, ML forecasting, and DL forecasting phases as shown in figure 2.

Phase1: Data Collection, this phase consists of data gathering, data cleansing, and data primary analysis steps to understand, the main input features of the datasets, discover pattern and trend in sales data, and analyze the relationship between sales volume and other inputs features of the datasets. Different data checks procedures such as accuracy, consistency, completeness, and integrity are done in this phase to assure the data quality and to be fit for use.

Phase 2: The Statistical Forecasting, the exponential smoothing and the exponential smoothing with seasonality forecasting algorithms are used in this phase. Each model is tuned by changing its hyper-parameter to achieve the forecast with the highest accuracy. The exponential smoothing with seasonality forecast achieved the best forecast accuracy in this phase.

Phase 3: Data Preparation for Advanced Analytics, this phase contains many processes to prepare datasets for advanced analytics (Machine Learning and Deep Learning forecasting techniques) such as feature selection, determine dependent and independent features, dealing with categorical values, dealing with outliers' values, transform data into vectors, and spilt correctly the dataset into test and train datasets.

Phase 4: The ML Forecasting, in this phase, the proposed framework performance will be enhanced using machine learning powered by artificial intelligence, to achieve the following benefits:

1) Extensive data analytics. 2) Generating a lot of useful features from the sales data. 3) Identifying patterns that humans and traditional forecasting tools could have missed.

The following are the steps used in this phase as figure 2 shows.

- The Random Forest (RF) and XGBoost algorithms will be used and compared to select the algorithm with the least error.
- In this phase, the forecast error of the statistical forecasting phase will be used as an input feature.
- Tuning each model by changing its hyper-parameters is a crucial step to get the best performance of the model. As usual, one can start with default values and change them until results are satisfied.
- Once the model and tuning parameter values have been defined, the type of resampling should also be specified (The bootstrap and *k*-fold cross-validation are the applied resampling type and method used in this research).

In this phase, the RF regression algorithm proved to give the highest accuracy of results compared to the XGBoost forecast algorithm. RF is a predictive modeling tool with an ensemble learning method for regression. It eliminates the decision trees' habit of overfitting to the training dataset. This is why it is selected.

Phase 5: The DL Forecasting, in this phase an artificial neural network (LSTM) is to be extensively trained by executing thousands of neural networks iterations. Defining the products' sales demand relationships available in each product and combining historical (static) and real-time (dynamic) data. Thus, achieving more accurate results.

Phase 5 in figure 2 shows the steps that will be followed.

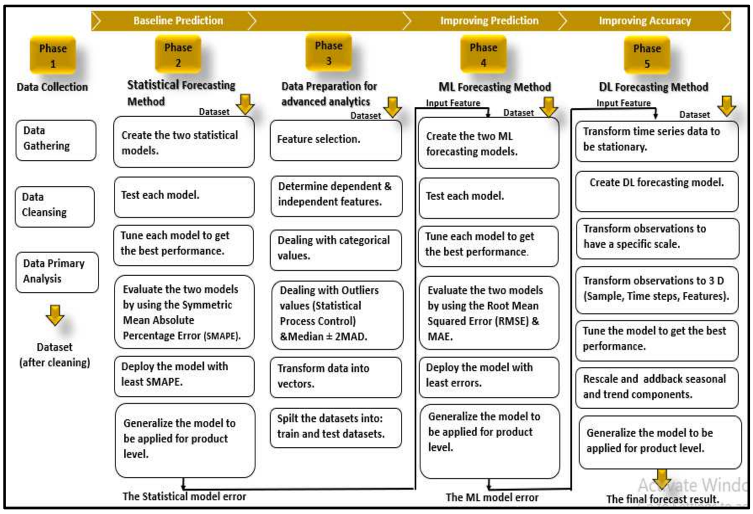

Figure 2: The cognitive analytics framework

1. Since, the neural network models cannot handle the non-stationary data. So, before creating the DL model, the sales demand data will be analyzed to configure its components by using the time series decomposition function (splitting the time series into level, trend, seasonality, and noise components).
2. The SARIMA model is used to remove the seasonal and trend components from the time series datasets. Differencing method used to stabilize the mean of the time series by removing changes at the level of a time series and eliminating trend and seasonality. The seasonal ARIMA model uses differencing at a lag equal to the number of seasons (s) to remove additive seasonal effects. Using lag equal 2 differencing will help to remove the trend and seasonality.
3. LSTM networks have the capabilities of learning long-term dependencies, especially in sequence prediction and autocorrelation problems. They are well-suited to making predictions based on time series data (sales demand).
4. A separate LSTM model for each category of products will be created. In this process, the time series grouping strategy will improve the LSTM performance.
5. The activation function to be used is the default one (tanh) and the loss function is the MAE (Mean Absolute Error) which is the sum of absolute differences between actual and predicted variables.
6. The optimizer that is used for training the LSTM model is Adam optimizer.
7. Tuning the LSTM model by changing its hyper-parameters is an important step to get the best accurate forecasting results.

B. The Framework Results

The validation of the proposed framework was done by implementing the framework design to a chosen mega retail store in Egypt aiming at properly manage its sales prediction in both levels: category and product. This means that forecasting models will be built considering category sales forecast then they are generalized to predict product sales as shown in the experiments' results. The framework is built using two consequent years (2015 and 2016). Table 1 shows the Key Performance Indicators (KPIs) of the Statistical, two-step approach and multi-step approach. It is observed that, the two-step approach which uses the machine learning "Random Forest" model remarkably improves all the KPIs over the baseline model (statistical method). The multi-step approach which used the deep learning LSTM model performs even better accuracy (lower MAE) than the two-step approach. Therefore, the multi-step approach's accuracy of all categories was improved.

Table 1: KPIs of the Statistical, Two-Step and Multi-Step Approach

	Statistical Forecasting Model Results					Two-Step Forecasting Approach Results					Multi-Step Forecasting Approach Results				
Category	Error RMSE	Scaled RMSE	Error (MAE)	Scaled MAE	Accuracy	Error RMSE	Scaled RMSE	Error MAE	Scaled MAE	Accuracy	Error RMSE	Scaled RMSE	Error MAE	Scaled MAE	Accuracy
Fast-Moving Consumer Goods	1,960,026	0.923	1,074,283	0.506	49%	965,237	0.488	707,832	0.358	64%	947,011	0.471	697,308	0.347	65%
Fresh food	22,407	0.186	15,064	0.126	87%	13,776	0.137	11,963	0.125	88%	16,888	0.157	11,765	0.109	90%
Non-Food	5,212	0.155	4,177	0.127	87%	4,138	0.132	3,674	0.117	88%	4,727	0.142	3,774	0.113	90%
Technology	9,803	0.422	6,301	0.271	73%	6,018	0.236	4,900	0.192	81%	4,798	0.220	3,823	0.175	83%
Average	499,362	0.422	274,956	0.258	74%	247,292	0.248	182,092	0.198	80%	243,356	0.248	179,168	0.186	82%

The highest accuracy rate of the proposed framework is 90%. It had an average accuracy value of 82% (over the base line model) which was the highest number compared to the ML forecast average accuracy (80%) and Statistical forecast average accuracy (74%). Finally, by giving indicating points on the achieved results, the following were achieved:

- The time sensitive "Fresh Food" and the static "Non-Food" categories achieved high accuracy of 90%.
- The proposed framework has improved the forecasting accuracy by more than 11% compared to the baseline models (statistical) and reduced the average forecasting error by more than 28%.
- According to our comprehensive experiments (6 algorithms and 24 models), the DL model can produce better outcomes, since one gets benefits from its recurrent architecture and optimal tuning.
- This multi-stage predictive framework proposed by the researcher pointed out that combination methods produced more accurate forecasts than a single method.

III. CONCLUSION

The sales demand's forecasting capability using the classical methods is insufficient to estimate future values because of the inefficiency in pattern recognition process. Shorter product life cycles and aggressive marketing, especially for the non-stationary data characteristic's factors have increased the complexity of sales forecasting. The proposed multi-stage forecasting approach which uses Machine Learning (ML) and Deep Learning (DL) networks produces better results using Neural Pattern Recognition. Thus, the proposed framework allows the identification of the hidden patterns in data, reflects the correlation between inputs, and provides a more accurate forecast in both the product and the category levels. It improves the prediction accuracy in different types of categories, those having time sensitive characteristics such as the Fresh Food category and those having static characteristics such as the Non-Food category. This shows that the framework proposed in this research has a greater advantage regarding the prediction accuracy, when implementing it to real data and reached an average accuracy of 82% with 90% as its highest accuracy.

A. Suggestion For Future Research

The work carried out in this research opened some more research points which can be addressed in the future, such as:

1) Using more historical data to run the experiments and spanning around an extended period of time.
2) The proposed framework could give better prediction in the Fast-Moving Consumer Goods category, if it breaks down into sub-categories.
3) Finally, this framework can be upgrading to real-time cognitive analytics by using the stream data management and the "Edge Analytics" to automatically collect data and analysis in an automated analytical computation at a network switch or devices instead of waiting for the data to be collected from the centralized data store.

REFERENCES

[1] Weng Chun Tan, Manjit Singh Sidhu, 2022. Review of RFID and IoT integration in supply chain management. Retrieved from https://reader.elsevier.com/reader/sd/pii/S2214716022000070?token=018AC4C77F5923B5D5CF237BD83C81195AA0DBF05FBDE4DEE73AEE0261E1784AAA69F23A1ACE738E75E84CAB53F93D9A&originRegion=eu-west-1&originCreation=20230113203636

[2] McKinsey Global Institute. AI THE NEXT DIGITAL FRONTIER, 2017. Retrieved from https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx

[3] PwC. Sizing the prize, 2020. Retrieved from https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf

[4] Prannoiy Chandran, 2018. Disruption in Retail: AI, Machine Learning & Big Data. Towards Data Science, Retrieved from https://towardsdatascience.com/disruption-in-retail-ai-machine-learning-big-data-7e9687f69b8f

[5] Hanan W. Elgezery, Mohamed M. Awny. AI for Retail industry in Egypt Challenges and Opportunities. Published in the 29th Annual Conference of the International Association for Management of Technology (IAMOT 2020)

[6] Judith Hurwitz and Daniel Kirsch. Machine Learning for Dummies. IBM Limited Edition. IMM14209USEN, 2018. Retrieved from https://www.ibm.com/downloads/cas/GB8ZMQZ3