iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38069

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Food Sales Forecasting Using Machine Learning Techniques: A Survey

Aaron Rodrigues[1], Cassandra Rodrigues[2], Royce Dcunha[3], Kevin Dsouza[4]
[1, 2, 3]*Department of Computer, St. Francis Institute of Technology, Mumbai, India*
[4]*Department of Electronics and Telecommunication Engineering, St. Francis Institute of Technology, Mumbai, India*

Abstract: Food sales forecasting is concerned with predicting future sales of food-related businesses such as supermarkets, grocery stores, restaurants, bakeries, and patisseries. Companies can reduce stocked and expired products within stores while also avoiding missing revenues by using accurate short-term sales forecasting. This research examines current machine learning algorithms for predicting food purchases. It goes over key design considerations for a data analyst working on food sales forecasting's, such as the temporal granularity of sales data, the input variables to employ for forecasting sales, and the representation of the sales output variable. It also examines machine learning algorithms that have been used to anticipate food sales and the proper metrics for assessing their performance. Finally, it goes over the major problems and prospects for applied machine learning in the field of food sales forecasting.
Keywords: Food, Demand forecasting, Machine learning, Regression, Timeseries forecasting, Sales prediction

I. INTRODUCTION

The precise and timely assessment of future sales, also known as sales forecasting, can provide essential knowledge to organizations involved in the manufacturing, wholesale, or retail of products in today's highly competitive and constantly changing business environment. Short-term forecasts are mostly useful for production planning and stock management; however, long-term forecasts can assist in corporate development decision-making (Doganis et al. 2006).

Due to the limited shelf-life of many items in the food sector, sales forecasting is especially crucial, as it leads to revenue loss in both shortage and surplus circumstances. Ordering too many things results in waste while ordering too few results in missed opportunities. Furthermore, due to factors such as pricing, promotions, changing consumer preferences, and weather variations (Van der Vorst et al. 1998), food consumer demand is continuously fluctuating. Managers generally forecast sales based on their whims. Skilled managers, on the other hand, are hard to come by and aren't constantly available. As a result, sales forecasting should be aided by computer systems that can fill in for a trained manager when she is unavailable and/or assist her in making the best decision possible by providing projections of future sales. One method to create such a system is to try to replicate the expert knowledge of experienced managers in a computer network. Alternately, machine learning techniques may be used to automatically create reliable sales forecast models using the abundance of sales data and associated information. The latter is a simpler procedure that is not influenced by the unique characteristics of a single sales manager and therefore is dynamic, meaning it can react to changes in the data. Moreover, it can outperform a human expert's forecast accuracy, which is generally flawed.

This paper looks at recent work on using machine learning to predict food sales, which is important for a range of companies such as supermarkets, restaurants, bakeries, and patisseries. Section 2 looks at how sales forecasting may be viewed as a machine learning task and presents learning techniques that have been used to estimate food sales previously. The final segment discusses how to evaluate sales forecasting systems. Finally, Section 4 concludes our research by outlining many interesting research avenues.

II. A MACHINE LEARNING TASK: FOOD SALES FORECASTING

Predicting food sales is a time series forecasting job. This issue may be tackled using traditional statistical approaches such as autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA). When dealing with time series forecasting, however, a machine learning method is typically more powerful and adaptable. It's powerful because it enables the employment of cutting-edge supervised learning techniques like regression support vector machines and model trees (Landwehr et al. 2005) (decision trees with linear regression functions at the leaves). It's versatile since it enables the addition of extra valuable input variables that aren't related to the time series in question.

The temporal granularity of sales data, the input factors to utilize for forecasting sales, and the specific form of the sales output variable are all discussed in this section. Finally, it goes through learning approaches that have been used to anticipate food purchases.

A. Temporal Granularity

Daily sales predictions are required in the case of short shelf-life products, like milk (Doganis et al. 2006) and fresh food (e.g. sandwiches and rolls) Chen et al. (2010), which are produced daily. For long shelf-life products, weekly sales predictions are adequate for stock management (Žliobaite et al. 2012). Independently of the lifespan of the products, quarterly predictions can help managers take business development decisions in the areas of financing, infrastructure planning, and marketing (Doganis et al. 2006).

B. Output Variable

Unsurprisingly, the number of sales for each product is the standard supervised information in a food sales prediction job. It's worth noting that certain food items are sold by the piece, while others are sold by weight. If the goal is to help a shop manager place suitable orders, specific sales statistics from that store should be utilized; if the goal is to plan output at a manufacturing plant, aggregate sales figures from all locations can be used. In other words, sales might be tracked at the shop level (retail) or the production plant level (manufacturing) (wholesale).

Unfortunately, the number of sales does not reveal how many sales chances were lost owing to out-of-stock items. Such information, on the other hand, is extremely difficult, if not impossible, to capture in stores. A client may not show interest in a product if it is obvious that it is out-of-stock, but even if she does, the salesperson will have to put up some effort to record it. In the case of bakeries and patisseries, there is also the issue that sales data for some goods is just approximate. One explanation for this is that salespeople are sometimes too busy to record the specific type of goods (for example, strawberry cupcake) and instead note the product's broad category (e.g., cupcake). Another cause is the mixed basket dilemma, which occurs when customers order a basket containing a range of goods with the same price per kilogram, such as pies. Weighting each product separately would take a significant amount of time, thus they are generally weighted all at once after the transaction and their broad category is noted. Bakeries and patisseries generally lack a stockholding unit as a result of these difficulties.

Using sales forecasting at the general category level is one way of addressing these issues. Another option is to allocate sales numbers to each of the individual items either evenly or depending on the shop's required quantity of these products in the same or subsequent time period. Sales prediction is represented as a regression job using sales data as the output variable. Bakker and Pechenizkiy (2009) provide an alternate representation that considers approximations of real sales figures to several sales levels, such as very low, low, medium, high, and very high. If the ordinal relationship between the values of the output variable is disregarded, sales prediction can be treated as an ordinal classification task (Frank and Hall 2001) or as a simple classification task.

C. Input Variables

The features utilized are perhaps the most critical element determining the success or failure of a machine learning project (Domingos 2012). It's crucial to look at the features that have been employed in previous food sales forecasting projects. Sales-related characteristics collected from the company's data warehouse (internal) and features received from external sources are two types of predictive variables. Lagged variables, such as product sales data for previous time units, are the most prevalent form of internal feature (days, weeks, etc). Lagged variables are the primary method through which propositional learning algorithms can record the connection between previous and current values of a series. They create a time-based window, often known as a snapshot. The number of lagged variables created determines the size of the window. Experimenting with different window sizes might help you find the right one. Based on the time granularity of the situation, reasonable default values can be chosen. If you have monthly sales data, for example, having delays up to 12-time steps in the past makes sense; if you have hourly data, you might want lags up to 24 or possibly 12. It is not required to include consecutive time units in the window; for example, for weekly data, it could make sense to include the weeks from the preceding three months, as well as the exact week from a year ago. Furthermore, consecutive lagged variables might be averaged into a single field to decrease the number of input variables, as high numbers of input variables can be detrimental to some learning methods. Time and date derived variables, such as whether it is before noon or not, the day of the week, whether it is a weekend or not, the month of the year, the quarter of the year, and the season of the year, are another popular sort of internal characteristics. Time and date derived variables, such as whether it is before noon or not, the day of the week, whether it is a weekend or not, the month of the year, the quarter of the year, and the season of the year, are another popular sort of internal characteristics. Product-related characteristics are an important category of internal features. The brand of the product, its package information, if it is on sale, its price elasticity, whether it has a short expiration date, and whether it is a holiday product are all examples of these. Weather data, financial indicators, and date-related factors (Liu and Ichise 2017) like as holidays or events attracting mass spending can all be utilized to anticipate sales(e.g. the Super Bowl in the US).

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

The sales of the previous 6 days, the sales of the corresponding day of the previous year (same day of the week, approximately same date), the sales of the previous 6 days of that day, and the percentile change in sales between the current year and the previous year were all taken into account by Doganis et al. (2006). The sales of this year with lag 1 (prior day) and 6 (the same day of the previous week, as outlets were open 6 days a week) and the corresponding sales of the previous year with lag 3, 5 and 6 were shown to be more useful empirically. In Meulstee and Pechenizkiy (2008), sales for the previous six time points (weeks in this example) of linked items were utilized as features. After an agglomerative hierarchical clustering procedure, related goods were determined to belong to the same cluster. Experiments using a variety of distance measurements revealed that Dynamic Time Warping and the Longest Common Subsequence are effective. Out-of-stock information is a fascinating piece of data that has been overlooked in the literature. The store manager does not need to put in a lot of work to record out-of-stock goods at the end of the day, and it may even be inferred automatically based on sales and stock data. This could then be used as an input feature.

D. Learning Techniques

The moving average is a basic prediction tool (MA). The mean of a target variable's value over the last n observations is the prediction. Doganis et al. (2006) employ a mixed computational intelligence method. The fuzzy means technique is used to train a radial basis function (RBF) network in an unsupervised way. For exploring the space of input variable subsets as well as the number of fuzzy sets for the input variables that are supplied as input to the fuzzy means method, a genetic algorithm is used (between 3 and 15). Meulstee and Pechenizkiy (2008) utilized an ensemble method.

Deep learning was recently used to forecast sales at a Japanese supermarket (Liu and Ichise 2017). A stacked denoising autoencoder was used to produce deep high-level features, which were then input into a long short-term memory network for forecasting future sales.

III.EVALUATION

The following experimental methodology is commonly used to evaluate a food sales prediction algorithm: each time series with sales of each product is split into two halves at the same time point, the first of which becomes the train set and the latter the test set. The algorithm is then trained and tested progressively, with all data up to the previous time point accessible for training the algorithm for each time point in the test set.

Consider a test set of n instances, (x_i, y_i), where x_i is the input vector and y_i is the target variable's value (sales). The mean value of the target variable in this test set will be denoted by y. Consider a model h: X → R, where X is the input space and R is the response space, which has been trained using the training set. The accuracy of a product's algorithm may therefore be assessed using the following criteria (Bakker and Pechenizkiy 2009; Aho et al. 2012):

1) Mean Squared Error (MSE): $\frac{1}{n}\sum_{t=1}^{n}(y_i - h(x_i))^2$

2) Root Mean Squared Error (RMSE): $\sqrt{MSE}$

3) Relative Root Mean Squared Error (RRMSE): $\sqrt{\frac{\sum_{t=1}^{n}(y_i-h(x_i))^2}{\sum_{t=1}^{n}(y_i-\bar{y})^2}}$

4) Mean Absolute Error (MAE): $\frac{1}{n}\sum_{t=1}^{n}|y_i - h(x_i)|$

5) Mean Absolute Percentage Error (MAPE): $\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_i-h(x_i)}{y_i}\right|$

6) Mean Absolute Scaled Error (MASE): $\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_i-h(x_i)}{MAE\,(Baseline)}\right|$

Calculating the prediction error for each of the items and aggregating the results is a common method of evaluating different food sales prediction algorithms. Aggregating unscaled metrics, such as MAE, MSE, and RMSE, is not a feasible option since the scale of the target variable may vary considerably between goods (e.g., low-volume versus high-volume items). MAPE adjusts the error based on the target variable's actual value. This produces data that can be aggregated over several products, but it has issues when the target variable's value is 0 or near zero. The scaled variants of the MAE and the RMSE, MASE and, RRMSE, respectively, do not have this issue. The MAE of a baseline approach, such as the moving average or the random walk, is scaled by MASE. RRMSE scales RMSE by comparing it to the RMSE of a baseline technique that forecasts the test set's mean value. Bakker and Pechenizkiy (2009) make the argument for adopting MASE with a distinct baseline for each product.

IV. CONCLUSIONS

Food sales forecasting is a critical responsibility for today's organization. It can help with both immediate and long-term decision-making, as well as cost-cutting and sales growth. Although it is essentially a time-series forecasting topic, it also presents a variety of intriguing research problems. The first is that successful outcomes necessitate the use of external factors with better predictive value (e.g. holidays, weather). The second is that it is made up of a number of distinct prediction jobs, one for each product, all of which might be interdependent. As Bakker and Pechenizkiy (2009) point out, this necessitates an armory of various learning algorithms that can fit the characteristics of each distinct time series.

Food sales prediction is a problem that hasn't gotten much attention in the machine learning literature. One of the primary causes for this is a lack of publicly available data, which is most likely related to business sensitivity and the desire to minimize potential competitor leaks. A recent Kaggle competition involving sales of a big supermarket shop in Ecuador is an exception to this rule. Many businesses conduct food sales forecasting daily. This work is still frequently relied exclusively on human knowledge. In other situations, food sales forecasting is supplied as a software module that focuses on basic business needs such as client transaction tracking. For example, FoodTec, a restaurant software system, has a module for sales forecasting that is based on the straightforward technique of using historical averages. Large food enterprises, such as supermarket chains, where sales forecasting is more important and financial discipline is required, might choose a more specific solution, such as Oracle's Retail Demand Forecasting solution. Typically, such software needs considerable modification, which is handled by consulting organizations. In both simple and advanced software solutions, the human decision-maker always has the final say and is usually allowed to enter adjustments in the form of an external variable to account for factors that are unknown to the (machine learning) system and may affect demand, such as a strike or an earthquake.

The addition of sales data from comparable goods as input variables for a feature expands the possibility of using multi-target regression techniques to handle groupings of similar products (Aho et al. 2012). According to the best of our knowledge, this has not been investigated in the past and might lead to better accuracy by exploiting dependencies among comparable goods (e.g., a rise in sales of one product could signal a similar rise in sales of another product a few days later).

REFERENCES

[1] Aho T, Ženko B, Džeroski S, Elomaa T (2012) Multi-target regression with rule ensembles. J Mach Learn Res 1:1–48

[2] Bakker J, Pechenizkiy M (2009) Food wholesales prediction: What is your baseline? In Foundations of Intelligent Systems. In: Proceedings of 18th International Symposium ISMIS 2009, volume 5722 of Lecture Notes in Computer Science, pp 493–502. Springer

[3] Chen C-Y, Lee W-I, Kuo H-M, Chen C-W, Kung-Hsing Chen (2010) The study of a forecasting sales model for fresh food. Expert Syst Appl 37(12):7696–7702

[4] Domingos P (2012) A few useful things to know about machine learning. Commun ACM 55(10):78–87

[5] Doganis P, Alexandridis A, Patrinos P, Sarimveis H (2006) Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. J Food Eng 75(2):196–204

[6] Frank E, Hall M (2001) A simple approach to ordinal classification. In: Proceedings of the 12th European Conference on Machine Learning, EMCL '01, pp 145–156, London, UK, UK. Springer

[7] Landwehr N, Hall M, Frank E (2005) Logistic model trees. Mach Learn 59(1–2):161–205 Liu X, Ichise R (2017) Food sales prediction with meteorological data A case study of a Japanese chain supermarket. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 10387 LNCS, pp 93–104

[8] Meulstee P, Pechenizkiy M (2008) Food sales prediction: if only it knew what we know. Proc IEEE Int Conf Data Min Workshops, ICDM Workshops 2008:134–143

[9] Žliobaite I, Bakker J, Pechenizkiy M (2012) Beating the baseline prediction in food sales: how intelligent an intelligent predictor is? Expert Syst Appl 39(1):806–815. https://doi.org/10.1016/j.eswa.2011.07.078

[10] Tsymbal A, Pechenizkiy M, Cunningham P, Puuronen S (2008) Dynamic integration of classifiers for handling concept drift. Inf Fus 9(1):56–68

[11] Van der Vorst JGAJ, Beulens AJM, De Wit W, Van Beek P (1998) Supply chain management in food chains: improving performance by reducing uncertainty. Int Trans Oper Res 5:487–499

[12] Žliobaite Indre, Bakker Jorn, Pechenizkiy Mykola (2009) Towards Context Aware Food Sales Prediction. In: 2009 IEEE International Conference on Data Mining Workshops, pp 94–99. IEEE, December