Temporal Convolutional Networks and Seasonal Decomposition: AI-Based Predictive Modelling for Retail Sales Volume Forecasting

Dr. Cristina Mateos, Professor of Human-Computer Interaction, Universidad Politécnica de Madrid (UPM), Spain

1. Introduction to AI-Based Forecasting in Retail Sales

AI-based forecasting systems have caught the attention of numerous retailers in recent years. An accurate and reliable prediction of sales is an important concern among business professionals because sales predictions are crucial in enhancing the decision-making abilities of any business. AI has a remarkable influence on various areas when integrated into prediction systems. This includes marketing, consumer behavior analysis, and much more. Modern AI has evolved into an efficient decision-making tool due to its ability to process vast datasets that were previously unattainable. Gone are the days when complicated mathematics and statistics were needed to forecast sales. Nowadays, machine learning algorithms can provide us with accurate sales predictions via AI solutions. Retailers are faced with numerous forecasting issues and are turning to the big data ecosystem to implement AI technology that provides the latest and most up-to-the-minute results. The fusion and integration of AI and retail/wholesale systems are reshaping how we interpret data. Why and how? A trader may forecast sales in several different time frames: weekly, monthly, quarterly, or even yearly. Statistical forecasting is the most common method employed today, though retailers face hurdles in this space. As a result, businesses seek AI solutions whereby sophisticated algorithms discover patterns and relationships within sales data. Predictions that are being calculated are relevant to both the growing demand in markets and the ongoing advances in sales prediction systems. Data is in a class of its own. Most forecasts anchor their value on data, and the greater the volume and value, the more accurate the forecasts tend to be.

2. Key Concepts in Machine Learning for Sales Forecasting

Key Concepts for Machine Learning in Retail Sales Forecasting - Supervised Learning Supervised learning is a machine learning technique for a model to learn to predict an outcome based on input labeled data. The model learns from past data and uses this knowledge to facilitate future decision-making or behavior. For example, think of a bank that wants to target new customers for a mortgage based on their risk of default on payments. The input to the model would, in this case, be labeled data of people who were clients in the past. It does not end with who should be targeted for a mortgage; the same model could be used for cross-selling a new product to existing clients. It helps to understand that the model or predictive model could estimate any value that is based on historic data if you give it the right features. Supervised models can be divided into linear and non-linear models. The big advantage of the first is that they are easy to interpret and computationally fast. On the other hand, non-linear models can detect more complex patterns, which can result in higher accuracy; the downside is that there is a potential for overfitting or having difficulty interpreting them. In this research paper, only classification models are tested against each other. Classification basically predicts either a binary or a multi-category class as the output. While it is essential to select an appropriate model, it will depend on the characteristics present in your data. In order to determine this, exploring and applying different models to your data is a time-consuming but essential task that needs to be done efficiently in terms of computational speed without sacrificing accuracy. It will and can vary and builds specifically on business problems and industry data dynamics, as will be showcased in this research paper. Key Concepts for Machine Learning in Retail Sales Forecasting - Unsupervised Learning Techniques Data represents an important source of information; hence, for retail data, clustering could enable retailers to segment their client base and find meaningful patterns that could be referred to the marketing department, understanding what drives sales and honing in on the relevant features that carry the heaviest weighting in the feature space of sales in terms of retail data analysis. This could be your underlying problem with sales forecasting. Supervised learning refers here more to classification, decision trees versus ensemble decision tree techniques, as well as the more exotic gradient boosting. The supervised learner will estimate our sales response. Continuing from the introduction on the clustering of unsupervised learning, it refers mainly to customers where you want to partition a client base into homogeneous subgroups. Clustering refers to all methods that partition the data set into subsets of observational units. A subgroup of customers is distinguished from other subgroups in order to understand and devise marketing strategies; a strategy could be, for example, to customize the retailer's store layout according to the different clusters of customers. Even in unsupervised machine learning techniques, we have different methods to test, including hierarchical, partition, and mixture models.

2.1. Supervised Learning Algorithms

When it comes to sales forecasting, supervised learning algorithms that are specifically designed for trained prediction of time-series data fit the task. Supervised learning allows developers to provide thousands or millions of known input/output pairs to the model to train on, and it conditions the model to predict future results given new data supplies as input. In the case of future sales, you can think of the training input as representing historical sales data and the output as the sales during the next hour, day, or month. A set of training data and labels/sales data is used to train a prediction model, and any following predictions are performed using new data that hasn't previously been seen before.

Various time series prediction methods are available. Sales usually have a form of trend (increases or decreases over time), seasonality (variation in sales caused by data-related or calendar-related effects), as well as correlational forecasting (variation in sales and sales-related data can have on other factors). The most common forecasting methods are Linear Regression, Polynomial Regression, Seasonal Decomposition of Time Series using loess, Random Forest, Gradient Boosting Trees, Regression SVM, Univariate Multilayer Perceptron, and Multivariate Multilayer Perceptron. Each of these methods has its respective strengths and weaknesses and is ideal for different types of forecasting tasks. Each model also has a number of hyperparameters that can be tuned to optimize the forecasting quality further. The hyperparameters are usually found by experimenting with a model using some already known data to see the predictions for future insights. Most machine learning models, including those used for forecasting, work best with large amounts of data. The model's ability to make accurate predictions is relative to the training data available, as more stable and complete data provides more accurate forecasting. Model hyperparameters need to be re-tuned when more training data is available.

Before actually using any discounting model, the user should decide on whether to use the default settings or fine-tune the hyperparameters manually. After deciding on the aforementioned, the model can be implemented to predict future instances of conditional sales using specific retail data, such as sales price, promotion indicators, and whether the item or store is affected by any price elasticity.

2.2. Unsupervised Learning Techniques

Unsupervised learning is a machine learning technique used in the analysis of data in the retail domain. It is said to be a more challenging field of machine learning compared to supervised learning because it is based on the input data and has no corresponding output label or values. Unsupervised learning analyzes complex data and the distribution of data to detect patterns, identify groupings, and structures in data patterns. Consequently, it consists of a group of techniques that cluster and organize related data. Techniques like clustering, hierarchical clustering, and autoencoders are used for detecting and grouping related customer segments in retail data. These techniques also help in anomaly detection, i.e., identifying the hidden patterns that could be indicators of unusual behavior and in finding outliers, irrelevant, and unexpected patterns that are uncommon in the data records.

In customer relationship management and retail sales, unsupervised learning is used for analyzing patterns and behaviors in point of sale and market basket data. One of the important applications of unsupervised learning in PoS helps in increasing the understanding of what and why customers purchase patterns. Also, the insights from the results of unsupervised learning are used to increase customer loyalty by providing satisfaction. Moreover, in the sales prediction model, the findings of unsupervised learning are used to determine input features for the model. By linking the output of unsupervised learning and then applying supervised learning to the output, the prediction increased considerably. More sophisticated sales prediction could be done using unsupervised learning, and the results are considered to be more reliable than only using supervised learning.

2.3. Feature Engineering and Selection

A significant effort in the success of predictive modeling lies in feature engineering and selection. In this subsection, we describe the approach that led us to most of our sold stock models. Feature engineering is the often-overlooked art of transforming raw information into more useful input for machine learning models. This means modifying existing features, extracting new features, or potentially removing noisy or highly biased input. A common adage states that "All models are wrong, but some are useful." A corollary may be stated similarly in machine learning: "All models are overfit, but some have interpretable feature importance."

Within the feature space, there are a number of strategies with specific features that could apply to forecast retail sales and do not necessarily translate to other cutting-edge deep learning model architectures that deal with raw sequential data. During the feature set creation stage, it is beneficial to consult domain experts who can suggest relevant features, feature transformations, and derived features. Some common set features are one-hot encoding of categorical variables, normalization of continuous variables such as temperature and humidity, and principal component analysis for dimensionality reduction. Care must be taken in this stage, as erroneous feature construction or irrelevant features can lead to strayed model performance or additional runtime, especially with very large or very sparse data. Careful feature selection can greatly improve this model's accuracy, but it takes some trial and error to determine the best features. With careful feature design, we can also easily interpret the model's importance around sale event forecasts.

3. Data Preparation and Preprocessing for Retail Sales Data

The preparation and preprocessing of raw data is an essential step in designing forecasting models for retail sales. The main purpose of this phase is to clean and organize the raw data, keeping in mind the next phase of analysis. Data preparation and preprocessing may involve different steps, including handling missing values, detecting outliers, data transformation, formatting temporal data, and scaling. The objective of data normalization is to modify the values of numerically based features, rendering them a standard representation. This technique has two key advantages: it develops in the model a clear understanding of the features and their association during the training process, and enables faster and more efficient classical algorithms. The significance and impact of each preprocessing stage is depicted by applying regression evaluation indices on serial models without that particular stage. The performance of models with different data preparation stages demonstrates the necessity of every phase in the proposed forecasting models.

Since sales forecasting critically depends on historical data, text summarization, cleaning, and preprocessing are some of the most critical phases in the forecasting pipeline. The findings revealed that the quality of input data deeply impacts forecast accuracy. Thus, based on the data used, a suboptimal forecasting model can still produce more accurate results than a sophisticated model trained on poor-quality data. Data preprocessing approaches facilitate the examination of internal data and the formatting of it into an appropriate format to further refine and utilize it for data analysis. The first fundamental stage in preparing data for forecasting is using techniques to refine input datasets by handling and eliminating any types of irregularities. Cleaning data substantially influences how well the learning algorithms work, and consequently the pattern identifications given to predict on.

4. Building and Evaluating AI-Based Retail Sales Forecasting Models

There is no one-size-fits-all forecasting model that works across all situations. In retail sales, selecting and developing the appropriate model depends on business requirements, available data, and the sales forecasting context. Expertise and iteration play a major role in developing new models. While a plethora of modeling options are available for AI-based sales forecasting, both classical and modern machine learning model evaluation metrics present a performance measure. This includes evaluation metrics such as accuracy, precision, recall, F1 measure, and the AUC-ROC method that measures success based on the true positive rate. A trading policy is important when properly evaluating the performance of a machine learning model. This includes the presence of a threshold that generates class predictions and a technique for selecting optimal or best trading policy models.

After building, training, and validating AI forecasting models, we might find optimal hyperparameters that allow the newly developed model to provide the best possible sales forecasting accuracy. This requires a technique commonly referred to as hyperparameter tuning or sampling. In a nutshell, a model's ability to generalize to fresh, unseen data is checked during cross-validation, which is generally applicable when machine learning and statistical models are actively developed and tested. Techniques including and beyond k-fold cross-validation and leave-one-out cross-validation can be implemented. Ensembling is a valuable method for enhancing prediction robustness and could be deployed for AI for retail forecasting to enable a collection of models to give us a more accurate AI-based retail sales forecast. In the cycle of build and test, evaluation feedback will guide the modification of a new uniquely developed model.

4.1. Model Selection and Hyperparameter Tuning

Model selection helps in identifying which forecasting algorithm is best suited to analyzing the available sales data and delivering accurate predictions based on the business context and the forecast horizon. The selection of the forecasting model should consider the nature of the sales data and the objective of the business, and whether the sales data is categorical, binary, ordinal, or continuous. Large datasets with a wide range of features may necessitate the use of more computationally expensive models. In retail, the availability of recent sales data positions deep learning networks such as LSTMs as ideal models for sales forecasting. However, these deep learning models require different hyperparameters to be set before training.

Hyperparameters are model parameters that cannot be estimated from the training data, and they are set before the model is trained. These hyperparameters help in determining the efficiency of the model and affect the model's ability to accurately forecast sales. The same deep learning models applied to the same sales data will deliver different forecasting performances with variations in hyperparameters such as the number of layers in LSTMs, the number of LSTMs in each layer, and the learning rates within the model. In practice, selecting the best set of values for multi-parameter models is done using hyperparameter optimization techniques such as grid search, random search, and Bayesian optimization. The determination of the best hyperparameter values has a significant effect on model forecasting accuracy, and a combination of hyperparameters that is not well tuned can lead to a low accuracy forecasting model. Given the complexity of training models, grid search and random search are the most appropriate methods to find the best hyperparameters. Grid search optimizes hyperparameters by trying every possible combination to find the best results. Random search optimizes hyperparameters by randomizing the hyperparameter combinations. Automating this hyperparameter search with Bayesian optimization is also useful in optimizing the hyperparameters within extremely complex models.

4.2. Cross-Validation Techniques

In order to properly evaluate forecasting models, we have to be able to determine how well the model performs on data it has not seen before. Cross-validation is a technique used in model selection to partition a sample of data into subsets in order to assess the extent that a model trained on some of the data can generalize to an independent dataset. This is especially important when dealing with relatively small sample sizes or when considering the application of models to testing datasets. The different strategies for performing cross-validation all consist of ways to divide this data into testing (or validation) and training sets. When considering a time series cross-validation strategy, samples are split using a particular timeframe, so that the training data comes before the validation data.

These cross-validation techniques can be contrasted with more traditional k-fold split strategies, which typically divide data randomly. When using a smaller validation data size, there may be a higher variance in the metrics used to measure performance. Due to this problem of sampling variance, two solutions exist: increase the size of the validation data or employ multiple randomly drawn validation datasets. Repeated stratified k-fold cross-validation can help to address the latter issue. The choice between cross-validation techniques should be informed by the number of samples available to train and/or test the forecasting models. When deciding on the alternating divided sample, it is important to consider how many training samples should be used contingent on the readily available number of samples.

5. Applications and Benefits of AI-Based Sales Forecasting in Retail

Implementing AI into sales forecasting allows retailers to address numerous tasks. Forecasting sales with greater accuracy enables precise inventory management, resulting in reduced carrying costs, stock-out prevention, and more efficient allocation. As part of a competitive intelligence system, AI-based predictive analytics aids in predicting customer purchasing patterns and responses to marketing activities, thus facilitating the optimization of marketing strategies. Accurate demand forecasting, realized with the help of AI, allows for product launch assessments, aiding with expectations, as well as product line management. As a result, they can reduce the necessity for unplanned and computationally demanding last call forecasts. Furthermore, AI enables personalized customer service by providing clients with unique and relevant recommendations to complement their specific needs based on a multitude of data sources, such as website interactions or items they added to their online baskets. This can be implemented in a variety of ways, such as product recommendations, flash sales, personalized email campaigns, and one-on-one offerings. Personalization is widely implemented, and some retail businesses have reported significant sales growth as a result of such initiatives.

The market and individual sales forecasting segment has already begun implementing AI-based solutions, and several instances of improved outcomes have been documented. A forecasting and analytics platform implemented an AI-based propagation algorithm and observed an improvement in sales forecasts. The manufacturing arm of a pet food business reduced worldwide safety stock inventory levels by over 10% after employing an AI-based predictive analytics platform, resulting in a 30% increase in stock availability across their three largest distribution centers. Despite these advantages, employing AI models comes with its fair share of limitations, first and foremost data quality issues. Research has shown that an estimated 25% of available sales, inventory, and supply chain related data is fit for the training of machine learning algorithms, thus narrowing the horizons for AI-based forecast improvements. Moreover, the success of AI-based forecasts can be heavily affected by data quality. Confidence is a more likely byproduct if sales data quality is excellent. Lower data quality, on the other hand, means there is a higher likelihood of uncertainty and an increased demand for human judgment.

6. Future Direction

6.1. Future Developments in Technology AI-based forecasting models have the potential to be developed further and coupled with the general turning towards big data. The already high level of accuracy associated with many types of advanced sales forecasting based on data analytics will still rise, especially because machine learning technologies will be implemented in more forecasting solutions. Moreover, bigger and more detailed data will also portend increasing accuracy. As more sales datasets become available in retail, the improvements in model performance, if accompanied by the increased availability and acceptance of new technologies, will yield better forecasting.

6.2. Automation of Decision Making In addition to improving accuracy, developments in machine learning and AI should also aid in the increased automation of decision-making in relation to supply chain management and bottom-up forecasting. Due to the increasing availability of real-time data sources and the integration of IoT devices into the retail environment, it will be easier to assess the effects of external trends and events. Furthermore, because much of the sales prediction process will fall to algorithm developers and will be based on AI, a greater level of accuracy in the reading of market trends will likely emerge. Lastly, economic modeling capability may also improve. While these developments have the potential to drive strong improvements in sales forecasting, several potential roles for abuse, bad practice, and ongoing challenges should also be considered in advance. Moreover, regulatory issues and ethical considerations may restrict some of the development of AI in sales prediction. Therefore, a focus on the proper use of forecasting, sales prediction, and other AI or machine learning models is advisable. Nevertheless, it is important to emphasize that while the future direction of forecasting, AI, and machine learning solutions as part of that is ripe for exploitation by retail organizations, especially combined with technologists who offer in-depth forecasting, there is a large number of digital solutions in the market. In conclusion, forecasting accuracy may significantly improve due to increasing developments in technology and integrated big data sources. However, ethical considerations and regulatory data barriers may hinder rapid progress. These built-in challenges are not new and are not easy to overcome. Neither are demands for large-scale cooperation between retail industry members. Resolving these issues will indeed require a combined effort between retailers and small tech provider companies.

7. Conclusion

The globalization and the information revolution have caused a fundamental change in consumer behavior. The internet has made economic and social interactions faster and more efficient by eliminating time and location constraints. With the increasing use of internet users in line with new technology as wearable smart devices, smart mobile devices that can access knowledge from anywhere have served as platforms for machine learning based applications. Quick access to information remotely affects shopping behaviors. Online shopping has become a significant shopping medium in which the majority of consumers can take advantage in the rapidly developing technology. Online shopping, which is in a leading position in retail sales, has brought huge commercial competition in the retail industry. With the increase of competition, commercial data have also increased production significantly. Generated data reveal important insights about the business, which is required to be evaluated and used for making accurate decisions. Predictive models help businesses to make decisions from large volume of data. In addition, the retail business also aims to have an indicator that will be valid in understanding the future market in order to take important decisions about its price, product, service and location. In the data-driven era, predictive models are among the critical, specialized tools retailers use for forecasting and marketing. Data-driven models such as machine learning provide much additional information on retail data such as positioning errors and the association of data with dependent data in each model. In addition, it also produces better predictions and has predictive performance compared to other statistical methods.