

# IMPLEMENTATION OF AUTO ARIMA ALGORITHM FOR PREDICTION OF SALES USING PYTHON

<sup>1</sup>K. Logesh, M.Tech.,(Ph.D)\*, <sup>2</sup>Sripadaraj.M.S,<sup>3</sup>Satyanarayana.P, <sup>4</sup>Saivarsha.M

<sup>1</sup>Associate Professor, <sup>2,3,4</sup>UG (B.Tech.,) Students  
Department of Computer Science and Engineering,  
Kuppam Engineering College, Kuppam, AP, India.

**Abstract:** This paper presents a use case of machine learning for sales forecasting in retail demand and sales prediction. In particular, the Auto Arima algorithm is used to design a prediction model to accurately estimate probable sales for retail outlets. The forecast of potential sales is based on a mixture of temporal and economical features including prior sales data, store promotions, retail competitors, location and accessibility of the store as well as the time of year. The model building process was guided by common sense reasoning and by analytic knowledge discovered during data analysis and definitive conclusions were drawn. The performances of the Auto Arima predictor were compared with those of more traditional regression algorithms like Linear Regression and Random Forest Regression. Findings not only reveal that the Auto Arima algorithm outperforms the traditional modeling approaches with regard to prediction accuracy, but it also uncovers new knowledge that is hidden in data which help in building a more robust feature set and strengthen the sales prediction model.

**IndexTerms - Prediction, Auto Arima Predictor, Traditional Regression, Data analysis.**

## I. INTRODUCTION

Sales prediction is an important part of modern business intelligence. It can be a complex problem. Especially, in the case of lack of the data, missing data, a lot of outliers. Sales can be treated as a 10-time series. At present time different time series theories and models have been developed. We can mention Holt-Winters model, ARIMA, SARIMA, SARIMAX, GARCH, etc. But their use in case of sales prediction is problematic due to several reasons. Here are several of them: We need to have historical data for a long time period to capture seasonality. But often we do not have historical data for a target variable, for example in case when a new product is launched. But we have sales time series for a similar product and we can expect that our new product will have a similar sales pattern. • Sales can have complicated seasonality - intra-day, intra-week, intra-month, annual. • Sales data can have a lot of outliers and missing data. We have to clean outliers and interpolate data before using a time series approach. • We need to take into account a lot of exogenous factors which have impact on sales.

### Existing system

Predicting the right demand at each retail outlet is crucial for the success of every retailing company because it helps towards inventory management, results in better distribution of produce across stores, minimizes over and under stocking at each store thereby minimizing losses, and most importantly maximizes sales and customer satisfaction. Due to the high stakes involved with demand prediction, it becomes a vital problem to solve for every retail company. Further, demand can depend on a variety of external factors like competition, weather, seasonal trends, etc. and internal actions like promotions, sales events, pricing, assortment planning etc., adding to the complexity of the problem. Consequently, the modeling of demand prediction taking into account all of the factors per retail outlet becomes essential for every retail company.

### Disadvantages

1. Accuracy and efficiency is less.
2. We don't have chances to plan for future business.

## II. Literature Survey

Literature survey is the most important step in software development process. Before developing the tool, it is necessary to determine the time factor, economy n company strength. Once these things are satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system, the above consideration is taken into account for developing the proposed system.

### The Walmart challenge: Modelling weekly sale

At writing, our internal competition at Bletchley has ended. Interestingly, the winning group had a different approach [1] then would be expected from an AI/Machine Learning bootcamp. Their forecasts were based simply on a median of the weekly sales grouped by the Type of Store, Store & Department number, Month and Holiday dummy.

Therefore, in my next approach, the goal will be to improve their results with the help of Neural Networks and other machine learning methods. In fact, the median will be computed similarly to how the winning group did, and a new variable, the difference to the median, will be computed. This difference will be the new dependent variable and will be estimated based on new holiday dummies, markdowns and info on lagged sales data if available.

### **Predict Sales Data**

This article is the missing link between knowing about machine learning and working with real data for the first time. So this article will **not** focus on the basic machine learning concepts but rather on the data science/data engineering part. In the end you will be able to get an intuition for datasets and transform data into something you can feed to an algorithm. You will have used cross-validation to check the performance of your predictions.

### **Predicting expected sales for Big mart's stores**

The Big Mart's sale prediction proposed by Analytics Vidhya. According to the information provided, Big mart is a big supermarket chain, with stores all around the country and its current board set out a challenge to all Data Scientist out there to help them create a model that can predict the sales, per product, for each store. Big Mart has collected sales data from the year 2013, [2] for 1559 products across 10 stores in different cities. With this information the corporation hopes we can identify the products and stores which play a key role in their sales and use that information to take the correct measures to ensure success of their business.

### **Practical Machine Learning Project in Python on House Prices Data**

The data set for this project has been taken from Kaggle's Housing Data Set Knowledge Competition. As mentioned above, the data set is simple. This project aims at predicting house prices (residential) in Ames, Iowa, USA. I believe this problem statement is quite self-explanatory and doesn't need more explanation. Hence, we move to the next step.

#### **1) Use Machine Learning to Predict the Quality of Wines**

Machine Learning (ML) is a subfield of artificial intelligence. It concerns giving computers the ability to learn without being explicitly programmed. Over the years, machine learning's popularity and demand has certainly been on the rise, as indicated by this hype curve.

### **III. Proposed system**

In this paper we have performing sales forecasting for stores using machine learning techniques. The task involved predicting the sales on any given month at any store. In order to familiarize ourselves with the task we have studied previous work in the domain including Time Series Algorithm as well as a spatial approach. A lot of analysis was performed on the data to identify patterns and outliers which would boost or impede the prediction algorithm. The features used ranged from store information to customer information as well as socio-geographical information. Machine learning methods Auto Arima which is an improved gradient boosting algorithm was observed to perform the best at prediction. With efficiency being the way forward in most industries today, we aim to expand our solution to help stores improve productivity and increase revenue by taking advantage of Data Analysis.

### **Advantages**

1. Accuracy and efficiency is high.
2. We can improve our sales with initial forecasting the sales.

### **IV. Architecture**

The master budget is the primary financial planning mechanism for an organization and also provides the foundation for a traditional financial control system. More specifically, it is a comprehensive integrated financial plan developed for a specific period of time, e.g., for a month, quarter, or year. This is a much broader concept than the first three types of budgeting. The master budget includes many appropriation budgets (typically in the administrative and service areas) as well as flexible budgets, a capital budget and much more.

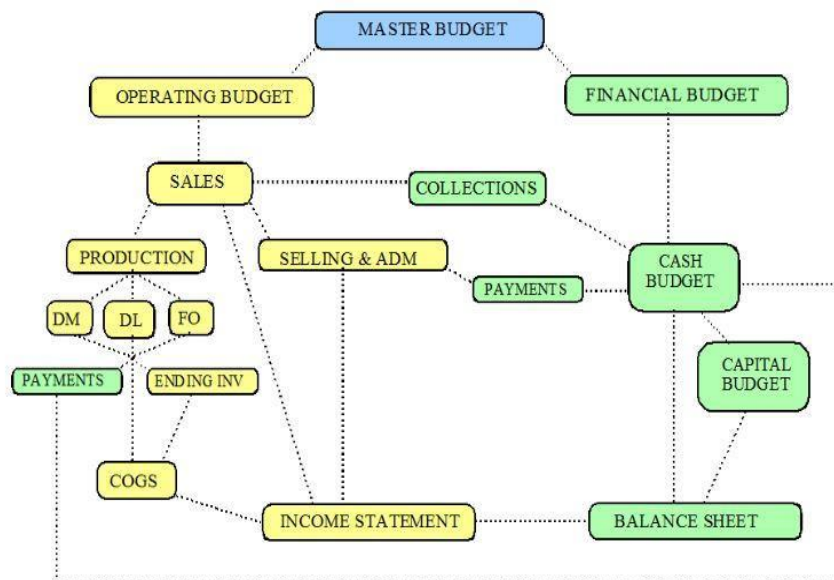


Figure 1. Diagram for Master Budget

The master budget has two major parts including the operating budget and the financial budget. The operating budget begins with the sales budget and ends with the budgeted income statement. The financial budget includes the capital budget as well as a cash budget, and a budgeted balance sheet. The main focus of this chapter is on the various parts of the operating budget and the cash budget. The budgeted balance sheet is covered briefly, but not emphasized. A detailed discussion of capital budgeting and investment management is provided after some other prerequisite concepts are introduced. In the next section, we consider the purposes, benefits, limitations and assumptions of the master budget.

### The Purposes and Benefits of the Master Budget

There are a variety of purposes and benefits obtained from budgeting. Consider the following:

#### A. Integrates and Coordinates

The master budget is the major planning device for an organization. Thus, it is used to integrate and coordinate the activities of the various functional areas within the organization. For example, a comprehensive plan helps ensure that all the needed inputs (equipment, materials, labor, supplies, etc.) will be at the right place at the right time when needed, just-in-time if possible. It also helps insure that manufacturing is planning to produce the same mix of products that marketing is planning to sell. The idea is that the products should be pulled through the system on the basis of the sales budget, rather than produced speculatively and pushed on the sales force. As discussed excess inventory and other resources hide problems and add unnecessary costs. The integrative nature of the budget provides a way to implement the lean enterprise concepts of just-in-time and the theory of constraints where the emphasis is placed on the performance of the total system (organization) rather than the various subsystems or functional areas.

#### B. Communicates and Motivates

Another purpose and benefit of the master budget is to provide a communication device through which the company's employees in each functional area can see how their efforts contribute to the overall goals of the organization. This communication tends to be good for morale and enhance jobs satisfaction. People need to know how their efforts add value to the organization and its' products and services. The behavioral aspects of budgeting are extremely important.

#### C. Promotes Continuous Improvement

The planning process encourages management to consider alternatives that might improve customer value and reduce costs. Recall that "Plan" is the first step in the Shewhart-Deming plan- do-check-action continuous improvement cycle discussed in Chapter 8. The PDCA cycle supports specific improvements in the company's processes. The financial plan and subsequent financial performance measurements reflect the financial expectations and consequences of those efforts.

#### D. Guides Performance

The master budget also provides a guide for accomplishing the objectives included in the plan. The budget becomes the basis for the acquisition and utilization of the various resources needed to implement the plan. Perfection of the guidance aspect of budgeting can significantly reduce the amount of uncertainty and variability in the company's operations. In a JIT environment, the budget can also serve as a guide to vendors. For example, suppliers to General Motors Saturn plant in Tennessee have access to Saturn's production schedule through an on-line database. This information allows Saturn's vendors to deliver the required parts in the order needed to precise locations just-in-time without a purchase order or delivery schedule.

## E. Facilitates Evaluation and Control

The master budget provides a method for evaluating and subsequently controlling performance. We will develop this idea in considerable detail in the following chapter. Performance evaluation and control is a very powerful and very controversial aspect of budgeting.

### V. Auto Arima predictor

Picture this – You've been tasked with forecasting the price of the next iPhone and have been provided with historical data. This includes features like quarterly sales, month-on-month expenditure, and a whole host of things that come with Apple's balance sheet. As a data scientist, which kind of problem would you classify this as? Time series modeling, of course. From predicting the sales of a product to estimating the electricity usage of households, time series forecasting is one of the core skills any data scientist is expected to know, if not master. There are a plethora of different techniques out there which you can use, and we will be covering one of the most effective ones, called Auto ARIMA.

### The Lead-Driven Model

This model involves analyzing each of your lead sources and creating a forecast based on the value of each source. Looking at the very beginning of the buyer's journey can tell us a lot about how that journey will end. It's like a bad romantic comedy; there are early tell tale signs of how the story will end if you're in any way perceptive. By assigning a value to each of your lead sources or types, you can get a better sense of the probability for each of those leads to turn into revenue.

$$\text{Leads Needed} = \text{Desired Revenue} / \text{Average Lead Value}$$

### The Opportunity Creation Driven Model

This model helps you predict which opportunities are more likely to close based on demographic and behavioral data. If we go back to the movie analogy, it's often easy to predict who is going to fall in love with who, who's the hero, and who's the villain based on what the characters look like and how they behave and interact with each other. Predicting an opportunity's likelihood to close is very much like this. By looking at demographic and behavioral data, we can get a better sense of the probability to close and the expected value of the deal.

$$\text{Expected Value of Opportunity} = \text{Average Sale Price} * \text{Close Rate}$$

### The Opportunity Stages Driven Model

This model predicts the probability of an opportunity to close based on where the prospect currently is in your **sales process**. This one is pretty logical. In sales, if you know the average sales cycle for your product and you have mapped out the stages involved for someone to get from early stage awareness to decision made, you can get a good sense for their likelihood to close within the current forecasting period.

$$\text{Expected Revenue} = \text{Deal Amount} * \text{Probability to Close}$$

## VI. Experiment and Evaluation

### Two modules

1. Producer
2. Consumer

#### Producer

In this module first producer can register the details using some required details like username, password, email, phone, date of birth, address, city etc... After that producer can login using username and password select type otherwise some message will be printed please enter required details. Producer can upload products with details, System will be shore uploaded products. Producer can view stock details and update and delete products. And producer view prediction of the stock that can available to view to choose product type and month.

## Consumer

In this module first consumer can register the details using some required details like username, password, email, phone, date of birth, address, city etc... After that consumer can login using username and password select type otherwise some message will be printed please enter required details. Consumer can view products and buy the product and available to no. of quantity choosing. And consumer view previews ordered products.

### Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

### Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.
- The entry screen, messages and responses must not be delayed.

### Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

### Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

**Test Results:** All the test cases mentioned above passed successfully. No defects encountered.

## VII. Conclusion

In our case study, we considered different machine learning approaches for time series forecasting. The accuracy on the validation set is an important indicator for choosing an optimal number of iterations of ML algorithm. The effect of machine learning generalization consists in the fact that a classifier captures the patterns which exist in the whole set of stores or products. It can be used for sales forecasting when there is a small number of historical data for specific sales time series in the case when a new product or store is launched. Using stacking model on the second level with the covariates that are predicted by machine learning models on the first level, makes it possible to take into account the differences in the results for machine learning models received for different sets of parameters and subsets of samples. For stacking machine learning models, the Lasso regression can be used. Using multilevel stacking models, one can receive more precise results in comparison with single models.

## References

- 1) The Walmart challenge: Modelling weekly sales, <https://www.kaggle.com/fernandol/cracking-the-walmart-sales-forecasting-challenge>
- 2) <https://www.databriefing.com/coding/predict-sales-1/>, Predict Sales Data
- 3) Predicting expected sales for Bigmart's stores, <https://medium.com/diogo-menezes-borges/project-1-bigmart-sale-prediction-fdc04f07dc1e>
- 4) Practical Machine Learning Project in Python on House Prices Data, <https://www.hackerearth.com/practice/machine-learning/machine-learning-projects/python-project/tutorial/>
- 5) How to Use Machine Learning to Predict the Quality of Wines.
- 6) Build High Performance Time Series Models using Auto ARIMA in Python and R, in <https://www.analyticsvidhya.com/blog/2018/08/auto-arima-time-series-modeling-python-r/>
- 7) <https://www.databriefing.com/coding/predict-sales-1/>  
 [1]-<https://www.kaggle.com/fernandol/cracking-the-walmart-sales-forecasting-challenge>  
 [2]-<https://medium.com/diogo-menezes-borges/project-1-bigmart-sale-prediction-fdc04f07dc1e>  
 [3]-<https://www.hackerearth.com/practice/machine-learning/machine-learning-projects/python-project/tutorial/>