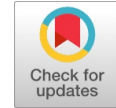


# Usage of Predictive Research on further Business

Amarendra Mohanty,Ranjana



**Abstract**— Predictive analytics is a group of methods that uses statistical and other empirical techniques to predict future events, based on past occurrences. Predictive analytics can generate valuable information for the management of a supply chain company to improve decision-making. This can be useful for demand forecasting, defect detection, maximizing equipment value, preventive maintenance, optimize marketing strategies, retain customer and connected aftermarket service in industry.

**Keywords**— Predictive Analysis; Big Data; Pattern; Data; Decision, Business; IoT('Internet of Things); R, Python

## I. INTRODUCTION

Predictive analytics incorporates a variety of statistical techniques from modeling, machine learning and data mining that analyze current and historical data in order to determine patterns; predict about future outcome and trends. Predictive analytic does not tell you what will happen in the future. It forecasts what might happen in the future with an acceptable level of reliability. It helps to identify potential risks and opportunities for a company. This also helps to better understand customers, products, partners and market. Often the unknown event of interest is in the future, but predictive analytics can be applied to any type of unknown whether it is in the past, present or future.

## II. BACKGROUND

The major problem for industry is to maintain older sales history. Organizations can't create more predictive models to forecast the demand due to lack of older sales history data. Organizations normally shut down their machines for maintenance. This machine downtime enacts a cost to firms due to declined productivity and can be mostly disruptive in both consumer products and complex manufacturing supply chains. Executives in asset-intensive industries often state that the primary operational risk to their businesses is sudden failures of their assets. Organizations usually failed in identifying not only faulty products but can also determine the important factors that influence the success or failure of the process. Retaining the customer is the major challenge for organizations. We didn't have any analytics related to customer behavior earlier. Organizations normally facing challenges in building perfect model and they don't have any perfect method to achieve optimized marketing

strategies. In this research we have explained how we can use predictive analytics by using Big Data Analytical tools R and Python.

Enterprise Business Analytics - Solution Coverage across the Manufacturing Value Chain

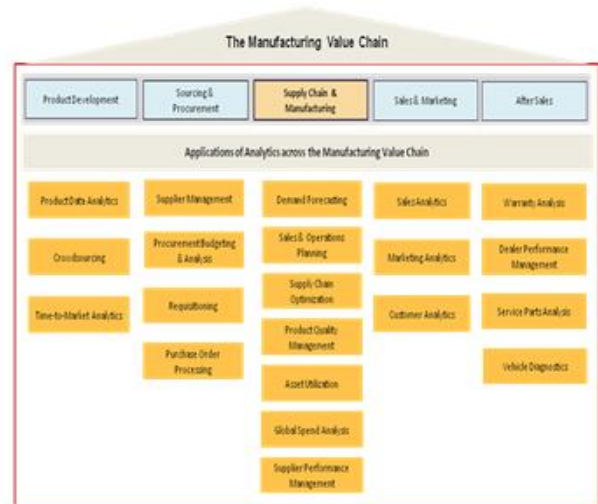


Figure 1: The Manufacturing Value Chain

## III. HIGH LEVEL FACTORS IN BUSINESS

### A. Demand Forecasting

Manufacturers need to sell their products. Demand is often seasonal or cyclical. In such cases, knowing how external factors such as oil prices, weather, the consumer price index and prime rate could affect your customer's sales demand can help in resource allocation in manufacturing. Predictive analytics takes historical sales data and applies forms of regression to predict future sales based upon past sales. Good predictive modelers find additional factors that influenced sales in the past and apply those factors to forecasted sales models.

### B. Preventive Maintenance

Efficient operating supply chains play vital role for the stable profits. A wave of new data generated by the "Internet of Things" (IoT) can provide real-time measurement and transmission of data from remote sources on detailed aspects of production processes. Existing IoT data is primarily analyzed by data scientists to prepare predictive models of different failure conditions. Those predictive models are then used to assess the incoming streaming data from the equipment. If a potential fault is detected, a message can be sent to the operator and maintenance staff, or an action can be created to immediately shutdown the machine to avoid damaging the capital asset and further disrupting production depending on the type.

Manuscript published on 30 September 2019.

\*Correspondence Author(s)

Amarendra Mohanty, Research Scholar, Dept. of MCA Hindustan Institute of Technology & Science, Chennai, India.  
(E-mail: mohanty.amarendra@gmail.com)

Dr.P.Ranjana, Associate Professor, School of Computing Sciences, Hindustan Institute of Technology & Science, Chennai, Tamilnadu, India.  
(E-mail:pranjana@hindustanuniv.ac.in)

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

C. Retain Customers

Predictive analytics can help not only to attract the new business, but also help to retain the customers. The first-time sales are converted into recurring revenue. In order to apply predictive analytics for customer retention, companies first need to collect data on their customers, including details about the products and services a customer purchased (e.g., price and brand), geographic information, demographic and whether they are first-time or returning customers. A 'score' can be produced by applying predictive analytics to these data which indicates how likely a customer is to make additional purchases. Based on consumer behavior, they can be targeted with personalized marketing campaigns, like specific product or service recommendations, special rebate and discount offers. Providing personalized messages on Birthdays, offers, discounts etc. can help foster loyalty and keep the customers away from your competitors.

D. Defect detection

Data mining can help in identifying the patterns that lead towards detecting the defects and potential failure in manufacturing equipment. This methodology helps in identifying not only faulty products but can also determine the significant factors that influence the success or failure of the process. The predictive analytics process for predicting failure and fault detection depends on data that is typically collected during the normal operation of machine tools. Examples of the kind of data collected in this scenario include temperature, vibration levels, acoustic information, measure of forces, deflections, and other similar technical inputs

E. Maximizing Equipment Value

Manufacturing engineers use much of their time optimizing the value of equipment in the factory. With predictive analytics and the kind of tools being developed to actualize it, companies can predict the points at which equipment begin to wear out, allowing them to implement preventive care sooner. This will allow the machinery to be at work for longer periods of time without intermittent shut downs for repair.

F. Optimize Marketing Strategies

Companies can increase their balance sheet by maximizing the profit on their marketing investments. An increasing number of companies are collecting data from their marketing efforts and applying predictive analytics to better understand their customers and how to communicate with them more effectively and efficiently.

Marketers today can collect valuable data about consumers through website, social media activities analytics tools, online forms, surveys and email campaign results. Companies can also track the purchase patterns of buyer or purchase similar data from third-party resources. Using predictive analytics, marketers can transform these data into valuable insights, like who are likely to purchase a particular product or service, consumer preference etc.

Companies can target the various segments in their target market with more effective personalized messaging by using this information. Predictive analytics can also indicate which marketing campaigns and channels are most effective

at driving sales. Ultimately, this can help to allocate more of marketing budget to efforts that yields a higher return.

Below are some of the predictive analytics algorithm models which can be used in the preparation of predictive models.

IV.SOLUTION & RESULTS

A.Time Series Using Python

The sample data are used to generate the Python Pandas Data Frame model to analyze 10 years Crude oil prices of WTI and Brent.

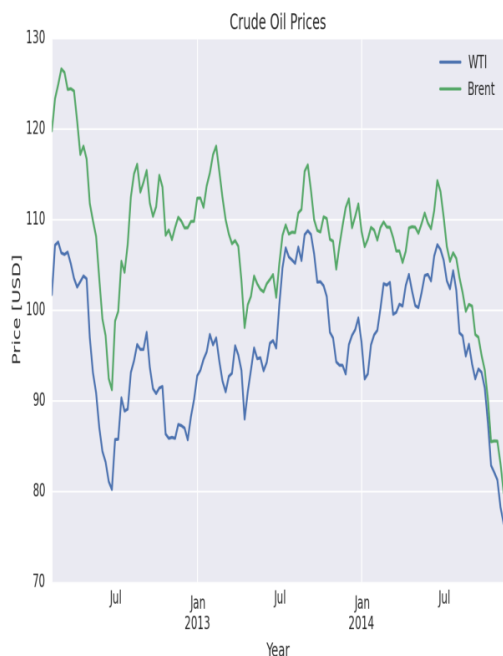
- Step 1: Create an excel file object.
- Step 2 - Parse the first sheet.
- Step 3- Rename the columns to of test file
- Step 4- Cut off the first 18 rows because these rows contain NaN values for the Brent prices.
- Step 5- Index the data set by date.
- Step 6- Remove the date column after re-indexing.
- Step 7- Visualize the data with seaborn.

	A	B	C
1		<b>Data 1: Crude Oil</b>	
2	Sourcekey	RWTC	RBRE
3	Date	Cushing, OK WTI Spot Price FOB (Dollars per Barrel)	Europe Brent Spot Price FOB (Dollars per Barrel)
4	Jan-1986	22.93	
5	Feb-1986	15.46	
6	Mar-1986	12.61	
7	Apr-1986	12.84	
8	May-1986	15.38	
9	Jun-1986	13.43	
10	Jul-1986	11.59	
11	Aug-1986	15.1	
12	Sep-1986	14.87	
13	Oct-1986	14.9	
14	Nov-1986	15.22	
15	Dec-1986	16.11	
16	Jan-1987	18.65	
17	Feb-1987	17.75	
18	Mar-1987	18.3	
19	Apr-1987	18.68	
20	May-1987	19.44	18.58
21	Jun-1987	20.07	18.86
22	Jul-1987	21.34	19.86
23	Aug-1987	20.31	18.98
24	Sep-1987	19.53	18.31
25	Oct-1987	19.86	18.76
26	Nov-1987	18.85	17.78
27	Dec-1987	17.28	17.05
28	Jan-1988	17.13	16.75

Figure 2: Sample Data of WTI and Brent  
Source: Prices for Crude Oil and Petroleum Products



**Figure 3: 10 Years Crude oil prices chart of WTI and Brent.**



**Figure 4: 3 Years Crude oil prices chart of WTI and Brent**

*B. Logical Regression Model using R*

This section explains how to predict the purchasing power of a customer using logistic regression in R.

Below are the steps required to achieve this setup.

Step 1: Import the data

Step 2: Check for class bias

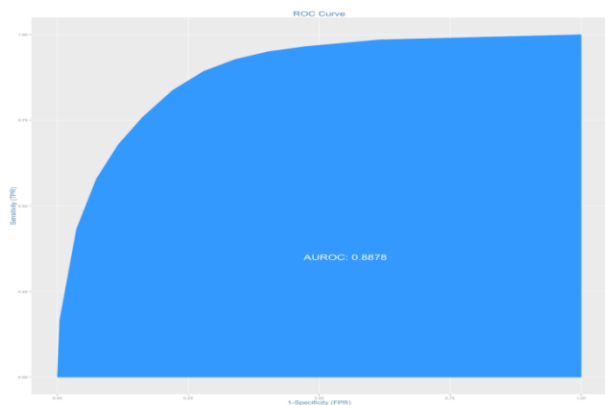
Step 3: Create training and test samples

Step 4: Compute information value to find out important variables.

Step 5: Build logit models and predict on test data

Step 6: Decide on optimal prediction probability cutoff for the model.

Step 7: Domodeldiagnostics.



**Figure 5: ROC curve**

ROC Curve (Receiver Operating Characteristics) - ROC curve, is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings.

TPR (True Positive Rate) - Sensitivity (also called the true positive rate, the recall, or probability of detection in some fields) measures the proportion of positives that are correctly identified as such (e.g., the percentage of sick people who are correctly identified as having the condition).

FPR(False Positive Rate) - The false positive rate is calculated as the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of actual negative events (regardless of classification)

**V. CONCLUSION & NEXT STEPS**

There is a buzz these days if these predictive methods really drive value or they are just an empty rhetoric. It seems clear that predictive analytics have had a positive impact on business since the early days. It is also clear that the impact will increase exponentially as data, models, methods and machine learning continue to improve based on maturity. The demand for predictive analytics will sweep across the industry and take the business to new heights. The market leaders continue to put forth their effort to transform data and apply analytics with increasing sophistication.

A brief description has given on how to develop predictive models in an industry using two of the existing algorithms (i.e. Time Series and Logistic regression algorithms) in Python and R. More predictive models can be developed based on the business scenarios by using different tools like SAS, Map/R, Python and R. The output of these predictive models can be compared with the real time data i.e. streaming from outside world into Big Data space. or batch data. Organization can use the tools like Apache Kafka and Strom for streaming process and can use any of the tools in Hadoop echo system for batch processing.

**REFERENCES**

1. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, May 2011.
2. David Turner, Michael Schroeck and Rebecca Shockley, "Analytics: The real-world use of big data in financial services", IBM Global Business Services Business Analytics and Optimization Executive Report, May 2013.
3. Vincent Charles, Tatiana Gherman, "Achieving Competitive Advantage Through Big Data", Middle-East Journal of Scientific Research 16 (8): 1069-1074, 2013.
4. Thomas H. Davenport, Jill Dyché, " Big Data in Big Companies ", International Institute for Analytics, May 2013.
5. Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey, " BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT, Business Intelligence Research
6. Elena Geanina ULARU, Florina Camelia PUICAN, Anca APOSTU, Manole VELICANU, " Perspectives on Big Data and Big Data Analytics", Database Systems Journal vol. III, no. 4/2012 3
7. Bernice Purcell, " The emergence of "big data" technology and analytics ", Journal of Technology Research.
8. Mastering Predictive Analytics with Python by Joseph Babcock Eaton, Deroos, Deutsch, Lapis, Zikopoulos, " Understanding big data: Analytics for enterprise class Hadoop and streaming data ",2012.
9. Modeling Techniques in Predictive Analytics: Business Problems and Solutions with R (FT Press Analytics) by Thomas W. Miller