What is Predictive Analytics

Edris Safari

Bellevue University, Nebraska U.S.A.

Abstract

Predictive Analytics is a popular topic of discussion and pursuit in many industries these days. The executives want to stay ahead of the game. They want to spend less money on marketing, advertising, and wasteful expenses and make more money on sales. Throw in the word Predictive followed by Analytics, and you have the attention of those executives nine out of ten times. This paper describes what PA is, what is does, what it can do and a little on how it can do what it does. It is worth noting that the how is covered in volumes of books, articles, and papers.

Keywords:

Predictive Analytics (PA)

Predictive Analytics

Analytics or mathematical analysis is a branch of mathematics that deals essentially with limits. Theories such as differentiation, integration, measure, infinite series, and analytic functions can be considered as related theories which have evolved from calculus. In analytics, we employ all the tools and techniques derived from these theories to arrive at a conclusion or prove something. This makes analytics useful in a wide range of applications.

In predictive analytics, the goal is to employ all the tools and techniques mentioned above to predict an outcome. The first question that comes to mind is how can we "predict" an outcome and then by what method and moreover using what data? These describe the predictive analytics life cycle. The answer to the first question is "we can predict an outcome by analyzing data from the past". The answer to the second question is "by employing all the tools and technique available to us (and we have many of them and the list is growing)". The answer to the final question is "your data".

With that bit of light shed on the topic, PA can be further described as employing statistical modeling and techniques to discover patterns in the historical data that with high level of confidence can predict future behavior. For example, if the historical data shows salary based on experience and the pattern shows that more experience earns more salary, the trajectory of pattern (say in a scatter plot) would show more experience predict higher salary. Throw in a few more variable such as age, education (no degree, undergraduate, graduate, etc.), number of awards, etc. and we have a project.

Predictive analytics is not an end-all part of a project, but it is an essential part of it. The identification and provision of data would precede analytics, with business intelligence

essentially beginning the project. While predictive analytics algorithms can contribute to

selection of data, its main power is utilized to predict an outcome based on the patterns that are

discovered by the algorithms.

The algorithms in predictive analytics are data-driven. In most cases, data is organized in

tabular form. The columns, features, or attributes represent the unit of analysis which address the

questions that are posed by business intelligence. The rows of the data are observations or

instances. For example, a bank customer table could have name, age, account balance, length of

membership, whether still a customer, etc. as columns. Then each row would describe a

customer. From these columns a target is selected that answers the question. This is called the

target or dependent variable. The rest of the columns are called independent variables. The

algorithms in predictive analytics can sift through the independent variables and find the best

variable(s) that can best predict the dependent variable. If the independent variable has

continuous value such as salary, or temperature and value of a stock, regression algorithms are

used. If it is categorical or classifiable, then classification algorithms are employed. During the

process of predictive analytics, models are selected that can produce the best results. There are

algorithms for that. Then the data is divided into two sets at 20 to 80, or 30 to 70 percent ratio.

One set is called the training set, or the model set and the other having a smaller portion of the

data (20-30%) is called the test set. Each set must be representative of the entire set. There are

algorithms for that as well. The selected models are "trained" using the training set, and then

using the test set a prediction is made. Algorithms to create confusion matrix that produces true

positive/negative and false positive/negative values. From these, accuracy, recall and precision of

the model is evaluated One might think that once we have our model and are confident of our

prediction, we can get paid and close the project. Well that depends entirely on the project. The

deployment and support of the solution aside, other considerations need to be made before calling the project done. In most cases results must be monitored and adjusted. In some case the analytics is an ongoing exercise due to the nature of the unit of analysis.

Conclusion

Predictive analysis is broad field of study. It addresses requirements from Aerospace and automotive to energy, retail, and financial sectors. With the advancements in computing power, PA algorithms can be implemented that can run faster and produce accurate results often in Realtime. With its footprint in pretty much any industry that has data, PA and Data Science promises to lead to advances that will provide, comfort, safety and security to the public.

References

1. Abbott, Dean. Applied Predictive Analytics . Wiley. Kindle Edition.

2. Mathematical analysis - https://en.wikipedia.org/

3. What is Predictive Analytics? Transforming Data Into Future Insights (2018). CIO

4. https://www.cio.com/article/3304301/7-secrets-to-predictive-analytics-success.html?nsdr=true#tk.cio_rs

5. https://www.cio.com/article/3343057/7-projects-primed-for-predictive-analytics.html