Time Machine: Can Data Predict the Future?

Kimberly Summerville
Bellevue University
ksummerville@my365.bellevue.edu

Cassandra Williams
Bellevue University
cwilliams@my365.bellevue.edu

Abstract
Ever wondered how an email service knows an email is spam before you've even opened it or marked it as such? Or how your credit card company can spot fraudulent charges and alert you before you're even aware? This is due to predictive models put in place that learn patterns and spot the outliers. This type of technology used to be exclusive to mathematicians and statisticians, but with the wave of big data and computerized transactions, it has branched out to new sectors, including, but not limited to email services and banks. We'll take a look at these old, yet new, algorithms and find out how it has affected our world today.

Author Keywords
Predictive analysis; Models; Time Series

Introduction
For many average investors, the stock market has been a territory of the fear and uncertainty due to the black void of a future each investment has. The ongoing joke that stock brokers are actually gamblers is not entirely far from the truth. When looking to the past, we see sudden sharp declines that very few saw coming, with the aftermath being seen as a national crisis. Despite the abundance of numerical patterns, the stock market

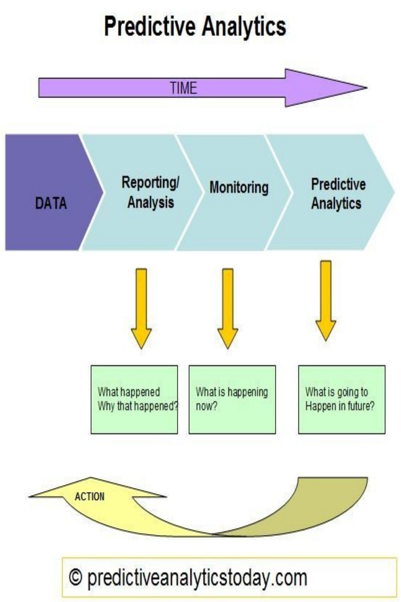

Figure 1.1 Process of Predictive Analytics

is one of the more difficult predictive challenges. The ability to accurately predict stock market activity has always been a goal of been of stock investors. The desire to avoid the ramifications of these dips, and possibly even benefit from them, has caused a surge of research into algorithms performing predictive analysis using data from the stock market. Algorithmic trading has already had an effect on the market, being directly responsible for the Flash Crash (a brief swing of over 1,000 points on May 6, 2010) in the Dow Jones Industrial Average. In this project, we will give an overview of some major popular models used in stock market predictive analysis and choose one to dive deeper into. Our goal is to analyze the accuracy and accessibility of these various methods and how it can affect an average investor's plans for investments.

Why Data Science?

Prediction Analysis is at the core of data science. It is taking the data we are given, analyzing the correlations and causes and uses that information to make decisions that are most likely to give future benefit to the user in a business project, large purchase, or even marketing strategy. The application in the stock market is just one example of how data science is becoming ingrained in society. The option for average consumers to take advantage of these advanced algorithms shows the analysis of data is transforming into an every day occurrence and accessible to all.

Starting the Predictive Analytics Modeling Process

Each predictive analytics model is composed of several predictors, or variables, that will impact the probability of various results. Before launching a predictive modeling process, it's important to identify the business objectives, scope of the project, expected outcomes, and data sets to be used.

Data Collection and Mining

Prior to the development of predictive analytics models, data mining is typically performed to help determine which variables and patterns to consider in building the model.

Prior to that, relevant data is collected and cleaned. Data from multiple sources may be combined into a common source. Data relevant to the analysis is selected, retrieved, and transformed into forms that will work with data mining procedures.

Mining Methods

Techniques drawn from statistics, artificial intelligence (AI) and machine learning (ML) are applied in the data mining processes that follow.

AI systems, of course, are designed to think like humans. ML systems push AI to new heights by giving computers the ability to "learn without being explicitly programmed," said renowned computer scientist Arthur Samuels, in 1959.

Classification and clustering are two ML methods commonly used in data mining. Other data mining techniques include generalization, characterization, pattern matching, data visualization, evolution, and meta rule-guided mining, for example. Data mining methods can be run on either a supervised or unsupervised basis.

- Also referred to as supervised classification, classification uses class labels to place the objects in a data set in order. Generally, classification begins with a training set of objects which are already associated with known class labels. The classification algorithm learns from the training set to classify new objects. For example, a store might use classification to analyze customers' credit histories to label customers according to risk and later build a predictive analytics model for either accepting or rejecting future credit requests.

- Clustering, on the other hand, calls for placing data into related groups, usually without advance knowledge of the group definitions, sometimes yielding results surprising to humans. A clustering algorithm assigns data points to various groups, some similar and some dissimilar. A department store chain in Illinois, for example, used clustering to look at a sale of men's suits. Reportedly, every store in the chain except one experienced a revenue boost of at least 100 percent during the sale. As it turned out, the store that didn't enjoy those revenue gains relied on radio ads rather than TV commercials.

The next stage in predictive analytics modeling involves the application of additional statistical methods and/or structural techniques to help develop the model. Data scientists often build multiple predictive analytics models and then select the best one based on its performance.

After a predictive model is chosen, it is deployed into everyday use, monitored to make sure it's providing the expected results, and revised as required.

Types of Prediction Analysis Models

When performing predictive analysis, there are five main types of models used:

- Classification Models, which are used for binary questions, where the answers are often “yes” or “no” (i.e. Will this stock go up?)
- Clustering Models that sort data into separate smart groups based on similar attributes
- Forecast Models which deal in metric value prediction, estimating numeric value for new

  data based on learnings from historical data (i.e. What will the price of a particular stock be in 3 months?)
- Outlier Models that are oriented around irregular data entries within a dataset and can identify anomalous figures either by themselves or with other numbers and categories
- Time Series Models that utilize a sequence of data points, using time as the input parameter and uses the last year of data to develop a numerical metric and predicts the chosen time period using that metric.

Each model has its merits and contribute different information to the overall analysis. The challenge for data scientists can be deciding which model will provide the most insight for the problem that they are trying to solve. A particularly interesting model is the Time Series Model due to the complex nature of it and more consideration is being paid to it in order to understand it better.

Time Series Model

Many people are looking at the problem. It is hugely complicated, and nobody really knows what type of sentiment analysis will work. Many companies are beginning to use artificial intelligence and machine learning to provide daily investment forecasts based on a combination of big data analysis, financial modeling, and sentiment analysis.

For example, the Israeli company I Know First's algorithm is based on artificial intelligence, machine learning and incorporates elements of artificial neural networks as well as genetic algorithms to model and predict the flow of money between markets. It monitors more than 10,000 assets on 6-time horizons spanning from 3-days to a year including stocks, ETF's, world indices, gold, currencies, interest rates, and commodities.

The system outputs the predicted trend as a number, positive or negative, along with the wave chart that predicts how the waves will overlap the trend. This helps the trader decide which direction to trade, at what point to enter the trade, and when to exit. The model is 100% empirical, meaning it is based on historical data and not on any human derived assumptions. The human factor is only involved in building the mathematical framework and initially presenting to the system the "starting set" of inputs and outputs. From that point onward, the computer algorithms take over, constantly proposing "theories", testing them on years of market data, then validating them on the most recent data, which prevents over-fitting. If an input does not improve the model, it is "rejected" and another input can be substituted.

Conclusion

Predictive models and analytics are powered by several different models and algorithms that can be applied to a wide range of applications. The stock market is a ripe opportunity for this technology, as there are no "sure things". There are many models and algorithms to use, but not all are accessible to the public. For a notice analyst, a classification model may be the way to go, to keep things simple. However, outlier models have proven to be more tailored for the finance and investing sector.

References

1. An Integrated Approach to Predictive Analytics for Stock Indexes and Commodity Trading Using Computational Intelligence. ;World of Computer Science & Information Technology Journal; 2015, Vol. 5 Issue 8, p142-148, 7p

2. Model and forecast stock market behavior integrating investor sentiment analysis and transaction data. ; Cluster Computing. Mar2017, Vol. 20 Issue 1, p789-803. 15p.

3. Trading Volume as a Predictor of Market Movement; International Journal of Finance & Banking Studies, Vol 8, Iss 2, Pp 57-69 (2019)

4. Analysis of temporal pattern, causal interaction and predictive modeling of financial markets using nonlinear dynamics, econometric models and machine learning algorithms. ; Applied Soft Computing; Sep2019, Vol. 82, pN.PAG-N.PAG, 1p

5. Fuzzy support vector regression model for forecasting stock market volatility. ; Journal of Intelligent & Fuzzy Systems. 2016, Vol. 31 Issue 3, p1987-2000. 14p.

6. Understanding the predictive power of social media. ; Internet Research; 2013, Vol. 23 Issue 5, p544-559, 16p

7. Detection of financial rumors using big data analytics: the case of the Bombay Stock Exchange. ; Journal of Organizational Computing & Electronic Commerce; 2018, Vol. 28 Issue 2, p79-97, 19p

8. A data analytic approach to forecasting daily stock returns in an emerging market. ; European Journal of Operational Research. Sep2016, Vol. 253 Issue 3, p697-710. 14p.

9. Brown, D. E. (2015). Predictive analytics: the macro view. Los Alamitos, CA: IEEE Computer Society.

10. Karim, M. R. (2017). Predictive analytics with TensorFlow: implement deep learning principles to predict valuable insights using TensorFlow. Birmingham: Packt Publishing.

11. Siegel, E. (2016). Predictive analytics: the power to predict who will click, buy, lie, or die. Hoboken, NJ: John Wiley & Sons.

12. What Is Predictive Analytics? 3 Real-World Examples of ... (n.d.). Retrieved from https://www.logianalytics.com/predictive-analytics/what-is-predictive-analytics/