History and Applications of Predictive Analytics

Jaden Phung
Alok Shah



APA Research

July 2025

Overview

Marketing Analytics is the process of tracking and analyzing data in order to reach a quantitative goal. These analytics provide companies with the evidence that shows their inner processes are actively achieving their goals. However, companies want concrete proof that their strategies will be successful before they put their time and energy into them. Predictive analytics allows for the use of data to predict these future trends and events.

History

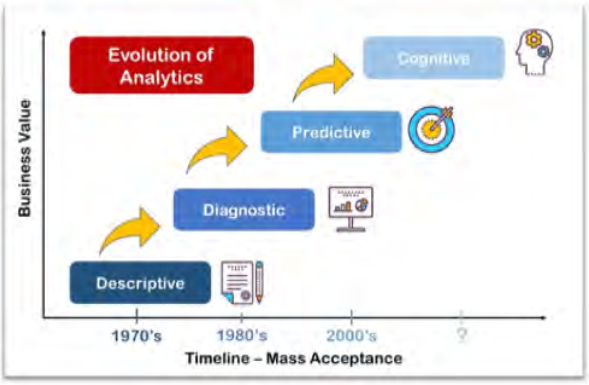
How did we get to predictive analytics? The history of predictive analytics can be seen through four distinct phases.

Phase 1: Descriptive analytics - comprised the creation of basic models of linear programming and computational models to help further understand “what happened.”

Phase 2: Diagnostics - investigated why something happened and helped analysts find important variables.

Phase 3: Predictive - investigated “what will happen,” which could only be done with the invention of mathematical models that allow for computers to predict outcomes based on data.

Phase 4: Cognitive - focused on increasing the accuracy of predictive models without human input.



Necessary Developments

Linear programming - A technique used in programming to find the best outcome while satisfying the given set of parameters.

Computational Modeling - Using computers to run complex systems using math and logic.

Monte Carlo simulation - A technique for estimating an outcome that can be altered due to random variables.

Nonlinear programming - Finding a solution for problems where the variables have nonlinear relationships.

Generative AI - Creates content based on recognizable patterns. Allows for quick predictions that continue to evolve with additional data.

Machine Learning Techniques – Techniques that allow for AI to learn from various sets of information.

Moore's Law - The number of components on a single chip doubles every two years at minimal cost. Led to increases in computing power that allow for analytical programs.

Regression – The statistical method used to determine the structure of a relationship between two or more variables.

Applications of Predictive Analytics:

Predictive analytics have been more widely used recently by businesses to predict the deterioration of a company's machines, forecast future cash flows, or estimate the stock price in five years. According to McKinsey, 64% of B2B companies are investing in predictive analytics software and tools to tackle some of their largest problems:

1. **Determining Staffing Needs:** Caesar's Entertainment used predictive analytics to determine staffing needs for a venue on any given day based on the events and activity level. This became fundamental to keeping their costs low, as overstaffing is a costly mistake and understaffing can cause a whole host of operational issues. The company developed a multiple regression model to predict the amount of foot traffic and based their staffing levels on this.
2. **Predicting Allergic Reactions:** AbbieSense uses predictive analytics that measure physiological signs and then automatically administers epinephrine if needed. The data is collected by a wearable that houses the medicine too. The algorithm can predict it much faster than human detection and could save many lives.
3. **Fraud Detection:** According to KPMG, financial institutions that utilize predictive analytics can detect fraud 60% more accurately than before. This improvement is equivalent to $15 billion in prevented false transactions. Predictive analytics help save these banks money and makes it more secure to open a credit card or bank account with a certain institution.

While these applications are wide-ranging and can be applied almost anywhere, one of the largest limitations of predictive analytics is the **quality of data**.

Analytics Horror Stories:

Out-of-control Home Buyer
Zillow Offers' iBuying platform, designed to automatically purchase homes for flipping, misvalued them, resulting in a $300 million loss in one quarter, and necessitated a 25% workforce reduction. Zillow purchased $2.8 billion worth of homes, and 60%+ were undervalued.

The Golden Water Bottle
A sports store sold a water bottle for $1 million, but the computer system still thinks someone will spend that much, and as a result, no water bottles were sold for a year. Another company's predictive pricing system decided to price thousands of products at $0.01, causing a rush of purchasing.

Don't Forget Units
In 1998, NASA's $125 million Mars Climate Orbiter was lost shortly after launch because of a miscalculation due to units. The Jet Propulsion Lab, who handled navigation, used metric units, while the construction team at Lockheed Martin used imperial units.

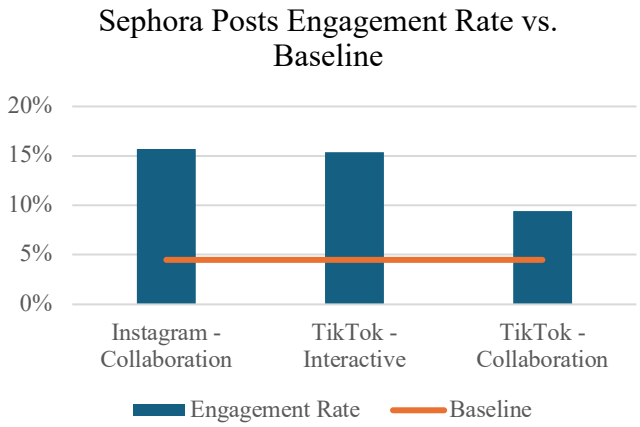
These examples show the importance of ensuring strong data to use predictive analytics properly. Since these models are built on historic information, data must be representative, accurate, and frequently updated to create realistic and useful predictions. One of the easiest places to get accurate data easily is social media. We hypothesize that this same predictive technology can be applied to social media marketing analytics to help businesses make decisions about content, language, and posting schedules.

Marketing Analytics Example:

Sephora has amassed nearly 50 million followers across social media platforms, more than three times the next closest competitor, and has engagement rates that reach over 15% per post. What does the beauty giant do differently than other companies? Sephora tracks social media trends (hashtags, mentions, product reviews), feeds the data into predictive algorithms, and then makes marketing decisions based on the insights. Using this information, they can adjust inventory, launch timely campaigns, and recommend trending products in-store and online.

1. Tracking What Works

Sephora's social media presence is full of different initiatives and ideas, from partnerships and influencers to memes and user-generated content. Each of these ideas is anticipated to attract a certain target audience; however, the actual success may be completely different. Based on historical data, Sephora stopped posting on X (Twitter) altogether and focused its efforts on TikTok, which is generating on average 12,500+ engagements per post. As shown to the right, Sephora has refined their posts to see certain collaborations. They are also using interactive filters and features to drive their engagement rates over 330% higher than the industry average. Tracking also shows what doesn't work, just like a meme-style TikTok they posted, garnering only a 0.023% engagement rate.



2. Optimizing for Certain Metrics

Sephora investigates and predicts each social media metric (i.e. likes, shares, comments, reposts, follows) separately as each has a different meaning and importance. The most important for their marketing strategy is shares, as it indicates virality. For example, beauty filter user-generated content was shared over 53,000 times, helping boost the product to more people.

3. Sentiment Analysis of Comments and Reviews

Finally, Sephora uses natural language processing and sentiment analysis to determine the cumulative opinion of its loyal customers. The model determines what phrases may attract

customers strongly, using verbiage on social media posts like “gentle but effective” or “glowy finish.”

[redacted] - Marketing Analytics Application:

Methodology:
To explore the potential effectiveness of these marketing techniques for [redacted], we downloaded the past three years of X (Twitter) data from the company, including information about the posts’ content, likes, shares, comments, and quote posts. We used this data to create a predictive model to see what components of a post could predict engagement.

Model: The model was created using RStudio by fitting a multiple regression predicting the number of engagements a post would receive based on the month, the hour, photos, tagged users, hashtags, URLs, and videos. The number of engagements was log-transformed to account for the skewed data. The model uses cross-validation to split the data into training and testing sets.

Results: The multiple regression model has an R^2 of 0.209 and a residual mean standard error of 1.73 engagements. This means that the model accounts for 20.9% of the variation within the data set and predicts engagements within a test set with an average error of 1.73. Considering that social media engagements are very random and are heavily dependent on the context and language used in the post, being able to include elements that drive engagement can help boost posts that the company wants.

Potential Recommendations: The most influential predictors of the model are photos (+1.49 engagements), external image URLs (+1.50 engagements), and videos (+2.00 engagements). Less significantly, having external URLs decreases engagement by 1.13. Considering that the average number of engagements is 19.61, these potential increases are notable.

1. **Include relevant images for posts that are positive.**

Both images and links to images have shown positive effects on engagement rate. Most analysis needs to be done to determine the type of image or elements within the image that drive this success. This insight does make sense as people who scroll will tend to linger on posts with images longer, especially on a primarily text-based social media platform like X.

2. **Regularly post videos, potentially with executives or other employees.**

Video posts are common for [redacted], with 44% of the posts from the sample data including video content. Due to this, continuing to regularly post content like this that has done well in the past is important. [redacted] tends to post interviews with executives, but more research is needed to see if this has a significant impact on engagement.

3. **While articles may decrease engagement, viewership is unknown.**

External URLs tend to decrease engagement most likely because users will click the link and forget to return to the app and like the post. As a result, this data is biased, as the true measure of success is clicks. By using trackers, the company can collect data to see where visitors are

finding the link. If this percentage is significantly high, posting external URLs will be advantageous. Since 52.2% of [redacted] posts include external URLs, it can be assumed that posting links drives viewership.

Closing Insights:

[redacted] **Recommendations**:

- Including relevant images and a regular schedule of posts is correlated with an increase in engagement.

- External URLs may influence the decrease in engagement as it does lead users away from the page.

Social Recommendations:

- Short-form content platforms such as TikTok do drive higher engagement as the average viewer's attention span continues to decline.

- Track social media trends (hashtags, mentions, product reviews) and feed the data into predictive algorithms.

Important Considerations:

- Generative AI is an important tool for predictive analytics and has replaced more traditional tools for data analytics.

- Models should be built on historic information, and data must be representative, accurate, and frequently updated to create realistic and useful predictions. One of the easiest places to get accurate data is social media.

.

- Predictive Analytics has saved companies money through fraud detection, forecasting future cash flows, and increasing social media engagement.

Sources

5 Ways Sephora is Using AI. (n.d.). Digitaldefynd.

https://digitaldefynd.com/IQ/sephora-using-ai-case-study/

7 real-world examples of how brands are using Big Data analytics. (2021, February 10). Bornfight.

https://www.bornfight.com/blog/7-real-world-examples-of-how-brands-are-using-big-data-analytics/

14 Data & Analytics Bloggers to Follow Today. (2021, July 22). Foundry.

https://foundryco.com/blog/data-analytics-bloggers-to-follow-today/

A Brief History of Predictive Analytics. (2019, May 6). Medium.

https://medium.com/@predictivesuccess/a-brief-history-of-predictive-analytics-f05a9e55145f

A Brief History of Predictive Analytics – Part 1. (n.d.). Afterinc.

https://afterinc.com/2018/12/28/brief-history-predictive-analytics-part-1/

Data quality horror stories and how to avoid experiencing them. (2024, May 1). Great Expectations.

https://greatexpectations.io/blog/data-quality-horror-stories-and-how-to-avoid-experiencing-them

How Sephora pairs individual, loyalty data to optimize segmentation. (n.d.). Retail Dive.

https://www.retaildive.com/ex/mobilecommercedaily/how-sephora-leverages-loyalty-data-to-optimize-segmentation?utm_source=chatgpt.com

Insights to impact: Creating and sustaining data-driven commercial growth. (2022, January 18). McKinsey.

https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/insights-to-impact-creating-and-sustaining-data-driven-commercial-growth

Just the Numbers: The Social Media Engagement Playbook. (2024, July 23). Beauty Matter.

https://beautymatter.com/articles/the-social-media-engagement-playbook

Sephora Social Media Strategy: Case Study. (2025, June 27). Enrichlabs.ai.

https://www.enrichlabs.ai/case-study/sephora-social-media-strategy

Sephora's Marketing Strategy: What Sets the Beauty Giant Apart in 2025. (2025, May 13). Rival IQ.

https://www.rivaliq.com/blog/sephora-marketing-strategy/?utm_source=chatgpt.com

The Power of Social Media Analytics: Case Study of Coca-Cola. (2023, August 10). Almabetter.

https://www.almabetter.com/bytes/articles/case-study-of-coca-cola

What Is Predictive Analytics? 5 Examples. (2021, October 26). Havard Business School Online.

https://online.hbs.edu/blog/post/predictive-analytics

Zillow will stop buying and renovating homes and cut 25% of its workforce. (2021, November 3). NPR.

https://www.npr.org/2021/11/03/1051941654/zillow-will-stop-buying-and-renovating-homes-and-cut-25-of-its-workforce