

Predictive analytics tools and their business applications

Originally published at <https://www.itransition.com/blog/predictive-analytics-tools>

Published: May 25, 2021 | By Martin Anderson Independent AI expert



Predictive analytics offers companies the ability to foresee future trends and events based on historical data from a number of possible sources, including metadata from communications systems, customer transaction histories, in-house reporting systems, customer demographics and, increasingly, a wide range of possible external or secondary data streams.

The advent of big data over the last twenty years has reinvigorated the predictive analytics sector: according to the 2021 Facts and Factors report, the predictive analytics market is set to grow at a 24.5% CAGR to around \$22.1 billion by 2026.

Arguably, this growth is occurring because enterprises are compelled to innovate in a market that has become more than usually political and economically volatile in the last 12-13 years since the advent of austerity—a climate that would normally incline a risk-averse attitude.

However, the pace of development and the level of innovation in the capabilities of [predictive analytics software](#) over the last 5-7 years have so transformed the potential of automated forecasting systems as to offer a potentially compelling advantage, notwithstanding the current market challenges.

In this article, we'll look at a few of the areas in which predictive analytics can help an organization to become more profitable and efficient. Since each industry and sector has diverse needs and diverse methods of obtaining the data that it needs for an effective predictive analytics system, we'll examine some example use cases from specific sectors, and derive some core lessons in each case.

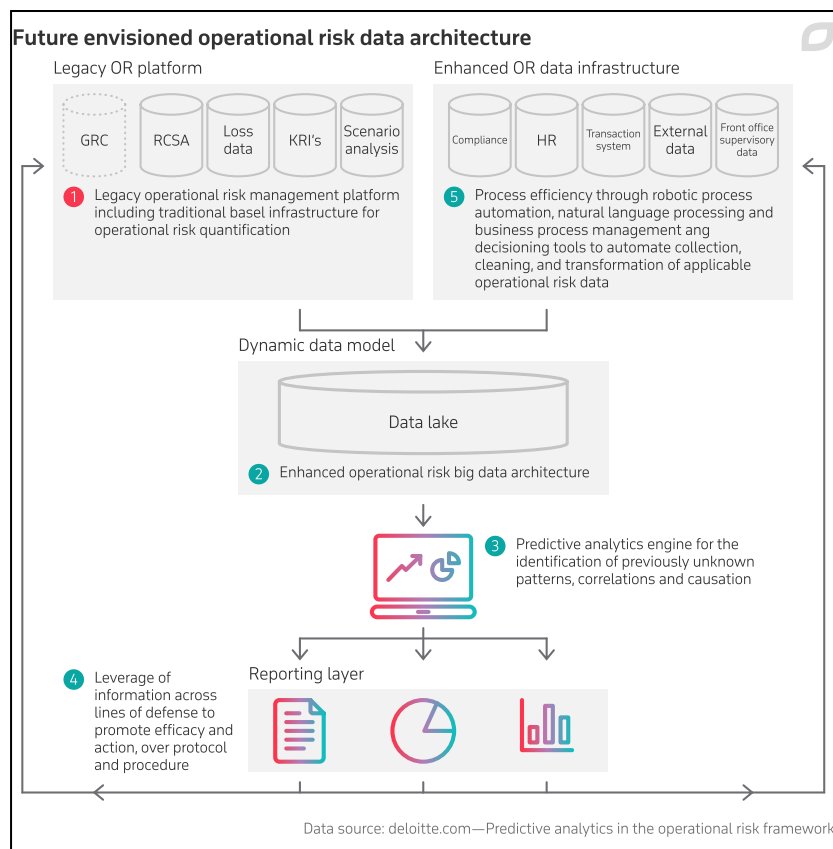
'Parallel' data sources in predictive analytics

In nearly all traditional architectures for predictive analytics systems, the process of refining data into insights involves the deduction of probabilities based on historical outcomes. This principle has become engrained over the centuries because, except in rare cases of 'inside information', it was all that was available. Though this is the established solution, it is, as of relatively recently, no longer the only solution.

The 'phantom menace' in analytics

In their report on predictive analytics in the operational risk framework, Deloitte characterizes less 'predictable' risk factors as a 'phantom menace' and suggests that the advantage of applying machine learning to analytics lies not just in distinguishing patterns from vast volumes of in-house business data (though this is useful), but also in side-loading other types of analytics, such as marketing statistics and openly available data sources, to build up a more complete picture of risk.

Deloitte suggest that operational risk models should be reassessed to account for parallel data streams from other areas, with both the 'traditional' model and a more modern data architecture contributing to a pooled resource that can be iterated through for a truly dynamic and versatile risk prediction system.

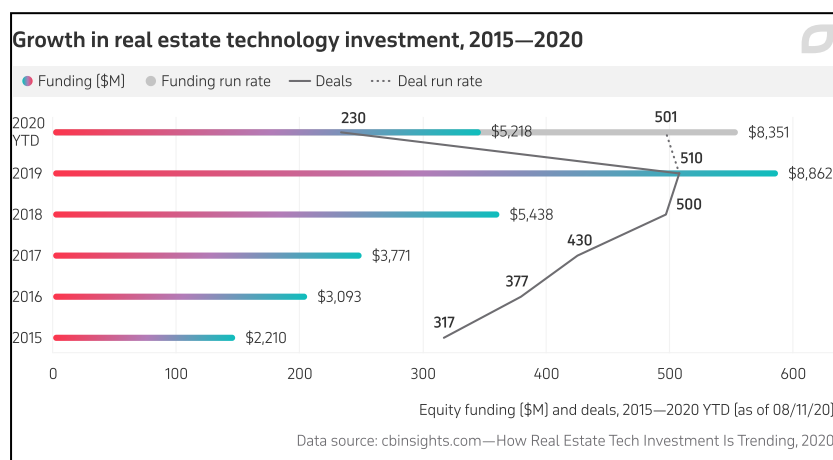


Though sectors with a requirement for high operational security (such as financial tech) maintain a circumspect attitude to such innovative approaches, less sensitive markets are able to experiment freely and adopt a vanguard position.

Predictive analytics for market insights

The advent of big data and the resurgence of machine learning have brought notable growth to the property technologies (PropTech) sector, with a new era of [predictive analytics in real estate](#) and new types of data that provide greater and more granular insight than their predecessors.

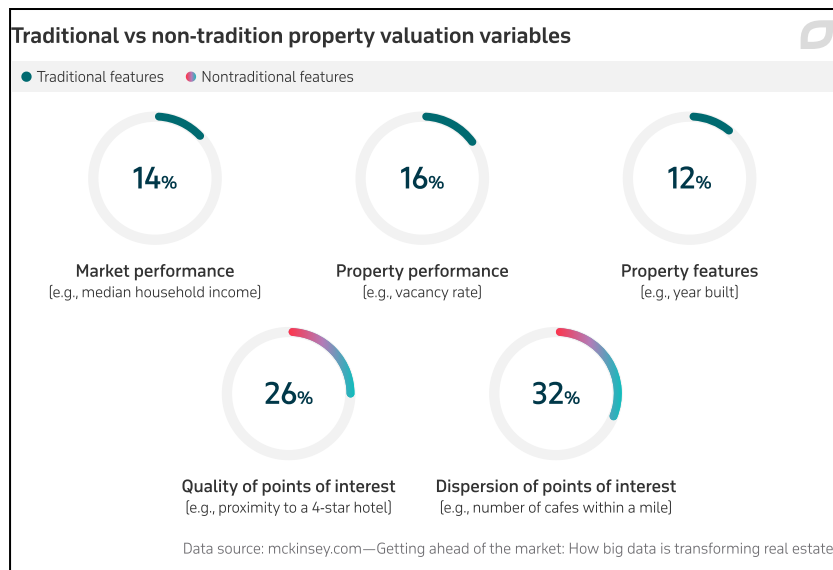
According to private equity company CB Insights, PropTech funding grew by 65% from \$5.4 billion in 2015 to \$8.9 billion in 2019, with startup investment quadrupling in the same period, and \$8.9 billion invested across 500 deals in 2019 alone.



Though the pandemic impacted this trend hugely from late winter of 2020, even this catastrophic event has given rise to new innovation.

Non-traditional variables on the rise in real estate predictive analytics

A report from McKinsey in 2018 found that nearly 60% of the predictive power underlying the new breed of analytics systems are coming from non-traditional variables.



The rise in [IoT use cases](#) and data source development over the last fifteen years offers a new wave of data streams that can feed into machine learning analytics systems. In the case of PropTech, examples of non-traditional data include the presence of local facilities, the proximity of cafes (or planning applications for cafes from major chains), rent per-square-foot measured against an average for a zip-code, availability of restaurants and gas stations, and many other data facets that have come online via map SDKs, as well as innovation in civic and commercial data sources.

AI addresses the 'X-factor' for real estate purchasers

New-York startup Localize.city employs hundreds of AI professionals, in addition to cartographers and urban planners, in order to provide a wide array of lateral information about potential home purchases and rentals in NYC.

Factors taken into consideration include the presence of vibrations from local train tracks and other installations; traffic throughput; and the likelihood of new construction projects, among many other apparently imponderable factors. The system can even provide data on access to sunlight, based on the height and disposition of adjacent buildings.



The app infrastructure of Localize has a direct effect on investment analytics in the NYC area, since it includes a CRM hook allowing realtors to visualize the current level of search activity. The proprietary service, free to NYC residents, performs real-time analysis across thousands of machine learning models.

Key consideration for developing market insights

Though predictive analytics approaches for generating market insights will vary greatly across each industry and sector, there are some common principles that we can observe from the PropTech sector:

- Re-evaluate the features derived from your existing internal data sources frequently to develop new approaches to developing insights from your data.
- Consider how proprietary company data could be correlated against valuable insights from external data sources, commercial and open-source (and, to this end, stay aware of new developments in data sources in your sector).

Consider how the addition of new data may change the consistency and applicability of year-on-year metrics that have become a staple of company reporting, and whether it is worth a 'reset' of the criteria for the company's business analytics.

Predictive analytics for fraud detection

Predictive analytics is widely used for fraud detection in financial processing architectures. Most of us have at some point experienced this directly, through triggering a fraud alert for a transaction that does not fit our bank's historical profile of how much we spend, what we spend it on, or in which locations we spend it.

Besides this highly personalized model of 'normal' behavior, predictive analytics systems may use heuristic approaches, where 'outlier' transaction events can indicate fraud, even though those actions may be relatively consistent with normal customer use. In general, this 'fuzzy logic' methodology is giving way to insights from [machine learning in banking](#), with each customer history generating a decision tree that will be run against their attempted actions.

This is a broad principle in fraud detection. It's applicable to a wide variety of use cases, such as pharmaceutical fraud, insurance fraud, deceptive retailer returns, tax fraud—and even in the fight against deepfakes.

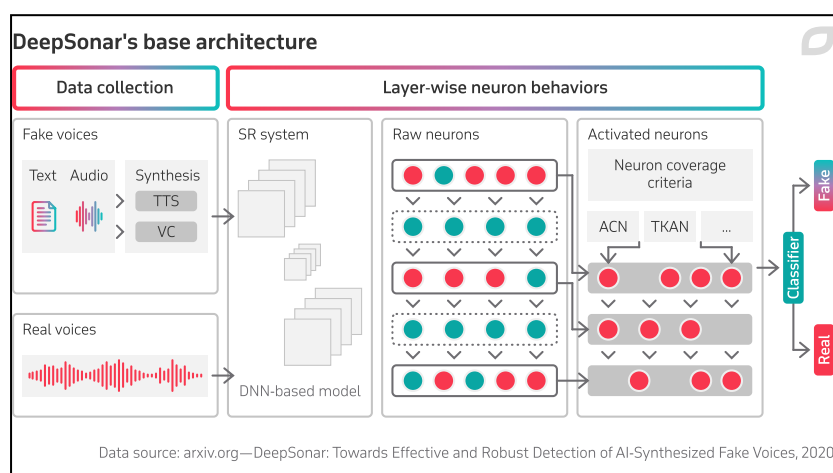
Combatting deepfake audio fraud

The potential use of deepfake technologies to clone a user's voice has garnered headlines in the last couple of years. A user's voice can be simulated by obtaining a variety of audio samples and extensively training a machine learning model so that it can be made to imitate them in audio conversations.

Deepfake audio fraud has undermined the recent trend in banking for user authentication via voice ID, and the fact that the analogue delivery channel is so susceptible to this kind of attack (as opposed to video deepfakes, which invite far more scrutiny and occur in a much more critical context) makes audio deepfake predictive analytics one of the hottest research sectors in this sphere.

In 2019, fraudsters used this technology to cheat the CEO of a UK energy firm out of €220,000 (\$243,000 USD), as reported in The Wall Street Journal. Though a later attempt in 2020 was less successful, the quality of deepfake audio is constantly improving, and end users are not yet adequately sensitized to the possibility of deception in this social context.

A number of initiatives are being developed to detect deepfake voices, but the development scene for deepfake audio is still quite nascent, and incidents low, so far. In 2020, a collaboration between Chinese universities resulted in DeepSonar, obtaining a 98.1% error detection rate for the English and Chinese languages.



The advent of audio imitation by machine learning has led to the creation of the Automatic Speaker Verification Spoofing and Countermeasures Challenge, which invites 'white-hat' deepfakers to fool the organization's AI-driven detection measures.

Key considerations for predictive analytics fraud detection systems

- Unless you have achieved compelling market capture, it's important to establish a balance between good security and customer experience. Consider an extended 'mirror' trial of any new system to ascertain an acceptable level of false positives.
- Check existing and new fraud detection predictive analytics systems for implementations that rely on cross-domain tracking, now arguably a dying technology.

In developing the project, familiarize yourself with the most common type of outlier detection methodologies and algorithms, such as standard deviation, DBScan Clustering, boxplots, Robust Random Cut Forest, and Isolation Forest. Some of these work better with high-dimensional data,

while others are more granular and dependent on data structure.

Predictive customer risk analytics

Viewed from the highest level of analysis, 'customer risk' encompasses a wide range of possible applications for predictive analytics systems, from credit-scoring, anti-money-laundering measures, and risk assessment through to security background checks and insurance underwriting. However, as with other types of predictive analysis systems, the core principles of [risk analytics](#) when it comes to customers center on projecting forward the results of historical analysis.

Since the field of healthcare takes in a wide range of possible applications, from medical risk assessment through to fraud and insurance, it's a good example of the benefits that [predictive modeling in healthcare](#) can bring to a diverse range of requirements.

Reducing rates of patient readmission

Despite criticism from Congress, in 2018 Medicare-imposed penalties to hospitals for patient readmission within a prescribed period from being discharged were estimated by the US government to exceed half a billion dollars. In 2021, it [was estimated](#) that Medicare penalizes half of all participating hospitals for cases where patients are readmitted within 30 days of leaving the hospital.

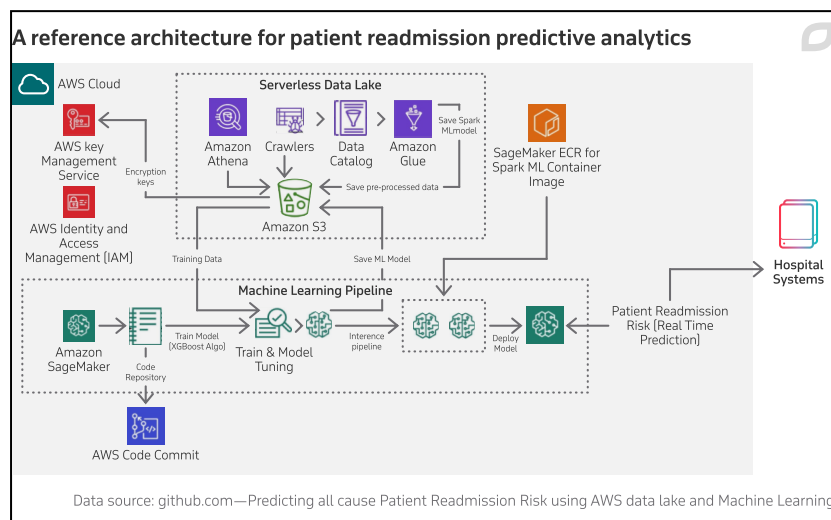
Additionally, the Agency for Healthcare Research and Quality (AHRQ) asserted in 2020 that episodes of readmission are among the most expensive to treat, estimating the annual cost in the United States at \$41.3 billion.

A research paper from the University of Michigan, released in February of 2020, found that risk prediction models based on a patient's electronic medical records [offer better performance](#) in predictive analytics systems than the use of administrative data, which has better current detail but no long-term context.

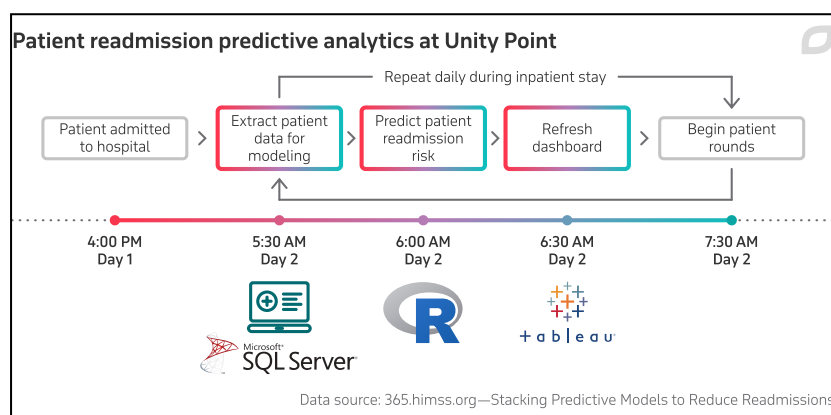
The good news about patient readmission figures is that the problem is such a critical focus of research that, unusually, addressing it with machine learning is becoming almost routine.

Synthetic data as an aid to predictive analytics in healthcare

AWS Senior Solutions Architect Anuj Gupta has demonstrated a HIPAA-compliant proof-of-concept workflow for a readmission predictive analytics architecture using synthetic data that could be substituted with real patient data in the development of healthcare frameworks:



The Iowa hospital network Unity Point realized a 40% improvement in 30-day patient readmission rates by modeling a batch-scoring predictive analysis pipeline upon a patient's admission to hospital. Written in R, and utilizing data from Tableau and Microsoft SQL server, the framework also re-calculates the risk factor when patients do not show up for scheduled follow-up appointments that were made prior to their discharge from hospital.



Key considerations for predictive customer risk analytics

- It's wise to place security and regulatory compliance above all other considerations during development of in-house systems. Where assembling cloud-based pipelines, prioritize providers that are capable of delivering geo-specific privacy and regulatory auditing, and which offer mature compliance frameworks for your sector.
- Choose machine learning methodologies and algorithms that support your data architecture and available methods of verification, rather than trying to adapt a popular new analytics framework to an unsuitable data model.

Resist deadline-driven rollouts, but rather trial the system in a mirrored virtual environment at least up to the point where repetition has proved its efficacy and its ability to at least match your current risk assessment solutions.