WHITE
PAPER

David Speights, PhD
Vishal Patel, PhD
Alan Gillette, PhD
Sruthi Shyamsunder, MS
Adi Raz, MBA
Appriss Retail.

In collaboration with Chris Hoffman,
Director of Corporate Investigations
& Loss Prevention Systems,
The Bon-Ton

"In our testing, we saw more than a 25% increase in fraud cases identified and a 25% decrease in time required to track down the fraudulent activity within many of the cases."

—Chris Hoffman, *The Bon-Ton*

A CASE STUDY WITH THE BON-TON

Machine Learning and the Evolution of Exception Reporting

Introduction

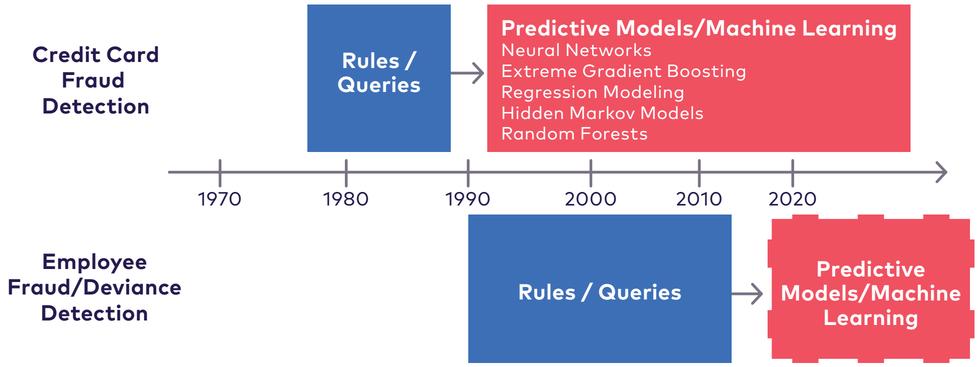
Recent years have seen an explosion of consumer products that incorporate artificial intelligence (AI). Self-driving cars, Alexa, Siri, chatbots, Google search, and smart traffic lights are examples of how artificial intelligence is impacting the world. While artificial intelligence has gained mainstream attention recently, AI applications have been "hidden" for many years in a variety of business applications such as: credit scores, credit card fraud detection, direct marketing, identity theft protection, oil exploration, and more.

Retail loss prevention has also seen applications of AI in technologies such as facial recognition and return fraud authorization. In this white paper, we extend the loss prevention professionals toolkit and show the results of a machine learning algorithm (a specific type of AI) developed to detect fraudulent and abusive employee activity for The Bon-Ton.

The ultimate objective for developing machine learning models is not to replicate a single retailer's successful investigations. Rather, it is to combine many learnings and obtain collective intelligence to curb employee deviance. Appriss Retail serves more than 90 retailers worldwide, and this experience provides as a solid foundation for such model development.

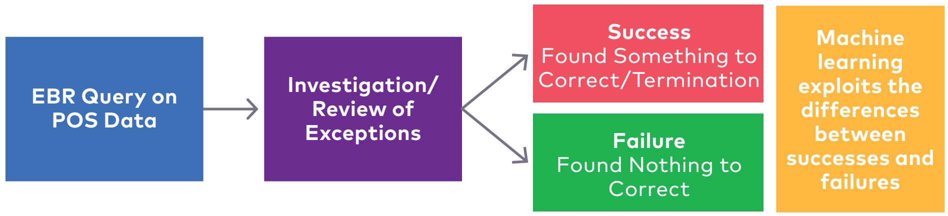
From Rules and Queries to Predictive Models and Machine Learning

Today, most Loss Prevention departments use traditional exception based reporting (EBR) tools to monitor their employees. Exception based reporting is usually viewed as a query run on point-of-sale (POS) data that flags suspicious activities. EBR tools used by retailers are similar to the tools used by the credit card industry in the 1980s to detect fraud. In the early 1990s, the credit card industry mostly abandoned the exception based reporting approach in favor of methods of detection based on predictive modeling and machine learning (see timeline below). A similar trend was seen with healthcare fraud about 15 years ago when the industry's tools converted from rules and queries to predictive modeling. Our prediction is that retail loss prevention will evolve in a manner comparable to credit card fraud detection.

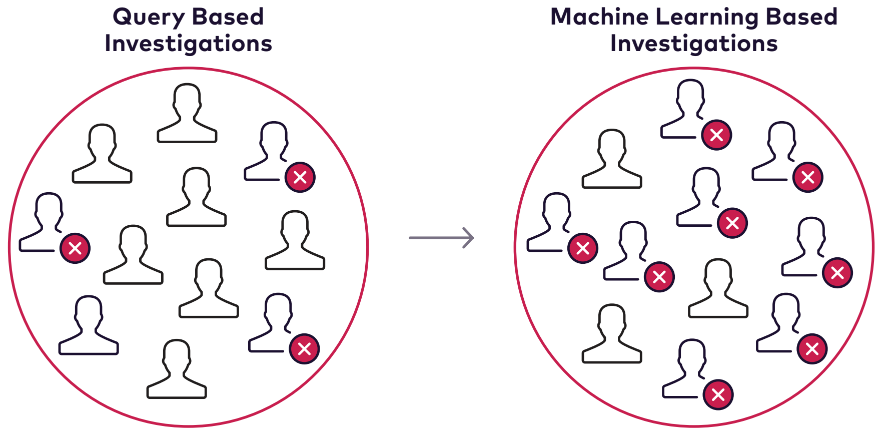


> “Appriss Retail’s machine learning-based models have shown a significant increase in the quality of fraud cases we can identify and the efficiency with which we are able to track down the source of fraud.“
>
> —Chris Hoffman,
> *The Bon-Ton*

A traditional exception based reporting process follows the model seen below. A query is run to look for anomalous behavior and its output is provided to an analyst in the form of a list of transactions or employees to investigate. Then, the analyst or investigator reviews the exceptions presented and determines, sometimes based on looking at additional information, if the transaction or employee in question warrants further review. Some of the exceptions will result in an action against the employee such as arrest or termination of employment, which we label as “success” below.

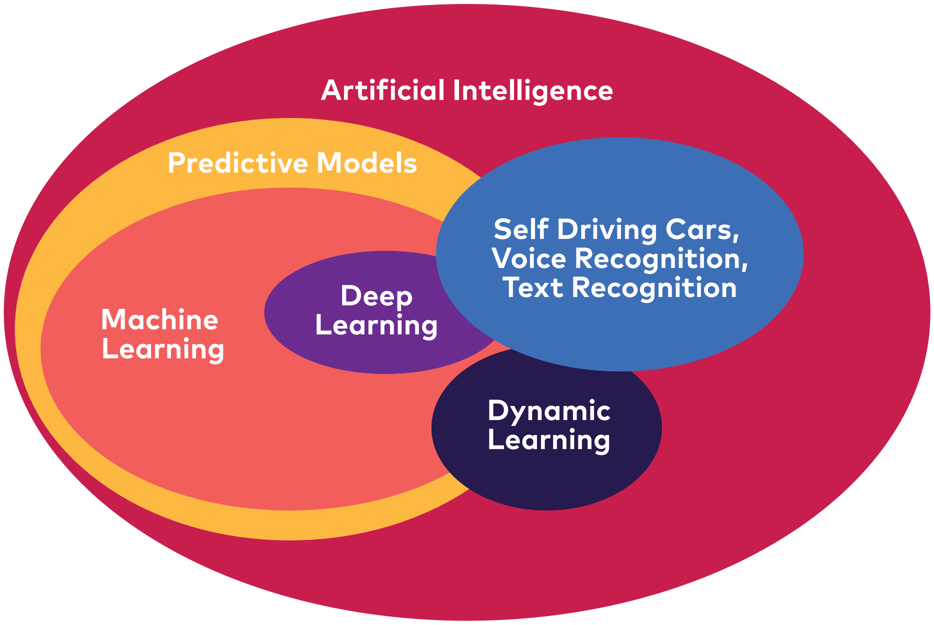


The goal of machine learning is to apply a complex mathematical model to the data and learn, using thousands of variables, the most useful elements in an employee’s transactions for identifying cases that are most likely to result in an action. With a machine learning model, the top ranked exceptions will have a much higher rate of success than those that are generated from traditional EBR (see the following diagram). Over time, an approach with a higher success rate will make loss prevention departments far more efficient and will decrease employee fraud rates.



Overview of Artificial Intelligence and Machine Learning

Artificial intelligence refers to intelligence exhibited by machines. It can cover a wide range of tasks such as authorizing a loan application or automatically driving a car. The next diagram illustrates some of the components of AI that are used frequently. Note that AI is the broad category that contains the other topics.



Predictive Models and Machine Learning

Predictive models and machine learning are shown in the diagram as almost completely overlapping circles and are very close in definition and purpose. The primary difference between the two is machine learning originated in computer science; whereas, predictive modeling was developed in the field of statistics. In the past decade or more, the two fields have essentially merged and now share a highly overlapping definition. Both are concerned with using data to predict or model outcomes, and both are tools used in artificial intelligence.

Deep Learning

To model relationships, a mathematical equation is defined that has variables that are known and parameters that are unknown. The unknown parameters are estimated using a mathematical optimization process to best fit the data. Typically, variables used in machine learning models are defined by a person as candidate variables for a prediction, and are derived from a raw set of data. For example, the raw data may be something like the point-of-sale data for a retailer and an example of a derived variable would be the number of line voids in the past 3 months. Deep learning is a process that uses highly complex model structures to learn patterns in the raw data without any human intervention, thus eliminating the need for derived variables. The difficulty in deep learning is that the more complex models are often more difficult to fit accurately.

Dynamic Learning

There are two very different types of model categories discussed in this section, static models and dynamic models. A static model is fit to a static pattern in data. The predictions from a static model can change as the data it is presented with changes, but if the model is presented with the same exact data, it will produce the same exact answer. By contrast, a dynamic model's prediction can change from day to day even if it is presented with the same information. Dynamic learning is a process by which models are re-fit or updated dynamically as new data presents itself. This approach is useful when a system is continually bombarded with additional information and the relationships in the data are dynamic and the model itself needs to change to adapt.

CASE STUDY: Applying Machine Learning to The Bon-Ton

Appriss Retail partnered with The Bon-Ton to apply machine learning methodology to detect employee deviance. The Bon-Ton is a department store chain headquartered in York, Pennsylvania and Milwaukee, Wisconsin. There are 267 stores, which includes nine furniture galleries and four clearance centers, in 26 states. For this project, we compiled more than five (5) years of historical transaction data and thousands of employee terminations—resulting from investigations—to train our machine learning models. To detect employee deviance, Appriss Retail identified more than 3,000 variables that describe an employee's pattern of activity. The resulting model was applied to recent employee activity and many cases were recommended to The Bon-Ton as candidates for further investigation. Two rounds of testing were performed to evaluate the system's ability to identify previously unidentified cases and to determine the quality of the cases being provided. Through two rounds of testing, there was at least a 25% increase in fraud-related terminations above what would normally be captured by EBR alone, and additionally, a 25% decrease in time required to track down the fraudulent activity within many of the cases.

Initial testing was done on data that was 12 weeks old to see how the normal investigation process compared to what the models were recommending for investigation. To conduct this investigation, the Appriss Retail team had no access to the case management notes after February 14, 2017. Only POS data through February 14, 2017, was used to recommend cases. The top 14 cases from this initial testing were reviewed. Three of the 14 employees were already terminated. Seven new investigations were opened and one re-opened. The remaining three cases were suspicious, but not fraudulent.

A second round of testing was performed with data through May 1, 2017, which was presented on May 26, 2017, allowing some time for the normal exception reporting process to identify cases. The following excerpts detail findings from top ranking cases that were not already identified during normal exception reporting. As of June 17, 2017, there have been nine terminations or likely terminations resulting from the model-based cases and several more cases are still under investigation. Four of the cases are shown on the right.

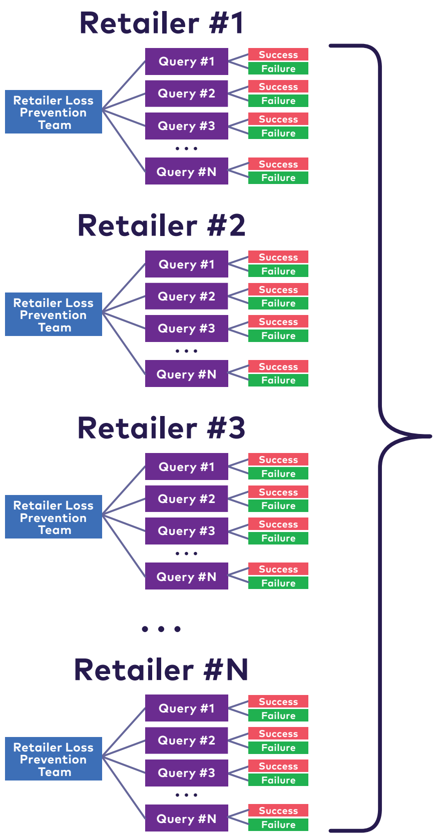
Case 1: Employee scored in the top 1% of the models. Employee is having people make non-receipted returns for him/her so the employee's ID doesn't get swiped. Employee also rang a return without receipt to his/her credit card. There was a customer and merchandise present during the return, however, the video clearly shows that a sales receipt was also present. The items were bought with discount and returned without one. The employee was terminated.

Case 2: Scored in the top 1% of the model. A follow-up investigation showed the employee processed fraudulent refunds and gave merchandise to his/her friends and family for free. When interviewed for merchandise theft and refund fraud, the employee admitted to $387 of merchandise theft and refund fraud. The employee was terminated.

Case 3: Scored in the top 1% of the model. The employee was identified using store credits that he/she also issued. Additional investigation showed the employee retaining sales receipts that were occasionally left behind by customers, and using the receipts to process fraudulent returns. The employee admitted to $213 in refund fraud. The employee was terminated.

Case 4: Scored in the top 1% of the model. Transactions with low sales amount but numerous line voids were noted. A follow-up investigation showed the cashier and a co-worker line voiding items in transactions for each other. The investigation is ongoing to ensure all transactions are reviewed. However, in the five sales already identified, over 36 designer clothing items were given out. The employee was terminated.

Results from the modeling effort at The Bon-Ton show much promise for using machine learning technology to identify employee deviance above and beyond what is identified by typical exception reporting. During the case review, the concentration of bad actors was much higher in the provided cases than through traditional EBR. In addition to the terminations, even employees who were not terminated showed significant issues that needed to be addressed.



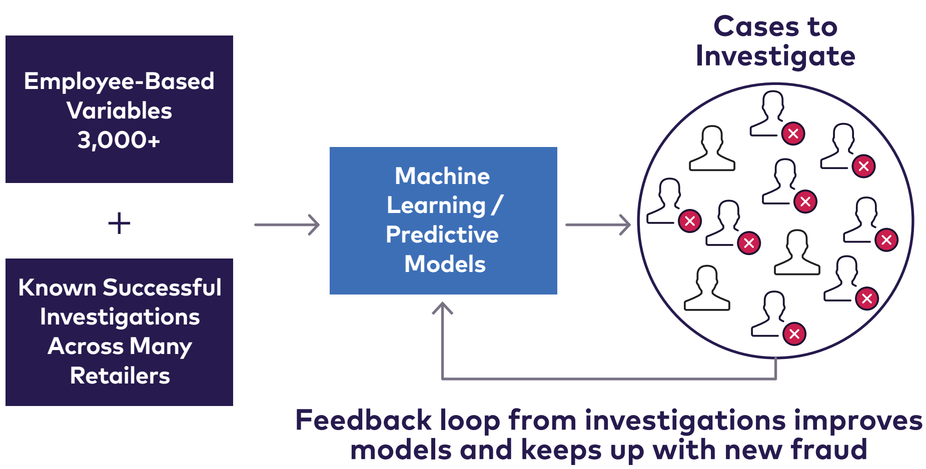
Collective Intelligence

Machine learning models allow developers to use data from many retailers to create a more powerful solution, and Appriss Retail's worldwide retailer base is a significant benefit. Each retailer is unique in the exception-based approaches and investigation techniques that result in employee terminations and prosecutions. The table on the left highlights the general structure across a set of retailers. While every retailer uses a variety of methods for pursuing investigations, the machine learning models collectively learn the most successful strategies.

Combining retailer strategies leads to better detection
- Variables which result from shared questions
- Models built on many successful strategies

Using Dynamic Learning and Feedback

Beyond the collective intelligence model described in the prior section, a final objective of an AI-based employee deviance detection system is a feedback loop from the investigations back into the models. During an investigation, the investigator will perform many actions that can be captured in a system and later used to improve system accuracy. For example, the investigator will click on related transactions and individuals, will open a case, will abandon a case or a transaction, and so on. The captured usage activity can be converted into information that a machine learning model can utilize to improve the models. This type of feedback loop ensures that as new fraud schemes are detected at one retailer, that pattern is captured in a broader model that propagates across all retailers.



Summary

As stated above, the ultimate objective for developing machine learning models is to combine many learnings and obtain collective intelligence to curb suspicious employee behavior. In this application of machine learning, The Bon-Ton has seen at least a 25% increase in fraud-related terminations above what would normally be captured by EBR alone.

Americas +1 949 262 5100
Europe/Middle East/Africa +44 (0)20 7430 0715
Asia/Pacific +1 949 262 5100

APPRISS® RETAIL
apprissretail.com