

EXPLAINABLE-AI FOR FINTECH



Overview

The ubiquitous collection and availability of big data offers tremendous opportunities for the financial sector. It allows highly personalized assessments of risks and prospects, which companies, large and small can then act upon to generate substantial profits.

However, raw financial data are typically rather noisy, with low veracity. However, to overlook or misinterpret even small nuances and trends in these data can be extremely costly. Designing sensitive predictive metrics from these data also requires substantial engineering, which ups the game even further.

The goal of financial analytics is predictive and even better, prescriptive recommendations on actions to mitigate risk and maximize profits. Both must be grounded in exquisite, ideally superior, knowledge of the domain at hand. Formulating this knowledge from data is the job of descriptive analytics.

A key to reliable descriptive data modeling is identifying subpopulations in the data that share certain common features and doing so in a statistically robust manner. No sane financial strategy should be based on learning predictive models from outlier data points. The challenge is to find these stable data regions.

One problem in this mission is the noisy nature of the data; another is the massive number of data attributes and the predictive metrics derived from them. Both lead to vast and unwieldy feature spaces where data are rarely homogenous, and subpopulations are difficult to separate.

Furthermore, many of the features and attributes may not be important for a particular task but this is difficult to assess beforehand. Data analysts are often confronted with the problem of selecting the right features for a specific analysis problem. Unable to cope, feature selection is reduced to a guessing game.

There are many FinTech application areas and many other business sectors where these critical challenges occur:

- Stock market prediction: predict stock prices from past data patterns and current observations
- Fraud detection and prevention: identify malicious entities that have common patterns
- Credit scoring: assess credit worthiness by estimating default risk based on past data
- Risk management: use past internal and external data to estimate risk and prevent future losses
- Personalized marketing: estimate a person's group to suggest purchases or investments

Challenges:

In any of these applications, as well as others, knowing a data point's membership in a concrete and statistically robust subpopulation is critically important. Determining an individual's subpopulation membership with high confidence drastically reduces risk and uncertainty in the decision's outcome.

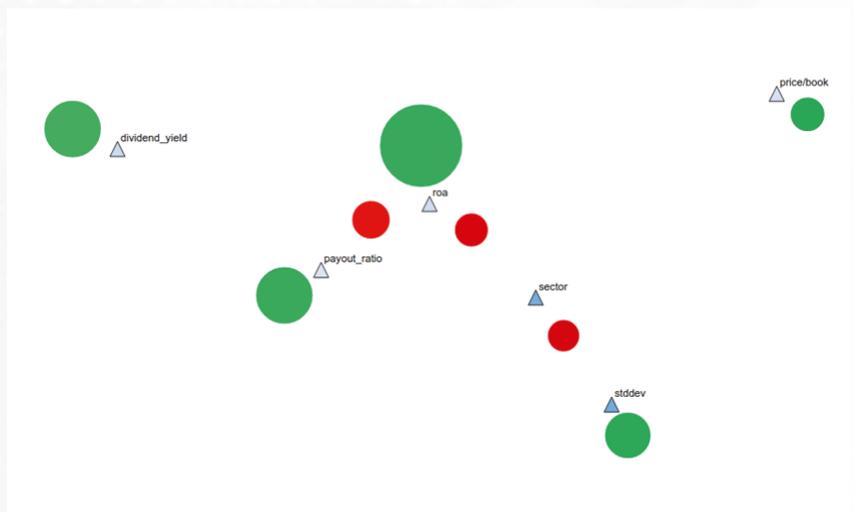
Moreover, many times it is not only important to know whether a particular data item will yield a certain outcome in a given response variable. Rather, it can be important whether it allows certain predictions with regard to the variable's growth or decline. For example:

- Will a certain stock rise or fall and what are the attributes and their ranges when this occurs?
- What are suspicious data signatures of possible fraud to inform on-site situational awareness?
- Will a loan applicant default and what are the specific attributes/values to look out for?
- What are specific customer profiles – attributes and value ranges – to use in targeted marketing?

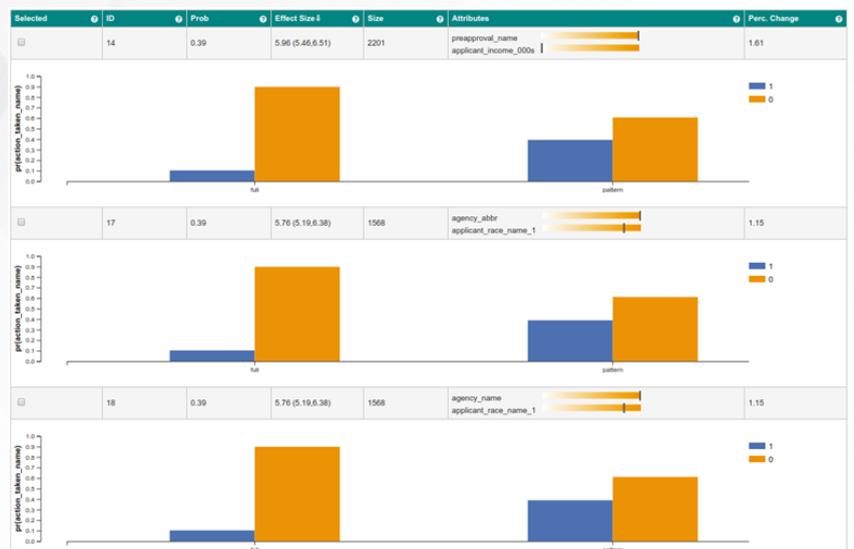
One typically assesses these types of predictions by statistically correlating the target variable with one or more attributes. However, correlation can be sporadic and confined to a single specific subpopulation. It may even vanish and remain undetected, when aggregated over loosely defined subpopulations.

Identifying robust and well-defined subpopulations in large datasets is hence mission-critical. However, the "data fog" of noisy data, in which these subpopulations are typically situated poses tremendous challenges. In other words, large data is a blessing and a curse at the same time.

Deep neural networks, random forests, etc. circumvent this problem in an elegant manner but this comes at the price of low explainability and subsequent potentially poorly informed decision making. This uncertainty is a fateful adversary in the high stakes effort of financial behavior modeling.



Contextual Pattern Visualization



Probability Bar Chart

How Explainable-AI can help:

Akai Kaeru's advanced pattern mining and discovery engine operates at the root of the problem. It directly decomposes high-dimensional input data into a reliable set of independent data patterns – the subpopulations –using a sophisticated, yet efficient suite of proprietary algorithms.

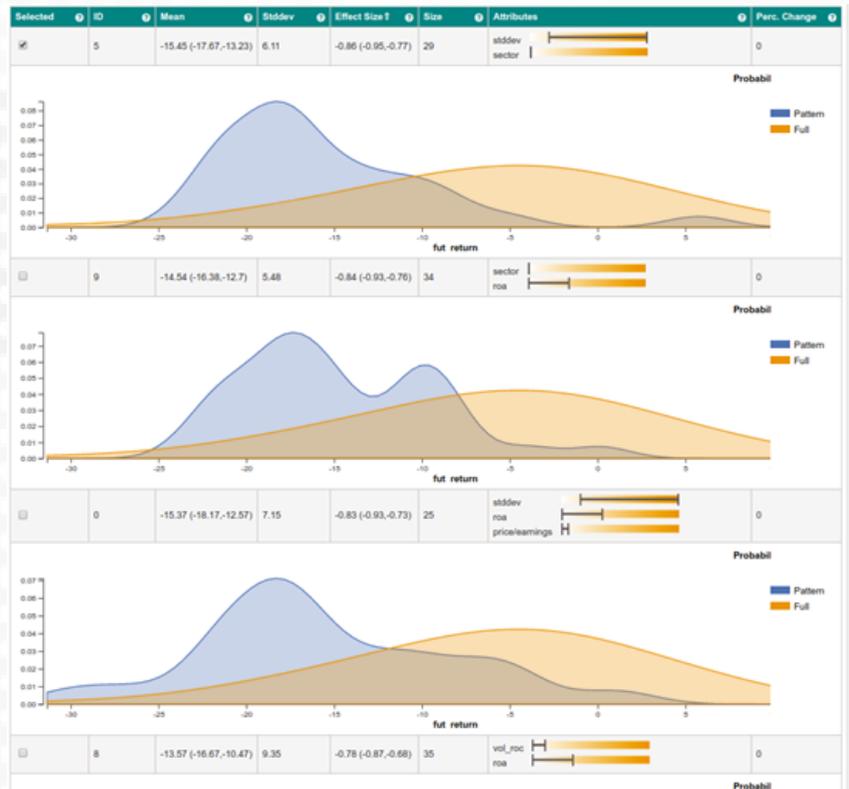
Each pattern consists of data items that behave similarly in terms of a given target variable and which are succinctly defined by just a small set of attributes. The succinct description of each pattern makes them easy to understand and appeals to the user's domain knowledge – a hallmark of explainable AI.

Our pattern mining engine delivers its products to an innovative, fully interactive visual interface. It alerts the analyst to a set of concrete and concise regions with unusually high or low target variable values and correlations. The patterns are explained by visualizing them in the context of their relevant attributes.

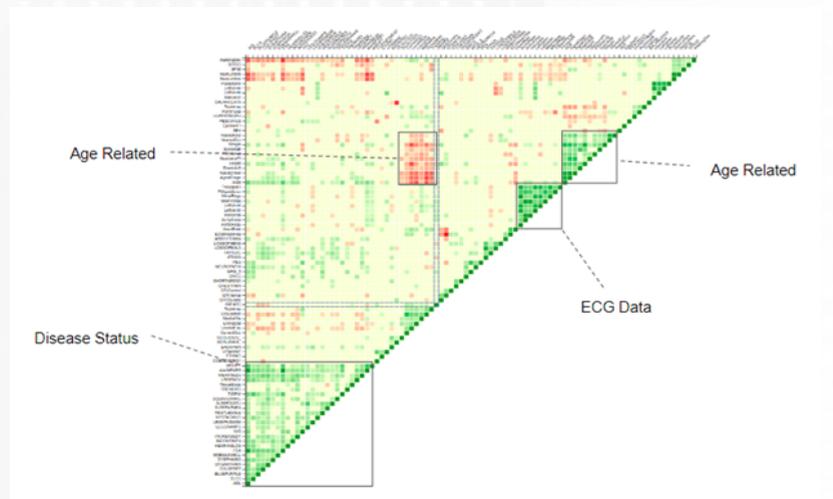
Insights like news or sentiments from textual data, such as twitter feeds, Google finance, etc., often form an important ingredient in FinTech analysis. They can be converted into sets of numerical attributes that can be integrated into the high-dimensional feature space for analysis within our software.

Also part of our software is a causal inference engine, which can establish true causal relationships among the patterns. It provides directional relationships such as "smoking causes cancer, but not vice versa". Our visualization interface conveys the intricate web of causal relationships directly to the pattern layout.

Via these capabilities, Akai Kaeru's software can help analysts and decision makers turn troves of data into actionable insight they can trust, justify, and explain to others. There are no magic black boxes.



Salient Pattern Mining Output



Correlation Heat Map

