Rapid Data Monetization

The Four Faces of Data Monetization

Data Monetization is the process of deriving calculable positive financial value from stored and available data. That value can come directly from vending access to valuable, proprietary data, or indirectly through the use of advanced analytics and external data to create and sell proprietary predictive models or scoring services.

There are four primary forms of data monetization.

- Intra-Business
  - Enterprise/Internet business value creation in Mktg., CRM, Mfg., and other functions
- Inter-Business
  - Internal business data sold or bartered to data providers, business partners, and others
- Market/Industry
  - Commercial data providers in Finance, Retail, Healthcare, Pharma and other segments
- Infrastructure/Environment
  - Commercial aggregation of signal, sensor, and social data for predictive IoT applications

Obstacles to Effective Data Monetization

Across all of these cases there are some common obstacles hindering the realization of the full business value of data assets. They include the following:

- Regulations governing data privacy, location and uses impede sharing across geo/biz boundaries
- Costly intermediary parties needed to assure legitimate data use by authorized data consumers
- Data acquisition, integration, analysis, and product discovery require complex, costly skills
- Analytical products derived from data assets are black box calculations, hard to sell and audit
- High volume prediction using conventional algorithms requires large, central compute resources

The Rulex Solution for Data Monetization

Rulex is a unique machine learning platform that makes data monetization easier and more profitable by addressing the critical drawbacks presented by conventional machine learning. Rulex enables data monetizers to:

- Extract more value from data assets with less regulatory and business risk
- Simplify, accelerate, and lower the cost of creating and delivering products/services
- Create new high value data and analytics products not possible with other tools

Rulex, Inc. | 75 Federal Street, Suite 920 | 02110 Boston, MA | USA
Office: +1 617 263 0080 | Fax: +1 617 263 0450 | www.rulex-inc.com |@Rulex_Analytics

Rapid Data Monetization

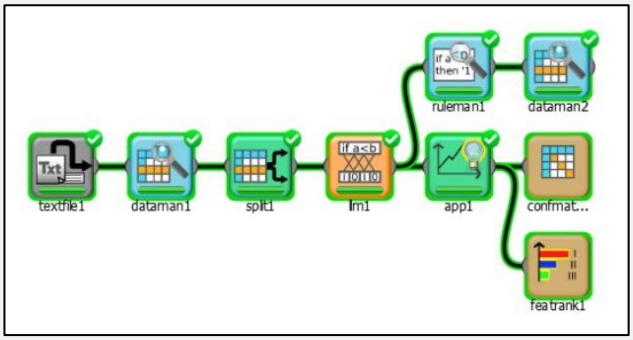
For data monetizers of all types, Rulex can dramatically increase productivity, competitiveness, and regulatory compliance. It can also reduce time to value for data assets and enable the rapid development of exciting new products.

- Quickly find, analyze, and share valuable, actionable data patterns without exposing valuable data assets to misuse or regulatory risk.
- Easily create and implement unique, high-value logic-based analytical products that are easy to explain, audit, and promote.
- Increase statistical analyst productivity with faster data discovery and enable a smooth, fast transition to machine learning with business rules-based predictive modeling.
- Offer faster, simpler, more efficient scoring services that can run closer to the network edge and can be easily implemented across geographic and organizational boundaries.

The Rulex platform is easy to download, install, and use and it is capable of processing a billion data items on a desktop PC, or processing much more data on a server or in the cloud. And, it has a short learning curve and it is easily implemented and integrated in any IT and network infrastructure.

Rulex is powered by unique technology that was developed through academic and government research in cognitive machine learning, and has been proven in commercial applications in financial services, healthcare, telecom, retail and other industries, in Europe and the US.

Rulex provides an easy-to-use, drag-and-drop graphical interface for data and business analysts.



Rulex protects data assets by eliminating the need to share valuable, sensitive source data with external parties like suppliers, contractors, or partners. Data owners can use Rulex to identify the desired data and then automatically share that data for machine learning in a form that cannot be reverse-engineered or misused, and is completely compliant with data privacy regulations. The following is a textual example of the learning data produced by Rulex.

Rulex, Inc. | 75 Federal Street, Suite 920 | 02110 Boston, MA | USA
Office: +1 617 263 0080 | Fax: +1 617 263 0450 | www.rulex-inc.com |@Rulex_Analytics

x_1	x_2	u_1	u_2	z	y	x_1	x_2	u_1	u_2	z	y
5100	2500000	1	1	011 0111	1	4000	4290000	1	3	011 1101	0
6200	3010000	1	1	011 0111	1	5800	4600000	1	3	011 1101	0
7100	3220000	1	1	011 0111	1	6000	4160000	1	2	011 1011	0
8400	2960000	2	1	101 0111	1	6700	3840000	1	2	011 1011	0
9300	3780000	2	2	101 1011	1	6900	5430000	1	4	011 1110	0
10600	4180000	2	3	101 1101	1	7800	4920000	1	4	011 1110	0
10900	3170000	3	1	110 0111	1	8700	5680000	2	4	101 1110	0
11300	2660000	3	1	110 0111	1	9800	5430000	2	4	101 1110	0
12400	3780000	3	2	110 1011	1	10600	5050000	2	4	101 1110	0
13500	4180000	3	3	110 1101	1	11500	5940000	3	4	110 1110	0

Rulex requires no programming or math skills to use. It automatically finds the most valuable data patterns and expresses them as actionable business rule. And it enables business experts to select the most valuable rules for use in predictive decisions.

Output	Cond 1	Cond 2	Cond 3	Cond 4
income = <=50K	relationship ∈ {Other-relative, O...			
income = <=50K	education ∉ {Bachelors, Doctorate...	relationship ∈ {Not-in-family}		
income = <=50K	education ∈ {10th, 11th, 12th, 1st...			
income = <=50K	education ∉ {10th, 5th-6th, 7th-8t...	occupation ∉ {Handlers-cleaners, ...	capital-gain ≤ 4243.500	hours-per-week ≤ 35
income = <=50K	age ≤ 29	relationship ∈ {Husband, Not-in-f...	capital-gain ≤ 5119.000	

Rulex provides data monetizers with these unique business benefits:

- Automatic Data Protection and Regulatory Compliance
- Intelligible, Auditable, Merchantable Predictive Models
- Complete, Automatic Audit Trail for Entire Analytics Process
- “Thin Prediction” on Small Distributed Systems at the Network Edge

The remainder of this document provides a detailed look at the way Rulex delivers these benefits to companies wanting to get the most value possible from their data, or other people’s data, or data from the Internet of Things.

If you searching for a solution to any kind of data monetization challenges, Rulex may be your solution. Please visit our web site, http://rulex-inc.com, or just drop an email to monetizeme@rulex-inc.com to find out more how Rulex can help you.

Rulex Data Monetization Benefits In-Depth

1. Automatic Data Protection and Regulatory Compliance

Rulex is unique in Machine Learning in that it does not require people building models to explore raw input data. With other algorithms, it is necessary to expose valuable data assets to theft, misuse, and regulatory risk by sharing them across company and country boundaries with external analysts.

With Rulex, the data owners can provide external analysts with "discretized" data for use in modeling. In this form, the source data has been reduced to its intrinsic patterns and no longer contains any metadata or identifiable data values. This is very different from awkward solutions like data masking and "anonymizing", where an altered form of the source data is passed to the analyst. Such schemes can help with regulatory risk, but they can still allow a range of off-label analytics that can't be controlled by the data owner, and they can't be applied to many kinds of data, such as geo-location data.

Rulex enables companies selling data to other companies to create "blind" data products that are automatically protected and compliant. Note that, in most cases, these data products will require the consuming company to use Rulex for modeling and scoring. This enables the creation of a company-to-company analytical ecosystem that allows a higher level collaboration and value exchange between the parties.

<u>Use Case: Retailers want to sell/barter POS data to their suppliers.</u>

Suppliers want to buy the data for analytics pertaining to branding, marketing, and other things that are of no consequence to the Retailer. The problem is that exposing raw POS data to the Suppliers gives them the ability to use analytics to gain leverage over the Retailer in relation to pricing, promotions, and other business dimensions. This problem creates the need for expensive intermediaries like 1010data and others to perform the prescribed analytics on behalf of the Supplier.

Rulex makes this unnecessary, allowing for a direct analytical relationship between the Retailer and the Suppliers where they both can get more value from the data without any imbalance in business leverage.

2. Intelligible, Auditable, Merchantable Predictive Models

Rulex is unique in that it produces predictive models that are based on conditional logic, rather than mathematical functions. Rulex models can be read, understood, explained, and certified by business people. The "solving functions" created by conventional machine learning algorithms for making predictions cannot be.

The following is a side by side comparison of the models produced by Rulex and an Artificial Neural Network algorithm for predicting voting choices in a US election, using a dataset taken from Kaggle.

Rulex Model

1. IF (White alone, percent, 2014 <= 71.100) **THEN party = Democrat**
2. IF (High school graduate or higher, percent of persons age 25+, 2009-2013 > 90.450) **THEN party = Democrat**
1. IF (High school graduate or higher, percent of persons age 25+, 2009-2013 <= 88.650 AND White alone, not Hispanic or Latino, percent, 2014 > 73.950) **THEN party = Republican**
2. IF (Persons per household, 2009-2013 > 2.615 AND White alone, percent, 2014 > 62.300 AND White alone, not Hispanic or Latino, percent, 2014 > 32.500) **THEN party = Republican**
3. IF (Persons 65 years and over, percent, 2014 > 19.850 AND High school graduate or higher, percent of persons age 25+, 2009-2013 <= 90.850 AND White alone, not Hispanic or Latino, percent, 2014 > 57.350) **THEN party = Republican**
4. IF (724908 < Manufacturers shipments, 2007 ($1,000) <= 16640803 AND Asian alone, percent, 2014 <= 6.350 AND White alone, not Hispanic or Latino, percent, 2014 > 63.000) **THEN party = Republican**
5. IF (Private nonfarm establishments, 2013 <= 289 AND Population per square mile, 2010 > 14.550 AND 56.150 < White alone, percent, 2014 <= 98.750) **THEN party = Republican**

As you can see above, the predictions are the result of the logical conditions inferred from the source dataset by Rulex. This is very different than the equivalent solving function produced by a conventional neural network algorithm, illustrated below.

Neural Network Model

$$
\begin{aligned}
f(\mathbf{x}) \;=\;& 0.293\, tanh(0.113\, x_0 + 0.337\, x_1 - 0.329\, x_2 + 0.251\, x_3 - 0.288\, x_4 - 0.297\, x_5 + 0.436\, x_6 + \\
+\;& 0.166\, x_7 - 0.184\, x_8 + 0.219\, x_9 + 0.483\, x_{10} - 0.222\, x_{11} + 0.173\, x_{12} + 0.012\, x_{13} + \\
+\;& 0.352\, x_{14} + 0.259\, x_{15} + 0.176\, x_{16} + 0.345\, x_{17} + 0.314\, x_{18} + 0.177\, x_{19} - 0.329\, x_{20} + \\
-\;& 0.363\, x_{21} + 0.216\, x_{22} - 0.148\, x_{23} - 0.043\, x_{24} + 0.316\, x_{25} - 0.068\, x_{26} - 0.421\, x_{27(0)} + \\
+\;& 0.15\, x_{27(1)} - 0.289\, x_{27(2)} - 0.241\, x_{28} + 0.16\, x_{29} + 0.199\, x_{30} - 0.111\, x_{31} - 0.164\, x_{32} + \\
+\;& 0.117\, x_{33} + 0.466\, x_{34} + 0.457\, x_{35} + 0.133\, x_{36} + 0.331\, x_{37} - 0.362\, x_{38} - 0.43\, x_{39} + \\
-\;& 0.491\, x_{40} - 0.155\, x_{41} + 0.371\, x_{42} - 0.05\, x_{43} - 0.177\, x_{44} - 0.044\, x_{45} + 0.225\, x_{46} + \\
+\;& 0.328\, x_{47} - 0.118\, x_{48} - 0.3) + \\
-\;& 1.934\, tanh(-0.233\, x_0 + 0.174\, x_1 - 0.252\, x_2 - 0.501\, x_3 - 0.125\, x_4 + 0.311\, x_5 - 0.573\, x_6 + \\
-\;& 0.299\, x_7 + 1.123\, x_8 + 0.318\, x_9 - 1.169\, x_{10} + 0.105\, x_{11} - 0.429\, x_{12} - 0.075\, x_{13} + \\
-\;& 0.143\, x_{14} + 0.146\, x_{15} - 0.531\, x_{16} + 0.077\, x_{17} - 0.133\, x_{18} - 0.122\, x_{19} + 0.162\, x_{20} + \\
-\;& 0.08\, x_{21} - 0.496\, x_{22} - 0.21\, x_{23} - 0.113\, x_{24} + 0.485\, x_{25} + 0.575\, x_{26} - 0.126\, x_{27(0)} + \\
+\;& 0.135\, x_{27(1)} + 0.022\, x_{27(2)} - 0.352\, x_{28} - 0.693\, x_{29} + 0.379\, x_{30} + 0.409\, x_{31} - 0.109\, x_{32} + \\
+\;& 0.228\, x_{33} + 0.292\, x_{34} + 0.161\, x_{35} - 0.086\, x_{36} - 0.3\, x_{37} - 0.089\, x_{38} + 0.163\, x_{39} + \\
-\;& 0.074\, x_{40} + 0.31\, x_{41} - 0.849\, x_{42} + 0.14\, x_{43} + 0.754\, x_{44} + 0.291\, x_{45} - 0.533\, x_{46} + 0.273\, x_{47} + \\
-\;& 0.285\, x_{48} - 0.286) + 0.252
\end{aligned}
$$

For companies wishing to monetize their data by selling analytical models or provide scoring services to other parties, as is widespread in Financial Services, this limitation makes it very difficult to create, value, market, and price-justify analytical products and services, and can also carry regulatory liabilities in many cases, where predictive calculations must be audited and certified.

Rulex enables companies wishing to sell analytical models or provide predictive scoring services to other parties to create, market, and certify new, higher value offerings faster, more easily, and with less required skills than conventional machine learning tools and techniques allow.

Use Case: Credit Bureau wants to offer new, high value analytics products and scoring services.

Many credit bureaus would like to create and sell new predictive models to financial analysts and scoring services to merchants and vendors using a combination of existing data sources and new data from social networks and other new Big Data sources.

With conventional machine learning, they would be forced employ a data scientist to disclose details of their data selection and preparation methods, parameter selection, and other things. They must do this because they cannot illustrate or explain the model. What could otherwise be trade secrets must become necessary disclosure.

Rulex makes this unnecessary by providing fully transparent prediction modeling and scoring that can be easily explained by business people without the need for any special tools or skills and without revealing anything about their source data and product development processes.

3. Complete, Automatic Audit Trail

Rulex is unique among machine learning solutions in that it automatically provides a complete, coherent audit trail for all software and human actions in the chain from data to prediction.

Data Monetization best practices place great importance on the analytical audit trail. The prevalence of multi-party relationships, regulatory constraints, and challenging value and cost accounting, makes it essential to fully account for all software and human action from data acquisition and discovery, to model development and testing, to predictive application deployment. And yet, conventional machine learning solutions make auditing very difficult and costly.

Conventional Machine Learning is usually accomplished in one of two ways, either through “big box” product suites from SAS, IBM, and others, or through an amalgamation of open source or point commercial solutions for data exploration, modeling, and scoring. In the first case, the tools often come with additional layers of control and management software for auditing, while the second case comes with virtually no built-in auditing features, forcing an ad hoc addition of procedures and tools to the machine learning pro-

cess. In both cases, each part of the process has its own semantics and way of reporting events. Thorough auditing requires a great deal of systems knowledge and additional labor.

Rulex makes this unnecessary. It doesn't simply provide added separate tools for auditing; it builds automatic auditing into the platform by recording every user and software action in a single, unified base, expressed in a complete, coherent language, according to a single set of process semantics. This allows for complete analytics auditing without added software layers, proactive processes, or forensics after the fact.

This examples show the form in which Rulex captures software and user actions for a few typical processes.

```
//Filter
col_0 = @set.data[indrow, findatt(@set, "State")];
a = (col_0 isin ["AR", "DE", "IL", "LA", "MA", "MD"]);
indrow = indrow[which(a)];
nind = size(indrow);

 //EDIT ATTRIBUTES
jatt = findatt(@set, "Area Code");
@set.att[jatt].datatype = "nominal";

 //CREATE ATTRIBUTES
jatt = size(@set.att)+1;
@set.att[jatt].name = "Ave Day Duration";
@set.att[jatt].datatype = "nominal";
@set.att[jatt].role = "input";
```

Notice that Rulex doesn't just record a log record of each operation, but rather captures calculations, parameter values, and functions performed in a complete procedural record of all platform activities.

Use Case: Consumer Data Provider needs to audit and report handling of personal data.

Data monetizers who use personal consumer data and other kinds of regulated data must be able to audit their data processing, handling and sharing, and anlytics actions to assure regulatory compliance.

Satisfying this requirement with conventional machine learning suites or toolchains requires one set of processes in relation to database and ETL tools, another for modeling, and another for prediction, each with its own interfaces and semantics. In addition to be complicated and error-prone, this approach suffers more for lacking the ability to provide a single, coherent, detailed, end-to-end expression of the entire history of the analytical process. To provide such a picture requires consultants to piece it together after the fact.

Rulex makes this unnecessary by providing the Rulex Language and using it on the Rulex Machine Learning platform to automatically capture a semantically complete record of all operations conducted on restricted data by the Consumer Data Provider.

It is worth noting that the Rulex audit trail also represents captured knowledge, creating a record of everything an analyst or data scientist has done; if he or she leaves the company, the knowledge they have gained about data or predictions is not lost. It is retained in a complete, coherent way.

4. Thin Scoring on Small Distributed Systems at the Network Edge

Rulex's unique, logic-based predictive models described earlier bring significant benefits of flexibility and efficiency to networked predictive applications that use the models to score new data, in batches or real time.

As you can see from the model comparison earlier, conventional machine learning algorithms produce predictive models taking the form of complex mathematical functions. Using such models to rapidly score large volumes of new data is computationally intensive and must typically be done on a large central system, rather than on smaller systems close to the places where new case data is captured.

For companies providing any kind of cross-border scoring services, this limitation makes it impossible to score new data where the new cases occur and introduces regulatory liability by forcing sensitive data across borders to complete the scoring on a central system.

For companies providing predictive IoT applications and services, this limitation makes it impossible to score new data on low-power devices at or near the edge of the network, as needed for real-time applications on mobile devices such as location-based marketing.

Using conventional prediction technology, enabling distributed scoring means added data processing steps for cross-border data protection and considerable duplicated processing hardware across multiple locations.

Rulex makes this unnecessary. Rather than mathematical functions, Rulex produces predictive models taking the form of conditional logic statements, which results in the simplest, most efficient scoring process possible. Logic rules can be used for scoring on devices as small as an Android phone or a System on a Chip (SoC) in a specialized IoT edge device.

<u>Use Case: Telco wishes to implement Predictive Maintenance for network and customer equipment.</u>

To do this, they must be able to capture data from various devices, build predictive models, and use those models to predict equipment failures across the network.

They must be able to search through an enormous volume of data for patterns that are rare and also important because they predict costly equipment service events. Once they have built the model, they must be able to use it to evaluate a huge volume of new network events and make predictions as quickly as pos-

sible. Also, as network conditions and configurations change, they must be able to refresh and redeploy the models as quickly as possible.

With Rulex, the Telco can continuously gather historical network signal data and automatically detect the patterns that predict equipment failure and create from them the conditional logic model defining the failure event. The model can then be distributed to small systems located throughout the network in switching nodes, where they can be used to rapidly identify new failure probabilities.

Conclusion

Rulex provides a unique solution to the most difficult problems of data monetizaton in the form of an easy-to-use machine learning platform for business and data analysts that empowers them to discover valuable, exciting new data and analytics products and services, and use, implement, and vend them in a safe, efficient way.

To find out how Rulex can help you obtain significant, measurable value from your data, other people's data, or IoT data, please visit http://rulex-inc.com or email us at monetizeme@rulex-inc.com.

You will find the path to more effective data monetization is shorter than you think.