Supply Chain Decision Automation

Applied AI for Planning Productivity

Tim Negris
SVP Marketing and Business Development
Rulex Inc.
www.rulex.ai
t.negris@rulex.ai

Introductions

Rulex - The AI that Tells You Why®

Italo-American AI software company providing unique proprietary, easy-to-use machine learning software for enterprise decision automation in Finance, Manufacturing, Utilities, and the IoT.

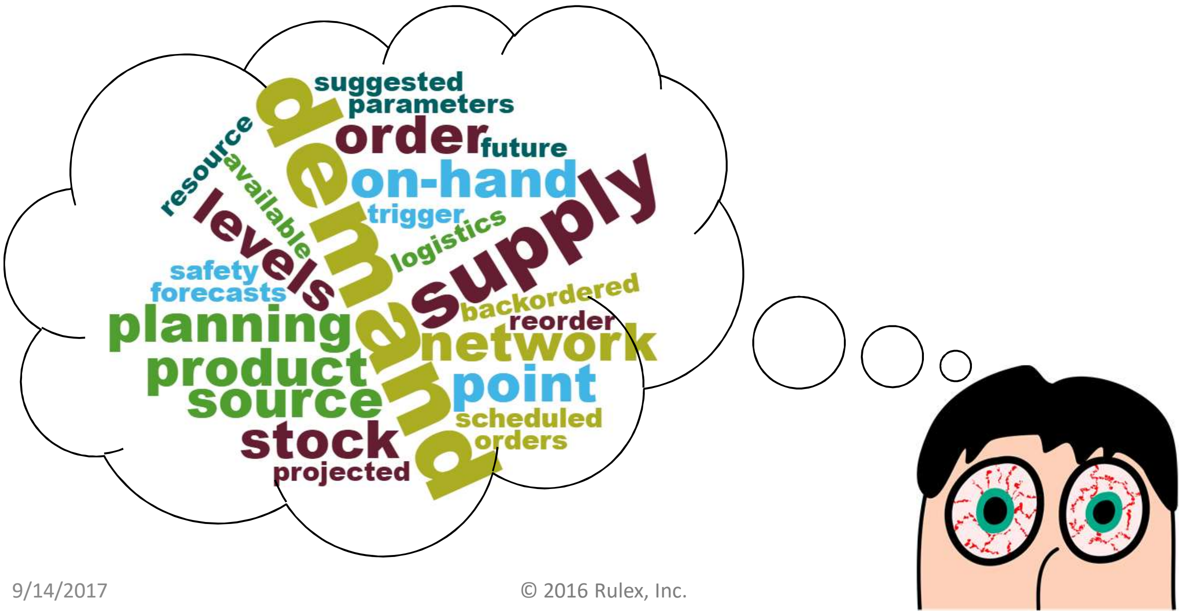
Tim Negris

First SQL Server Product Manager, Oracle VP Server Marketing, IBM VP Software Marketing and Sales, Strategic Advisor to HP and Dell, VP numerous AI startups.

Topics

- The ERP Decision Explosion
- AI Is the Answer, What Are the Questions?
- Automated Decision Use Cases
- Common Mistakes to Avoid

ERP Decision Explosion



Today's planners are facing a decision explosion. Especially in manufacturing, where the business is becoming more complex and dynamic, they face a growing number of decisions that must be made ever more quickly, and they also face the possible competitive and economic consequences of incorrect or inconsistent decisions.

In this presentation we will examine the causes and effects of this explosion and how Artificial Intelligence can improve and simplify ERP decision-making for planners of all experience levels.

Decision Explosion Causes

Global Sourcing

Brand Expansion

Accelerated Mfg.

Distribution Diversification

9/14/2017 5

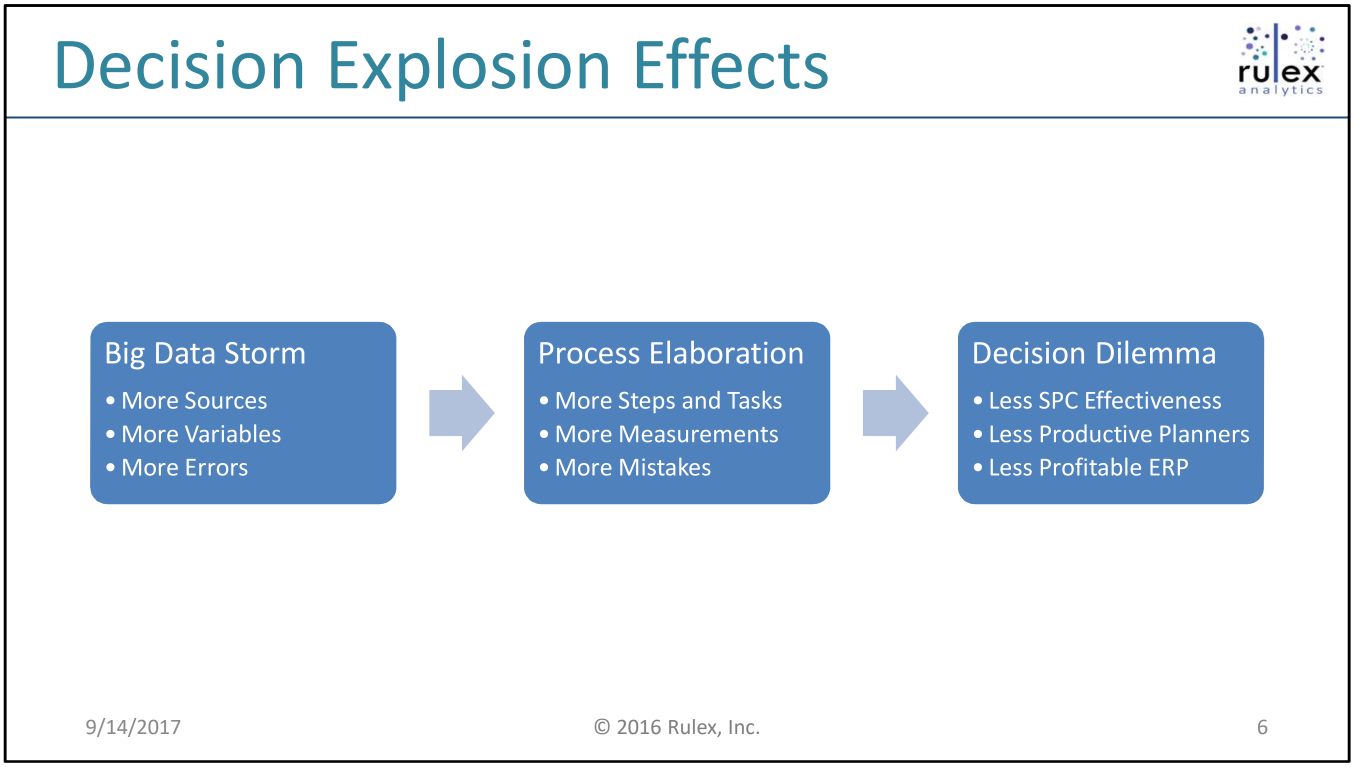
What is causing this decision explosion? There are many contributing factors, but here we see some the most important ones.

Global Sourcing. In years past, the supply chain was characterized by relatively fixed and localized sourcing schemes for manufacturing inputs, but rapid advances in technology for logistics and transportation, coupled with the emergence of highly price-competitive new suppliers all around the world, have created a wide range of new decisions concerning supplier selection and materials scheduling.

Brand Expansion. Especially in CPG, manufacturers are working ever harder to leverage successful brand investments and to serve increasingly diverse and individualized consumer demand, creating an explosion of product variations and compression of delivery time frames, creating new decisions for effective demand planning.

Accelerated Manufacturing. New technologies are enabling manufacturing to be more nimble, changeable, and distributed, resulting many new planning decisions across a range of fulfilment execution functions.

Finally, distribution channels are diversifying rapidly and straining the ability to put the right products in the right places at the right time using statistical methods, creating a new planning decision burden in demand planning in the form of the need for “rule-of-thumb” procedures to assure effective distribution resource functions.



That explosion in ERP planning decisions has started a chain reaction that leads to a decision dilemma.

All of the new business processes coming from the explosion are creating huge amounts of data, the ERP Big Data Storm. We are seeing rapid growth in volume and velocity of new data coming from more sources with more variables, and we are seeing even faster growth in the rate of new errors in critical ERP applications data.

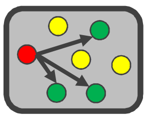
And more data means more process. That is, more steps and tasks for collecting, integrating, and analyzing data; more measurements to make and interpret, and more mistakes in process design and execution due to data errors, statistical forcasting inaccuracy and other factors.

All of this leads up to a decision dilemma.

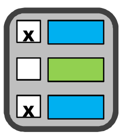
The statistical process control and optimization methods and software in wide use today are falling out of sync with the increasingly dynamic and variable nature of ERP, and the growth in data and process that comes with. Planners must spend an increasing amount of time validating automated forecasting output manually adjusting it according to process rules or rules of thumb. And the combination of incorrect forecasts and the added labor needed to try and prevent them, can produce an enormous negative business impact across the entire ERP environment.

Planning Decision Types

- Correction
  - Data Errors, Human Mistakes, Condition Changes



- Guidance
  - Stocking Levels, Shipment Routing, Materials Substitution



- Selection
  - Suppliers, Distribution Lanes, Inventory Assignment

9/14/2017 7

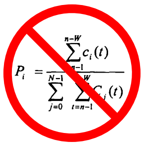
To appreciate the magnitude of the planning dilemma, it may help to consider the breadth of decision types in the planning explosion. There are many, but these three represent the ones with the greatest business impact.

Correction – Deciding how to fix data errors, correct process mistakes, and respond to sudden or unexpected changes in system or business conditions.

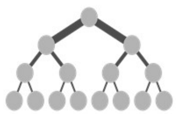
Guidance – Deciding how best to hold, move, and utilize raw materials, work in process, and finished goods.

Selection – Deciding which is the right choice at the right time to get the best result in materials sourcing, distribution direction, and effective inventory placement.

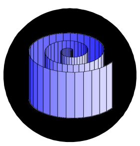
Planning Decision Dilemma



- Statistical Methods (SPC) Not Dynamic
  – Humans must second-guess the system, e.g. SAP APO



- Low-Value Decisions Multiplying Rapidly
  – New tasks created by new data and business processes



- Dark Data Patterns Difficult to Discover
  – Hidden relationships, anomalies, errors, missing data

9/14/2017 8

As we look across the growing scope of the ERP planning decision dilemma and its impact on the business, these three challenges emerge as the most important.

Statistical process methods are not dynamic. When the business changes, those methods must be changed and tested, which is not a trivial process. In practice, SPC is rapidly falling behind the rate of business change and cannot catch up. As a result, planners must fill in this intelligence gap by applying expert judgment to optimizer outputs and such to assure their correctness and consistency.

The greatest financial impact of the decision dilemma comes from the vast multiplication of planning process steps and variations which are creating a corresponding multiplication in low value decisions, also called "decisions that should not have to be made". Todays planner is making at least two orders of magnitude more individual decisions that a decade ago, many of them artifacts of system and network design weaknesses, with little business value.

Finally, ERP systems contain vast amounts of dark data, system, machine, and application data that is not normally exposed to end-users. Dark data errors can cause problems which are hard to diagnose and correct, and yet it is also a vast untapped analytical asset for improved decision automation. More on this shortly.

Decision Automation to the Rescue

- Use Machine Learning to Predict Best Actions
  - Improve Forecasting Accuracy
  - Eliminate Low-Value Decision
  - Correct and Exploit Dark Data

In the face of all of the factors above, human decision making using rules of thumb gained from experience or formal rules gained through operational research is not enough. And it is possible to use machine learning to enable automated decision making.

That is, to use machine learning to predict the best action to be taken, by a person or a computer program to improve forecasting, eliminate human tasks, and to correct and exploit dark data.

Let's see how this works.

AI Is the Answer; What are the Questions?

- Which decisions to automate?
- What data will be needed?
- What kind of AI will be needed?
- Which uses cases to target?

9/14/2017 10

In the tech, business, and popular media, Artificial Intelligence (AI) is a popular topic and presented as the solution to many different problems. But, if AI is the answer, for the user, what are the most important questions.

This section seek to answer these questions.

Which are the best decisions to automate?

What the data that will be needed to build predictive models?

What kind of AI software should be used?

And finally most important: What is the business problem we wish to solve?

Which Decisions to Automate?

- Costly Decisions
  - High levels of time, skills, cost, or execution risk
- New Decisions
  - Manual planning steps, new business capabilities
- Dark Decisions
  - Data correction, real-time event response

Here are the key types of candidates for decision automation.

Costly decisions are ones that take a lot of time or require a lot of skill to make, and ones where a bad decision can induce financial risk as with unfulfilled service level agreements.

The addition of new business lines, suppliers, materials, brands, products, channels, and customers can create numerous new, often manual planning steps in the ERP workflow. Look to the new business processes for decisions that can be automated to support new capabilities.

And then there are dark decisions, places in the operations of the ERP data and applications where errors can automatically be prevented and corrected and the system can make adjustments in real time for optimum performance.

What Data Is Needed?

- Existing ERP Data
  - Master data, applications output, spreadsheets, etc.
- Derived Data
  - Aggregations, calculations, statistics, etc.
- Dark Data
  - Transient process variable, real-time events, etc.

Analytical workers spend as much as 80% of their time gathering, exploring, cleaning, and integrating data. Many users make the mistake of looking to the data to tell them what they can predict, when they should consider the prediction first.

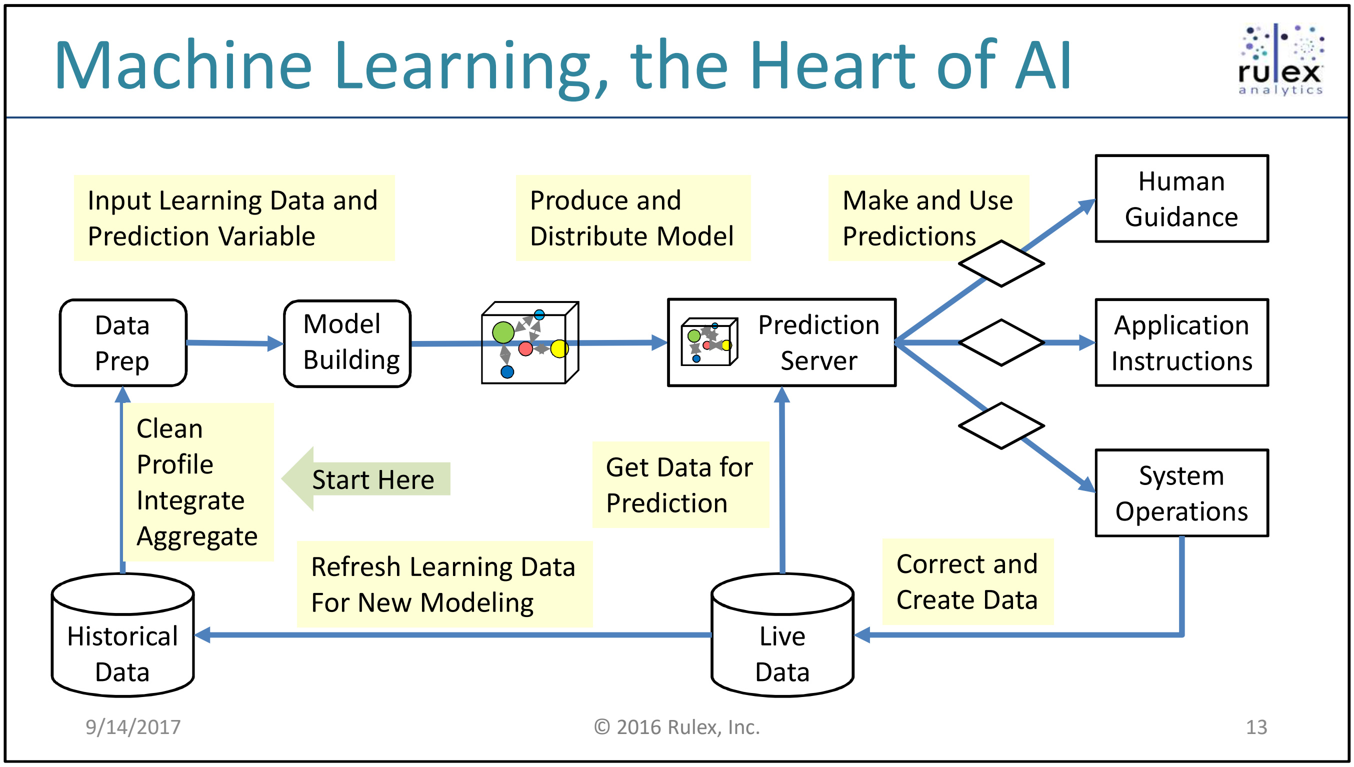
Once the candidate decisions have been identified then it is time to find the data that can support them, that is, what data is currently used to make those decisions manually and what other data may make those decisions easier to automate.

The primary data sources for automating ERP planning decisions are shown here.

The first is the obvious one, the data that supports the ERP planning applications – master data, applications output, spreadsheets and so forth.

However, there is also a wealth of intermediate data aggregations, calculations, and statistics in the ERP applications and middleware that can enable richer predictions on a shorter duty cycle.

And finally, again, there is Dark Data. This is hidden applications data, machine data from factory equipment, network activity logs, sensor signals, and many other things. Dark Data can provide deep mechanical insights and enable unassisted, automated real-time prediction directly in the IT infrastructure.



Here is a general illustration of all the parts of the application of Machine Learning in ERP.

Starting at the green arrow, the process begins with gathering and preparing historical data to produce input data for the machine learning algorithm and identify the variable to be predicted.

The machine learning algorithm uses that input data to produce and distribute predictive models to prediction servers.

Then, new data from the live system is submitted to the prediction server which compares it to the predictive model and produces predictions, which can be used by humans, applications, and systems to make the digital decisions.

Notice that predictive system operations can be used to correct and create data in the live ERP system and the live data can be used to refresh the historical data used for refreshing the models for consistent accuracy.

As you can see, it is not very mysterious, just a continuous data/knowledge circuit that makes the entire ERP system more responsive and effective through decision automation.

What Kind of Machine Learning to Use?

	Best Datatypes	Skills Required	Modelling Methodology	Model Transparency
Deep Learning e.g. Watson, Google, etc.	Images, Video, Unstructured Data, NLP	Advanced Data Science, Math, Programming	Experimental Iteration	None
Math Learning e.g. Neural Nets, SVM, etc.	Database, text, events, time series, signals	Medium Data Science, Math, Programming	Experimental Iteration	Requires Forensic Math Procedures
Logic Learning e.g. Rulex, Xpert Rule, etc.	Database, text, events, time series, signals	Business, Data, and Process Expertise	Knowledge Extraction	Self-Explanatory

9/14/2017 14

There are several different kinds of machine learning and it is very important to select the technology that best suits the desired predictions and available data types.

Shown here by line are three different kinds of machine learning used in AI, Deep Learning, Math-based Learning, and Logic-based learning.

For each, the chart shows the datatypes, required skills, modeling methodology, and model transparency.

Deep Learning systems like IBM Watson are good for data like images and natural language, they require very advanced Data Science skills and use experimental methods for model building, and provide no way to understand why the model works as it does.

Math-based learning methods like Neural Nets and others are good for textual and numerical data, require medium skills, and like Deep Learning, build models through experimental iteration. The only way to understand why a prediction is made requires forensic mathematics to reverse engineer the model.

Finally there is logic-based learning as found in our product, Rulex. It addresses the same datatypes and math learning, but does not require data science skills, automatically extracts models from data with no experimentation, and its models are fully self-explanatory.

About Model Transparency

Math-based methods produce models that cannot be understood by business people.

$$
\begin{aligned}
f(\mathbf{x}) \; = \; & 0.901\, tanh(0.084\, x_{0(0)} + 0.200\, x_{0(1)} - 0.373\, x_{0(2)} + 0.069\, x_{0(3)} + \\
- \; & 0.305\, x_{0(4)} - 0.271\, x_{0(5)} + 0.140\, x_{0(6)} + 0.055\, x_{0(7)} - 0.168\, x_{0(8)} + \\
+ \; & 0.138\, x_{0(9)} + 0.215\, x_{0(10)} - 0.218\, x_{0(11)} + 0.046\, x_{0(12)} + 0.122\, x_{0(13)} + \\
+ \; & 0.284\, x_{0(14)} + 1.071\, x_{0(15)} + 0.279\, x_{0(16)} + 0.171\, x_{0(17)} + 0.087 + \\
- \; & 0.347\, x_{0(18)} + 0.215\, x_{0(19)} + 0.121\, x_{0(20)} + 0.401\, x_{0(21)} - 0.075\, x_{0(22)} + \\
+ \; & 0.223\, x_{0(23)} + 0.204\, x_{0(24)} - 0.379\, x_{1(0)} - 0.032\, x_{1(1)} - 0.409\, x_2 + \\
+ \; & 0.340\, x_3 - 0.312\, x_4 - 0.520\, x_5 + 0.902\, x_6 + 0.459\, x_7 - 0.282) + \\
- \; & 0.013\, tanh(0.452\, x_{0(0)} + 0.449\, x_{0(1)} + 0.001\, x_{0(2)} + 0.338\, x_{0(3)} + \\
- \; & 0.326\, x_{0(4)} - 0.477\, x_{0(5)} - 0.433\, x_{0(6)} - 0.137\, x_{0(7)} + 0.352\, x_{0(8)} + \\
- \; & 0.050\, x_{0(9)} - 0.148\, x_{0(10)} - 0.065\, x_{0(11)} + 0.180\, x_{0(12)} + 0.323\, x_{0(13)} + \\
- \; & 0.097\, x_{0(14)} - 0.401\, x_{0(15)} - 0.503\, x_{0(16)} - 0.216\, x_{0(17)} - 0.184 + \\
+ \; & 0.193\, x_{0(18)} - 0.056\, x_{0(19)} + 0.368\, x_{0(20)} - 0.512\, x_{0(21)} - 0.502\, x_{0(22)} + \\
- \; & 0.221\, x_{0(23)} + 0.457\, x_{0(24)} - 0.501\, x_{1(0)} + 0.108\, x_{1(1)} - 0.218\, x_2 + \\
- \; & 0.201\, x_3 - 0.293\, x_4 + 0.151\, x_5 - 0.274\, x_6 - 0.116\, x_7 + 0.198) - 0.563
\end{aligned}
$$

Accurately predicts problems but provides no knowledge to fix defective business processes.

Logic-based methods produce models in self-explanatory business language.

IF (customer_province in {A, B, C, D} AND damage_class in {1} AND Number of days between policy start and date of accident <= 371) THEN **Fraud = Yes**
IF (Number of days between date of accident and policy end <= 2) THEN **Fraud = No**

Makes accurate predictions and provides business rules that explain the problems and corrections

Here we may dwell on the difference in model transparency between math and logic based machine learning

On the left is the text of a math-based predictive model from a Neural Net system. This is a small portion of a much larger model. As you can see it is a mathematical function that explains nothing. It may make accurate predictions but provides no knowledge for eliminating process flaws.

On the right is an equivalent logic model, here, for easy understanding from a fraud detection applications. The model is self explanatory, showing in simple, first order conditional logic terms exactly why it makes positive and negative decisions.

Decision Automation Solutions

- Recommendation Engine
  - Tell a human what is the right thing to do.
- Application Integration
  - Tell software to do what a human used to do.
- Autonomous Operation
  - Tell software to do what a human can't do.

Here we see the various ways in which automated decisions can be implemented.

In the first, AI is telling a human what is the best thing to do. Everyone who uses Netflix or Amazon's streaming music service has used an AI-based recommendation engine which predicts the movies or music they don't already know about but will probably like. In ERP, similar capabilities are possible to recommend many different kinds of optimum decisions, like materials substitutions, supplier selection, and many others.

In the second automated decision case, AI is telling ERP applications what to do, enabling them to make decisions that were previously performed by humans. This can remove considerable latency from the daily planning workflow and free planners from many tedious, low-value tasks.

Finally, there are autonomous operations are where AI is used to direct software functions that humans cannot perform. This could be because of the required decision speed or complexity, or because of decisions concerning dark data, like ERP master data creation and correction.

Lets look at some use cases that will illustrate some of these capabilities.

Decision Automation Use Cases

- A sampling of Rulex use cases from one of the largest manufacturers in the world
  - Inventory Assignment Snapshot
  - Master Data Creation Snapshot
  - Self-Correcting Supply Chain Data Deep Dive

For more than a year now, Rulex has been working with one of the world's largest manufacturers to implement automated decisions across a wide array of ERP planning functions around the world.

Here we will take a quick look at two familiar planning cases, Inventory Assignment and Master Data Creation, and make a more detailed examination of an unprecedented capability where active master data corrects itself.

Automated Inventory Assignment

Select and ship critical materials for which inventory will go negative in the near future at a certain location.

Assign the priority to a shipment in order to maintain an acceptable level of safety stock.

Select and substitute interchangeable materials to reduce negative inventory and keep good levels of safety stock.

Eliminate thousands of planner decision to save hundreds of planner hours.

The placement of inventory in the right place at the right time is a perpetual planning challenge. It must be fast, accurate, and consistent, even when business conditions change. Like many other manufacturers, this company had been using statistical optimization software to guide the process. Sometimes the results were good, and other times the recommendations resulted in SLA penalties. To remedy this inconsistency, it required advanced skills to apply corrective rules of thumb to the optimizer output.

Using Rulex AI, they achieved greater speed and accuracy by automating decisions that included the ones shown ere.

As a result, they are eliminating thousands of planning decsions and saving hundreds of hours every month.

Automated Master Data Creation

Find anomalies in the master data and provide the right values to fix them.

Find the logic behind the way master data values are assigned and express it in an understandable way.

Predict the value of new materials which will be activated in the master data database.

Eliminate mistakes and optimize materials value in master data activation.

9/14/2017 19

For most manufacturers, creating Master Data records for new materials, packaging, and other things, is a problem-prone process. Missing information, regional variations in naming and coding conventions, subjective value selection, and other factors contribute to the creation of incomplete, incorrect, and incompatible records. As the atoms of the ERP information processes, faulty master data can cause numerous problems, costs, and inefficiencies throughout the entire system.

This manufacturer is employing Rulex AI to create master data records that are accurate and complete by finding and fixing missing and incorrect data field values, exposing and explaining the logic of correct and incorrect record creation processes, and predicting the value of new materials as they are activated.

The results are cleaner master data, smoother data utilization, and optimized materials value, savings millions of dollars every year.

Self-Correcting Supply Chain Data Deep Dive

- Use AI to automatically find and fix errors in active supply chain data with no a priori knowledge of data or errors.
  - Find errors and fix them as fast as possible.
  - Don't impact production system performance.
  - Be able to audit and explain all corrections.

Now we will make a deep dive into a revolutionary ERP capability enabled by Rulex – self-correcting supply chain data.

The solution operates on active, in-process ERP data and uses Rulex AI to automatically find and fix data errors with no prior knowledge of the data or error conditions.

The solution is fast, transparent, and all corrections are fully explainable.

Let's look at the details.

Data Errors Can Be Costly

- Excessive inventory and transportation costs.
- Inefficient manufacturing production processes.
- Order fulfillment errors violating customer SLAs.
- Inefficient use of business expert resources.

These are just a few of the ways in which live data errors impact supply chain enterprises.

Revenue, Profitability, Customer Satisfaction, and Competitiveness, are all places where these errors can have material, measurable negative business impact.

One error in the wrong place at the wrong time can cost millions of dollars.

Data Correction Challenges

- Too much data for human sample identification.
- Data too dynamic for long modeling cycles.
- Too much data for record-wise anomaly review.
- Predictive models don’t give reasons for errors.

We submit that this traditional approach is unworkable in a large supply chain.

Even the most expert planner will only be able to find and tag a fraction of the kinds of anomalies that exist and building a sufficiently large sample of example cases is an extraordinarily time consuming process, taking as much as 80% of the time needed by the data science process.

Also the data is always changing throughout the day and the process cycle. This doesn’t bode well for machine learning algorithms that can take a day or two to build a single model.

And when the model is used to predict anomalies in new data, then what? Each one must be reviewed and possibly corrected by a business expert. There is just too much data for that to be practical in a large or complex supply chain.

But, the biggest problem by far is that these algorithms produce predictive models that cannot be explained or understood. They are math functions. They can tell you that they think a record is an outlier, but they cannot tell you why they think that.

Types of Supply Chain Data Errors

- Descriptions
- Identifiers
- Values
- Relationships
- Limits
- Thresholds
- . . .

Supply chain data error can be found in descriptions, identifiers, classifications codes, control variables, and many other places.

Some of these errors may be the result of incorrect human input, incorrect relationships between data elements and data sets, incorrect output from statistical optimizers and other software, mangled EDI transmissions, and any number of other causes.

And, because of the interconnected nature of supply chain data, an error in one field can create errors in another field.

Supply chain errors can also be subtle, involving obscure applications data not regularly examined by any person or process.

Based on our research, for every known type of error, there will be hundreds of unknown types, and for every known error, there may be thousands of others going undetected until they cause problems.

Data Anomalies

- Field values inconsistent with “normal” values.
- Can be Outliers (just unusual) or Errors (wrong).
- Occur infrequently and are hard to find.
- Only humans can discriminate between the two.

9/14/2017 24

Let’s look more closely at the nature of these errors.

Terminology is very important here. A data anomaly is no the same thing as a data error.

An anomaly is just a case where the value in a data record field is not like the others in that field. It may or may not be an error.

The field may contain a value that is unusual compared to the others, but not incorrect. For example, this could be an unusually high or low temperature reading that is still within an allowable range. These values are outliers and should be left alone.

And then there values that are truly wrong. These must be found and fixed.

The problem is that anomalies in general, and errors in particular occur infrequently, making them hard to find. They can also be very subtle and hard to identify.

Only an expert human, like a supply chain planner, can tell you if a given anomaly is an outlier or an error.

Typical Anomaly Detection/Correction Process

- Planners and other subject matter experts identify sample errors for use by machine learning algorithms.
- Data scientists build model to predict anomalies in new data.
- SMEs eliminate outliers and correct erroneous records.
- SMEs define and submit database record corrections.

9/14/2017 25

You will see an increasing number of solutions purporting to correct supply chain data using conventional machine learning techniques in the hands of trained data scientists.

The machine learning finds the suspect cases and then planners and others who understand the data my separate the errors from the outliers and then correct the errors manually, one by one.

The models find the anomalies but they don't correct the errors. People must do that.

From Data Inspection to Pattern Evaluation

Data Inspection

Var1	Var2	Var3	Var4	Normal
…	…	…	…	…
■	✪	⊙	★	✗
■	❖	⊙	★	✓
⊙	◆	■	✪	✓
⊙	◆	■	★	✓
⊙	❖	⊙	✪	✗
■	◆	⊙	★	✓
⊙	❖	■	✚	✗
■	◆	⊙	✪	✓
■	◆	⊙	✪	✓
…	…	…	…	…

- Lengthy, costly procedural process.
- Difficult to understand why anomalies are generated.

Pattern Evaluation

If Var2 = ✪ then Anomaly
If Var1= ⊙ AND Var3≠ ■ then Anomaly
If Var4= ✚ then Anomaly
….

- Only a few patterns to review (resource and time saving)
- Clear meaning of the anomaly (logic approach).

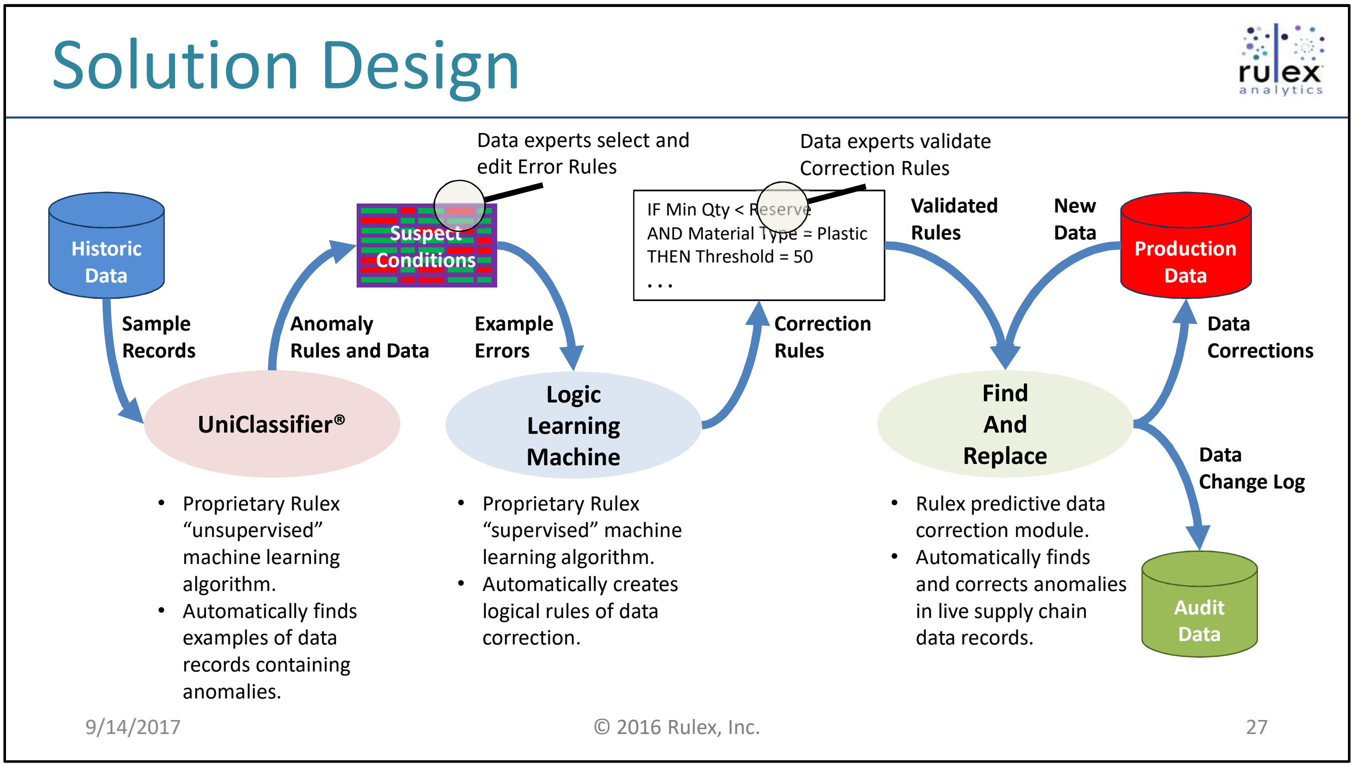
9/14/2017 26

Here on the left is a symbolic representation of the data inspection methodology we discussed a moment ago, and on the right we see the way SC Squared deals with the problem of supply chain data errors.

In the traditional method, individual records show differences from the rest that indicate that they are suspected anomalies. These must ultimately be qualified and corrected if needed by planners and other experts.

Rules SC Squared works differently. It finds unusual patterns that exist in multiple records in the raw data and expresses them in the form of conditional logic statements that say exactly why the various records are or are not anomalies.

On the left we have a record-by-record process. On the right a short review of the logic governing all the anomalous record.



Here is a complete, high level picture of the Rulex SC Squared solution. The system comprises three primary software modules.

The UniClassifier is new proprietary Rulex machine learning algorithm. It is said to be an unsupervised algorithm because it does not require any a priori knowledge of the data, nor does it require pre-identification of example cases. In SC Squared, its job is to find examples of anomalies in historic data and to produce anomaly rules for expert review.

The Logic Learning Machine then takes the suspect records tagged by the UniClassifier and builds the logical model for correcting the data errors.

Experts can then validate predictions and corrections to configure the third software component, the Find and Replace module. This component then automatically, continuously examines new data in the system and identifies and corrects erroneous records in the life system.

For effective governance and business alignment, the system produces an audit log of all automated changes to the supply chain data.

Operational Workflow

- Planners and other process experts use Rulex Desktop to extract “rules of wrongness” from historical data.
- Planners and experts validate rules and configure Rulex Find and Replace module for replacement behaviors.
- New record instances are submitted to F&R module in batches or streams.
- F&R detects new errors and performs predictive data replacement and logs all changes.

9/14/2017 28

Here is a summary of the human side of what we just looked at.

Planners and others use desktop software to discover, qualify, and edit the anomaly rules from historical data and configure the corrective rules to be applied to new data.

Then data automation is used to submit new applications data to the Find and Replace module, which automatically finds and fixed new errors and and provides a variety of outputs for business users, including the audit log, statistical analysis, and event signals for initiating other automated processes.

Solution Summary

- Automatically detects anomalies and automatically corrects data errors.
- Simplifies and accelerates outlier elimination.
- Operates on live applications data.
- Non-intrusive, non-disruptive, easy to use.
- Leverages existing data/process expertise.
- Works as an "Intelligent Assistant" for planners.

9/14/2017 29

Rulex SC Squared is a decidedly different solution for supply chain data correction.

It automatically detects anomalies and automatically corrects data errors. Remember we said that only experts can distinguish between anomalies.

However, with Rulex, the experts don't need to look at hundreds or thousands of suspect data records. They only have to examine a handful of rules that qualify all the anomalies in the data. Eliminating one anomaly rule automatically eliminate thousands of corresponding actual anomalies. More on this shortly.

Rulex SC Squared operates on live data. This is the data in applications databases, process workflows, and so on.

The solution is non-intrusive, non-disruptive, and it is very easy for business people to use without leaning any new skills.

It leverages the knowledge, experience, and skills of planners and other experts in a highly efficient and effective way, working like an intelligent assistant that helps them do their normal jobs much more effectively.

Common Mistakes to Avoid

- Data
- Modeling
- Prediction

We will now close with a look at some common mistakes we have learned to avoid in implementing AI for ERP planning decision automation – data mistakes, modeling mistakes, and prediction mistakes.

Data Mistakes

- Machine Learning from a Data Lake
  - Weak/missing signals, no dark data, lost data associations
- Underestimating Data Governance challenges
  - Data silo politics, governance, security, access impact
- Insufficient Data Hygiene
  - Subtle errors and inconsistencies, missing values

9/14/2017 31

Data Mistakes to avoid.

First, don't use a BI data lake for machine learning. Most data lakes contain data that has been processed, aggregated, and altered for ease of access and simplicity of human analysis. For machine learning this means that many valuable signals have bee removed or weakened, there is no dark data, and the original associations between data sources is often lost. Use raw applications data instead. It contains all the facts needed to build effective predictive models.

Second, don't underestimate the challenges of getting the data you need for AI. In addition to routine data silo politics between the planning and IT organizations, the need for best practices for security and master data governance in data sharing, and the potential impact of new, unfamiliar processes touching live system data, mean that you should budget considerable resources and time for identifying and acquiring the data you will need for effective machine learning.

Finally, don't assume that your data is clean and correct. In the master data creation use case, the company was quite sure that their error rate was less than 5%. After using Rulex to discover the actual errors, they were stunned to discover the errors rate was over 25%. Don't neglect the certification and correction subtle errors and inconsistencies and the potential negative impact of missing data.

Modeling Mistakes

- Choosing the wrong machine learning method
  - Mismatched to goals and available data or skills
- Underestimating skills needed for Math Learning
  - Python/R programming, statistics, mathematics, etc.
- Poor model testing, validation, and management
  - False positives, hidden inaccuracy, missed system changes

9/14/2017 32

Modeling Mistakes to avoid.

Choosing the wrong algorithms. As we saw earlier, there are various kinds of machine learning with differing requirements and capabilities. Take care to align your choice with the kind of data you have, the predictions you want to make, and your requirements for explainable predictions.

Next, if you do select Math-based machine learning algorithms, like Neural Nets, don't underestimate the level of data science skills you will need to make them do anything useful. Good data scientists command high salaries, they are hard to find and harder to retain, and it takes a long time for them to gain the needed ERP data expertise to be fully productive.

Finally, recognized that models can make the right decisions for the wrong reasons, or can lose accuracy over time. Many new users of machine learning do not spend sufficient time testing, validating, and refreshing their models, resulting in unknown errors and missing the effect of changes in business conditions. Select machine learning tools that make it easy to visualize data and models, verify model accuracy and coverage, and to share and maintain models in a production ERP environment.

Prediction Mistakes

- Disrupting/Impacting Business Processes
  - Prediction “design” forces changes to workflows
- Underestimating Implementation Complexity
  - Data access, application integration, file management
- Not Updating Models Frequently Enough
  - Changing data, business conditions, mfg. processes

Prediction Mistakes to avoid.

Good decision automation simplifies and eliminates planning decision tasks, while incorrect, discovery-oriented use of machine learning can force the need for new and different workflow tasks and procedures. Start with the current decision workflow and look for the tasks that can be eliminated or simplified, rather than starting with the data and looking for new things to predict.

Especially for math-based AI, implementation can create many new requirements for IT support, including new data sources, middleware operations, file sharing, and many other things. It is absolutely essential that IT stakeholders are included in the project early and that they will have the budget to do their part.

Once you have built an accurate predictive model, it will only stay accurate if it is continuously rebuilt. Especially with math learning, building a model can take a very long time and consume considerable compute resources, and as a result, it is not done as frequently as it should be. Over time data, business conditions, and production processes all change, sometimes in ways that are very subtle and that have a huge impact on prediction accuracy. You must be diligent in keeping models up to date and correctly aligned with data and prediction goals.

Rulex - AI That Tells You Why®

Thank You

To receive a copy of this presentation including full speaker notes, email t.negris@rulex.ai

For more information about Rulex, the AI that tells you why, please visit www.rulex.ai.