compellon™

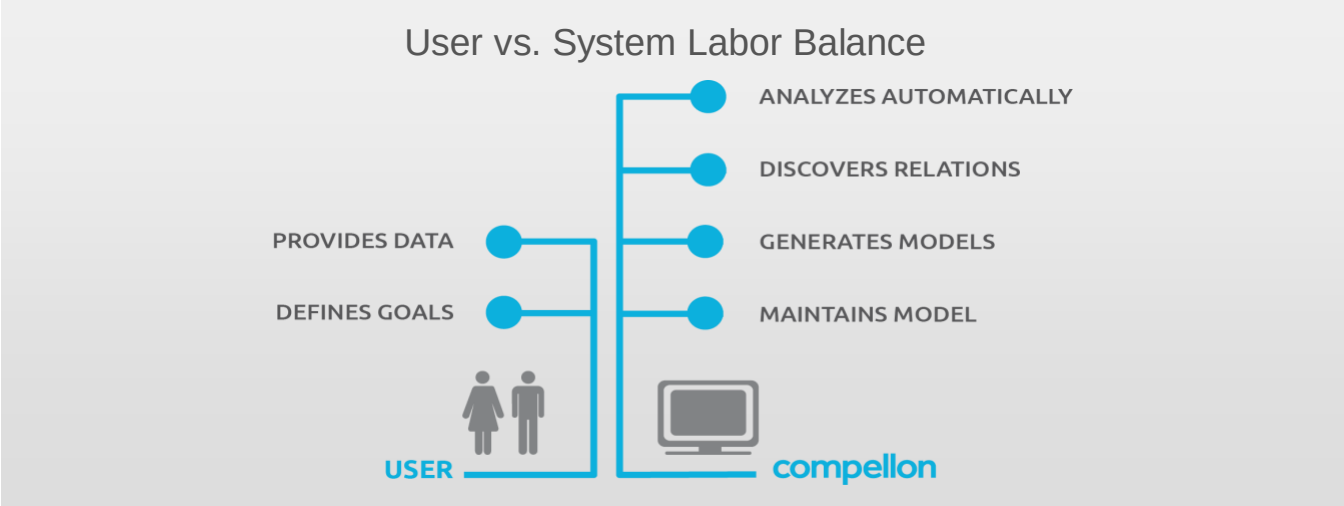
Technical White Paper

Overview

Compellon 20|20 is a clear-box analytics platform that automatically identifies the relationships in your data while learning and adapting over time, so you can continuously optimize your business processes to drive the outcomes that matter most. Today, Compellon 20|20 is helping companies in a range of industries reduce failures on manufacturing lines, improve the customer experience and the employee experience, and target and personalize communications and offerings to customers most at risk or likely to buy. The common thread across all of these use cases is the ability to automatically identify key drivers, predict with confidence using a custom model, and effectively change business outcomes using strategic and tactical actions recommended from the evidence in the data, even in the most dynamic environments. All of this is powered by Compellon's Intelligent Agent that autonomously learns and understands information and relationships in the data. By using Compellon 20|20, teams save time and money by using the right information to focus their efforts on the areas that matter most to their business outcomes.

Compellon's clear-box machine learning is an innovative approach to artificial intelligence pioneered by Compellon's Chief Scientist, Dr. N. N. Lyashenko (Liachenko). His focus is advancing AI along a path that *augments* human expertise, rather than attempting to imitate it. When applied to using data-informed decisions and models to improve business outcomes, his vision is to shift the balance of effort so that the human focuses only on defining the business challenge to be explored and resolved, while the machine does all the analytical work, such as finding relationships and proposing hypotheses. Over years of solving problems otherwise intractable using existing methods, Dr. Lyashenko developed his vision into an implementation that is practical for everyday operation. The result is an approach that can be truly automated from end to end because knowledge discovery and predictive model construction are *based solely on the evidence in the data*, without any user-imposed analytic assumptions, or bias, unrelated to the business goals.

Compellon has incorporated this approach into a clear-box analytics platform that makes sophisticated capabilities accessible to subject matter experts and data analysts, regardless of statistical or data science skills. With Compellon 20|20, the user need only provide the data and their domain specific goals.



In the typical data science practice of today, just figuring out which data matters before even starting to build predictive models requires extensive judgement, bias, and iteration by an expert across a complicated workflow and a range of tools. Compellon 20|20 removes this manual effort to transform and distill down inputs into a few variables. Instead, Compellon20|20 uses sophisticated algorithms that *automatically* identify which variables matter, including their hidden patterns and relationships, and discovers those most predictive along with their combination effects. Compellon 20|20 can do this even on data sets with thousands of columns.

Instead of the typically extensive trial-and-error effort to fit data into preconceived model types, Compellon 20|20 generates *custom* predictive models every time, to further simplify and speed the effort involved, and to avoid the pitfalls of constrained or faulty hypotheses. The platform fully assesses the characteristics of the data before making any analytical decisions. This notably contrasts the de facto analytical incumbents that force the user to make many choices regarding data transformation and statistical or machine learning methods just to get started.

Finally, this unique approach powers delivery of prescriptive outputs that go beyond just predicting outcomes to explaining *why* and advising *what to do*. With this unprecedented intelligent and bias-free approach, Compellon 20|20 dramatically broadens accessibility of advanced analytics to more users and organizations across a much broader scope of problems than previously possible.

Technology

At the heart of Dr. Lyashenko's approach are techniques and algorithms to find dependence patterns in data. For example:

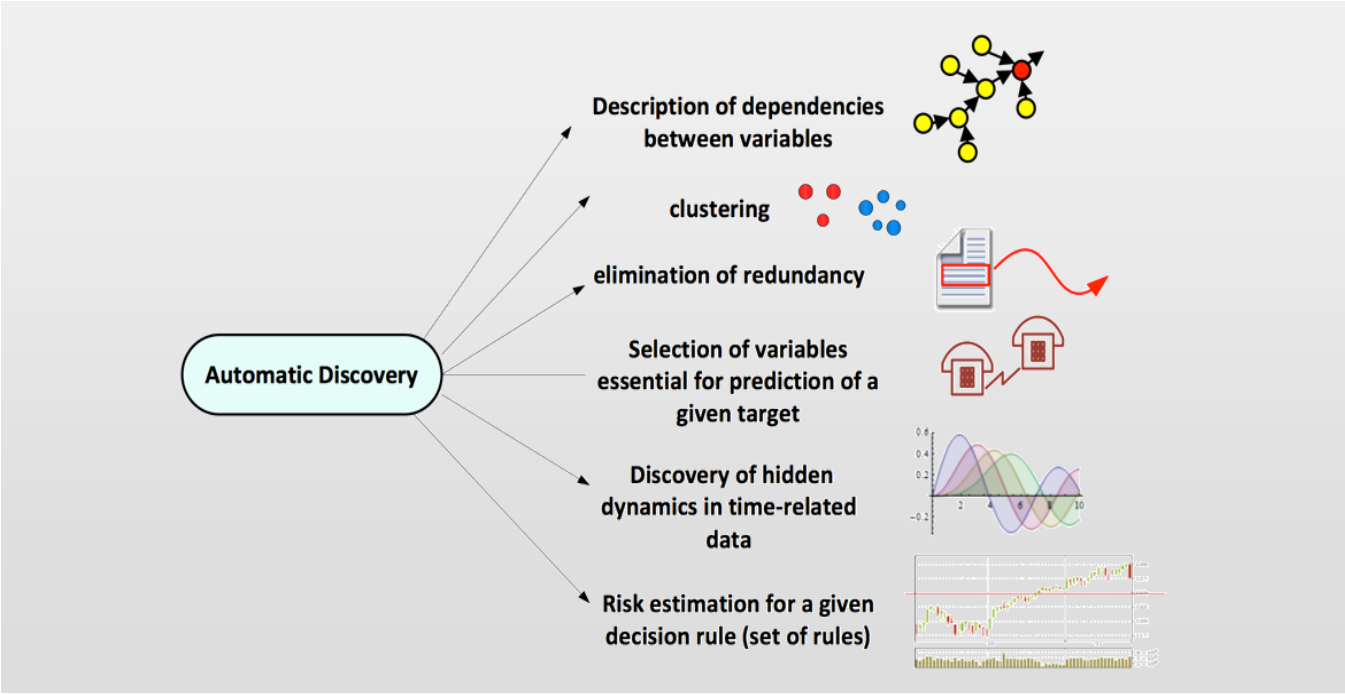
- Automatic assessment of the quality of a data set for building predictive models
- Rapid variable reduction and identification of the "driving forces" behind outcomes
- Fast development of custom models capable of predicting outcomes and providing forced ranked drivers for why outcomes are predicted
- Ability to autonomously monitor and adapt predictive models to maintain relevancy as data changes

The algorithms heavily apply "information theory" to uncover high-dimensional relationships. This approach is especially powerful in revealing hidden "group effects," in which variables interact and collaborate in interesting ways, uncovering important patterns typically overlooked

(or found only after a lengthy process) with traditional data science methods. It also allows Compellon 20|20 to overcome limitations present in many other methods reliant on simplifying assumptions such as variable independence, commonality of scale, and linearity.

Compellon's method for processing data occurs in several stages. The first is the "data health" test to ensure the data representation in the user data set does not exhibit flaws that would damage the quality of analysis. For example, if it is obvious that a constant variable (i.e., the same in all observations) has no predictive qualities, it is automatically removed from the analysis.

The next stage of the process is the reduction of the number of variables in the input data set, analogous to the "feature selection" step in other machine learning approaches. Here, Compellon 20|20's powerful discovery methods are novel in that they do not use standard model-based measures of dependence – or any assumptions, as are typical in ordinary data science methodology – that might otherwise miss discovering the multi-dimensional relations among variables. In fact, the engine obtains all of its information about the variables early in the process, before any decisions about eventual model type or variables have been made. Consequently, discoveries made at this intermediate step of finding the ultimate drivers of the predictive model also reveal extremely important information about variables and relationships that help the line of business or analyst user diagnose and solve the business problem at hand. Examples include guidance on improving data collection, understanding the themes and chains of factors or events that contribute to outcomes, and anticipating the side effects of different decisions.



The final stage of the process is the construction of a custom predictive model. Again, this aspect of Compellon 20|20 is distinct from the mainstream approach that relies on fitting data to existing algorithms (for example logistic regression, neural networks, or SVM) through either iterative handcrafting by expert data scientists or via automation that silently tests pre-canned models to find the assumed "best fit." The Compellon engine custom constructs a model for each data set that immediately has an optimized balance of predictive accuracy with stability. It also pays

special attention to group effects in which drivers combine; more often than not, variables that have weak predictive power on their own make a disproportionally stronger predictor as a group.

Throughout the process, Compellon 20|20 focuses on incorporating the relevant business goals, enabling users to tune models to best serve their goals. For example, the user can inform Compellon20|20 that there is a high cost to missing a rare diagnosis relative to the cost of a potentially extraneous follow-up test, or that there is a high lifetime cost of actually losing a customer versus the modest cost for incentives to mitigate the risk. Compellon20|20 automatically tunes the drivers selected and the model structure to produce a model that appropriately balances the cost impact of an erroneous prediction.

Equally as important as automated modeling, this approach also delivers unique capability advantages. Models go beyond just predicting positive or negative outcomes by providing the confidence for the prediction and the reasons, in order of their contribution, for each individual prediction. Models also generate quantitative prescriptive advice to identify the optimal best next actions to take to most efficiently achieve the desired business goal.

Finally, another exciting direction enabled by this unique process for constructing predictive models will build on the engine's ability to know *why* it created the model in the way that it did and leverage this inherent self-awareness. For example, the engine knows when *changes* in the data require adjustments to the model, and those changes can be made automatically, without human intervention or assumptions. This positions the Compellon 20|20 platform for further breakthroughs in speed and simplicity in the deployment of predictive solutions to meet the demands of increasingly dynamic environments and use cases.

USER INTERFACE

Compellon's core technology requires no human guidance and assumptions for the underlying analytics, enabling Compellon 20|20 to deliver bias-free insights that can be trusted more than any other predictive modeling platform. The Compellon 20|20 UI is composed of three sections: "Review", "Model", and "Action". Review focuses on data preparation, Model focuses on defining the business goal and exploring the automatically-generated model and discoveries, while Action focuses on using the model to drive decisions and actions.

Review focuses the user on exploring the input data, confirming or correcting the automatic assessments of data types, and optionally applying business judgment to focus results later in the process. Because Compellon 20|20 retains data in native human readable form throughout, categorical and ordinal data require no numeric encoding or splitting into multiple columns. Even column headers can retain a full text description, such as the actual text of a survey question. In addition to simplifying data preparation, this makes outputs easily consumable by line of business users and their stakeholders.

Another way in which Compellon 20|20 simplifies data preparation is by processing many input columns as-is, while other systems would require cleaning and possibly removing them. Examples include columns with missing values, text inadvertently mixed into a column of numeric data, and columns that later prove to be a perfect proxy for the target variable.

Model focuses on detecting relationships and patterns in the data and developing the models to represent the business goal. One of the defining characteristics of the Compellon 20|20 user

interface is focusing the user on defining the business outcome of interest in the context of their data (for example, a KPI or metric, or combination of both), and this is done in the Model area. All models and outputs are generated automatically based on the objective, plus information contained in the data. (Though Compellon 20|20 builds predictive models based on this one step, advanced features are available for experienced users to refine and tune to their business goal.)

Once the target is identified, the system performs the analysis and generates the model and the outputs. Variables relevant to the target, and the predictive model and its quality, can all be explored. During its automatic discovery process, Compellon 20|20 uncovers information not generally associated with standard modeling techniques that are delivered in unique and valuable outputs. For example, the relationship graph shows how variables in the data set that are relevant to the target directionally relate to each other and the target outcome. An easily explainable visualization of the model structure illustrates the relationship of drivers to predictive outcomes, including the non-obvious "group effects."

Because of the powerful capabilities for rapidly exploring any outcome in the data, some use Compellon 20|20 primarily for discovery. For those building models that will be used to make predictions on new data observations, users can assess the quality of the model with common industry metrics around predictive accuracy, including easily validating with a test set.

Action focuses on applying the models developed to drive action. One action is Bulk Predict. Here, new data observations are uploaded and "scored" using the predictive model. For each outcome, the output includes the prediction, the confidence, and the reasons in order of precedence to enable personalizing offers or prioritizing targets of interest. These results can be downloaded or queried through API calls to incorporate into operational workflows.

Another action is to get right to the answer many Compellon 20|20 users start with, which is "how do I achieve my goal?" For this, the platform delivers an interactive Advisor built on the predictive model. For example, if the current achieved outcome is 10% of the population, and the goal is 15%, the Advisor will propose several scenarios, in order of effort, for achieving the target improvement by changing the input amounts.

Finally, Profiles shows the unique combinations of inputs that represent a high probability of meeting the targeted outcome, for example, as a persona of a likely buyer.

PLATFORM & INTEGRATION

The Compellon 20|20 platform is architected as a modern SaaS platform. All functionality is exposed using a simple yet powerful REST API that allows for easy integration with application and data sources. Compellon 2020 utilizes Apache Spark for horizontal scalability and industry standard cloud environment to scale to the largest of datasets. Trained models can be integrated into the production environment through the platform's APIs. Additionally, trained models can be exported as Java objects where ultra-high speed predictions are required (for example in manufacturing, web analytics, and financial services).

APPLICATIONS

Compellon 20|20 is ideal for environments in which organizations need explainable and bias-free insights to know what to do and why, getting answers rapidly enough to act even in changing

environments, and doing so with diverse data, even with thousands of input columns. Examples today include reducing failure on manufacturing lines, identifying how to improve customer satisfaction on survey data, targeting marketing spending to increase engagement and program effectiveness, and reducing customer churn by understanding the causes and what personalized actions to take to prevent it.

In general, many use cases substantially benefit from Compellon 20|20 if the desired business scenario can be framed through a binary or numeric target present in the data. Additionally, the platform's automatic discovery can also be used as a productivity booster for data scientists to speed their exploration of a data set's features and help them to optimize data-collection. For example, the Compellon engine will easily reveal variables in the input set that are proxies to the target that will confound regression analyses.

CONCLUSION

More and more businesses recognize the criticality of applying data driven insights to optimize and improve everyday operations such as manufacturing, customer experience, marketing, and research and development. This criticality will only increase as more data is created every day and a broad range of industries undertake the digitization of almost everything. Yet the combination of the dynamic environment with rapidly changing data, shortage of data science talent and skills, and limitations of today's tools make progressing from using data for rear-view reports to guiding where to go nearly impossible for many companies.

Compellon believes solving these challenges requires a revolutionary approach; an approach that empowers those closest to the business and the data to achieve better business outcomes utilizing advanced machine discovery and machine learning – without assumptions. Leveraging the power of the breakthrough algorithms and methodologies developed and tested over 30 years of practice by Dr. Lyashenko, only the Compellon 20|20 platform provides the unique combination of clear-box results and bias-free insights needed to ensure businesses can truly use the power of their data to save time and money by focusing efforts on the areas that matter most to their business outcomes.