



COMPLEXUS ANALYTICS

Fundamentos conceptuales y metodológicos del sistema Complexus Analytics

Marzo 2021



Contenido

Introducción	3
Resumen	3
Procesos de Adaptabilidad en los Sistemas Complejos	4
El azar salvaje.....	5
¿Cuánta información aporta y cuánta información oculta un KPI?	7
Ejemplo 1) Los promedios no aportan información válida para detectar cambios de comportamiento	7
Ejemplo 2) Los valores globales pueden ocultar información sobre comportamientos desagregados	8
Ejemplo 3) Oportunidades o Ineficiencias que se presentan bajo situaciones que involucran condiciones múltiples.....	9
Analítica de Sistemas Complejos	10
Análisis Exploratorio de Datos	11
Técnicas utilizadas por el EDA	12
Subjetividad	12
Generalidad	13
Utilización de toda la información disponible.....	13
Inteligencia Artificial Interpretable (XAI)	13
Problema desafío.....	14
Modelo de IA.....	15
Perfilamiento genético.....	15
Árbol de decisión para agrupar los micro-patrones	16
Clustering de mini-patrones	17
Interface explicativa	18



Introducción

Complexus Analytics es un software de Inteligencia Artificial que analiza datos provenientes de organizaciones complejas para descubrir oportunidades y problemas que son difíciles de detectar mediante métodos convencionales.

Facilita la adaptabilidad de la organización al cambio permanente. Fomenta el desarrollo de ventajas diferenciales y comparativas. Descubre situaciones y comportamientos relevantes antes de que sus efectos produzcan impactos visibles sobre los indicadores clave de performance (KPI).

El logro de estos beneficios es posible gracias al uso de una innovadora metodología orientada a la creación de conocimiento que rompe con los paradigmas convencionales del análisis de datos.

Este documento describe sus fundamentos teóricos y los criterios utilizados para su implementación práctica.

Resumen

El fundamento central de Complexus Analytics es que el contexto globalizado, innovador y competitivo ha transformado a las organizaciones en sistemas verdaderamente complejos.

La complejidad de las organizaciones se aprecia en aspectos cuantitativos como la creciente diversidad de productos y servicios o la cantidad de datos que se registran y se almacenan. También se manifiesta en la aparición de una gran cantidad de técnicas para afrontar los cambios del entorno como por ejemplo: tercerización, joint venture, alianzas, partners, clusters, ecosistemas, interoperabilidad, auto-organización, colaboración, co-creación, reestructuración, reorganización, reingeniería, resizing, rightsizing, downsizing, as a service, etc.

En estas condiciones el desafío de la gestión ya no se limita a tener que realizar un efectivo **control** de metas y objetivos. Se añaden nuevas exigencias entre las que se destacan: 1) detectar **ineficiencias** que se **ocultan** bajo la complejidad de los procesos, y 2) asegurar la **adaptabilidad de la organización** a un contexto en permanente cambio.

Estas nuevas exigencias de gestión requieren información específica. La mayoría de las organizaciones cuenta con registros de datos que les permitirían obtenerla mediante los cálculos apropiados. El problema es que las metodologías y herramientas de análisis disponibles en su mayoría han sido concebidas para optimizar el control de metas y objetivos, y resultan muy ineficientes cuando se las quiere usar para obtener información con mayores niveles de anticipación y desagregación.

Es así, que la finalidad central de Complexus Analytics es encontrar eficientemente información a tiempo y detallada para detectar oportunidades e ineficiencias ocultas, y para asegurar la adaptabilidad de la organización.

Pero para lograr estos fines no alcanza con aportar nuevas herramientas, sino que es necesario revisar los fundamentos de la gestión cuantitativa tradicional. Por eso Complexus Analytics parte de un conjunto de innovaciones conceptuales y metodológicas de tipo disruptivo. En primer lugar utiliza un enfoque analítico diferente (**analítica de sistemas complejos**) que en lugar de poner el énfasis en encontrar regularidades (o patrones) procura detectar las irregularidades relevantes. En segundo lugar hace uso de una innovadora metodología de análisis (**Análisis Exploratorio de Datos**), que permite investigar esa gran cantidad de situaciones irregulares para descubrir nuevos conocimientos aplicables. En tercer lugar implementa estos enfoques metodológicos mediante un nuevo abordaje de la Inteligencia Artificial denominado XAI (**Explainable Artificial Intelligence**) que permite combinar



interactivamente la potencia de análisis del software con la experiencia y los conocimientos subjetivos que aportan los seres humanos.

A continuación se describen en detalle estas metodologías.

Procesos de Adaptabilidad en los Sistemas Complejos

Desde el punto de vista sistémico el *control de Indicadores organizacionales* se basa en el **control por realimentación** creado por Norbert Wiener a mediados del siglo XX como parte de la Cibernética (o ciencia del control y la comunicación en máquinas y animales).

La Cibernética surgió para asegurar el funcionamiento estable de los sistemas frente a las variaciones circunstanciales del entorno. Su punto de partida es que los sistemas dinámicos aseguran su funcionamiento estable mediante un ciclo de realimentación que consta de dos fases: la primera se encarga de **comparar los resultados obtenidos con los propósitos establecidos**, y la segunda de desencadenar **acciones correctivas** en caso de ser necesarias.¹ Este ciclo fue la base conceptual sobre la que se desarrolló la automatización industrial en la segunda mitad de SXX. A su vez fue la base matemática para la **gestión de indicadores** y los **Tableros de Mando** organizacionales.

Si bien los aportes de la Cibernética fueron imprescindibles para el diseño de sistemas más efectivos y eficientes, a mediados de los 80 la **Teoría de la Complejidad** amplió los desafíos y añadió el propósito de asegurar su **evolución adaptativa**.² Esto significa que se pasó a considerar que no era suficiente con asegurar que los sistemas respondieran adecuadamente a los cambios previsibles del entorno, sino que también era necesario (y posible) instrumentar formas de aprendizaje que procuraran asegurar la continuidad adaptativa del sistema frente a los cambios no previstos.

En el plano industrial la incorporación de esas capacidades de adaptación y aprendizaje permitió que los autómatas cibernéticos se transformaran en las máquinas inteligentes que hoy constituyen la incipiente **cuarta revolución industrial** (4IR). Pero existe bastante atraso en la incorporación de estos conceptos en el plano de la gestión cuantitativa de las organizaciones.

En general el “estado del arte” de la gestión cuantitativa de las organizaciones aún se asocia con los **Tableros de Mando**, pero en esencia estos están concebidos para asegurar el control del cumplimiento de los planes establecidos, y su diseño no contempla la necesidad de **gestionar la adaptabilidad**. Si bien es fundamental que la gestión de las organizaciones realice un control efectivo mediante el cálculo y evaluación de sus KPI, la realidad actual exige ir más allá. Si la información se limita a sustentar el control solo permitirá reaccionar frente a hechos consumados. El **control** se ocupa de reaccionar frente hechos totalmente constatados e ignora potenciales indicios derivados de cambios incipientes que aún no se manifiestan con total claridad. Para que las organizaciones

¹ Wiener brinda un sencillo ejemplo sobre cómo operaría el control por realimentación en la calefacción de una vivienda: “Se parte de una posición para la temperatura deseada y, si la temperatura de la vivienda desciende por debajo de lo establecido, se conecta un dispositivo que abre el regulador o aumenta la combustión para que la temperatura suba hasta el nivel determinado. Si por el contrario, la temperatura excede el nivel deseado, el regulador se cierra o la combustión desciende o se interrumpe. De este modo, se mantiene la temperatura de la casa a un nivel constante aproximado.” Wiener, N. (1985). *Cibernética o El control y las comunicaciones en animales y máquinas*. Barcelona: Tusquets (p. 135)

² Murray Gell-Mann, creador de los **Sistemas Adaptativos Complejos** (CAS por sus siglas en inglés) lo expresó así: “en los comienzos de la cibernética los autómatas no eran capaces de crear un esquema susceptible de evolución... Sólo ahora estamos entrando en una era de autómatas que pueden considerarse verdaderos sistemas complejos adaptativos” Gell-Mann, M. (2003). *El quark y el jaguar: Aventuras en lo simple y lo complejo*. Barcelona: Tusquets



puedan gestionar su adaptabilidad en forma ágil necesitan métodos cuantitativos que vayan más allá del control y que aporten información útil para detectar los cambios cuando aún son incipientes y todavía no llegan a impactar notoriamente sobre los KPI.

A continuación se exponen algunas de las causas por las que los KPI no aportan información suficiente para gestionar la complejidad.

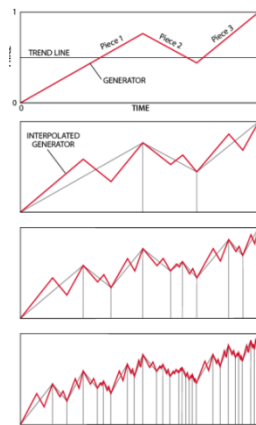
El azar salvaje

Una de las razones por las que los indicadores no aportan información suficiente para gestionar la complejidad es que su cálculo se realiza en base al supuesto de que la realidad en la que se mueven las organizaciones está gobernada por una forma de azar **suave** o **benigno**. Pero los procesos complejos suelen estar gobernados por un azar de tipo **salvaje** que altera el sentido y significado de los KPI como se mostrará a continuación.

En la década del 60, el matemático Benoit Mandelbrot analizó en detalle los datos de la *cotización bursátil del precio del algodón en los mercados centralizados del Medio Oeste norteamericano a lo largo de un siglo*. A partir de ese análisis realizó importantes descubrimientos entre los que se encuentra la concepción de la matemática fractal y la diferenciación de dos formas de azar a las que denominó respectivamente **benigno** y **salvaje**.

Para llegar a estos descubrimientos Mandelbrot empezó por constatar que los métodos de la estadística tradicional no se podían aplicar a los datos que estaba analizando, porque en ellos no se cumplían las restricciones básicas que esos métodos exigen.

Una de las restricciones de los métodos estadísticos básicos es que los eventos que se analizan sean independientes entre sí. Esta condición se verifica en varios fenómenos físicos, pero no es común que se cumpla en las actividades humanas (a excepción de algunas situaciones muy particulares como los juegos de azar). Por este motivo los estadísticos se las han ingeniado para levantar estas exigencias desarrollando métodos que solo requieren que los eventos sean estables y su interdependencia sea acotada en el tiempo. A estos procesos estables se los denomina estacionarios.³ En los procesos estacionarios la influencia de un evento sobre sus vecinos disminuye rápidamente con el tiempo.⁴ Así, por ejemplo, el efecto de cualquier cambio puntual de los precios se supone que se tendría que atenuar rápidamente con el correr de los días. Esta es una exigencia que intuitivamente resulta muy razonable y todos los pronósticos estadísticos la consideran válida⁵. Sin embargo Mandelbrot encontró que sorprendentemente los datos que analizaba no cumplían con ese



supuesto:

“...el aspecto visual de las gráficas comprimidas a la escala logarítmica de la evolución de los precios del día era asombrosamente parecido al de las gráficas de los precios del mes, de los precios del año, del decenio y del siglo”⁶

A esta extraña rareza Mandelbrot la denominó **auto- semejanza** y es una propiedad característica de los fractales. Significa que la serie de datos se parece en cierto sentido a las másmushkas (o muñecas rusas): su forma se compone de partes de distinta duración que son muy similares entre sí y que

³ Nos referimos a la estacionariedad débil que exige que la media sea constante y que la correlación dependa solo de la diferencia de tiempo.

⁴ Una propiedad de los procesos estacionarios es que la autocorrelación es máxima en 0.

⁵ Cuando los datos presentan tendencias o ciclos son sometidos a procesos de desestacionalización de forma que se transformen en estacionarios.

⁶ Ver <http://revistas.uned.es/index.php/empiria/article/download/707/636> pag. 73.



parecen estar metidas unas dentro de otras.⁷ Pero, si la forma en que evolucionan los precios de un día es similar a la de un mes, o un año, o un decenio, entonces hay una extraña relación entre los eventos que no disminuye rápidamente a medida que pasa el tiempo, o sea no ajustan a las exigencias de los procesos estacionarios.

Se carece de métodos estadísticos para hacer estimaciones en procesos auto-semejantes. **¡En estos procesos ni siquiera se puede estar seguro de que el promedio de los datos históricos se acerque a una estimación razonable del valor futuro más probable!**⁸

Como si esto no fuera suficientemente sorprendente, Mandelbrot también detectó otros desajustes entre los supuestos habituales de la estadística tradicional y el comportamiento real de los precios del algodón. Así por ejemplo encontró que la serie de datos no presentaba continuidad a lo largo del tiempo sino que tenía “fluctuaciones muy bruscas”. Esto lo llevó a concebir la existencia de una forma de azar “que con toda legitimidad se podría calificar de **“salvaje”**”. Este azar salvaje es radicalmente distinto del **azar benigno** que supone la estadística tradicional. En presencia del **azar salvaje** no son aplicables las hipótesis de estabilidad y regularidad que son el fundamento de la mayoría de los métodos estadísticos clásicos.⁹

Por ejemplo una de las características del azar salvaje es que ya no se puede asegurar que se cumpla el Teorema Central del Límite, y por lo tanto tampoco se puede afirmar que la distribución de probabilidad de los acumulados tienda necesariamente a una Normal. Mandelbrot encontró que en la serie de datos que estaba analizando estas acumulaciones producían una distribución de tipo Pareto (a la que denominó distribución de Lévy-Pareto). Investigaciones posteriores en otros campos, como los Sistemas Adaptativos Complejos¹⁰ y la topología de redes¹¹ han reforzado la idea de que los sistemas complejos producen distribuciones similares a la de Pareto.

Ahora se verá como inciden estos conceptos en el cálculo de los indicadores organizacionales.¹² La mayoría los KPI se basan en la acumulación de valores. Por ejemplo: total de ventas por artículo, total de ventas por cliente, total deudas, etc. Si se dan las condiciones de **azar benigno**, entonces se cumplen los supuestos que aseguran que la acumulación de eventos estables y regulares tiende a formar distribuciones Normales, y por lo tanto su gráfica tiene forma de una campana.¹³

⁷ Ver <https://www.scientificamerican.com/article/multifractals-explain-wall-street/>

⁸ Un proceso en el que la media temporal tiende a la media se dice que es **ergódico**. Si los eventos son independientes la **Ley de los Grandes Números** asegura el promedio histórico tiende al valor esperado. En los procesos estacionarios no tiene por qué haber independencia, pero como su media es constante se sabe que se cumple una condición necesaria para la ergodicidad, que generalmente se supone válida. Pero si un proceso no es estacionario no se cuenta con ningún elemento que permita suponer su ergodicidad.

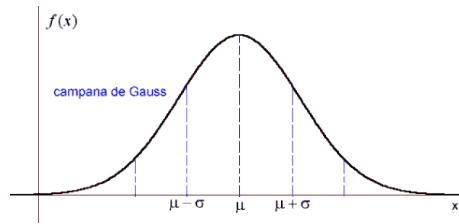
⁹ Ver Del Azar Benigno al Salvaje. B. Mandelbrot (1996)

¹⁰ Investigaciones del Instituto Santa Fe – lideradas por el premio Nobel Murray Gell-Mann – concluyeron que en los Sistemas Adaptativos Complejos “la totalidad del sistema podría tender a evolucionar hacia la condición en que se aplican dichas leyes de potencia [generalización de Pareto]” Ver Gell-Mann, M. (2003). El quark y el jaguar: Aventuras en lo simple y lo complejo. Barcelona: Tusquets. p. 337

¹¹ El estudio de la Ley Potencial (o Libre de Escala) en redes ha puesto al descubierto la incidencia de factores como “Conexión preferencial” “Coeficiente de Agrupamiento” y “Asortatividad”

¹² Las investigaciones de Mandelbrot ha sido usadas por Nasim Taleb quien afirma que le llevó quince años dar con ellos para poder reducir las incertidumbre de algunos Cisnes Negros financieros transformándolos en Grises.

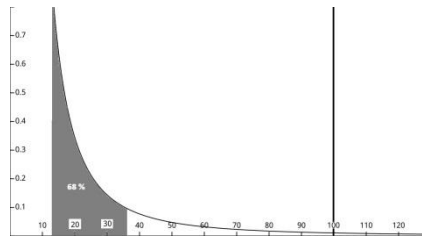
¹³ Esto es por el Teorema Central de Límite, que afirma que la suma de variables aleatorias independientes y de varianza no nula pero finita se aproxima a una distribución normal



Cuando los indicadores tienen forma de campana se dan situaciones muy convenientes para la gestión, por ejemplo el valor medio coincide con el valor más probable, y la gran mayoría de los valores se encuentran rodeándolo (Por ejemplo el 68% de los valores se encuentran en un intervalo de amplitud dos desviaciones estándar y centrado en la media).

Los KPI más sofisticados se suelen obtener mediante operaciones aritméticas sobre los indicadores fundamentales.¹⁴ Y en estos también se sabe que cuando impera el **azar benigno** sus distribuciones con forma de campana.¹⁵

El problema es que cuando no impera el **azar benigno** las distribuciones de acumulados y promedios adquieren otras formas. Como se ha dicho antes una de las formas más estudiadas hasta el momento son las de Pareto (que se conocen como funciones potenciales o libres de escala). Y estas distribuciones presentan características muy peculiares como que la varianza es infinita o muy grande y que el promedio no se encuentra cerca de los valores más probables.



En estas condiciones los promedios y acumulados no se pueden interpretar tan fácilmente como en el caso de azar benigno. A continuación se muestran ejemplos reales que muestran cómo los indicadores tradicionales no aportan toda la información que se necesita para analizar en profundidad situaciones y comportamientos complejos.

¿Cuánta información aporta y cuánta información oculta un KPI?

Si bien los KPI resultan muy efectivos para el control de objetivos no aportan información desagregada que permita encontrar situaciones ocultas bajo condiciones complejas ni detectar comportamientos incipientes antes de que afecten los resultados. A continuación se brindan tres ejemplos que muestran cómo el análisis de acumulados y promedios no revela toda la información que se necesitaría para detectar comportamientos incipientes o ineficiencias ocultas en situaciones complejas.

Ejemplo 1) Los promedios no aportan información válida para detectar cambios de comportamiento

Es común encontrar indicadores organizacionales que siguen una distribución de Pareto. Aquí se mostrará un ejemplo sobre cómo su promedio no siempre refleja toda la información relevante. Para

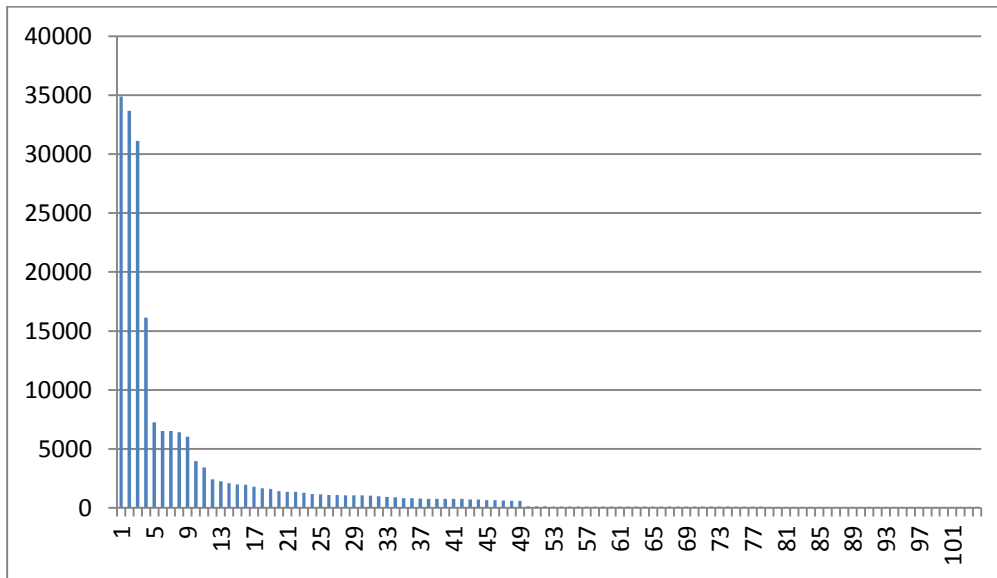
¹⁴ Por ejemplo el **Índice de fidelización se calcula como** (Nº de clientes x nº de compras totales) / 100.

¹⁵ El Teorema Fundamental de la Distribución Normal afirma que cualquier combinación lineal de variables aleatorias normales independientes es una variable aleatoria normal. Se sabe que el cociente de normales sigue una distribución de Cauchy, que es una distribución campaniforme. Por otra parte para muchas variantes del producto de variables Normales se conocen distribuciones con forma de campana https://en.wikipedia.org/wiki/Product_distribution



este ejemplo se utilizan datos del registro de quejas de clientes de entidades financieras de EEUU durante el 2019.¹⁶

En estos datos se tiene que el total de quejas por empresa sigue una distribución que sigue la regla de Pareto. La siguiente gráfica muestra la cantidad de quejas en el eje Y y las empresas en el eje X:



En estos datos apenas el 1% de las empresas (31 sobre un total de 2902) totalizan el 80% de las quejas (185.899 sobre un total de 229.989). El promedio de quejas por empresa de 79,25. Además a estas 31 empresas le siguen otras 100 (las primeras de la larga cola), cuyo acumulado es de 21.504 quejas, lo que representa casi un 10% del total (exactamente un 9,35%).

Supongamos que esas 100 empresas presentaran un aumento de sus quejas de un 25% cada una. Esto sin duda sería un cambio de comportamiento relevante que podría estar anunciando un problema incipiente que debería recibir atención. Pero si el control de quejas se realiza a partir de alertas sobre cambios en el promedio esta situación pasaría desapercibida. La razón es que el hipotético aumento de quejas que estamos suponiendo apenas subiría el promedio de 79,25 a 81,10, o sea menos de 2 unidades, lo que representa un insignificante 2.3%.

Es natural que una variación de apenas el 2.3 % en un indicador de este tipo no llame la atención. Entre otras cosas porque se suele considerar natural que variaciones de este tipo ocurran por el “ruido” habitual del azar benigno. Pero en este caso la variación no se origina en el ruido, sino en un aumento del 25 sobre 100 empresas.

En suma la información que aporta el promedio de por sí no alcanzaría para detectar algunos cambios potencialmente relevantes.

Ejemplo 2) Los valores globales pueden ocultar información sobre comportamientos desagregados

Hay otras situaciones en que los acumulados esconden información. Una de ellas es la que se conoce como **Paradoja de Simpson**. Esta paradoja muestra que un conjunto de datos sumados puede presentar una característica totalmente contraria a la que presentan los mismos datos desagregados. Esta paradoja es muy conocida por un caso muy sonado que involucró a la universidad de Berkeley. Los datos de ingresos de estudiantes acumulados mostraban que esa

¹⁶ Datos publicados por el **Consumer Financial Protection Bureau** sobre 238.433 quejas registradas desde el 01/01/2019 hasta el 19/11/2019



universidad en el año 1973 había aceptado un porcentaje de postulantes masculinos superior al femenino, en cambio cuando se analizaban los datos por departamento se veía que ocurría lo contrario.¹⁷

Esta misma paradoja aparece frecuentemente en la gestión, produciendo contradicciones entre lo que se desprende de los valores globales acumulados y lo que se desprende de los datos desagregados.

Por ejemplo, usando los mismos datos sobre quejas financieras del ejemplo anterior se puede ver que la variación de tendencia del total de quejas del año 2019 es de 2.1%¹⁸ Esta variación parece indicar cierto nivel de estabilidad en la cantidad de quejas. Sin embargo al abrir esos cálculos por empresa se tiene que algunas presentan una variación de tendencia muy positiva y otras muy negativas.

Empresa	Quejas	%	Variación Tendencia
EQUIFAX, INC.	34879	14.6%	+53.10%
TRANSUNION INTERMEDIATE HOLDINGS, INC.	33663	14.1%	+67.77%
EXPERIAN INFORMATION SOLUTIONS, INC.	31107	13%	+45.69%
JPMORGAN CHASE & CO.	7260	3%	-23.33%
WELLS FARGO & COMPANY	6526	2.7%	-40.20%
CAPITAL ONE FINANCIAL CORPORATION	6517	2.7%	+15.36%

Si se quiere conocer lo que ocurre realmente (para tomar medidas correctivas o para aprovechar oportunidades) es imprescindible desagregar la información por las diferentes variables o campos. Incluso no alcanza por hacerlo por una sola variable. Por ejemplo si se está analizando ventas se debería abrir los totales por sucursal, vendedor, forma de entrega, rango de precios, etc., y también por cada combinación de dos y más variables (Ej. Sucursal-vendedor, sucursal - forma de entrega, etc.). Esta forma de proceder se relaciona con las situaciones relevantes que implican condiciones múltiples como la que se muestra en el siguiente ejemplo.

Ejemplo 3) Oportunidades o Ineficiencias que se presentan bajo situaciones que involucran condiciones múltiples

Otro caso de pérdida de información en los indicadores globales son las situaciones o los cambios de comportamiento que se presentan bajo múltiples condiciones. Este tipo de información es muy difícil de encontrar pues implica evaluar muchas combinaciones de variables, pero en contrapartida deben tenerse en cuenta porque pueden suelen poner al descubierto oportunidades de mejora o pérdidas de eficiencia.

Por ejemplo tomando los mismos datos de quejas financieras que en los ejemplos anteriores se puede encontrar esta situación muy particular:

¹⁷ https://es.wikipedia.org/wiki/Paradoja_de_Simpson

¹⁸ Esto surge de que al principio del año se tenía una tendencia de 552 quejas por día y al final una tendencia de 564.



Las quejas “Cerradas con explicación” debidas a problemas en el “Reporte de Crédito” con motivo “El reporte pertenece a otra persona” de las empresas EQUIFAX, INC. TRANSUNION INTERMEDIATE HOLDINGS, INC. y EXPERIAN INFORMATION SOLUTIONS, INC. representan el 80% de las quejas del Estado de Florida.

Como se ve esta situación particular involucra 5 variables (respuesta dada al cliente, tipo de servicio, tipo de problema, empresa y Estado), y es relevante porque estaría indicando que las tres empresas citadas tienen un serio problema en la entrega de correspondencia en el Estado de Florida. Esta situación es muy difícil de encontrar porque los datos tienen 18 campos, y esto implica que para hallar esta combinación sería necesario evaluar 262.143 combinaciones entre esos campos.¹⁹

En síntesis

Para encontrar todos los problemas y oportunidades que encierran los sistemas complejos es preciso analizar en detalle los datos desagregando sus valores para cada valor, y para cada combinación de valores. Para lograrlo Complexus Analytics implementa una analítica compleja que es capaz de identificar una gran cantidad de micro-situaciones (o micro-segmentos). Los fundamentos básicos de esa analítica compleja se describen en el siguiente punto.

Analítica de Sistemas Complejos

Los datos provenientes de sistemas complejos presentan algunas características que dificultan su análisis. Por ejemplo contienen “discontinuidades, saltos y rupturas”, así como “eventos únicos y discretos, que no siguen un patrón universal claro que pueda decodificarse matemáticamente de manera inmediata”. Además se producen bajo la influencia de un contexto no determinista debido a la existencia de “incertidumbres” y “subjetividad”, que a veces no llega a contener una “racionalidad coherente y estricta”.²⁰

A continuación se mostrará que para afrontar estas características la primera condición que debe cumplir una herramienta analítica es que debe ser capaz de detectar, evaluar y mostrar en forma ágil y amigable una gran cantidad de irregularidades específicas que a su vez pueden involucrar varias variables cada una.

La afirmación anterior se basa en que un proceso analítico consiste - desde el punto de vista de la teoría de la información - en una transformación de un conjunto extenso de símbolos (datos originales) difíciles de interpretar, en otro conjunto más reducido de símbolos (regularidades) que son más fáciles de entender y manipular. Obviamente este proceso de transformación debe asegurar que la síntesis que se construye no pierde información relevante

Por otra parte, la teoría de la complejidad aporta el concepto de **complejidad efectiva** que resulta útil para dimensionar las características que debe cumplir esa transformación. Gell-Man & Lloyd definen la **complejidad efectiva** de un conjunto de datos como “la longitud de la descripción más comprimida de sus regularidades”. De acá se desprende que sería una contradicción suponer que un sistema tiene una complejidad efectiva alta, y que a la vez es posible describirlo adecuadamente mediante una síntesis breve. Si se encontrara una síntesis breve de regularidades para describir

¹⁹ Nota este cálculo surge que la cantidad de combinaciones surge de sumar las posibles combinaciones de 1, 2, 3,... n variables, siendo n la cantidad total de variables. Esa suma da como resultado $2^n - 1$.

En este caso es 2 elevado a la 18, -1. Evaluar esta cantidad de combinaciones es imposible mediante métodos manuales, porque implicaría que el analista ejecutara al menos una consulta para cada combinación. Tampoco es razonable utilizar técnicas de “fuerza bruta”. Por eso es imprescindible utilizar técnicas de IA que se describen más adelante.

²⁰ A COMPLEXITY APPROACH FOR PUBLIC POLICIES, IPEA, 2015, p 28



adecuadamente un sistema,²¹ entonces por definición ese sistema no poseería una complejidad efectiva alta. O lo que es lo mismo (por el contra-recíproco) si el sistema tuviera una alta complejidad efectiva, entonces no sería posible encontrar un conjunto breve de regularidades que lo describan adecuadamente.

En términos generales: la síntesis de regularidades que puede encontrar un proceso analítico será necesariamente tanto más extensa cuanto más complejo sea el sistema de donde surgieron los datos originales. Obsérvese que esa síntesis podrá ser extensa porque contiene muchas regularidades, y/o porque contiene regularidades que involucran muchas variables.

En base a lo anterior se tiene que la efectividad de una herramienta analítica para encontrar todas las propiedades y comportamientos relevantes de un sistema muy complejo no solo dependerá de su capacidad detectar y manejar regularidades, sino que también debe ser capaz de detectar las irregularidades que producen la complejidad.

Complexus Analytics parte de la base de que el sistema analizado es efectivamente complejo y que por lo tanto necesariamente deberá detectar, evaluar y permitir manipular de manera ágil y amigable irregularidades que pueden involucrar muchas variables.

Una vez planteado este desafío, el problema metodológico es cómo explorar una gran cantidad de irregularidades para evaluarlas y así determinar cuáles aportan los impactos prácticos más relevantes. La metodología exploratoria que utiliza Complexus Analytics se explica en el siguiente punto.

Análisis Exploratorio de Datos

Desde el punto de vista metodológico el análisis de irregularidades que realiza Complexus Analytics se basa en el **Análisis Exploratorio de Datos** (EDA por sus siglas en inglés). Este es un enfoque radicalmente diferente al enfoque estadístico clásico que se emplea comúnmente.

El EDA es una metodología creada especialmente para descubrir nuevos conocimientos. Fue concebida por el estadístico Jhon W. Tukey en la década de los 70. El EDA se puede describir como un “trabajo de detectives - detective numérico o detective de cuentas o detective de gráficos”.²² Su propósito es ayudar a descubrir pistas o indicios que permitan formular nuevas hipótesis que a su vez permitan crear nuevos conocimientos. De acuerdo a Tukey el descubrimiento de nuevos conocimientos no se produce con la mera aplicación de técnicas o herramientas sino que también requiere la comprensión e interpretación del analista: “Un detective que investiga un crimen necesita herramientas y comprensión. Si no tiene polvo de huellas dactilares, no podrá encontrar huellas dactilares en la mayoría de las superficies. Si no entiende dónde es probable que el criminal haya puesto sus dedos, no buscará en los lugares correctos. Igualmente, el analista de datos necesita herramientas y comprensión (p 1: Tukey (1977))

Hasta el momento el Análisis Exploratorio de Datos se ha utilizado principalmente en la ingeniería y especialmente para actividades de I&D. Así por ejemplo el EDA forma parte del repertorio

²¹ Acá el término “adecuado” se refiere al nivel de adecuación entre la descripción y los fines perseguidos. Claramente se trata de un criterio subjetivo. De hecho el concepto mismo “complejidad efectiva” en subjetivo como lo exponen Gell-Man & Lloyd: “... la CE de una entidad depende del contexto, y en gran medida es subjetiva. Depende de la granulación (nivel de detalle) con que se describa la entidad, del lenguaje utilizado para describirla, de los conocimientos y la comprensión previos que se asuman y, por supuesto, de la naturaleza de la distinción que se haga entre regularidad y aleatoriedad.” <https://www.santafe.edu/research/results/working-papers/effective-complexity>

²² Tukey es ampliamente reconocido por ser el creador del algoritmo FFT (Fast Fourier Transformation)



metodológico compilado por el *Engineering Statistics Handbook* del *National Institute of Standards and Technology* (NIST) de EE.UU.²³

El hecho de que todavía no sea común que el EDA se aplique en otras áreas de las organizaciones se debe a que hasta el momento su uso exige que los analistas tengan mucha práctica en el uso de métodos y gráficos matemáticos bastante complejos. Pero en la actualidad la Inteligencia Artificial permite automatizar estos procesos de exploración con lo que se posibilita extender su aplicación a todas las áreas de la empresa.

A continuación se describen las características fundamentales del EDA y los criterios que utiliza Complexus Analytics para implementarlos.

Características fundamentales del Análisis Exploratorio de Datos

De acuerdo al *Engineering Statistics Handbook* el EDA se caracteriza por *propiciar el descubrimiento de secretos estructurales insospechados, que incluso pueden llegar a cambiar radicalmente la forma en que se comprende el objeto estudiado.*²⁴ Para lograr estos descubrimientos esta metodología **no se limita a una mera colección de técnicas**, sino que se trata de una nueva **filosofía de análisis** que cambia: la forma clásica en que se diseccionan y se miran los datos, lo que se busca en ellos, y el modo en que se interpretan.²⁵

Técnicas utilizadas por el EDA

El *Engineering Statistics Handbook* expone que la mayoría de las técnicas utilizadas por el EDA son de naturaleza gráfica. La razón para que esta metodología haga un uso intensivo de los gráficos estadísticos es que estos tienen la virtud de incentivar la capacidad humana de reconocer patrones a la vez que no imponen condicionamientos para su aplicación porque muestran todo el universo de valores.²⁶

Pero la interpretación de gráficos estadísticos tiene el inconveniente de que requiere un gran entrenamiento previo. Por este motivo Complexus Analytics reemplaza el análisis de gráficos por algoritmos de Inteligencia Artificial que son capaces de explorar todos los valores de las variables y de sus combinaciones para identificar rarezas. Además Complexus Analytics presenta esas rarezas de manera simple, clara y ordenada a través de informes y cuadros comparativos. De esta forma el analista evita tener que explorar “manualmente” grandes volúmenes de datos representados en gráficos complejos, y puede dedicar todo su tiempo a interpretar el impacto y la significancia de las rarezas que fueron encontradas y presentadas por el sistema.

Subjetividad

Una de las características más relevantes del EDA es el papel que se le otorga a la subjetividad.

Las técnicas de análisis clásicas “sirven para elaborar los fundamentos probabilísticos de la ciencia y la ingeniería”. Por este motivo deben asegurar que sus conclusiones son absolutamente firmes, lo que a su vez exige que sean totalmente “rigurosas, formales y objetivas”.

En cambio las técnicas exploratorias del EDA procuran incentivar el descubrimiento de nuevas hipótesis y por ello su prioridad es ser “sugestivas, indicativas y perspicaces”. En estas condiciones la interpretación subjetiva del analista juega un rol fundamental, y por ello los resultados “pueden diferir de un analista a otro, aunque los analistas experimentados suelen llegar a conclusiones idénticas.”²⁷

²³ <https://www.itl.nist.gov/div898/handbook/eda/eda.htm>

²⁴ <https://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm>

²⁵ <https://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm>

²⁶ <https://www.itl.nist.gov/div898/handbook/eda/section1/eda11.htm>

²⁷ <https://www.itl.nist.gov/div898/handbook/eda/section1/eda124.htm>



Complexus Analytics implementa el carácter subjetivo del EDA proporcionando mecanismos ágiles para la interacción entre el analista y los algoritmos. Esta interacción permite que el analista aporte su capacidad interpretativa basada en su experiencia y conocimientos personales (subjetivos y contextualizados) mientras que los algoritmos aportan su potencia de cálculo sistemático y objetivo.

Generalidad

Otra característica esencial del EDA es su generalidad derivada de la ausencia de exigencias de supuestos estadísticos

La validez de las técnicas estadísticas clásicas depende de que se verifiquen supuestos como por ejemplo: distribuciones normales, independencia, varianza finita, etc. El problema es que en la práctica ocurre que a veces “los analistas desconocen los supuestos subyacentes exactos, o en el caso de que los conozcan no siempre los comprueban”. Además como se ha visto antes las distribuciones de Pareto (de aparición muy frecuente) no cumplen estos requisitos.

En cambio las técnicas del EDA “hacen pocas o ninguna suposición” por lo que su uso resulta válido para la mayoría (o la totalidad) de las situaciones.

Como se verá más adelante Complexus Analytics cumple con esta característica pues no impone ninguna hipótesis sobre las características estadísticas de los datos que analiza.

Utilización de toda la información disponible

El EDA no se basa exclusivamente en el cálculo de indicadores de resumen.

Las técnicas clásicas de estimación suelen “cartografiar” los datos mediante indicadores como la media, dispersión, entropía, etc. De esta forma el analista trabaja sobre indicadores que reflejan ciertas características que se consideran fundamentales. Esta forma de proceder presenta la ventaja de que ordena y acelera el análisis, pero también acarrea una importante desventaja pues al “concentrarse en estas pocas características puede filtrar otras (como por ejemplo asimetría, long tail, autocorrelación, etc.) lo que provoca pérdida de información”

En contraposición “el enfoque EDA utiliza (y muestra) todos los datos disponibles. En este sentido no hay pérdida de información”. O sea el EDA no apela a ningún resumen basado en resúmenes contruidos sobre índices estadísticos, sino que utiliza toda la información tan desagregada como sea posible.

Complexus Analytics analiza todos los valores de variables de enumerados y todas las categorías de los datos numéricos continuos. Uno a uno, para asegurar que puede encontrar aquellos que producen desvíos notorios. Esto lo hace tanto para los valores individuales de cada variable como para todas sus combinaciones. La cantidad enorme de cálculos que esto implica exige el apoyo de Inteligencia Artificial. A continuación se describe el enfoque y los algoritmos de IA que utiliza Complexus Analytics

Inteligencia Artificial Interpretable (XAI)

La arquitectura de cálculo que permite cumplir con las características y desafíos antes expresados es la que se conoce como **Inteligencia Artificial Interpretable (Explainable Artificial Intelligence o XAI)**.

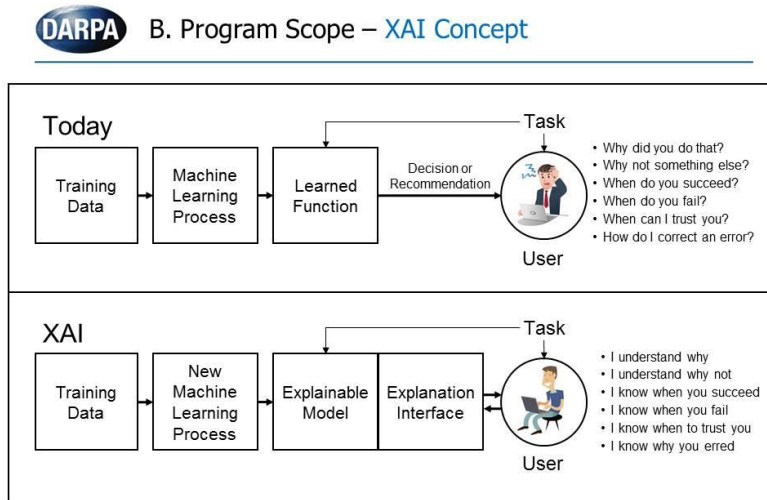
La XAI refiere a una forma de abordar la Inteligencia Artificial que asegura que los resultados obtenidos sean entendibles y comprensibles para los seres humanos. Este marco contrasta con la idea de que la IA se puede comportar como una caja negra que no necesita proporcionar explicaciones sobre los resultados que produce.

Según la **Agencia de Proyectos de Investigación Avanzada para la Defensa (DARPA)** el objetivo de la XAI es que el usuario final tenga un manejo real del sistema que administra, para lo que es



indispensable que los programas de IA hagan explícita la racionalidad de sus decisiones, sugerencias y recomendaciones.

La figura siguiente sintetiza las diferencias básicas entre la IA habitual y la XAI:²⁸



Distribution Statement "A" (Approved for Public Release, Distribution Unlimited)

5

Por otra parte de acuerdo a los documentos de DARPA una solución XAI necesita tres niveles de descripción para quedar bien determinada:

- Problema desafío
- Modelo de IA
- Interface explicativa

A continuación se describe la especificación de cada uno de estos bloques para Complexus Analytics.

Problema desafío

Sintetizando todo lo expuesto hasta el momento, el problema desafío de Complexus Analytics es: **proveer una herramienta para analizar datos provenientes de sistemas complejos que permita encontrar pistas e indicios que conduzcan a descubrir oportunidades e ineficiencias ocultas por situaciones complejas y comportamientos incipientes, incluso antes de que se reflejen en los indicadores clave de performance.**

Como se ha visto esto requiere tener en cuenta algunas capacidades imprescindibles:

- Poder descubrir y manejar irregularidades)
- Evaluar esas irregularidades para descubrir las que aportan conocimiento útil de acuerdo a su impacto. De acuerdo al EDA esto exige:
 - Integrar al proceso el conocimiento subjetivo que aportaran los expertos en el tema
 - Evitar el uso de supuestos estadísticos que arrojen sospechas sobre la validez de los resultados
 - Procurar que los cálculos se realicen sobre toda la información disponibles evitando el uso de indicadores estadísticos que de alguna forma filtran parte de la información que contienen los datos

²⁸ <https://www.darpa.mil/program/explainable-artificial-intelligence>



- Realizar todo el proceso con una demanda de tiempo que resulte aceptable para la gestión de organizaciones

Modelo de IA

El modelo de IA de Complexus Analytics se compone de diferentes procesos (o algoritmos) que se describen a continuación.

Perfilamiento genético

Para asegurar que se tiene en cuenta la totalidad de las irregularidades), Complexus Analytics comienza por construir la máxima desagregación posible de agrupamientos. Para esto utiliza un concepto introducido por John Holland – creador de los algoritmos genéticos- quien propuso tratar los diferentes campos que componen los datos como si fueran “genes”. Holland propone mirar los datos que se analizan como **bloques de construcción**, por ejemplo si fueran datos de una cara humana estos bloques serían pelo, frente, cejas, ojos, etc. Estos **bloques de construcción** se pueden pensar como si fueran genes que describen las “características genéticas” del objeto que representan:

Este enfoque, que trata los valores en las posiciones individuales como la bloques de construcción, se corresponde estrechamente con el enfoque clásico para evaluar la los efectos de los genes individuales en un cromosoma. Cada gen tiene varias formas alternativas, llamadas alelos. Los diferentes alelos del gen humano para el color de ojos, por ejemplo, producen ojos azules, ojos marrones, ojos verdes, y así sucesivamente.²⁹

Ejemplo:

Datos Originales sobre compras			
	Inciso	Moneda	Tipo Resolución
1	Universidad de la República	DOLAR PIZARRA VENDEDOR	Adjudicada totalmente
2	Universidad de la República	PESOS URUGUAYOS	Adjudicada totalmente
3	Universidad de la República	PESOS URUGUAYOS	Adjudicada totalmente
4	Ministerio de Salud Pública	PESOS URUGUAYOS	Adjudicada parcialmente
5	Ministerio de Salud Pública	PESOS URUGUAYOS	Adjudicada totalmente
6	Ministerio de Salud Pública	PESOS URUGUAYOS	Adjudicada totalmente
7	Ministerio de Salud Pública	PESOS URUGUAYOS	Adjudicada totalmente

Perfilamiento Genético Resultante				
	Inciso	Moneda	Tipo Resolución	Cantidad
1	Universidad de la República	DOLAR PIZARRA VENDEDOR	Adjudicada totalmente	1
2 y 3	Universidad de la República	PESOS URUGUAYOS	Adjudicada totalmente	2
4	Ministerio de Salud Pública	PESOS URUGUAYOS	Adjudicada parcialmente	1
5, 6 y 7	Ministerio de Salud Pública	PESOS URUGUAYOS	Adjudicada totalmente	3

Esta forma de agrupar los datos produce la máxima cantidad de agrupaciones que se puede obtener a partir de ellos. El mayor problema práctico de este proceso es evitar que el usuario se vea obligado a precargar un diccionario con los valores posibles y rangos de cada tipo campo. Esto implica utilizar heurísticas que determinen ese diccionario automáticamente.

²⁹ Holland, J. *Hidden Order: How Adaptation Builds Complexity*, 1995, Helix Books, pág. 62.



Los valores de los campos con tipo de datos **enumerado** (por ejemplo Turno, Origen, País, Sucursal, etc.) se detectan en base a los diferentes valores que toma. El sistema los detecta automáticamente en base a la proporción de repeticiones en relación a la cantidad total de registros. Hay un tipo especial de enumerados que presentan una dificultad adicional, se trata de los que presentan un conjunto de valores muy grande. Por ejemplo es el caso de códigos de identificación como Cliente, Empresa, Artículo, etc. En estos casos suelen aparecer varios valores del enumerado con frecuencias de aparición insignificantes. Estos valores en lugar de aportar información útil generan una nube de perfiles con poco sentido que afecta el resto de los cálculos y sobretodo la interpretación final del usuario.

Para evitar esto el sistema utiliza una heurística que procura crear categorías independientes solo para los ítems que tienen frecuencias significativas, a la vez que agrupa el resto quintiles o porciones de “long tail” de acuerdo a la distribución de frecuencia que se observa que caracteriza a la población analizada.

Por su parte los campos numéricos admiten diferentes tipos de categorizaciones, por ejemplo un campo numéricos se puede categorizar como una lista de valores (1,2,3), rangos de valores de igual tamaño (0 a 100, 101 a 200, ... 401 a 500) o quintiles (rangos diferentes pero con igual cantidad de ítems cada uno.) Complexus Analytics utiliza una heurística para asignar la categorización más conveniente en base a criterios empíricos.

Una vez que el sistema ha construido las categorizaciones automáticas el usuario puede revisarlas por ejemplo para cambiar los rangos de un campo numérico.

Este proceso de **perfilamiento genético** asegura que se obtiene el máximo conjunto posible de micro-patrones, pues cada “cromosoma” contendrá todos los registros que poseen iguales categorías. En contrapartida es poco probable (aunque no imposible) que esta división en micro-patrones arroje agrupamientos de impacto relevante, por eso es necesario realizar procesos de re-agrupamiento. El algoritmo que se describe a continuación detecta el conjunto mínimo de reglas que permiten realizar un primer nivel de agrupamiento de estos micro-patrones.

Árbol de decisión para agrupar los micro-patrones

El uso habitual de los algoritmos del tipo **árbol de decisión** es el de automatizar la aplicación de reglas de decisión. Estos algoritmos extraen el conjunto mínimo de reglas de decisión que se encuentra implícito en un conjunto de datos. O sea detecta las reglas que dieron lugar a las decisiones que se han tomado hasta el momento y que están registradas en los datos.

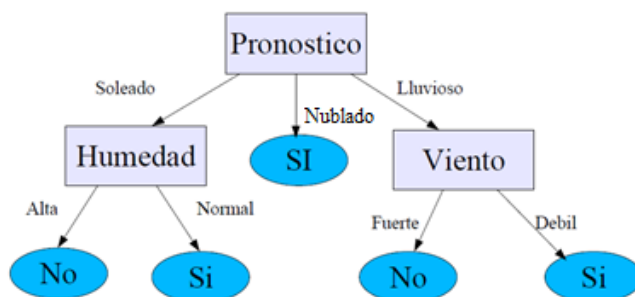
Si el conjunto de datos es suficientemente amplio, estas reglas que fueron automáticamente extraídas permitirán determinar cuál es la decisión que le correspondería a cualquier combinación de datos. En este sentido los árboles de decisión se consideran algoritmos del tipo de aprendizaje supervisado.

A continuación se muestra un ejemplo sobre cómo opera este algoritmo. Supongamos que tenemos datos que muestran la forma en que se ha decidido si las condiciones eran adecuadas para practicar deportes al aire libre.



Pronóstico	Temperatura	Humedad	Viento	¿Adecuado?
Soleado	Alta	Alta	Flojo	No
Soleado	Alta	Alta	Fuerte	No
Nublado	Alta	Alta	Flojo	Si
Lluvia	Moderada	Alta	Flojo	Si
Lluvia	Baja	Normal	Flojo	Si
Lluvia	Baja	Normal	Fuerte	No
Nublado	Baja	Normal	Fuerte	Si
Soleado	Moderada	Alta	Flojo	No
Soleado	Baja	Normal	Flojo	Si
Lluvia	Moderada	Normal	Flojo	Si
Soleado	Moderada	Normal	Fuerte	Si
Nublado	Moderada	Alta	Fuerte	Si
Nublado	Alta	Normal	Flojo	Si
Soleado	Moderada	Alta	Fuerte	No

Un proceso de árbol de decisión toma estos datos y construye el siguiente árbol:



Una vez obtenido el árbol es fácil traducirlo en reglas. Por ejemplo si está soleado y la humedad es normal se decide jugar. Los algoritmos más comunes para construir estos árboles son el ID3 y el C4.5, y cuando la cantidad de ítems a clasificar es muy grande se complementa con estrategias como “Random Forest”.

Complexus Anlytics utiliza estos algoritmos para un fin totalmente diferente, lo usa para **identificar la agrupación mínima de perfiles genéticos de acuerdo a las reglas que clasifican el logro de un objetivo**.

Por ejemplo si se analiza el “árbol de decisión” de datos de **compras** tomando como objetivo la **demora del proceso**, entonces las ramas del árbol serán las combinaciones de características que arrojaron como resultado compras de duración breve, mediana, larga, etc. Y en ese caso cada rama del árbol será una regla de clasificación que permite agrupar las compras con ciertas características comunes y que producen un determinado valor del objetivo (compra breves, larga, etc.).

El uso de los algoritmos de árboles de decisión aseguran que obtienen mínimo árbol posible, por lo tanto se tendrá que los agrupamientos se construyen a partir de las reglas tendrán las mínimas exigencias posibles. Por eso este proceso transforma los micro-patrones del perfilamiento genético en mini-patrones agrupados. Pero si bien estos los mini-patrones suelen arrojar algunos resultados con impacto relevante, igual se hace necesario explorar la existencia de otros agrupamientos más amplios que puedan arrojar resultados relevantes. Ese es el propósito de la siguiente fase que consiste en un proceso de **Clustering**.

Clustering de mini-patrones

Los algoritmos de clustering o clusterización son capaces de organizar los datos en conjuntos o agrupamientos. Estos agrupamientos se construyen para asegurar que los datos que forman un conjunto sean más similares entre sí de lo que son comparados con los datos de los otros conjuntos.



COMPLEXUS ANALYTICS

Generalmente, el agrupamiento de datos se considera como un algoritmo de aprendizaje "no supervisado", ya que todas las variables son de entrada y no existe una variable de salida como punto de referencia para supervisar el aprendizaje.

En general estos algoritmos tratan de minimizar la cantidad de grupos. O sea parten de la idea de que es posible encontrar un conjunto reducido de patrones que arrojen información relevante. Como se ha dicho antes este supuesto reduce artificialmente la complejidad efectiva. Pero Complexus Analytics lo utiliza solo como complemento para agregar posible marco-hallazgos a los que aportan los micro y mini patrones de regularidad.

Complexus Analytics utiliza un algoritmo de clustering de tipo "hasing", que va filtrando los grupos en función de la significación de su frecuencia.

Cabe acotar que se utiliza una información que surge secundariamente del proceso anterior. Los árboles de decisión permiten calcular la frecuencia con que cada variable aparece en las reglas de clasificación y esto se puede tomar como un indicador de que tan significativa es cada variable en la clasificación de los datos. En el ejemplo del punto anterior la variable Pronóstico se usa en el 100% de los casos, Humedad solo en 4 de 14, Viento en 5 de 14, y la Temperatura en ninguno. Este orden le permite a Complexus Analytics direccionar en forma eficiente la búsqueda de agrupamientos significativos en el proceso de clustering.

Interface explicativa

Esta es la interface que permite que el usuario aporte su conocimiento y capacidad de interpretación para realiza la selección de los indicios con impacto más relevante. Para esto combina dos grandes funcionalidades que interactúan entres si:

- 1) Visualización de los desvío de acuerdo a su relevancia
- 2) Refinamiento interactivo de las condiciones de cada agrupamiento

Obsérvese que para implementar estas prestaciones es necesario contar con una forma de comparar la incidencia que tiene cada agrupación. A continuación se explica cuál es el fundamento del criterio de ordenamiento que utiliza Complexus Analytics.

Como se ha dicho antes el sistema se basa en el Análisis Exploratorio de Datos que es "trabajo de detectives", y una premisa clásica de la investigación detectivesca es que **los indicios más relevantes son los que menos encajan con el contexto**. *Tratándose de datos objetivos que se pueden contar estos se traduce en que la relevancia de una pista viene dada por el nivel de desproporción que presente.*

La primera magnitud que permite empezar a medir la relevancia de un grupo de datos está dada por su cantidad de elementos. Por ejemplo es natural que se considere más relevante una agrupación que incluye al 25% de los trámites analizados que una que solo incluye al 3%. Lo mismo ocurre si el análisis se realiza en términos de un valor acumulable, por ejemplo es más relevante una agrupación que implica el 30% del importe vendido que otra que implique solo un 10%. Y como consecuencia natural se desprende un posible orden por la tasa (acumulado/cantidad). Por este motivo Complexus Analytics realiza los cálculos y permite obtener informes en base a estos tres indicadores básicos: cantidad, acumulado y tasa (acumulado/cantidad).

Pero este acercamiento no alcanza para comparar completamente el impacto de diferentes agrupaciones de datos. Por ejemplo si en un departamento trabajan solo el 3% de los empleados pero la proporción de sexos es de 7 a 1 hay una rareza que puede valer la pena investigar.



Para resolver este problema de comparación se tomó el siguiente postulado: ***en general se espera que las magnitudes presenten comportamientos linealmente proporcionales y una medida de su desproporción está dada por el alejamiento de ese comportamiento ideal.***

Por ejemplo si la mayoría de las “compras directas” demoran menos de 7 días, entonces las “compras directas” que demoren más de 7 días serán las que merezcan atención. Este principio resulta también válido para combinaciones de campos. Por ejemplo

- si los reclamos de los vecinos de la **Zona A** son el 10% del total de reclamos
- y además los reclamos por **Contenedor Roto** son el 5% del total de reclamos,
- entonces se suele esperar que los reclamos por **Contenedores Rotos de la Zona A** sean un 0,5% del total. (o sea la multiplicación 10% por 5%)
- Y por último lo llamativo sería que los reclamos por **Contenedores Rotos de la Zona A** fueran muy superiores o muy inferiores a 0,5%

Cabe acotar que este postulado está inspirado en afirmaciones de Jay W. Forrester - ingeniero del MIT que entre sus méritos cuenta con ser el fundador de dinámica de sistemas, inventor de la memoria RAM en 1950, y haber creado la primera imagen animada sintética – quien expresó que “La mente humana aún no está adaptada para interpretar el comportamiento de los sistemas sociales. Los sistemas sociales pertenecen a la clase llamada sistemas de retroalimentación no lineal de multi-lazo”³⁰

En base a esta forma de medir las desproporciones Complexus Analytics brinda un camino para evaluar el impacto de cada agrupación. Además mediante funciones de filtrado y ordenamiento permite que el usuario refine y reagrupe los perfiles hasta encontrar los que tienen más sentido para generar acciones concretas de mejora.

³⁰ Forrester, J. (2000). Comportamiento contraintuitivo de los sistemas sociales. Recuperado de <http://master.edgarmorin.mx/file.php/14/Comportamientocontraintuitivodelossi.pdf>