

Transcripción Coloquio de Transformación Digital:

Ciencia de datos en el sector público

23.11.2020

-José Inostroza: Muy buenos días. Estamos muy contentos de estar reunidos en este nuevo coloquio. Estamos transmitiendo por nuestro canal de [YouTube](#). Pareciera que con datos se pueden hacer cosas muy bien. Vamos a aprender en esta conversación que para eso hay que tener metodología, estrategia. No todo es tan trivial como a veces parece. Y tenemos dos expositores de lujo, dos ingenieros. En primer lugar Sebastián Ríos que es académico de la Facultad de Ciencias Físicas y Matemáticas, un experto en ciencia de datos. Sebastián es doctor en Ingeniería en Información de la Universidad de [Tokio](#) y también es Ingeniero Industrial con una especialización en Tecnologías de la información y minería de datos de la Universidad de Chile. Tiene una larguísima trayectoria en materia de ciencia de datos y ha sido muy famoso. Y obviamente, nos va a comentar sobre los aportes que ha hecho con algunas aplicaciones muy específicas en el ámbito de la salud, lo cual nos llena de orgullo a todos. Además Sebastián, como ustedes pueden ver elocuentemente es músico, rockero, famoso por su actividad rockera en la Universidad, así que estamos muy contentos de que estés presente.

(02:49) Además de Sebastián, está Charango Munizaga, que también es músico, justamente especialista en el charango. Pero además, él tiene otra gran virtud: venir del futuro. Está en Australia, donde está terminando el día. Charango, además, es Ingeniero Matemático de la Universidad de Chile y tiene un doctorado en Ingeniería en Recursos Naturales de la [Laurentian](#) University de Canadá, y además tiene un Máster en Gestión de Operaciones de la Universidad de Chile. Y actualmente en Australia ejerce funciones de docencia y también consultoría, especialmente en empresas mineras. Y además, en este último período ha trabajado con distintas instituciones públicas de Chile. Probablemente lo más destacado en ese sentido es que Charango es la persona que estuvo detrás del famosísimo caso de la SUSESO, de aplicación de análisis de datos, inteligencia artificial, en el Proyecto de Modernización, del cual hemos hablado bastante en este coloquio. Charango es el hombre que metió las manos e hizo cosas bastante notables.

Sebastián va a disponer de 30 minutos, luego expondrá Charango y abriremos la conversación en esta comunidad.

-Sebastián Ríos: (05:15) Muy buenos días a todos. Gracias por la invitación.

CIENCIA DE DATOS EN EL SECTOR PÚBLICO: MÁS ALLÁ DE LA MODA. ¿COMO CREAR VALOR PÚBLICO EN LA PRÁCTICA?

Sebastián Ríos - srios@dii.uchile.cl

Lo que me interesa de esta lámina para que les queden mis datos, es el equipo. Esos chiquillos que están ahí son parte del Centro de Investigaciones de Inteligencia de Negocios, [CEINE](http://ceine.cl), que yo dirijo, y que son los que están detrás de mí en todos los proyectos que hago.

PRESENTACIÓN

➤ Sebastián A. Ríos

- Prof. Asociado, depto. de Ing. Industrial, U. de Chile
- Fundador y Director, Centro de Investigación en Inteligencia de Negocios (2011 - presente)

- Más referencias en mi sitio personal:

<http://sebastian.rios.tv>



Your Profile
[linkedin.com/in/sebarios](https://www.linkedin.com/in/sebarios)

Websites
ceine.cl (Company Website)



<http://sebastian.rios.tv/>

<https://www.linkedin.com/in/sebarios/?originalSubdomain=cl>

Están los estudiantes de magíster y los de pre-grado, los que llevan conmigo muchos más años, para que los conozcan, porque éstos no son esfuerzos individuales, al final. Todas estas cosas son equipos de trabajo que están ahí, empujando a que las cosas salgan.

Primero les voy a hablar de una motivación cortita respecto a problemáticas que podemos resolver con Ciencia de Datos, después ¿qué es Ciencia de Datos? y les voy a mostrar un caso de éxito que es el Almohadita-Ámbar, que lo hicimos en un Hospital pediátrico.

HAY PROBLEMAS COTIDIANOS QUE NOS AFECTAN A TODOS!

SE DEMORA UNA HORA EN PASAR
CUANDO LLEGA VIENEN 3 JUNTAS

Podríamos hacer otras cosas

Mala calidad de Vida!

ceine

Primero, en lo cotidiano, vivimos con un montón de problemas que nos afectan en el diario vivir, en nuestra ciudad, con nuestro trabajo. Acá hay un macro problema: todo el tema de la basura, la gestión de los residuos, el reciclaje, ¿dónde colocamos los puntos limpios?, ¿cada cuánto se pasa a recolectar la basura? También tenemos todo el tema de la congestión vehicular. El tema de atención en recintos cerrados, que ahora con el tema del COVID es peor, porque la gente hace la misma fila, pero afuera, en la calle. O el tema de la gestión de energía.

Todos esos son macro temas que tienen pequeñas áreas, donde podemos hacer algo con Ciencia de Datos. Entonces, si me quedo en el tema del tráfico, uno va a internet y salen

un montón de cosas al tiro. Algo anda mal, como este que dice que la micro se demora una hora en pasar y cuando llega, llegan tres juntas. Eso lo hemos vivido todos en algún momento. Y eso también afecta porque uno se queda esperando en la fila a la intemperie. Ahora con el corona virus, esa fila es peor, porque es más larga. Tampoco pueden entrar todas las personas al microbús, así que realmente es horrible si pasa esto. O cuando uno va en un taco, ahí estamos perdiendo vida, tiempo en que podría ser productivos, estar haciendo algo o con nuestras familias. Nos da mala calidad de vida al final.

ESTADÍSTICA BÁSICA BIEN PENSADA.. YA APORTA VALOR!

Monitoreo en Tiempo Real



Podríamos hacer otras cosas

ceine

Monitoreo en Tiempo Real

Report event

Police

Traffic jam

Accident

Hazard

Speed cam

Chit chat

Update map

Sugiere una ruta alternativa

(08:12) Y esto se podría resolver, y se ha resuelto con Ciencia de Datos. Les voy a mostrar la aplicación que es obvia, que conocen: Waze. Esta aplicación yo la mostraba desde el 2008 cuando hacía cátedras. Llegaba a Chile de vuelta de Japón y se la mostraba a mis alumnos y les decía: "Ojalá que más gente la use". Y claro, ahora todos la usan y es una maravilla porque podemos entregar información, hacemos tagging colaborativo del sistema vial de la ciudad. Eso es lo que estamos haciendo: entregamos en tiempo real, con el GPS, el tiempo promedio de desplazamiento del celular. Como somos muchos, la estadística que se obtiene es mucho más precisa. Y tenemos estadísticas de todas las calles, no sólo de uno o dos celulares que antes se conectaban y que andaban dando vueltas, pues eso tampoco le sirve a nadie. Entonces, tiene utilidad en la medida en que todos lo estamos usando. Ahora, si nos vamos a la información que está detrás, no hay ningún algoritmo complejo aquí. Lo único que se está haciendo es procesamiento masivo y sacando estadísticas básicas. Se llama estadística descriptiva en el contexto de un fenómeno que antes era impensado hacer. Y tenemos un muy buen delivery, una muy buena entrega de esta información al cliente final, a las personas, a través de una aplicación móvil, en la que podemos ver la ruta que más nos convenga. Y tenemos información valiosa, simplemente con el GPS sabemos si hay un súper taco, sabemos dónde parte y dónde termina, para tomar la decisión en forma oportuna, en el momento correcto. Esa información me evita a mí estar una hora pegado en el tráfico. O tenemos

información de accidentes. La maravilla está que hay un algoritmo atrás que está generando las mejores rutas dado el destino al que yo quiero llegar. Me muestra opciones de ruta, ni siquiera tengo que pensarlas, la puedo seguir. Y toma en cuenta las características del tráfico en ese momento. Y ahí sí hay un algoritmo, una cosa más sofisticada que me entrega la mejor ruta, para poder cumplir con mi meta de demorarme menos. Incluso, puedo indicar a la máquina que evite los peajes, si es que no quiero pagar. Mi función objetivo puede cambiar.

(10:33) Pero el punto es que el estado cero es la estadística descriptiva, esos datos granulares del fenómeno que yo quiero encontrar es lo más importante. Y es siempre la barrera que uno se topa cuando quiere hacer Ciencia de Datos, uno habla de los modelos estocásticos, pero si no tengo el sustrato básico, es imposible construir cualquier aplicación.

PERO HAY MUCHOS OTROS PROBLEMAS...



Problema: Inventario

- No se sabe que productos hay disponible ni que significan los diferentes códigos

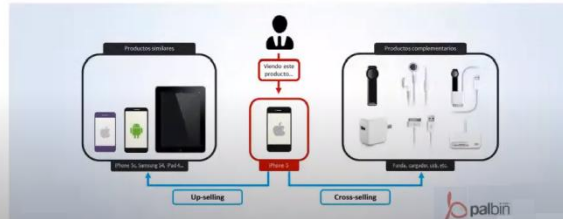


Y así hay muchísimos otros problemas de los cuales me podría explayar mucho rato. Por ejemplo el tema de la imagen que tienen las empresas en internet. La gente reclama de los servicios de las empresas en internet, queda por escrito y afecta la imagen de la empresa. Entonces, procesar esos textos, procesar las opiniones en el diario, los blogs, Facebook, es un tema. Es información masiva, es texto, que es una variable muy subjetiva de procesar, pero hay técnicas, se puede hacer cosas con eso.

El tema del inventario, probablemente muchos de acá están en organizaciones que tienen bodega, que se ven como esta, probablemente están más desordenadas, otras pueden ser más ordenadas, pero los problemas persisten: el tema del sobre stock, el tema de los quiebres de stock, ¿cierto?, el tema de las mermas, el tema de la obsolescencia. Y todo eso antiguamente era registrado en papel. Hoy día está en gran parte está digitalizado, pero no se utilizan los datos de forma correcta para mejorar los procesos en la bodega. Tengo una cátedra de procesos en ingeniería industrial, son 30 proyectos por año, y por lo menos 10 de esos proyectos son en bodegas. Entonces, imagínense, llevo 24 semestres dictando el curso y les puedo decir que este es un problema súper común, que está resuelto hace 20 años.

EJEMPLO 1. UPSELLING & CROSS SELLING

► Típicamente en una empresa de retail es interesante caracterizar los clientes que son potenciales candidatos para poder hacer una campaña de up-selling o cross-selling.



EJEMPLO 2: DETECCIÓN DE FUGA DE CLIENTES

1. Se sabe que es más barato retener los clientes que captarlos
2. Sin embargo, por mala calidad de servicio, o por ofertas de la competencia, entre otros, los clientes pueden decidir dejar nuestra empresa
3. La pregunta que surge es: ¿Seremos capaces de detectar a estos clientes que se desean cambiar a otra empresa?

(12:36) Y aquí hay más temas, el tema del Upselling (venta adicional) y el Cross selling (venta cruzada de complementarios) de productos que puedo vender, dado el perfil de compra de cada cliente. Porque puedo vender una línea de productos mejor, superior, puedo vender accesorios. ¿Qué accesorios le vendo? El tema de la segmentación de clientes, un tema de clusterización, muy tradicional, se enseñan todas las maestrías de Ciencia de Datos, en minería de datos se enseña esto. El tema de la fuga de clientes, cuando ustedes tienen un servicio mensual. ¿Quiénes se escapan?, ¿Quiénes se cambian? la fuga, también un problema típico de Ciencia de Datos.

EJEMPLO 3: DETECCIÓN DE FRAUDE

Robo de Electricidad

Robo de Agua

Robos en Banca: Poco conocido pero más frecuente de lo que se piensa. Personal de un banco, roba fondo desde ciertas cuentas. (ej. Ancianos)

EJEMPLO 3: DETECCIÓN DE FRAUDE

Evasión de Impuestos: Esta es una de las aplicaciones más famosas en el mundo de BI.

Supermodelo fue detenida en Israel por evasión de impuestos

Fusión Cuprum-Argentum al rojo vivo: Contraloría refuta por segunda vez a Super de Pensiones y ordena abrir proceso formal de invalidación

Y el tema del fraude: robo de energía, robo de televisión, robo de agua. En servicios es un tema que es súper importante para aplicaciones de Ciencia de Datos. Fraudes financieros, evasión de impuestos en el Servicio de Impuestos Internos, el tema de fusión de las AFP. Y así podría poner muchos más ejemplos. Traté de poner estos que son como más visibles, para motivarlos un poco. Pero claro, ustedes al interior de su organización tienen muchos otros problemas y son cosas mucho más específicas que estos grandes y macro problemas.

Y ¿cuál es el común denominador? las personas que trabajamos en una organización pública o privada, necesitamos tomar decisiones de forma ágil, informadas, oportunas.

CUAL ES EL COMÚN DENOMINADOR...

- En cualquier organización **publica o privada** la toma de decisiones debe ser **ágil, informada y oportuna** para generar una ventaja competitiva.
- Sin embargo, hoy en día la **cantidad de datos** con los que contamos en la empresa es tan **grande** que muchas veces no es posible cumplir con lo anterior



Necesitamos mejorar los procesos productivos para lograr ventajas competitivas, lo que implica funciones y objetivos distintos: bajar los costos, aumentar el ingreso, mejorar la satisfacción de mis clientes. Alcanzar más de uno, varios juntos, pero necesito hacerlo como parte de una organización.

Y sin embargo, la cantidad de datos que hay, por la escala de Santiago, donde ya tenemos 7 millones de personas, significa en cualquier servicio miles de clientes: 100 mil, 200 mil. Entonces, la cantidad de datos asociados a los servicios en general, la escala que tienen, las magnitudes son gigantes, para esas personas y toda esa actividad, multiplicado en varios órdenes de magnitud. Entonces, ¿Cómo logro procesar esos datos rápido? ¿Cómo logro generar información valiosa para tomar decisiones ágiles, informadas y oportunas, logrando esa ventaja competitiva? Ahí es donde aparece la Ciencia de Datos, como una respuesta metodológica a cómo hacer ese procesamiento, que no es trivial.

1. ¿QUÉ ES DATA SCIENCE?

DEFINICIÓN DE DATA SCIENCE

- La Ciencia de Datos es un área multidisciplinaria mediante la cual se busca analizar **grandes volúmenes de datos** de una empresa u organización, con el fin de **generar una cierta información/conocimiento...**[para una **mejor toma de decisiones.**]- (sríos)
- “mejor toma de decisiones”:
 - En el momento oportuno - ojalá proactivamente
 - Que sea la mejor decisión posible (dada la información disponible)
 - Depende de nuestro propósito: ya sea menos costos, más ingresos, o mejorar la calidad de la atención, etc.

ENORME CANTIDAD DE DATOS.. PERO ASÍ NO SIRVEN!

- Multitud de Plataformas
 - SAP, excel, e-mail, etc
- Qué problemas con los Datos?
 - están distribuidos
 - en formatos distintos
 - duplicados
 - desactualizados



- Problema de Calidad de Datos
- Problema de la Integración de Datos
- Problema de Capacidades Humanas
- Problema de Regulación (privacidad)
- Muchos otros Problemas...



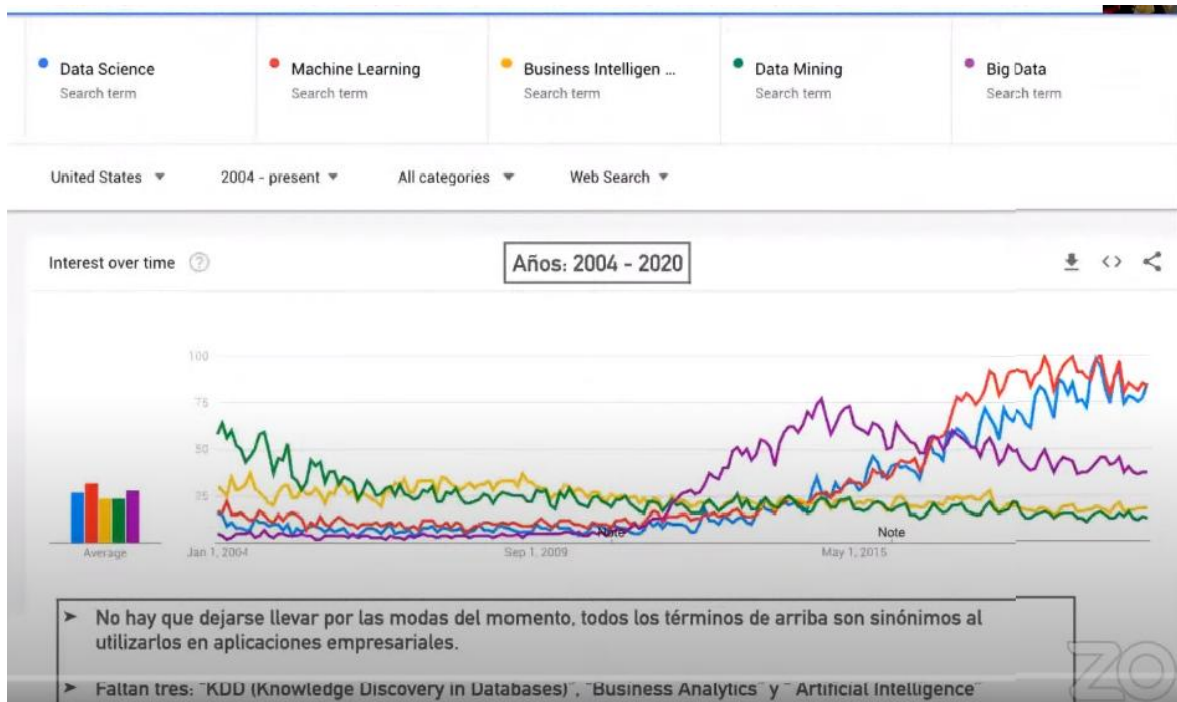
Entonces, ¿Qué es la Ciencia de Datos? un área multidisciplinaria que lo que busca es procesar grandes volúmenes de datos de una empresa u organización, con el fin de generar cierta información o conocimiento. A veces le colocan “valioso”. Pero en la Ciencia de Datos, lo que a veces se olvida o deja fuera de la definición es el propósito:

“para una mejor toma de decisiones”. Tal vez, la Ciencia de Datos siempre fue empujada por las universidades, partió el año 2009. Y claro, para los académicos, para gente como yo, trabajar en el algoritmo porque sí está bien. Yo publico un artículo, ahí está la técnica, el método que les va a ayudar a mejorar su clusterización, su segmentación de mercado, bien, aplausos! Pero cuando uno está en un ambiente empresarial, no basta el mejor método, porque si el mejor método no apoya un proceso productivo real, concreto, con problemas, que se quiere mejorar, no tiene mucho sentido hacer el procesamiento. Entonces se los doy como una insinuación, nunca olviden: “superpongan el proceso productivo que ustedes quieren apoyar, a la técnica” No tiene sentido que lo hagan a la inversa. A menos que sean académicos.

(16:40) Y lo que pasa cuando uno quiere hacer esto, es que uno se encuentra con este cuadro desolador, pero que es súper común, en servicios públicos sobre todo: toneladas de papeles, fichas que están metidos en bodegas, en un closet, con la lluvia se les mojaron. Yo lo he visto. Y en el mejor de los casos, cuando no tenemos papeles, tenemos softwares y pasa lo de la imagen que está abajo: que tenemos un montón de softwares por todos lados, que tienen formatos distintos, están distribuidos, la data está duplicada, no sé a cuál creerle, están los datos desactualizados. Y entonces todo esto genera un montón de problemas. Problemas de la calidad de los datos, problemas de integración, problemas de las capacidades humanas, porque ya al tipo no le basta con saber Excel, tiene que procesar el SAP, tiene que conectarse a otros, tiene un MySQL, etc. Hay problemas de regulación. El tema ético es súper importante, siempre se dejan de lado los temas que los tratan. Y así surgen una infinidad de problemas que no son sólo técnicos. Es un tema más cultural, una barrera que está ahí frenando todo el avance. Muchas veces es más cultural que técnico, porque la tecnología está ahí, y ahora está empaquetada, maravillosa, para llegar y usarla. Y hay evidencia empírica de esto. ¿Por qué no se hace? Sistemas ERP para manejo de bodega existen hace 20 años, ¿Por qué no los usamos en Chile? Yo he ido a empresas privadas con camiones, empresas grandes de alimentos, tienen el ERP y no lo usan o usan el 20% de la funcionalidad del monstruo que tienen, ¿Por qué? no lo puedo entender, pero bueno. Temas culturales en nuestro país que tenemos que mejorar.

Y ahora me voy a la parte que decía “más allá de la moda”, para explicarles un poco que esto no es nuevo. También se le conoce como Inteligencia de negocios, análisis de negocios, Analítica, Minería de datos, entre otros. Tal vez el nombre, Ciencia de Datos, es algo que está súper de moda ahora, pero no hay que desesperarse. Hay un sólo propósito, hay muchos nombres y muchas modas, pero ese gran propósito se ha mantenido en los últimos 20 o 30 años. Y el propósito es lo que les decía: cómo logro tomar decisiones de mejor manera, de forma más informada, oportuna y ágil, para mejorar mi proceso

productivo. Entonces si yo me remontara en el pasado, “Knowledge Discovery in Databases” o Descubrimiento de conocimiento en bases de datos, es como el término más viejito, con el que yo podría vincular data con procesos productivos, es del año 1992.



Me metí al [Google Trends](https://trends.google.com/) y busqué la evolución de los términos desde el año 2004. De ahí para atrás no hay datos. Y claro, el año 2004 estaba fuerte la línea verde que es Minería de Datos, y en estado emergente la línea amarilla que eso es Inteligencia de Negocios. Ambas son lo mismo, es lo que les he dicho: procesamiento de los datos masivos con técnica congelada, para poder congelar información y tomar decisiones. Lo mismo. Entonces, el año 2004 comenzó a bajar la popularidad de Minería de datos, porque empezó a subir la popularidad de Inteligencia de negocios. Y hacia el año 2008, digamos, Inteligencia de negocios se instaló fuerte, al menos en Chile. Y las empresas, sobre todo las TELCO, que son las que tímidamente empezaron con tecnología y son más flexibles, crearon áreas de inteligencia de negocio, después siguieron la banca, y así sucesivamente vino una lluvia de áreas de inteligencia de negocio el año 2008, que se correlaciona con este gráfico. Y eso duró como hasta el 2010, donde ahí viene una línea morada, de BIG DATA. Y hay una explosión ahí, y la única diferencia que tiene cada una de estas líneas es el énfasis que hacen. BIG DATA es el mismo procesamiento, pero el énfasis que hace es en el fierro, en el hardware. En qué tecnología voy a usar para acopiar la data masiva y que sea eficiente, que la pueda recuperar rápidamente, para que los algoritmos que vienen después la puedan utilizar. O en el Panel de control: que los gráficos se puedan generar de forma rápida y no se quede pegado el relojito y que no pueda tomar ninguna decisión. Y tiene

sentido, porque en empresas que tienen papeles y todo, llegan las consultoras a hablar de Big Data y los mecanismos que vamos a hacer eficiente, vamos a digitalizar las cosas. Y empieza y cobra una importancia brutal, se hacen proyectos de BIG DATA, pero eso en realidad dura como 4 años más o menos y empieza a bajar. Claro, ya resolví el problema de los fierros, de cómo guardo. Ahora necesito concentrarme en cómo los procesos. De hecho entre el 2010 y el 2015 se instala el tema de la computación en la nube. Al principio causaba un poco de temor, un poco de pudor de “quién va a tener mis datos” ¿Quién va a ser el dueño de mis datos? Que no los voy a tener yo. Ahora, si ustedes van a cualquier empresa grande, parte de su proceso productivo está en la nube y ni se preocupan de quiénes son los datos, quién los procesa. Ahí mismo se instaló, se resolvió el tema del almacenamiento, etc. Y entonces empezó a surgir la línea roja y la línea celeste, que son de procesamiento. Empezó a hacerse popular el Aprendizaje automático, línea roja, y Ciencia de Datos, celeste, cómo áreas. Ahora, si uno se pone a pensar en Aprendizaje automático o inteligente, no está mal hablar de eso. Cuando uno habla en un contexto empresarial, lo que está diciendo es que esas técnicas las va a utilizar para tomar decisiones, rápidas, informadas, ágiles, para producir una ventaja competitiva. Pero el Aprendizaje automático, la verdad, viene del año 60, de la mano de la inteligencia artificial, y es una sub área de la inteligencia artificial, y se utilizaba para tratar de hacer pensar o simular inteligencia humana a una pieza de software, que ojalá tuviera cuerpo humano. Pero la robótica es otra área, más electrónica.

(23:47) Pero claro, después de que hablamos del Big Data nos empezamos a preocupar de la parte del procesamiento. y por eso hoy si ustedes se fijan, ya nadie habla mucho del Big Data, pero sí hablan de Aprendizaje automático, si se habla de Ciencia de Datos, como los algoritmos que van a consumir esa data y van a generar información valiosa, para tomar decisiones. Pero esa definición es la misma de la inteligencia de negocios del año 2004, es la misma de la Minería de Datos del año 98. Claro, la diferencia está que el año 98, Big Data era el Excel que aguantaba cien mil filas y se moría. Ahora no, ahora Excel tiene un complemento, similar al Power BI, y ustedes pueden usar Power BI, entre otras cosas más y hacer maravillas con Excel. Entonces, ojo, para que no les de susto, porque son modas, lo importante es tener claro lo que ustedes pueden hacer con las técnicas.

2. CÓMO SE CREA VALOR A PARTIR DE LOS DATOS

El valor se crea y se produce cuando la tecnología logra apoyar un proceso productivo concreto. No es la tecnología por la tecnología. Entonces, no tengo que pensar en la red neuronal que voy a aplicar, no tengo que pensar en algoritmos super especiales, en

Aprendizaje profundo. Hay empresas que me han llegado diciendo que quieren aplicar Deep learning para mejorar sus procesos productivos. Es un algoritmo, es una técnica. Y ante la pregunta: ¿Será la mejor herramienta para resolver problemas? o ¿Cuál es el problema productivo que tienen? No tienen idea de cuál es el problema productivo que tienen, qué quieren usar, a qué quieren aplicar las técnicas. Entonces, ese es el peor problema. Entonces, mientras ustedes tengan claridad en el problema, después se verá la técnica que se utiliza. Puede que sea incluso estadística descriptiva, que sean cosas súper simples, y que no sea necesario aplicar Ciencia de Datos.

ROBOT PROCESS AUTOMATION

- ▶ Es en este contexto que surge como la IA afecta los procesos productivos.
- ▶ Surge el término Robot Process Automation o RPA
 - ▶ Y Aquí se habla de BOTS que tienen IA.. y se caricaturiza con un robot.
 - ▶ Sin embargo, en la inmensa mayoría de las soluciones, se refiere al SOFTWARE que hace alguna tarea automáticamente. No tienen un cuerpo físico.
 - ▶ Se utilizan para dar respuesta a clientes porque usualmente hay preguntas MUY frecuentes.
 - ▶ Se utilizan para contactar clientes antiguos en un CRM y darles promociones
 - ▶ Contactar prospectos de clientes nuevos
 - ▶ Contactar clientes con alta probabilidad de cambiarse de servicio



INDUSTRY 4.0: LINEAS PRODUCTIVAS MÁS INTELIGENTES

- ▶ En el fondo es similar a RPA, pero acá si hay máquinas (robots). y se le puso un nombre (que viene de los manufactureros).



(25:39) Entonces hay dos opciones: el RPA (Automatización de procesos robotizados), donde se habla de cómo podemos automatizar los procesos productivos usando inteligencia artificial, que es llevar los algoritmos a la última consecuencia. Que los procesos sean lo más automáticos posibles y dejar personas fuera de esos procesos productivos, que lo hagan las máquinas, típicamente softwares. O Industria 4.0 que dice lo mismo pero en un ámbito más industrial, donde tengo una línea productiva, una cadena de producción. Y las cadenas productivas típicamente tienen maquinarias.

INDUSTRY 4.0: LINEAS PRODUCTIVAS MÁS INTELIGENTES



3. Es posible por ejemplo:

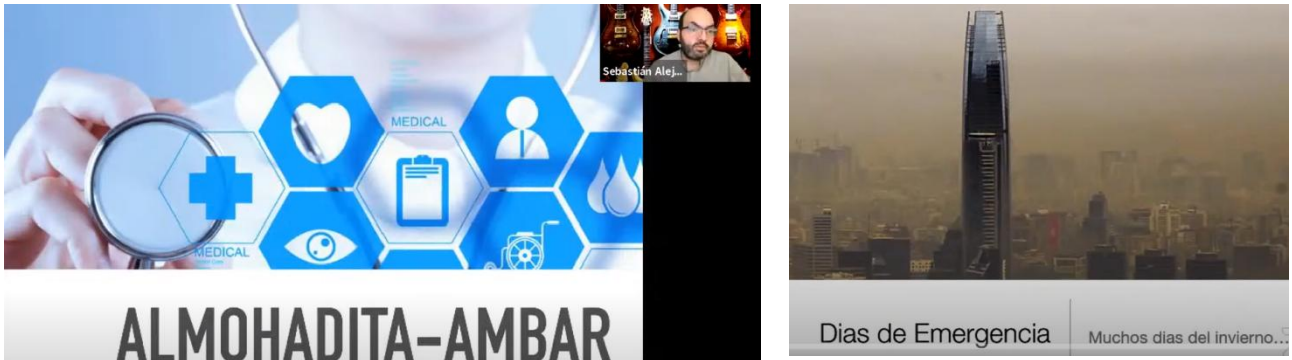
- Predecir fallas en uno o más equipos para generar mantenimiento proactivo de los equipos.
- Encontrar patrones de fallas, es decir, si una máquina A falla, esto se produce porque las máquinas B y C (previas) fallaron.
- Planificar mejor la mantención de una planta

1. Las máquinas transmiten datos de sus sensores a un centro de datos.
2. Los datos son procesados por algoritmos de Ciencia de Datos para encontrar Información/Conocimiento valioso



Se llaman máquinas herramientas, son robots, producen cigüeñales, que tienen su panel de control y tienen su estatus. Dicen si están operando mal, si están operando más o menos, todas tienen sensores y todas reportan estatus. En Industria 4.0 lo que se les ocurrió fue conectar toda la data, todos los sensores y mandar la data a un único lugar.

Se guardan esos datos, se procesan y tratan de generar información valiosa. Por ejemplo el tema de predicción de fallas. En mantenimiento predictivo es un tema súper importante dentro de la Industria 4.0. Porque parar una línea productiva por accidente va a tardar mucho más en ser reparada y por lo tanto, estoy perdiendo plata como empresa. Entonces si yo pudiese predecir cuándo va a fallar una máquina y planifico su mantención, voy a perder menos. Ese es un propósito de negocio súper concreto donde se trabaja con Ciencia de Datos.



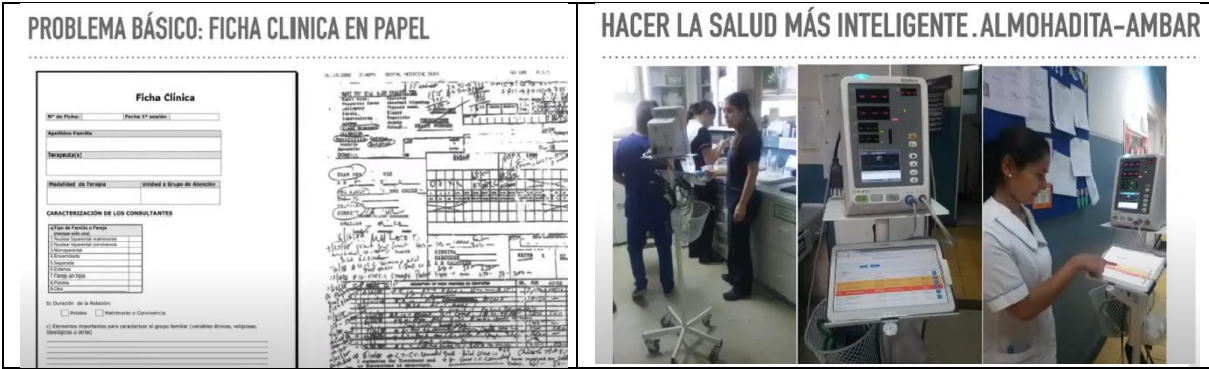
(27:34) Bueno y así llego a mi proyecto, al Almohadita-Ámbar, que se trata de monitoreo de pacientes que tienen enfermedades respiratorias crónicas a los que debemos predecir sus riesgos de ponerse mal, peor de lo que están en un momento determinado del tiempo. Surge como inspiración el que en Santiago, en todos los inviernos, recurrentemente, está súper contaminado. Y cuando vamos a un hospital público, también en los privados, pero en los públicos es peor, uno se encuentra con ese panorama: las familias llevando a los niños a la urgencia. Y el problema es que los niños con enfermedades respiratorias crónicas tienen muchísimas posibilidades de hacer una secuela grave. Entonces tenemos que asegurarles cama y asegurarles una buena atención, ojalá más rápido que a los niños normales.



Entonces el proyecto consistió en tener una plataforma de sensores y con esos dispositivos conectar al niño y medir las señales biométricas: temperatura, presión,

saturación de oxígeno, frecuencia cardíaca, y esos datos comunicarlos a un servidor central. Ahí se hace el procesamiento de esa data en tiempo real. Y estamos buscando patrones. Y si encontramos algún patrón, y si encontramos que se va a poner peor, hacer una predicción de ese estado y desplegarlo al equipo médico, a las enfermeras que están a cargo de esos niños, para que ellos, puedan tomar alguna acción.

Pero nos encontramos con un panorama que es muy común en los hospitales públicos: que la ficha está en papel. A la izquierda la ficha teórica. A la derecha, la realidad: es una ficha que está rayada por todos lados, que sólo ellos la entienden. Tiene números y signos biométricos. Si es que tiene números cuesta mucho tomar y reconstruir un caso a partir de esto. Piensen que los crónicos están varios días, no son cualquier niño. No van dos horas al hospital, están tres, cuatro o diez días. Los monitorean cada dos horas y registran en estas fichas en papel. Después de 3 días, su ficha es un alto de papeles y la revisan varios médicos y con esos papeles el médico tiene que reconstruir la historia. Entonces, dado que es muy inseguro el papel, queríamos mejorar la seguridad de la atención.



(30:46) Para eso creamos una plataforma, que está finalmente en esa tableta. Pueden ver los colores que genera el algoritmo. La gente del hospital ya medía los signos biométricos con estos monitores hospitalarios que tienen sensores. Y esas medidas las colocaban en el papel.

DIGITALIZAMOS EL PROCESO DE CONTROL DE SIGNOS VITALES. . .



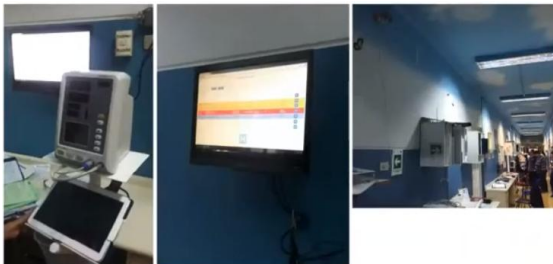
¿Qué hicimos? creamos un software que transmite a una tableta o celular, lo que tengan a mano, donde ellos puedan revisar estas variables: temperatura, presión cardíaca, etc., de forma muy sencilla, muy rápida. Pueden meterlo ahí con estos enlaces. ¿Cuál es la gracia? Ellos pueden reconstruir los estados aquí, se generan los gráficos de las variables biométricas. Pueden realizar anotaciones, todo en un solo registro.

REVISIÓN DEL ESTADO ACTUAL Y PREDICCIÓN



Pero lo más importante, es que hay un algoritmo detrás que está generando una predicción de riesgo intra hospitalario. Y esa es la línea rosada que está ahí. En este caso, la línea oscila, no puede generar nada, no le cree a su predicción. Porque tenemos varios tipos de predicción. Cuando no logramos tener historia suficiente para generar una predicción, queda en blanco como está en esa pantalla que está al centro, el algoritmo no le puede dar información extra al doctor.

Pero cuando tenemos historial suficiente y logramos poner un punto que le cree al algoritmo, podemos pintar en celeste, que es riesgo bajo, en amarillo que es riesgo medio o en rojo que es riesgo alto. El riesgo alto, cuando en algún niño su estado es rojo, eso quiere decir que en dos horas más ese niño va a estar peor de lo que está ahora, o la probabilidad de que esté peor es mucho mayor. Entonces, en vez de chequearlo, de irlo a ver cada dos horas, lo van a ver cada media hora. Entonces, cambiamos el protocolo clínico de atención hospitalaria, gracias a la información del sistema.

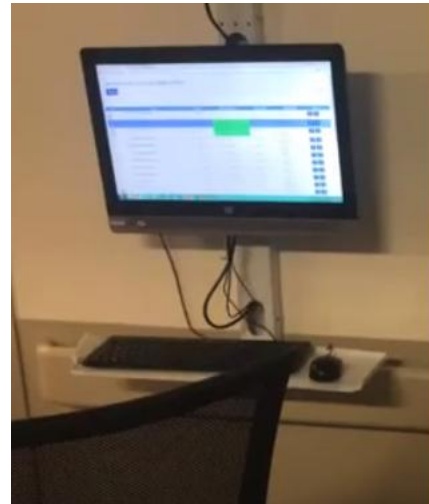


Y al que está celeste en vez de chequearlo cada dos horas, lo chequeo cada tres, entonces libero un poco de horas de las enfermeras y los técnicos con los que están menos grave, pero aumento los chequeos con los que están peor. Y eso es un chequeo que lo puedo hacer de forma dinámica.

De hecho el software tiene los nombres de las personas que atienden, de las enfermeras, y automáticamente asigna a una persona para revisarlo en un tiempo determinado. Entonces, con el algoritmo se cambia el proceso productivo completo, se está accionando con algo que no se podía hacer antes. Antes, cuando teníamos puros papeles y cuando estábamos en la época en que no había tantos computadores, es lógico pensar que si tengo que revisar a 30 pacientes, bueno, a las 9, 11, 15 y 18 vamos a revisar a todos. Esa es la forma más simple de trabajar. Pero no es la óptima. Con el algoritmo, con toda esta data que tenemos, con todos estos estados predictivos, el algoritmo planifica en tiempo real, que en media hora más tengo que hacer una revisión de ciertos pacientes, en una hora y media más a estos otros, y en tres horas a los que están menos grave. Y lo hace

dinámico, porque tengo la data, porque el proceso ahora está digital, y gracias a eso yo puedo mejorar mi atención.

EL NUEVO EXEQUIEL... FINALES 2017



Este sistema lo empezamos a pilotear el año 2015 y nunca lo bajamos, esa es la verdad. Lo empezamos a pilotear con tres cunas, después pasamos a 33 cunas de la Unidad de lactantes del Hospital Exequiel y después pasó a todo el hospital. Y el año 2017, el Exequiel se cambió de lugar físico a un hospital nuevo, y estas son fotos del hospital nuevo. Es un hospital lindo, con equipos nuevos, y ahora están monitoreando 160 cunitas, todos los días. La gente trabaja con esto, ahí está la tableta, el historial, el índice de riesgo. Y bueno, tenemos estos dos monitores que son táctiles grandes, interactivos, por pasillos, son varios pisos. Aquí los doctores pueden tocar las pantallas y juntarse en el pasillo, lejos del paciente, para discutir el caso. Hasta octubre de 2018 lo usaron más de 3.800 pacientes, se monitorearon en el sistema 160 cunitas, y está dentro de todo el hospital. Muchas gracias.

- José: muchísimas gracias Sebastián. Un caso realmente maravilloso. Muchos me habían comentado lo emotivo que resulta comprobar cómo se puede aumentar valor con la ingeniería, con la ciencia y sobre todo con el compromiso de académicos como tú. Damos la palabra a Charango, para su exposición.

Desmitificando la Ciencia de Datos

¿Es la Ciencia de Datos para Todos?

José Charango Munizaga-Rosas
13 Noviembre 2020



-Charango Munizaga: (36:37) Buenos días a todos. Hoy estoy con la camiseta de la U de Chile en honor a [Carlos Campos](#) que nos dejó esta semana, gran goleador, que nunca abandonó a la U.

El tema que voy a abordar es similar al de Sebastián, pero el ángulo es probablemente bien distinto. Tengo una opinión, que es personal y que está basada en cierta vivencia empírica, de que la ciencia de datos puede ser una cosa bastante más masiva de lo que uno piensa. Y de hecho, lo que vamos a mencionar hoy día, hay una introducción, vamos a explicar lo que es, lo que no es, porque también es importante, un tema de manejo de expectativas, y vamos a dar algunos detalles de un entrenamiento que hicimos en la SUSESO y que después yo lo he repetido en Perú y en Australia. Y hay cosas comunes que caen o que aparecen como factores a favor y otras cosas que son potenciales obstáculos, pero hay un proceso de aprendizaje y una forma de enfrentar el problema de que si la ciencia de datos es para todos o no. Porque aparentemente es como una cosa súper etérea, está súper encima, súper volada, es como que unos pocos escogidos pueden hacer uso de esas herramientas, pero he empezado a aprender que ese no es tan necesariamente el caso.

Como contenidos, voy a decir cuál es mi prontuario criminal, algunas definiciones básicas, el proceso de la ciencia de datos, algunas consideraciones prácticas que no son menores, algunos mitos comunes (encontré un listado y hay varios mitos y los vamos a explicar bien cortito), y básicamente esta es una propuesta de modelo de entrenamiento de desarrollo de competencias. Y vamos a explicar por qué arma un montón de sentido, porque básicamente desde el punto de vista de la función pública, es posible crear estas capacidades. Ese es el ángulo que le quiero dar. O sea, estoy contando un poco cuál es la conclusión. ¿Es posible hacerlo? Sí, y voy a explicar cómo se llega a ello.

¿Quién es José Charango Munizaga-Rosas?

- Ingeniero Matemático y Magíster en Gestión de Operaciones (Universidad de Chile)
- Ph.D. in Natural Resources Engineering (Laurentian university, Sudbury, ON, Canada)
 - Ex profesor de Planificación Minera, Estimación de Recursos e Investigación de Operaciones Mineras en WASM, Kalgoorlie.
 - Ex Senior Lecturer en DMEE, Curtin University, Perth, Australia.
- Actualmente Principal Consultant and Chief Data Scientist en Coalesce Group y Chief Data Scientist en BlueSky Labs, Perth, Western Australia
- Fundador y CEO de Minformatics SpA, Santiago, Chile
- Profesor Adjunto en el Departamento de Mineral and Energy Economics, Curtin University, Western Australia, Australia
- Profesor Adjunto, Departamento de Ingeniería de Minas, Director Académico Diplomado de Economía de Minerales, Universidad de Chile, Santiago, Chile
- Profesor Extraordinario, Facultad de Ingeniería de Minas, Universidad Nacional del Altiplano, Puno, Perú



Este soy yo, en una foto decente. Básicamente me dediqué a trabajar en minería, fui profesor de la Escuela de Minas en Australia, me aburrí un poco de la academia y después me fui a la industria.

Soy profesor adjunto del Departamento de Ingeniería de Minas de la Universidad de Chile, profesor extraordinario de la Facultad de Ingeniería de Minas de la Universidad Nacional del Altiplano, Puno, Perú. Y tengo mi consultora en Chile y también acá en Australia que es dónde vivo. En el pasado he hecho un montón de cosas en minería, pero también empecé a trabajar en Ciencia de datos. Las cosas que hacía en minería son planificación minera, geo estadísticas, todas estas cosas como súper matemáticas, porque es un problema de formación profesional. Eso fue lo que estudié como pre-grado, pero empecé a aplicar a un montón de problemas: logísticos, de seguridad social con la SUSESO, en salud, finanzas, distribución eléctrica y últimamente problemas de planeamiento y control de riego, todo el tema agroindustrial, sobre todo por Perú.

¿Qué he Hecho en el Pasado? (Prontuario Criminal)

- Modelos de Automatas Celulares para Flujo Gravitacional
 - Simulación de método Block caving, simulación de subsidencia para operaciones de carbón long-wall
- Investigación de Operaciones Aplicado a Minería
 - Algoritmos genéticos, programación entera mixta, planificación minera
- Geoestadística y Análisis de Riesgo en Minería
 - Uso de simulaciones condicionales para incorporar incertidumbre de leyes a la planificación minera, localización óptima de drillholes
- Ciencia de Datos en Distintos Dominios:
 - Salud
 - Seguridad Social
 - Logística
 - Finanzas
 - Distribución Eléctrica,
 - Planeamiento y Control de Riego (Agroindustria)

(40:30) Veamos algunas cosas básicas de datos. Estos son datos del 2016. Habían proyecciones en 2016, se estimaba que para el 2020, aproximadamente 1,7 megabytes de nueva información sería creada cada segundo por cada humano en el planeta. Y esto son proyecciones basadas a nivel de lo que tienen los data centers de Facebook, de Google, etc. De las estadísticas de aquella época, menos del 0,5% de todos los datos que creamos son alguna vez analizados y usados. Han pasado cuatro años, pero no han cambiado mucho las cosas. El 70% de los datos son creados por individuos, sin embargo, las empresas son responsables de almacenar y gestionar el 80% de dicho volumen. O sea, los que crean los datos no son necesariamente los dueños de los datos, terminan en manos de corporaciones. Lo que además tiene todas esas consecuencias de coletazos morales o éticos que mencionaba Sebastián en su presentación. En resumen, los datos están todos lados, y se recolectan indiscriminadamente, y es tan indiscriminado que muchas veces se recolectan datos por recolectar. Ejemplos clásicos, hay uno en minería, un camión minero tiene 800 sensores, sensores para todo, pero el tipo que ocupa el camión ocupa cero. Pero

la minería se precia de decir que es una industria de punta, avanzada, y que está ocupando tecnologías.

Acerca de los Datos (circa 2016, USA)

- Menos del 0.5% de todos los datos que creamos son alguna vez analizados y usados
- Hacia fines de 2016, un 73% de las organizaciones estaban ya invirtiendo o habían invertido en **big data**
- Google utiliza alrededor de 1,000 computadores para responder una única consulta
- Se proyectaba que para 2020, habrían más de 50.000 millones de dispositivos inteligentes conectados en el mundo recolectando, analizando y compartiendo datos (probablemente el número quedó chico...)
- En 2015, un estimado de 1 trillón de fotos fueron tomadas y miles de millones de ellas fueron compartidas online
- En 2016, 40,000 consultas de búsqueda fueron hechas cada segundo solamente en Google alone, lo que da del orden de 1.2 trillones de búsquedas por año
- Se estimaba que para el 2020, aproximadamente 1.7 Mb de nueva información sería creada cada segundo por cada humano en el planeta
- Los datos deficientes le cuestan solamente a los negocios de USA del orden de US\$600.000 millones anualmente
- 70% de los datos son creados por individuos, sin embargo, las empresas son responsables de almacenar y gestionar 80% de dicho volumen
- Hay aproximadamente tantas piezas de información digital como estrellas hay en el universo

Los datos **están en todos lados** y se **recolectan indiscriminadamente**

(42:08) Algunos conceptos generales. Esto es muy similar a lo que estaba hablando Sebastián. Una compañía o una organización racional quisiera ahorrar dinero, aumentar la eficiencia, reducir los costos, etc. Quisiera hacerlo bien. Para satisfacer cualquiera de estos deseos, se necesita crear una estrategia, crearla y aplicarla. Una estrategia inaplicable sería una tontera. Ahora, esta estrategia puede ir desde la aplicación de reglas simples a modelos súper complejos. Y ojo que más no es mejor en este caso. Existe este concepto de que mientras más rimbombante sea el nombre de la técnica que uno va a ocupar, mejor van a ser los resultados que uno va a obtener. Y la verdad es que no es necesariamente el caso.

Conceptos Generales (1)

- Todas las compañías/organizaciones "racionales" quisieran:
 - Ahorrar dinero
 - Aumentar la eficiencia
 - Reducir los costos
 - Etc.
- Para satisfacer cualquiera de esos deseos, una *estrategia* es necesaria y debe ser aplicada:
 - Esta estrategia puede ir desde la aplicación de reglas simples a modelos super complejos
- ¿Qué se necesita?
 - Primero, una **definición apropiada del problema**:
 - El hecho de que una compañía quiera hacer más dinero no garantiza que ese dinero sea obtenible
 - También es **esencial definir un mecanismo que permita discriminar buenas soluciones** (bajo el supuesto de que el problema a resolver vale la pena, i.e., hay un problema)

"Si tienes un problema que puede ser resuelto, no tiene sentido preocuparse. Si tienes un problema que no puede ser resuelto, no tiene sentido preocuparse" – Proverbio budista

(42:47) ¿Qué es lo que necesita uno en general? Una **definición apropiada del problema**, y esto es súper importante, mucha gente no define bien los problemas. El hecho de que una compañía, por ejemplo, quiera hacer más dinero, no garantiza que ese dinero sea obtenible. Y lo que muchas veces uno si puede hacer, es ver síntomas de problemas, pero no los problemas en sí mismos. Así que todo el proceso de la problemología, en teoría de sistemas, todo el proceso de tratar de entender ¿qué es lo que está pasando y qué es lo que hay que resolver?, es tal vez mucho más importante que la forma en que se resuelve. Y de la misma manera, hay que definir un mecanismo que permita discriminar buenas soluciones, una vez que ya tengo un problema definido. Debo buscar dentro del espacio todas las posibles soluciones y definir cuál es la mejor. Eso bajo el supuesto de que vale la pena resolver ese problema.

Conceptos Generales (2)

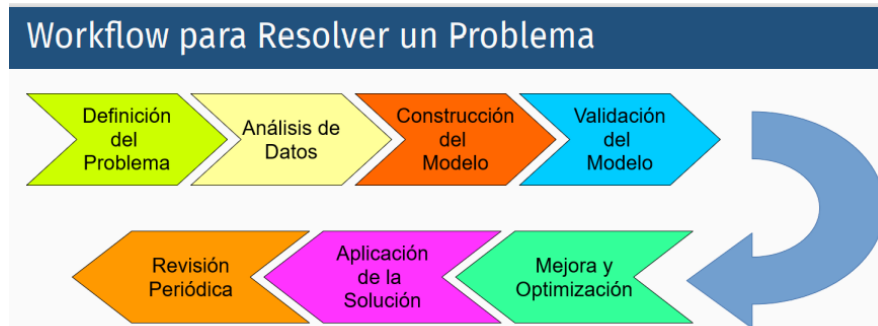
- Todo esto suena muy bonito, sin embargo:
 - Las compañías/organizaciones no tienen información completa, perfecta y transparente
 - El futuro es difícil (de hecho **imposible**) de predecir
 - Las compañías/organizaciones usualmente ni siquiera conocen los datos que poseen
 - Menos aún conocen los problemas que tienen
 - Hay **problemas metodológicos** relacionados a la recolección de datos y el como se usan
 - El proceso de curatoria de datos debe garantizar que la recolección y almacenamiento no produce sesgos
 - Cualquier error en este aspecto inhabilita el uso de los datos (**Rubbish In = Rubbish Out**)
- Los datos por sí mismos son de poco valor:
 - Para entregar valor, los datos tienen que ser usados de manera inteligente
 - Usualmente los datos deben ser transformados para ser utilizados por los modelos matemáticos
 - Y esta es la parte *aburrida pero fundamental* del trabajo donde se gasta del orden de un **80%** del tiempo y que aparentemente no produce resultados vistosos

Como comentario aparte, hay un proverbio budista que dice: “Si tienes un problema que puede ser resuelto, no tiene sentido preocuparse. Si tienes un problema que no puede ser resuelto, no tiene sentido preocuparse” Tal vez esto apunta a no obsesionarse demasiado con tratar de tener una definición perfecta de problema, porque en este caso en particular, definiendo problemas, lo perfecto es enemigo de lo bueno.

Todo suena súper bonito, conceptualmente hablando, pero como hecho general las compañías o las organizaciones no tienen información completa, perfecta y transparente, no la tienen. Cualquiera que diga y ese es el error frecuente que muchas veces he escuchado en clientes, o en potenciales clientes: “tengo una base de datos y la he estado recolectando en los 3 últimos años”. Lo primero que a uno le pasa es que se da cuenta de que la base de datos tiene faltantes, tiene problemas de codificación, hay falta de uniformidad. Yo me gasto un 80% del tiempo solamente lidiando con los datos. Y en

general en cualquier estudio de Ciencia de datos, la parte más engorrosa, fome, latera y larga es lidiar con los datos. El futuro es difícil, de hecho es imposible de predecir. Esa es otra cosa importante de anotar: las compañías u organizaciones usualmente ni siquiera conocen los datos que poseen, saben que tienen datos, pero no saben lo que tienen. No saben cómo se miden. Uno les pide un diccionario de definición de datos y mandan cualquier cosa. Entonces, ojo, que el tener datos no significa nada. Y menos aún si no conoces los datos, no conoces los problemas que tienes.

(45:15) Hay problemas metodológicos relacionados a la recolección de datos y el cómo se usan. Y el proceso entonces de curatoría de datos debiese garantizar que la recolección y almacenamiento no produce sesgos, como principio general. Nos vamos a dar cuenta de que metodológicamente hablando es muy fácil equivocarse. Cualquier error en este aspecto inhabilita el uso de los datos. Y como regla a recordar: la basura que entra es igual a la basura que sale. Puedo tener un modelo perfecto, pero si le estoy dando datos malos, no voy a obtener nada bueno. En conclusión, los datos por sí mismos son de poco valor. Tener datos es lo mismo que tener nada. Para entregar valor, los datos tienen que ser usados de manera inteligente. Hay que tomar decisiones basados en ellos, hay que hacer algo con ellos, no solamente almacenarlos o recolectarlos. Y usualmente los datos deben ser transformados para ser utilizados por los modelos matemáticos. Y eso es otra cosa: hay requisitos bastante especiales en ciertos casos con respecto a ciertos datos que uno puede incorporar o no, dentro de ciertos algoritmos. Y esta es la parte aburrida, pero fundamental, donde como ya dijimos, se gasta del orden del 80% del tiempo.



El Workflow general para resolver un problema parte por definirlo, lo que pareciera trivial, pero es la parte menos trivial. Y de aquí para adelante, si está mal hecho, todo el resto falla. Luego sigue analizar datos, construir un modelo, validar el modelo, mejorar y optimizar, aplicar la solución y después revisar periódicamente. O sea, una vez que uno parte, el problema está definido, el resto fluye. Y en general, entre la parte de definición del problema y análisis de datos se gasta harto tiempo. Construir el modelo y validarlo, es menos complicado, comparativamente hablando.

El Proceso de la Ciencia de Datos

La definición que encontré en Wikipedia, dice que la Ciencia de Datos, también conocida como la ciencia conducida por datos, es un campo interdisciplinario, que usa el método científico, procesos y sistemas, para extraer conocimiento o visión desde los datos en varias formas, ya sean estructurados o no estructurados, similar al descubrimiento de conocimiento en bases de datos, [KDD](#), por sus siglas en inglés, que también lo nombró Sebastián en sus presentación. Ahora, la Ciencia de datos está en la intersección de tres dominios distintos: matemáticas, ciencias de la computación y conocimiento del negocio. Matemáticos y estadísticos, hay en general. El conocimiento es especializado, pero hay. Gente que entienda los modelos computacionales y que sea capaz de implementarlos de manera eficiente, esto está bien facilitado últimamente con las librerías existentes, tal como señalaba Sebastián. Tal vez hace 15 años atrás como que no había nada. Hoy día hay un montón de librerías y un montón de marcos operacionales, es decir, herramientas hay, y se ha eliminado bastante del problema de enfrentar estas cosas.

¿Qué es el Data Science?

Definición adaptada de Wikipedia:

“El Data Science, también conocido como la ciencia conducida por datos, es un campo interdisciplinario que usa el método científico, procesos y sistemas para extraer conocimiento o visión desde los datos en varias formas, ya sean estructurados o no estructurados, similar al descubrimiento de conocimiento en bases de datos (KDD sus siglas en Inglés)”

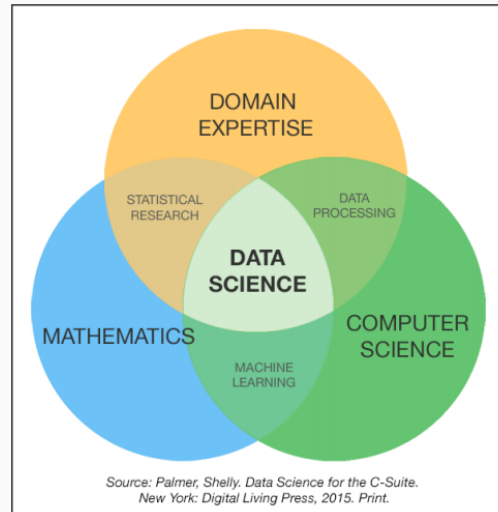
- La ciencia de datos está en la intersección de tres dominios distintos:
 - Matemáticas, Ciencias de la Computación y Conocimiento del Negocio
- Matemáticos y estadísticos hay disponibles en general, si bien el conocimiento es especializado
 - Muchos de los algoritmos están encapsulados en librerías y se pueden utilizar con algunas limitaciones dadas por los supuestos de los modelos que implementan
- Gente que entienda los modelos computacionales y que sea capaz de implementarlos de manera eficiente se facilita con las librerías existentes para ingestión y transformación de datos así como de algoritmos de aprendizaje
- **La comprensión de los aspectos del negocio, en particular los relacionados a la forma en que este opera y todos los detalles asociados a él, es la parte difícil**

(48:06) El problema, es que el tercer componente, que es la comprensión de los aspectos del negocio, en particular los relacionados a la forma en que este opera y todos los detalles asociados a él, es la parte difícil. Uno puede saber matemáticas y uno puede saber programar. Eso no quiere decir que uno pueda terminar entendiendo bien un negocio. Hay negocios con gente que lleva trabajando 10 años y el negocio es tan retorcido y tiene tantas excepciones y tantas otras cosas, que a esa persona le cuesta comunicar de manera simple de qué se trata el negocio. Entonces, ojo, esta es la parte difícil. El expertise de dominio, o en este caso el conocimiento del negocio, es la parte más difícil. Claro, las

matemáticas son complejas y si uno no tiene dedos para el piano está súper sonado, pero es algo que eventualmente uno puede subsanar. Y también la parte de Ciencia de computación, porque hay en el fondo librerías y hay herramientas. No hay nadie que tenga el 100% de todo, siempre hay gente que tiene más de uno y menos de otro.

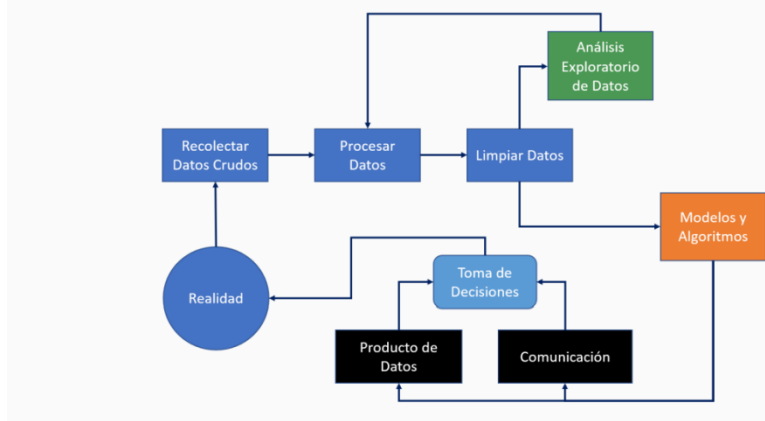
El Rol del Científico de Datos

- Dar con una persona que reúna las tres habilidades simultáneamente es difícil
 - En mi opinión no existe nadie que sea 100% bueno en cada una de las áreas:
 - Siendo el dominio de conocimiento del negocio es el que más complica ¿Porqué no mejor **entrenar** a quienes conocen el negocio para así poder identificar los problemas que hay que resolver? Sobre esto un poco más adelante...
 - Siempre es posible complementar debilidades a través de apoyo de un equipo
- La ciencia de datos **no** es **Analítica de Datos**, tampoco **Machine Learning**, menos **data mining**

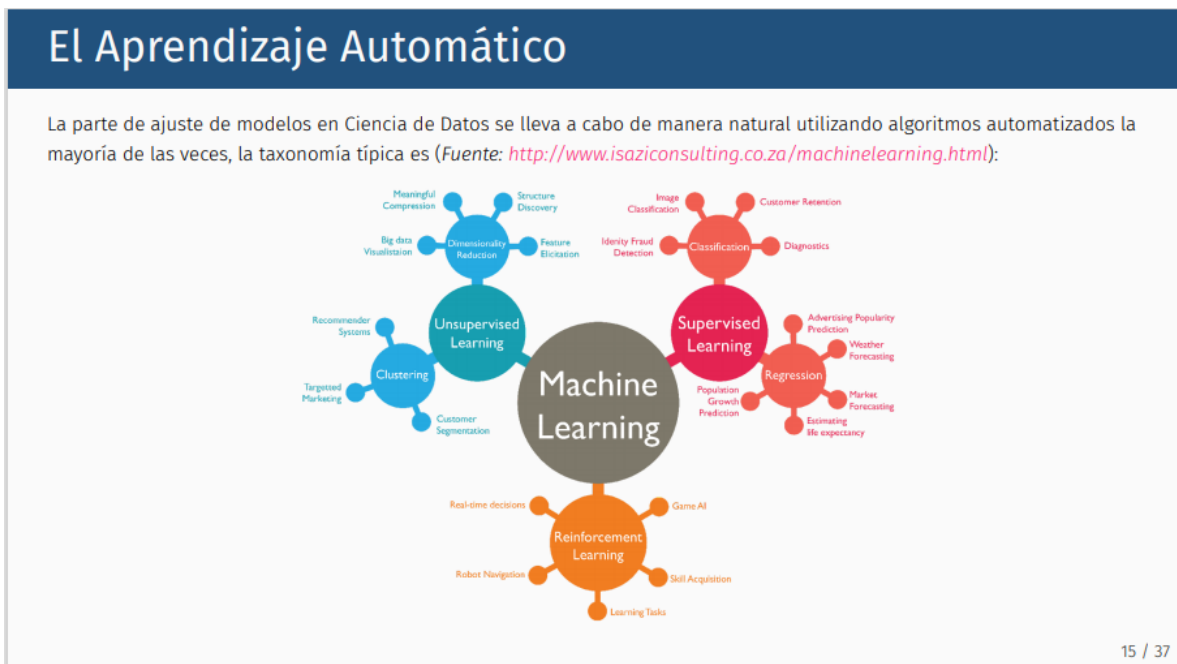


(49:13) Entonces, acá es donde parte el punto de vista más o menos filosófico. Si la parte más difícil de esta historia es el conocimiento del negocio, ¿por qué entonces no entrenamos a las personas que sí conocen el negocio? Más que tratar de enchufarle conocimiento de cómo opera el negocio y de manera completa a un tipo que sabe matemáticas y que sabe programar, ¿por qué no enseñamos los rudimentos del razonamiento matemático necesario y del uso de librerías o de herramientas a gente que sí conoce muy bien el negocio? Entonces, la ciencia de datos no es Analítica de Datos, tampoco es Aprendizaje automático y menos Minería de datos. De hecho lo vamos a ver después, porque este es uno de los mitos.

El Proceso de Ciencia de Datos



¿Cuál es el proceso? Si este círculo es la realidad, de esa realidad lo que uno hace es recolectar datos crudos. Esos datos crudos, usualmente no sirven tal y como vienen, hay que procesarlos. Una vez que uno los procesa, eventualmente uno los tiene que limpiar, ocupando herramientas en este proceso iterativo de análisis exploratorio de datos, hasta que uno puede generar una base de datos que se le pasa a los modelos y algoritmos que uno entrena. Y una vez que uno tiene el resultado de este modelo entrenado, esta capacidad de predecir, clasificar o hacer cosas con este modelo, uno recién ahí comunica o genera un producto de datos que se puede utilizar en el proceso de toma de decisiones. Cuando uno toma decisiones, uno afecta la realidad, por tanto este es un proceso iterativo, es un proceso de mejora continua, en la medida en que voy conociendo la realidad y capturo datos. Voy a tomar decisiones que van a afectar a su vez a la realidad que es la que me generó los datos en primera instancia. Entonces, esto es tratar de apuntar a un objetivo móvil, porque siempre va cambiando.



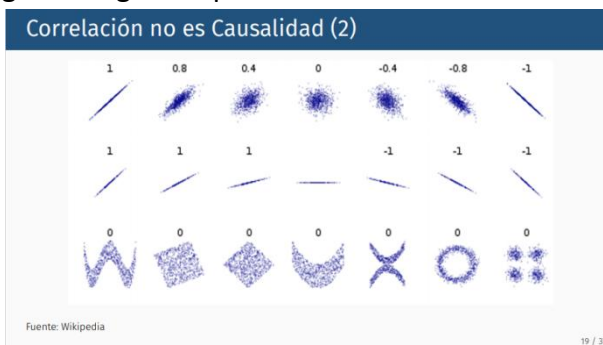
(50:54) El [aprendizaje automático](#) que tiende a confundirse con Ciencia de datos, es básicamente el conjunto de algoritmos, no más. La Ciencia de datos es un proceso que mostré en la lámina anterior, que ocupa el método científico. O sea, uno parte de un problema, elabora una hipótesis, genera un experimento, recolecta los datos y hace el modelo. A diferencia de los aprendizajes que son básicamente las técnicas matemáticas o computacionales. Hay tres grandes áreas: el **aprendizaje supervisado**, que es básicamente el típico modelo mental de regresión, donde tengo variables que son independientes, que las ocupo para explicar una variable dependiente. Luego el **aprendizaje no supervisado**, en el que tengo básicamente la tarea de encontrar algún tipo de estructura de los datos,

donde no existe ninguna estructura predefinida. Y finalmente los que tienen menos uso, pero igual existen, que son los métodos de **aprendizaje reforzado**, donde básicamente uno va tratando de capturar, por la vía de varias interacciones del usuario con el sistema, el cómo el usuario va respondiendo.

Algunas Consideraciones Prácticas

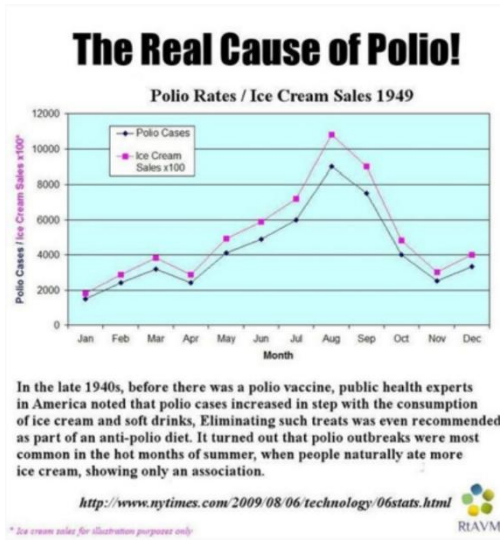
Número 1, los promedios son el enemigo. No puedo insistir más en esto. De hecho hay dos frases famosas. Una de Des McHale, de la que no pude encontrar el año, que dice, “El ser humano promedio tiene un seno y un testículo”, pero eso en realidad no lo he visto nunca. Si uno toma el promedio, uno no encuentra a una persona que sea así. Esto básicamente trata de apuntar a que el promedio no es representativo de la realidad. Es un número que está comprimiendo la variabilidad, la riqueza, del fenómeno en cantidad. Nicanor Parra, lo dijo también de una manera similar: “Hay dos panes. Usted se come dos. Yo ninguno. Consumo promedio: un pan por persona”.

(52:45) Segunda cosa importante. La **correlación no es relación de causalidad**. Básicamente uno quisiera tener una expresión $Q = f(x,y,z)$ Alguna cosa como función de otras variables, donde Q es la variable dependiente y (x, y, z) son las variables independientes de las que puedo obtener un resultado estadísticamente significativo. No implica que estas variables x, y, z que estoy midiendo causen Q , originen o den lugar a Q . La causalidad es una relación entre causa y efecto, básicamente en que un evento A ocasiona B o viceversa. Usualmente, en matemáticas uno lo mide a través del uso de correlaciones. Ahora, hay que notar que esa correlación es solamente lineal. Y acá, mi gran amigo Wikipedia me salva de nuevo.



Típica cosa que uno ve cuando está haciendo cursos de estadística: distintos diagramas de dispersión reflejan distintos niveles de correlación, desde la correlación perfecta que se presentan por una línea ya sea con pendiente 1 ó -1, independiente si es directa o inversa, pasando por la nube amorfa donde no hay relación.

Todos estos conjuntos de abajo tienen estructura claramente definida y correlación cero. El problema es que mucha gente habla en lenguaje de correlaciones: “hemos encontrado una correlación entre una cosa y otra”. Eso es lo mismo que nada, porque primero: correlación no es relación de causalidad. Segundo, hay estructuras en las que hay relaciones entre variables que no se pueden medir a través de correlaciones.



Hay un ejemplo clásico. En 1949, los brillantes de EEUU, vieron que el consumo de helado, esta línea rosada, era casi exactamente igual al número de casos de poliomielitis. Entonces dijeron, la poliomielitis está originada por el consumo de helado, entonces al año siguiente los brillantes dijeron: “nadie va a comer helado”. Adivinen qué pasó? Hubo los mismos casos de polio, ¿por qué? Porque la poliomielitis se transmite por vía aérea.

Los niños, cuando salen a jugar en el período de julio a octubre, que es cuando más calor hace, están compartiendo, transmiten la poliomielitis y además consumen helado. Entonces, hay un factor subyacente común, pero eso no quiere decir que una cosa cause a la otra. Este es el mejor ejemplo de los clásicos de que correlación no es relación de causalidad. Hay otros más. No sé si ustedes han escuchado a la iglesia del monstruo espagueti volador, el [Pastafarismo](#), ellos dicen que el calentamiento global es producido por la extinción de los piratas, y muestran un gráfico donde se ve la variación del número piratas y la temperatura global, que muestran una correlación inversa.

(55:31) Así como este hay un montón de [correlaciones espurias](#), las que pueden consultar en el sitio web que se dedica a mostrar este tipo de correlaciones que no tienen sentido. Por ejemplo, la línea verde es el gasto en Ciencias, Espacio y Tecnología, que se correlaciona muy bien con los suicidios por colgamiento, estrangulación y sofocación. Si aplicáramos la misma idea del polio, diríamos, no hay que gastar plata en Ciencia, Espacio y Tecnología para poder salvar vidas. Hay otro que muestra la tasa de divorcio en el Estado de Maine que se correlaciona muy bien con el uso per cápita de margarina. Otro de correlación inversa de colonias de abejas que producen miel con los arrestos juveniles por producción de marihuana. Todo lo cual obviamente no tiene ningún sentido, pero matemáticamente hablando, si hay una correlación.

Existencia de Posibles Sesgos

El escoger una técnica de recolección de datos condiciona las conclusiones posibles de obtener desde esos datos

Otra variable a considerar es que existen posibles sesgos. Escoger una técnica de recolección de datos condiciona las conclusiones posibles de obtener desde esos datos.

Ejemplo típico, este es del profesor Engel: si ustedes van a pescar con una malla de 10 por 10 cm, la conclusión es que no hay peces de menor tamaño. Otra más: hagan una encuesta en un supermercado a las 11 am de un día lunes. ¿A quién va a representar esa encuesta? Probablemente a las dueñas de casa, a estudiantes haciendo la cimarra y jubilados. Entonces, cómo uno elije la toma de muestras, la toma de los datos, efectivamente puede indicar o guiar la obtención de resultados. Medir las características de un sistema dado lo afecta. Por ejemplo, si uno le quisiera preguntar a una familia ¿Cuánta plata gana?, ¿Qué creen que va a pasar? Si yo soy la oficina de Impuestos, esa familia gana súper poca plata. Y si soy un Banco y esta familia está pidiendo un préstamo, probablemente la familia va a decir que gana mucha plata. O sea, el pedirle o preguntarle directamente a la gente ¿Cuánto gana? ya sea por distintos motivos, por uno u otro, la gente miente, tiende a mentir. Entonces ¿Qué es lo que se puede hacer? Por ejemplo, el Fondo Monetario Internacional o el Banco Mundial hacen encuestas de caracterización económica. ¿Dónde? Siguen a la gente, le siguen los gastos, aproximan el ingreso por la vía de ver cuánto gastan. Llega el momento en que no tienes más crédito, no tienes más plata, entonces metodológicamente hablando eso es un proxy de cuanto ganas.

El Problema con Los Problemas (1)

“Si me dieran una hora para salvar el planeta, gastaría cincuenta y nueve minutos definiendo el problema y un minuto resolviéndolo” Albert Einstein? (Fecha desconocida)

“Dame seis horas para cortar un árbol y gastaré las primeras cuatro afilando el hacha.” Abraham Lincoln? (Fecha desconocida)

La definición de problemas es clave en todo el proceso de la Ciencia de Datos, la siguiente lista muestra los aspectos mínimos que se cree deben ser cubiertos al identificar problemas:

- Hay que establecer la necesidad de solución
- Justificar la necesidad
- Hay que contextualizar el problema
- Escribir la declaración del problema

(57:47) Insisto, esto todo parte del problema. Si el problema está bien definido, el resto fluye más naturalmente. Acá hay dos citas que se le atribuyen, una a Albert Einstein: “Si me dieran una hora para salvar el planeta, gastaría cincuenta y nueve minutos definiendo el problema, y un minuto resolviéndolo” Y en la misma línea, Abraham Lincoln, aparentemente dijo: “Dame seis horas para cortar un árbol y gastaré las primeras cuatro afilando el hacha” Entonces, el saber definir problemas es una cosa clave. Da lo mismo que tan bueno sea yo para hacer magia matemática o magia computacional, porque si no entiendo el problema estoy jodido. Entonces, como es tan clave, para definir problemas, hay que fijarse en cuatro puntos: Primero, hay que **establecer una necesidad de solución**. Hay problemas que si no se necesita resolverlos, es una pérdida de tiempo hacerlo. Y esto

habitualmente pasa en la típica estructura piramidal de compañías u organizaciones donde el jefe dice: “yo creo que esto es un problema” y los de abajo boicotean la solución del problema, porque para ellos no es un problema. Hay que **justificar la necesidad**. Hay que **contextualizar el problema** pues el problema existe dentro de un contexto. Para una empresa u organización lo que es un problema, para otras no lo es. Y finalmente, no menor, **escribir la declaración del problema**, porque esta es la promesa, esto es lo que uno va a entregar. Acá está el contrato entre lo que hay que resolver contra el eventual entregable.

Algunos Mitos Comunes

(59:25) “La Ciencia de Datos es para genios matemáticos o PhD”. La verdad es que el gran requerimiento es saber algo de estadísticas, algo, ni siquiera la parte complicada. O, razonar en términos estadísticos. No es tanto, es la parte conceptual.

“La Ciencia de Datos es sólo aprender herramientas”. No, el científico de datos debe pensar fuera de esquemas tradicionales para obtener soluciones y además debe conocer el negocio. No solamente herramientas. Es súper fácil calentarse con el algoritmo, nombres rimbombantes, con la librería bonita, con la visualización simpática, pero eso son fuegos artificiales, no más. La esencia del cuento, es que uno tiene que pensar el problema, familiarizarse con los datos, y sobre todo, entender el negocio.

“El científico de datos será reemplazado muy pronto por la inteligencia artificial”. Tercer mito. No, a pesar de que las máquinas se llevan el peso más fuerte del trabajo no son capaces de interpretar con criterio o sentido común los resultados. Acá les voy a mencionar algo que mi profesor de física de enseñanza media siempre nos decía: “Si los tontos pudieran volar, todos los días estaría nublado”. Y esto aplica en particular en este sentido. Hay gente que no tiene sentido común. Entonces el sentido común es más importante, y eso no se le puede enseñar a una máquina. La máquina no tiene sentido común, hace lo que el programador o el algoritmo le dice.

“Los científicos de datos trabajan en herramientas sofisticadas todo el tiempo”. No puede ser más lejano a la verdad. A uno le gustaría ocupar herramientas sofisticadas, sí, pero la verdad es que muchas veces la herramienta no hace al científico de datos, sino que es el proceso que uno sigue. Partir de una definición de problema, con una hipótesis, diseñar un experimento, recolectar los datos y dar una respuesta. El proceso de la ciencia, el método científico aplicado a la resolución de problemas, es lo que caracteriza a un científico de datos.

“Los científicos de datos sólo trabajan con volúmenes grandes de información”. Si bien es cierto que los volúmenes grandes de información tienen potencial para reproducir mejor

el modelo, la estrategia para enfrentar un problema no está sólo circunscrita al uso de grandes volúmenes de información. También tenemos velocidad, variedad y veracidad. Estas son las cuatro V que básicamente justifican el uso de ciencia y modelos de datos.

Otro mito más: “si sabes programar puedes ser científico de datos”. Mentira. El científico de datos está en la intersección de tres disciplinas. Conocimientos matemáticos y del negocio también son requeridos, la programación por sí sola no te hace un científico de datos.

“Ciencia de datos e inteligencia de Negocios son lo mismo”. No, este es uno de los mitos más extendidos. A pesar de ciertas similitudes, la inteligencia de negocio se concentra más en los aspectos operacionales y contextuales de una organización. Es más descriptiva y se centra en los proyectos de la audiencia. Por otro lado, la ciencia de datos se concentra mucho más en la analítica predictiva, la identificación de patrones y revelaciones. Es mucho más predictivo. Es un pequeño cambio de foco y yo sé que por ahí de repente en un ratito más nos vamos a enfrascar en una discusión, pero no quiero ir para allá. Desde mi punto de vista la ciencia de datos es el proceso más que el conjunto de herramientas en particular.

“Una mayor cantidad de datos se traduce en mayor precisión”. Más datos, con el análisis erróneo, no implican mejor precisión. Puedo tener un montón de datos y estar aplicando mal la herramienta, y eso me va a dar malos resultados. De hecho hay una ley que se conoce como la Ley del instrumento, que es un dicho muy antiguo: “si uno tiene un martillo, todos los problemas parecen un clavo”. En este caso aplica al estar ocupando el mismo caballito de batalla. Aunque uno se lo aplique a todos los datos del universo, va a producir malos resultados.

(01:03:09) “La ciencia de datos no entrega beneficios económicos”. Ese es un mito. Lo ven como algo científico y la ciencia no tiene habitualmente esta buena reputación comercial. El proceso en sí, no solo permite caracterizar fenómenos de la organización sino que además permite una mejor toma de decisiones, que naturalmente debiese verse reflejada en mayor eficiencia y utilidades.

Finalmente, “los científicos de datos no son científicos en ningún sentido razonable”. Eso también es falso. Los científicos de datos ocupan el método científico en el desarrollo del trabajo investigativo. Lo que pasa es que las herramientas enfrentan la esencia de una manera distinta. Acá ya no es el tipo de “ajustemos el modelo” u “obtenemos la ecuación”. Acá hay un algoritmo, eventualmente lo que uno hace es ajustar un algoritmo, para dar una respuesta a un problema.

¿Ciencia de Datos para Todos?

Con todo lo que se ha discutido podemos darnos cuenta de los siguientes hechos fundamentales, que son un resumen a modo de conclusión de esta presentación

Primero: la ciencia de datos ha llegado para quedarse. En eso, yo creo que estamos todos de acuerdo.

El proceso no puede ser automatizado en un 100%, aún hay elementos de buen juicio que no pueden ser reemplazados por máquinas o algoritmos.

Tercero, se requiere un gran conocimiento del negocio donde se aplicarán los modelos, conocimiento muchas veces basado en la experiencia y difícil de replicar por herramientas automatizadas.

Cuarto, hay herramientas de fácil acceso que pueden ser utilizadas o en su defecto capacidad de desarrollo disponible.

Quinto, muchos modelos que agregan valor no son demasiados complicados. De hecho hay toda una tendencia que apunta al uso de lo que se llama modelos explicables, y eso es un tema de otra presentación completa. Pero para decirlo de manera simple, si usted tiene un problema y hace un modelo de red neuronal, tal vez el modelo va a ser súper preciso, pero les deseo la mejor de las suertes tratando de explicar qué es lo que pasa ahí. Nadie puede hacerlo. Sin embargo, el simple modelo de árboles de decisión, te entrega cosas que te permiten entender y relacionar súper bien con el dominio de aplicación del problema. Entonces, algo más complejo no necesariamente es mejor.

Lo más importante al aplicar los modelos es una buena comprensión conceptual de los mismos, más que las matemáticas detrás, entender qué es lo que uno está haciendo. Y hay un montón de librerías que básicamente encapsulan los detalles matemáticos institucionales. Así que básicamente lo que yo estoy diciendo es: dado que lo más importante es el conocimiento del negocio y dado que hay herramientas por la vía librería, y dado que el sentido común no se puede reemplazar, propongo que lo que uno puede hacer es tomar gente que entienda el negocio y entrenarla.

(01:05:47) Y ahí es donde apareció esta primera experiencia, que la hicimos en la SUSESO. Le pusimos “Aprender Haciendo”. Lo que hicimos básicamente es decir: bueno, si esto tiene apoyo y encapsula los detalles matemáticos, o sea la parte de matemáticas está de alguna forma cubierta y la parte computacional también, por qué no enseñamos a la gente que sí conoce el negocio. Era la mejor forma de desarrollar competencias dentro de la organización. Este tipo de desarrollo de competencias lo hicimos primero en 8 sesiones

de entrenamiento en 2 meses, una por semana. En paralelo y extendiéndose más allá del término de las clases formales, se crearon grupos, de preferencia heterogénea.

Aprender Haciendo

- Basado en lo anteriormente discutido, se cree que aprender haciendo es la mejor forma de desarrollar competencias dentro de las organizaciones
- Un esquema de desarrollo de competencias de este tipo comprende los siguientes elementos:
 - Sesiones de entrenamiento
 - En paralelo y extendiéndose más allá del término de las clases, se crean grupos, de preferencia heterogéneos y/o transversales a la organización
 - Cada grupo definirá un problema, recolectará datos, ajustará modelos y finalmente presentará los resultados de su investigación en un foro público
 - Cada grupo es asesorado por el consultor mediante lo que se denomina "Acompañamiento Experto" para reuniones semanales de avance y apoyo en la programación de soluciones
 - La actividad de cierre cuenta con un jurado conformado por el equipo de liderazgo de la organización quien determina la mejor solución de entre todas las presentadas
- Duración total del proceso: **6 meses en promedio**

Está por ahí Luis Báez, de asistente. Él acuñó el concepto de juntar gente de distintos caminos en la vida, con distintas competencias: uno que sea bueno para computación, otro para esto o esto otro. La armada de equipo, fue lo que hizo que esto funcionara súper bien. Grupos heterogéneos y transversales a la organización. Cada grupo definió un problema, recolectó datos y ajustaron en el fondo el modelo. Y presentaron los resultados, al final, en un foro público.

(01:07:23) Cada grupo es asesorado por el consultor, lo que denominamos "acompañamiento experto", una cosa un poco soberbia de llamarlo así, pero básicamente, el que da las clases tiene reuniones semanales con los grupos, para irlos guiando y apoyando en la implementación de las soluciones. Y la actividad de cierre contó con un jurado. De hecho en el primero que hicimos Adrián Medrano fue parte del jurado y estuvo el Superintendente también, Claudio Reyes. Básicamente fue el equipo de liderazgo de la organización quien determina la mejor solución dentro de todas las presentadas. La duración total del proceso fue de seis meses.

Esto lo he repetido en otros lados, en Perú y en Australia. Lo he hecho dos veces más y ahora estoy con la posibilidad de hacerlo otras dos veces en Perú. Invariablemente observo los siguientes beneficios. Primero: la conformación de grupos heterogéneos permite que áreas que en general son disímiles, puedan interactuar y conocerse. Este es el mejor ejercicio de [Kumbayá](#) que puede tener una organización. A las organizaciones habitualmente les cuesta hacer que gente de distintos departamentos se entiendan. Estos grupos transversales, hacen y permiten que uno pueda entender lo que el otro tipo, que

está sentado en el otro piso, hace. En la práctica, el modelo de aprendizaje se constituye en un proceso de consultoría efectiva. Imagínense lo que es tener a tres, cuatro o cinco grupos resolviendo problemas con la ayuda de un consultor. Permite básicamente tomar problemas que son definidos desde las bases, para los cuales existe sentido de que se puedan resolver y que de hecho no constituyen una amenaza directa para ellos. Porque ellos los tienen identificados y además representan dolores que existen en la organización, lo que remueve un montón de barreras de opción. Parte de todo este problema es la gestión del cambio humano. Uno puede llegar con una solución maravillosa, pero si la organización la rechaza estamos jodidos.

No mucho tiempo después de terminada la actividad comienzan a aparecer los campeones. ¿Quiénes son? Son aquellos que quedaron súper prendidos con los cursos, con la actividad. En el caso de la SUSESO aparecieron varios campeones. Y ellos empiezan a buscar otros problemas, empiezan a estudiar por su cuenta, empiezan a tomar cursos [Edex](#), en línea, y terminan haciendo hartas cosas y se mantiene la relación en el tiempo. Todavía nos siguen preguntando cosas. Incluso un par de ellos, terminaron escribiendo un paper donde estaban tratando de mostrar sus avances y estaban súper entusiasmados. Todavía conversamos a menudo. Algunos se motivan lo suficiente para aprender nuevas herramientas, incluso completar estudios de diplomados o magister. Esto es una increíble oportunidad de desarrollo de capital humano dentro de las organizaciones. O sea, ¿por qué ir donde un consultor y darse toda la lata de explicarle cómo funcionan las cosas, si uno puede empezar a generar soluciones desde la misma organización, al menos para la fruta que está colgando bajo? Ese es el concepto.

(01:10:28) ¿Cuáles son los principales obstáculos de implementación? Porque no todo es un mar de rosas rojas. El primer problema y tal vez el más importante que hay que sortear, tiene que ver con la organización semana a semana, los espacios para el acompañamiento experto. Y acá también está Saby Vargas de SUSESO. Él tenía que enviar las invitaciones, todas las semanas. Lo cual era bastante desgastante. Siempre pasan cosas de última hora, en que entran a otra reunión y es necesario reagendar.

El equipo de liderazgo de la organización debe entender esta actividad como algo que forma parte de la jornada de trabajo y no como algo adicional o que ya se esté haciendo. De lo contrario tiene cero posibilidades que se pueda hacer. El compromiso en el tiempo hace que haya posibilidades reales de que algún estudiante no termine la actividad, algunos se enferman, otros asumen responsabilidades laborales nuevas, a otros simplemente no les interesa. O sea, hay cambios porque es una actividad que dura seis meses.

Y el último problema no menor, tiene que ver con las personalidades de los integrantes de cada grupo. En ocasiones simplemente no es posible que trabajen juntos. Y en esos casos, lo mejor es separar al grupo en subgrupos. Y siempre uno se encuentra con gente super brillante, eficiente, pero que no quieren trabajar con nadie más. Eso está en todos lados, en todas las organizaciones.

Conclusiones

Hay datos en todas partes, primera conclusión. Pero no significa que podamos hacer frente a este océano de datos, todo depende del problema que queremos resolver. El problema es lo más importante. Apretar botones en los paquetes de software no es difícil, sin embargo, entender lo que se está haciendo y las implicancias de los resultados que salen del paquete de software es usualmente no trivial. Es ahí donde se requiere el conocimiento conceptual. La capacidad interpretativa.

Los datos por sí solos no tienen valor alguno, lo que hacemos con ellos es lo que agrega valor. Hay una gran oportunidad que sin embargo requiere información transparente y modelos o herramientas especializadas. El elemento más importante que no puede olvidarse es la gente. Primero: ellos son los que en definitiva ayudan a implementar las mejoras de los procesos y los que pueden boicotear todo el proceso, si es que no están de acuerdo. Ellos son los que entienden el corazón del negocio en profundidad. Ese viejito que lleva 20 años haciendo el trabajo sabe más que el joven que llegó hace un par de años. Y ellos son los que saben dónde está la piedra en el zapato de la organización. Con esto termino.

-Rafael: (01:13:20) Muchas gracias Sebastián y Charango. Increíbles las presentaciones. Tenemos mucha participación tanto en YouTube como en Zoom. Partiremos con Verónica Valle, pidiéndole a Sebastián si puede dimensionar un poco cuáles fueron los resultados para los pacientes y para el Hospital Exequiel González Cortés, si es que existe alguna evaluación sistemática de esta tremenda experiencia que nos relató.

-Sebastián: Gracias por la pregunta. Hay un montón de beneficios, pero lo primero que nos ocurrió cuando estábamos haciendo el piloto, fue muy importante. Cuando uno monitorea en papel, lo que ocurre con los técnicos paramédicos cuando son tantos niños, se cansan como todo ser humano pues el trabajo es non stop, 24 horas. Cuando tenemos episodios críticos en invierno eso colapsa. Ahora se ha visto con el Corona Virus. Ahora la gente entiende un poco mejor lo que es colapsar una urgencia. Entonces, muchas veces se les pasaba el medir a los niños y anotar, lo cual era un error involuntario. O se demoraban más de una hora o dos horas en monitorearlos, cuando se daban cuenta y lo iban a revisar. Eso sumado a que los registros estaban en papel y era difícil de reconstruir el

estado. Ello generaba un riesgo en los niños. Entonces, de hecho el proyecto se llama Almohadita-Ámbar porque una bebita falleció producto de que no se dieron cuenta a tiempo que se estaba poniendo grave. Y en las auditorias se dieron cuenta de que eso lo podrían haber evitado, sin que la hubieran cambiado de unidad de lactantes a urgencia, a tiempo. Al implementar este proyecto ha permitido que eso no pase de nuevo. Que no pase que se te olvide monitorear a los niños. ¿Por qué? Porque el sistema avisa. Y los niños que tenían que ser chequeados a las 15:00 si no los chequeaste en 10 minutos suben de prioridad y empiezan a hacer un parpadeo de pantalla, y hay un responsable de eso. Entonces uno sabe que se te pasó, por último uno dice: “no puedo ahora, estoy haciendo otra cosa” y llamas al colega para que lo vaya a monitorear por ti. Entonces, ha permitido que no se pasen los chequeos, lo que aumenta la seguridad. Al año usualmente hacían paro cardíaco de tres a cuatro niños en la unidad. En el año que nosotros lo vimos funcionando no hizo paro ninguno. Entonces, ahora es una anomalía, porque son tres a cuatro casos, pero eso pasa porque lograron cambiar de unidad a tiempo a los niños, por las alarmas que genera el sistema. El sistema no toma decisiones solo, el sistema simplemente muestra un estado de alerta y lo que pasa es a criterio del técnico paramédico, de la enfermera, decir: ¿Sabes qué? Yo le tomé los signos vitales. Y mentalmente ellos tienen unas reglas y el sistema lo que hace es confirmar el estado que la persona tiene en la cabeza, lo apoya. Por eso se llama un sistema de apoyo a la toma de decisiones, con un algoritmo, esa es la verdad. Entonces, cuando el modelo marca un riesgo alto y la persona que está midiendo considera que para él no es tan alto, entonces lo que hace es pedir opinión, re chequear los signos vitales y prestarle más atención. Entonces, ese beneficio también está implícito ahí. Medirlo es muy difícil, porque los parámetros nuestros son que los niños no hagan un paro cardíaco en la unidad. Pero son pocos casos, dos o tres. Pero todo eso está ocurriendo ahora y no ocurría antes. Y así hay un montón de otros beneficios: se puede hacer mejor gestión de los mismos técnicos paramédicos, se pueden generar mejor los turnos, un montón de mejoramientos operacionales, sólo a partir de digitalizar la ficha y generar los modelos.

-Rafael: (01:18:13) Muchas gracias Sebastián. Tenemos a Álvaro Villablanca que está preguntando desde YouTube, respecto a la posibilidad de aplicación de Ciencia de datos en la Tesorería General de la República, en el área de cobranzas, porque tienen mucha información y quieren cambiar una estrategia de análisis caso a caso, por una estrategia de análisis basado en data. Y en la misma línea, está Paola Riveros dentro de nuestro grupo Zoom, para los efectos de entender esto como un cambio cultural. Si quieres nos comentas y pedimos a Charango responder.

-Paola Riveros: Hola buenos días. Sí a mí me motivó cuando escuché a Charango. Bueno, en la Universidad de Talca hace dos años el Vicerrector nuevo que llegó implementó la

plataforma SAP. Nosotros trabajábamos con un workflow, que era como un traje a la medida de la cultura de la Universidad de Talca, que es una Universidad regional, que funciona de una forma más lento a nivel de procesos. Pero al instalar SAP se hizo una inducción de que nos iba a arreglar la vida. Pero resulta que SAP no es fácil de manejar, menos para un funcionario administrativo que a lo mejor no ha tenido una formación de otro tipo que no sea administrativa. A veces se han formado en la Universidad. Entonces fue crítico porque hubo muchos errores, hubo una pérdida de dinero garrafal. De hecho en ese tiempo quien era Vicerrector era Raúl Salgado y le pusieron “TranSalgado” al proceso. Te explicaban una cosa, después llegaba una inducción. Al final, después de casi un año tuvimos una inducción real los funcionarios que usábamos los módulos, porque además te enseñaban módulos que a lo mejor ni usabas. Y eran tan complejos. Yo de hecho tuve el módulo de MM y yo no lo uso. Yo uso el módulo de contratos de recursos humanos, porque yo trabajo en el área de generar contratos honorarios a los profesores. Y hago seguimiento de análisis de datos de todos los académicos. Y fue crítico, porque yo tenía colegas de edad muy resistentes al cambio. Para mí, no hubo un acompañamiento, una buena planificación de un acompañamiento y me hizo mucho sentido cuando se habló de estas reuniones, de estas capacitaciones heterogéneas de varias áreas que involucraba a las jefaturas. Lamentablemente en la institución siempre hay capacitaciones donde todos tenemos que aprender, desde la jefatura. Pero las jefaturas, no sé si por egos académicos, no sé la verdad, simplemente no participan. A ellos tienen que hacerles la inducción aparte. Entonces ahí tú cortas un poco la forma del proceso porque la jefatura tiene que saber cómo funciona en general, no digo que en el detalle técnico...

-Rafael: (01:22:07) Gracias Paola. Está claro el punto. Le pasamos a Charango para hacer un abordaje a nuevas entidades que tengan este tipo de problemáticas.

-Charango: Se ha hablado harto de la SUSESO, tal vez va a ser el ejemplo de libros por años, en temas de cómo se implementó este cambio. Yo llegué al final del proceso. La Superintendencia de Seguridad Social tenía todas las semanas reuniones, Rodrigo Moya, con los doctores, con los abogados, manejaron el proceso de tal forma y escucharon a la gente antes de imponer cosas. La preocupación por el usuario final es ciertamente un estilo de gerenciamiento que esa organización tuvo, que no lo he visto en otros lados. He visto fracasos similares al que menciona de la Universidad de Talca. Sobre todo porque llega alguien y dice el tema del martillo. “Este es el martillo que tengo y todos los problemas parecen un clavo”, porque es el martillo que sé ocupar. Empieza a imponer herramientas y se olvidan que al final el éxito de implementación de esas herramientas es llamar a la gente. Ese es un tema súper personal, no puedes obligar a alguien por decreto a que trate de ser empático con la gente que trabaja con él o para él, también depende de cuál es la visión de la gente. Entonces, el tema acá del cambio cultural, la gestión del

cambio, es la parte totalmente horrorosa de esta historia, es lo que puede hacer el gran éxito o el gran problema. Si uno comprende al usuario final, muy probablemente va a ser capaz de representar sus deseos y sus necesidades. Pero es fácil meter las patas sobre todo si es que uno no ha estado haciendo el trabajo. Si uno es consultor, por muy experto que sea, si no ha tenido que llenar la ficha por tres o cuatro años, te cuesta dimensionar que aunque lo pases a versión electrónica es la misma ficha. Entonces, por ahí, la comprensión profunda del negocio es la que hace que las cosas funcionen bien. Y en el caso particular de la SUSESO ellos se hicieron parte y gestionan desde adentro, con apoyo de consultores externos. Pero ellos identificaron la necesidad y desde dentro se gestiona el cambio. Entonces, es un caso bastante único, porque en otros casos ciertamente no les funcionaba igual, y bueno, paciencia Paola, qué más te puedo decir.

-Rafael: (01:24:35) Sebastián, pensando que debemos implementar la ley de transformación digital en toda la administración pública y particularmente en el mundo de la salud, hay distintas consultas respecto a costo-implementación para otro hospital, más allá del Exequiel, de este modelo de cambio de trabajo en papel a un trabajo digital, con aplicaciones en distintas áreas e incluso con monitoreo de enfermos crónicos a través de la red de atención primaria. ¿Cómo ves tú y qué recomendaciones de política pública harías en esa línea?

-Sebastián: Es complicado, porque las políticas públicas no son mi tema. Eso se lo dejo a José, pero si uno habla de lo que yo he visto, la realidad de los CESFAM y de los hospitales, hay disparidad de grados tecnológicos en todos ellos. Algunos usan ficha médica electrónica. Fichas electrónicas además hay de muchas empresas. En el Exequiel nosotros ayudamos a construir la ficha médica electrónica también, partimos construyendo la ficha del Exequiel hace como 10 u 11 años. Y gracias a eso nos pudimos conectar en las almohaditas. Como universidad hemos ido a ayudar a construir softwares, a costa de algunos tesisistas que han hecho los pilotos. Entonces hay distintas formas de abordarlos, pero ciertamente hay que hacer las inversiones, hay que destinar los recursos para poder hacerlo. Y hay que hacerlo cuanto antes, porque realmente los beneficios son muy grandes, pero yo creo que hay que buscar hacerlo de forma correcta. Creo que no hay que desesperarse. De pronto se invita a tremendas empresas que llevan un montón de software y a veces es mejor ir más despacio. Hay productos locales que son súper buenos, pero yo creo que nada de eso va a funcionar si desde el gobierno central no se fija una estrategia única, un modelo de cómo se van a estructurar los datos o qué es un protocolo de almacenamiento de los datos. Entonces, independiente de la empresa que yo use, el formato de los datos va a ser el mismo y vamos a poder interoperar sin ningún problema. O el tema de cómo vamos a hacer el almacenamiento de los resultados de los modelos, para hacerlos auditables. Lo que decía Charango, que es importante poder auditar los

modelos. O la misma ficha médica electrónica, las prescripciones médicas. Yo tengo un proyecto para analizar los riesgos de medicación en las recetas médicas. Entonces claro, me encuentro con las fichas en papel, con las fichas en Excel, con las fichas en sistema, está todo distribuido. Entonces, si desde el gobierno del sistema central no se fija una estructura, porque no puede ser por ley, por encima. Tiene que dar una bajada, generar una especie de librería. Éstas van a ser las formas de comunicación. Puede haber un equipo de ingenieros de software fijando estándares y protocolos. Y ni siquiera tenemos que inventarlos, hay un montón de protocolos para medicina, salud, ya hechos. Y decir: nos vamos a ir por este protocolo, por este mecanismo y de ahí para adelante normar la actividad de desarrollo. Pero si eso no aparece, no ocurre, lo que va a pasar es que vamos a quedar con esa imagen que les mostré de muchos sistemas distribuidos, por todos lados con data desactualizada. Y es así como vamos a gastar miles de millones de dólares, como lo hemos hecho hasta ahora y no vamos a resolver el problema. Tal vez hasta lo agrandemos.

-Charango: ¿Puedo agregar una cosita? La ley viene con un presupuesto que es súper ratón, perdonen que lo diga así. Pero cuenta con 5 millones de dólares y en uno de estos foros el ex Superintendente de Seguridad Social mencionó que el puro proceso de transformación digital de la SUSESO fue del orden de 8 millones de dólares. Entonces, para entender la naturaleza del problema, el crear una ley no es nada. Y cuando hay que financiar, y ni siquiera estamos hablando de que esto sea horrorosamente caro, sino de que son soluciones que afectan la vida de todos los ciudadanos del país, entonces es como poner las cosas dentro de contexto. Creo que está súper mal calculado, es mi opinión personal.

-Rafael: (01:29:38) Charango, y en la línea de lo que podemos colaborar en las capas medias de las organizaciones para la mejora de los datos, la curatoría de la que tú hablabas. En el caso específico de SUSESO entendemos que hicieron un trabajo muy sistemático, tú lo comentabas de esta formación y auto aprendizaje del equipo durante seis meses. ¿Eso está documentado, lo pueden compartir? Hay preguntas en esa línea.

-Charango: No lo tengo documentado. O sea, preparé el material para las clases, se hizo el proceso, se hizo la actividad, algunos de ustedes participaron. Adrián me acuerdo que estaba ahí como jurado, estaba el Superintendente también, estaba el Fiscal, entre nosotros escogimos quien se llevaba el premio. Pero la verdad fue un poco un experimento porque apostamos a que la gente podía aprender. Pero no es algo que nosotros supiéramos que se hubiese hecho. Entonces nos lanzamos para adelante para ver hasta dónde íbamos a llegar y afortunadamente llegamos bien y llegamos al final, con resultados bastante mejores de los que pensábamos. El grupo que ganó, la calidad de su

trabajo era bastante bueno, publicable. Entonces el tema de la capacitación es algo que, tal y como lo veo, si bien hay problemas que son muy difíciles de resolver, creo que en esos problemas Sebastián entra o puedo entrar yo, pueden entrar consultores, agentes expertos para resolver esos problemas difíciles. Pero hay un montón de cosas fáciles, fruta que está colgando abajo, cosas chicas, que se podrían hacer desde la misma organización y evitarse en el fondo todo el proceso largo. Entonces, el tema de capacitación lo ven como un costo cuando en realidad es una inversión que tiene un retorno positivo en el tiempo.

-Rafael: Sí, no nos cabe la menor duda, y en esa línea, Sebastián, Pilar Díaz nos plantea la posibilidad de aplicar Ciencias de Datos a estudios históricos, a partir de la data contenida en el Archivo Nacional, que ahora no sólo es física, sino también digital, no sé si Pilar nos quieres complementar la pregunta.

-Pilar: Buenos días, antes que todo, felicitarlos por la presentación. En realidad lo mío era una reflexión más bien, porque soy muy novata en estos temas. Pero veo que en los documentos electrónicos que se conservarán en el Archivo Nacional, a partir del próximo año en adelante, a corto o a mediano plazo ya van a ser archivos históricos, van a ser una fuente importante de información. Yo me imagino solamente, si pensamos así a largo plazo en los expedientes judiciales que pueden estar en los archivos nacionales, que estuvieran en formato electrónico. Toda la data importante que se puede hacer para generar análisis, pensando en futuras reformas judiciales de proceso, etc. etc. Y eso extrapolarlo a los diferentes ámbitos de la administración pública. Creo que se abre un panorama muy interesante. Y en realidad veo que hay historiadores que van a buscar documentos físicos, siempre se habla del testamento de la Quintrala y ese tipo de documentos más antiguos. Yo me imagino que en 10 o 15 años más habrá investigadores de diferentes ámbitos, pero que con manejo de datos, que puedan sacar una información muy relevante para una toma de decisiones, no necesariamente estudios históricos, sino que una toma de decisiones importante para generar políticas públicas, revisar como ha sido la evolución de la institución y la toma de decisiones en el pasado, para tomar decisiones en un momento determinado. Solamente lo planteo, como les digo. Yo soy novata, estoy maravillada con el mundo de los datos. Creo que con el mundo de los datos y los archivos podríamos generar unas muy buenas alianzas para la administración pública especialmente. Entonces, era más que todo un comentario. Pero yo en realidad, ni siquiera me atrevo a preguntar.

-Sebastián: (01:34:02) Hay un montón de información que uno puede sacar a partir de los documentos que están escritos. Por ejemplo, tengo una publicación reciente donde procesamos todas las publicaciones científicas médicas de Chile, desde parte del año

2014, 2015 y 2016, con minería de textos, que son técnicas de algoritmos para procesar textos. Y por ejemplo, descubrimos que el 86% de las investigaciones que estaban financiadas por algún fondo estatal estaban concentradas en sólo cinco áreas o problemas médicos y el resto había poquito. Entonces la pregunta es: tal vez debiéramos crear una política para poder descentralizar y si los gastos se concentran sólo en ciertas áreas preguntarnos: ¿estamos gastando bien? Por lo menos, ahora con estas técnicas podríamos monitorear, revisar qué es lo que está pasando en ese aspecto. Pero claro, en los textos escritos uno puede hacer descubrimientos. Trabajamos con SERNAC, los años 2011 y 2012 procesando los reclamos. Y eran 350 mil reclamos en un sólo año. Y entonces detectamos si había problemas que fueran comunes para la ciudadanía. Cuando uno reclama por internet o por teléfono la persona que te atiende clasifica en algunas categorías de problema, pero son finitas, son 7 u 8 categorías, no es mucho. Pero claro, la cantidad de problemas que hay son infinitos, entonces esas categorías normalmente agrupan demasiados reclamos y los analistas no pueden revisar esos 50 o 60 mil reclamos de forma rápida, para tomar una decisión. Entonces, por ejemplo, corremos los algoritmos y detectamos varias familias de reclamos y, por ejemplo, salía un clúster con una matemática específica que era el tema de La Polar y uno veía como había un pick en enero y febrero del año, parece 2011, y luego bajaba. Entonces si uno se empieza a fijar en las puntas, uno decía: “aquí hay mucha gente reclamando de algo súper específico, debo revisar”. Y lo que hacía SERNAC era organizar demandas colectivas. En este caso amerita que SERNAC intervenga y haga algo más proactivo y coordinado. Estas herramientas les dieron esa visión que no tenían antes. Entonces claro, lo que uno puede hacer es infinito.

-Pilar Díaz: (01:36:56) Solamente un ejemplo y saliendo un poco del ámbito del Archivo Nacional. Por ejemplo, toda la correspondencia que recibe la Presidencia de la República. Imagínate lo que significa para un gobierno, cualquiera sea el signo, tener esa información, esos datos trabajados, analizados en un momento oportuno. Porque pensemos que la correspondencia que recibe el Presidente ha llegado a ser 600 mil, y todo se correlaciona con la situación política del país. O sea, mientras más álgido sea el momento, más escriben. Y a la vez con ciertas solicitudes de habitación, educación, lo que sea. La riqueza que esa información le puede dar pistas al grupo que está gobernando para tomar decisiones en un momento determinado. Realmente esto abre un panorama que debiese ser utilizado, claro, cuidando el tema de datos personales, como corresponde, pero realmente es una maravilla.

-Rafael: En la misma línea, Sebastián, pregunta Ye-seng Pineda, si el caso del hospital de niños lo piensan replicar en regiones para que no sea sólo una acción centralizada.

-Sebastián: Claro, “Almohadita” es un proyecto de investigación de CONYCID, hoy la [ANID](#). Con esos recursos pudimos construirlo, porque la Directora del hospital nos tenía fe, sabía que nosotros íbamos a hacer algo concreto. Y lo que quedó es público. Obviamente el software que se usa en el Exequiel es de allí y ahí está funcionando, pero todo el conocimiento nuestro, cuando nos invitan de otros hospitales lo damos a conocer en forma gratuita, les decimos lo que tienen que hacer, les mostramos los algoritmos, las metodologías. Y lo que hay que hacer es recalibrar los modelos con la data de cada hospital. Y claro, ahí nosotros les pasamos el conocimiento a las personas del área TI que hay en el hospital. Obviamente, llegar a implementarlo propiamente tal va a requerir inversión, va a requerir un proyecto, que haya un jefe de proyecto. Ahí probablemente cada hospital tiene que hacer sus inversiones, pero al menos la participación nuestra para poder hacer que salga, siempre ha sido gratis.

-Charango: (01:39:42) Quisiera agregar una cosita. me quedé pegado con lo que estaban discutiendo recién y creo que hay que notar algo que es importante. Parte de la cosa educacional. Los algoritmos de procesamiento de lenguaje natural no comprenden, son algoritmos bastante tontos, solamente generan descripciones estadísticas de lo que los textos contienen. Entonces el caso de uso que tú mencionabas Pilar del historiador que quiere ver el testamento de la Quintrala, esa comprensión de lo que dice el texto, solamente lo puede hacer un humano. Tal vez el computador lo que más puede decir es este testamento se parece a todos estos otros testamentos que tengo en mi base de datos. Entonces, hay que tener esa precaución, porque o sino uno puede generar tal vez la falsa expectativa de que básicamente uno le inyecta documentos a este cerebro artificial que me va a responder preguntas, cuando la verdad no es así. El computador puede almacenar hechos, puede identificar secciones de los documentos, pero hasta por ahí no más llega.

-Pilar Díaz: Sí, más que todo me refería al símil del historiador más tradicional de hoy. En 20 años más va a ser alguien que interprete datos, que pueda hacer una revisión y que a la vez pueda dar otros usos a los documentos electrónicos. Y fíjate que sería ideal, pero ni siquiera estoy pensando en documentos digitalizados, que fueron generados en papel, porque ese ya es otro tema, con otro presupuesto. Sino que del próximo año en adelante, cuando el Archivo Nacional empiece a recibir documentos electrónicos que ya vengan con metadatos con cierta información estandarizada que se le pueda sacar para analizar grandes grupos de documentos, para sacar ciertas conclusiones, ciertos análisis, ciertas tendencias, que creo que serían buenos insumos para políticas públicas a futuro.

-Roxana: Dos temas distintos. Hace algunos años, con esta inquietud de extraer datos de documentos, que es una cosa muy compleja, hicimos un experimento con un grupo de

Wikipedistas, de ingresar a Wikisource, que es uno de los tantos proyectos de la Wiki, los cuerpos legislativos que se habían producido desde el primer gobierno de O'Higgins, hasta 1870 más o menos. ¿Cuál era nuestro experimento? Dado que en los cuerpos legislativos hay muchos datos, hasta la rendición de cuentas de O'Higgins en un viaje a Chillán, donde había gastado tanta plata y rendía. Es real, les puedo enviar el link del proceso, está ahí, es una cosa que puede ser sumamente interesante. Y ¿qué hicimos con los Wikipedistas? Contratamos un grupo de historiadores y desarrollamos un conjunto de marcas en que ellos iban identificando cuáles eran datos importantes. Porque es interesante saber cuánto costaba el tabaco, el alcohol y como fue evolucionando eso. Y llegamos a bastante, porque la idea era después construir un motor que leyera las marcas. Entonces yo quería saber si para este tipo de datos, el motor iba a extraer. No me iba a traer todo el documento, me iba a traer sólo la marca, pero ahí quedó. No lo pude terminar y yo creo que ahora, con herramientas actuales se podría hacer algo bastante más interesante, podríamos usar micro etiquetas de los documentos, ya que van a ser todos electrónicos, para poder identificar data, ese es una idea no más.

Y lo segundo es una pregunta que le quería hacer a Charango. En gestión del conocimiento se dice que uno no sabe lo que sabe. Tu conocimiento se va instalando a través de tu trayectoria, tus estudios, y voy a poner dos ejemplos que pueden parecer muy elementales. El típico caso es el del mecánico que escucha el ruido del motor y sabe que hay un problema y sabe cuál es el problema. Y otro muy curioso, que vi un día en mantención del Metro, en que con un palo golpeaban los neumáticos y sabían si estaban bien o no. Y esas cosas hay que enseñarlas a otras personas. Entonces ¿cómo te transfiero mi oído y te enseñó a escuchar, cómo te enseñó algo? La pregunta de fondo es en este proceso de [Maker movement](#) que tú haces ¿es posible extraer ese conocimiento y transmitirlo y sistematizarlo para otras personas?

-Charango: (01:45:22) Parte de uno de los fundamentos filosóficos del Maker Movement, es compartir, está articulado en torno a la enseñanza al compartir, a la cosa más abierta. Entonces, el compartir es súper bueno, pero dentro de esta historia hay que reconocer que hay gente que está dedicada a la consultoría y que lo que quiere hacer es ganar plata y que lo que menos quiere hacer es compartir. Pero el equipo interno se puede implementar y de todas maneras debiese pasar. La tragedia griega acá es cuando alguien ha estado trabajando 20 años en una compañía y justo tuvo un accidente, esa persona se enferma, desaparece por seis meses, y queda el desastre. Y eso pasa todo el tiempo.

-Roxana: Se fuga el conocimiento.

-Charango: Y es irremplazable. Entonces, ahí es donde el tema del compartir frecuentemente y el hacer que nadie sea insustituible, que todos sean parte del equipo,

que colaboren como equipo, es un tema de manejo de personas, pero un tema que tiene que ver con la cultura organizacional. Y eso es tal vez fuera de la piscina donde a mí me toca nadar, pero entiendo que tiene una dimensión. Si yo soy re malo para manejar gente, todo lo contrario, me mando condoros todo el tiempo, pero entiendo que hay gente que tiene las habilidades para entender, canalizar y generar sinergia entre equipos diversos e interdisciplinarios. Y eso es algo a lo que en Chile no se le pone en general mucha atención. Mi percepción de Chile, es que casi uno debiera estar agradecido de que tiene trabajo, es así como a uno lo tratan, cuando de verdad es al revés. La institución debiese estar feliz de que uno pueda colaborar. Vivir en el extranjero te entrega una visión distinta. Parte de la discusión de hoy, que tiene que ver con el tema de los datos, de unas API, de unos servicios REST, dependiendo del país dónde uno esté, todo lo que se genera con plata del contribuyente australiano, es público, y se expone, porque está pagado por todos, está pagado con impuestos de todos. Y creo que ahí Chile está un poco al debe, todavía tiene estos silos funcionales, donde la gente de Impuestos Internos es el dueño y el carcelero de sus datos, la gente de Registro Civil lo mismo. Cada institución tiene sus datos y en el fondo cree que al compartirlo va a dejar de ser fundamental, o sea, trata de entender el pequeño pedazo de peñasco. Y la información no sirve de nada, sólo hay que usarla bien.

-Rafael: (01:48:13) Bueno, le damos las gracias particularmente a Sebastián y a Charango. Esta visión de compartir frecuentemente es lo que transformamos en evidencia, en Transformación Pública. Hemos participado de 15 coloquios en el 2020, hemos contado con 24 relatores que hoy integran nuestra galería de la fama. Y los 7 coordinadores: Roxana, Macarena, José, Adrián, Paloma, Andrés y quien les habla, estamos más que orgullosos de este esfuerzo conjunto que hemos hecho. El compromiso de los relatores y de los coordinadores es estar disponibles para colaborar en distintas iniciativas de la red y hoy eso se está dando en el día a día. Estamos muy contentos de que más de 600 personas han participado este año, y las presentaciones de YouTube tienen cientos de visitas.

Tenemos muchas propuestas para el 2021 y en particular queremos focalizarnos en la implementación de la gestión documental y en decisiones basadas en datos, particularmente en las Superintendencias, en aquellas instituciones que tienen un impacto sobre el resto de la institucionalidad pública. Nuestro compromiso es colaborar en una buena implementación de la Ley de transformación digital. Insistimos en hacer la consulta pública, se hizo y participamos de ella. Y además, como Centro de Sistemas Públicos de Ingeniería Industrial de la Universidad de Chile, estamos colaborando con muchas instituciones públicas en Chile y otros países de América Latina y vía la Agencia de Cooperación Internacional, incluso llegamos a países africanos. Nos proponemos entonces

en 2021 iniciar este trabajo colaborativo interinstitucional a la brevedad. Contamos con la colaboración de todos ustedes.

Y una pregunta final para José: ¿Qué opinas del [decreto de gradualidad](#) que el Presidente firmó ayer?

-José Inostroza: (01:50:53) Tal vez nos salimos un poco del tema en específico. Sobre el decreto, no sé si alguien lo conoce, yo no lo conozco, pero no hay mucho que opinar. Si uno no lo conoce uno no puede opinar, así que no sé qué más decir. Lo que sí, como aprendizaje comunitario, estos son esfuerzos mayores. Lo ideal es esperar más tiempo para hacer una discusión más profunda sobre esa regulación. Se hizo esta consulta, lamentablemente no fue sobre el reglamento, sino sobre ciertas ideas generales sobre el tema y el reglamento salió sin haber sido consultado. Yo creo que estamos al debe como país en esa materia, porque la transformación digital es un proceso mayúsculo, importante, tiene que ser articulando muchas voluntades y capacidades de mucha gente, no solamente de un grupo pequeño. Y creo que en ese sentido el proceso podría haber sido mejor y más participativo. Simplemente destacar, para conectar y tratar de concentrarnos en los temas, porque esto obviamente es muy disperso o podría ser muy disperso, lo que dice Sebastián y lo que dice Charango: necesitamos una capacidad de normalización en materia de datos en el país. Es muy importante. Sin eso vamos a estar limitados respecto de las posibilidades, de las soluciones y de estas metodologías. En Chile en este momento no tenemos una autoridad suficientemente institucionalizada, con capacidad para empujar esa normalización. Yo ahí simplemente invito a la reflexión: si tal vez el INE, dado que tiene ciertas facultades legales muy importantes en términos de generar instrucciones, que tiene una institucionalidad que genera respeto desde el punto de vista del buen cuidado de los datos personales, y que en general en el mundo los INE cumplen una función muy importante. El caso de Nueva Zelanda en esta materia es bien notable. Tal vez debiéramos como país fortalecerlo, a efecto de generar estas normalizaciones. Eso es distinto a la interoperación de datos que es otra temática, donde tampoco tenemos una institucionalidad suficientemente fuerte. Pero en la línea de ciencias de datos, es decir de disponer de muchos datos bien curados, como plantea Charango, y que sea un sistema bien rutinario de acumulación de datos, en eso me parece que el INE podría ser una institución importante. Lo dejo como reflexión, pero claramente hay muchas oportunidades. No sé si tenemos a tantos Sebastián Ríos, o Charango Munizaga, pero hay gente, hay mucha inteligencia en Chile, hay Adrián Medrano, Claudio Reyes, tenemos gente. Lo que nos falta, como decían los expositores, son los sistemas más normalizados y eso sí tiene que ver con política pública, y ahí tenemos un desafío importante. Muchas gracias por sus presentaciones, lo que más destaque es su motivación. Yo creo que hay mucha gente en Chile con muchas capacidades, lo que nos falta de

repente es más motivación, más coraje. Tenemos más gente y tenemos que fomentar eso, hay que ir a golpear las puertas de las universidades, hay que ir a buscar a la gente, pero se pueden hacer muchas cosas. Y con la Ley de transformación digital yo creo que vamos a tener más oportunidades, porque la digitalización en general va a estar mucho más disponible.

-Rafael: (01:54:50) Nos despedimos hasta marzo del próximo año. Un gran abrazo a todos.