

Implementing an agile lifecycle model for Data Science project.

Roberto Ernesto Jovel Barrera, Rodolfo Francisco Toche Hernandez

Maestria en Arquitectura de Software, Universidad Don Bosco

Antiguo Cuscatlan, El Salvador

Mayo de 2023

rej.barrera@gmail.com

esfrantoche@gmail.com

Abstract — This document shows an investigation to determine the efficiency or benefits of implementing an agile model according to data science projects, helping to companies or departments focused on information technologies to improve the project management processes in this subject. The investigation performed a comparative analysis with three agile frameworks designed for data science, due to it fits better for SCRUM methodology Data Driven Scrum (DDS) was chosen; nevertheless, to measure the benefits when the proper method is applied three criteria have been considered: solution quality, team autonomy and customer satisfaction after applying the model into a company specialized in data science solutions. Basically, the strategy of implementing provided to the team the mechanisms of functionality of this model, knowing previously the current situation of how the company managed the projects in order to measure the three metrics mentioned before. As a general conclusion, it was verified that there is a support considerable substantial to the software development teams, an improvement in the quality of delivered product that the companies with this type of projects deliver to the customer and finally keep the customer experience.

Keywords; *data science, scrum, agil methodology, project management, structured data, data driven scrum, DDS.*

I. INTRODUCCIÓN

La administración de proyecto actualmente se ha enfocado en modelos de desarrollo de software por medio de metodologías que contemplan una ejecución de procesos diferentes a los necesarios para los campos del Data Science y los requerimientos corporativos que maximicen el valor de sus productos o servicios. Cada parte del ciclo de vida de un proyecto de Data Science suele ser un experimento. Todo proyecto de Data Science comienza definiendo el planteamiento del problema, para tener una idea clara de lo que estamos tratando de resolver. Esto también incluye el establecimiento de los criterios de éxito. Puede tratarse de una métrica comercial, por ejemplo, el aumento de las ventas, o de una métrica de rendimiento del modelo.

II. DESARROLLO DE CONTENIDOS

En la actualidad las compañías salvadoreñas utilizan la información recolectada de sus clientes para construir recursos estadísticos que les permitan tomar decisiones que potencian su desarrollo en el mercado, mantener en el negocio e incluso posibles aperturas de negocios. Esto ha empujado a las áreas de tecnología de las empresas que incluyan en sus proyectos un área de la tecnología conocida como Ciencia de Datos.

A. *¿Que son los datos?*

Los datos se crean constantemente^[3] por diferentes recursos tecnológicos como teléfonos móviles, redes sociales, tecnologías de imagen para determinar un diagnóstico

médico, sistemas de facturación; todo esto y más crean nuevos datos, y deben almacenarse en algún lugar para algún propósito. Los dispositivos y sensores generan automáticamente información de diagnóstico que debe ser almacenado y procesado en tiempo real. El almacenamiento, así como la estructura de datos son importantes para la transformación de lotes de datos para identificar patrones significativos y extraer información útil.

B. Estructura de datos

Los datos masivos pueden presentarse en múltiples formas, incluidos datos estructurados y no estructurados^[4], como datos financieros, archivos de texto, archivos multimedia y mapeos genéticos. A diferencia de gran parte del análisis de datos tradicional realizado por las organizaciones, la mayor parte de Big Data es de naturaleza no estructurada o semiestructurada, lo que requiere diferentes técnicas y herramientas para procesar y analizar. Los entornos informáticos distribuidos y las arquitecturas de procesamiento paralelo masivo (MPP) que permiten la ingesta y el análisis de datos en paralelo son el enfoque preferido para procesar datos muy complejos.

Los cuatro tipos principales de estructuras de datos:

- Datos estructurados: datos que contienen un tipo, formato y estructura de datos definidos (es decir, datos de transacciones, cubos de datos de procesamiento analítico en línea [OLAP], RDBMS tradicional, archivos CSV e incluso hojas de cálculo simples).
- Datos semiestructurados: archivos de datos textuales con un patrón discernible que permite el análisis (como los archivos de datos del lenguaje de marcado extensible [XML] que se describen por sí mismos y están definidos por un esquema XML).
- Datos casi estructurados: datos textuales con formatos de datos erráticos que se pueden formatear con esfuerzo, herramientas y tiempo (por

ejemplo, datos de flujo de clics web que pueden contener inconsistencias en los valores de datos y formatos).

- Datos no estructurados: datos que no tienen una estructura inherente, que pueden incluir documentos de texto, PDF, imágenes y videos.

C. Arquitectura Analítica Tipica de un proyecto de Ciencia de Datos

Los proyectos de Ciencia de Datos necesitan espacios de trabajo diseñados específicamente para experimentar con datos, con arquitecturas de datos flexibles y ágiles. La mayoría de las organizaciones todavía tienen almacenes de datos que brindan un soporte excelente para los informes tradicionales y las actividades de análisis de datos simples, pero desafortunadamente tienen más dificultades para respaldar análisis más sólidos.

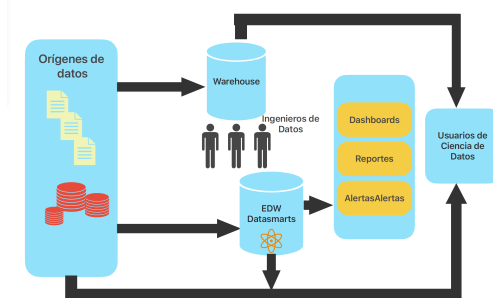


Figura 1. Descripción de la arquitectura analítica en la ciencia de datos

Aunque la Ciencia de Datos y la ingeniería de software son campos diferentes, muchas organizaciones las tratan de la misma manera. Por lo tanto, no es sorprendente que muchas organizaciones también estén impulsando la adopción de Agile en la Ciencia de Datos. Sin embargo, los resultados son mezclados, a menudo debido a intentos fallidos de aplicar prácticas de software Agile a la Ciencia de Datos.

Como resultado de una investigación comparativa con tres principales frameworks ágiles para la administración de proyectos de ciencia de datos se considera que Data Driven Scrum(DDS) cumple con más afinidad a un ciclo de vida de proyectos administrados por una metodología inapropiada.

Tabla 1

Comparación en base a pros y contra de cada framework y las herramientas compatibles con la metodología

Metodología	Pros	Contras	Herramientas
Data Driven Scrum (DDS)	<ul style="list-style-type: none"> Iteraciones basadas en capacidad. Se ajusta con organizaciones SCRUM. Flexible para varios ciclos de vida. Revisiones y retrospectivas en cada iteración. 	<ul style="list-style-type: none"> No es comprensible, es necesario un entrenamiento. Cuando un recurso se libera antes de tiempo debe esperar próxima iteración para tomar una asignación si el backlog está vacío. 	<ul style="list-style-type: none"> Jira. Rally. Herramientas que soporten Kanban.
Crisp-DM	<ul style="list-style-type: none"> Generalizable. Perceptible o intuitivo. Fácil de utilizar. De inicio confiable. Resultados seguros. Flexible. 	<ul style="list-style-type: none"> Demasiada documentación. Reuniones generales difíciles de dominar. 	<ul style="list-style-type: none"> Monday.com. Wrike. Project Manager. Jira
Team Data Science Process (Microsoft (DSP))	<ul style="list-style-type: none"> Comprensible Inclusión opcional de Scrum Mantenibilidad 	<ul style="list-style-type: none"> Ciertas inconsistencias. Empinada ruta de aprendizaje. Aspectos específicos de Microsoft 	<ul style="list-style-type: none"> GitHub Azure Pipelines Azure Boards Azure Monitor Visual Studio AKS

Nota: Información obtenida de los sitios web oficiales de las metodologías.

D) Introducción al agilismo

Agile^[1] es una metodología de desarrollo de software para construir un software de forma incremental utilizando iteraciones cortas de 1 a 4 semanas para que el proceso de desarrollo esté alineado con las necesidades cambiantes del negocio. En lugar de un desarrollo de un solo paso de 6 a 18 meses donde todos los requisitos y riesgos se predicen por adelantado, Agile adopta un proceso de retroalimentación frecuente donde se entrega un producto viable después de una iteración de 1 a 4 semanas.

E) Manifiesto ágil

El surgimiento de este manifiesto^[2] tuvo origen a principios del 2001 a raíz de los inconvenientes encontrados hasta ese tiempo relacionado con calidad del producto que los equipos de desarrollos de software entregaban, se elaboró un documento que aportó una mejora a los ciclos de desarrollo de software. Hoy en día existen una variedad de frameworks basados en la metodología ágil, pero todos buscan la misma finalidad convertir piezas o elementos tangibles que

sean útiles, sin embargo, es necesario considerar los doce principios del manifiesto:

1. Nuestra mayor prioridad es satisfacer al cliente a través de la entrega temprana y continua de software valioso.
2. Los requisitos cambiantes son bienvenidos, incluso tarde desarrollo. Los procesos ágiles aprovechan el cambio para la ventaja competitiva del cliente.
3. Entregue software que funcione con frecuencia, desde un par de semanas a un par de meses, de preferencia una escala de tiempo más corta.
4. Los empresarios y los desarrolladores deben trabajar juntos diariamente a lo largo del proyecto.
5. Construir proyectos en torno a personas motivadas.
6. Bríndeles el entorno y el apoyo que necesitan, y confiar en ellos para hacer el trabajo.
7. El método más eficiente y eficaz de transmitir información hacia y dentro de un equipo de desarrollo es una conversación cara a cara.
8. El software que funciona es la medida principal del progreso.
9. Los procesos ágiles promueven el desarrollo sostenible. Los patrocinadores, desarrolladores y usuarios deben poder mantener un ritmo constante indefinidamente.
10. Atención continua a la excelencia técnica y un buen diseño mejora la agilidad.
11. Simplicidad: el arte de maximizar la cantidad del trabajo no hecho--es esencial.
12. Las mejores arquitecturas, requisitos y diseños. surgen de equipos autoorganizados.
13. A intervalos regulares, el equipo reflexiona sobre cómo para volverse más efectivo, luego afina y ajusta su comportamiento en consecuencia.

F) Data Driven Scrum (DDS)

Data Driven Scrum (DDS) es un marco ágil diseñado específicamente para equipos de Data Science. En resumen, DDS tiene como objetivo mejorar la colaboración y la comunicación de un equipo de Data Science.

La Data Science Process Alliance (Alianza de Procesos de Data Science) creó Data Driven Scrum para abordar el hecho desde su perspectiva que otros enfoques ágiles bien conocidos, como a menudo Scrum y Kanban, a menudo no cumplen con las necesidades únicas de los proyectos de Data Science [1]. Bajo el esquema del DDS hay tres conceptos clave que permiten a un equipo obtener los beneficios de la agilidad dentro de un proyecto de Data Science. Esta agilidad ayuda a garantizar que un equipo se centre en tareas de máxima prioridad, al tiempo que permite que las tareas futuras se reordenen según sea necesario. Estos tres conceptos ágiles son:

1. Agile pretende ser una secuencia de ciclos iterativos de experimentación y adaptación.
2. El objetivo de cada ciclo debe ser tener una idea o experimento en mente, construirlo, observar el análisis, y luego analizar esas observaciones para crear la siguiente idea o experimento.
3. Pasar de una idea inicial, a la implementación, y al análisis de los resultados debería ser la base para una iteración. La finalización del proceso empírico debe marcar el final de esa iteración y no un número predeterminado de horas transcurridas.

G) Ciclo de vida Propuesto

Descripción del flujo de trabajo de alto nivel del DDS

- En primer lugar, los equipos de DDS hacen una lluvia de ideas sobre posibles preguntas para responder o experimentos para realizar.
- Luego, el equipo prioriza esas preguntas y experimentos, eligiendo el ítem de mayor prioridad en el que trabajar como equipo.

Esto incluye identificar los datos a utilizar y los modelos que deben crearse.

- A continuación, el equipo interpreta colectivamente los resultados del trabajo. En función de los resultados, el equipo despliega los resultados y prioriza el trabajo futuro.

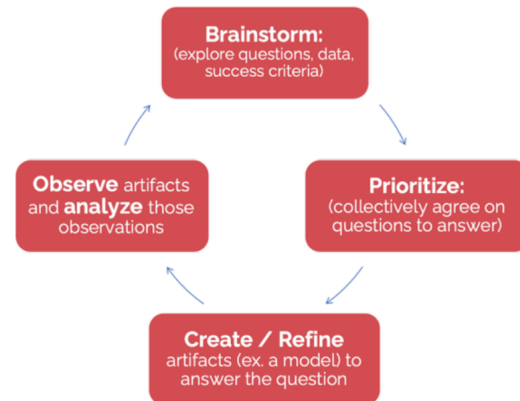


Figura 2. Descripción del flujo de trabajo de alto nivel del DDS

Roles del DDS

Al igual que Scrum, hay tres funciones clave en DDS:

Product Owner: El Product Owner en DDS es el punto central empoderado del liderazgo del producto la "voz del cliente", la persona que decide sobre los incrementos del producto. Prioriza qué características y funcionalidades construir, el orden en el que construirlos y qué aspectos de ellos observar y analizar. En resumen, el Product Owner es el propietario del Backlog y prioriza sus elementos, asegurando que cada uno esté claramente definido y que el trabajo próximo y las prioridades del equipo sean visibles y transparentes.

Experto en procesos: el experto en procesos actúa como entrenador, facilitador y eliminador de impedimentos. El experto en procesos también ayuda al equipo a comprender y adoptar los valores, principios y prácticas de DDS para ayudar a la organización a obtener resultados excepcionales de la aplicación de DDS.

Miembros del equipo DDS: Al igual que Scrum, cada equipo DDS es un grupo de tres

a nueve personas. El equipo de DDS está compuesto por un grupo multifuncional de profesionales (científicos de datos, ingenieros de software, ingenieros de pruebas, etc.) que tienen todas las habilidades necesarias para crear artefactos, por ejemplo, modelos, para responder a las preguntas / experimentos, es decir, para diseñar, construir, probar e implementar el producto deseado.

III. IMPLEMENTACION

Para conocer las mejoras o beneficios de implementar una metodología apropiada a un proyecto de Ciencia de Datos^[6] se realizó una encuesta con el objetivo conocer el estado actual de tres variables que miden la efectividad percibida por los diferentes roles (ingenieros de datos, administradores de proyectos y el cliente/producto owner). Las variables por considerar dentro de equipos de tecnología con soluciones de Ciencia de Datos son:

A. Calidad de la Solucion

Los equipos son más efectivos cuando tienen una fuerte preocupación por la calidad. Esto significa que con frecuencia hablan de ello y exploran formas de mejorarlo. La calidad desde la perspectiva de los stakeholders, Los equipos son más eficaces cuando las partes interesadas aprecian la calidad que ofrecen. Esto significa que su trabajo está libre de problemas graves y que las partes interesadas pueden obtener apoyo cuando sea necesario.

B. Bienestar de equipo de trabajo

Seguridad Psicológica

Los equipos son más efectivos cuando operan en entornos que hacen que sea seguro para los miembros asumir riesgos interpersonales. Por ejemplo, admitir que no saben algo, ofrecer retroalimentación, ofrecer ayuda o pedir ayuda.

Moral del Equipo

La moral del equipo es la voluntad de los equipos de seguir adelante incluso si las cosas se ponen difíciles. La moral alta es un amortiguador contra los contratiempos y el

agotamiento y un gran indicador de la atmósfera en un equipo.

C. Satisfaccion de cliente

Capacidad de Respuesta (Perspectiva de los Stakeholders)

Los equipos son más efectivos cuando son capaces de responder rápidamente a las necesidades emergentes de las partes interesadas (Stakeholders). Esto incluye nuevas ideas, oportunidades comerciales y también problemas o errores.

Felicidad de los Stakeholders

Se definen equipos efectivos como aquellos que logran satisfacer a sus stakeholders con resultados valiosos y que tienen miembros del equipo satisfechos.

Valor del equipo (Perspectiva de los Stakeholders)

Los equipos son más efectivos cuando las partes interesadas consideran valioso y útil el trabajo que se realiza.

El instrumento empleado para la medición de los parámetros descritos anteriormente es la **Scrum Team Survey (STS)** una encuesta diseñada para tal fin.

La encuesta es una herramienta para apoyar equipos ágiles en su ciclo de mejora continua, crear transparencia sobre el estado actual y ayuda a los equipos a identificar mejoras y hacer adaptaciones.

Las métricas empleadas en esta investigación están basadas en los estudios realizados por Christiaan Verwijfs y Daniel Russo a partir de su modelo de efectividad de equipos Scrum.

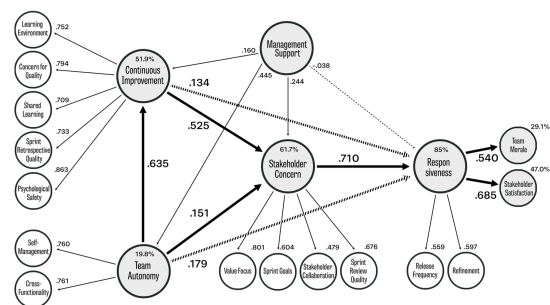


Figura 3. Ejemplo de un pie de figura

La encuesta realiza una evaluación de seis áreas:

1. Mejora Continua
2. Preocupaciones de los Stakeholders
3. Efectividad del Equipo
4. Capacidad de Respuesta
5. Autonomía del Equipo
6. Apoyo del área de Gestión

Sin embargo, la medición para el experimento se limitará estas áreas del modelo:

1. Mejora Continua
2. Preocupaciones de los Stakeholders
3. Efectividad del Equipo

1. Primera Iteración

La organización del equipo de trabajo previo a la implementación del ciclo de vida propuesto consta de un grupo de ingenieros de datos que pertenecen a la compañía donde se implementa el proyecto piloto trabaja de manera conjunta con el administrador de proyecto asignado a cada iniciativa de negocio, este a su vez vigila que el producto de ciencia de datos pase por el equipo de control de calidad y después de las sesiones de presentación de avances obtienen la aprobación del cliente. Respecto a la metodología y marco de trabajo utilizado estos equipos implementan una metodología SCRUM apegando las piezas de ciencia de datos como una pieza de software tradicional.

Roles del equipo a evaluar:

- 2 dueños del producto (Product Owner).
- 1 administrador de proyecto.
- 2 analistas de datos.
- 2 ingenieros de Machine Learning para creación de modelos de datos.
- 1 ingeniero y científico de datos.
- 2 ingenieros de revisión de la calidad.

Luego de la primera corrida de la encuesta para establecer la línea sobre la cual se evaluará el impacto en ambos equipos los resultados de las encuestas son los siguientes:

Caso de Estudio A

Los resultados iniciales muestran un puntaje relativamente alto en los componentes a evaluar sin embargo ambos componentes de Bienestar del Equipo de Trabajo junto con el de Felicidad de los Stakeholders requieren mayor atención, mientras que en el Caso de Estudio B ambos componentes Bienestar del Equipo de Trabajo y la Atención por la Calidad destacan como áreas de oportunidad.

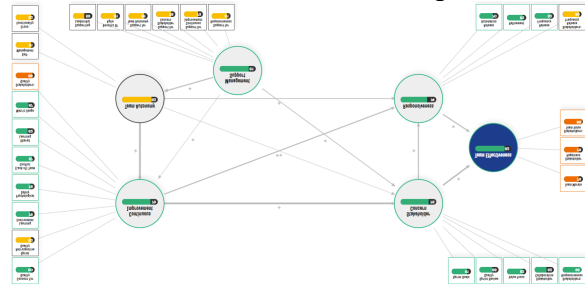


Figura 4. Resultados del caso de estudio A

Tabla 2

Linea base Caso de Estudio A	Puntaje
1) Satisfacción del Cliente	
Capacidad de Respuesta (Perspectiva de los Stakeholders)	96
Felicidad de los Stakeholders	78
Valor del equipo (Perspectiva de los Stakeholders)	96
2) Calidad de la Solución	
Atención por la Calidad	80
Calidad (Perspectiva de los Stakeholders)	89
Bienestar del equipo de trabajo	
Seguridad Psicológica	78
Moral del Equipo	78

Caso de Estudio B

Del mismo modo los resultados iniciales para este Caso de Estudio también muestran un puntaje relativamente alto y ambos componentes sobre la Bienestar del Equipo de Trabajo y la Atención por la Calidad destacan como áreas de oportunidad.

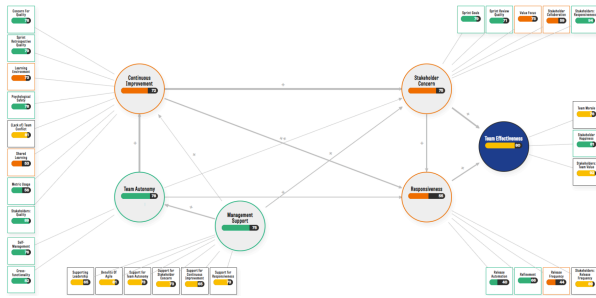


Figura 5. Resultados del caso de estudio B

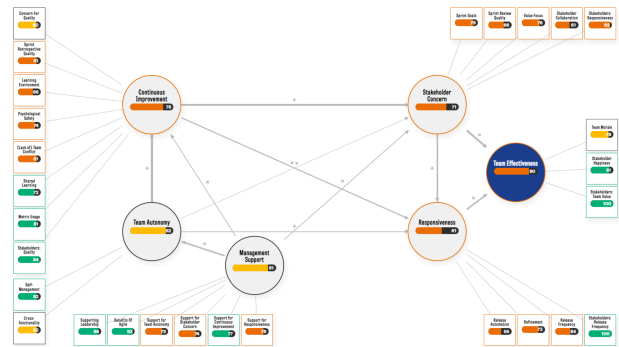


Figura 6. Resultados Segunda Iteración Caso de Estudio A

Tabla 3

Linea Base Caso de Estudio B	Puntaje
1) Satisfacción del Cliente	
Capacidad de Respuesta (Perspectiva de los Stakeholders)	94
Felicidad de los Stakeholders	81
Valor del equipo (Perspectiva de los Stakeholders)	92
2) Calidad de la Solución	
Atención por la Calidad	78
Calidad (Perspectiva de los Stakeholders)	89
3) Bienestar del equipo de trabajo	
Seguridad Psicológica	78
Moral del Equipo	78

2. Segunda Iteración

Cuatro semanas después de la primera iteración se realizó nuevamente la encuesta para medir el impacto de la incorporación de las sugerencias basadas en el Ciclo de Vida propuesto en las formas de trabajo de los equipos, los resultados fueron los siguientes:

1) Caso de Estudio A

Los resultados de la segunda iteración de la encuesta corresponden a las respuestas de estos 10 participantes que contestaron la primera vez. Se obtuvo un mejor puntaje en el promedio total del componente de Satisfacción del Cliente al igual que en el de Calidad de la Solución, sin embargo, el componente de Bienestar del Equipo de Trabajo tuvo una disminución.

Las siguientes tablas muestran el detalle de los subcomponentes evaluados:

Satisfacción del Cliente	Puntaje	Variación
a) Capacidad de Respuesta (Perspectiva de los Stakeholders)	92	-4
b) Felicidad de los Stakeholders	81	+3
c) Valor del equipo (Perspectiva de los Stakeholders)	100	+4
Promedio	91	

Tabla 4. Resultados del Componente Satisfacción del Cliente, Segunda Iteración Caso de Estudio A

Calidad de la Solución	Puntaje	Variación
a) Atención por la Calidad	80	0
b) Calidad (Perspectiva de los Stakeholders)	94	+5
Promedio	87	

Tabla 5. Resultados del Componente Calidad de la Solución, Segunda Iteración Caso de Estudio A

Bienestar del Equipo de Trabajo	Puntaje	Variación
c) Seguridad Psicológica	74	-4
Moral del Equipo	78	0
Promedio	76	

Tabla 6. Resultados del Componente Bienestar del Equipo de Trabajo, Segunda Iteración Caso de Estudio A

2) Caso de Estudio B

Asimismo, los resultados de esta segunda iteración corresponden a las respuestas de las 8 personas que contestaron la primera vez. En este caso también se obtuvo un mejor puntaje en el promedio total del componente de

Satisfacción del Cliente y en el de Calidad de la Solución, el componente de Bienestar del Equipo de Trabajo tuvo una disminución.

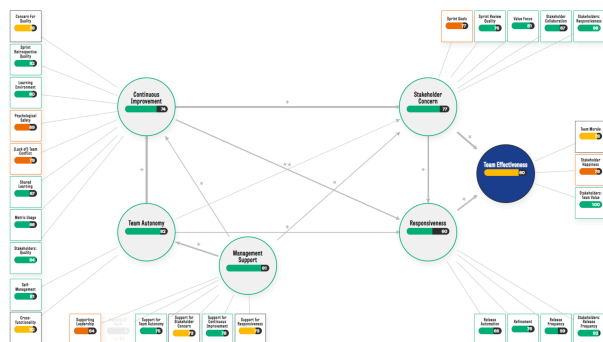


Figura 7. Resultados Segunda Iteración Caso de Estudio B

Las siguientes tablas muestran el detalle de los subcomponentes evaluados:

Satisfacción del Cliente	Puntaje	Variación
a) Capacidad de Respuesta (Perspectiva de los Stakeholders)	98	+4
b) Felicidad de los Stakeholders	78	-3
c) Valor del equipo (Perspectiva de los Stakeholders)	100	+8
Promedio	92	

Tabla 7. Resultados del Componente Satisfacción del Cliente, Segunda Iteración Caso de Estudio B

Calidad de la Solución	Puntaje	Variación
a) Atención por la Calidad	78	0
b) Calidad (Perspectiva de los Stakeholders)	94	+5
Promedio	86	

Tabla 8. Resultados del Componente Calidad de la Solución, Segunda Iteración Caso de Estudio B

Bienestar del equipo de trabajo	Puntaje	Variación
c) Seguridad Psicológica	69	8
Moral del Equipo	78	0
Promedio	73.5	

Tabla 9. Resultados del Componente Bienestar del Equipo de Trabajo, Segunda Iteración Caso de Estudio B

IV. ANALISIS DE RESULTADOS

Al analizar los resultados obtenidos en ambos equipos luego de implementar las adaptaciones con base al ciclo de vida propuesto, estos mejoraron sus puntajes para los componentes de Satisfacción del Cliente y Calidad de la Solución, buena parte de estos resultados se deben a que el equipo incorporaba ya en su forma de trabajo muchas de las etapas del ciclo de vida, esto facilitó en gran manera que la adopción fuese más efectiva en los equipos y explica los altos puntajes en los resultados de la encuesta desde la línea base.

En ambos casos de estudio un sólido dominio de Scrum como framework de trabajo, su enfoque en la frecuente entrega de valor al cliente y la perspectiva de producto en el desarrollo, habilita de manera positiva la adopción del ciclo de vida.

El componente de *Bienestar del Equipo de Trabajo* se ve influenciado en gran medida por la relación entre el equipo y los Stakeholders y la manera en que ambas partes perciben el valor generado, en cada caso de estudio hubo motivos distintos por los que este componente se vio afectado, sin embargo, el ciclo de vida como tal estima eventos de inspección y adaptación de la efectividad del equipo, puntualmente las Retrospectivas, que permiten dar respuesta a estas preocupaciones.

Caso de estudio A

El ciclo de vida permitió al equipo realizar una Primera Retrospectiva en la que se identificaron los retos que se presentaban en cuanto a los componentes a mejorar. Parte de las preocupaciones tenían que ver con la forma en la que se esperaba que los usuarios utilizaran la herramienta, comprender el flujo de trabajo que los usuarios manualmente realizan para automatizar estos procesos se definió como una de las principales oportunidades de mejora.

Adicionalmente el feedback provisto por los Stakeholders desde la perspectiva del equipo podía ser aún mejor para ayudar a definir más claramente los objetivos del producto y de esa

manera enfocar sus esfuerzos y evitar reprocesos.

Las acciones tomadas al respecto de estos puntos de mejora permitieron al equipo mejorar en los componentes de *Satisfacción del Cliente* y *Calidad de la Solución* al final de la segunda iteración de la encuesta.

En cuanto al componente de *Bienestar del Equipo de Trabajo* una Segunda Retrospectiva reveló que el equipo manifiesta una desalineación de la comunicación con los Stakeholders uno de los factores por lo que existe esa percepción es que los eventos Reviews al momento de inspeccionar los incrementos las conversaciones giran en torno a los detalles técnicos lo que podría confundir a los Stakeholders al ser su enfoque de producto. La forma en que se comunican los avances y la reacción de los Stakeholders ante estos logros, al mejorarse la comunicación generarían un impacto positivo en la moral del equipo y permitirían aumentar su puntaje en este componente.

Caso de estudio B

En este caso los retos para el equipo identificados en la Primera Retrospectiva se centraron en la claridad de los requerimientos puntualmente en la mejora de los criterios de aceptación de las historias de usuario. La necesidad de eventos de refinamiento más efectivos en los que se desglosen de manera más detallada estos requerimientos en una necesidad expresada por el equipo para lograrlo.

Del mismo modo la identificación de dependencias entre las Historias de Usuario permitiría al equipo a realizar una mejor estimación de los esfuerzos y planear con mayor precisión los posibles incrementos al final de la iteración.

De esta manera queda de manifiesto la relación estrecha que existe entre la transparencia de los requerimientos y como esto impacta de manera positiva en los componentes *Satisfacción del Cliente* y *Calidad de la Solución* al trabajarse en estas preocupaciones el equipo obtuvo un mejor puntaje en la segunda iteración de la encuesta.

Con respecto al componente de *Bienestar del Equipo de Trabajo* durante la Segunda Retrospectiva el equipo manifiesta que los retos se enfocan en la capacidad del equipo para poder completar las metas definidas al principio de su iteración.

Superar las dificultades en cuanto que suponen realizar despliegues de manera más frecuente y de esta forma entregar avances significativos a los Stakeholders durante las Reviews mejoraría el bienestar del equipo al generar un ambiente de éxito al final de cada iteración y en general de progreso hacia los objetivos.

V. CONCLUSIONES

Como ya se ha explicado en esta investigación el Data Science es una disciplina que emplea métodos estadísticos y de aprendizaje automático para convertir los datos en conocimiento práctico. Estas tecnologías están actualmente en un proceso de transición de su etapa de innovación a convertirse en herramientas a emplear con fines comerciales cada vez más accesibles para los mercados regionales.

A diferencia de las disciplinas tradicionales de desarrollo de software, el Data Science adopta el patrón de pensamiento centrado en datos, reconoce que la propiedad de los datos es más activa que pasiva, convierte estos datos en inteligencia, resuelve tareas intensivas de datos, mejora y enriquece las experiencias de los usuarios y abre las posibilidades a nuevos productos.

Bajo esta perspectiva al realizar este estudio en equipos que desarrollan proyectos de Data Science resulta valioso resaltar los aspectos identificados que influyen en el éxito de estos proyectos.

La adopción del agilismo y sus prácticas dentro de la organización permite que los equipos puedan afrontar los retos que los proyectos de Data Science presentan de forma efectiva, teniendo espacios para la experimentación y la reflexión periódica en la que se identifican oportunidades de mejora y se adapta la forma de trabajo. La flexibilidad que permiten estos marcos de trabajo habilita la creatividad en la resolución de problemas complejos.

Por otro lado, el éxito comercial de este tipo de proyectos va de la mano en mantener un enfoque de producto donde las tecnologías del Data Science son una herramienta y no el fin en si mismo. Más allá de las capacidades técnicas de los modelos generados estos deben ser capaces de responder a las necesidades de los usuarios que utilizarán el producto, por lo que la visión de producto debe estar presente desde el inicio y a lo largo de todo el ciclo de vida.

Esto también implica saber comunicar de manera precisa en términos de negocio los progresos en cada incremento del producto, una explicación técnica profunda de las aproximaciones estadísticas o matemáticas de los modelos empleados podría ser adecuada en un contexto diferente, un enfoque en cuanto al valor generado a los usuarios y al negocio a partir de los datos enriquece la conversación y potencia el interés e involucramiento de los Stakeholders lo que nos lleva al siguiente punto.

El éxito en general de los proyectos y los proyectos Data Science no son la excepción pasa necesariamente por el involucramiento de los Stakeholders en todo el ciclo de vida. La interacción de los equipos de trabajo con los Stakeholders debe ser una relación en donde la retroalimentación fluya en ambos sentidos de manera constante, sin eso resulta muy difícil que pueda entregarse valor a los usuarios y al negocio al final de cada iteración.

Para concluir los resultados de los casos de estudio muestran que un ciclo de vida agile impacta de manera positiva en la Satisfacción del Cliente y en la Calidad de las Soluciones y abre la posibilidad a tener conversaciones sobre cómo mejorar en otros aspectos como el Bienestar del Equipo de Trabajo. El Scrum Team Survey como herramienta de los equipos integrada dentro del ciclo de vida permite cuantificar estos componentes. Para la investigación se realizaron dos iteraciones y los resultados fueron positivos, más iteraciones de este ejercicio significan más oportunidades para mejorar por lo que es recomendable mantener estas revisiones de

manera continua para así avanzar en los retos que aún quedan pendientes.

El desarrollo de software es complejo, la incorporación de nuevas tecnologías aumenta aún más esta complejidad y nos empujan a buscar nuevos caminos, la agilidad no nos muestra un camino como tal, pero es la brújula que nos permite crear nuestro propio camino en la incertidumbre.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Kobielus, J. (Mayo de 2017). *Agile Development in Team Data Science*. (Wikibon) Recuperado en Enero de 2023, de <https://wikibon.com/agile-development-in-team-data-science/>
- [2] Agile Alliance. (2001). *The Agile Manifesto*. Recuperado en Enero de 2023, del sitio web oficial <https://agilemanifesto.org/principles.html>
- [3] John Wiley, Data Science & Big Data Analytics, «Discovering, Analysing Visualizing and presenting Data» (2015), Indiana, United States
- [4] Data Science Process Alliance, «A Jumpstart to Data Science Project Management,» Data Science Process Alliance, 2022. [En línea]. Available: <https://www.datascience-pm.com/wp-content/uploads/2022/11/Jumpstart2022.pdf>. [Último acceso: Enero 2023].
- [5] Nabati, E. G., & Thoben, K.-D. (2016). On Applicability of Big Data Analytics in the Closed-Loop Product Lifecycle: Integration of CRISP-DM Standard. *13th IFIP International Conference on Product Lifecycle Management (PLM)*.
- [6] Pinto, N. S. (Julio de 2020). *Framework para la evaluación de calidad de proyectos ágiles de software*. Recuperado en Enero de 2023, del sitio web http://sedici.unlp.edu.ar/bitstream/handle/10915/109769/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y
- [7] Verwijs, C., & Russo, D. (2022). A Theory of Scrum Team Effectiveness. *ACM Trans. Softw. Eng. Methodol.*
- [8] Vorhies, W. (Julio de 2016). *Data Science Central*. Recuperado en Enero de 2023, de <https://www.datasciencecentral.com/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome/>