

EDISON Data Science Framework: A Foundation for Building Data Science Profession For Research and Industry

La Idea Principal: Poner Orden en el Caos de la “Ciencia de Datos”

Imagina que todo el mundo empieza a llamarse “doctor”, pero no existe una carrera de medicina estandarizada. Algunas universidades enseñan biología, otras química, y otras solo primeros auxilios, pero todos los graduados salen con el título de “doctor”. Las empresas que los contratan no saben qué esperar, y los estudiantes no saben qué deben aprender. Sería un caos, ¿verdad?

Eso es exactamente lo que está pasando con la **Ciencia de Datos (Data Science)**. Es un campo nuevo y muy popular, pero:

- Las universidades están creando másteres y cursos improvisados, juntando materias que ya tenían.
- Las empresas no saben exactamente qué habilidades debe tener un “Científico de Datos”.
- Los estudiantes se gradúan con grandes vacíos en su conocimiento.

El proyecto **EDISON** y su marco de trabajo (**EDISON Data Science Framework - EDSF**) nacen para solucionar este caos. Su objetivo es crear una “receta” o un “plano” **estándar** para definir la profesión del Científico de Datos y guiar la creación de programas educativos de alta calidad.

Parte 1: El Problema - ¿Por qué Necesitamos un Marco de Trabajo?

El artículo empieza explicando por qué la situación actual es un problema:

1. **Es un Campo Multidisciplinario:** La Ciencia de Datos no es solo programación, ni solo estadística, ni solo conocimiento de negocio. Es una mezcla de todo, y es difícil enseñarlo bien.
2. **Las Universidades Están Improvisando:** La mayoría de los programas actuales simplemente empaquetan cursos viejos. El resultado son graduados que pueden ser muy buenos en una cosa (por ejemplo, Machine Learning), pero no saben nada de otras áreas cruciales (como la gestión de datos o la ingeniería de software).
3. **La Industria Necesita Algo Más:** Las empresas necesitan profesionales que puedan manejar el **ciclo de vida completo de los datos**: desde su recolección y limpieza hasta la creación de modelos y la comunicación de resultados que generen valor. Los graduados actuales a menudo no tienen esta visión completa.

En resumen, hay una **gran brecha entre lo que las universidades enseñan y lo que la industria realmente necesita**. El objetivo de EDISON es cerrar esa brecha.

Parte 2: La Solución - El Marco EDISON (EDSF) y sus 4 Piezas Clave

El EDSF es la solución propuesta. Piensa en él como un kit de herramientas para construir un buen programa de Ciencia de Datos. Tiene cuatro componentes principales que trabajan juntos:

1. **Marco de Competencias (CF-DS: Competence Framework)**
 - **¿Qué es?:** Es la “**descripción del puesto de trabajo**” para un Científico de Datos. Define todas las **habilidades y capacidades** que un profesional debe tener. No se trata de lo que *sabe*, sino de lo que es capaz de *hacer*.
 - **Grupos de Competencias:** El marco las agrupa en 5 áreas:
 1. **Análisis de Datos:** Estadística, Machine Learning, etc.
 2. **Ingeniería de Datos:** Ingeniería de software, infraestructura, Big Data.
 3. **Gestión de Datos:** Curación, preservación y gobierno de datos (esta es una de las áreas que EDISON identificó como crucial y que a menudo se olvida!).
 4. **Métodos de Investigación/Científicos:** Saber formular una hipótesis y diseñar experimentos (¡la otra gran área olvidada!).

5. **Conocimiento del Dominio:** Entender el negocio o el campo de investigación en el que se aplican los datos (finanzas, biología, marketing, etc.).

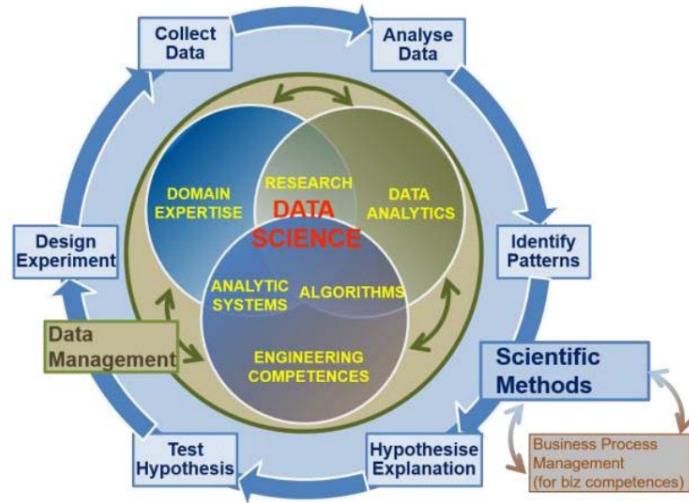


Figure 2. Data Science competence groups

2. Cuerpo de Conocimiento (DS-BoK: Data Science Body of Knowledge)

- **¿Qué es?:** Es la “biblioteca” o los “libros de texto” de la Ciencia de Datos. Define todos los **temas y conocimientos teóricos** que un estudiante debe aprender para adquirir las competencias del CF-DS.
- **Ejemplo:** Si la *competencia* (CF-DS) es “ser capaz de construir un modelo predictivo”, el *conocimiento* (DS-BoK) necesario incluye temas como “regresión lineal”, “árboles de decisión”, “redes neuronales”, etc.

3. Modelo de Currículo (MC-DS: Model Curriculum)

- **¿Qué es?:** Es la “plantilla del plan de estudios” para las universidades. Es la pieza que conecta las dos anteriores. Le dice a una universidad cómo estructurar sus cursos para enseñar el conocimiento (DS-BoK) que lleva a las competencias (CF-DS).
- **Características clave:**
 - Define **Resultados de Aprendizaje (Learning Outcomes)**: Qué debe ser capaz de hacer un estudiante al final de un curso.

- Se basa en la **Educación Basada en Competencias (CBE)**: Esto es muy importante. Significa que la evaluación no se basa en las horas que pasas en clase (“horas-silla”), sino en si puedes demostrar que has dominado una habilidad. Si ya sabes programar, podrías demostrarlo y saltarte esa parte.
- Organiza los temas en **Troncales** (obligatorios para todos) y **Electivos** (para especializarse).

4. Perfiles Profesionales (DSP: Data Science Professional Profiles)

- **¿Qué es?**: Reconoce que no todos los “Científicos de Datos” son iguales. Esta pieza define los diferentes **roles o especializaciones** que existen en el campo.
 - **Ejemplo de perfiles**: Un *Data Engineer* (más enfocado en la infraestructura), un *Data Analyst* (más enfocado en el análisis de negocio) y un *Machine Learning Scientist* (más enfocado en la investigación) son todos perfiles diferentes, y cada uno requiere un nivel distinto de dominio en cada una de las 5 competencias.
-

Parte 3: ¿Cómo se Usa Todo Esto en la Práctica?

El artículo explica cómo este marco puede ser utilizado en el mundo real.

1. **Para Crear Nuevos Cursos**: Una universidad que quiere lanzar un nuevo “Máster en Ciencia de Datos” puede usar el EDSF como una guía paso a paso:
 - Decide qué **perfil profesional (DSP)** quiere formar.
 - Usa el **modelo de currículo (MC-DS)** como esqueleto para diseñar las asignaturas.
 - Consulta el **cuerpo de conocimiento (DS-BoK)** para asegurarse de que cubre todos los temas teóricos necesarios.
 - Define los objetivos de cada asignatura basándose en el **marco de competencias (CF-DS)**.
2. **Para Evaluar Cursos Existentes**: Una universidad puede comparar su programa actual con el marco EDSF para ver qué “agujeros” tiene. Quizás descubren que están enseñando mucho sobre análisis pero casi nada sobre gestión de datos.

3. **Validación con “Universidades Campeonas”:** El equipo de EDISON no creó esto en una torre de marfil. Están colaborando activamente con universidades reales (las “Champion Universities”) para probar el marco, aplicar sus recomendaciones y obtener feedback para mejorarlo. Esto asegura que el marco sea práctico y no solo teórico.

Conclusión y Pasos Futuros

- **La Misión:** El EDSF busca estandarizar la profesión de Científico de Datos, creando una base sólida para la educación y la formación.
- **El Beneficio:** Ayudará a las universidades a crear mejores programas, a las empresas a saber qué habilidades esperar de los candidatos, y a los profesionales a planificar sus carreras.
- **El Futuro:** El proyecto continuará recogiendo feedback, expandiendo el marco para incluir conocimientos específicos de diferentes dominios (ej: Bio-Data Science) y trabajando con organismos de estandarización para que este marco sea ampliamente aceptado.

En resumen, el marco EDISON es un intento ambicioso y muy necesario de **profesionalizar la Ciencia de Datos**, asegurando que la próxima generación de expertos esté verdaderamente preparada para los desafíos del mundo real.