

DATA SCIENTIST: LOS MATEMÁTICOS HÍBRIDOS

M^a Cruz Gaya López
mcruz@universidadeuropea.es
Universidad Europea de Madrid. España

Núcleo temático: I. Enseñanza y aprendizaje de la Matemática en las diferentes modalidades y niveles educativos.

Modalidad: CB, T, MC, P, F, CP, CR

Nivel educativo: Unviersitario

Palabras clave: "formación universitaria", "profesión matemático", "análisis de datos", "data scientist"

Resumen

En la última década los matemáticos han visto incrementadas su empleabilidad de forma muy significativa. La aparición de profesiones como los data scientist (analistas de datos) han hecho que las empresas acudan a las universidades para captar talento en las áreas de las matemáticas y la tecnología. ¿Cómo son actualmente los profesionales que cubren estas vacantes? Normalmente son ingenieros informáticos con lagunas importantes en matemáticas, o matemáticos que necesitan actualizarse en tecnología.

Se hace necesaria formación específica para estos perfiles profesionales. Como muestra de ello en el “Libro Blanco para la definición de titulaciones universitarias en el ámbito de la Economía Digital” el Ministerio de Industria puso de manifiesto esta necesidad y proporcionaba las guías para el desarrollo de este tipo de titulaciones, tanto a nivel de grado como a nivel de máster.

En esta comunicación presentamos un ejemplo, el grado en Ingeniería en Matemática aplicada al Análisis de Datos. Este grado unifica conocimientos de matemáticas, informática y empresa, condimentados con metodologías de aprendizaje basadas en proyectos, con el objetivo de crear los profesionales que demanda el entorno profesional.

Introducción

En los últimos años han irrumpido en nuestra sociedad términos como el Internet de las Cosas, la industria 4.0, las Smart Cities, y el Big Data. Todos estos términos tienen en común la necesidad del almacenamiento masivo de grandes cantidades de datos en tiempo real y del análisis automático de los mismos con el objetivo de obtener información útil como base en la toma de decisiones a niveles estratégicos. Los ámbitos de aplicación son, además, muy extensos como Salud, Educación, Industria, Energía, y un largo etcétera. Los efectos que está teniendo en los resultados de las organizaciones son tales que en algunos casos se están

tomando decisiones de gran nivel basándose únicamente en los resultados de estos procesos automáticos. Un ejemplo de ello son los nuevos seguros de automóviles que están basando el cálculo de las cuotas en los resultados de analizar los datos de la trazabilidad recopilados a través de todos los sensores que hoy en día disponen los automóviles. Es tal la influencia que el análisis de datos está generando en la sociedad y, en particular, en la industria, que ya se habla de la 3° revolución industrial.

Este reto requiere de profesionales que estén bien formados y perfectamente preparados para generar valor en torno a esta gran oportunidad.

El Data Scientist

El profesional del análisis de datos, o “data scientist” como se conoce internacionalmente, se encuadra dentro del ámbito de conocimiento de las ciencias y de las tecnologías de la información. También, aunque en menor medida, tiene que tener conocimientos de Empresa para ser capaz de aportar valor dentro de la entidad de la que se desarrolle.

Un analista de datos tiene que ser capaz de abordar funciones de Explotación, Desarrollo, Diseño y Administración de sistemas de gran volumen de datos, y en un nivel superior, las funciones de Análisis, Arquitectura y Dirección de los mismos. Para ello tiene que dominar las técnicas y herramientas de las matemáticas y la estadística, así como, las técnicas y herramientas software que se utilizan para el almacenamiento de grandes cantidades de datos en tiempo real, para el análisis y procesamiento de los mismos (utilizando modelos predictivos e inteligencia de negocio), así como para la visualización de forma adecuada para mejor entendimiento del encargado de tomar la decisión final.

Actualmente la formación utilizada para este perfil profesional ha estado centrada principalmente en nivel de postgrado. Existen numerosos másteres tanto nacionales como internacionales que ofrecen esta formación. Sin embargo, este tipo de enseñanzas, en la mayoría de los casos de 60 ECTS no son suficientes. Los perfiles de ingreso de estos másteres son graduados del sector TIC que carecen de formación suficiente en matemáticas y en empresa o matemáticos que carecen de formación en tecnologías de información. Se hace necesaria una formación de grado de 240 ECTS que permita formar al estudiante de forma integral en estas 3 ramas de conocimiento.

Un profesional con mucho presente y más futuro

Se prevé una alta empleabilidad para los egresados de este grado. Para ello se ha tenido en cuenta varios estudios, entre ellos los realizados por organizaciones como: Adecco, Ametic, la CODDII (conferencia de directores y decanos de ingeniería informática) y otros.

En cuanto a la situación de empleabilidad actual se puede mencionar:

- Las titulaciones técnicas han aumentado su demanda en el mercado en los últimos años. Las ingenierías son las mejor posicionadas, de hecho, casi la mitad de las ofertas de empleo se centran en perfiles cualificados con carreras técnicas. De entre todos ellos los profesionales de las Tecnologías de la Información y las comunicaciones son los que mayor tasa de ocupación, el 94%.
- Los datos del informe Carreras (Adecco, 2014) reflejan el ascenso de ofertas de empleo para titulaciones tecnológicas, encontrándose 14 disciplinas técnicas y de Ciencias entre las 20 más demandadas.
- Las tasas de Afiliación a la Seguridad Social también son muy altas. A destacar los ingenieros informáticos, que alcanzan el 78% durante su 4º año.
- Como aspectos negativos se encuentran los salarios (entre los 18000 y los 30000 euros al año para los recién egresados) y el hecho de que es un mundo muy masculinizado. Además este interés profesional, queda patente en las ofertas de empleo en los grandes portales dentro de este sector. Ejemplos son 82 puestos en Infojobs (consulta de octubre de 2014), 63 puestos en AT&T. Y en portales más específicos tenemos 300 puestos en Big data Careers y 226 ofertas en Big data Spain.

Las previsiones de empleo futuras indican que se crearán casi medio millón de puestos de trabajo en empleos relacionados con la Economía Digital en el conjunto de los países miembros de la UE para el año 2020. Más en concreto, algunos estudios en Europa prevén que, solo en el Reino Unido, el número de personal especialista en datos trabajando en grades empresas se incrementará en más del 240% en los próximos cinco años.

Cabe destacar también el autoempleo como alternativa al empleo por cuenta ajena. Se han registrado 3550 start-ups en España en el pasado 2014 (estudios de la cámara de comercio del año 2014), la mayoría de ellas con base tecnológica.

En España no se ha encontrado ningún título específico que plantee formación para este perfil profesional, el de analista de datos (data scientist). Sin embargo, fuera de España, tanto en Europa como en Estados Unidos, sí que existen, manteniéndose la denominación de Data

Science como la más común. La traducción de este término al español, “ciencia de los datos” hemos considerado que no identifica la formación que provee puesto que una parte esencial del plan de estudios, con un peso del 22,5 % sobre el total, son los conocimientos de matemáticas. Por ello consideramos que la denominación de Grado en Ingeniería en Matemática Aplicada al Análisis de Datos se justifica por el contenido de la misma y el perfil profesional que cubre. El egresado de este grado es un matemático y tecnólogo que va a aplicar su conocimiento en el análisis de los datos, de ahí su título.

Grado en Ingeniería Matemática aplicada al Análisis de Datos.

Perfil de egreso.

La descripción del estudiante tipo que finaliza sus estudios en el Grado en Ingeniería Matemática aplicada al Análisis de Datos se define a través de las competencias generales relacionadas con la profesión que se muestran a continuación:

CG1. Capacidad para recopilar e interpretar datos e informaciones y extraer conclusiones reflexionando sobre asuntos de índole social, científica o ética en el ámbito del análisis de datos.

CG2. Conocimiento de las herramientas matemáticas básicas, principalmente de cálculo, álgebra lineal y probabilidad, para su aplicación rigurosa y fiable que permita modelizar problemas reales complejos.

CG3. Conocimiento y aplicación de las tecnologías y herramientas informáticas, principalmente las bases de datos, la programación de algoritmos y la inteligencia artificial, para construir, analizar e interpretar fuentes de datos incluyendo su obtención, preprocesado, almacenamiento, análisis y visualización de resultados, que ayuden en la toma de decisiones en campos diversos.

CG4. Conocimiento de los fundamentos de empresa que permitan entender, interpretar y mostrar los datos de la forma más adecuada a cada organización.

CG5. Capacidad para colaborar con profesionales de otros campos (financiero, marketing, sanidad, etc.), trabajando en equipo, participando en la organización y la gestión de proyectos, atendiendo a las normas de ética profesional y las relativas a la protección y seguridad de datos..

CG6. Capacidad de aprender de forma autónoma nuevas técnicas y herramientas, así como defender la necesidad de mantener, a lo largo de su vida profesional, un aprendizaje continuado y abordar problemas nuevos con nuevas herramientas.

CG7. Capacidad para comunicar a todo tipo de audiencia de manera clara y precisa, conocimientos, metodologías, ideas, problemas y soluciones en el ámbito del análisis de datos.

Estas competencias se materializan en la adquisición a lo largo de las distintas materias del grado en las siguientes competencias específicas:

BÁSICAS.

CE1 Comprensión del concepto de empresa, así como las áreas funcionales de la misma, incluyendo las relaciones entre ellas, y aplicación de las distintas herramientas disponibles en cada una de ellas.

CE2 Comprensión de los fundamentos básicos, los principios y las aplicaciones de los sistemas informáticos, el desarrollo software y las bases de datos.

CE3 Conocimiento y aplicación de forma eficiente los modelos de tipos de datos y los algoritmos para diseñar soluciones a problemas.

CE4 Comprensión de las técnicas de diseño, implementación, captación, almacenamiento y explotación de bases de datos y los sistemas de gestión de bases de datos, tanto estructuradas como no estructuradas, monolíticas y distribuidas

CE5 Comprensión del lenguaje matemático y su aplicación para enunciar proposiciones y transmitir los conocimientos adquiridos en los distintos campos de las matemáticas

CE6 Aplicación de los conocimientos sobre: álgebra lineal, cálculo diferencial e integral, métodos numéricos, estadística y optimización para la resolución de problemas.

CE7 Conocimiento y aplicación de las herramientas informáticas de análisis estadístico, cálculo numérico y simbólico, visualización gráfica, optimización y otras para experimentar en Matemáticas y resolver problemas.

AVANZADAS

CE8 Capacidad para evaluar las tendencias en el mercado de la Economía digital, así como su impacto en el desarrollo social, económico y cultural.

CE9 Análisis de las técnicas de marketing digital valorando el impacto de sus decisiones en los beneficios, el mercado, las personas y la sociedad.

- CE10 Análisis de las técnicas de ingeniería de sistemas de información a los procesos de negocio.
- CE11 Conocimiento y aplicación de los principios fundamentales y técnicas básicas de los sistemas inteligentes y su aplicación prácticas.
- CE12 Conocimiento del Ciclo de vida de los datos, desde la operación hasta la visualización. De los datos al conocimiento y del conocimiento a la estrategia de negocio.
- CE13 Análisis de las técnicas de replicación, conservación, restauración y anonimización de los datos.
- CE14 Conocimiento de los modelos y formas de evaluación de servicios en base a criterios de capacidad de utilización y calidad de servicio.
- CE15 Conocimiento de la legislación en materia de datos personales, privacidad y derechos fundamentales de las personas, así como los criterios y mecanismos de evaluación y certificación de la seguridad vigentes en la actualidad.
- CE16 Aplicación de las metodologías, arquitecturas y técnicas propias de Big Data para el almacenamiento y la gestión de los datos.
- CE17 Aplicación de los modelos y estándares del ámbito de los sistemas de grandes volúmenes de datos.
- CE18 Aplicación de las técnicas de aprendizaje computacional para diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.
- CE19 Comprensión de las técnicas de interoperabilidad de sistemas e integración y agregación de datos.
- CE20 Capacidad para diseñar interfaces eficientes en el contexto del Big Data que garanticen la accesibilidad y usabilidad, utilizando.
- CE21 Aplicación del pensamiento estadístico y tener la capacidad de enfrentarse a las distintas etapas de un estudio estadístico (desde el planteamiento del problema hasta la exposición de los resultados).
- CE22 Conocimiento y aplicación de las técnicas y modelos, matemáticos, estadísticos y de optimización, aplicados al procesamiento de datos, los sistemas de ayuda a la decisión, la búsqueda de relaciones entre variables y la realización de predicciones.

CE23 Capacidad para participar de forma activa en proyectos que incluyan el uso de datos abiertos y herramientas de análisis estadístico en entornos distribuidos.

CE24 Capacidad para participar de forma activa en proyectos en el ámbito de sistemas de gran volumen de datos que requieran el conocimiento, la evaluación, la selección y la utilización de herramientas soporte para el desarrollo de proyectos de big data.

CE25 Ejercicio original a realizar individualmente y presentar y defender ante un tribunal universitario, consistente en un proyecto en el ámbito de la ciencia de los datos de naturaleza profesional en el que se sinteticen e integren las competencias adquiridas en las enseñanzas.

Plan de estudios.

El plan de estudios del grado en Matemática Aplicada en la Universidad Europea de Madrid, combina los conocimientos de 3 materias principales, matemáticas, informática y empresa con el desarrollo de proyectos reales y la formación en competencias transversales.

| Grado en Ingeniería Matemática Aplicada al Análisis de Datos | | | | | | | | | | |
|--|---|--|--|---|---|-------------------------------------|-------------------------|------------------------|---|------------------------------------|
| | TRIMESTRE 1 | | | | TRIMESTRE 2 | | | TRIMESTRE 3 | | |
| PRIMERO | Fundamentos de empresa | Principios básicos de la estadística | Fundamentos de programación y computadores | Eficacia personal y profesional | Álgebra | Programación orientada a objetos | Marketing y Ventas | Bases de datos | Cálculo numérico | Proyecto de Sistema de Información |
| SEGUNDO | Operaciones y RRHH | Matemática discreta | Estructuras de datos | Impacto e influencia relacional | Inferencia estadística | Inteligencia Artificial | Proyecto de Open Data I | Gestión de proyectos | Estadística computacional | Proyecto de Open Data II |
| TERCERO | Dirección estratégica y Análisis Financiero | Lenguajes de programación estadística | Almacenamiento o masivo de datos | Proyecto de Big Data I | Análisis multivariante de datos (Econometría I) | Aprendizaje Automático | Proyecto de Big Data II | Visualización de datos | Liderazgo emprendedor | Proyecto de Big Data III |
| CUARTO | Virtualización y seguridad | Estudio de datos de panel (Econometría II) | Sistemas de información empresarial | Aplicaciones y tendencias del análisis de datos | Economía y marketing digital | Seguridad y Legislación profesional | Trabajo fin de grado | Prácticas externas | Ampliación de prácticas /actividades universitarias | Trabajo fin de grado |

Figura 10 Plan de estudios del Grado en Ingeniería Matemática aplicada al Análisis de Datos.

El estudiante adquiere conocimientos en matemáticas desde sus bases a través de asignaturas como Álgebra, Cálculo o Matemática discreta para profundizar en la estadística adquiriendo competencias en inferencia estadística, estadística computacional o econometría. En informática estudia desde asignaturas básicas como programación o bases de datos para especializarse en inteligencia artificial, big data, aprendizaje automático o visualización de datos. En la materia de empresa el estudiante comienza con una asignatura de fundamentos que le da una visión genérica de la estructura y organización de una empresa para conocer los procesos que se desarrollan en cada departamento de la misma haciendo especial hincapié en la Economía y el Marketing Digital.

Un grado de Project Based School

Siguiendo la metodología de aprendizaje basado en proyectos (Project Based School) que se utiliza en la Escuela de Arquitectura, Ingeniería y Diseño de la UEM, el estudiante tendrá la oportunidad de realizar un proyecto real en cada curso académico propuesto por empresas y organizaciones externas donde aplicará los conocimientos adquiridos en las asignaturas de ese año. Así el proyecto de sistema de información, que se realiza en primero, el alumno tendrá la oportunidad de desarrollar proyectos que incluyan sistemas sencillos para la adquisición de datos y su almacenamiento, cálculos estadísticos sobre los mismos y el diseño e implementación de aplicaciones que muestra los resultados. En segundo se realiza el Proyecto de Open Data donde el alumno implementará un proyecto que incluya el uso de datos abiertos y herramientas de análisis estadístico en entornos distribuidos. En tercero realizará la implementación y puesta en marcha de un proyecto en el ámbito de los sistemas de gran volumen de datos que requieran el conocimiento, la evaluación, la selección y la utilización de herramientas soporte para el desarrollo de proyectos de big data.

Conclusiones.

- El Data Scientist es un perfil profesional muy necesario ya en el entorno profesional ya en la actualidad y, de forma más pronunciada en un futuro a corto, medio y largo plazo.
- La formación actual no cubre las necesidades: formación en 3 áreas principales: matemáticas (estadística), informática y empresa, esta última en menor medida.
- Es necesario adaptar la formación universitaria a las nuevas profesiones que genera la Economía Digital, el científico de datos, o data scientist, es el ejemplo más claro de ello.
- El Grado en Ingeniería Matemáticas aplicadas al Análisis de Datos cubre esta necesidad.
- Actualmente esta titulación ya tiene los permisos de las agencias de acreditación y está prevista su puesta en marcha en el curso 2017/2018. Los trabajos futuros consistirán en realizar una evaluación de la misma realizando un análisis tanto cuantitativo como cualitativo.

Referencias bibliográficas

- ACM-AIS. (2010). Information Systems 2010. Curriculum Guidelines for Undergraduate Degree Programs in Information Systems. Association for Computing Machinery (ACM) and Association for Information Systems (AIS).
- Adecco. (2015). Informe Carreras. Obtenido de <http://www.adecco.es/data/NotasPrensa/pdf/639.pdf>
- Ametic, R. (2011). Perfiles profesionales más demandados en el ámbito de los contenidos digitales en España 2012-2017.
- Conferencia de Directores y Decanos de Ingeniería Informática (Coddii), C. d. (2013). Informe de Empleabilidad. Obtenido de <http://coddii.org/wp-content/uploads/2013/04/coddinforme-empleabilidad-2013.pdf>
- Ministerio de Industria, Energía y Turismo. Libro Blanco para el diseño de las titulaciones universitarias en el marco de la economía digital. (2015) Obtenido de <http://www.agendadigital.gob.es/planes-actuaciones/Bibliotecacontenidos/Material%20Formaci%C3%B3n%20de%20excelencia/Libro-Blanco.pdf>