

ANTEPROYECTO

HIFP

Herramienta para la Identificación de Fuga de Personal

Nombre del equipo: **Grupo 2**

Miembros del equipo:

Alumno	email
Álvaro Bru Serrano	alvaro_bs95@hotmail.com
Antonio M. Cubero Lastra	acubelas@yahoo.es
Marco Benito Kaesbach	marcobenito97@gmail.com
Manuel Fernández Oria	manu.oria@hotmail.com
John Jairo Orozco Giraldo	jjog22@hotmail.com

Índice

1. Identificación del proyecto.....	3
2. Resumen.....	3
3. Antecedentes y estado actual del proyecto.....	4
3.1 ¿Qué es la inteligencia artificial?.....	4
3.2 Estado actual de la inteligencia artificial.....	4
3.3 Antecedentes al proyecto	5
3.4 Casos similares	5
3.5 Estado del proyecto.....	6
4. Objetivos del Proyecto	6
4.1 Objetivo general.....	6
4.2 Objetivos específicos.....	6
5. Plan de trabajo y presupuesto del proyecto	7
5.1 Plan de trabajo y distribución de tareas	7
5.2 Presupuesto	9
6. Viabilidad y plan de recursos	9
6.1 Estudio de viabilidad técnica.....	9
6.2 Plan de recursos	10
6.2.1 Desarrollo del proyecto.....	10
6.2.2 Mantenimiento del servicio	10
6.3 Estudio de viabilidad económica.....	10
7. Referencias.....	11

1. Identificación del proyecto

Esta propuesta hace referencia al proyecto “Herramienta para la identificación de fuga de personal”, (HIFP).

Dicha herramienta permitirá la identificación de posibles intenciones de cambios de trabajo entre los empleados pertenecientes al departamento de Data Science, con el objetivo de poder tomar decisiones anticipadas para cubrir el posible gap que pueda dejar el empleado, bien sea lanzando nuevas ofertas de trabajo, formando a otros trabajadores para el puesto o buscando soluciones que mejoren la satisfacción del empleado para lograr retenerlo.

El proyecto está enmarcado dentro del ámbito de la inteligencia artificial (UNESCO-120304), tratándose de una herramienta inteligente capaz de predecir la posibilidad de cambio de trabajo de cualquier empleado.

Otros campos involucrados en el proyecto son:

- Análisis de errores (120602)
- Construcción de algoritmos (120601)
- Análisis de datos (120903)
- Técnicas de predicción estadística (120914)

2. Resumen

La fuga de talento puede suponer un gran problema para las empresas por varias razones:

- Pérdida de un profesional que conoce bien los sistemas y la forma de trabajar de la empresa.
- Pérdida de tiempo y dinero en buscar nuevos candidatos para cubrir un rol que posiblemente tarden tiempo en ejecutar al mismo nivel que el anterior empleado.
- Gasto en formación del nuevo empleado.
- Posible traspaso de talento a una empresa de la competencia.

En muchos casos, un simple movimiento por parte del empleador podría hacer cambiar la opinión del empleado y evitar los problemas mencionados.

HIFP propone un sistema de identificación de fuga de talento para la empresa, haciendo uso de la inteligencia artificial para predecir la probabilidad de que un empleado del departamento de Data Science esté buscando un nuevo empleo. El uso de la inteligencia artificial permitirá que el proceso sea sencillo, ágil y automático.

Para ello, se ofrece una herramienta que permite generar encuestas de satisfacción para los empleados a través de un formulario web. Una vez recolectada la información, nuestro algoritmo devolverá la probabilidad que tiene cada empleado de querer cambiar de puesto. Esta información se presentará en forma de panel de control o dashboard que permitirá mantener una monitorización constante e incluso alertará cuando la probabilidad de que uno de los empleados quiera cambiar de trabajo aumente por encima de cierto umbral. Para ello, la herramienta reevaluará a cada empleado de manera automática cada cierto tiempo.

Esta solución ofrece múltiples posibilidades a los altos cargos o al departamento de recursos humanos en la hora de gestionar el capital humano. Permite la anticipación ante la fuga de talento, con el objetivo de tomar decisiones anticipadas, como pueden ser: mejorar las condiciones del empleado en cuestión; buscar nuevos roles y responsabilidades; y en caso de que realmente el empleado deje la empresa: preparar ofertas de trabajo para cubrir la futura vacante; formar a otros empleados para cubrir su puesto; o cualquier otra acción que minimice la pérdida de productividad en relación con ese puesto.

Además, funcionalidades más avanzadas (premium) incluyen un análisis de sentimientos, proporcionando un rating de satisfacción del empleado a partir de un campo adicional de observaciones o comentarios en el formulario de satisfacción. Esta funcionalidad ofrece la posibilidad de conocer el nivel de satisfacción general dentro del departamento con el fin tomar decisiones más transversales en términos de organización interna.

3. Antecedentes y estado actual del proyecto

3.1 ¿Qué es la inteligencia artificial?

Antes de hablar de la situación actual de la Inteligencia Artificial (IA), conviene conocer en que consiste. Aunque no existe una definición exacta, se considera la disciplina científica que busca la creación de máquinas o programas capaces de imitar operaciones realizadas por la mente humana, tales como el razonamiento, el aprendizaje, la creatividad o la capacidad de planear. Actualmente encontramos tres niveles de inteligencia artificial: la inteligencia artificial débil, la inteligencia artificial general, y la inteligencia artificial fuerte.

La inteligencia artificial débil sirve para resolver problemas concretos delimitados, los cuales una máquina es capaz de resolver mejor que una persona. La inteligencia artificial general es capaz de realizar cualquier tarea realizada por una persona, y consiste en la agrupación de muchas inteligencias artificiales débiles. Finalmente, la inteligencia artificial fuerte consiste en la creación de una máquina consciente, con sentimiento e inteligencia propia [1].

La realidad es que, aunque el término de inteligencia artificial nos pueda llegar a sonar nuevo o incluso lejano, lleva en nuestras vidas más de lo que creemos, y esto no ha hecho más que comenzar.

3.2 Estado actual de la inteligencia artificial

Actualmente estamos viviendo uno de los mayores momentos de crecimiento de la IA. Y es que esta técnica es aparentemente imparable y próximamente influirá en cada una de nuestras decisiones [2]. El principal motivo de este crecimiento es la gran cantidad de datos que se manejan hoy en día en todas las empresas, gracias a la digitalización del mundo.

La realidad es que actualmente quedan pocas actividades humanas que sean ajenas a las tecnologías de IA. Gracias a los continuos avances tecnológicos encontramos nuevas formas de alcanzar nuestros objetivos a través de nuevas técnicas como son: Machine Learning, Deep Learning, Reinforcement Learning, y Natural Language Processing [3].

En nuestro caso vamos a centrarnos en los algoritmos de Machine Learning (ML), ya que a través de ellos somos capaces de abordar problemas como el propuesto en este proyecto. El ML se basa en técnicas estadísticas que permiten a la máquina adquirir información de los datos de entrada y aprender de ellos para realizar una actividad cada vez de manera más efectiva [3]. Dentro de este grupo de algoritmos encontramos los de clasificación, los cuales son capaces de predecir a que clase pertenece un input concreto en función de sus variables.

Para nuestro proyecto nos centraremos en este tipo de algoritmos ya que nos encontramos ante un problema de clasificación.

3.3 Antecedentes al proyecto

La inquietud por saber si un empleado de tu empresa se va a ir o se va a quedar no es algo nuevo. Existen diferentes casos en los que diferentes empresas han realizado algoritmos para intentar predecir con que probabilidad sus empleados podrían irse, y de esta manera ayudar al área de Recursos Humanos (RRHH) para anticiparse a dicho suceso.

Para poder llevar a cabo este tipo de proyectos es necesario tener los datos suficientes para realizar un buen análisis de estos. De manera, que estos deben tener no solo una alta calidad, sino también deben ser completos. Es importante que las variables tengan que ver con el tipo de persona, como el género o la edad, y con el tiempo que lleva la persona en la empresa. De esta manera seremos capaces de conocer que variables afectan a la fuga de trabajo [4].

Como veremos en la siguiente sección, existen diferentes empresas que ya han llevado a cabo este tipo de proyectos con éxito, basándose en las variables anteriormente comentadas. Esto nos da un plus a la hora de entender y aplicar posibles soluciones previas a nuestro caso.

3.4 Casos similares

Como hemos comentado en el apartado anterior existen diferentes casos similares en los que diferentes empresas buscan predecir tanto la fuga como el absentismo laboral. Aquí vamos a exponer algunos de esos casos ya que guardan similitud con nuestro proyecto.

El primer caso que vamos a comentar es el de AIS Group, consultora que fue pionera en el uso de la IA para detectar el absentismo laboral. El principal motivo para realizar esta detección era para permitir que el departamento de RRHH se adelantase a los hechos y mitigase los efectos negativos. Ya que como comentaba la empresa, “El absentismo laboral es uno de los problemas más graves a los que se enfrentan las áreas de RRHH de las empresas y su gravedad es proporcional al tamaño de aquellas”. Para solucionarlo se desarrolló un modelo predictivo basado en machine learning, capaz de analizar miles de variables y encontrar relaciones entre ellas [5].

El segundo caso es el de Wal-Mart o Credit Suisse, empresas que analizan grandes cantidades de datos para obtener la probabilidad de que un empleado abandone la compañía. Para hacer este tipo de predicciones tienen en cuenta docenas de factores que pueden influir, desde los años que llevan en la empresa, hasta test de personalidad. Aunque cada empresa es un mundo distinto, desde Credit Suisse apuntan que “por cada punto que se reduce la deserción no deseada, la entidad ahorra entre 75 y 100 millones de dólares” [6].

El último caso del que vamos a hablar es el de Sinnetic, esta consultora reveló un algoritmo capaz de predecir con tres meses de anticipación, cuando un cliente puede irse. Aunque existen diferentes algoritmos que podrían llevar a cabo la detección de fuga de clientes, la consultora propone el algoritmo de Satch con el que obtienen un 90% de confianza para predecir en un trimestre la fuga de clientes. Al estar centrado específicamente en el tema, dicen que es posible no solo conocer cuántos clientes se irán sino los motivos potenciales [7].

3.5 Estado del proyecto

Gracias al momento actual que vive la IA, podemos enfrentarnos a proyectos de este tipo. Tenemos los medios necesarios para abordar de manera satisfactoria el proyecto que se nos propone.

Nuestra propuesta tiene como base la eficacia y la eficiencia, todo el equipo está cualificado en el ámbito de la IA con experiencia previa en el desarrollo de modelos, así como en su puesta en producción. Nuestro modelo estará basado en un algoritmo de clasificación de ML, el cual será elegido una vez llevemos a cabo las pruebas pertinentes hasta encontrar el que mejor se adapte a las necesidades. Además, HIFP no sólo permite anticipar la salida de un empleado, sino que se diferencia de las soluciones propuestas hasta el momento ya que propone un sistema de monitorización constante e incluso puede proporcionar información adicional relevante para la empresa para poder conocer los motivos por los que un empleado puede querer dejar su actual puesto. De este modo, sirve como apoyo en la toma de decisiones estratégicas que afecten a todo el departamento con el fin de mejorar la satisfacción general.

4. Objetivos del Proyecto

4.1 Objetivo general

Como equipo de proyecto y como individuos tenemos la obligación ante la Universidad Europea de Madrid, de comprobar y aplicar nuestros conocimientos adquiridos a lo largo del máster, con profundización en técnicas y metodologías aplicadas a la Inteligencia Artificial y Machine Learning, utilizando casos de la vida real o de la industria. En este caso, buscamos implementar una solución de IA/ML para predecir/clasificar si un empleado con rol de Data Science está buscando un nuevo empleo o no, de modo que sea práctico y utilizable para las empresas que tengan tal reto.

4.2 Objetivos específicos

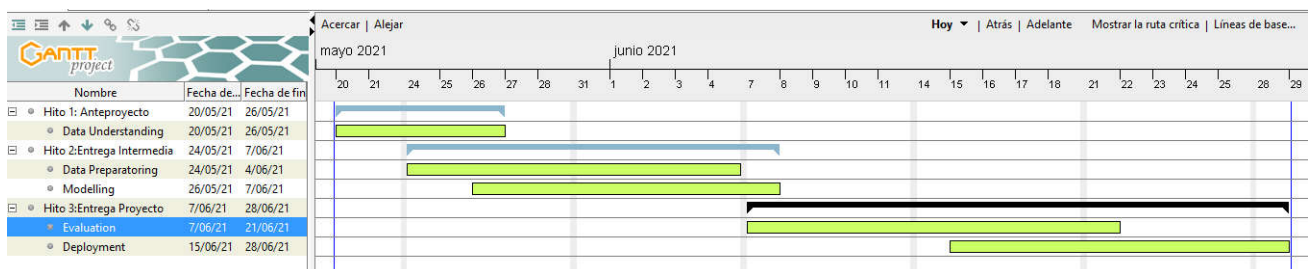
- Aplicar la metodología CRISP-DM en toda la implementación del proyecto, con actividades que son similares a las etapas de modelamiento de Machine Learning y despliegue de modelos.
- Utilizar técnicas avanzadas de análisis de datos, para definir características importantes en los datos y así confirmar las variables que mejor predicen la probabilidad que un empleado con un rol definido, busque nuevo trabajo.
- Aplicar técnicas de definición de calidad de datos e inclusión de datos nuevos que pueden

reforzar o mejorar la predicción que puede cambiar en el tiempo.

- Implementar algoritmos ML de clasificación, comparando y optimizando sus hiperparámetros para encontrar el modelo que mejor realice predicciones de acuerdo con las métricas definidas.
- Desplegar el modelo en un entorno ejecutable en Cloud, con una API que permita consumir el modelo sin requerir de un experto en desarrollo.
- Aplicar una solución de entendimiento de lenguaje natural, leyendo comentarios de los empleados para confirmar su sentimiento o insatisfacción.
- Presentar resultados en un tablero, las predicciones entregadas por los modelos aplicados.

5. Plan de trabajo y presupuesto del proyecto

5.1 Plan de trabajo y distribución de tareas



1. Hito Anteproyecto:

a. Data Understanding:

- Comprensión del Negocio: En esta primera fase nos enfocamos en la comprensión de los objetivos y exigencias del proyecto.
- Comprensión de los datos: En esta etapa nos familiarizaremos con los datos, identificando los problemas de calidad, así como su estructura y las dependencias entre variables, correlación, etc. para así poder generar hipótesis.

Esta fase será llevada a cabo por todos los miembros del equipo, ya que es necesario que todos los integrantes entiendan el problema a resolver, y se pueda definir una primera estrategia entre todos.

b. Configuración de entorno de trabajo: Configuración entorno de trabajo colaborativo, Watson Studio, repositorio de Github, instalación de software (Anaconda, PyCharm), etc.

Cada miembro del equipo debe encargarse personalmente de tener todo correctamente configurado.

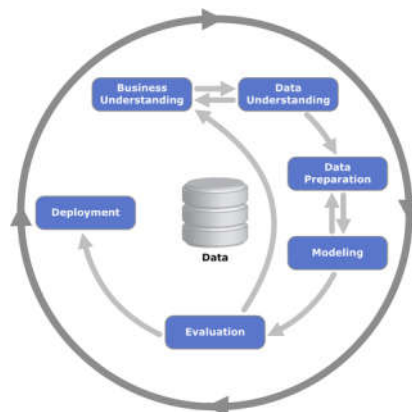
2. Hito Entrega Intermedia:

- a. Data Preparation: Técnicas aplicadas para la preparación y limpieza de datos que se va a necesitar para nuestro conjunto de datos final.
 - Limpieza de datos (tratamiento de nulos y outliers): Álvaro
 - Construcción de nuevas variables (Feature engineering): John
 - Estudio de importancia de variables, reducción de dimensionalidad: Marco
 - Integración de los datos para que sean consumibles por el modelo, creación de pipeline automatizada: Álvaro
- b. Modelling: Fases de entrenamiento del modelo, técnicas utilizadas y comparativa entre ellas.
 - Análisis de modelos de clasificación: Manuel y Antonio
 - Configuración de hiperparámetros hasta alcanzar los valores óptimos: Manuel y Antonio

3. Hito Entrega Proyecto:

- a. Evaluation: Pruebas para comprobar el rendimiento y efectividad del modelo.
 - Generación de métricas y gráficas donde se muestre la evolución de las variables e información relevante para el cliente para posterior inclusión en interfaz de la herramienta: Álvaro y Manuel
 - Evaluación de modelos anteriores: Manuel
 - Exploración de la calidad desde una perspectiva de análisis de datos: Manuel
- b. Deployment: Despliegue del modelo para ser consumido.
 - Puesta a punto del entorno de producción: John
 - Explotación de los modelos dentro del entorno de Producción: John
 - Frontend para la inclusión de datos y panel de control: Marco y Antonio

La planificación se ha realizado según el Modelo CRIPS-DM [8], que define el ciclo de vida del proyecto.



5.2 Presupuesto

Los recursos empleados para el proyecto son humanos (Data Scientist, Data Engineer, Web developer...), todos aquellos perfiles necesarios para el desarrollo de este.

Los recursos de software son la adquisición de una licencia IBM Cloud, para poder desplegar el proyecto, y herramientas para la planificación y estructura de los diferentes plazos del proyecto.

Recurso	Unidad	Precio (€)
Ordenador personal	5	5.000
Licencia Watson Studio IBM (Standard v1)	1	80,80
Data Scientist, Data Engineer, Web developer...	(150 h x 25€) 5	18.750,00
Licencia Gantt Project	1	0,00
Coste Total Proyecto €		23.830,80

6. Viabilidad y plan de recursos

6.1 Estudio de viabilidad técnica

La viabilidad técnica permite obtener la información necesaria respecto a si existe, o está al alcance, la tecnología necesaria para el sistema a implementar. La tecnología necesaria para la realización del proyecto se cita a continuación:

- Jupyter Notebook para el análisis
- IDE PyCharm para el desarrollo
- Python 3.x como lenguaje de programación del desarrollo
- Librerías para análisis, tratamiento y visualización de los datos:
 - Scikit-learn
 - Pandas
 - Numpy
 - Matplotlib
- GitHub como repositorio de código
- IBM Cloud Foundry como plataforma de despliegue
- IBM Watson Studio como gestor de modelos
- Página HTML como frontend

Todas las tecnologías mencionadas son herramientas open source, a excepción de la licencia Standard v1 de IBM Watson Studio, la cual entra dentro del presupuesto del proyecto. Por tanto, la tecnología necesaria para poder desarrollar el proyecto existe y está al alcance del equipo para poder desarrollar el proyecto.

6.2 Plan de recursos

El plan de recursos se divide en dos partes principales: el desarrollo del proyecto y el mantenimiento del servicio posterior.

6.2.1 Desarrollo del proyecto

Recursos técnicos	Coste estimado (€)
Hardware	
Personal computer (x5)	5.000
Licencias de software	
IBM Watson Studio (Standard v1)	80,80
IBM Cloudant	Free
Gantt Project	Free
Total	5.080,80

Recursos humanos	Esfuerzo estimado (horas)
Data scientist (x2)	350
HTML depelover	150
Data engineer	100
Test automation specialist	150
Total	750

6.2.2 Mantenimiento del servicio

Recursos técnicos	Coste estimado (€/mes)
Licencias de software	
IBM Watson Studio (Standard v1)	99
Total	99

Recursos humanos	Esfuerzo estimado (horas/mes)
Data scientist	4
HTML depelover	2
Data engineer	5
Test automation specialist	5
Total	16

6.3 Estudio de viabilidad económica

El presupuesto inicial es de 20.000€ para la puesta en producción de la aplicación. A continuación, se incluye el detalle del coste estimado de puesta en producción:

Recursos técnicos	Coste estimado (€)
Personal computer (x5)	5.000
IBM Watson Studio (Standard v1)	80,80
Data scientist (x2)	350h x 25€/h = 8.750
HTML depelover	150h x 25€/h = 3.750
Data engineer	100h x 25€/h = 2.500
Test automation specialist	150h x 25€/h = 3.750
Total	23.830,8

En cuanto al servicio postproducción, supondría el siguiente coste mensual:

Recursos técnicos	Coste estimado (€)
IBM Watson Studio (Standard v1)	99,00
Data scientist (x2)	4h x 25€/h = 100
HTML depelover	2h x 25€/h = 50
Data engineer	5h x 25€/h = 150
Test automation specialist	5h x 25€/h = 150
Total	549,00*

*Precios del mantenimiento del soporte. Para peticiones de cambio en la aplicación/modelo, será necesaria una estimación y aprobación de costes.

7. Referencias

- [1] *Master UEM. Inteligencia Artificial: Tema 1. Introducción a la Inteligencia Artificial.*
- [2] S. Advice, «estado-del-arte-de-la-inteligencia-artificial-en-2021,» [En línea]. Available: <https://www.sage.com/es-es/blog/estado-del-arte-de-la-inteligencia-artificial-en-2021/>.
- [3] Thinkopen, «Artificial Intelligence: state of the art,» [En línea]. Available: <https://www.thinkopen.it/en/project/artificial-intelligence-state-of-the-art/>.
- [4] I. Analítica, «MODELOS DE PROBABILIDAD DE FUGA DE EMPLEADOS,» [En línea]. Available: <https://inteligencia-analitica.com/modelos-probabilidad-fuga-empleados/>.
- [5] p. L. (interempresas.net), «Inteligencia Artificial contra el absentismo laboral,» [En línea]. Available: <https://www.interempresas.net/Proteccion-laboral/Articulos/262453-Inteligencia-Artificial-contra-el-absentismo-laboral.html>.
- [6] Expansion, «El algoritmo que revela cuándo un empleado quiere irse,» [En línea]. Available: <https://www.expansion.com/emprendedores-empleo/desarrollo-carrera/2015/03/19/550b1a74ca4741c5688b4575.html>.
- [7] Revistapym, «¿Pueden las empresas predecir la fuga de sus clientes?,» [En línea]. Available: <https://revistapym.com.co/pueden-las-empresas-predecir-la-fuga-clientes>.
- [8] B. Smartup, «Crisp-DM: los 6 pasos del proceso de Data Mining,» [En línea]. Available: <https://blog.smartup.es/crisp-dm-6-pasos-proceso-data-mining/>.