**Memòria tècnica**

**Cupons Indústria 4.0 (2021)**

**(EMC/778/2021)**

**Sistema expert amb Intel·ligència Artificial per automatitzar el procés de monitorització, captura i selecció de la normativa afecta al procés de HSE Compliance**

**Cupó Diagnosi Indústria 4.0**

**Cupó Implantació Indústria 4.0**

Asesores Corporativos S.A (ASECORP)



Nom i data: Xavier Dominguez, 24/3/2021

# Empresa

* Descripció de l’empresa

ASECORP és una consultoria especialitzada a optimitzar i garantir el compliment legal en matèria de medi ambient, prevenció de riscos laborals i seguretat industrial. Per a això hem dissenyat una base de dades, www.asecorp-online.com, que constitueix la clau i fet diferencial del nostre servei.

* Descripció de la situació actual i àmbit de l’actuació.

Actualment la BBDD, que és el actiu principal de la companyia, s’alimenta de manera manual amb un procés en el que intervé un equip de 15 persones que revisen cada dia els 23 sites que corresponen a els butlletins a monitoritzar, repassen la totalitat de disposicions, reglaments, anuncis, lleis, etc. que es publiquen diàriament per tal de detectar aquells que són objecte de l’interès del nostre àmbit d’operació, de manera que puguem completar i tenir actualitzada la nostra BBDD amb l’objecte d’assessorar adequadament als nostres clients.

Aquesta BBDD és un repositori que atresora tota la normativa identificada en un període de 20 anys i conté més de 40000 registres.

* Identificar reptes de futur.

El repte d’aquest projecte, es centra en replicar de manera automàtica i amb la mínima assistència humana el procés anteriorment descrit.

La implementació proposada s’abordarà amb tècniques del processament del llenguatge Natural, abreujat NLP. Aquesta és un àrea de la Intel·ligència artificial que no sols es centra en la comprensió del llenguatge sinó a aspectes cognitius generals humans. En el nostre cas s'emprenen models de detecció de expressions regulars (REGEX) per a establir les coincidències entre les normes publicades i les d’interès contingudes en la BBDD.

# Servei subvencionable

# Servei

* Descripció del servei sol·licitat a contractar.  
  + Abast.   
    El projecte es centrarà en el procés concret de la revisió dels 23 butlletins i la identificació de les normes que son objecte del servei d’ASECORP.
  + Planificació.

Les actuacions corresponents a la fase d’implementació, està previst que comencin a finals de Març del 2021, donat que la fase de prevalidació de la proposta ja està tancada i validada. Aquesta actuació es preveu que s’extengui fins a finals de Juny.

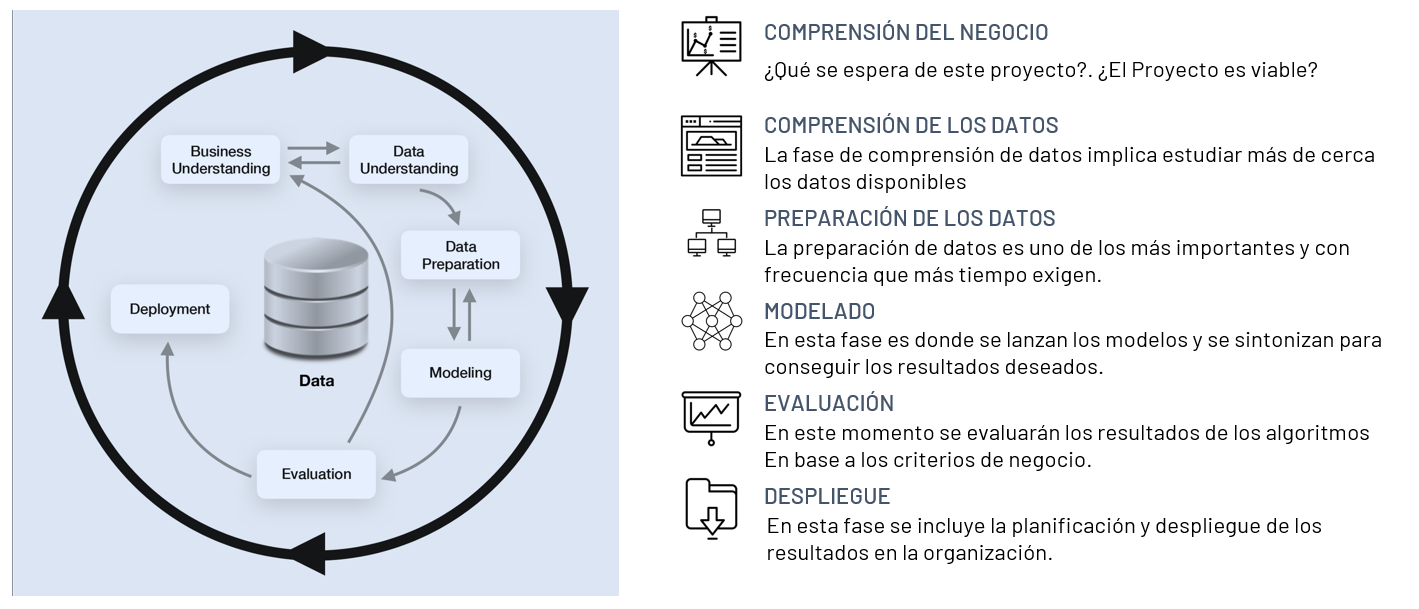
* + Objectius del servei (qualitatius i quantitatius).

L’objectiu principal és aconseguir replicar de manera automàtica i amb la mínima assistència humana el procés anteriorment descrit, per tal de lliurar d’aquesta tasca administrativa i rutinària als nostres consultors i que puguin augmentar el seu temps de dedicació a la digitalització de la empresa.

* Descripció del projecte on s’emmarca el servei sol·licitat i principals reptes empresarials a abordar.

Com s’ha exposat anteriorment, el projecte consisteix en la implementació d’un sistema de rèplica de manera automàtica i amb la mínima assistència humana en el procés de detecció d’aquelles normes publicades que són objecte de l’interès del nostre àmbit d’operació, de manera que puguem completar i tenir actualitzada la nostra BBDD.

La estructura d’un projecte d’aquest tipus es pot definir de la següent manera

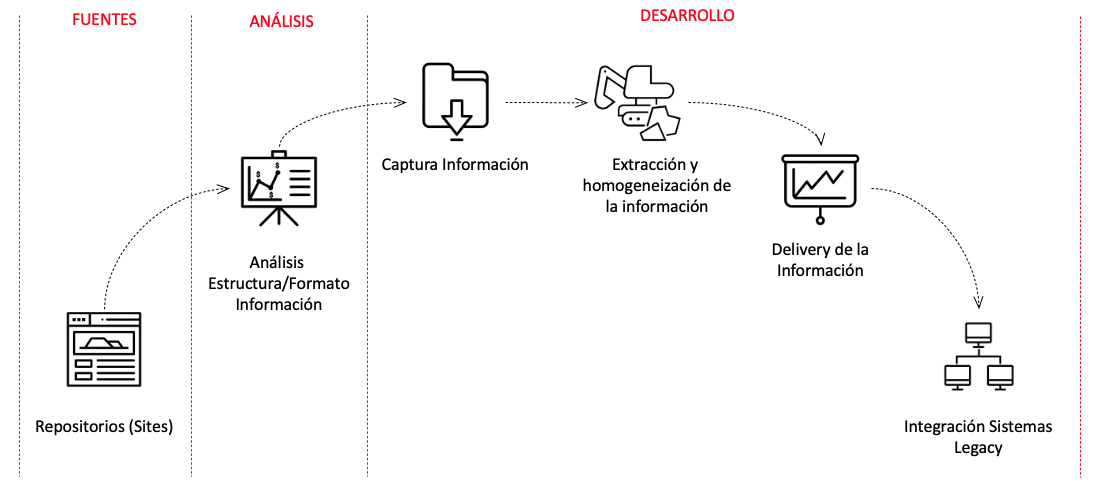


En el nostre cas el projecte es centrará a completar les quatre fases finals del procés representat, donat que s’ha fet un projecte anterior de validació en el que s’ha desenvolupat la metodologia necessaria per automatitzar el procés i s’han definit les regles d’assignació automàtica.

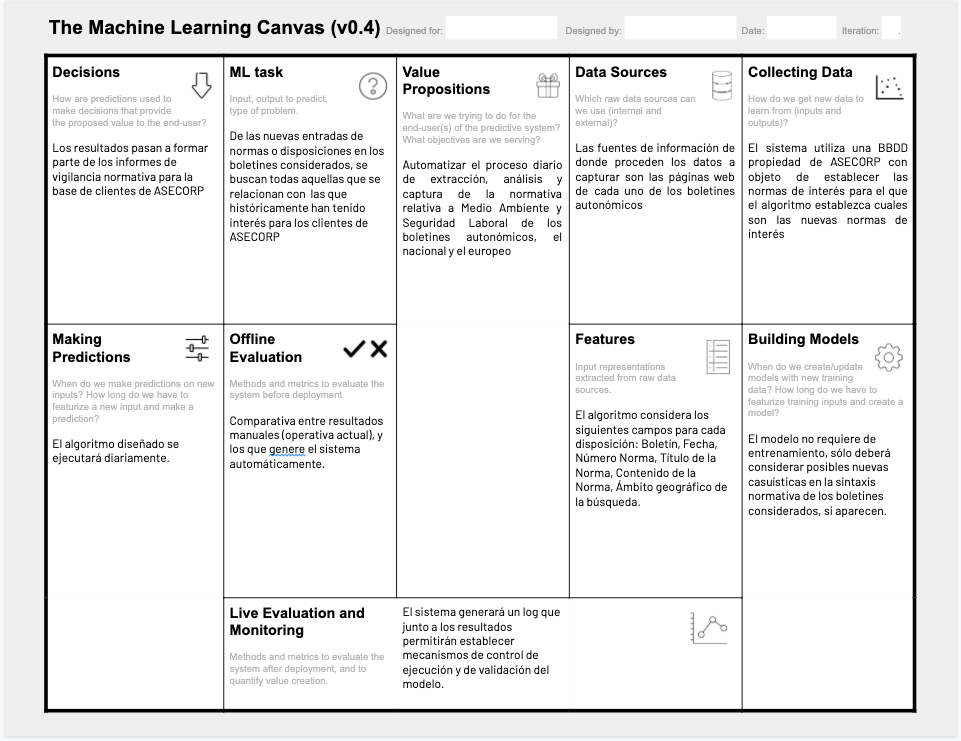
La primera complexitat del procés és la de capturar de manera automàtica tota la informació continguda en els 23 butlletins a monitioritzar, atès que l'estructuració i el format de lliurament de la informació són heterogenis, havent de treballar amb els formats PDF, XML, HTML i Javascript.

A partir de la captura e indexació de la informació, el mecanisme d’acció incorporarà suficient intel·ligència com per a identificar i proposar aquelles disposicions que siguin d'interès per a ASECORP d'entre totes les publicades a cada moment.

D'altra banda, no sols ha d'indexar-se la informació capturada en els sites on es publiquen els butlletins, sinó que la BBDD o repositori de ASECORP on es continguda la informació identificada en el passat també ha d'indexar-se, atès que és la que conté la informació que permetrà discriminar que informació és rellevant o no.

****

En definitiva, el projecte consistirà a programar els sistemes de scraping de cadascun dels butlletins, consolidar la informació i sintonitzar l'algorisme de IA per a implementar el procés automàtic. A banda d’aixó s’hauran de crear el mecanismes d’entrega e integració dels resultats amb la BBDD. El projecte s’enmarca dintre de les eines de Dades i Connectivitat, i Inteligència perque el projecte contempla la utilització de tècniques de Big Data així com d’Inteligència Arrtificial.

****

* Justificació de la necessitat i oportunitats per l’empresa del servei sol·licitat.

La digitalització de certes tasques dins del nostre procés permetrà principalment dos beneficis: un l’alliberament de temps de les tasques més administratives i rutinàries dels consultors i un increment en l’ eficiència de la realització d’aquestes tasques, estandarditzant el procés i reduint errors.

* Resultats esperats i indicadors del projecte.

Donat que el procés actualment s'executa de manera manual per un equip de moltes persones, els beneficis són clars:

* Eficiència, ja que passem a un procés automàtic sense intervenció humana.
* Homogeneïtat, atès que una vegada establertes les regles de selecció i sintonitzat l'algorisme l'execució no contindrà les discrepàncies i biaixos que inevitablement introdueix el factor humà
* Precisió, atès que el procés de captura i registre de la normativa es produeixen errors que podran evitar-se.
* Increment Valor afegit, ja que els consultors queden alliberats d'aquesta tasca i poden emprar aquest temps a aportar major valor als clients.
* S’estima que amb la implementació d’aquest sistema s’alliberaran 636 hores anuals dels nostres consultors dedicades exclusivament a aquestes tasques. Aquest temps es podrà destinar a altres tasques que aportin més valor al nostre servei personalitzat i col·laborant amb conjuntament amb els nostres proveïdors a la digitalització de l’empresa.
* **Només** en el cas de que l’ajut sol·licitat sigui per la **categoria d’Implantació Indústria 4.0** cal **aportar la següent informació:**
* Descripció de les principals línies d’actuacions contemplades en el pla de transformació digital de la empresa sol·licitant del que deriva aquesta actuació i previsió d’actuacions a mig termini que es derivin del projecte sol·licitat.
* Les principals línies d’actuacions contemplades en el pla de transformació digital de la empresa son les següents:
* Digitalització del procés de l’entrada de la normativa a la base de dades d’ASECORP ON LINE. Aquest procés es divideix en diferents fase: lectura i identificació de les normes que són objecte del nostre servei, anàlisis de la norma i identificació dels requisits legals i enviament d’alertes.
* Digitalització de les formacions que es realitzen a ASECORP.
* Digitalització de la comunicació d’ASECORP.

El servei sol·licitant formaria part de la línia d’actuació de la digitalització del processos que estan relacionats amb la base de dades d’ASECORP ON LINE. En concret, amb la fase de la lectura i identificació de les normes que son objecte del nostre servei.

* Justificació dels beneficis que aportarà el projecte d’implantació acord amb l’estratègia de transformació digital definida prèviament per l’empresa.

La implementació d’un sistema de rèplica de manera automàtica i amb la mínima assistència humana en el procés de detecció de normes que són objecte de l’interès del nostre àmbit d’operació, és el primer pas per desenvolupar una digitalització de tot el procés d’alimentació de la nostre base de dades.

* Descripció del repte principal i reptes associats que pretén resoldre la empresa amb la solució que proposa.

El principal repte d’ASECORP és implantar una digitalització en el seu servei sense perdre la seva personalització. ASECORP ofereix un servei i tracte principalment personalitzat que proporciona gràcies al seu equip de consultors. Existeixen altres serveis que estan totalment automatitzats i no tenen aquesta principal característica. ASECORP vol intensificar la seva digitalització del servei atacant principalment aquells processos on no és necessari que la feina sigui personalitzada i que a més pot aportar una millora en la eficiència del procés. Aquesta eficiència vindria donada la reducció del temps dedicat a aquesta tasca i la disminució d’errors comesos.

Descripció de la tecnologia principal i tecnologies associades a Indústria 4.0 que es desenvolupen en el projecte.

Dins d’aquest projecte com s’ha esmentat anteriorment es desenvoluparà tècniques del processament del llenguatge Natural, abreujat NLP, dins l’àrea de la Intel·ligència artificial. En el nostre cas s'emprenen models de detecció de expressions regulars (REGEX) per a establir les coincidències entre les normes publicades i les d’interès contingudes en la BBDD.

* Sectors d’aplicació de la solució a implementar.

La solució proposada és d’aplicació al sector de la consultoria de conformitat i altres relacionades, que requereixin del coneixement actualitzat de normatives legals publicades.

# Proveïdor/s subcontractat/s i pressupost del servei

* En aquest apartat cal fer una descripció dels proveïdors **(\*)** que realitzaran el servei sol·licitat. En el cas dels assessors acreditats per ACCIÓ cal indicar el **nom complert** de l’assessor.

***(\*) Important:*** *cal tenir present que la subcontractació dels serveis cal fer-la a través d'algun d'aquests proveïdors acreditats per ACCIÓ:*

* + - *Agents tecnològics amb el segell TECNIO (*[*+ informació*](https://www.accio.gencat.cat/ca/serveis/processos-acreditacio/tecnio/)*)*
    - *Assessors acreditats per ACCIÓ en l'àmbit sectorial RIS3CAT d'Indústria 4.0, i en almenys una de les següents àrees d’expertesa: estratègia, tecnologia, operacions i organització (*[*+ informació*](https://www.accio.gencat.cat/ca/serveis/processos-acreditacio/assessors/)*)*

Per **cadascun dels proveïdors subcontractats** que realitzaran el servei sol·licitat cal descriure la següent informació:

* Descripció del proveïdor..

Nom i cognoms de l’assessor: Alfredo Arias Santaella

Facturarà com autònom?:

SI;

NO (si està en nòmina d’una empresa proveïdora de serveis, indicar el nom de l’empresa a través de la que facturarà el servei: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_)

* Justificació de la necessitat i idoneïtat del proveïdor i descripció de les activitats que realitzarà en el marc de l’actuació subvencionada.
* L’assessor ha d’estar acreditat per ACCIÓ en una de les següents àrees d’expertesa: (Estratègia / Tecnologia / Operacions i organització) i en l’àmbit RIS3CAT (Indústria 4.0). Per més informació consulteu “[processos d’acreditació](https://www.accio.gencat.cat/ca/serveis/processos-acreditacio/assessors/)”. Si el proveïdor està acreditat per ACCIÓ en i amb data d’acreditació vigent en el moment de fer la present sol·licitud, marqueu “SI”. En cas contrari indiqueu “NO”.

SI

NO

* Explicació i justificació del pressupost sol·licitat.

El projecte d’implementació consistirà en el disseny dels algorismes de detecció de correspondències entre les disposicions d’interès del client ja recollides a la base de coneixement que forma part de la seva proposta de valor, i les mencions a les disposicions, normes, o anuncis que van apareixent diàriament als 21 butlletins a considerar entre els del estat espanyol, les diferent comunitats, i la unió europea.

El “breakdown” de activitats del projecte inclou 4 grans capítols, que son els que s’esmenten a continuació:

1. Recollida i processament de la BBDD del client: aquest capítol fa referencia a la recollida de la informació de la base de coneixement del client en el format adequat, la estandardització de la informació de manera que sigui possible la cerca de correspondències en les noves disposicions publicades d’acord a la tipologia sintàctica feta servir de manera normativa en aquestes publicacions, i la indexació mitjançant la generació de “TAGS” corresponents a la identificació de les disposicions, normes o anuncis d’interès.
2. Recollida i processament dels 21 butlletins: en aquest grup de tasques l’objectiu és recollir de manera personalitzada a cada butlletí la informació d’interès, fer la indexació generant els “TAGS” de les disposicions esmentades, i cercar la correspondència als registres de la base de coneixement del client. Els reptes a aquesta fase rau en la captura de la informació dels butlletins de manera automàtica, accedint a les seves planes web, i mitjançant algorismes de ‘Scraping’ per la recollida de dades que es presenten de maneres molt diverses: planes web estàtiques amb formats estructurats tipus HTML o XML, planes web dinàmiques que generen la informació en el moment del consum fent servir tecnologies JavaScript, i que requereixen d’estratègies de “Scraping” més avançades, o entrega de la informació en formats no estructurats tipus PDF que també compliquen la captura i estructuració de la informació recollida. Una vegada recollida la informació cal normalitzar-la i cercar la correspondència dels TAGS trobats amb la seva aparició en la base de coneixement del client. La estratègia que millor funciona per fer aquesta correspondència de manera unívoca, es la de la utilització de expressions regulars “REGEX”, i és per això que es fonamental la normalització de la base de coneixement a les formes sintàctiques fetes servir als diferents butlletins.
3. Entrega de resultats: Generats els resultats, s´ha de gestionar l’entrega de resultats per tal de minimitzar la intervenció humana en tot el procés. En aquesta fase, s’haurà de gestionar l’execució periòdica dels algorismes, i dissenyar la interfície de la sortida de resultats amb els sistemes de informació del client.
4. Validació i ajustos finals: Donada la varietat de fonts d’informació que intervenen en el projecte, i el fet que l’èxit del projecte depèn en gran mesura de la precisió de detecció de les disposicions d’interès publicades, caldrà estar molt atent a la generació de resultats durant un cert temps comparant la cerca manual amb l’automàtica dels algorismes per tal de detectar casuístiques que corregir.

|  |  |  |  |
| --- | --- | --- | --- |
| Definició de les activitats subcontractades | Breu descripció | Hores previstes | Import (€) |
| Activitat 1 - Captura BBDD ASECORP | Captura, anàlisi, i estandardització de la BBDD entregada per ASECORP. | 16 hores | 1.280 € |
| Activitat 2 – Processament dels 21 butlletins | Implementació dels processos de captura, anàlisi i cerca de correspondències de les disposicions trobades, a la BBDD de ASECORP, per cadascun dels 21 butlletins d’interès. | 21 butlletins x 8 hores = 168 hores | 13.440 € |
| Activitat 3 - Entrega resultats | Definició, e implementació de mecanismes d’entrega del resultats generats. | 20 hores | 1.600 € |
| Activitat 4 - Validació final | Validació i ajustos finals. | 40 hores | 3.200 € |
|  |  | **(30, 5 dies) 244 hores** | **19.520 €** |

El pressupost sol·licitat s’emmarca dintre dels projectes d’Industria 4.0 dels Cupons d’Innovació i Estratègia 2020. L’import total sol·licitat és de 19.520€, contant un cost de 80€/hora, desglossat de la següent manera en relació a les activitats subcontractades:

* Per la **categoria d’Implantació Indústria 4.0**, que permet la contractació de més d’un proveïdor acreditat (si el projecte ho requereix), caldrà definir un quadre d’activitats subcontractades per proveïdor participant, en funció de les actuacions a desenvolupar per cadascun d’ells.

**Annex 1 - Breu descripció del cas a implementar**

Nota. A diferència de la informació facilitada prèviament en aquesta memòria sobre el projecte (considerada de caràcter confidencial a efectes d’ACCIÓ i només per a la finalitat d’avaluació tècnica del projecte), per a la categoria de cupons d’Implantació Indústria 4.0, la informació sobre el projecte que l’empresa beneficiària faciliti en aquest Annex 2,  serà utilitzada amb finalitat de difondre públicament el cas d’ús que plantegeu en aquesta actuació d’implantació, prèvia aprovació dels continguts finals per part de l’empresa beneficiària abans de la difusió final. L’objecte, és considerar-la com una mesura més de sensibilització a altres pimes que s’estiguin plantejant iniciar aquest camí de la transformació digital perquè puguin conèixer diversos casos d’ús i oportunitats. I arribat el moment, i un cop finalitzada la vostra actuació, disposar també del vostre cas d’èxit en el portfolio de casos d’èxit en Indústria 4.0 d’ACCIÓ. Com empresa, vostès també us podríeu beneficiar de la resta de casos d’ús que difondrem i pels quals estem demanant també la corresponent autorització.

Títol del projecte

**Automatització del procés de detecció de disposicions normatives d’interès, als butlletins oficials de l’estat espanyol, les comunitats, i la unió europea**

Repte Principal (Extensió màxima 2 línies)

* ***Disseny de un programari de cerca i detecció de disposicions normatives d’interès en butlletins oficials.***

Reptes associats (Extensió màxima 1 línia per punt)

* ***Normalització de la base de coneixement de referencia***
* ***Disseny de la interfície de entrega de resultats als sistemes d’informació del client***

Descripció de la solució (Extensió 10 línies)

Disseny d’un algorisma de detecció de correspondències entre les disposicions normatives d’interès del client, que ja hi son recollides a la base de coneixement que forma part de la seva proposta de valor, i les mencions a les disposicions, normes, o anuncis que van apareixent diàriament als diferents butlletins a considerar entre els del estat espanyol, les diferent comunitats, i la unió europea.

Focus Principal (Extensió màxima 2 línies)

* ***Solució automatitzada per incrementar la eficiència del procés manual actual***

Focus associats (Extensió màxima 1 línia per punt)

* ***Alliberament dels consultors per dedicar-hi el seu temps a feines de més valor afegit***
* ***Repetibilitat o homogeneïtat que elimini el biaix que necessàriament introdueix la intervenció humana***

Beneficis esperats (Extensió màxima 1 línia per punt. Es recomana aportar els KPI)

* ***Reducció temps dedicació per consultor a descobrir noves normatives. KPI: grau d’automatització del procés (ara és 0 i ha de tendir al 100%).***
* ***Reducció temps transcorregut entre publicació de nova normativa i entrega a client. KPI : velocitat del procés (ara son hores i ha de tendir a minuts).***
* ***Increment normatives descobertes respecte a les normatives per descobrir. KPI: eficiència del procés (ara és alt a costa de molta dedicació de consultors però hauria de tendir al 100% amb una molt baixa dedicació de consultors).***

Tecnologia Principal (Extensió màxima 2 línies)

* ***IA***
* ***Dig Data /Data Analytics***

Altres tecnologies associades (Extensió màxima 1 línia per punt)

* ***Python***
* ***Natural Language Processing (NLP)/Regex***

Descripció del proveïdor (Extensió 10 línies)

Consultor acreditat per ACCIO amb àrees d’expertesa en Tecnologia i Industria 4.0