

Grau en Enginyeria Informàtica de Gestió i Sistemes d'Informació

**DORA: Desenvolupament d'un prototip pel diagnòstic automàtic de
l'esquistosomiasi basat en Transformers**

Estudi de la viabilitat

Antoni Jordi Noguera Recasens

TUTOR: Elisa Sayrol Clois

5è any - Doble Grau Enginyeria Informàtica de Gestió i Sistemes de Informació i Disseny i
Producció de Videojocs

Índex

1.	Planificació	1
1.1.	Planificació inicial	1
1.2.	Desviacions.....	2
2.	Anàlisi de la viabilitat tècnica.....	3
2.1	Anàlisi inicial	3
2.2	Anàlisi posterior	3
3.	Anàlisi de la viabilitat econòmica	5
3.1.	Costos de producció. Pressupost.....	5
3.1.1	Infraestructura.....	5
3.1.2	Hardware	5
3.1.3	Cost de personal	6
3.1.4	Pressupost final	7
4.	Anàlisi de viabilitat mediambiental	9
5.	Aspectes legals.....	11
6.	Gestió de la diversitat i perspectiva de gènere	13

Índex de figures

Fig. 1.1. Imatge del cronograma que el projecte continua realitzat amb un diagrama de

Gantt2

Índex de taules

Taula 3.1.1. Taula creada pel càlcul del pressupost final.....7

1. Planificació

1.1. Planificació inicial

Inicialment, el projecte, parteix d'un estudi teòric previ pel qual fan les parts tècniques i els coneixements requerits per a poder dur a terme el projecte. Com a pas inicial l'estudiant ha de fer els cursos i llegir la documentació requerida i esmentada en la metodologia. S'estima que aquest estudi previ que el desenvolupador ha de dur a terme ocuparà una durada de dos mesos com a màxim sent la data de començament el 19 de setembre de 2023 i la finalització el 20 de gener de 2024.

Posteriorment a l'estudi de la part teòrica del projecte, l'estudiant realitzarà la documentació teòrica requerida en aquest projecte en la qual podrà reflectir els coneixements recollits sobre la malaltia, els coneixements tècnics requerits i les implicacions ètiques amb el que, també, podrà fer unes conclusions respecte a com procedir en el projecte sent la data de començament el 25 de gener de 2024.

La següent passa a realitzar és el desenvolupament del projecte, i per això el laboratori amb el qual es fa conjuntament el projecte facilitarà a l'alumne l'accés a la base de dades amb les mostres amb la seva pertinent anotació per a poder realitzar l'entrenament del model de IA d'aquest TFG. Amb les dades, l'estudiant durà a terme el desenvolupament pertinent de l'eina. En aquesta fase, és on els esforços de seguiment i control seran més crítics, ja que el projecte es basa en aquesta fase i l'èxit d'aquest dependrà de l'èxit obtingut en aquesta fase.

Finalitzat el desenvolupament de la tecnologia, s'arribarà a l'última fase del projecte on es podrà veure reflectit l'anàlisi prèvia de viabilitat, l'èxit del projecte en la tasca assignada i la implementació realitzada final de la qual es poden extreure conclusions sobre el mètode i la reproductibilitat que té el projecte en qüestió a l'hora de voler fer noves aplicacions d'aquest amb diferents objectius, siguin dintre de l'àmbit mèdic o un altre.

Es facilitarà al client un prototip final amb el qual pot fer servir l'algorisme desenvolupat i pot comprovar el seu funcionament. Aquest prototip és presentat al client i aquest pot determinar continuar amb el desenvolupament d'aquest en col·laboració amb el desenvolupador o fer-ho ell mateix.

A la figura 6.1.1, es mostra un cronograma amb la planificació d'entrega i resolució del projecte en cada iteració seguint la metodologia explicada en l'apartat cinc.

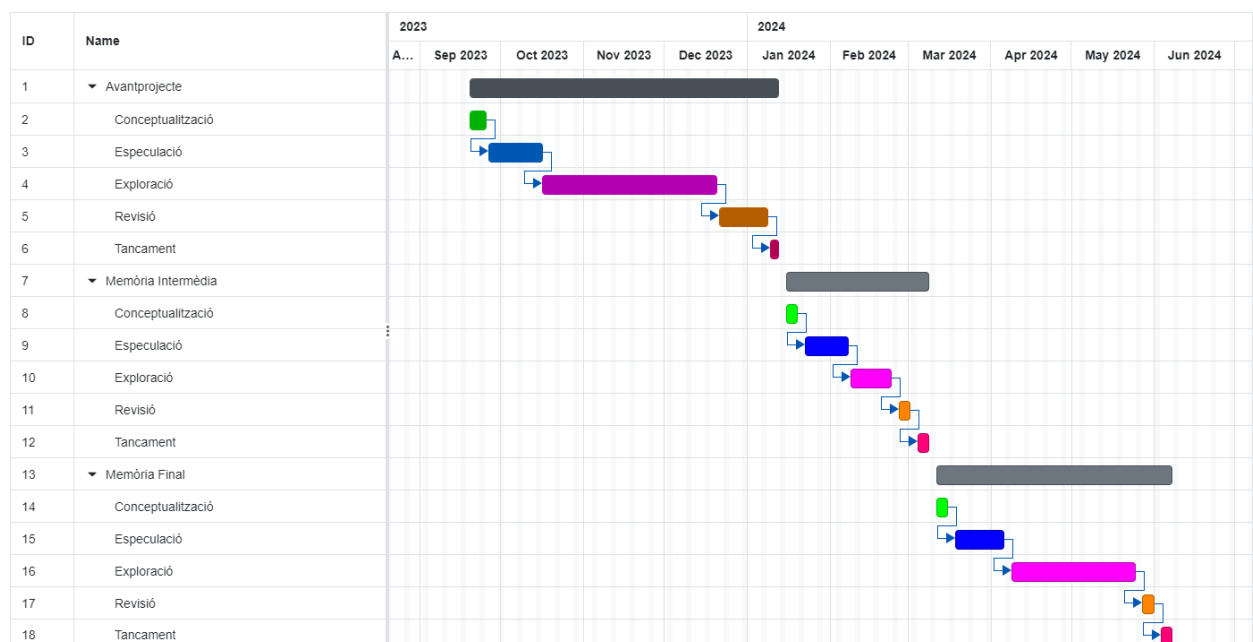


Fig. 1.1. Imatge del cronograma que el projecte continua realitzat amb un diagrama de Gantt

1.2. Desviacions

Durant la etapa del avantprojecte, seguint la planificació inicial, no hi va haver cap desviació sobre la planificació i es va poder seguir la guia.

Durant la etapa del desenvolupament de la memòria intermèdia, si que hi ha hagut desviacions sobre la idea original. En aquesta fase, s'ha focalitzat molt més la especulació sobre l'exploració, la qual ha quedat rellevada a un segon plànol i no s'ha pogut obtenir un resultat similar al esperat per a aquesta etapa. També, la revisió amb el laboratori no ha sigut possible de realitzar-la degut a la manca d'una versió de la solució capaç de poder fer una demostració de la seva tasca.

2. Anàlisi de la viabilitat tècnica

2.1 Anàlisi inicial

Els antecedents que es poden trobar sobre una aplicació semblant de la IA en malalties mitjançant la classificació d'imatges, poden demostrar que aquest producte és possible de realitzar. L'única qüestió que es podria posar en dubte és la viabilitat del desenvolupament d'aquest amb només un sol desenvolupador.

D'una banda, un dels antecedents al projecte desenvolupat pel mateix laboratori amb el qual es col·labora, va poder desenvolupar fent servir una aplicació similar de la intel·ligència artificial que es voldrà implementar en aquest i amb un sol desenvolupador al càrrec. Tenint en compte les capacitats requerides i el desenvolupament a realitzar, sent tan similar, no implica una gran dificultat. D'altra banda, l'ús dels models transformers és innovador i les referències indiquen la necessitat de grans bases de dades per millorar algorismes basats en CNN s'ha decidit prèviament estudiar si la base de dades permet l'ús d'aquests models. Finalment, es pot deduir que hi ha viabilitat en el desenvolupament d'aquest projecte per la part tècnica i, tot i que poden aparèixer inconvenients, no presenta una dificultat excedent que impedeixi el seu desenvolupament amb els recursos actuals.

2.2 Anàlisi posterior

Realitzada una implementació del model escollit per a la tasca i un estudi exhaustiu sobre les capacitats d'aquests i les seves capacitats de rendiment en comparació d'altres, es pot deduir que tècnicament el projecte es viable però es requereix d'una base sòlida en els coneixements de *computer vision*.

La necessitat d'aquests coneixements no són un impediment però sí que es tracta d'un requeriment obligatori previ al desenvolupament d'aquesta tasca. El projecte tècnicament parteix d'una documentació i d'un concepte nou i el qual continua en estudi i investigació però existeixen suficients recursos per a sostenir-lo i desenvolupar-lo.

Pel que fa la infraestructura requerida, com en l'anàlisi inicial i en la planificació inicial del projecte, es continua afavorint la plataforma Google Colab per realitzar la tasca encomanada.

3. Anàlisi de la viabilitat econòmica

Tenint en compte el pressupost calculat en l'anterior grup i la tendència del preu de venda que tenen productes similars basats en IA en el mercat podem veure que sí que hi ha una viabilitat econòmica en el desenvolupament del projecte a petició del client. El finançament del projecte de cara a una possible continuació i desenvolupament a llarg termini dependrà del laboratori de la Vall d'Hebron – el client – el qual pot decidir si continuar el desenvolupament posterior al prototip. Per tant, el projecte és viable a curt termini, és a dir, fins que el prototip sigui desenvolupat i a llarg termini pot arribar a ser viable també tot i que es depèn de la decisió del client.

3.1. Costos de producció. Pressupost

3.1.1 Infraestructura

Per la part d'infraestructura, es basarà en un servei al núvol el qual podrà proporcionar una infraestructura adequada pel desenvolupament i el funcionament de la tasca. En aquest cas, s'ha escollit fer servir Google Colab, una eina que permet la programació i execució de Python en un navegador i accés a GPUs sense cost.

A més, dona també l'opció d'escalar el projecte a preus com 11,19 € per 100 unitats computacionals i 51,12 € per 500 unitats computacionals els quals no són a subscripció sinó per ús i també ofereix l'opció d'una subscripció Pro d'11,19 € al mes. En un principi es dubta la necessitat de més unitats computacionals, però s'opta per escollir l'opció de la subscripció mensual.

Pel que fa a l'oficina des d'on es treballa, es té en consideració el cost del preu del lloguer el qual són 1.500 € al mes, incloent-hi costos de llum, internet i aigua. Tenint en compte la durada del projecte, uns sis mesos, el cost total que s'estima són d'uns 9.000 € per l'espai.

3.1.2 Hardware

El maquinari que es farà servir en aquest projecte no requereix una sèrie de requisits computacionals, ja que amb l'opció escollida en la infraestructura se supleix la manca que pugui tenir l'ordinador on es du a terme la tasca. L'opció més adient per a portar a cap el projecte en qüestió serà un portàtil Dell Inspiron 15. Aquest portàtil entra dintre dels

estàndards corporatius en empreses informàtiques i disposa d'un maquinari adequat per a la feina:

- Processador i5 de 12^a generació
- Windows 11
- Targeta gràfica Intel UHD
- Memòria DDR4 de 8GB
- SSD de 512 GB

Aquest portàtil té un cost de 498,99 € al portal oficial de l'empresa. És un portàtil de baix cost en comparació d'altres més potents, però serveix per a la tasca a realitzar en el projecte.

3.1.3 Cost de personal

D'una banda, en el cost personal es té en consideració el del desenvolupador. Aquest compleix el perfil de Data Analyst, el rol que sol ser un desenvolupador d'un model de IA i l'entrenament d'aquest. Un sou mitjà d'una Data Analyst és de 27.000 €/any pel qual són 12,98 €/hora durant 20 hores setmanals, el que finalment és en un total de 6.230,50 € en els sis mesos de projecte. D'una altra banda, el cost humà inclou també del tutor. Aquest compleix el perfil de Consultor Sènior, aquest rol ocupa l'aconsellament i la guia pel desenvolupador. El sou mitjà d'un Consultor Sènior és de 40.000 €/any pel qual són 19,23 €/hora durant 20 hores setmanals, el que finalment és un total de 9.230,40 € en els 6 mesos de projecte.

Tenint en compte el model empresarial, sent aquest el d'Autònom, a part del cost del treballador, es té en compte també els costos de l'empresa conseqüents d'impostos. La quantitat final a pagar per part de l'autònom sobre l'IRPF sobre els sis mesos de projecte és de 1.869,12 €.

3.1.4 Pressupost final

A continuació, es facilita una taula amb el càlcul dels costos del projecte:

Concepte	Quantitat	Unitat	Cost unitari	Cost mensual	Cost final
Costs Fixes					
Salari Consultor Sènior	1		1.538,40€	1.538,40€	9.230,40€
Salari Data Analyst	1		1.038,40€	1.038,40€	6.230,50€
Lloguer Oficina	1	Oficina	1.500€	1.500€	9.000€
Subtotal	3		4.076,80€	4.076,80€	24.460,90€
Costs Variables					
Subscripció Google Colab	1	Subscripció	11,19€	11,19€	67,14€
Ordinadors	1	Hardware	498,99€	-	489,99€
Subtotal	2		510,18€	11,19€	557,98€
Costs Totals	5	-	4.586,98€	4.087,99€	25.018,88€

Taula 3.1.1. Taula creada pel càlcul del pressupost final

En conclusió, el pressupost final que assolirà el projecte com a mínim és de 25.018,88 € com en el mostrat en la taula anterior. Aquest valor pot veure's afectat a causa d'un increment dels recursos requerits o una reducció en les hores requerides en el desenvolupament del projecte, es planeja revisar-ho en la memòria intermèdia. El cost plantejat és inferior al cost mitjà d'un projecte de desenvolupament de programari, el qual habitualment és igual o superior als 30.000 €.

4. Anàlisi de viabilitat mediambiental

Aquest projecte té una empremta de carboni que es pot estimar en 0,11 tones de carboni. Aquest càlcul ha sigut possible a una estimació dels costos i del consum mitjà per més que hi haurà durant els sis mesos de desenvolupament. En un primer moment, s'han escollit totes les opcions possibles que assegurin que el projecte tingui l'impacte mínim possible sobre l'empremta de carboni digital i, veient el resultat final dels càlculs, es pot dir que aquest projecte té un impacte mediambiental adequat i no s'excedeix en gran manera.

Pel que fa la prevenció d'averies que poden impactar negativament a la empremta de carboni del projecte i, per tant, un impacte major sobre el medi ambient, es segueix la prevenció de riscos i les mesures de seguretat que ofereix la plataforma de Google en l'ús de les seves màquines virtuals i el seu magatzem d'arxius en el núvol. Aquestes mesures es basen en el traspàs dels serveis al servidor més proper al del incident per tal de consumir la menor energia possible.

5. Aspectes legals

Els aspectes legals que s'involucren en el projecte, només hi ha dos dels quals es pugui aportar un relleu major. D'una banda, hi ha un acord de confidencialitat el qual que la base de dades amb la que es desenvoluparà, la qual és propietat del laboratori, no podrà ser filtrat el seu contingut. En aquest acord, tant el desenvolupador del projecte com totes les institucions que estan involucrades, han firmat i han acceptat aquestes condicions de desenvolupament. D'altra banda, la base de dades disposada pel laboratori de la Vall d'Hebron ha passat pel comitè ètic d'aquesta institució.

6. Gestió de la diversitat i perspectiva de gènere

El producte a desenvolupar no fa distinció de gènere ni de diversitat segons les necessitats de cada cas. En aquesta ocasió, es tracta d'un projecte que ajuda a diagnosticar amb més facilitat i eficiència una malaltia tropical la qual afecta una majoria de persones que no tenen els recursos necessaris per a poder tractar-la ni diagnosticar-la dintre de la seva capacitat econòmica. La majoria d'aquestes persones es troben en zones de l'Àfrica i, per tant, és on tindrà més impacte.

La solució basada en IA no tindrà un biaix de gènere conegut, ja que totes les dades són anònimes i no se'ns pot facilitar la persona darrera de cada mostra per a poder fer un entrenament equitatiu sense discriminació. Per part del laboratori, se'ns assegura que no hi ha cap biaix ni discriminació en les mostres.