



ACB Manager: plataforma per a l'assessorament en jocs de tipus Màner

Xavier Garcés Martí
Màster en Enginyeria Informàtica

Samir Kanaan Izquierdo

23 de Desembre de 2015



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

Llicències alternatives (triari alguna de les seqüents i substituir la de la pàgina anterior)

A) Creative Commons:



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-SenseObraDerivada 3.0 Espanya de Creative Commons](#)



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-CompartirIgual 3.0 Espanya de Creative Commons](#)



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial 3.0 Espanya de Creative Commons](#)



Aquesta obra està subjecta a una llicència de [Reconeixement-SenseObraDerivada 3.0 Espanya de Creative Commons](#)



Aquesta obra està subjecta a una llicència de [Reconeixement-CompartirIgual 3.0 Espanya de Creative Commons](#)



Aquesta obra està subjecta a una llicència de [Reconeixement 3.0 Espanya de Creative Commons](#)

B) GNU Free Documentation License (GNU FDL)

Copyright © ANY EL-TEU-NOM.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free

Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

C) Copyright

© (l'autor/a)

Reservats tots els drets. Està prohibit la reproducció total o parcial d'aquesta obra per qualsevol mitjà o procediment, compresos la impressió, la reprografia, el microfilm, el tractament informàtic o qualsevol altre sistema, així com la distribució d'exemplars mitjançant lloguer i préstec, sense l'autorització escrita de l'autor o dels límits que autoritzi la Llei de Propietat Intel·lectual.

FITXA DEL TREBALL FINAL

Títol del treball:	<i>ACB Manager: plataforma per a l'assessorament en jocs de tipus Mànager</i>
Nom de l'autor:	<i>Xavier Garcés Martí</i>
Nom del consultor:	<i>Samir Kanaan Izquierdo</i>
Data de lliurament (mm/aaaa):	<i>12/2015</i>
Àrea del Treball Final:	<i>Intel·ligència Artificial</i>
Titulació:	Màster en Enginyeria Informàtica
Resum del Treball (màxim 250 paraules):	
<p>L'objectiu d'aquest projecte és desenvolupar una plataforma web basada en la Intel·ligència Artificial per assessorar a jugadors de jocs de tipus <i>mànager</i>, més concretament el SuperManager ACB de bàsquet.</p> <p>La plataforma, serveix com a base per desenvolupar assessors similars per altre jocs, sigui de bàsquet o d'altres esports.</p> <p>S'ha fet un estudi de la regressió per tal de predir valoracions, i s'utilitzen algoritmes genètics de cerca per a l'optimització de plantilles de jugadors.</p> <p>S'utilitza Ruby on Rails (RoR) com a marc de treball, juntament amb diverses llibreries RubyGems:</p> <ul style="list-style-type: none">- LibLinear per a la generació de prediccions.- Una implemetació pròpia de l'algoritme genètic per a la optimització d'equips.- Nokogiri per a l'obtenció de dades mitjançant el rastreig de codi html d'altres pàgines.	

Abstract (in English, 250 words or less):

The aim of this project is to develop a web platform based on artificial intelligence to advise players of manager type games, specifically the SuperManager ACB basketball .

The platform serves as a basis for developing other similar games advisers either basketball or other sports.

It has made a study of regression to predict ratings, and it uses genetic algorithms for optimization of search templates players.

Using Ruby on Rails (RoR) as a framework, along with several libraries RubyGems :

- LibLinear for generating predictions.
- A itself to implementing the genetic algorithm for optimization of equipment.
- Nokogiri for obtaining data by tracking other html pages.

Paraules clau (entre 4 i 8):

Intel·ligència Artificial, regressió linial, prediccions, optimització, algoritmes genètics, jocs mànager

Índex

1. Introducció.....	1
1.1. Context i justificació del Treball.....	1
1.2. Objectius del Treball.....	1
1.3. Enfocament i mètode seguit.....	2
1.4. Anàlisi de Riscos.....	3
1.5. Pressupost.....	3
1.6. Planificació del Treball.....	4
1.7. Breu sumari de productes obtinguts.....	5
1.8. Breu descripció dels altres capítols de la memòria.....	6
2. Pla detallat del Projecte.....	7
2.1. Diagrama de Gant.....	7
2.2. Informe de Seguiment del Projecte.....	7
2.3. Pla de Gestió de Riscos.....	9
2.4. Resum d'incidències.....	10
3. Visió General.....	11
3.1. Entorn de Desenvolupament.....	11
3.2. Repositori de Dades.....	11
3.3. Font de Dades.....	12
3.4. Funcionalitats.....	13
3.5. Model de Base de Dades.....	16
4. Funcionalitats de la Plataforma.....	18
4.1. Funcionalitats Back-end.....	18
4.2. Funcionalitats Front-End.....	19
5. Prediccions i SVR.....	20
5.1. Dades Utilitzades.....	20
5.2. Normalització de Dades.....	20
5.3. SVM i SVR.....	21
5.4. LIBLINEAR.....	21
5.5. Estudi de Regressió.....	22
5. Optimització.....	24
5.1. Algoritme Genètic.....	24
6. Disseny i Publicació.....	26
6.1. Disseny.....	26
6.2. Publicació.....	27
7. Conclusions.....	28
8. Glossari.....	29
9. Bibliografia.....	30

Llista de figures

Imatge 1. Diagrama de Gantt	7
Imatge 2. Exemple d'estadística per al recull de dades	12
Imatge 3. Estructura de la pàgina inicial	14
Imatge 4. Pàgina detall de les prediccions d'un partit	15
Imatge 5. Model de Base de Dades, Versió 3.0	16
Imatge 6. Disseny Final de la Home	26
Imatge 7. Disseny Final de la fitxa de Prediccions	27

1. Introducció

1.1. Context i justificació del Treball

Avui dia els jocs on-line de tipus mànager per a esports són una tendència, cada temporada en surten de nous (especialment en el futbol) i amb ells també apareixen noves pàgines web o aplicacions que donen suport a aquest tipus de pàgines.

Entenent els jocs mànager en l'esport com aquells jocs en què el jugador ha de gestionar algun tipus de plantilla d'un esport determinat, en què els membres de la plantilla són jugadors de la vida real i d'equips reals.

Crec que aquest és el principal atractiu d'aquest tipus de jocs, que es depèn totalment dels jugadors reals (no de cap simulador) cosa que pot portar ha estar mirant un Levante - Getafe que ni et ve ni et va un diumenge a la nit, o pots desitjar que aquella jornada marqui Cristiano Ronaldo sent seguidor del Barça. És addictiu vaja.

A ram d'aquest tipus de jocs, hi ha cada dia més pàgines que els complementen, siguin fòrums per equips o jugadors on es fan prediccions del tipus, Messi (futbol) acostuma a marcar contra l'Atlético de Madrid, o Navarro (bàsquet) acostuma a jugar bé contra el Madrid), o pàgines on pots veure estadístiques o càlcul de punts. La majoria de jocs es financen gràcies a la publicitat en la mateixa pàgina.

Aquest projecte va encaminat a utilitzar tècniques d'Intel·ligència Artificial per a fer prediccions de valoracions de jugadors i optimització de plantilles per tal d'alinejar la millor relació valoració/punts.

En el meu cas em decantaria pel bàsquet, ja que, tot i que és menys utilitzat i té menys afició, només hi ha un joc a Espanya: el *SuperManager ACB*. Així doncs la plataforma utilitzarà les dades estadístiques de l'ACB per tal de fer prediccions i optimitzacions de plantilles.

1.2. Objectius del Treball

L'objectiu final del treball és obtenir una plataforma capaç de donar suport a jugadors d'un joc online de tipus mànager, en aquest cas concret el Supermanager ACB de bàsquet.

Dins d'aquest objectiu s'enumeren els diferents objectius per ordre d'importància:

- Implementar un sistema de predicció de valoracions, punts, rebots i assistències de jugadors per a cada jornada.
- Implementar un sistema d'optimització de plantilles en funció de les característiques establertes per les regles del joc Supermanager ACB.
- Capturar de dades estadístiques per partit, jugador, temporada.
- Analitzar els paràmetres i els atributs a l'hora de fer les prediccions a partir de la regressió, avaluant amb quins obtenim els millors resultats.
- Oferir una plataforma accessible i fàcil d'utilitzar per tal de guanyar la fidelitat dels usuaris/jugadors.
- Implementar una plataforma flexible per a poder reutilitzar la mateixa solució en altres jocs, per exemple de futbol.
- Integrar els coneixements que he adquirit durant el màster en relació a les tècniques de Intel·ligència Artificial.
- Aprofundir i consolidar coneixements en el desenvolupament en *RoR*.

1.3. Enfocament i mètode seguit

Aquest tipus de jocs, tal i com he comentat anteriorment, són molts freqüents avui en dia, i hi ha múltiples aplicacions per a mòvil i pàgines web que fan recomanacions en quant a compra de jugadors, alineacions, etc. Si bé en el fútbol hi ha moltes possibilitats, en el bàsquet les pàgines que hi ha són contades, i normalment són blogs on algun periodista o seguidor al bàsquet fa recomanacions, a partir del seu coneixement.

En cap dels dos esports hi ha cap plataforma que ofereixi dades a partir del tractament de dades amb tècniques d'Intel·ligència Artificial, així que l'estratègia que és seguirà serà la **diferenciació**, ja que s'ofereix un servei que actualment no existeix per a un tipus de joc que està en constant creixement.

Aquest any, al joc de SuperManager ACB hi ha registrats més de 150.000 equips, el que no vol dir que el nombre de jugadors sigui el mateix, ja que un mateix jugador pot tenir els equips que vulgui. Si d'aquests 150.000 jugadors es suposa que cada jugador té 10 equips de mitja, tenim que 15.000 usuaris cada setmana necessitaran informació per a fer els seus fitxatges o per a fer les seves alineacions.

A més, tenint en compte que el que interessa és generar visites i suposant que un usuari que tingui 10 equips, n'abandoni la meitat durant la temporada, igualment voldrà consultar informació dels seus 5 equips durant tota la temporada. Això suposa 75.000 consultes a la setmana.

Amb aquestes dades de visites es pot finançar perfectament la pàgina web mitjançant publicitat.

1.4. Anàlisi de Riscos

Tot i tractar-se d'un projecte peculiar, pels recursos dels que es disposa i al tractar-se d'un projecte acadèmic, hi ha una sèrie de riscos a tenir en compte i dels quals es farà el corresponent seguiment.

Codi	Descripció	Probabilitat	Impacte	Nivell
R01	No disposar de les dades adequades ni suficients per aplicar les tècniques de regressió i optimització.	Baixa	Alt	Mig
R02	Escollir una plataforma no adequada per a treballar amb tècniques de regressió i optimització.	Baixa	Mig	Baix
R03	Els resultats de les prediccions són incoherents i no ofereixen als usuaris informació útil per al seu joc.	Mitja	Mig	Mig
R04	La plataforma és insuficient i l'usuari no pot accedir a la informació de forma ràpida i fàcil.	Baixa	Mig	Baix
R05	La pàgina web no està ben estructurada a nivell SEO i no rebem visites a Internet, som invisibles per als usuaris. ¹	Mitja	Alt	Baix
R06	No disposar del temps necessari per finalitzar el treball, per factors inesperats, com per exemple malaltia o feina.	Baixa	Alt	Mig

Es pot notar que la última columna, nivell, fa referència a la relació entre la probabilitat i l'impacte del risc, i que el color de la cel·la correspon al què hi ha escrit. Es pot veure doncs que no hi ha un risc amb nivell alt, és a dir amb alta probabilitat i impacte alt, així que principalment s'estarà alerta als riscos amb nivell mig.

1.5. Pressupost

Per al desenvolupament del projecte no serà necessària cap tipus d'inversió econòmica, ja que el programari que s'utilitza tot és lliure o s'utilitza la versió gratuïta.

En cas que aquest s'acabés publicant a Internet sí que s'haurien de tenir compte les despeses del servidor per fer funcionar *RoR*. Hi ha múltiples opcions, on les més econòmiques van d'entre 5 i 10 euros al més, i poden arribar a 400 euros al més quan estem parlant de pàgines amb gran volum de tràfic.

¹ El posicionament no és un aspecte que entri dins del projecte, tot i això ho menciono com a risc per tenir en compte que en el desenvolupament de la plataforma s'ha de tenir en compte aquest aspecte.

Personalment em decantaria per *Heroku*, que té un preu d'una mica més de 6 euros al més, això vol dir un total de 76 euros l'any.

Si es volgués disposar d'una llicència de *Github*, el preu seria el mateix, 76 euros l'any.

1.6. Planificació del Treball

La planificació del projecte, en aquest cas, vindrà donada per les entregues de l'assignatura que ja vénen marcades amb unes dates determinades.

Seguint una metodologia *Agile* com *Scrum*, i tenint en compte que no es disposa de client i que els principals actors són un mateix i el consultor, es dividiran les entregues que són aproximadament cada mes, en esprints d'un mes dividits en 4 setmanes-tasques cada un, i així és com s'anomenaran els següents apartats posteriors a la planificació.

Així doncs, a continuació es presenta la taula de fites, que s'ha actualitzat, que s'ha anat actualitzant en les entregues inicials, i on es poden identificar les fites parcials de cada entrega.

Data	Setmana	Descripció
29/09/2015	0	Proposta inicial del treball: Introducció, Objectius, Abast, Estudi de Mercat, Planificació, Font de dades, Plataforma
30/09/2015		Entrega PAC 1
07/10/2015	1	Instal·lació i configuració de Plataforma <i>RoR</i> . Creació del Projecte Web.
14/10/2015	2	Selecció de llibreries a utilitzar.
21/10/2015	3	Obtenció de dades en el format correcte i classificació. Definició de les tècniques d'Intel·ligència Artificial a utilitzar per a les prediccions i optimitzacions.
27/10/2015	4	Estructura de la web i modelatge de la base de dades del Joc. Definició de l'estructura de la plataforma web (opcions de menú, disseny, opcions que s'oferiran, etc.)
28/10/2015		Entrega PAC 2
08/11/2015	5	Aplicació de les tècniques de regressió per prediccions.
15/11/2015	6	Funcionalitats back-end: - Actualització de dades.

		- Actualització de resultats.
18/11/2015	7	Funcionalitats front-end: - Visualització de partits i partits a la home. - Visualització de prediccions de partits.
24/11/2015	8	Test i Documentació Planificació de l'anàlisi de tècniques de regressió.
25/11/2015		Entrega PAC3
02/12/2015	9	Aplicació de les tècniques d'optimització en equips per valoració, punts, rebots i assistències.
09/12/2015	10	Estudi de Regressió.
16/12/2015	11	Funcionalitats front-end: - Compra/venda de jugadors. - Configuració de plantilles. - Registre d'equips i usuaris (login)
22/12/2015	12	Test + Documentació + Anàlisi de Resultats
23/12/2015		Entrega PAC 4
30/12/2015	13	Publicació de les plataformes (si s'escau).
05/12/2015	14	Presentació Power Point.
06/01/2016		Presentació

Més endavant es tractarà amb més detall la planificació del projecte.

1.7. Breu sumari de productes obtinguts

Com a resultat del projecte s'obindrà una plataforma web capaç de:

Pel què fa a l'administració (bak-office):

- Captar dades estadístiques d'anys anteriors de la lliga ACB de bàsquet.
- Captar dades estadístiques cada setmana (de cada jornada).
- Calcular o capturar el valor dels jugadors (preu) que anirà variant en funció de les seves estadístiques.

Quant al que què podran accedir els usuaris:

- Consulta de prediccions de valoracions/punts/rebots/assistències per partit i jugador de cada setmana.

- Assessorament de configuració de plantilles obtenint la més òptima en funció de les valoracions/punts/rebots/assistències resultants de les prediccions i de les mitges de la temporada fins al moment.
- Registre d'usuari i creació d'equips.
- Configurador de plantilles.

1.8. Breu descripció dels altres capítols de la memòria

Si bé en aquest primer capítol es presenta una de quin és el motiu del projecte, els objectius, l'enfocament i els resultats que s'esperen, i una mica per sobre la planificació; en els capítols que vénen es tractaran dos aspectes principalment:

- Per una banda en el capítol 2 s'aprofundeix en la planificació del projecte, ja que en tractar-se d'un projecte de diverses entregues en relativament poc temps, el més probable és que es prenguin decisions que afectin tant a les funcionalitats com a la planificació, i per tant al producte final.

No té per què ser una cosa dolenta, al contrari, quan més aviat es detectin desviacions, modificacions, nous requeriments o requeriments que no són necessaris, millor serà el resultat final.

- En segon lloc, i sent el motiu del projecte, es troben la resta de capítols (3, 4 i 5) que contenen el contingut del projecte en sí

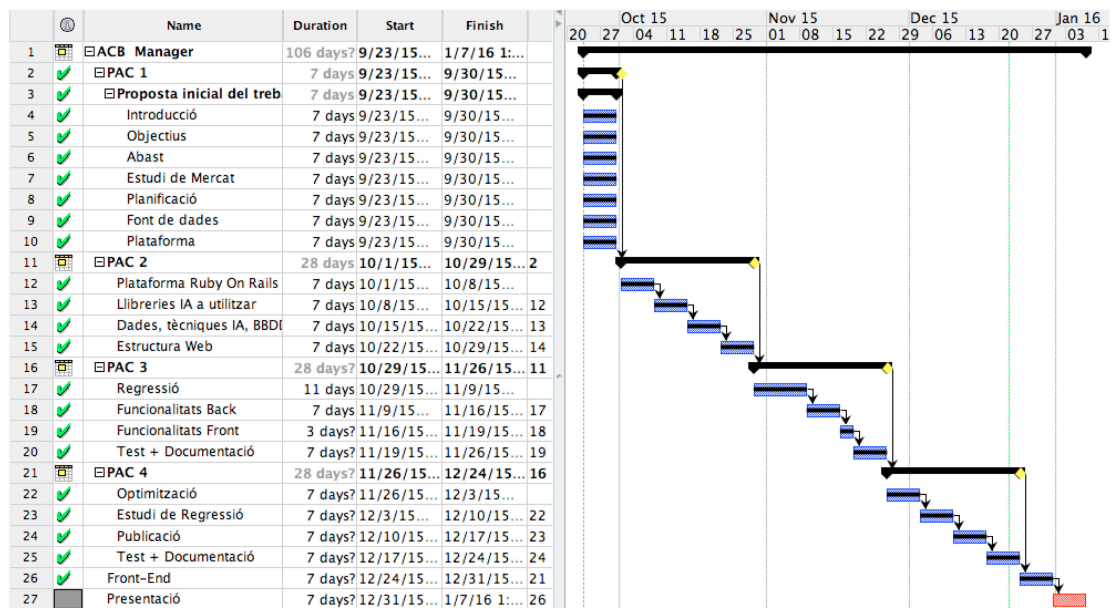
S'han anat implementant els capítols a partir dels diferents sprints realitzats (corresponents a les PAC 1, 2 i 3) i en ells hi podem trobar el relacionat amb la Intel·ligència Artificial, l'obtenció de dades, un estudi de d'atributs i paràmetres sobre la regressió, l'aplicació d'algoritmes genètics de cerca; juntament amb tots els aspectes que sorgeixin relacionats amb el framework de RoR.

2. Pla detallat del Projecte

2.1. Diagrama de Gant

A continuació es presenta el cronograma o diagrama de *Gantt*. Està directament relacionat amb la taula presentada anteriorment, i mostra d'una forma visual la distribució de les tasques. Cal tenir en compte que tant aquest diagrama com la resta de documents presentats i relacionats amb la planificació s'han anat actualitzant al llarg del projecte.

El número que identifica a cada tasca es farà servir en la resta de documents per referir-se a cada tasca en qüestió:



Imatge 1. Diagrama de Gannt

El Diagrama de Gannt ha anat patint certes modificacions en les entregues tot i que cap d'elles ha estat quelcom destacable que hagi fet que no es pogués fer les entregues parcials i la final a temps i complint els requeriments inicials.

2.2. Informe de Seguiment del Projecte

Un altre document que resulta molt interessant i que permet a qualsevol persona que no tingui cap tipus de coneixement del projecte fer-se una idea de l'estat del projecte en molt poc temps és l'informe de seguiment. Aquest document s'ha anat actualitzant en cada entrega i s'ha adaptat al projecte en qüestió.

Estat del Projecte	
Situació General del Projecte	
El projecte està a un 95% de la seva execució.	

El projecte està pràcticament acabat, simplement falta la última entrega que correspon al video presentació. S'han complert amb les objectius inicials plantejats i les dates d'entrega i la la part relacionada amb la publicació de la pàgina també s'ha fet, tot i estar prevista per la setmana següent a aquesta entrega.

Aspectes a destacar del Període

Durant el període del 23 de Setembre fins al 23 de Novembre (PAC1, PAC2 i PAC3 i PAC4), cal destacar:

- S'han definit la major part de les funcionalitats de captura de dades estadístiques i es poden aplicar filtres per obtenir dades de jugadors, partits, equips; juntament amb els corresponents acumulats.
- Es disposa de la pàgina web (ja publicada) podent consultar prediccions de la jornada corresponent i podent consultar les millors plantilles.
- Pel què fa a funcionalitats, la pàgina permet el registre d'usuari i desar plantilles per usuari.
- Es disposa de la versió definitiva del model dades..
- No hi ha cap retard. Els retards en les entregues anteriors s'han compensat en aquesta última entrega.

Activitats Completades

Tasca	Activitat	Data de Compliment	Situació
3	Proposta inicial del treball.	23/09/2015	Acabada, a temps.
12	Instal·lació i configuració de la plataforma RoR. Creació del Projecte Web.	01/10/2015	Acabada, a temps. Acabada, a temps.
13	Selecció de llibreries a utilitzar.	08/10/2015	Acabada, a temps.
14	Estructura i modelatge de la base de dades del Joc. Definició de l'estructura de la plataforma web (opcions de menú, disseny, opcions que s'oferiran, etc.).	22/10/2015	Acabada, a temps. Acabada, a temps.
15	Estructura web.	15/10/2015	Acabada, a temps.
17	Aplicació de les tècniques de regressió per previsió de valoracions.	09/11/2015	Acabada 4 dies tard.
18	Funcionalitats Back-End.	16/11/2015	Acabada, a temps.
19	Funcionalitats Front-End.	19/11/2015	Acabada, a temps.
20	Test i Documentació.	26/11/2015	Acabada, a temps.

Activitats a començar en el següent període

Tasca	Activitat	Data Inici	Data Fi
22	Aplicació de les tècniques d'optimització en equips per punts,	26/11/2015	Acabada, a temps.

	rebots i assistències		
23	Estudi de Regressió.	03/12/2015	Acabada, a temps.
24	Publicació de les plataformes (si s'escau).	10/12/2015	Acabada, abans de temps.
25	Test i Documenació.	17/12/2015	Acabada, a temps.
26	Funcionalitats Front-End.	24/12/2015	31/12/2015
27	Presentació Power Point / Video.	31/12/2015	01/7/2016

2.3. Pla de Gestió de Riscos

Un altre informe que s'inclou dins de la pla de seguiment i que s'ha naat actualitzant durant el projecte és l'Informe de Seguiment de Riscos, que parteix dels riscos definits en la primera part del projecte.

Situació			
Sumari			
	Risc	Accions de Mitigació	Data límit
	R01. Dades inadequades	Avaluació de les dades capturades (Fet)	22/10/2015
	R02. Plataforma inadequada	Avaluació de RoR i la disponibilitat de llibreries que resolguin tècniques de Regressió i Optimització (Fet)	14/10/2015
	R03. Resultats inesperats	En curs	14/10/2015
	R04. Plataforma insuficient	En curs	24/11/2015
	R05. Posicionament SEO	En curs	Durant tot el projecte i posterior posta en marxa
	R06. Manca de temps	En curs	Durant tot el projecte
Justificacions			
<p>Risc 01: les dades trobades es consideren suficients i adequades per a portar a terme el projecte. Es tenen dades històriques i el càlcul d'acumulats s'ha fet a partir d'aquestes dades sense cap problema.</p> <p>Risc 02: tot i que inicialment la llibreria plantejada no ha estat l'adequada, s'han trobat alternatives que compleixen els requeriments del que es necessita pel</p>			

projecte.

Risc 03: els resultats han permès fer un anàlisi de l'estudi de regressió. El risc s'ha mantingut com a Mitjà durant tot el projecte perquè és un factor important i s'ha d'estar alerta.

Risc 04: igual que l'anterior, però es considera de Baix Risc.

Risc 05: se seguiran directrius bàsiques per un bon posicionament però no se sabrà fins a la pujada de la web i la posterior avaluació. Com ja es va comentar queda fora del projecte però es menciona perquè també és un factor important.

Risc 06: portant el projecte el màxim al dia possible s'evitaran o es mitigarán possibles retards per factors inesperats interns o externs.

2.4. Resum d'incidències

Finalment, també es mostra una taula on s'han anat anotant les incidències que han sorgit durant el projecte:

Resum d'Incidències		
Incidència	Accions de Mitigació	Data tancament
I001: Retard en la finalització de la tasca 14	S'integrarà la tasca en la en el següent esprint, però no afectarà la durada del projecte.	30/10/2015 (prevista)
I002: Retard en la tasca 17	S'han ajustat les tasques de la PAC4 per tal que no afecti la durada del projecte.	09/11/2015 (4 dies de retard)
I003: Priorització de la publicació (tasca 24) respecte la front-end (Tasca 26)	S'han intercanviat les dues tasques en la última fase perquè s'ha considerat més important tenir una plataforma accessible que no pas la implementació de certes funcionalitats no relacionades amb la Intel·ligència Artificial.	

3. Visió General

3.1. Entorn de Desenvolupament

Per al desenvolupament del projecte s'ha escollit **Ruby o Rails**, un *framework open source* creat el 2004 i que està basat en el Model Vista Controlador (MVC), com la majoria de plataformes per a desenvolupament web o software avui dia.

Actualment és una plataforma molt estesa i utilitzada per grans portals com són *Airbnb*, *Github*, *Basecamp*, entre d'altres.

Després de treballar durant una mica més d'un any en aquesta plataforma, es considera que és una eina molt adequada per desenvolupar aquest tipus de projecte, en què es disposa d'un temps limitat i la flexibilitat i la velocitat a l'hora de desenvolupar és un element molt important.

Un element molt important de *RoR* són els *RubyGems*, que són extensions/paquets/l·libreries instal·lables que donen suports a certes funcionalitats. Així doncs, en aquest cas s'utilitzarà els *RubyGems LibLinear* per tal d'implementar la funcionalitat d'Intel·ligència Artificial que necessitem per al projecte (SVR), mentre que per a l'algoritme genètic s'implementarà una versió adaptada al que necessitem.

3.2. Repositori de Dades

Tot i que es tracta d'una eina força més útil quan hi ha més d'una persona treballant en un projecte, amb un entorn de desenvolupament, test i producció, és important mencionar que s'utilitzarà *Github* com a repositori.

Github és un repositori web que permet allotjar projectes i controlar les versions d'aquests, per tal de treballar diferents persones a la vegada sobre branques creades a partir d'una branca principal (anomenada màster) i tenir la possibilitat després d'unir diferents branques.

És molt útil també per tal de detectar errors. Sovint passa que es el desenvolupador implementa alguna funcionalitat i s'adona que deixa de funcionar alguna altra part del projecte que no s'ha tornat a testejat. Si es treballa correctament i s'actualitzen i comenten els canvis que es van fent es pot retrocedir a versions anteriors (a *comits* concrets) per descobrir els errors.

Per aquest projecte s'ha utilitzat principalment una branca màster, però s'han treballat amb branques paral·leles per al desenvolupament de la part d'optimització o de les funcionalitats de login i registre d'usuari.

Github, d'altra banda, és una eina molt útil per pujar els canvis i recollir-los en un entorn de desenvolupament diferent. Amb un parell d'instruccions (*push* per pujar i *pull* per baixar) i menys d'un minut el repositori està actualitzat a la última versió.

El repositori es pot consultar a la següent direcció:
<https://github.com/xavi666/manageracb>

3.3. Font de Dades

L'obtenció de dades s'ha fet a través de la pàgina web www.acb.com.
 Gràcies a un RubyGem anomenat *Nokogiri*, es captura l'html de les pàgines que contenen les dades estadístiques i s'emmagatzemen a la base de dades. Posteriorment es llegeix aquest html per tal de capturar les dades que interessin..

Un exemple de la informació capturada és:

UCAM MURCIA CB 85										REB				TAP				FP				
D	Nombre	Min	P	T2	T2 %	T3	T3 %	T1	T1 %	T	D+O	A	BR	BP	C	F	C	M	F	C	+/-	V
5	Martín, Alberto																					
6	Antelo, J. A.	30:22	18	4/10	40%	2/6	33%	4/4	100%	9	5+4	1	1	0	0	0	0	0	3	4	10	20
8	Neto, Raulzinho	26:22	13	5/9	56%	1/1	100%	0/1	0%	2	1+1	6	2	3	2	1	0	0	3	3	10	16
10	Cabezas, Carlos	31:17	7	0/0	0%	1/2	50%	4/4	100%	1	0+1	5	1	1	0	0	0	0	1	3	10	14
11	Radovic, Nemanja	22:10	9	4/6	67%	0/1	0%	1/1	100%	3	2+1	0	1	3	2	0	0	0	1	2	0	8
15	Wood, Scott	20:22	8	1/1	100%	1/3	33%	3/3	100%	0	0+0	0	2	0	0	0	0	0	2	1	-1	7
22	Kelati, Thomas	32:5	4	1/3	33%	0/1	0%	2/2	100%	1	1+0	0	1	2	0	0	0	0	4	1	3	-2
23	Lima, Augusto	27:28	22	10/11	91%	0/0	0%	2/6	33%	7	3+4	0	4	2	0	2	0	4	1	4	8	31
27	Rojas, Sadiel	9:54	4	2/2	100%	0/0	0%	0/0	0%	2	1+1	0	1	1	0	0	0	1	5	2	5	3
35	Arteaga, Victor																					
	Equipo		0	0/0	0%	0/0	0%	0/0	0%	2	2+0	0	0	0	0	0	0	0	0	0	0	2
	Total	200:0	85	27/42	64%	5/14	36%	16/21	76%	27	15+12	12	13	12	4	3	0	5	20	20	9	99
E	Ocampo, Diego																					
5f	Rojas, Sadiel																					
VALENCIA BASKET CLUB 76										REB				TAP				FP				
D	Nombre	Min	P	T2	T2 %	T3	T3 %	T1	T1 %	T	D+O	A	BR	BP	C	F	C	M	F	C	+/-	V
5	Ribas, Pau	19:5	12	3/4	75%	2/4	50%	0/1	0%	3	1+2	3	1	2	0	0	0	0	1	2	-6	14
9	Van Rossom, Sam	26:43	10	1/2	50%	2/4	50%	2/2	100%	1	0+1	3	0	2	0	0	0	0	2	2	-1	9
10	Sato, Romain	15:38	4	2/4	50%	0/0	0%	0/1	0%	2	1+1	0	0	2	0	0	1	0	4	3	-13	-1
12	Lishchuk, Serhiy	17:51	8	3/5	60%	0/0	0%	2/2	100%	0	0+0	0	0	2	0	0	1	1	4	2	0	1
13	Lucic, Vladimir	27:23	10	3/5	60%	0/1	0%	4/4	100%	5	2+3	0	0	2	0	0	0	0	3	3	-6	10
16	Vives, Guillem	13:44	1	0/1	0%	0/0	0%	1/2	50%	2	0+2	1	2	1	0	0	0	0	0	2	-5	5
17	Martínez, Rafa	5:58	0	0/1	0%	0/1	0%	0/0	0%	1	1+0	2	0	1	0	0	1	0	0	1	3	0
23	Buycks, Dwight	15:6	1	0/1	0%	0/0	0%	1/2	50%	0	0+0	0	1	1	0	0	0	0	0	2	-7	1
25	Loncar, Kresimir	11:1	6	3/5	60%	0/0	0%	0/0	0%	0	0+0	0	1	3	0	0	0	1	1	0	-9	1
34	Aguilar, Pablo	24:26	15	4/6	67%	2/4	50%	1/3	33%	7	5+2	1	2	0	1	0	0	2	2	2	-5	19
44	Harangody, Luke	23:5	9	0/1	0%	3/6	50%	0/0	0%	6	3+3	0	0	3	0	0	0	0	3	1	4	6
	Equipo		0	0/0	0%	0/0	0%	0/0	0%	0	0+0	0	0	0	0	0	0	0	0	0	0	0
	Total	200:0	76	19/35	54%	9/20	45%	11/17	65%	27	13+14	10	7	19	1	0	3	4	20	20	-9	65

Imatge 2. Exemple d'estadística per al recull de dades

Aquesta imatge correspon a l'estadística d'un partit i per a cada jugador es tenen els minuts jugats, punts de 2/3/1 punts, rebots, pilotes perdudes, recuperades, etc. Finalment tenim la valoració del partit (columna V) que es calcula a partir de la resta de camps.

Actualment el joc de SuperManager ACB té competició per valoració, punts, rebots i assistències, així que aquests 4 valors són els que es volen predir.

Tenint en compte que per temporada hi ha unes 7.300 valoracions, que surten de 306 partits cada temporada (9 partits cada jornada i 34 jornades) amb un màxim de 24 jugadors per partit, i que tenim les dades de 30 anys (des de la temporada 1985-1986), disposaré de més de 200.000 valoracions de jugadors. Tot i així, es treballaran només amb dades de la temporada 2014-2015 per a la creació del model de regressió, i les prediccions es faran sobre cada jornada de la temporada actual.

Arribat a aquest punt en que es disposen d'unues 7300 valoracions de la temporada anterior, es necessita calcular els valors acumulats d'equips i de jugadors, de tal manera que per una valoració d'un jugador contra un equip rival concret seran necessàries:

- Les dades mitjanes acumulades d'aquest jugador fins el dia d'aquest partit.
- Les dades mitjanes acumulades de l'equip contra el qual juga aquest jugador el d'aquest partit. En aquest sentit es necessiten les dades acumulades que podriem denominar positives (punts fets, triples fets, assistències d'equip, etc.) però també les negatives (valoració de l'equip rival en un partit, punts que anota el rival, assistències, triples anotats i rebots que agafa).

Per fer aquest tipus de càcul s'executarà una funció que va generant registres per jugador, equip i jornada de cada temporada, acumulant els valors ja sigui del jugador o de l'equip.

3.4. Funcionalitats

Com que uns dels objectius d'aquest projecte és disposar d'una pàgina web útil per als usuaris i que realment sigui una eina usable, hi ha aspectes no relacionats amb la Intel·ligència Artificial que s'han inclòs dins del projecte. Algunes d'aquestes funcionalitats ja s'han anat comentant però es comentaran perquè quedi clar el què quedarà inclòs i el que no.

Tal com s'ha especificat en els objectius, es vol aconseguir una plataforma fàcil d'utilitzar per a l'usuari, amb informació útil a simple vista per tal que l'usuari tingui la necessitat de consultar la informació.

S'ha fet un petit esquema de com serà la pàgina d'inici, on clarament es poden identificar les dues parts sobre les quals es treballarà: predicció de valoracions i optimització de plantilles.

LOGOTIPO
JUGADORES
EQUIPOS
ESTADISTICAS

PREDICCIONES DE PARTIDOS

OPTIMIZA TU EQUIPO

Nº JORNADA		PUNTOS TOTALES:		DINERO EN CAJA:
JUGADOR/EQUIPO	COTIZACIÓN	VALORACIÓN	COSTE	
BASES	Salgado, Javier <small>Moviestar Estudiantes</small>	345.000	3,00	456.780
	<input type="text"/>	-----	---	-----
	<input type="text"/>	-----	---	-----
ALEROS	<input type="text"/>	-----	---	-----
	<input type="text"/>	-----	---	-----
	<input type="text"/>	-----	---	-----
	<input type="text"/>	-----	---	-----
PIVOTS	<input type="text"/>	-----	---	-----
	<input type="text"/>	-----	---	-----
	<input type="text"/>	-----	---	-----
	<input type="text"/>	-----	---	-----

OPTIMIZAR

USUARIO

CONTRASEÑA

CREAR USUARIO
LOG IN

RECOMENDACIONES DE COMPRA

- 1 - González, C.
- 2 - Sastre, J.
- 3 - Pustovyi, A.
- 4 - Adama, D.
- 5 - Martínez, R.
- 6 - Samuls, S.
- 7 - Rojas, S.
- 8 - Bourousia, I.
- 9 - Hanga, A.
- 10 - Gomes, B.
- 11 - Bellas, T.
- 12 - Omic, A.
- 13 - Dubljevic, B.
- 14 - Musli, D.

Imatge 3. Estructura de la pàgina inicial

3.4.1. Predicció de valoracions

A la pàgina inicial es troba una taula amb els partits de la pròxima jornada. Aquest tipus d'informació ha d'estar emmagatzemada a la base de dades i s'haurà d'anar actualitzant, sigui de forma automàtica o manualment, tot i que no és un aspecte que es comentarà en el projecte ja que dependrà del temps del que es disposi.

Un cop s'accedeix als partits, es mostra una taula amb les prediccions del partit. Aquestes prediccions seran, tal com s'ha ja comentat, de valoració, punts, rebots i assistències. L'última columna (Cotización) fa referència al preu del jugador, però com que es tracta d'un valor calculat a partir de la valoració no es comenta.

A continuació es mostra una fitxa d'exemple:

LOGOTIPO
JUGADORES
EQUIPOS
ESTADISTICAS

PREDICCIONES DE PARTIDOS

GCA	VALORACIÓN	PUNTOS	REBOTES	ASISTENCIAS	COTIZACIÓN
Kuric, Kyle	30	15	7	3	1.900.000
Omic, Alen	2	0	1	1	500.000
Aguilar, Pablo	7	2	0	1	900.000
Pangos, Kevin	20	15	7	2	1.700.000
Newley, Brad	10	6	4	0	1.100.000
Oliver, Albert	30	12	9	1	800.000
Rabaseda, Xavi	10	8	2	3	700.000
Paseniks, Anzejs	15	10	7	2	1.300.000
Báez, Eulis	3	0	0	3	1.200.000
Paulí, Oriol	7	2	7	0	1.800.000
Salín, Sasu	9	2	3	0	1.350.000
Savané, Sitapha	30	18	7	2	1.670.000

SEV	VALORACIÓN	PUNTOS	REBOTES	ASISTENCIAS	COTIZACIÓN
Sánchez, Alfonso	10	8	2	3	700.000
Slokar, Uros	15	10	7	2	1.300.000
Balvin, Ondrej	3	0	0	3	1.200.000
Rodríguez, Berni	7	2	7	0	1.800.000
Henton, LaDontae	9	2	3	0	1.350.000
Anderson, Jermaine	30	18	7	2	1.670.000
Bamforth, Scott	30	15	7	3	1.900.000
Cate, Emanuel	2	0	1	1	500.000
Oriola, Pierre	7	2	0	1	900.000
Rigo, Tomeu	20	15	7	2	1.700.000
Mijenovic, Nenad	10	6	4	0	1.100.000
Nachbar, Bostjan	30	12	9	1	800.000

USUARIO

CONTRASEÑA

CREAR USUARIO
LOG IN

RECOMENDACIONES DE COMPRA

- 1 - González, C.
- 2 - Sastre, J.
- 3 - Pustovyj, A.
- 4 - Adama, D.
- 5 - Martínez, R.
- 6 - Samuls, S.
- 7 - Rojas, S.
- 8 - Bourousia, I.
- 9 - Hanga, A.
- 10 - Gomes, B.
- 11 - Bellas, T.
- 12 - Omic, A.
- 13 - Dubljevic, B.
- 14 - Musli, D.

Imatge 4. Pàgina detall de les prediccions d'un partit

Aquest apartat es basarà en les tècniques de regressió ja mencionades anteriorment i s'implementarà mitjançant la col·lecció *LibLinear*, que permet utilitzar diverses tècniques de regressió automàtica (SVR) per a generar prediccions de punts, valoracions, rebots i assistències.

Es farà un estudi d'atributs i de paràmetres tal i com es veurà en el següent apartat.

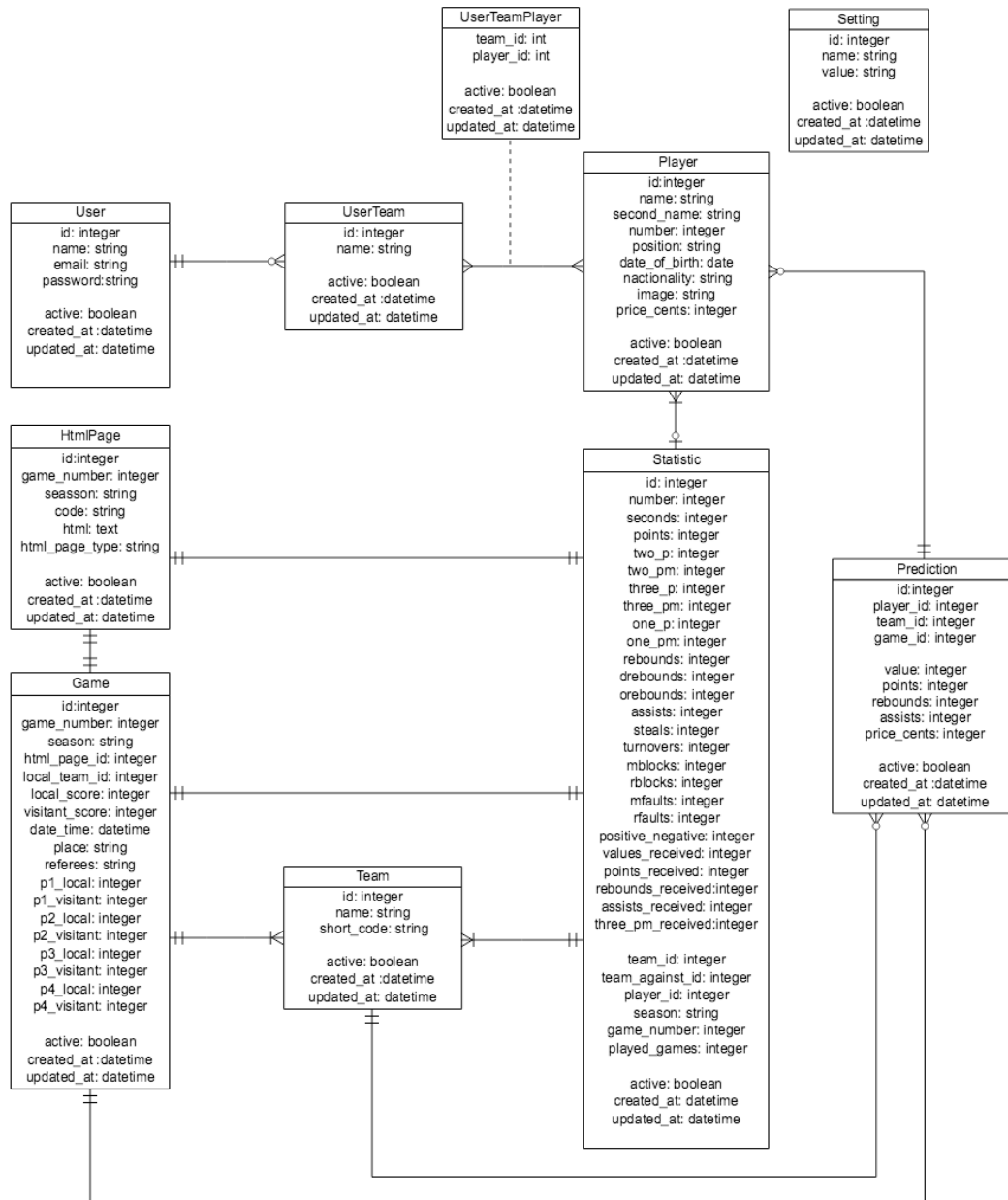
3.4.2. Optimització de plantilles

En la Figura 1, a la part inferior, es pot veure com funcionarà l'optimitzador de plantilles. Bàsicament un usuari des d'un primer moment ja podrà confeccionar la seva plantilla, però a la vegada oferirà la possibilitat d'introduir el pressupost i obtenir el millor equip en quant a valoració, punts, rebots, assistències o triples; sempre partint de les prediccions de cada jornada generades per la pròpia aplicació.

S'implementarà un Algoritmes Genètics de Cerca per tal de solucionar el problema que volem resoldre i se'n comentaran els aspectes més rellevants.

3.5. Model de Base de Dades

A continuació es mostra la proposta de model de dades, que s'està utilitzant ja per implementar el projecte.



Imatge 5. Model de Base de Dades, Versió 3.0

Es comenten les taules més importants i els elements a tenir en compte de cascuna d'elles:

- **HTMLPage**: emmagatzema el codi html de les pàgines de les quals es s'obtidran les dades. L'atribut *html_page_type* pot ser game (quan es tracta dels partits de cada jornada) o statistic (quan es tracta d'una estadística d'un partit concret).

- **Statistic:** a partir de les HTMLPage de tipus statistic es generen estadístiques per a un jugador amb el jugador, jornada i temporada, equip en el qual juga, equip contra qui juga, i totes les dades estadístiques del jugador (segons jugats, punts, intents, faltes, etc.).

Aquesta taula també un atribut *type_statistic* que pot ser *game* (estadística d'un jugador en un partit concret), *team* (estadístiques acumulades d'un equip en una jornada concreta), i *player* (estadístiques acumulades d'un jugador en una jornada concreta).

Dels equips, es guarden les dades de punts fets o faltes fetes, però també es guarden les dades dels punts, rebots, assistències i triples rebuts. A partir d'aquestes dades es construirà el model de regressió.

- **Game:** es generarà com en el cas anterior a partir del codi html, i conté informació dels partits de cada jornada i temporada.
- **Team:** els diferents equips de la lliga. Actualment són els mateixos 18 equips en les dues últimes temporades.
- **Player:** a mesura que un jugador intervé en un partit es va creant a la base de dades. Es disposa d'un camp *active* que ens determina si l'usuari està actiu (aquesta temporada) o no.
- **Prediction:** es desa la informació de les prediccions per jugador i partit. Així doncs, tindrem que un jugador concret en un partit contra un rival concret, tindrà unes prediccions de punts, rebots, assistències i rebots, guardats en aquesta taula.
- **Price:** tot i que es guarda la informació del preu actual del jugador a la taula *Player*, es guardarà un històric dels preus dels jugadors de cada jornada.
- **User:** correspon als usuaris de la web i tenim la informació bàsica: nom, contrasenya encriptada i email.
- **UserTeam i UserTeamPlayer:** permet guardar la informació dels jugadors que formen part d'un equip i dels propietaris de l'equip.
- **Setting:** es guarda informació de configuració molt concreta. Ara mateix hi ha tres registres: *season* (la temporada actual, 2015), *game_number* (jornada actual, que es va actualitzant cada jornada) i *season_data* (tempora que s'utilitza per a generar el model de regressió).

Així doncs, com que ja tenim els partits de cada jornada i la temporada actual (2015), actualitzant la variable *game_number* al final d'una jornada per la següent jornada, ja mostrarà els partits de la jornada corresponent.

4. Funcionalitats de la Plataforma

Les funcionalitats s'han dividit en dos grups:

Funcionalitats de Back-End

Inclouen totes aquelles opcions que permetran capturar les dades i automatitzar informació. Sense aquestes opcions, la captura d'informació resultaria difícil i costosa, i ha de servir per poder generar un volum de dades òptim per a treballar-hi i també per poder anar recollint informació a mesura que avancen les jornades.

Tal com es comprovarà al llarg del projecte, aquest és un aspecte clau per a les prediccions com a les optimitzacions.

Funcionalitats de Front-End

Inclouen totes les funcionalitats per a l'usuari, visualització d'estadístiques, de prediccions, registre d'usuari, creació d'equips, etc. S'intentarà arribar al màxim d'opcions per tal d'oferir una plataforma atractiva per a l'usuari, útil i fàcil d'utilitzar.

4.1. Funcionalitats Back-end

Importació de Pàgines Html

S'utilitza la llibreria Nokogiri per llegir pàgines html completes i emmagatzemar-les a la base de dades local. Com que les dades no variaran podem fer una sola crida per capturar la informació i posteriorment treballar de forma local sobre el codi html guardat.

Mitjançant la parametrització podem capturar dades de temporades concretes, i es capturen dades de 3 tipus:

- Estadístiques de partits. Amb la informació de cada jugador.
- Partits per temporada, per tenir informació dels partits passats i dels pròxims, bàsicament el calendari.

Creació de dades

A partir de la informació captura en l'apartat anterior, s'alimenta tota la base de dades, creant diferents tipus d'informació.

- Jugadors.
- Equips.
- Estadístiques de jugador en cada partit contra un equip rival concret.
- Partits entre dos equips
- Preus dels jugadors. Aquesta informació es carrega un cop a la setmana o es recalcula cada setmana a partir d'una càrrega inicial i es fa a partir de fitxers html descarregats del propi SupermanageACB.com ja que es necessita accés amb usuari i contrasenya per accedir a la informació.

Creació d'acumulats

Un dels elements més importants és la creació de les dades acumulades, tant de jugadors com de partits. Així doncs, un equip tindrà unes dades acumulades parcials per a cada jornada, de la mateixa manera que ho tindrà un jugador.

El que interessa per a fer les prediccions és tenir les dades mitjanes d'un jugador en la jornada 10, per exemple, i encreuar-les amb les dades mitjanes de l'equip contra el qual juga. Totes aquestes dades permetran crear un model d'anàlisi que després en una jornada concreta ens permetrà fer una predicció d'un jugador que té unes dades acumulades en l'actual temporada, i que juga contra un equip amb unes dades acumulades fins al moment.

4.2. Funcionalitats Front-End

Com en qualsevol pàgina web, el fet que cada cop que un usuari consulta la pàgina web (un cop a la setmana o esperem que molt més freqüentment) per veure les recomanacions, ha d'introduir tot el seu equip, pot resultar una mica molest. És per això que hi haurà la possibilitat de registrar els usuaris per poder crear els equips i consultar la informació més ràpidament.

Registre d'usuari

Simplement demanarem un correu electrònic i una contrasenya. Es tracta del típic registre d'usuari, amb una validació per correu electrònic. Inicialment no es contempla demanar més dades ni fer un perfil d'usuari més complex.

Iniciar Sessió

Amb l'usuari anterior s'ha de poder iniciar sessió (introduint el nom d'usuari i la contrasenya a la part superior dreta de la Figura 1) i accedir al llistat dels nostres equips (poden ser 1 o varis).

El funcionament a l'hora d'introduir equips i jugadors en aquests equips serà exactament tal com es veu a la Figura 1. Òbviament es controlarà que no es passa el pressupost i que no s'introdueix el mateix jugador diverses vegades en un mateix equip.

A banda d'aquest tipus de funcionalitats, hi ha altres opcions que poden ser interessants però que quedaran fora del projecte, com per exemple un enviament de Butlletí un cop finalitzada una jornada i quan comença la següent, oferint informació parcial de la jornada, o un resum dels encerts, o no, en les prediccions (això millor ho deixem per una fase més avançada en funció dels resultats). En qualsevol cas, l'objectiu únic d'aquestes opcions és generar visites, així que quedarà fora del projecte.

5. Prediccions i SVR

En l'apartat anterior s'ha explicat com funcionaria l'encreuament de dades de forma general, i en aquest s'aprofundirà sobre les dades utilitzades, s'analitzarà la tècnica de regressió lineal i els seus paràmetres, i s'analitzaran els resultats.

5.1. Dades Utilitzades

Les prediccions que volem obtenir són en relació a valoració, punts, rebots i assistències. A continuació tenim una taula on podem veure quins són els atributs que s'introduiran en cadascun dels models de regressió per a cadascuna de les prediccions:

Predicció	ATRIBUTS			
Valoració	Posició Jugadors	Valoració Jugador	Valoració Equip	Valoració Equip
Punts	Posició Jugadors	Punts Jugador	Punts Equip	Punts rebuts Equip
Rebots	Posició Jugadors	Rebots Jugador	Rebots Equip	Rebots rebuts Equip
Assistències	Posició Jugadors	Assistències Jugador	Assistències Equip	Assistències rebudes Equip

5.2. Normalització de Dades

Un dels passos previs a la regressió lineal, és la normalització de les dades, que en el cas que ens ocupa consisteix en:

Eliminar dades irrellevants

En la creació del model de regressió, s'han eliminat les estadístiques d'aquells jugadors que no disputen cap minut/segon en un partit.

Equiparar les magnituds

Per tal que un atribut no tingui més influència en les dades que un altre i tenint en compte que hi ha atributs en unitats diferents, ja que hi ha valors de mitja de punts de jugador per partit, però també hi ha valors de mitja de punts per judaor, es disposa d'una funció per tal de normalitzar les dades (a partir de la resta de la mitjana i la divisió per la desviación típica de cada columna).

Càlcul de les mitjanes

A mesura que es desenvolupava s'ha trobat que les mitjanes calculades no corresponien a les dades oficials de ACB, i això era perquè quan un jugador no participa en un partit (per exemple, per lesió), no es comptabilitza com a partit jugat (i per tant no influeix en la mitja).

És per això que s'ha utilitzat el camp *played_games* per comptabilitzar el nombre de partits acumulats de cada jugador, i així obtenim el valor mig reals

dels jugadors que, per exemple, han estat lesionats, o s'han incorporat en un partit a mitja temporada o han deixat l'equip a mitja temporada.

Remenar els valors

Per tal de no treballar sempre amb les mateixes dades de tipus test, i agafar valors de tota la temporada es va servir la funció `suflé` per "remenar" els registres i agafar valors aleatoris. Això farà que obtinguem valors diferents en cada execució.

5.3. SVM i SVR

Les Màquines de Vectors de Suport (SVM) són una sèrie d'algoritmes d'aprenentatge supervisat que resolen problemes de classificació i regressió. En aquest cas s'utilitzarà per fer prediccions mitjançant la regressió, és a dir, que donat un conjunt de dades d'entrenament (creat a partir de les estadístiques de la temporada 2014-2015) es construirà un model de regressió capaç de fer una o diverses prediccions.

Posteriorment a les MVS apareix el concepte de SVR (Regressió per Màquina de Vectors) que té com a objectiu trobar una funció que sigui capaç de fer prediccions, però afegeix un marge d'error (en forma de tub al voltant de la funció) que equival al marge de les SVM, que determinarà si inclou o no la predicció dins del conjunt d'entrenament si està o no dins del marge d'error.

Això traduït a les estadístiques de jugadors, farà que en funció del marge establert es considerin o no dins del conjunt d'entrenament certs elements. Per exemple, podem incloure en el conjunt d'entrenament una predicció de valoració de 15 punts quant a la valoració real és de 18. En canvi, com que el nombre de rebots d'un jugador en un partit és inferior, entre 1 i 12 per exemple, pot ser necessari modificar el marge en els rebots a 1 o 2, ja que no considerarem com a bona una predicció de 3 rebots si després el jugador n'agafa 0.

5.4. LIBLINEAR

LIBLINEAR és una llibreria de codi obert de classificació lineal a gran escala, que suporta tècniques regressió logística i màquines de vectors, i que és molt eficient en grans conjunts de dades disperses.

Aquesta llibreria ens proporciona tres maneres diferents de fer l'anàlisi, i el que farem és testejar les tres funcions per tal i els diferents paràmetres per utilitzar el que tingui millor resultats (si és el cas).

Així doncs, els tres mètodes disponibles, són:

Liblinear::L2R_L2LOSS_SVR

L2-regularized L2-loss support vector regression (primal)

Liblinear::L2R_L2LOSS_SVR_DUAL

L2-regularized L2-loss support vector regression (dual)

Liblinear::L2R_L1LOSS_SVR_DUAL

L2-regularized L1-loss support vector regression (dual)

L1-loss Vs L2-loss

La funció L1-norm loss es coneix com a la desviació mínima absoluta (LAD) o error absolute mínim (LAE), és a dir el valor mínim de la suma de les diferències absolutes entre el valor objectiu i l'estimat (la predicció).

La funció L2-norm loss es coneix com a error mínim quadràtic (LSE), és a dir el valor mínim de la suma del quadrat de les diferències entre el valor objectiu i la predicció.

A efectes pràctics tenim que:

- La funció L1-loss és més robusta que la L2-loss. La robustesa fa referència a la resistència als valors atípics en les dades, és a dir, si els valors dels extrems poden ser ignorats o no per seguir mantenint seguretat i eficàcia en les estimacions.

- La funció L2-loss és més estable que la L1-loss. Això vol dir que per una banda la funció L2-loss ofereix una única solució, mentre que L1-loss pot oferir múltiples solucions. A més, l'estabilitat també fa referència al fet que si es modifica alguna de les dades, la línia de regressió sempre es mou només lleugerament; és a dir, els paràmetres de regressió són funcions contínues de les dades.

La configuració Dual o Primal defineixen diferents maneres d'arribar al mateix resultat, i poden influir en el temps de resposta (sent Primal més lent) i no pas en el valor de la predicció. Així doncs, a no ser que hi hagi molta diferència no es comentaran els resultats, ja que aquesta part queda fora de l'anàlisi.

A banda d'aquestes tres funcions, LIBLINEAR presenta diferents paràmetres, dels quals se n'analitzaran 2:

Paràmetre C:

El paràmetre C controla l'equilibri entre els errors de la predicció i la maximització del marge, de tal manera que a mesura que C augmenta, el marge disminueix (el marge tendeix a l'anomenat hard-margin).

epsilon-SVR:

El valor de epsilon determina el nivell de precisió de la funció obtinguda a través del model de regressió. Així doncs, l'augment d'aquest valor redueix la precisió.

5.5. Estudi de Regressió

El codi relacionat amb la regressió es troba al controlador PredictionsController:

https://github.com/xavi666/manageracb/blob/master/app/controllers/predictions_controller.rb

Per tal d'avaluar els resultats s'han fet servir dues dades:

- Mean Squared Error: l'error quadràtic mig, mesura la mitjana dels errors al quadrat, és a dir, la diferència entre l'estimador i el que s'estima. Es tracta d'un valor sense unitats per tant es compararan els resultats, sent millor quant més baix és aquest valor.
- Squared Correlation Coefficient: el coeficient de determinació, determina la qualitat d'un model per replicar resultats i la proporció de variació dels resultats que es pot explicar pel model. Es tracta d'un valor que va entre 0 i 1 sent millor quant més s'apropa a 1.

Un dels primers aspectes a tenir en compte és amb quin nombre de registres es crearà el model de regressió per avaluar els resultats. Tenint en compte que amb les dades actuals tenim més de 5000 dades estadístiques s'ha de trobar un valor adequat que ens permeti tenir bons resultats però que tampoc consumeixi massa temps.

En aquest cas s'han fet les proves amb 500 registres, ja que a mesura que he anat pujant aquest valor els resultats no variaven gaire, i s'ha considerat una mostra representativa suficient per analitzar els resultats.

En l'anàlisi dels diferents mètodes que disposa la llibreria LibLinear el mètode L2R_L2LOSS_SVR ha presentat resultats clarament pitjors respecte els altres en totes les execucions, així que l'hem descartat d'entrada. Pel què fa als altres dos mètodes, L2R_L2LOSS_SVR_DUAL i L2R_L2LOSS_SVR_DUAL els resultats han estat més o menys similars.

Un altre dels aspectes que s'han analitzat és el fet de si l'atribut position (posició en què juga el jugador: base, aler, pivot) modifica (millora o empitjora) els resultats del model de regressió. En aquest sentit s'ha detectat que amb l'atribut position s'obtenen millors resultats, la qual cosa significa que aquest atribut aporta informació al model i les nostres prediccions són millors amb la seva presència.

En relació al paràmetre C, els millors resultats que s'han obtingut han estat amb valors situats entre 1.5 i 0.5, així que deixant-lo amb el seu valor defecte (1) és suficient.

I finalment, amb el paràmetre epsilon-SVR s'ha pogut comprovar que a mesura que augmentava aquest valor la precisió disminuïa, amb la qual cosa s'ha optat per mantenir aquest valor per defecte de 0.1.

5. Optimització

En els jocs de tipus manager, una de les situacions més comunes com a jugador es produeix quan es vol tenir el màxim rendiment amb els jugadors més econòmics, és a dir, comprar els jugadors amb millor rendiment en relació al seu preu.

En aquest apartat es vols solucionar aquest problema, ja que són diverses les competicions per les quals es poden competir al SuperManager ACB, com són per valoracions, punts, rebots, assistències o triples.

Per tal de resoldre aquest problema s'implementarà un Algoritme Genètic que ens permetrà obtenir el millor equip dins d'un pressupost especificat.

5.1. Algoritme Genètic

Un Algoritme Genètic (GA) és un mètode de computació que simula l'evolució biològica d'un individu. Aquest algoritme fa evolucionar una població d'individus aplicant-li accions aleatòries similars a les que actuen en l'evolució biològica (mutacions i recombinacions genètiques), així com una selecció en funció d'un criteri, que ens permetrà determinar quins són els individus més adaptats, que sobreviuen, i quins no sobreviuen i són descartats.

Aquest mètode normalment s'utilitza per optimitzar funcions que són intractables o amb grans o desconeguts espais de cerca.

La implementació de l'optimització s'ha fet íntegrament al controlador UserTeamsController, i està basat en la implementació feta per *Jason Brownlee* a *Clever Algorithms: Nature-Inspired Programming Recipes* [7].

- La funció *search*, encarregada d'inicialitzar el nombre de poblacions especificades. Entenem població com a un equip format per 11 jugadors (3 bases, 4 alers i 4 pivots) on no hi hagi jugadors repetits i no se superi el pressupost especificat.

Dins d'aquesta funció també es fa evolucionar tantes vegades com s'hagi especificat cadascuna de les poblacions.

- La funció *tournament*, on es realitza la selecció basant-se en comparacions directes dels individus.
- La funció *reproduce*, s'encarrega de reproduir l'individu seleccionat en la funció anterior.
- La funció *crossover*, s'utilitza dins de la funció anterior i correspon a la funció d'encreuament, que consisteix a fragmentar dos equips i intercanviar cadascun dels fragments.

- La funció *point_mutation*, correspon a la funció de mutació, i el que fa és mutar un equip (modificar els jugadors) tenint en compte que no es poden repetir jugadors ni superar el pressupost màxim.

Finalment, una de les decisions a tenir en compte és determinar quina serà la mida de la població i el nombre de generacions (les variables *pop_size* i *max_gens* al codi). Aquests dos valors determinaran uns millors o pitjors resultats però sempre aniran en detriment del temps.

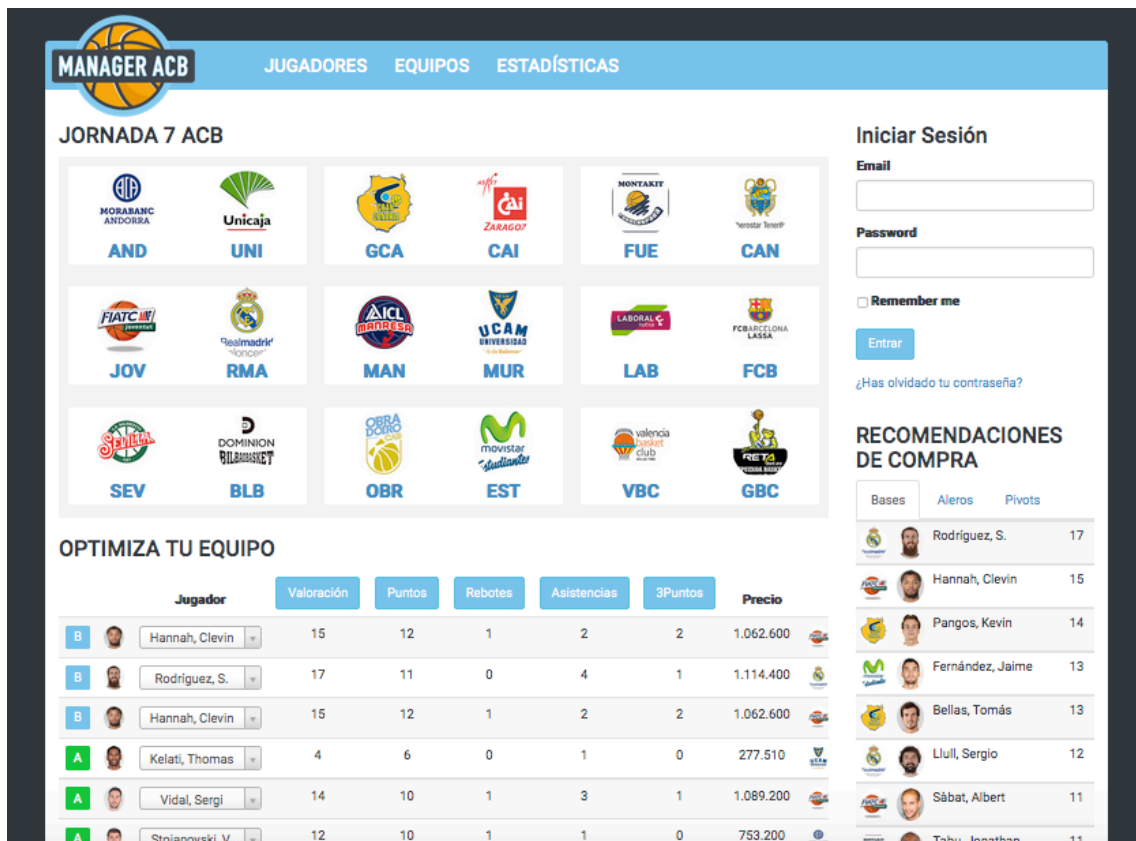
Així doncs, després de testejar diferents valors s'ha determinat que s'obtenen solucions òptimes amb una mida de població de 100 i 40 generacions.

6. Disseny i Publicació

6.1. Disseny

Partint de l'estructura inicial presentada anteriorment, per a tot el relacionat amb el disseny s'ha utilitzat Bootstrap, un dels frameworks gratuïts (d'HTML, CSS i JavaScript) més populars que permet desenvolupar projectes web responsive, de forma molt fàcil i intuïtiva.

Així doncs, el disseny final de la pàgina inicial ha quedat:



Imatge 6. Disseny Final de la Home

Mentre que la fitxa de prediccions té el següent aspecte:

The screenshot shows the Manager ACB website interface. At the top, there is a navigation bar with the logo and links for 'JUGADORES', 'EQUIPOS', and 'ESTADÍSTICAS'. The main content area is titled 'Predicciones de Partidos' and displays two tables of player statistics for ICL Manresa. The first table lists players like Grigonis, Marius and Sánchez, Ricky with their respective values for Valoración, Puntos, Asistencias, Rebotes, and 3PM. The second table lists players like Antelo, J. A. and Bamforth, Scott. To the right, there is a login section 'Iniciar Sesión' with fields for Email and Password, a 'Remember me' checkbox, and an 'Entrar' button. Below the login section is a 'RECOMENDACIONES DE COMPRA' section with tabs for 'Bases', 'Aleros', and 'Pivots', listing recommended players and their values.

Imatge 7. Disseny Final de la fitxa de Prediccions

6.2. Publicació

Finalment s'ha prioritzat la publicació del projecte per tal de poder accedir-hi i visualitzar tant les prediccions com l'optimització d'equips.

Per a fer-ho, com ja s'havia previst, s'ha utilitzat Heroku, una plataforma com a servei de computació al núvol que suporta diferents llenguatges de computació, entre ells Ruby on Rails.

Es pot accedir públicament a la plataforma a través de la següent url:
<http://manageracb.herokuapp.com/home>

En aquest cas s'ha utilitzat la versió gratuïta, que té certes limitacions, i una d'elles és que la base de dades només pot tenir un màxim de 10.000 registres. Això era un problema pel fet que en aquest projecte es treballa amb grans quantitats de dades, ja que només amb les estadístiques hi ha més de 30.000 registres.

Per a solucionar-ho, tenint en compte que les dades estadístiques només es fan servir per a la creació de les prediccions i que realment no es necessita al servidor Heroku, s'ha generat un script que exporta les dades que realment necessitem, és a dir:

- Jugadors actius (uns 200 registres)
- Equips (18 registres)
- Partits (306 registres per una temporada)
- Prediccions (uns 200 registres per jornada).
- Settings (3 registres de moment).

7. Conclusions

Considero que s'han assolit els objectius i que s'ha obtingut un producte que pot ser desenvolupat i millorat, i on s'ha aprofundit en els dos aspectes d'Intel·ligència Artificial proposats com a objectius, com són les prediccions a partir de models de regressió i l'optimització a partir d'Algoritmes Genètics de cerca.

Quant a la planificació considero també que s'ha utilitzat una metodologia àgil adequada al tipus de projecte, dividint el projecte en sprints per cada entrega, i on principalment el més important ha estat portar el projecte al dia i actualitzar els documents de seguiment proposats.

Si bé hi ha hagut alguns canvis en la planificació, aquests no han estat important i no hi ha hagut retards en entregues i pràcticament totes les funcionalitats que s'havien proposat s'han acabat implementant.

També faig una valoració molt positiva a la plataforma utilitzada i als coneixements adquirits. Si bé utilitzo Ruby on Rails com a eina de treball, el projecte m'ha permès treballar en un entorn des de zero, amb tot el que això suposa quant a problemes i cerca de solucions.

Considero que de tot el projecte el què ha estat més complicat ha estat la captura i preparació de les dades, un element que considero molt important avui dia en l'àmbit de la Intel·ligència Artificial. El fet de treballar amb grans volums de dades fa que el testeig i l'anàlisi de resultats sigui un procés lent i laboriós, amb la qual cosa cal una molt bona planificació.

M'hauria agradat aprofundir una mica més en l'anàlisi d'atributs, per tal d'intentar obtenir millors resultats en les prediccions amb altres dades, així com treballar més en la normalització de dades.

Tanmateix, l'anàlisi del model de regressió s'ha fet amb valors comparatius perquè no he estat capaç de treballar amb la NRMSE (Normalized Root Mean Square Error) i seria bo obtenir valors numèrics de com són de bones les prediccions obtingudes.

Dins de les possibles millores de futur, inclouria aspectes d'automatització de captura de dades i generació de prediccions, intentar incloure informació més rellevant per als jugadors o aspectes d'usabilitat de la pàgina web.

8. Glossari

RoR: Ruby on Rails

Ai4r: Artificial Intelligence for Ruby

SVM: Supports Vector Machine

SVR: Supports Vector Regression

LAD: Least Absolute Deviation

LAE: Least Absolute Error

MSE: Mean Squared Error

NRMSE: Normalized Root Mean Square Error

GA: Genetic Algorithm

PaaS: Platform-as-a-Service

9. Bibliografia

- **[1]: LIBLINEAR: A Library for Large Linear Classification**,
Machine Learning Group at National Taiwan University Contributors
[Última Modificació: Juliol del 2015]
URL: www.csie.ntu.edu.tw/~cjlin/papers/liblinear.pdf
- **[2]: Liblinear-Ruby**, *kei500*
[Última Modificació: Novembre 2015]
URL: <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>
Repositori: <https://github.com/kei500/liblinear-ruby>
- **[3]: RubyGems: Ai4r**, *Sergio Fierens*
[Última Modificació: Novembre 2014]
URL: <http://www.ai4r.org/>
Repositori: <https://github.com/SergioFierens/ai4r>
- **[4]: Differences between l1 and l2 as loss function and regularization**,
[Última Modificació: Desembre 2013]
[URL:www.chioka.in/differences-between-l1-and-l2-as-loss-function-and-regularization/](http://www.chioka.in/differences-between-l1-and-l2-as-loss-function-and-regularization/)
- **[5]: What is the difference between L1 and L2 regularization?**,
Xavier Amatrian
URL: <https://www.quora.com/What-is-the-difference-between-L1-and-L2-regularization>
- **[6]: A tutorial on support vector regression***, *Alex J. Smola i Bernhard Scholköpfung*
[Acceptat: Novembre 2013]
- **[7]: Clever Algorithms: Nature-Inspired Programming Recipes**,
By Jason Brownlee PhD