

Les claus d'èxit dels videojocs

Òscar Urenda Moix



Abstract—This study aims to uncover historical success indicators in the gaming industry by analyzing various parameters such as development teams, launch platforms, branding, and critic/user ratings. Employing data handling techniques and classification models, notably RandomForest, the research predicts game success based on sales. Results demonstrate reasonable predictive accuracy, emphasizing the significance of developers, publishing companies, and platforms across different regions (North America, Europe, and Japan). It reveals varying player preferences influenced by cultural and geographical factors, showcasing differences in the impact of developers and platforms between Western and Eastern markets.

Keywords— classification, noisy web data, CNNs, ubiquitous reweighting

1 INTRODUCCIÓ

La indústria del videojoc es una de les indústries més aclamades del món. Cada any milions de persones disfruten dels productes d'aquesta indústria, ingents quantitats de diners. Aquesta indústria no només ha experimentat un creixement impressionant, sinó que també ha evolucionat constantment, adaptant-se a les noves tecnologies i demandes dels jugadors. En aquest entorn altament competitiu, identificar quines han estat les claus del èxit en un joc és crucial per a les empreses desenvolupadores i distribuïdores. Per això hem estudiat diferents paràmetres com pot ser per exemple el quip desenvolupador del joc, la plataforma sobre la que es llença el producte, la marca que el publica, la puntuació de la crítica d'usuaris i crítics professionals...

2 PROPOSTA/METODOLOGIA

Al principi de tot, s'ha fet una visualització de les dades on es pot veure que partim amb un total de 16 columnes i 16719 files. També s'ha evidenciat que les dades contenen molts nans, per tant el següent pas ha estat veure quants nans tenen les dades. En aquest pas hem vist que algunes de les mètriques que creiem que serien més importants contenen una quantitat molt elevada de nans com pot ser el cas de la puntuació dels crítics

amb un 51%, la puntuació dels usuaris amb un 40% o el equip desenvolupador del videojoc amb un 39.6% que haurien de ser tractats. Per fer el tractament de nans vaig pensar en mirar de fer una Mitjana dels valors, però la idea es va descartar ràpidament degut a que hi havia una quantitat massa elevada. Donat que la base de dades sobre la que hem treballat és molt gran, vaig provar a eliminar les files que contien nans i el resultat va ser un dataframe amb 6825 files, cosa que considerarem que té un tamany acceptable per realitzar l'entrenament del nostre classificador. A continuació hem fet un tractament dels atributs continus i catagorics i hem decidit que per determinar si un joc és o no èxitós ho fariem a partir de les vendes realitzades. Per això hem fet la Mitjana tant de vendes parcials com globals i hem determinat que un joc és exitós si el seu número de vendes està per sobre de la mitja. Al determinar aquest target, hem comprovat com de balancejades teniem les dades i el resultat ha estat que en les vendes globals, vendes a nord Amèrica i Europa estan bastant ben balancejades, però les vendes al Japó tenen un fort desbalanceig degut a que la majoria eren jocs no exitosos. Per arreglar-ho he fet un oversample que ha nivellat la quantitat de dades. A continuació hem provat diferents tipus de classificadors amb un cross validation on hem vist que obtenen millors resultats amb el RandomForest, per tant hem decidit entrenar el nostre model amb aquest algorisme. Després, un cop hem entrenat el model hem validat els seus resultats amb les gràfiques ROC i PR, i les matrius de confusió. Finalment hem mostrat els atributs més importants per a predir si un joc serà èxitós de forma global, a Nord Amèrica(NA), a Europa (EU) o a Japó (JP). S'ha decidit no tenir en compte les vendes d'altres països degut a que estava representada una varietat massa elevada de països, amb cultures molt diferents entre elles, cosa que fa que les dades no siguin representatives per a tots ells

3 EXPERIMENTS, RESULTATS I ANÀLISI

Amb aquest data set, hem pogut fer els 4 classificadors segons on estiguessim mirant les vendes, en aquesta memòria exposaré les imatges del test global i explicaré els resultats obtinguts pels altres 3, ja que les imatges són sempre bastant semblants.

Amb les vendes globals, obtenim un 73% d'exactitud, amb un 73% de precisió de jocs no exitosos i un 74% en jocs exitosos

Vendes globals	precisio n	recall	F1-score	Support
False	0.73	0.76	0.74	862
True	0.74	0.71	0.72	845
Accuracy			0.73	1707
Macro avg	0.73	0.73	0.73	1707
Weighthed avg	0.73	0.73	0.73	1707

Taula 1. Valors de les mètriques del classificador de vendes Globals

Amb les vendes a NA obtenim un 74% de exactitud, un 77% de precisió de jocs no exitosos i un 71% amb els exitosos, amb EU obtenim un 72% de exactitud, precisions de joc exitosos i

no exitosos i finalment a JP obtenim un 81% d'exactitud, precisions de joc exitosos i no exitosos.

Hem fet la corba ROC de cada un dels models seleccionant AUC com a mesura de fiabilitat. AUC indica l'àrea per sota de la recta. Un AUC = 1 indica que la classificació es perfecta i un AUC < 5 indica que la predicció és pitjor que una classificació aleatòria

La corba ROC de les ventes globals té un AUC de 0.81, per a ventes en NA tenim un AUC = 0.83, per a EU es de 0.79 i per JP és 0.89

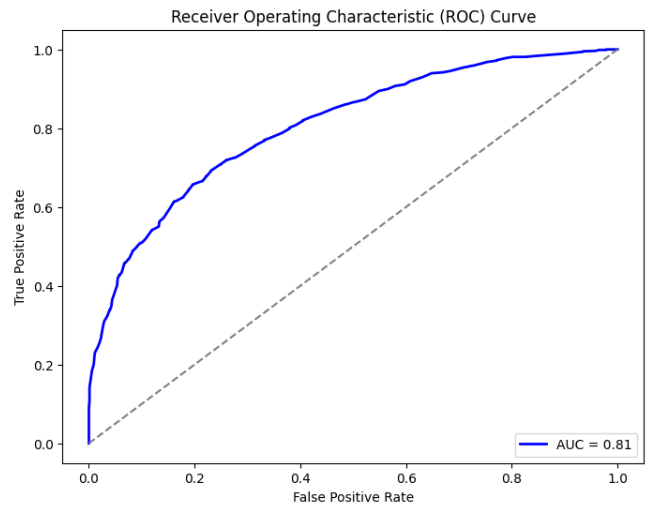


Fig. 1 Gràfic de la corba ROC del classificador de ventes globals

Aquests valors ens indiquen que les nostres prediccions en els 4 models fets son bastant correctes degut a que es manté un bon equilibri entre la tasa de true positives i False negatives, sobretot amb el classificador de JP

Tambe hem realitzat la corba PR utilitzant AUC on per a les ventes globals AUC és 0.81, per a les ventes de NA es de 0.84, per EU es de 0.79 i per JP es de 0.89

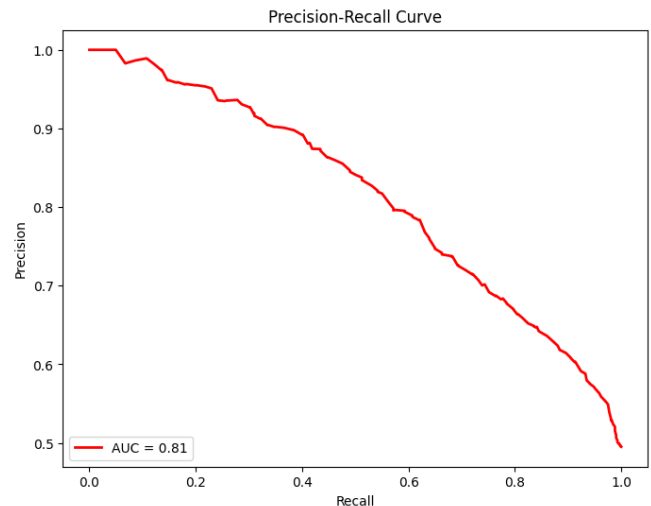


Fig. 2 Gràfic de la corba PR del classificador de ventes globals

Finalment hem utilitzat com a tècnica de validació per tal de veure com de be estava classificant totes les dades del conjunt del test.

El classificador de ventes globals té els resultats mostrats en la figura 3

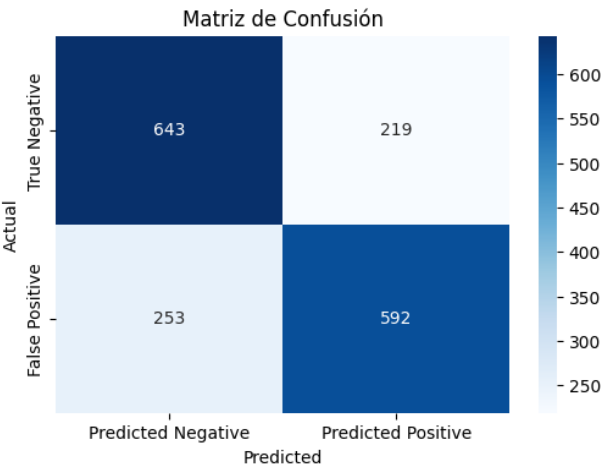


Fig. 3 Matriu de confusió del classificador de ventes Globals

La matriu de confusió de ventes a NA conté 670 True Negatives, 188 False negatives, 595 True Positives i 254 False Positive

La matriu de confusió de ventes EU conté 655 True Negatives, 222 False negatives, 585 True Positives i 245 False Positive

La matriu de confusió de ventes a conté 417 True Negatives,96 False negatives, 398 True Positives i 95 False Positive

Un cop vistes les mètriques de validació, tenim la importància de cada atribut per tal de poder predir el resultat.

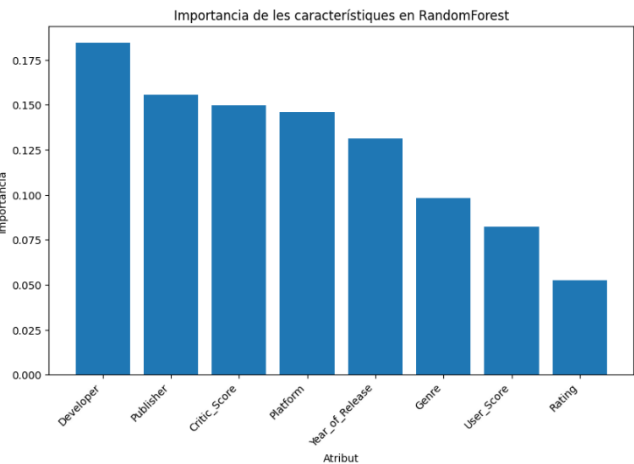


Fig. 4 Gràfic de la importància de cada atribut per al clasificador de ventes globals

El clasificador de ventes globals podem veure que sobretot prioritza el quip que desenvolupa els jocs,seguit de la empresa que el publica i les crítiques que rep el joc per part de

professionals. El resultat total de la importància per a aquest cas és el següent:

Developer	18.44%
Publisher	15.56%
Platform	14.98%
Critic Score	14.57%
Year of Release	13.12%
Genre	9.79%
User Score	8.24%
Rating	5.26%

Taula 2. Valors de les mètriques del classificador de vendes Globals

Ara bé això varia una en funció de de la regió con es venguin els jocs.

A NA, importa molt més l'equip que desenvolupa el joc que la resta d'atributs, seguit de la empresa que els publica i la plataforma sobre la que es llença el videojoc. Els seus resultats son els següents

Developer	23.55%
Publisher	16.18%
Platform	14.96%
Year of Release	13.01%
Critic Score	12.34%
User Score	7.79%
Genre	7.63%
Rating	4.50%

Taula 3. Valors de les mètriques del classificador de vendes NA

A EU, també se li dona sobretot importància a la empresa que desenvolupa el joc, però Qui publica el joc importa una mica més que en el model anterior. Els resultats son els següents

Developer	22.3%
Publisher	16.02%
Year of Release	14.84%
Platform	13.21%
Critic Score	11.50%
Genre	9.21%
User Score	8.01%
Rating	4.76%

Taula 4. Valors de les mètriques del classificador de vendes a EU

En canvi a JP, les importàncies canvien bastant. En aquest classificador la plataforma sobre la que surt el joc és la més important, seguit de qui publica el joc i qui el desenvolupa. Els resultats son els següents:

Platform	18.44%
Publisher	16.64%
Developer	16.06%
Year of Release	14.54%
Critic Score	11.77%
Genre	8.62%
Rating	7.01%
User Score	6.87%

Taula 5. Valors de les mètriques del classificador de vendes a JP

5 CONCLUSIONS

El classificador RandomForest ha demostrat ser un model que és capaç de predir d'una forma bastant certera quan un joc, donades diferents característiques d'aquest, serà un joc exitós o no, per tant es donen per vàlides les relacions de importància alhora de fer la predicció. Això es pot mostrar en els elevats percentatges obtinguts tant al Accuracy i recall com F1-score.

Inicialment el projecte només estava plantejat per tal de fer una predicció sobre les dades globals, però al provar els diferents sectors de vendes, s'ha apreciat clarament que no és un enfoc del tot precís, degut a que les preferències dels videojugadors poden arribar a ser molt variades en funció del seu lloc d'origen.

Donats els resultats tant semblants entre les prioritats alhora de comprar jocs que hi ha entre NA i EU i les grans diferències que hi ha entre aquestes dues últimes i JP, dona a entendre que les prioritats poden arribar a dependre de si vius en una zona amb una cultura Oriental o en una zona amb cultura Occidental.

Seguint amb aquest raonament, podem apreciar que al Occident, els equips desenvolupadors de joc s'han guanyat una gran reputació, ja que en moltes ocasions la predicció es basa en aquests per fer la predicció. Traduit al món real, implica que els jugadors occidentals es basen molt en aquests equips alhora d'adquirir els seus productes.

En canvi els jugadors Orientals no es basen tant com els occidentals en l'equip desenvolupador, tot i que igualment es un factor important alhora de decidir, si no que també els hi importa la plataforma en la que es facin els jocs.

Ara bé, quan es llença un joc que té com objectiu a tots els possibles mercats de la indústria, com a aspectes generals quan més renom tingui l'equip desenvolupador, juntament amb la empresa que publica el joc i la plataforma seran factors determinants de cara a l'èxit del joc

BIBLIOGRAFIA

- [1]El cas Kaggle utilitzat per obtenir totes les dades es: [Video Game Sales with Ratings \(kaggle.com\)](https://www.kaggle.com/datasets/robikscube/video-game-sales-with-ratings)