

Analyticsify

El poder de los datos

Sistema de Recomendación Video juegos

Elaborado por: Rafael Eduardo Diaz



UNIVERSITAT DE
BARCELONA

INDICE

- ❑ Antecedentes
- ❑ Sistemas de Recomendación
- ❑ Teoría de los SR
- ❑ Extracción de los datos API Steam
- ❑ Análisis Exploratorio
- ❑ Ajuste del sistema de recomendación
- ❑ Modelo de Negocio
- ❑ Viabilidad Financiera
- ❑ Conclusiones

ANTECEDENTES

- ❑ Cifras de la firma Statista revelan que la industria de los videojuegos alcanzo un valor aproximado de 198 mil 500 millones de dólares se prevé que para 2027, su valor incremente 141 mil 600 millones mas, un punto donde las empresas están apuntando para aumentar su mercado.
- ❑ En los 90 Valve se enfrentaba a la mentalidad de un mercado en donde los juegos se pirateaban con facilidad, y en donde todo el mundo era libre de hacer lo que quería con ellos, sin tener que crear cuentas ni conectarse online. Pero esa visión no era rentable económicamente para las compañías.
- ❑ Como hizo Nintendo en los 80 cuando salvó el mercado de videojuegos en Estados Unidos con Super Mario Bros y la NES, hundido tras el fracaso de Atari y el escándalo de E.T., Steam salvó a los videojuegos de PC a principios del milenio. Aún con sus fallos y sus polémicas, es justo reconocer su inmensa aportación a la historia de los juegos de PC.

ANTECEDENTES

- ❑ En 2003, se lanza Steam, pero no como una tienda de videojuegos, sino como un software para instalar parches automáticamente en los juegos, y con medidas anti-cheaters. Aunque al principio Steam no la tuvo fácil poco a poco, se fue ganando a los usuarios por sus servicios y comodidad.
- ❑ Esto se convirtió en un éxito lo que llevó a Valve a plantearse usar Steam para vender juegos. Actualmente Steam tiene unos 325 millones de usuarios y más de 50.000 juegos. Se lanzan más de 8.000 juegos año.
- ❑ Debido a esta gran cantidad de oferta de videojuegos, y de venta online, es difícil para el usuario encontrar productos que satisfagan sus necesidades.
- ❑ Nacen los **sistemas de recomendación (SR)** que ayudan a los usuarios a encontrar los productos de su interés, mejorando su satisfacción y generando rentabilidad para las empresas que lo implementan.

Como surgen los SR

En los años **80** nacen los primeros estudios se originaron en varias áreas, como las ciencias cognitivas, la teoría de la aproximación, la recuperación de información, la teoría de la previsión, la administración y el marketing.

Tapestry en el 93 elabora un sistema que almacena el conocimiento del usuarios. En 1996 MoviLens crea a Movifinder.

Netflix emplea un modelo de transfer learning, que aprende de una fuente y mejora con ello los resultados: estudia qué títulos de su catálogo conectan con los contenidos originales de la plataforma y qué tipo de visión generaran esos contenidos.

La explosión máxima de la investigación en SR se produjo cuando Amazon lanzó su método Filtros Colaborativos (FC) a fines de la década de 1990, aumentando con éxito sus ventas

El Premio Netflix1 en 2006 y otorgo 1 millón de dólares estadounidenses al ganador de la competencia que proporcionó el mejor sistema de recomendación para películas.

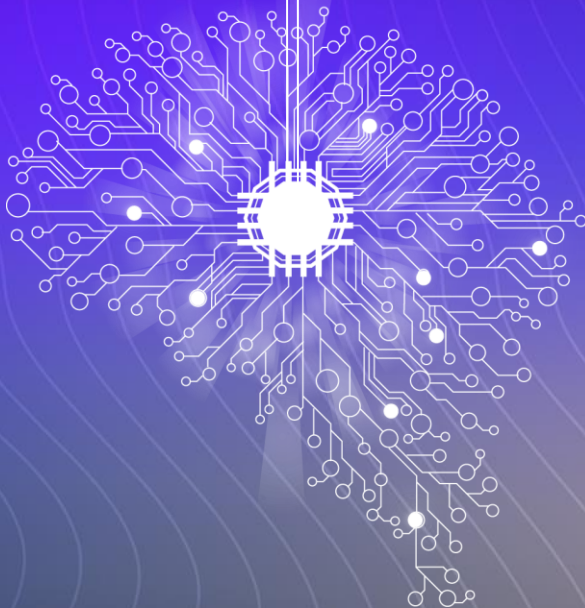
Sistema de Recomendación Basado en Filtros Colaborativos

Popular (POPULAR): Es un algoritmo no personalizado que recomienda a todos los usuarios los elementos más populares que aún no han calificado.

Aleatorio (RANDOM): Crea recomendaciones aleatorias que se pueden utilizar como referencia para la evaluación del algoritmo de recomendación.

Alternative Last Square (ALS): El algoritmo ALS busca relaciones de forma colaborativa, sea de forma user – user o ítem – ítem.

Recomendaciones híbridas (HybridRecommender): agrupa las recomendaciones de varios algoritmos ponderándolas por un vector de pesos.



El recomendador basado en reglas de asociación (AR): usa reglas de asociación para encontrar elementos recomendados.

Singular Value Decomposition (SVD): Este algoritmo se basa en la Factorización de Matrices.

El filtrado colaborativo basado en usuarios (UBCF): Encontrar usuarios similares y recomendar en base a los vecinos más cercanos.

El filtrado colaborativo basado en elementos (IBCF): Encontrar similitud entre ítems y realizar recomendaciones en base a los ítems más similares a los preferidos por el usuario.

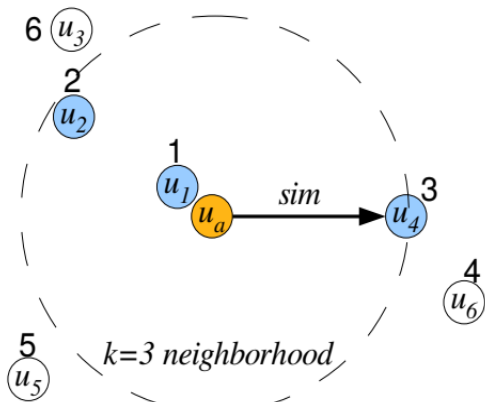
Teoría Filtros Colaborativos

R	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	?	4.0	4.0	2.0	1.0	2.0	?	?
u_2	3.0	?	?	?	5.0	1.0	?	?
u_3	3.0	?	?	3.0	2.0	2.0	?	3.0
u_4	4.0	?	?	2.0	1.0	1.0	2.0	4.0
u_5	1.0	1.0	?	?	?	?	?	1.0
u_6	?	1.0	?	?	1.0	1.0	?	1.0
u_a	?	?	4.0	3.0	?	1.0	?	5.0
\hat{r}_a	3.5	4.0		2.3		2.0		

(a)

S_a	u_a
u_1	0.3
u_2	1.0
u_3	0.2
u_4	0.3
u_5	0.1
u_6	0.1

(b)



(c)

$$sim_{Jaccard}(x, y) = \frac{|x \cap y|}{|x \cup y|}, \quad (7)$$

Las reglas de asociación son de la forma $x \rightarrow y$ donde $x, y \subseteq \mathcal{I}$ y $x \cap y = \emptyset$.

$$soporte(x \rightarrow y) = soporte(x \cap y) = \frac{Freq(x \cap y)}{|\mathcal{D}|} = \hat{P}(E_x \cap E_y)$$

$$confianza(x \rightarrow y) = \frac{soporte(x \cap y)}{soporte(x)} = \hat{P}(E_y | E_x)$$

Extracción de los Datos API Steam

API Steam



Información video juegos

3.691 juegos de la
ultima decada, que
sean pagos y tengan
almenos 25 reseñas.

01

02

Reseñas

1.933.168 reseñas.
Que sean en español.
Le gusto o No el video
juego.
Foco LATAM y España

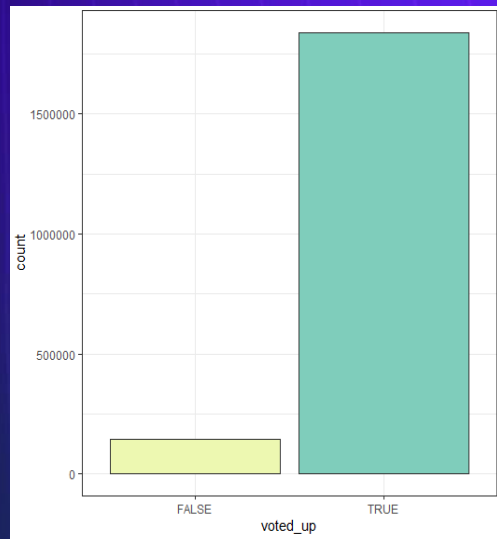


Usuarios

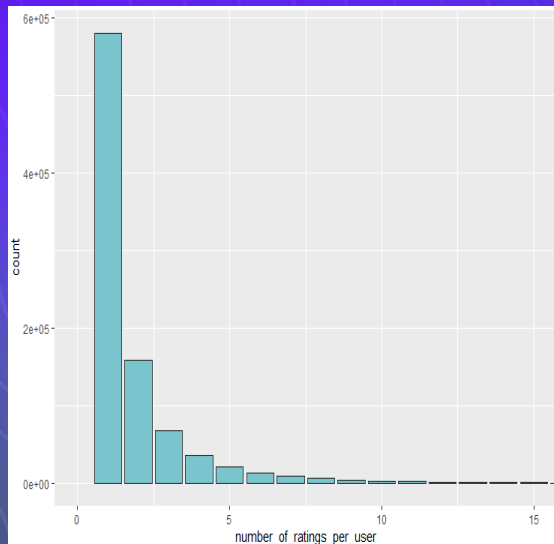
14.058 Usuarios
que tienen
almenos 10
reseñas

03

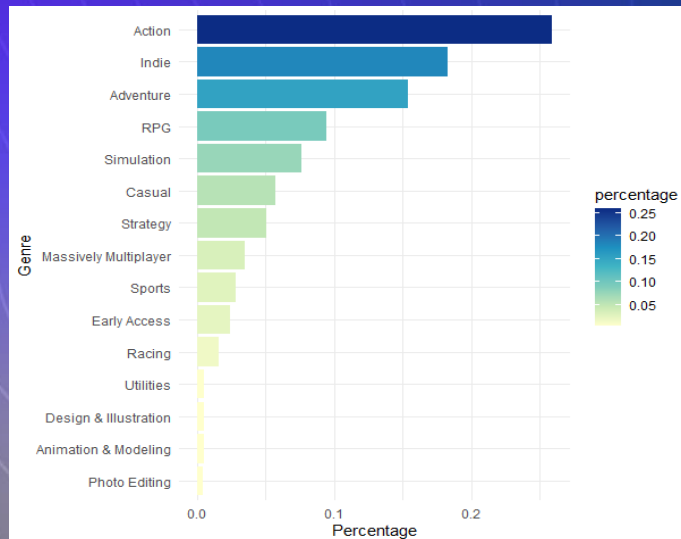
Análisis Exploratorio de Datos



El 93% de las reseñas son positivas mientras que el 7% son negativas.



El grueso de los usuarios se concentra en 5 calificaciones o menos.



En la Figura 4.5 se visualiza la proporción de juegos por género, donde el 26.2% son de acción, seguido de indie con un 18% y aventura con el 15.4%, completan en top 3.

Análisis Exploratorio de Datos

	TP	FP	FN	TN	N	Precision	Recall	TPR	FPR	N
RANDOM										
	0.0060412	2.993959	8.449325	3669.551	3681	0.0020137	0.0007807	0.0007807	0.0008152	3
	0.0102345	4.989765	8.445132	3667.555	3681	0.0020469	0.0012477	0.0012477	0.0013587	5
	0.0159204	6.984080	8.439446	3665.561	3681	0.0022743	0.0019942	0.0019942	0.0019017	7
	0.0223881	9.977612	8.432978	3662.567	3681	0.0022388	0.0028225	0.0028225	0.0027168	10
	0.0339730	14.966027	8.421393	3657.579	3681	0.0022649	0.0042684	0.0042684	0.0040751	15
POPULAR										
	0.2212509	2.778749	8.234115	3669.766	3681	0.0737503	0.0355614	0.0355614	0.0007565	3
	0.3379531	4.662047	8.117413	3667.883	3681	0.0675906	0.0536785	0.0536785	0.0012692	5
	0.4410803	6.558920	8.014286	3665.986	3681	0.0630115	0.0697177	0.0697177	0.0017857	7
	0.5881308	9.411869	7.867235	3663.133	3681	0.0588131	0.0924902	0.0924902	0.0025624	10
	0.8218195	14.178180	7.633547	3658.366	3681	0.0547880	0.1271445	0.1271445	0.0038601	15
ALS										
	0.3927505	2.607249	8.062615	3669.937	3681	0.1309168	0.0651118	0.0651118	0.0007097	3
	0.5927505	4.407249	7.862616	3668.137	3681	0.1185501	0.0955206	0.0955206	0.0011997	5
	0.7622601	6.237740	7.693106	3666.307	3681	0.1088943	0.1212204	0.1212204	0.0016980	7
	0.9841507	9.015849	7.471215	3663.529	3681	0.0984151	0.1545182	0.1545182	0.0024543	10
	1.2977967	13.702203	7.157569	3658.842	3681	0.0865198	0.1991590	0.1991590	0.0037302	15
AR										
	0.3273632	1.411585	8.128003	3671.133	3681	0.1833382	0.0523784	0.0523784	0.0003843	3
	0.3743426	1.651883	8.081024	3670.893	3681	0.1790511	0.0576685	0.0576685	0.0004497	5
	0.3887704	1.729567	8.066596	3670.815	3681	0.1778799	0.0590435	0.0590435	0.0004709	7
	0.3987918	1.772566	8.056574	3670.772	3681	0.1775978	0.0598789	0.0598789	0.0004826	10
	0.4049751	1.803909	8.050391	3670.741	3681	0.1774924	0.0603385	0.0603385	0.0004911	15

	TP	FP	FN	TN	N	Precision	Recall	TPR	FPR	N
UBCF										
	0.0482587	2.951741	8.407107	3669.593	3681	0.0160862	0.0057518	0.0057518	0.0008037	3
	0.0820896	4.917910	8.373276	3667.627	3681	0.0164179	0.0100028	0.0100028	0.0013390	5
	0.1196162	6.880384	8.335750	3665.664	3681	0.0170880	0.0150997	0.0150997	0.0018734	7
	0.1766880	9.823312	8.278678	3662.721	3681	0.0176688	0.0220515	0.0220515	0.0026747	10
	0.2764037	14.723596	8.178962	3657.821	3681	0.0184269	0.0360639	0.0360639	0.0040089	15
IBCF										
	0.3038380	2.696162	8.151528	3669.848	3681	0.1012793	0.0475632	0.0475632	0.0007340	3
	0.4611230	4.538877	7.994243	3668.006	3681	0.0922246	0.0725236	0.0725236	0.0012356	5
	0.5953802	6.404620	7.859986	3666.140	3681	0.0850543	0.0928030	0.0928030	0.0017436	7
	0.7772566	9.222743	7.678109	3663.322	3681	0.0777257	0.1208777	0.1208777	0.0025108	10
	1.0253731	13.974627	7.429993	3658.570	3681	0.0683582	0.1578295	0.1578295	0.0038045	15
HYBRID										
	0.4117271	2.588273	8.043639	3669.956	3681	0.1372424	0.0664518	0.0664518	0.0007045	3
	0.5670931	4.432907	7.888273	3668.112	3681	0.1134186	0.0899248	0.0899248	0.0012067	5
	0.6972992	6.302701	7.758067	3666.242	3681	0.0996142	0.1092622	0.1092622	0.0017158	7
	0.8686567	9.131343	7.586709	3663.413	3681	0.0868657	0.1347919	0.1347919	0.0024859	10
	1.1004264	13.899574	7.354940	3658.645	3681	0.0733618	0.1687342	0.1687342	0.0037840	15

Análisis Exploratorio de Datos

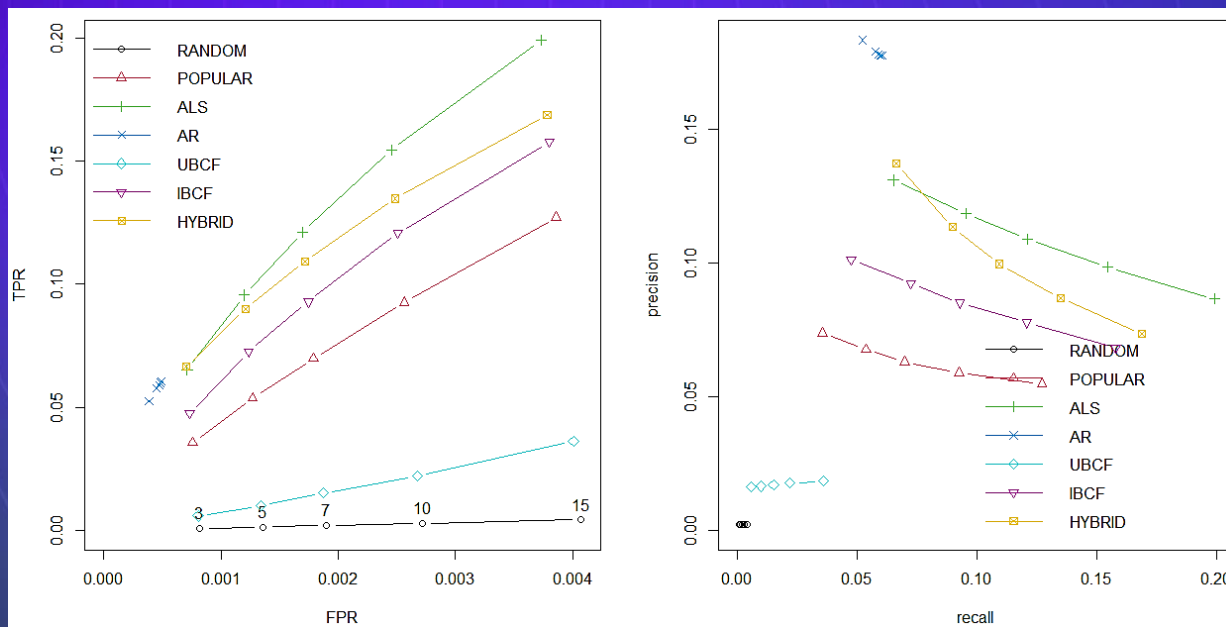


Figura 4.7: Curva ROC (izquierda) y precisión vs recall (derecha).

ML-OPS (Puesta en Producción)



Modelación

- Base de 259k ratings
- Se particiona en train (80%) y test (20%)
- Validación cruzada de 10 capas.
- 7 Algoritmos diferentes
- Con $n=[3,5,7,10 \text{ y } 15]$



Predicción

- Predicción con el modelo híbrido de top 10 de recomendaciones



Despliegue

Se despliega el modelo con el framework R Shiny



- 76561198069541163
- 76561197960284889

Análisis de viabilidad

Modelo de Negocio

Crear una empresa de soluciones analíticas

Conclusiones

Que aporte hizo este proyecto



Benchmarking

Comparación profunda de la competencia

Análisis Viabilidad

Determinar si el proyecto es rentable

Modelo de Negocio

Paso 1

Matricula mercantil y RUES

Paso 2

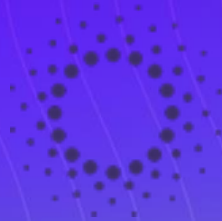
Se ubica en Bogotá, edificio BlueWork

Paso 3

Inversion de los socios y prestamo con el banco

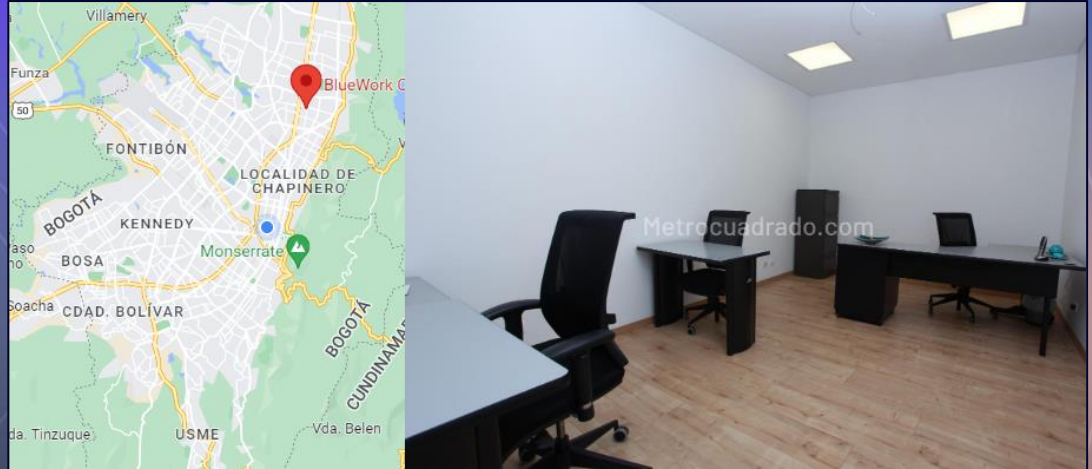
Paso 4

Contratación personal y puesta en marcha de los proyectos



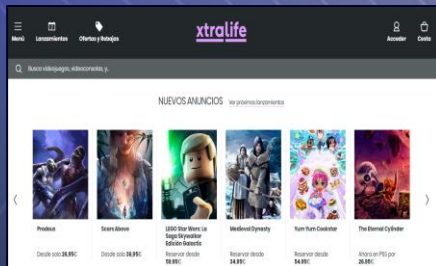
Analyticsify

El poder de los datos



Benchmarking

Plataformas de videojuegos



Juegos Digitales Colombia

Xtralife

Empresas de Analítica



Análisis de Viabilidad

Cargo	No	Sueldos	Prestaciones	Total Sueldo	Año 1	Año 2	Año 3	Año 4	Año 5
Data Scientist	1	\$ 1.300	\$ 674	\$ 1.974	\$ 23.690	\$ 30.797	\$ 40.035	\$ 52.046	\$ 67.660
Desarrollador Web	1	\$ 800	\$ 415	\$ 1.215	\$ 14.578	\$ 18.952	\$ 24.637	\$ 32.028	\$ 41.637
Call Center & Support	1	\$ 450	\$ 233	\$ 683	\$ 8.200	\$ 10.660	\$ 13.858	\$ 18.016	\$ 23.421
Total	3	\$ 2.550	\$ 1.322	\$ 3.872	\$ 46.468	\$ 60.409	\$ 78.531	\$ 102.091	\$ 132.718

Costos Fijos					
Conceptos / Año	Año 1	Año 2	Año 3	Año 4	Año 5
Salarios	\$ 46.468,1	\$ 60.408,6	\$ 78.531,2	\$ 102.090,5	\$ 132.717,7
Arriendo	\$ 4.440,0	\$ 4.688,6	\$ 4.951,2	\$ 5.228,5	\$ 5.521,3
Servicio de Agua	\$ 180,0	\$ 190,1	\$ 200,7	\$ 212,0	\$ 223,8
Servicio de Luz	\$ 360,0	\$ 380,2	\$ 401,4	\$ 423,9	\$ 447,7
Servicio de Internet + Telefonía	\$ 421,8	\$ 445,4	\$ 470,4	\$ 496,7	\$ 524,5
Posicionamiento web	\$ 213,6	\$ 225,6	\$ 238,2	\$ 251,5	\$ 265,6
Subtotal	\$ 52.083,5	\$ 66.338,4	\$ 84.793,1	\$ 108.703,1	\$ 139.700,6

Costos Variables					
Conceptos / Año	Año 1	Año 2	Año 3	Año 4	Año 5
Uso de Google Cloud Computing	\$ 7.200,0	\$ 7.603,2	\$ 8.029,0	\$ 8.478,6	\$ 8.953,4
Transportes	\$ 100,0	\$ 105,6	\$ 111,5	\$ 117,8	\$ 124,4
Mantenimiento PC	\$ 200,0	\$ 211,2	\$ 223,0	\$ 235,5	\$ 248,7
Comisiones ejecutivo comercial	\$ 3.900,0	\$ 4.118,4	\$ 4.349,0	\$ 4.592,6	\$ 4.849,8
Subtotal	\$ 11.400,0	\$ 12.038,4	\$ 12.712,6	\$ 13.424,5	\$ 14.176,2

Nota: Todos los valores están dados en USD con una TRM de \$4.600 COP

Costos Totales	\$ 63.483,5	\$ 78.376,8	\$ 97.505,6	\$ 122.127,6	\$ 153.876,8
-----------------------	--------------------	--------------------	--------------------	---------------------	---------------------

Ingresos por soluciones

01

Sistema de Recomendacion

Para mejorar las ventas y fidelización de los clientes

\$10.000

02

Consultoria & Modelos ML

Consultorias de todos los temas relacionados con Datos

\$500

03

Página Web

Hechas a la medida del cliente

\$1.500

ESTADO DE RESULTADOS						
	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5
Ventas		\$ 50.000,0	\$ 88.492,8	\$ 157.234,2	\$ 227.391,4	\$ 240.125,3
-Costo de Ventas		\$ 63.483,5	\$ 78.376,8	\$ 97.505,6	\$ 122.127,6	\$ 153.876,8
=Utilidad en ventas		-\$ 13.483,5	\$ 10.116,0	\$ 59.728,5	\$ 105.263,8	\$ 86.248,5
- Egresos operacionales		\$ 32.828,5	\$ 34.141,7	\$ 35.507,3	\$ 36.927,6	\$ 38.404,7
		\$ 32.083,5	\$ 33.396,7	\$ 34.762,3	\$ 36.182,6	\$ 37.659,7
Depreciación		\$ 745,0	\$ 745,0	\$ 745,0	\$ 745,0	\$ 745,0
Amortización		\$ -	\$ -	\$ -	\$ -	\$ -
=Utilidad operacional		-\$ 46.312,1	-\$ 24.025,7	\$ 24.221,2	\$ 68.336,2	\$ 47.843,8
+Otros ingresos		\$ -	\$ -	\$ -	\$ -	\$ -
-Otros Egresos		\$ -	\$ -	\$ -	\$ -	\$ -
=Utilidad Antes de impuestos		-\$ 46.312,1	-\$ 24.025,7	\$ 24.221,2	\$ 68.336,2	\$ 47.843,8
-Impuesto de renta 30%	30%	\$ -	\$ -	\$ 7.266,4	\$ 20.500,9	\$ 14.353,1
=Utilidad neta		-\$ 46.312,1	-\$ 24.025,7	\$ 16.954,8	\$ 47.835,3	\$ 33.490,6
BALANCE GENERAL						
	Año 0	Año 1	Año 2	Año 3	Año 4	Año 5
ACTIVO	2.510	- 26.796	- 7.932	36.390	76.401	50.344
Caja	500	- 45.067	- 68.348	- 43.382	18.433	46.521
Bancos	2.000	2.000	2.000	2.000	2.000	2.000
Oficina						
Muebles y enseres	-	-	-	-	-	-
Maquinaria y equipo	-	-	-	-	-	-
Equipo de computo	3.750	3.750	3.750	3.750	3.750	3.750
Dep.acum activo	-	745	1.490	2.235	2.980	3.725
Cargos diferidos preoperativos	-	-	-	-	-	-
PASIVO		17.016	13.594	9.668	5.565	-
Obligaciones financieras	-	17.016	13.594	9.668	5.565	-
PATRIMONIO	2.510	- 43.812	- 21.526	26.721	70.836	50.344
Impto por pagar		-	-	7.266	20.501	14.353
Derechos de Matricula + RUES	10					
Capital	2.500	2.500	2.500	2.500	2.500	2.500
Utilidades	-	- 46.312	- 24.026	16.955	47.835	33.491
Utilidad Neta Acumulada	-	- 46.312	- 70.338	- 53.383	- 5.548	27.943

Conclusiones

```
graph TD; A[Conclusiones] --- B[Los sistemas de recomendación basados en filtros colaborativos permiten a partir de información historica realizar recomendaciones personalizadas que mejoran las ventas y experiencia del usuario.]; A --- C[El software R, junto con el paquete recommenderlab, ofrece una estructura flexible y completa para evaluar todo tipo de modelos de recomendación. Y Shiny facilita el despliegue.]; A --- D[El mercado creciente de video juegos en LATAM y España ofrece una excelente oportunidad de negocio, creando soluciones analiticas que pueden ser muy rentables para las tiendas de viedeojuegos y para los que ofrecen estos sistemas.];
```

Los sistemas de recomendación basados en filtros colaborativos permiten a partir de información historica realizar recomendaciones personalizadas que mejoran las ventas y experiencia del usuario.

El software R, junto con el paquete recommenderlab, ofrece una estructura flexible y completa para evaluar todo tipo de modelos de recomendación. Y Shiny facilita el despliegue.

El mercado creciente de video juegos en LATAM y España ofrece una excelente oportunidad de negocio, creando soluciones analiticas que pueden ser muy rentables para las tiendas de viedeojuegos y para los que ofrecen estos sistemas.

GRACIAS

