

## PROPUESTA DE MONOGRAFÍA

Título del proyecto	Proyección de ganancias para una empresa de apuestas	
Estudiante 1		
Nombre completo	Andrés Julián Espinal Benjumea	e-mail: <a href="mailto:ajulian.espinal@udea.edu.co">ajulian.espinal@udea.edu.co</a>
		GitHub: <a href="https://github.com/AndresEspinal">https://github.com/AndresEspinal</a>
Nombre asesor	Yony Fernando Ceballos	e-mail: <a href="mailto:yony.ceballos@udea.edu.co">yony.ceballos@udea.edu.co</a>

### 1. Descripción del problema

El mercado de apuestas online en Colombia ha sido uno de los mayores sectores que ha crecido en el país en los últimos años, este está regulado por Coljuegos y se estima que para el año 2022 se reportaron apuestas por 26 billones de pesos, 67% más que el año anterior<sup>1</sup>, esta torta debe ser repartida por 16 casas de apuestas que operan en Colombia legalmente<sup>2</sup>.

Debido a que en la actualidad se encuentra una gran competencia y licencias de juegos muy costosos incluyendo impuestos, es importante que la empresa obtenga márgenes de ganancias alrededor del 8% mensual para que el negocio tenga viabilidad financiera por lo cual se desea conocer la proyección de estas ganancias de los próximos dos a tres trimestres (según los resultados obtenidos) y de manera indirecta conocer esos factores que logran o evitan cumplir con dichos objetivos.

### 2. Descripción del dataset

La base de datos de la empresa se obtiene de un sistema propio que controla todas las transacciones de apuestas. A continuación, se presentan las columnas que se van a manejar en la monografía:

- **Month:** Corresponde al mes y año en que se realizó la apuesta.
- **Group:** El cual es un resumen del deporte, país y liga. (Esta columna puede ser separada en 3 ya que muestra diferente información).
- **Sport:** Hace alusión al deporte.
- **Result type:** Me muestra en qué mercado se realizó la apuesta (ej.: Handicap, goles, triunfo).
- **Bet slip count:** Cantidad de tickets que se apostaron. (Dato no es 100% confiable ya que el sistema hace un aproximado).
- **No. of bets:** Dato no es 100% confiable.
- **Bet count:** Se refiere al número de apuestas que se realizó en el mercado en particular de result type.
- **Paidback:** Se refiere a la cantidad de dinero cancelado e invalidado al cliente por algún tipo de situación ya sea porque el evento así lo requiere o por algún tipo de error.
- **Cancelled:** Son apuestas que son canceladas por el mismo cliente en un lapso de 10 minutos desde que fue hecha ya sea porque se equivocó u otra situación particular.

---

<sup>1</sup>Obtenido de <https://www.elpais.com.co/economia/cuanto-han-crecido-las-apuestas-en-linea-colombia.html>

<sup>2</sup> Obtenido de <https://www.coljuegos.gov.co/publicaciones/301841/juegosonline/>

- **Won:** Se refiere a la cantidad de dinero ganado por parte del cliente contando con lo que apostó.
- **Total stake:** Es la cantidad de dinero que el cliente apuesta.
- **Profit:** Son las ganancias obtenidas por el cliente en caso de que la apuesta es ganadora y se podría decir que es lo ganó (Won) menos lo que apostó (Total stake).
- **Profit-rate:** Que sería el Profit dividido el Total stake.
- **Bonus:** Pequeño porcentaje que el cliente obtiene al hacer uso de apuestas combinadas.
- **Paidout:** Es lo pagado en dinero real al cliente que se desglosa de Won más Bonus.

Por ahora se ha decidido que estas variables pueden ser de interés para la obtención de la proyección, pero es posible usar nuevas variables que el sistema permita arrojar según se vaya desarrollando el proyecto.

La base de datos cuenta con 15 columnas y 262.516 filas

### 3. Métricas de machine learning:

La base de datos actual presenta muchas diferencias entre la cantidad de dinero que los clientes depositan para hacer sus apuestas. Esto genera muchas diferencias a la hora de sacar una buena base de datos y es posible que se entiendan como datos atípicos.

Por tal motivo la MAE<sup>3</sup> será usada por ahora como la métrica de predicción dado que no es tan sensible a los datos atípicos y puede entender mejor el modelo de predicción sin ir a hacer métricas desajustadas.

### 4. Métricas de negocio:

Se espera que el modelo arroje una predicción cercana al 80% ya que se requiere preparar estrategias para contrarrestar pérdidas obtenidas en diferentes meses.

### 5. Referencias

Coljuegos. (23 de 01 de 2023). [www.coljuegos.gov.co](http://www.coljuegos.gov.co). Obtenido de [www.coljuegos.gov.co](http://www.coljuegos.gov.co): <https://www.coljuegos.gov.co/publicaciones/301841/juegosonline/>

El País S.A. (01 de 04 de 2023). [www.elpais.com.co](http://www.elpais.com.co). Obtenido de [www.elpais.com.co](http://www.elpais.com.co): <https://www.elpais.com.co/economia/cuanto-han-crecido-las-apuestas-en-linea-colombia.html>

sitiobigdata. (27 de 05 de 2019). [sitiobigdata.com](http://sitiobigdata.com). Obtenido de [sitiobigdata.com](http://sitiobigdata.com): <https://sitiobigdata.com/2019/05/27/aprendizaje-automatico-seleccionando-metricas-regresion/#>

---

<sup>3</sup> Obtenido de <https://sitiobigdata.com/2019/05/27/aprendizaje-automatico-seleccionando-metricas-regresion/#>