

Métodos de clasificación

ANALÍS DE DATOS

LUCIANO ALFONSO DENIS ARREDONDO

Contenido

Introducción.....	2
Selección de datos	3
Categorías y Umbrales.....	4
Método de clasificación	6
Regresión logística.....	6
Gráfico Cartesiano.....	7
Matriz de confusión.....	8
Conclusión.....	10

Introducción

En el vasto mundo de los videojuegos competitivos, pocos títulos han alcanzado la magnitud y la popularidad de League of Legends (LoL). Con millones de jugadores participando en emocionantes batallas estratégicas, LoL se ha convertido en un fenómeno global que va más allá del simple entretenimiento. La competitividad feroz, las tácticas elaboradas y la toma de decisiones instantánea han elevado a este juego a un nivel donde cada movimiento puede cambiar el rumbo de una partida.

En este contexto, el análisis de datos se ha erigido como una herramienta esencial para los jugadores y estrategias de LoL. En particular, el estudio de métricas específicas durante los primeros 10 minutos de una partida puede proporcionar *insights* cruciales sobre el rendimiento de los equipos y sentar las bases para el desarrollo del juego. En este trabajo, nos sumergimos en la riqueza de datos generada por las partidas de League of Legends, centrándonos en dos métricas clave: "blueKills" y "blueExperienceDiff"

Selección de datos

La elección de los datos en nuestro análisis se centra en dos métricas cruciales dentro del juego League of Legends: "blueKills" y "blueExperienceDiff". Estas dos columnas, extraídas del conjunto de datos, no son meros indicadores superficiales, sino ventanas reveladoras hacia las dinámicas complejas y estratégicas que definen el desarrollo de una partida.

- **"blueKills"**: Más allá de ser un simple recuento de eliminaciones, "blueKills" encapsula la esencia de la destreza individual y la coordinación del equipo. Cada asesinato no solo representa una ventaja numérica momentánea, sino también un paso hacia el control del mapa y la consolidación de la posición estratégica.
- **"blueExperienceDiff"**: La acumulación de experiencia se convierte en un recurso invaluable. Esta métrica, al medir la disparidad en la experiencia entre los equipos azul y rojo, se convierte en un termómetro para evaluar la eficacia en el control del mapa y la recolección de recursos. Un liderazgo en experiencia no solo representa un avance numérico, sino también una posición estratégica superior que puede inclinar la balanza a favor del equipo azul.

Al seleccionar estas dos categorías, no solo buscamos analizar el desempeño cuantitativo, sino también capturar la esencia táctica y estratégica que impulsa el competitivo universo de League of Legends. Enfocándonos en "blueKills" y "blueExperienceDiff", buscamos desvelar los matices que delinean las victorias y derrotas, y cómo estas variables pueden servir como faros para futuras decisiones tácticas.

Categorías y Umbrales

La creación de categorías y la definición de umbrales no son solo pasos técnicos en nuestro análisis; son la esencia misma de la interpretación y contextualización de los datos. En nuestro estudio sobre las partidas de League of Legends, las métricas de "blueKills" y "blueExperienceDiff" se vuelven más significativas cuando las clasificamos en categorías distintivas.

```
# Primero, calcularemos los promedios de 'blueKills' y 'blueExperienceDiff'
# y luego definiremos umbrales para categorizar las partidas.
kills_mean = all_sheets['blueKills'].mean()
kills_std = all_sheets['blueKills'].std()
exp_diff_mean = all_sheets['blueExperienceDiff'].mean()
exp_diff_std = all_sheets['blueExperienceDiff'].std()

# Definimos los umbrales para las categorías elegidas
thresholds = {
    'kills': {
        'Excelente': kills_mean + kills_std,
        'Medio': kills_mean,
        'Bajo': kills_mean - kills_std
    },
    'exp_diff': {
        'Excelente': exp_diff_mean + exp_diff_std,
        'Medio': exp_diff_mean,
        'Bajo': exp_diff_mean - exp_diff_std
    }
}
```

Código 1 – Umbrales

En el código presentado, calculamos los promedios y desviaciones estándar de ambas métricas para establecer umbrales que delinean el rendimiento de las partidas. Estos umbrales son puntos de referencia que nos permiten categorizar las partidas en segmentos significativos. Por ejemplo, si un equipo logra un número de asesinatos superior al promedio más una desviación estándar, consideramos esa partida como "Excelente". De manera similar, si la diferencia de experiencia acumulada supera la media más una desviación estándar, también la clasificamos como "Excelente". Estos umbrales no son simplemente números; representan los puntos donde el rendimiento se destaca y se vuelve digno de una atención especial.

	blueKills	blueExperienceDiff	performance_category
0	6	2378	Excelente
1	3	-539	Medio
2	12	359	Excelente
3	4	-2052	Medio
4	7	-864	Medio
5	7	-139	Medio
6	6	2242	Excelente
7	4	-875	Medio
8	5	939	Medio
9	4	-3582	Medio

Tabla 1 - Tabla ejemplo de umbrales

Las categorías de **"Excelente"**, **"Medio"** y **"Bajo"** van más allá de ser simples etiquetas. Reflejan el rendimiento relativo de un equipo en comparación con la media de la población. Una partida clasificada como "Excelente" no solo significa que un equipo ha alcanzado un alto número de asesinatos o una gran diferencia de experiencia, sino que también ha superado significativamente a la mayoría de las partidas analizadas. Estas categorías se convierten en herramientas interpretativas poderosas que permiten una comprensión rápida y clara del rendimiento.

En resumen, la definición de umbrales y categorías no es solo una tarea técnica; es el proceso de traducir números fríos en narrativas estratégicas. Estos umbrales nos permiten identificar no solo el rendimiento excepcional sino también las áreas que necesitan mejora. Las categorías son los puentes que conectan la abstracción de los datos con la realidad táctica de League of Legends.

Método de clasificación

La selección de un método de clasificación adecuado es fundamental para extraer insights significativos de los datos. En este estudio, hemos aplicado cuatro métodos distintos: Regresión Logística, Análisis de Discriminante Lineal (LDA), Naive Bayes y k-Nearest Neighbors (k-NN). Sin embargo, en este análisis, se centrará netamente en la **Regresión Logística**, un método poderoso y versátil.

```
## Visualizamos los límites de decisión para Regresión Logística en los datos reducidos a 2D
plt.figure(figsize=(10, 8))
plot_decision_boundaries_very_low_res(X_train_r, y_train_r, LogisticRegression(max_iter=1000),
"Regresión Logística")
plt.show()

# Además, calculamos las métricas para el modelo de Regresión Logística
logistic_model = LogisticRegression(max_iter=1000)
logistic_model.fit(X_train_r, y_train_r)
y_pred_logistic = logistic_model.predict(X_test_r)

# Calculamos las métricas de evaluación
logistic_accuracy = accuracy_score(y_test_r, y_pred_logistic)
logistic_f1 = f1_score(y_test_r, y_pred_logistic, average='weighted')
logistic_cm = confusion_matrix(y_test_r, y_pred_logistic)
logistic_roc_auc = roc_auc_score(y_test_r, logistic_model.predict_proba(X_test_r), multi_class='ovr')

# Mostramos las métricas
logistic_metrics = {
    'Accuracy': logistic_accuracy,
    'F1 Score': logistic_f1,
    'Confusion Matrix': logistic_cm,
    'ROC AUC': logistic_roc_auc
}
```

Código 2 - Regresión Logística

Regresión logística

La Regresión Logística se destaca por su capacidad para modelar y predecir la probabilidad de que una variable categórica binaria ocurra. En nuestro contexto, hemos aplicado la Regresión Logística para categorizar las partidas de League of Legends en las etiquetas de rendimiento que definimos anteriormente ("Excelente", "Medio", "Bajo"). Aunque la Regresión Logística comparte su nombre con la regresión lineal, su aplicación difiere significativamente. En lugar de ajustar una línea recta a los datos, modela la probabilidad logarítmica de una variable dependiente binaria.

Las razones principales del por qué se utilizó la regresión logística como método son

- Proporciona coeficientes que representan la relación entre cada variable independiente y la probabilidad logarítmica de la variable dependiente. En nuestro caso, estos coeficientes podrían revelar cómo "blueKills" y "blueExperienceDiff" influyen en la probabilidad de clasificación en diferentes categorías de rendimiento.
- En el gráfico de límites de decisión, podemos observar cómo la Regresión Logística demarca las regiones en el espacio de características donde se predice una categoría específica. Este método es fundamental para la interpretación visual de cómo las variables de entrada contribuyen a la clasificación.
- Utilizamos métricas como la precisión del modelo para evaluar su rendimiento en el conjunto de prueba. La precisión nos dice cuántas predicciones fueron correctas en comparación con el

total de predicciones. Esta métrica es crucial para entender la utilidad práctica de nuestro modelo de Regresión Logística en el contexto de League of Legends.

Las métricas de este análisis para el modelo de regresión logística son:

- **Accuracy (Precisión):** 93.75% - Esta métrica indica la proporción de predicciones correctas en comparación con el total de predicciones realizadas por el modelo.
- **F1 Score:** 93.81% - Un promedio ponderado que combina precisión y exhaustividad, proporcionando una medida equilibrada del rendimiento del modelo.
- **ROC AUC:** 98.64% - El área bajo la curva ROC (Receiver Operating Characteristic) indica la capacidad del modelo para distinguir entre las clases.

En conjunto, estas métricas y la matriz de confusión nos ofrecen una visión completa del rendimiento del modelo de Regresión Logística en la clasificación de partidas de League of Legends. Se destacan tanto las fortalezas como las áreas de mejora, proporcionando una guía valiosa para futuras iteraciones y ajustes.

Gráfico Cartesiano

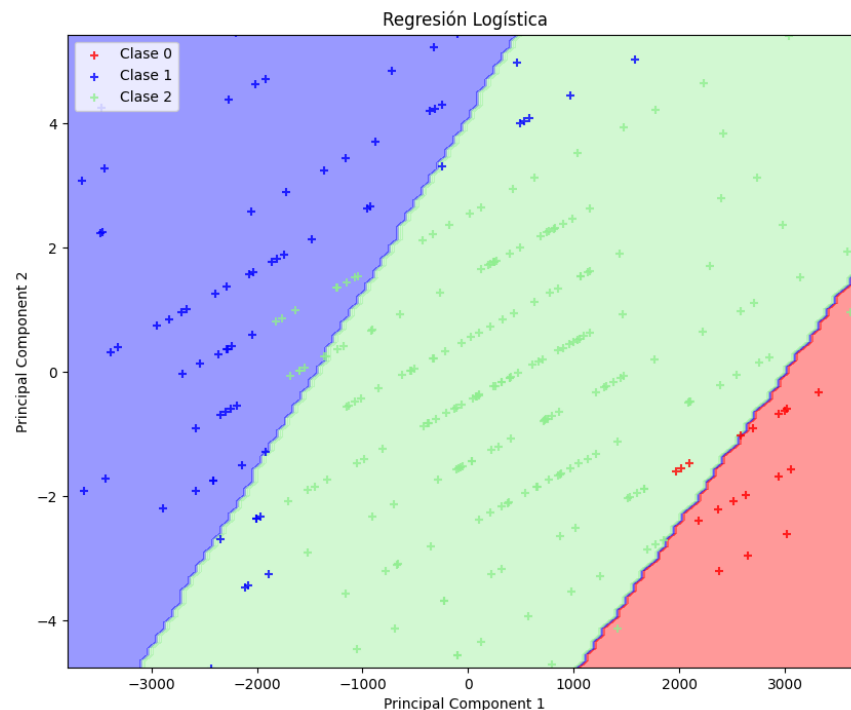


Figura 1 - Resultado Regresión Logística

En el gráfico cartesiano visualizado para el modelo de Regresión Logística, observamos los límites de decisión trazados en un espacio bidimensional. Cada punto en el gráfico representa una partida, proyectada en los componentes principales obtenidos mediante PCA. Los límites de decisión demarcan las regiones en las que el modelo clasifica las partidas en diferentes categorías de rendimiento, como "Excelente", "Medio" o "Bajo". Este gráfico no solo proporciona una visión visualmente atractiva del rendimiento del modelo, sino que también ilustra cómo la Regresión Logística segmenta el espacio de características para realizar predicciones precisas.

Clasificación de Datos:

- **Clase 0 (color azul):** Las observaciones clasificadas en la Clase 0 están principalmente en el lado izquierdo del gráfico. Esta clase puede representar un tipo específico de partidas según las métricas subyacentes (por ejemplo, partidas con un bajo número de asesinatos o una gran diferencia negativa de experiencia).
- **Clase 1 (color verde):** Las observaciones clasificadas en la Clase 1 ocupan la región central del gráfico, lo que podría interpretarse como partidas con estadísticas moderadas o promedio en cuanto a asesinatos y diferencia de experiencia.
- **Clase 2 (color rojo):** Las observaciones en la Clase 2 se encuentran a la derecha, lo que podría indicar partidas con un alto número de asesinatos o una gran diferencia positiva de experiencia.
- **Fronteras de Decisión:** Las líneas que separan los colores representan las fronteras de decisión del modelo de regresión logística. Estas líneas son el lugar donde el modelo estima una probabilidad igual entre las clases adyacentes.
- **Etiquetas de Observaciones:** Las cruces representan las observaciones individuales proyectadas en este espacio de componentes principales. Su ubicación en relación con las fronteras de decisión muestra la clasificación predicha por el modelo de regresión

Matriz de confusión

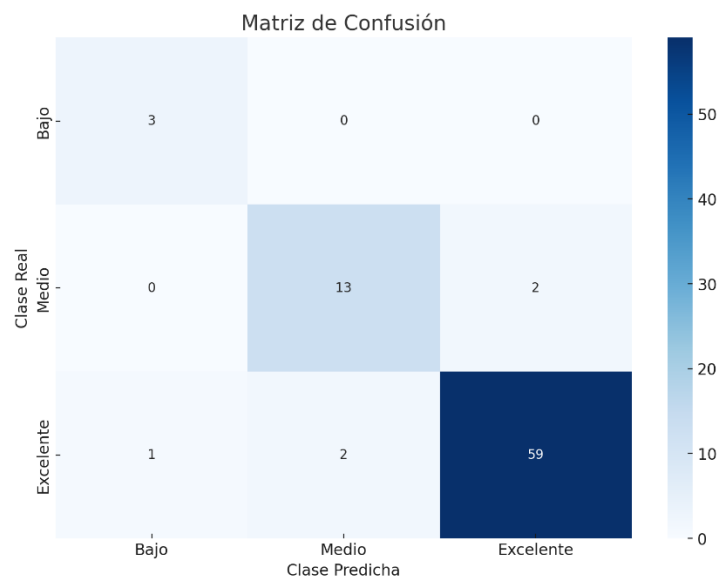


Figura 2 - Matriz de confusión

La matriz de confusión es una herramienta valiosa para entender la efectividad del modelo de Regresión Logística en términos de clasificación. En la matriz de confusión graficada, observamos que el modelo ha realizado predicciones precisas en la mayoría de las instancias de las cuales podemos interpretar lo siguiente:

- En la clase "Bajo", el modelo clasificó correctamente 3 partidas y no cometió errores de clasificación.

- Para la clase "Medio", el modelo acertó en 13 partidas, pero se equivocó en 2, clasificándolas como "Excelente".
- En la clase "Excelente", el modelo acertó en 59 partidas, pero cometió un error al clasificar una partida como "Bajo" y dos partidas como "Medio".

Estas cifras son fundamentales para evaluar la capacidad del modelo para distinguir entre las diferentes categorías. Además, las métricas adicionales, como la precisión (Accuracy) y el F1 Score, proporcionan una visión holística del rendimiento del modelo.

Conclusión

En el dinámico reino competitivo de League of Legends, donde la estrategia y la habilidad convergen en cada enfrentamiento, el análisis de datos se erige como un faro de conocimiento, iluminando patrones y tendencias que trascienden la superficie del juego. Al seleccionar cuidadosamente las variables "blueKills" y "blueExperienceDiff", no solo cuantificamos la destreza individual y la disparidad de experiencia, sino que también exploramos facetas tácticas fundamentales que definen el desempeño en el campo de batalla.

La definición de categorías ("Excelente", "Medio", "Bajo") y la asignación de umbrales estratégicos basados en promedios y desviaciones estándar agregan profundidad al análisis. Estas categorías no son etiquetas arbitrarias; son ventanas a la excelencia, el rendimiento medio y áreas de oportunidad. Los umbrales, actuando como puntos de referencia tácticos, permiten una clasificación más significativa y contextualizada, trazando un mapa del terreno estratégico de League of Legends.

La elección de la Regresión Logística como método de clasificación revela su poder interpretativo. Más allá de su capacidad para clasificar partidas, la Regresión Logística desglosa la influencia de las variables en la probabilidad logarítmica de rendimiento. Visualizar los límites de decisión en un gráfico cartesiano proporciona una comprensión intuitiva de cómo se forjan las decisiones estratégicas en el calor del enfrentamiento.

Los resultados, evidenciados en la matriz de confusión y métricas de evaluación, desglosan el rendimiento del modelo. Cada número en la matriz representa un acierto o una oportunidad de mejora, ofreciendo una visión valiosa para ajustes futuros. La precisión en las clases "Excelente", "Medio" y "Bajo" no solo es una métrica; es una ventana a las dinámicas tácticas que dan forma al juego.

Este análisis no es simplemente una instantánea del presente; es una brújula para futuras partidas. La comprensión de las estrategias que subyacen en el rendimiento no solo informa las tácticas actuales, sino que también proyecta luz sobre áreas de mejora y adaptación para el futuro.

En este cruce entre análisis de datos y estrategia en League of Legends, este estudio no solo clasifica partidas; desvela las complejidades tácticas que moldean el curso del juego. A medida que exploramos las posibilidades y aplicaciones de métodos analíticos, nos sumergimos más profundamente en la riqueza estratégica de este emocionante universo competitivo.