

# Análisis de Reglas de Asociación en Partidas de Ajedrez

## Implementación del Algoritmo Apriori

Alberto Bartolomé, Mario Marín

11 de junio de 2025

# Agenda

- 1 Introducción
- 2 Metodología
- 3 Resultados
- 4 Verificación de Hipótesis
- 5 Interpretación
- 6 Conclusiones

- **Implementación del Algoritmo Apriori**
  - Aplicación práctica para descubrir reglas de asociación
- **Análisis de Patrones de Juego**
  - Identificación de relaciones entre variables categóricas
- **Verificación de Hipótesis**
  - Validación estadística de reglas propuestas
- **Comparación de Modalidades**
  - Análisis diferencial entre ritmos de juego

## Características principales:

- **121,332 partidas** analizadas
- **11 variables** originales
- Datos de lichess.org (enero 2013)
- Sin valores faltantes

## Variables clave:

- Resultado
- Elo jugadores
- Control tiempo
- Terminación
- Movimientos

**Foco:** Ajedrez relámpago (600+0) - 2,452 partidas

## 1. Tratamiento de valores especiales:

- Conversión  $\dot{\text{L}}^{\text{en}}$  Elo  $\rightarrow$  900 puntos
- 78 valores en WhiteElo, 140 en BlackElo

## 2. Sistema de categorización:

### Categorías Elo:

- Principiante: 0-1199
- Intermedio: 1200-1599
- Avanzado: 1600-1999
- Experto: 2000-2399

### Duración partida:

- Corta:  $\dot{\text{J}}$ 20 movimientos
- Media: 20-39 movimientos
- Larga: 40-59 movimientos
- Muy larga: 60+ movimientos

## Generación de nuevas variables analíticas:

- **EloDiff**: Diferencia absoluta de Elo
- **EloDiff\_Cat**: Diferencia en categorías de Elo
- **JugadorMasFuerte**: Identificación del jugador dominante
- **Variables categóricas**: WhiteElo\_Cat, BlackElo\_Cat, MovesCount\_Cat

**Resultado:** Dataset procesado con **17 columnas** listo para análisis

## Principio básico:

- Todos los subconjuntos de un conjunto frecuente son frecuentes
- Búsqueda de patrones en datos transaccionales

## Métricas de evaluación:

- **Soporte:** Frecuencia de aparición del conjunto
- **Confianza:** Probabilidad condicional de ocurrencia
- **Lift:** Medida de independencia estadística

## Configuración utilizada:

- Soporte mínimo: 0.01 (1 %)
- Confianza mínima: 0.1 (10 %)

## Transformación a formato transaccional:

```
def preparar_datos_apriori(df_subset):  
    cols_categoricas = [  
        'Result', 'WhiteElo_Cat', 'BlackElo_Cat',  
        'MovesCount_Cat', 'ECO', 'Termination',  
        'JugadorMasFuerte'  
    ]  
  
    transacciones = []  
    for _, row in df_subset.iterrows():  
        transaccion = []  
        for col in cols_disponibles:  
            transaccion.append(f'{col}_{row[col]}')  
        transacciones.append(transaccion)
```

**Resultado:** 2,452 transacciones con 6.0 ítems promedio por transacción



## Características generales:

- **2,452 partidas** (2.02 % del total)
- Elo promedio: **1553** (ambos colores)
- Movimientos promedio: **32.9**
- Diferencia Elo promedio: **145**

## Distribución resultados:

- 1-0: **49.2 %**
- 0-1: **47.9 %**
- Tablas: **2.9 %**

**Observación:** Equilibrio entre victorias de blancos y negros, pocas tablas

## Ejecución exitosa:

- **873 conjuntos frecuentes** encontrados
- **9,346 reglas de asociación** generadas
- Matriz de codificación:  $2452 \times 20$

## Proceso de análisis:

- Codificación binaria de transacciones
- Aplicación de Apriori con soporte mínimo 0.01
- Extracción de reglas con confianza mínima 0.1
- Ordenamiento por lift descendente

# Top 3 Reglas Más Significativas

## 1. Regla de Tablas en Partidas Largas

- `MovesCount_Cat_Muy larga, Termination_Normal` → `Result_1/2-1/2`
- Confianza: **21.7 %** — Lift: **7.39**

## 2. Regla Inversa de Tablas

- `Result_1/2-1/2` → `MovesCount_Cat_Muy larga, Termination_Normal`
- Confianza: **38.9 %** — Lift: **7.39**

## 3. Asociación Tablas-Duración

- `Result_1/2-1/2` → `MovesCount_Cat_Muy larga`
- Confianza: **48.6 %** — Lift: **7.01**

## Reglas relacionadas con diferencias de Elo:

### 4. Intermedio vs Principiante

- `WhiteElo_Cat_Intermedio, JugadorMasFuerte_Blanco` → `BlackElo_Cat_Principiante`
- Confianza: **15.7 %** — Lift: **4.42**

### 5. Patrones de Victoria por Tiempo

- `JugadorMasFuerte_Negro, WhiteElo_Cat_Avanzado` → `Termination_Time forfeit`
- Confianza: **34.6 %** — Lift: **4.39**

**Interpretación:** Los jugadores más fuertes tienden a ganar, especialmente por tiempo

# H1: Diferencia 1 Categoría → Victoria del Más Fuerte

## Análisis estadístico:

- **1,014 partidas** con diferencia 1 categoría
- Representa el **41.4 %** del subconjunto

## Resultados por color:

### Jugador fuerte con blancas:

- 508 casos aplicables
- 349 victorias (**68.7 %**)
- Soporte: 0.207

### Jugador fuerte con negras:

- 506 casos aplicables
- 347 victorias (**68.6 %**)
- Soporte: 0.206

**Conclusión:** Regla MODERADA - Ventaja clara pero no absoluta

## H2: Diferencia 2 Categorías → Victoria del Más Fuerte

### Análisis estadístico:

- **31 partidas** con diferencia 2 categorías
- Casos muy específicos pero reveladores

### Resultados por color:

#### Jugador fuerte con blancas:

- 16 casos aplicables
- 15 victorias (**93.8 %**)
- Soporte: 0.0065

#### Jugador fuerte con negras:

- 15 casos aplicables
- 14 victorias (**93.3 %**)
- Soporte: 0.0061

**Conclusión:** Regla FUERTE - Diferencias grandes predicen victoria con alta confianza

# H3 y H4: Partidas Entre Grandes Maestros

## Análisis de disponibilidad:

- **0 partidas** GM vs GM en el subconjunto
- El ajedrez relámpago (600+0) no atrae a GMs de élite
- Niveles más comunes: Intermedio y Avanzado

## Distribución real por categorías:

- |                                      |                                       |
|--------------------------------------|---------------------------------------|
| • Intermedio: <b>57,571</b> partidas | • Experto: <b>3,550</b> partidas      |
| • Avanzado: <b>57,540</b> partidas   | • Principiante: <b>2,670</b> partidas |

**Conclusión:** Hipótesis no verificables en este subconjunto

## 1. Patrón de Tablas:

- Las tablas están fuertemente asociadas con partidas muy largas (60+ movimientos)
- Lift de 7.39 indica asociación muy significativa

## 2. Diferencia de Elo Predictiva:

- Diferencias 1 categoría: Victoria en 68.7 % de casos
- Diferencias 2 categorías: Victoria en 93.8 % de casos

## 3. Equilibrio por Color:

- Sin ventaja significativa de blancas en ajedrez relámpago
- Distribución casi 50-50 en resultados



## Para jugadores:

- La diferencia de nivel es predictiva del resultado
- Partidas largas aumentan probabilidad de tablas
- El tiempo de juego nivela las diferencias tradicionales

## Para plataformas de ajedrez:

- Sistemas de emparejamiento pueden usar estas reglas
- Predicción de duración de partidas
- Balanceado de competiciones

## Para investigación:

- Validación de algoritmos de minería de datos
- Patrones replicables en otros deportes/juegos

## Limitaciones técnicas:

- Solo una modalidad de tiempo analizada (600+0)
- Ausencia de partidas de élite (GM vs GM)
- Datos de un solo mes (enero 2013)

## Limitaciones metodológicas:

- Categorización simplificada de variables
- No se consideran factores como aperturas específicas
- Análisis descriptivo, no causal

## Oportunidades de mejora:

- Análisis longitudinal con múltiples meses
- Comparación entre diferentes modalidades
- Inclusión de más variables contextuales

## Sobre el algoritmo Apriori:

- **Implementación exitosa** en dataset real de ajedrez
- **9,346 reglas** generadas con patrones interpretables
- Eficiente para datasets de tamaño medio (2,452 registros)

## Sobre las reglas descubiertas:

- **Reglas estadísticamente significativas** con lift  $\geq 7$
- Patrones coherentes con conocimiento experto de ajedrez
- **Verificación parcial** de hipótesis propuestas

## Metodología validada:

- Preprocesamiento robusto de datos categóricos
- Sistema de verificación de hipótesis implementado
- Métricas de evaluación apropiadas

## **Extensiones inmediatas:**

- Análisis de modalidad bala (60+0) para comparación
- Implementación de otros algoritmos (FP-Growth, Eclat)
- Análisis temporal de evolución de patrones

## **Investigación avanzada:**

- Incorporación de análisis de aperturas (códigos ECO)
- Predicción de resultados usando machine learning
- Análisis de redes sociales entre jugadores

## **Aplicaciones prácticas:**

- Sistema de recomendaciones para jugadores
- Herramientas de análisis para entrenadores
- Algoritmos de emparejamiento inteligente

**¡Gracias por su atención!**