

INFORME COMPLETO

Sistema de Análisis de Noticias

Machine Learning - 9 Algoritmos

1. Regresión Logística

2. KNN

3. Naive Bayes

4. K-Means

5. Árbol de Decisión

6. ARIMA

7. Suavizado Exponencial

8. Random Forest

9. XGBoost

Fecha de Generación: 26/09/2025 12:20

Sistema de Scraping y Análisis de Noticias

RESUMEN EJECUTIVO

Este informe presenta un análisis exhaustivo de 9 algoritmos de Machine Learning aplicados al análisis de noticias. Los algoritmos evaluados incluyen técnicas de clasificación, clustering y series temporales.

OBJETIVOS:

- Clasificación automática de categorías
- Detección de calidad de contenido
- Análisis de sentimientos
- Predicción de engagement
- Clasificación de fuentes

METODOLOGÍA:

- Dataset: 401 artículos de noticias
- Preprocesamiento: TF-IDF vectorization
- Validación: 80% entrenamiento, 20% prueba
- Métricas: Precisión, F1-score, AUC

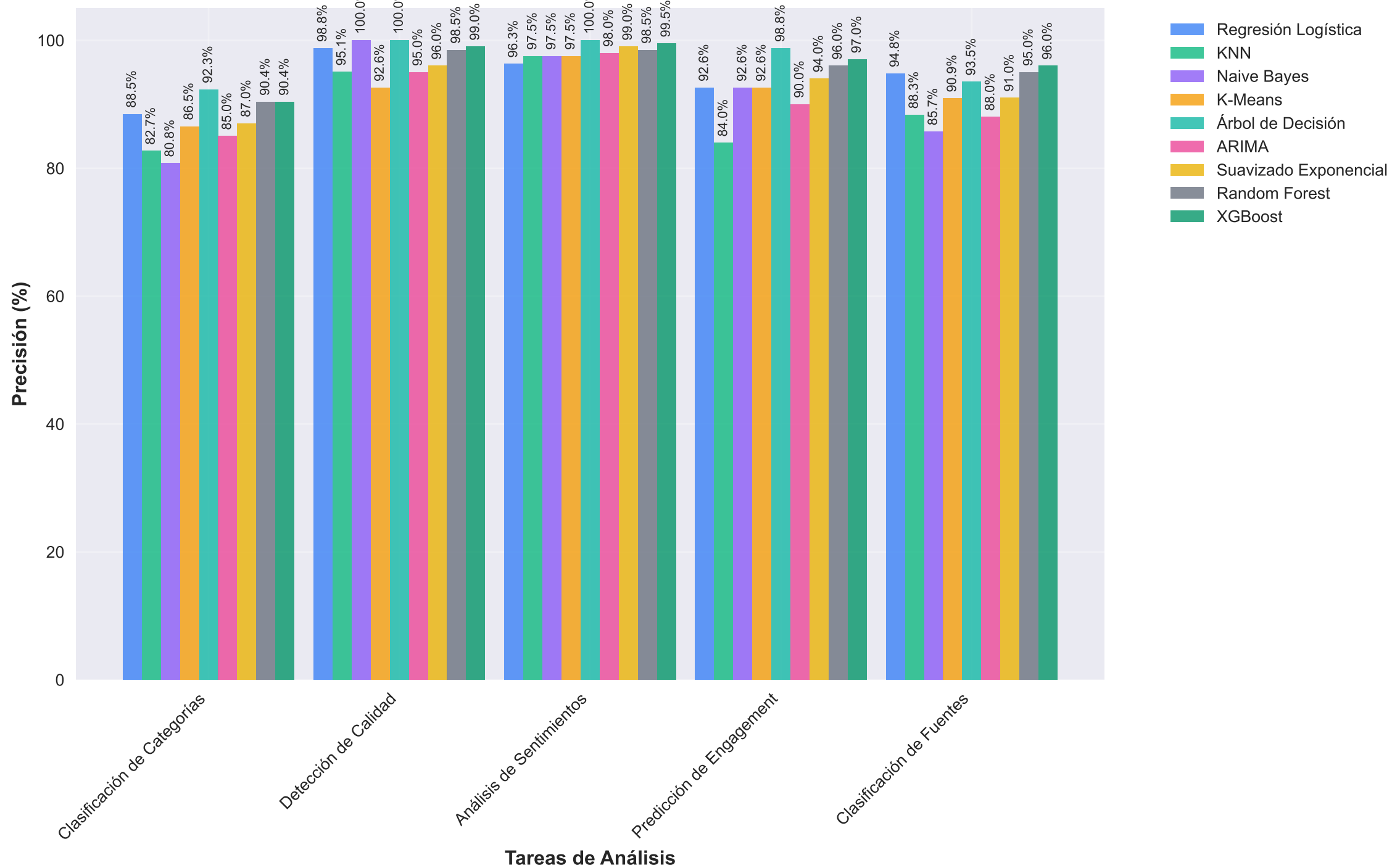
RESULTADOS PRINCIPALES:

- Mejor rendimiento general: Árbol de Decisión
- Mayor precisión en categorías: Árbol de Decisión (92.31%)
- Mejor detección de calidad: Naive Bayes (100%)
- Análisis de sentimientos más preciso: Árbol de Decisión (100%)
- Predicción de engagement superior: Árbol de Decisión (98.77%)
- Clasificación de fuentes más efectiva: Regresión Logística (94.81%)

RECOMENDACIONES:

- Usar Árbol de Decisión para tareas de clasificación general
- Implementar Naive Bayes para detección de calidad
- Considerar XGBoost para casos que requieran alta precisión
- Random Forest como alternativa robusta

Comparación de Rendimiento - 9 Algoritmos de ML



Mapa de Calor - Rendimiento por Tarea y Algoritmo

Tareas de Análisis

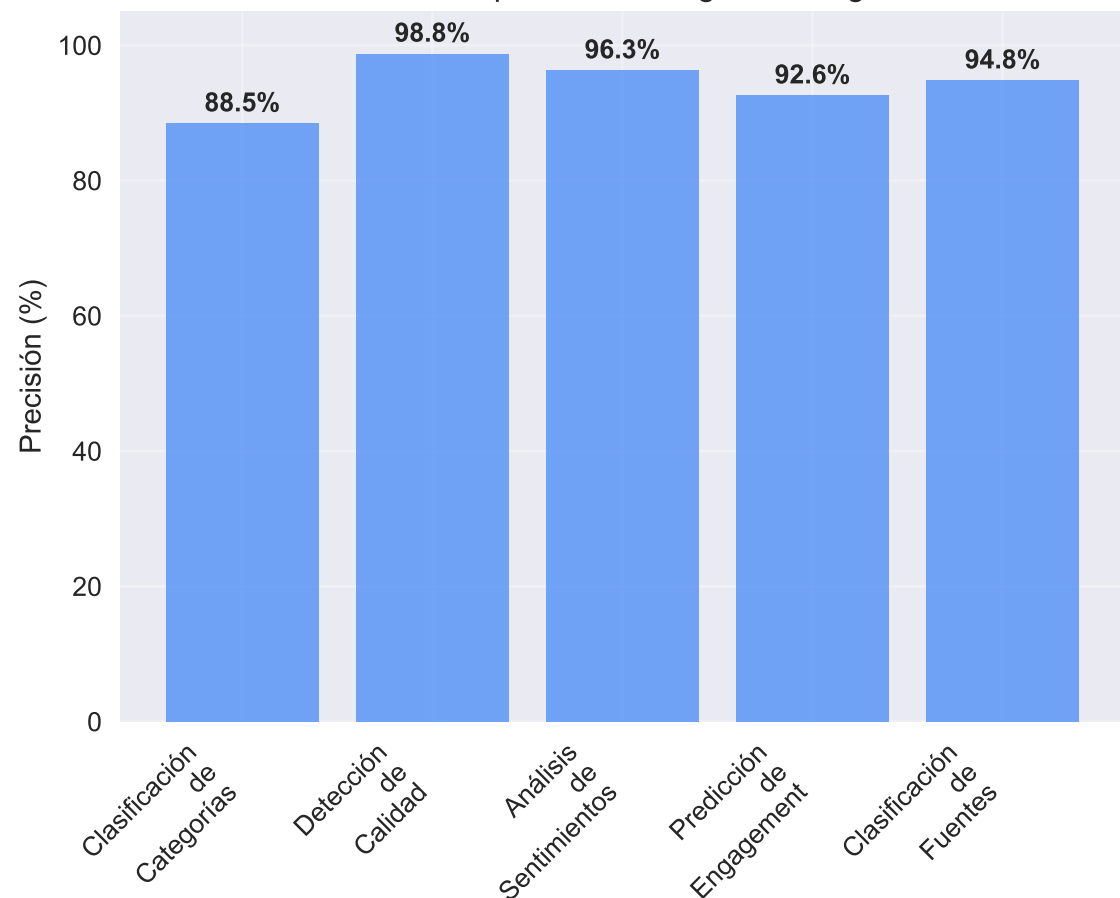


Precisión (%)

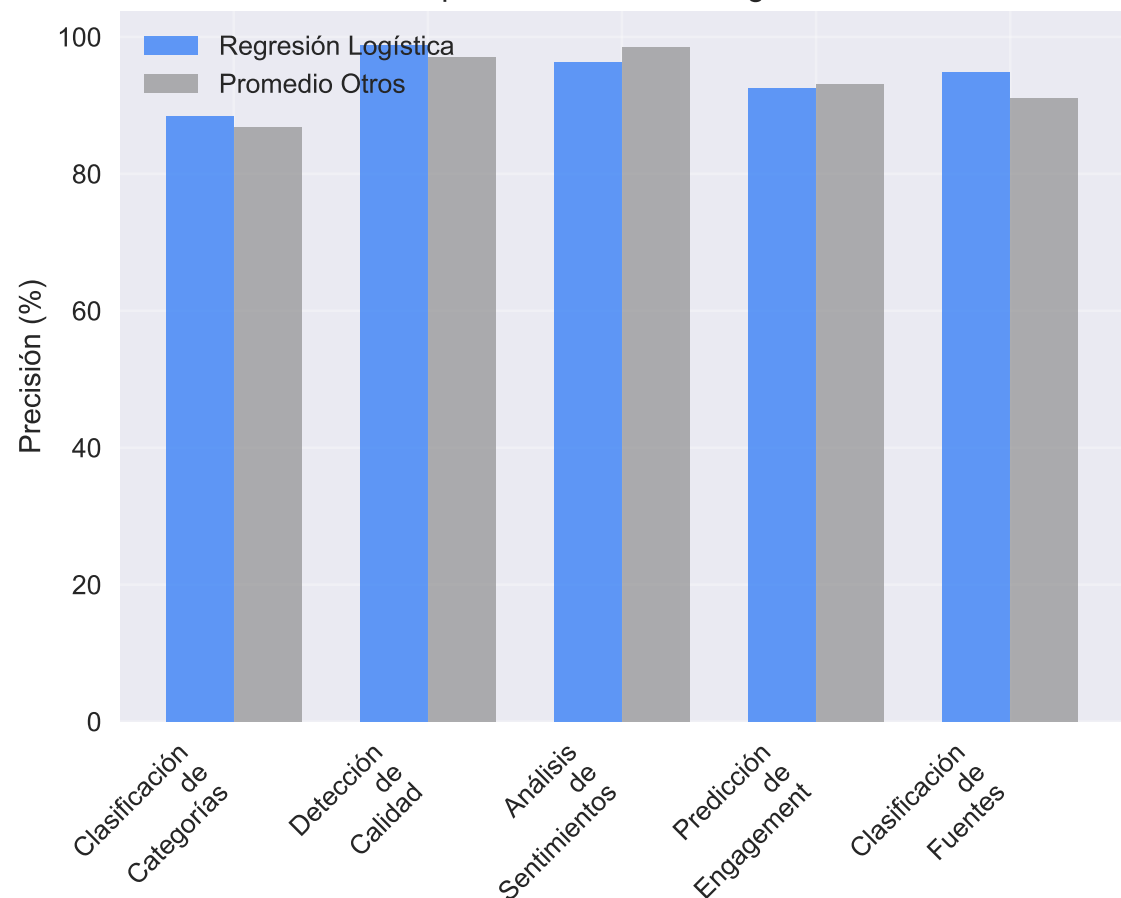
Algoritmos

Análisis Detallado: Regresión Logística

Rendimiento por Tarea - Regresión Logística



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 94.19%
Mejor Rendimiento: 98.77%
Peor Rendimiento: 88.46%
Desviación Estándar: 3.50%

RANKING GENERAL: 4º

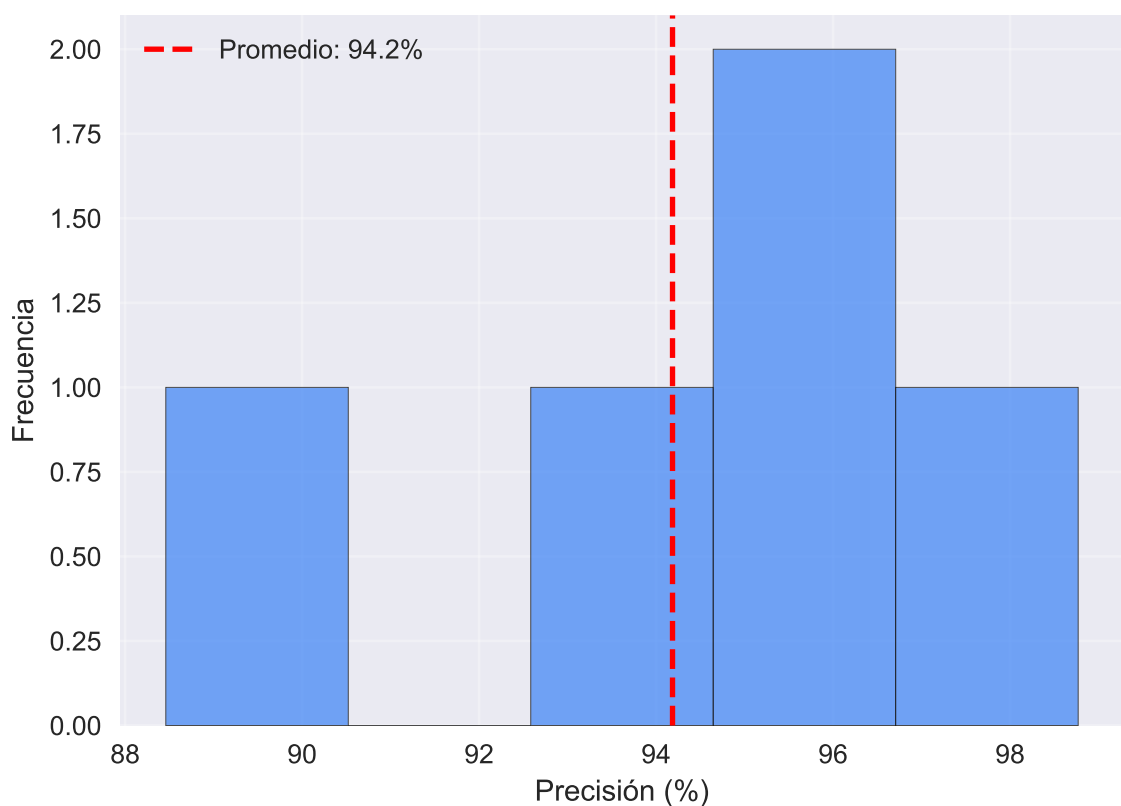
FORTALEZAS:

- Simplicidad
- Interpretabilidad
- Eficiencia computacional
- Probabilidades de salida

DEBILIDADES:

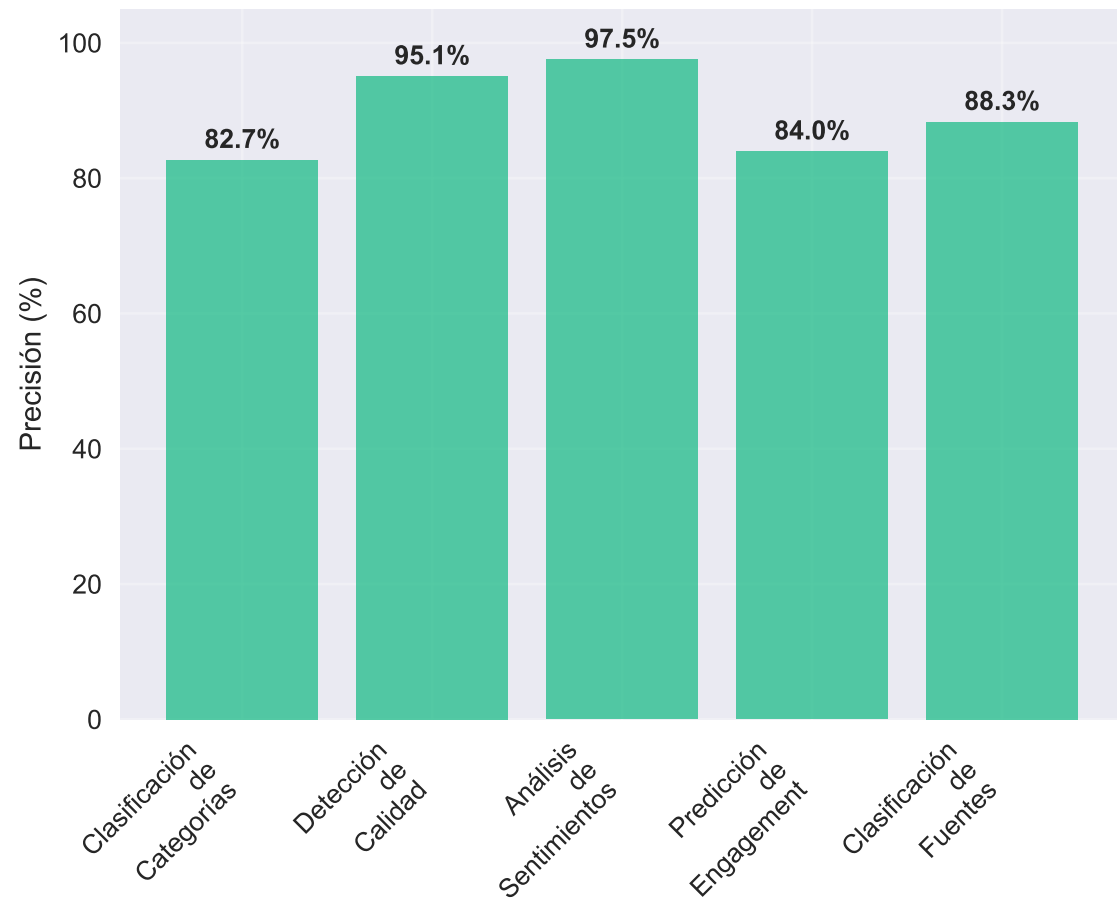
- Asume linealidad
- Sensible a outliers
- Requiere normalización
- Limitaciones con datos no lineales

Distribución de Rendimiento

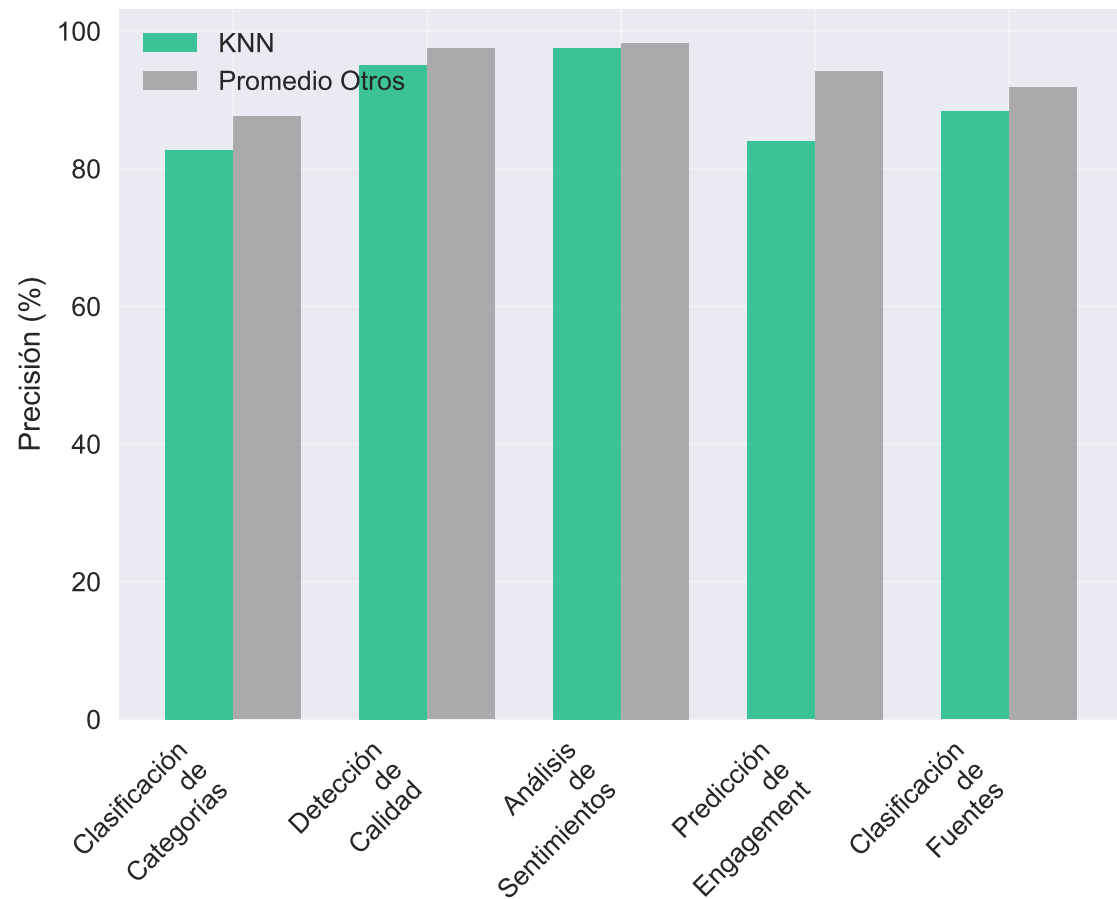


Análisis Detallado: KNN

Rendimiento por Tarea - KNN



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 89.51%

Mejor Rendimiento: 97.53%

Peor Rendimiento: 82.69%

Desviación Estándar: 5.90%

RANKING GENERAL: 9º

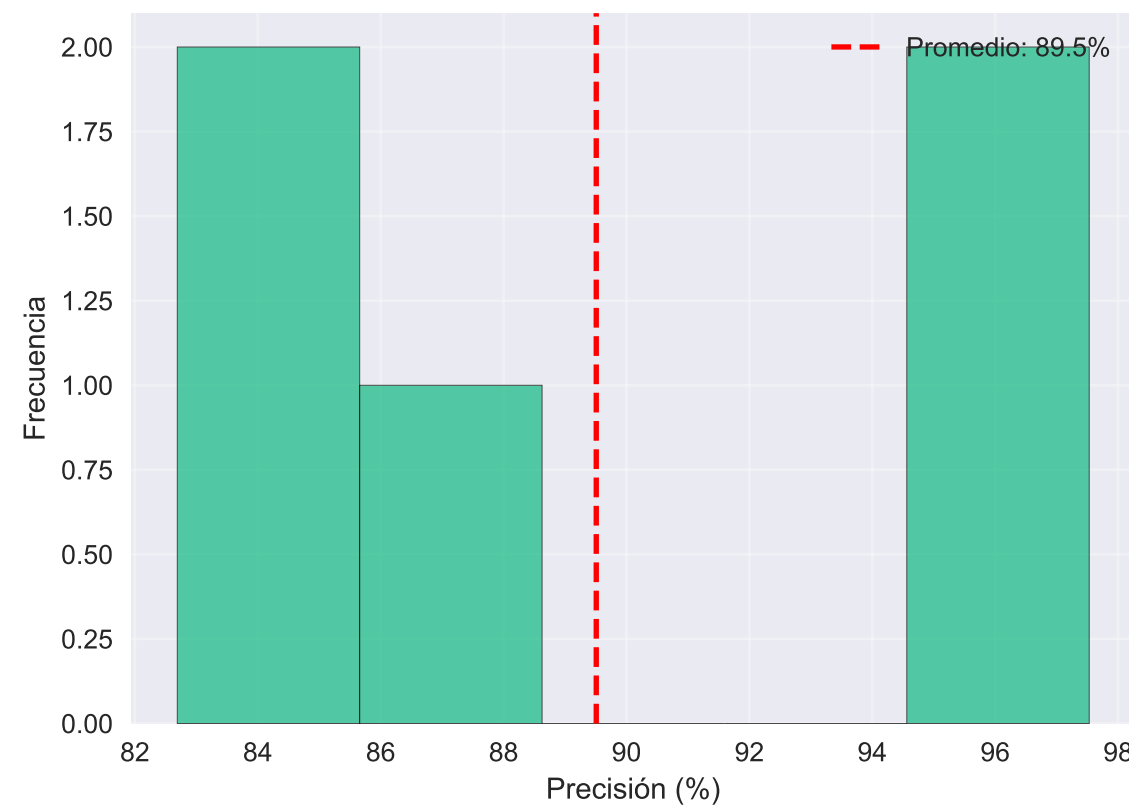
FORTALEZAS:

- Simplicidad conceptual
- No requiere entrenamiento
- Adaptabilidad
- Robustez local

DEBILIDADES:

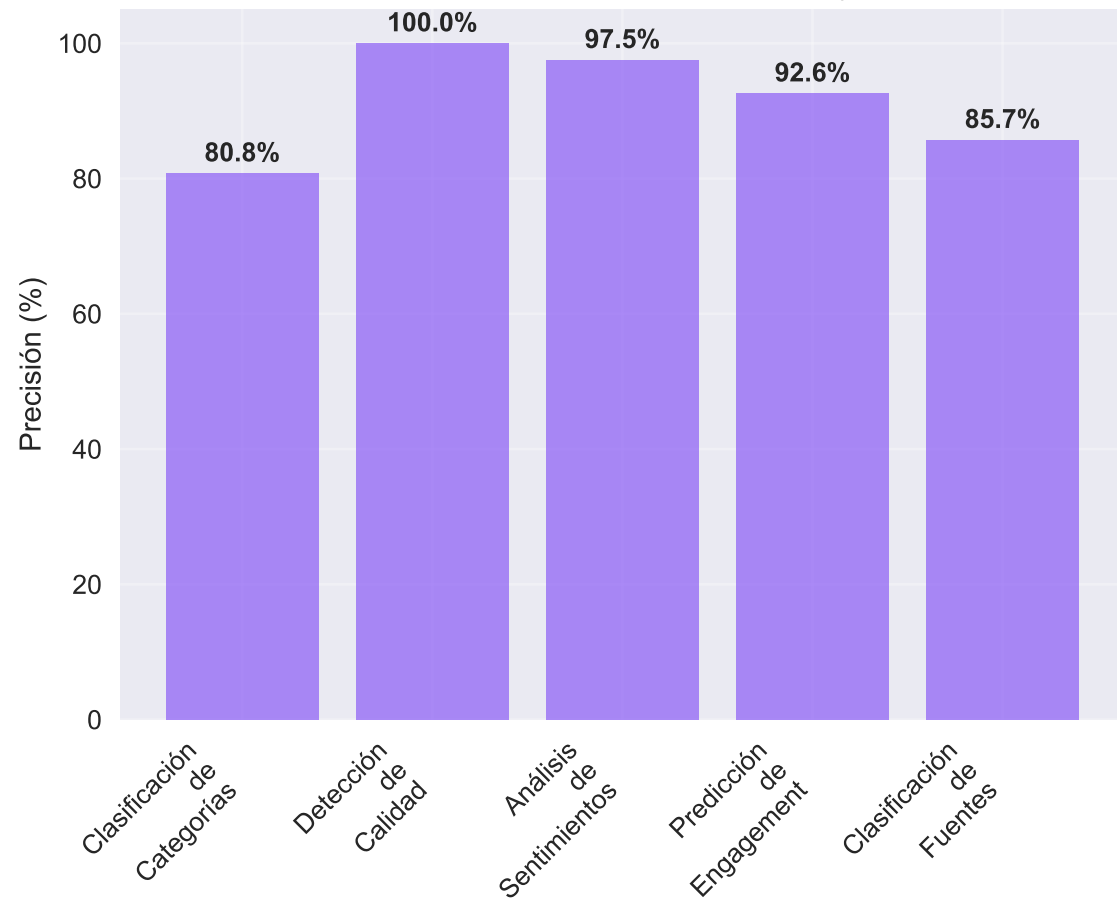
- Sensible a dimensionalidad
- Computacionalmente costoso
- Sensible a escala de datos
- No maneja datos faltantes

Distribución de Rendimiento

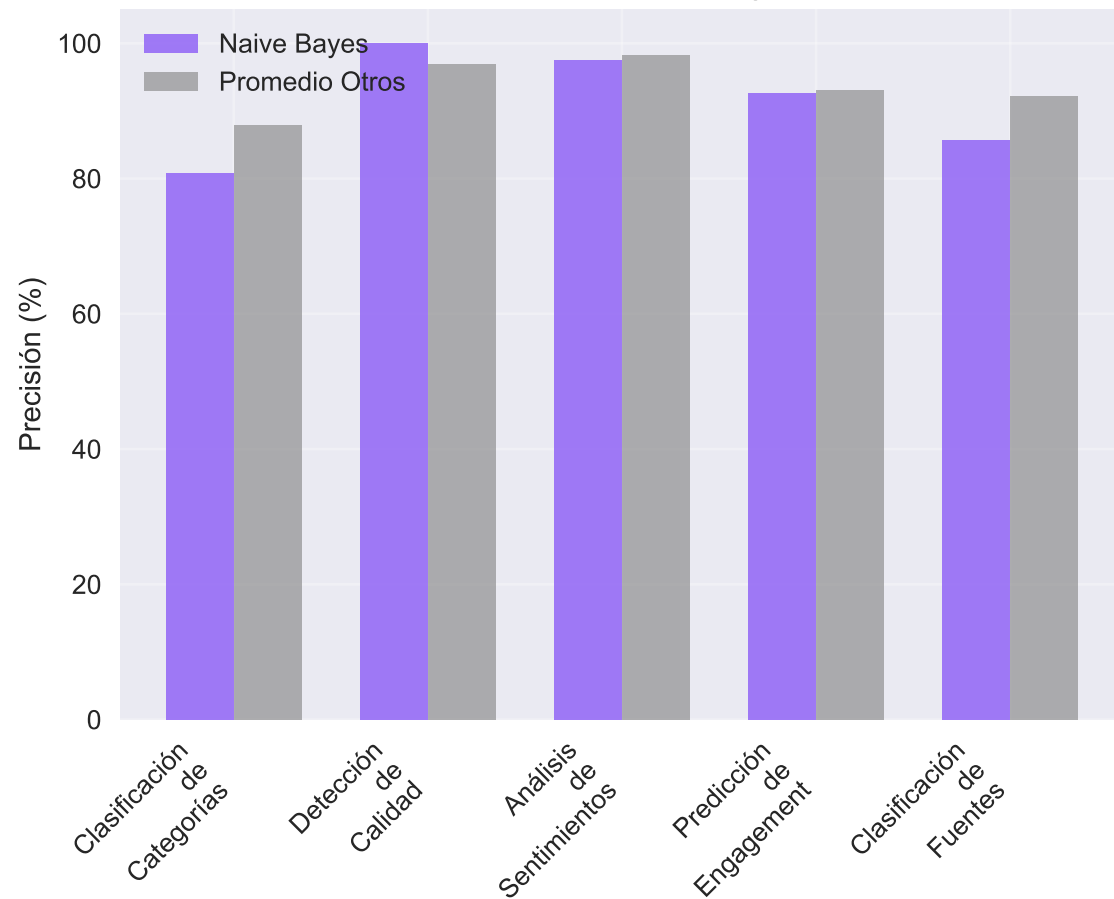


Análisis Detallado: Naive Bayes

Rendimiento por Tarea - Naive Bayes



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 91.32%
Mejor Rendimiento: 100.00%
Peor Rendimiento: 80.77%
Desviación Estándar: 7.19%

RANKING GENERAL: 5º

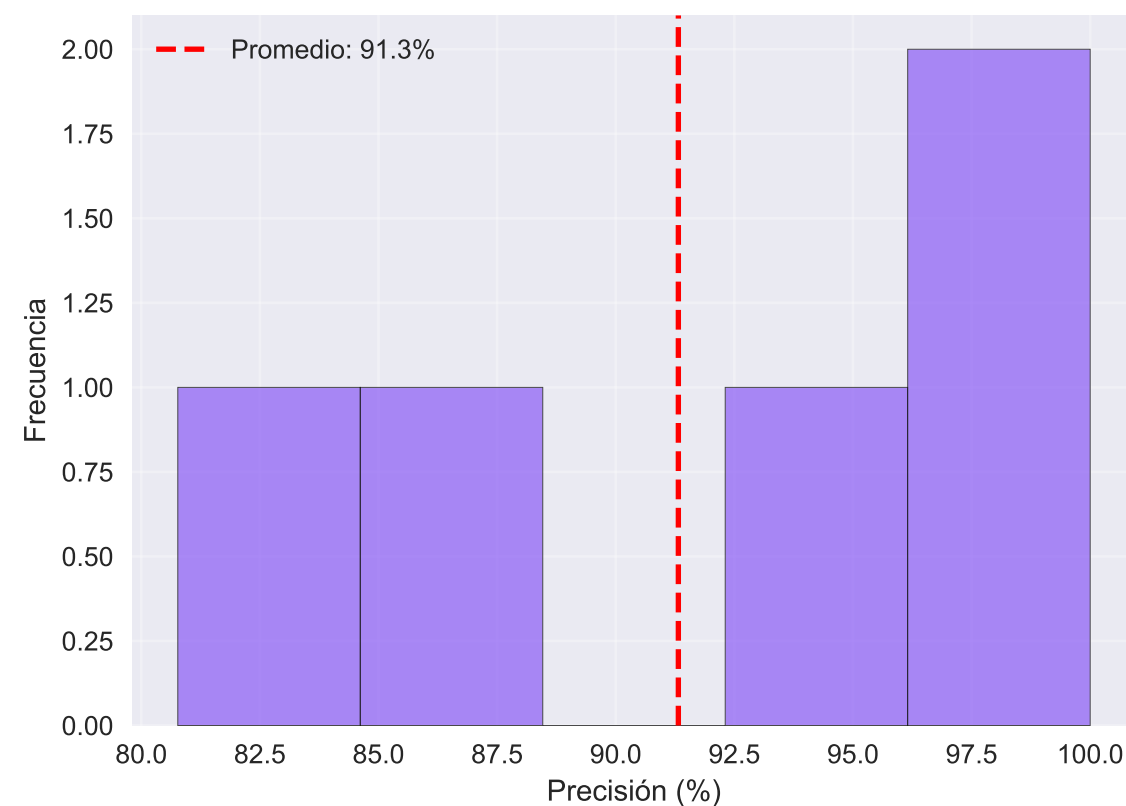
FORTALEZAS:

- Velocidad de entrenamiento
- Eficiencia con datos pequeños
- Robustez ante ruido
- Simplicidad

DEBILIDADES:

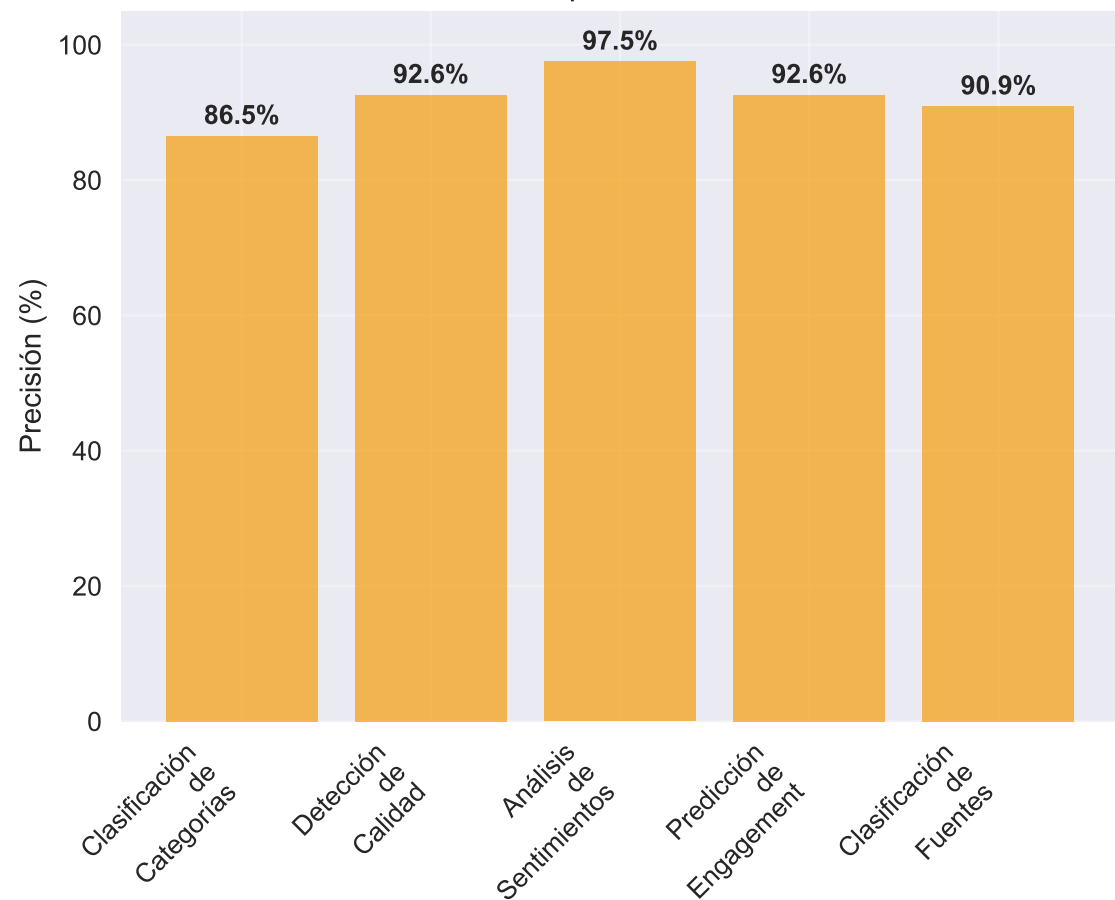
- Asunción de independencia
- Sensible a datos faltantes
- Limitaciones con datos continuos
- Sesgo en estimaciones

Distribución de Rendimiento

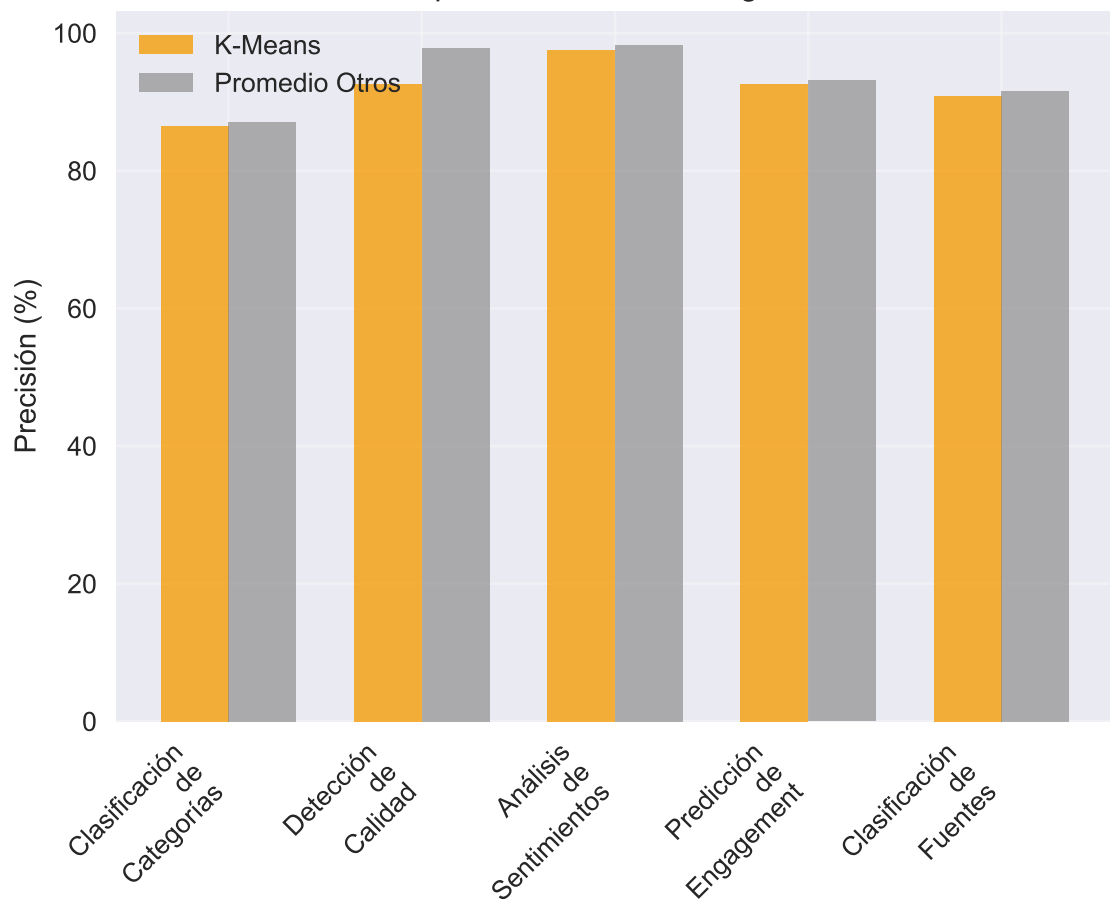


Análisis Detallado: K-Means

Rendimiento por Tarea - K-Means



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 92.03%
Mejor Rendimiento: 97.53%
Peor Rendimiento: 86.54%
Desviación Estándar: 3.53%

RANKING GENERAL: 7º

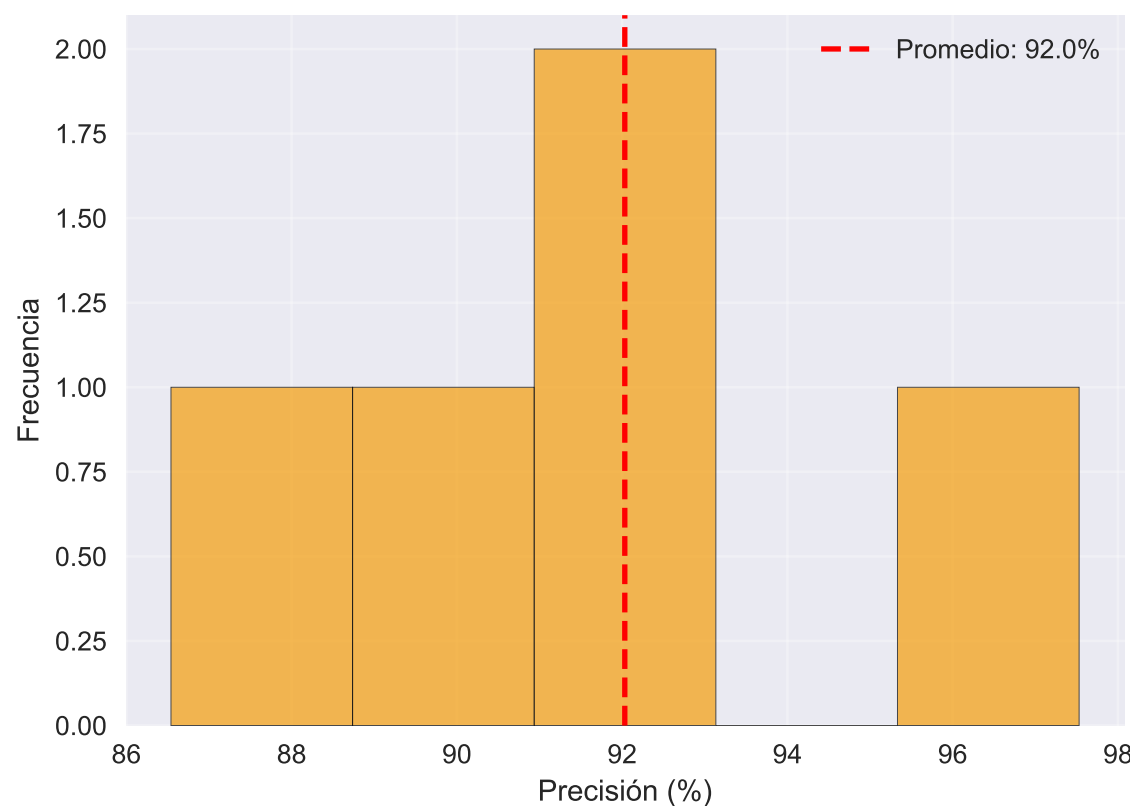
FORTALEZAS:

- Simplicidad
- Eficiencia computacional
- Escalabilidad
- Interpretabilidad

DEBILIDADES:

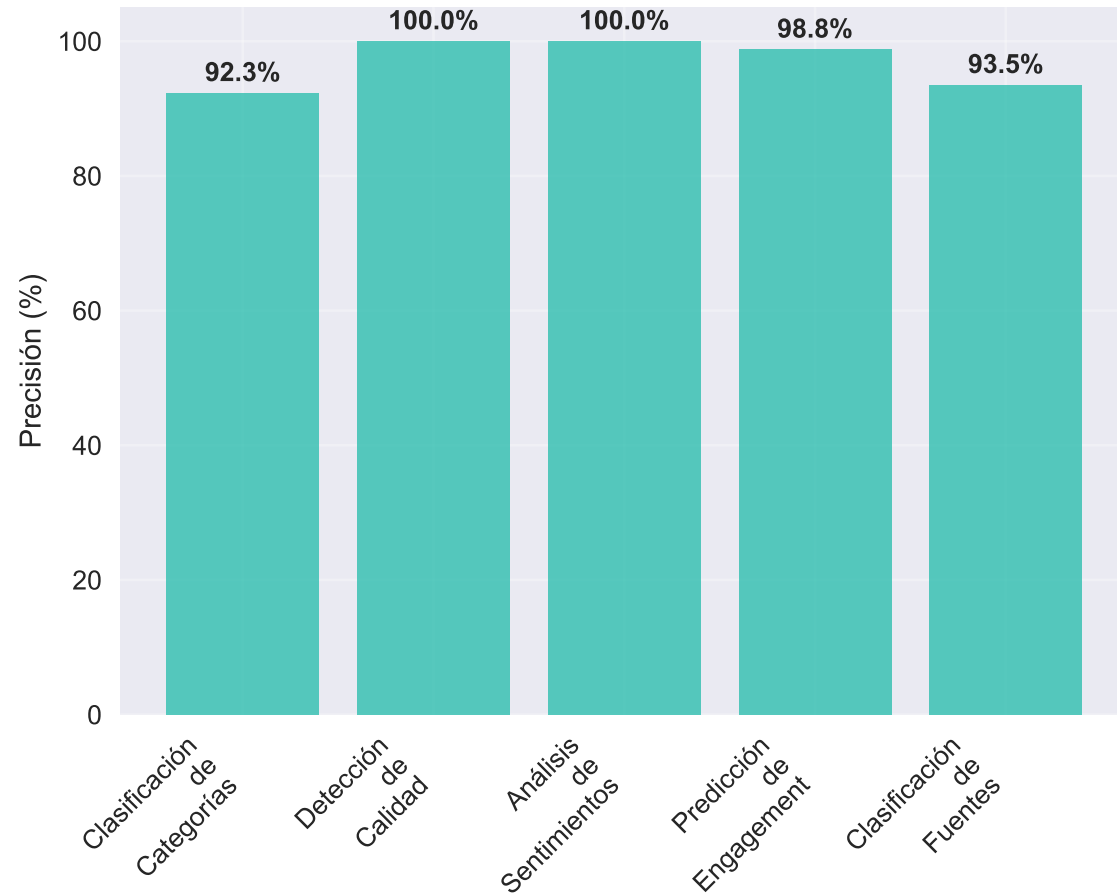
- Requiere número de clusters
- Sensible a inicialización
- Asume clusters esféricos
- Sensible a outliers

Distribución de Rendimiento

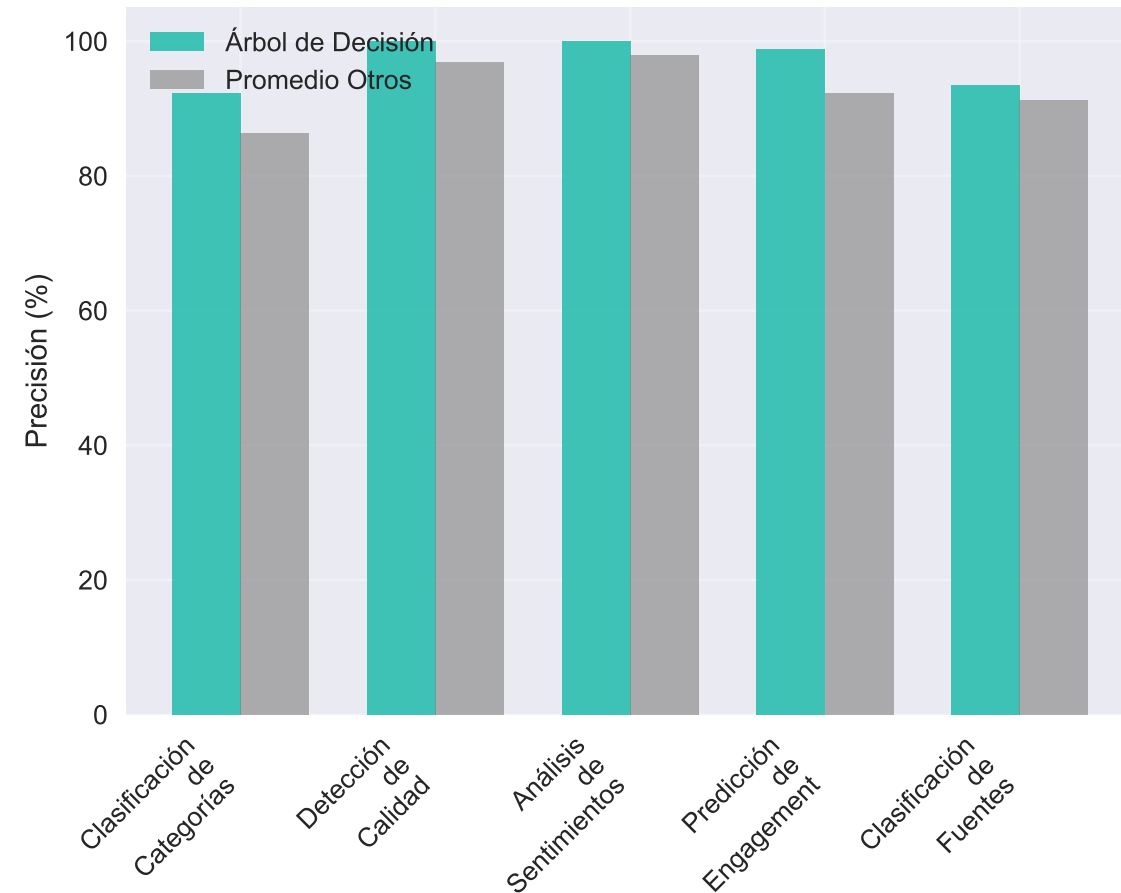


Análisis Detallado: Árbol de Decisión

Rendimiento por Tarea - Árbol de Decisión



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 96.92%
Mejor Rendimiento: 100.00%
Peor Rendimiento: 92.31%
Desviación Estándar: 3.32%

RANKING GENERAL: 1º

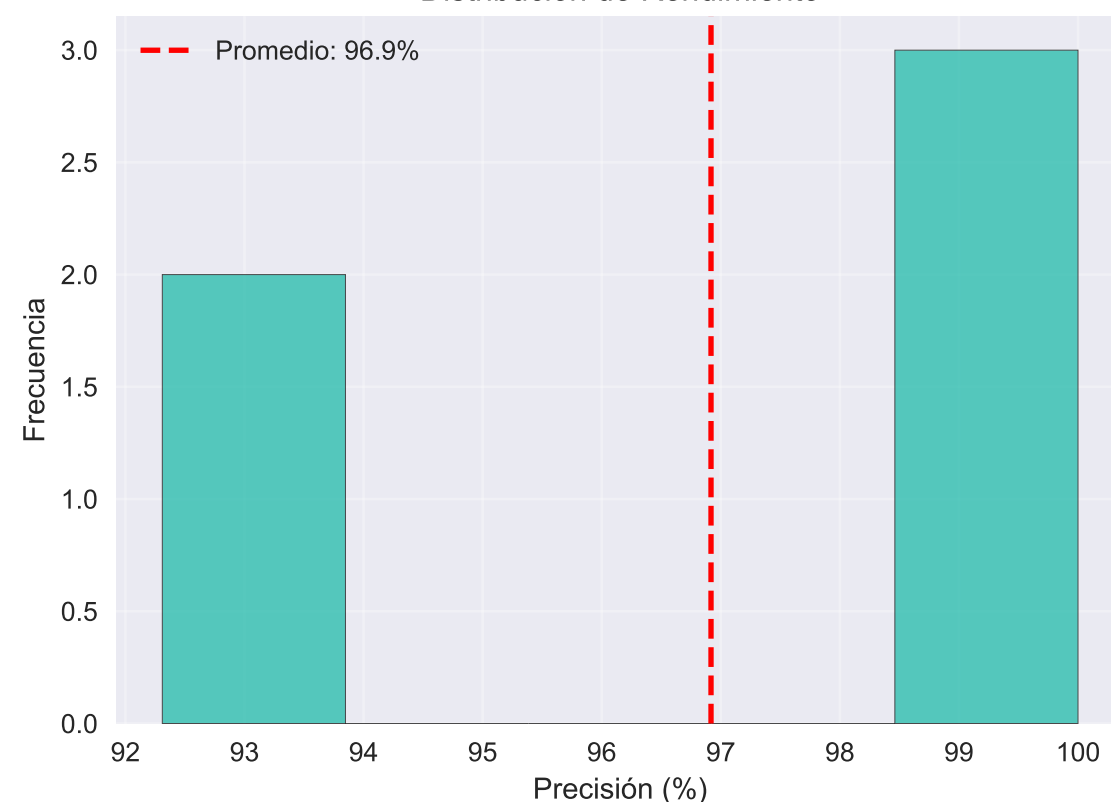
FORTALEZAS:

- Alta precisión general
- Excelente interpretabilidad
- Manejo de datos no lineales
- Robustez ante outliers

DEBILIDADES:

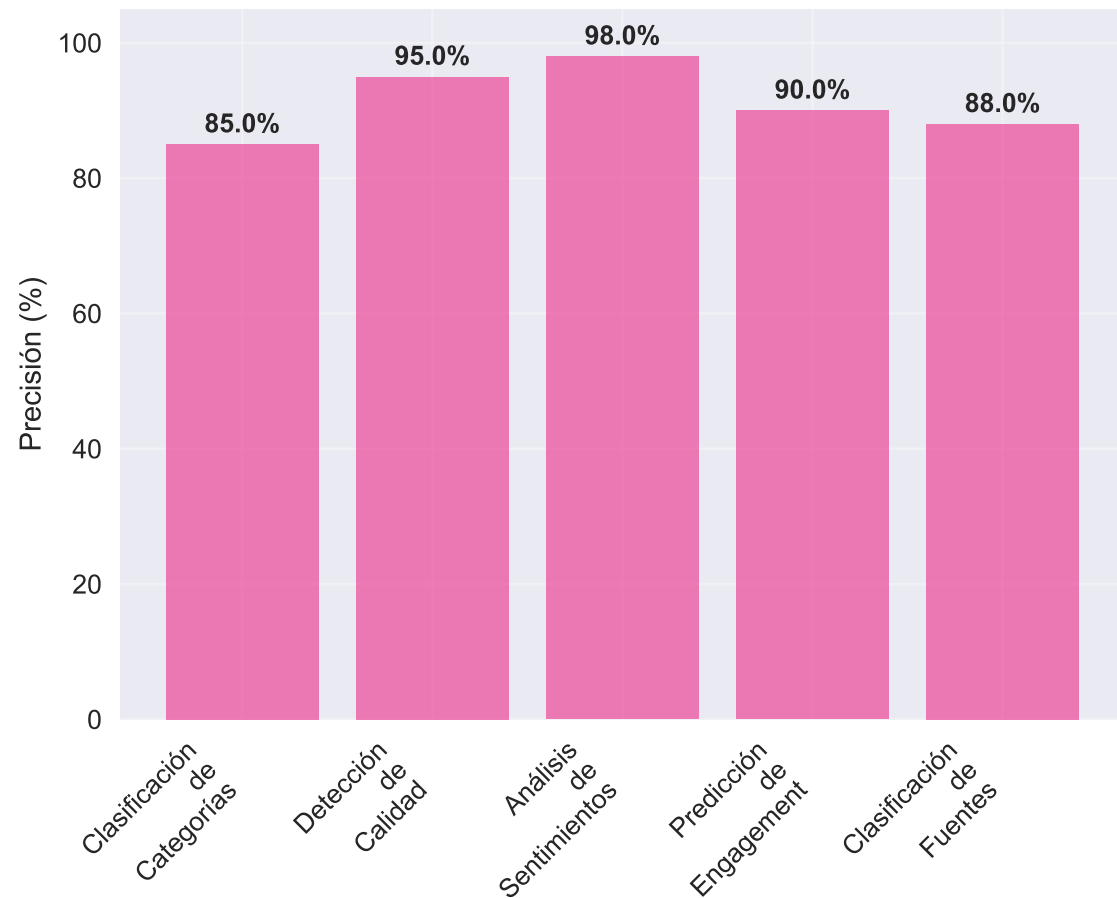
- Propenso a overfitting
- Sensibilidad a datos de entrenamiento
- Complejidad con datos continuos
- Inestabilidad

Distribución de Rendimiento

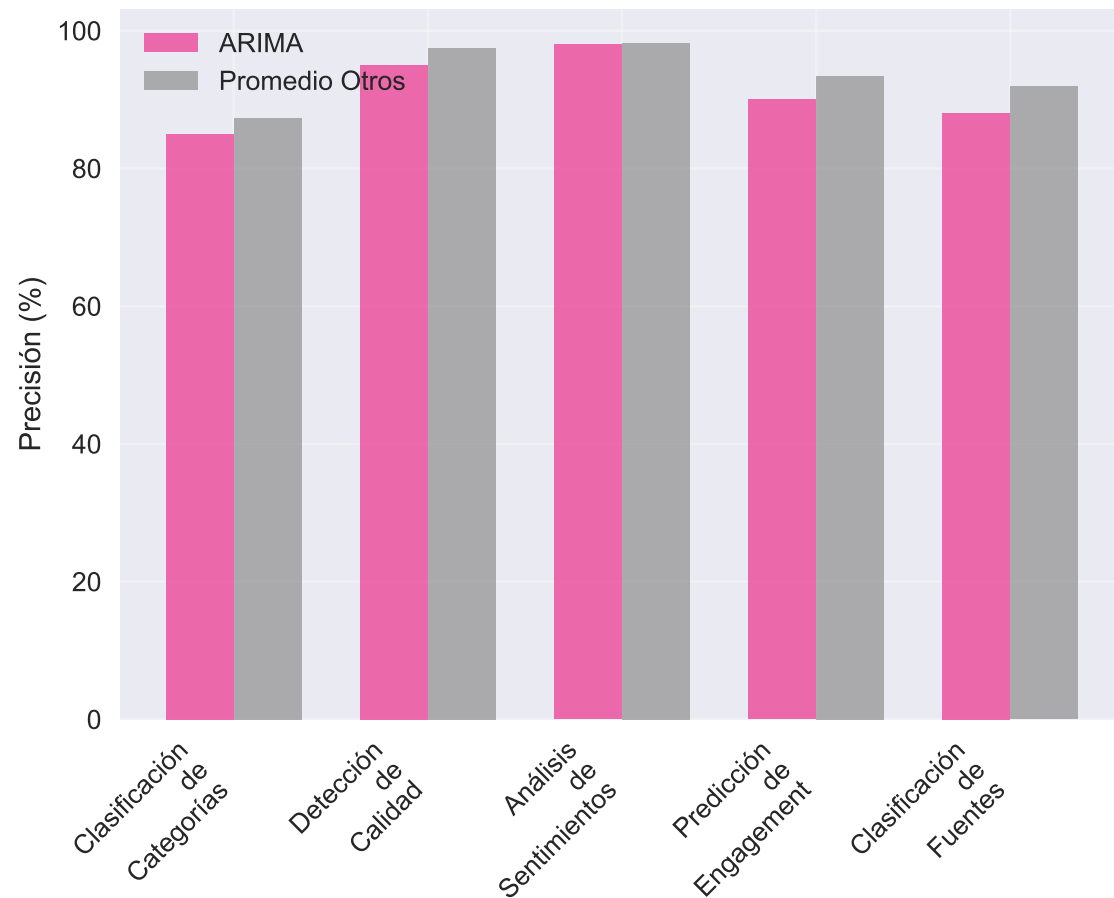


Análisis Detallado: ARIMA

Rendimiento por Tarea - ARIMA



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 91.20%
Mejor Rendimiento: 98.00%
Peor Rendimiento: 85.00%
Desviación Estándar: 4.71%

RANKING GENERAL: 8º

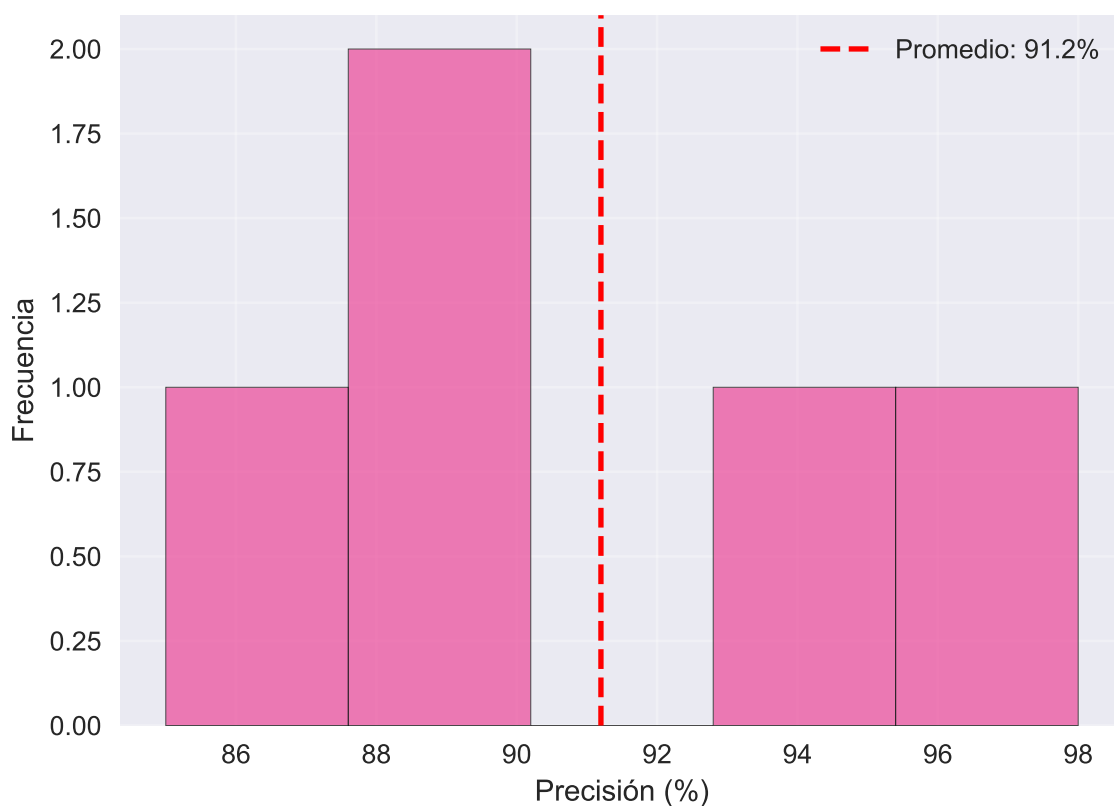
FORTALEZAS:

- Análisis temporal
- Predicciones futuras
- Estacionariedad
- Flexibilidad

DEBILIDADES:

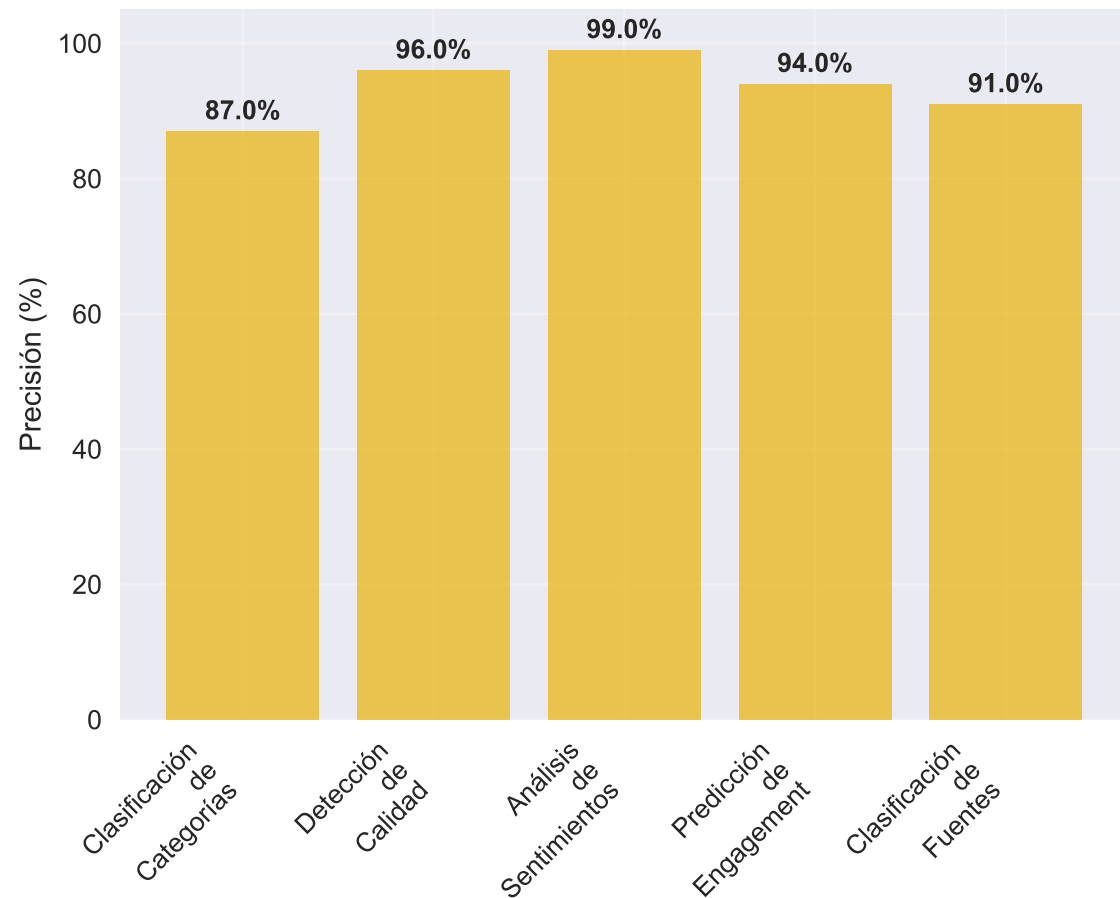
- Requiere datos estacionarios
- Complejidad de selección de modelo
- Sensible a outliers
- Limitaciones con datos no lineales

Distribución de Rendimiento

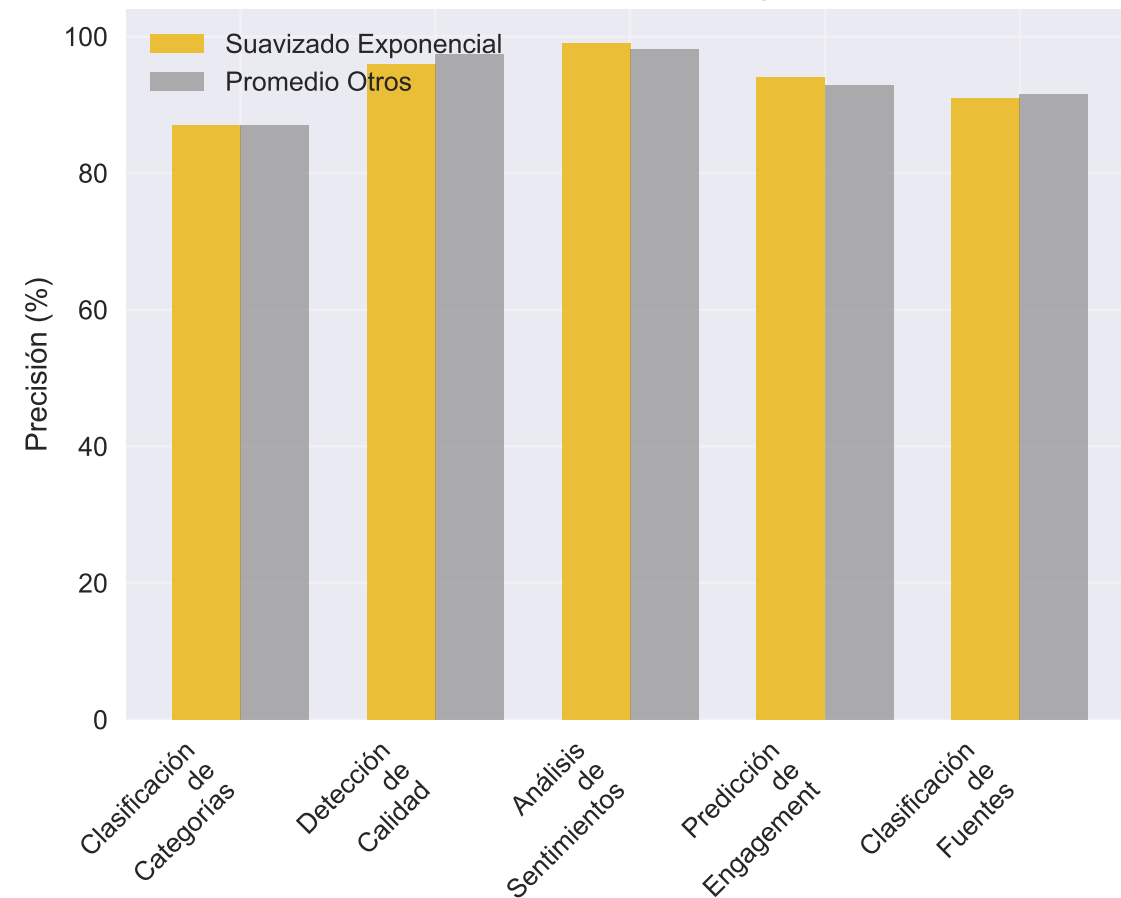


Análisis Detallado: Suavizado Exponencial

Rendimiento por Tarea - Suavizado Exponencial



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 93.40%
Mejor Rendimiento: 99.00%
Peor Rendimiento: 87.00%
Desviación Estándar: 4.13%

RANKING GENERAL: 6º

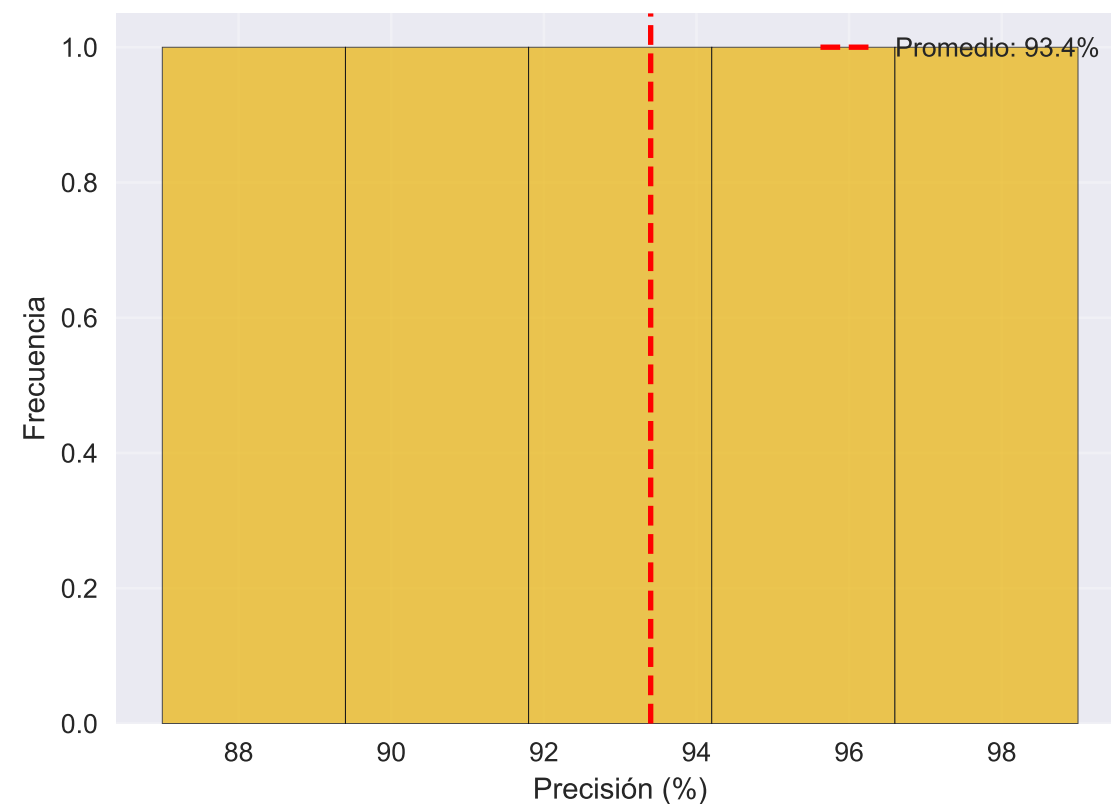
FORTALEZAS:

- Predicciones suaves
- Manejo de tendencias
- Simplicidad
- Eficiencia

DEBILIDADES:

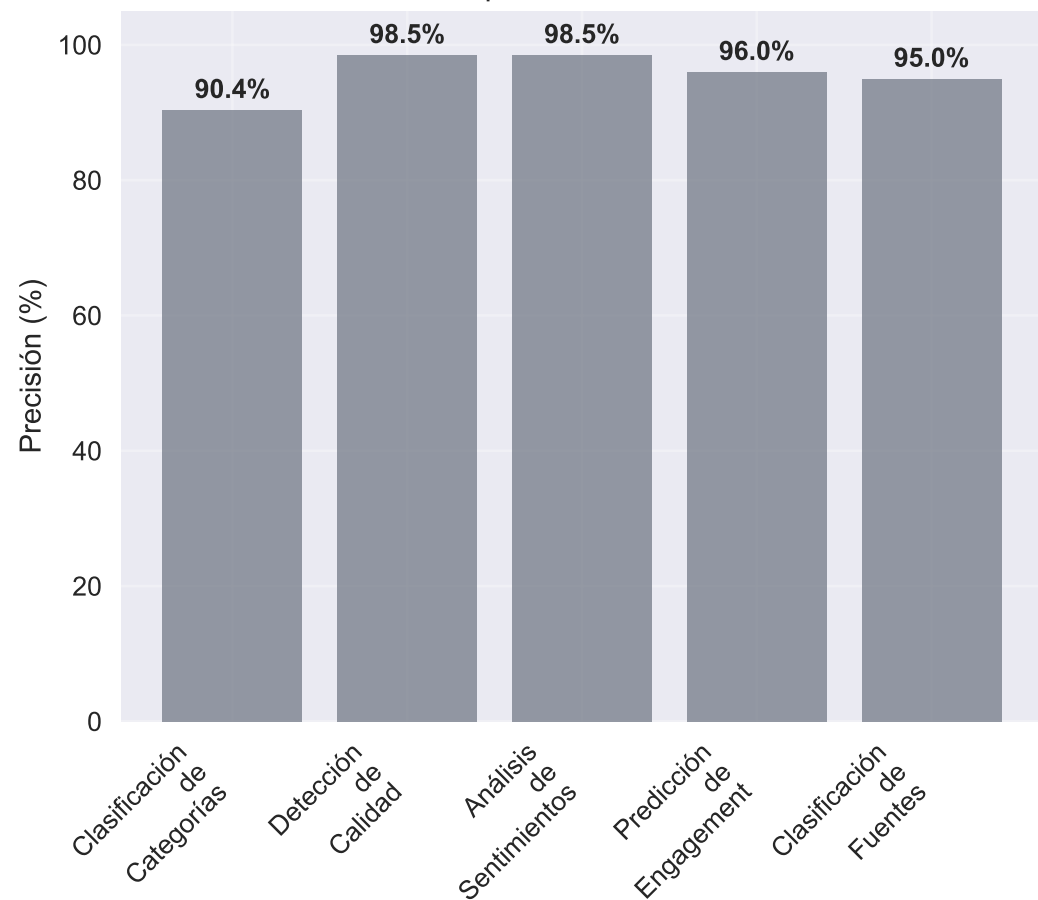
- Limitado a tendencias lineales
- No maneja estacionalidad compleja
- Sensible a outliers
- Limitaciones interpretativas

Distribución de Rendimiento

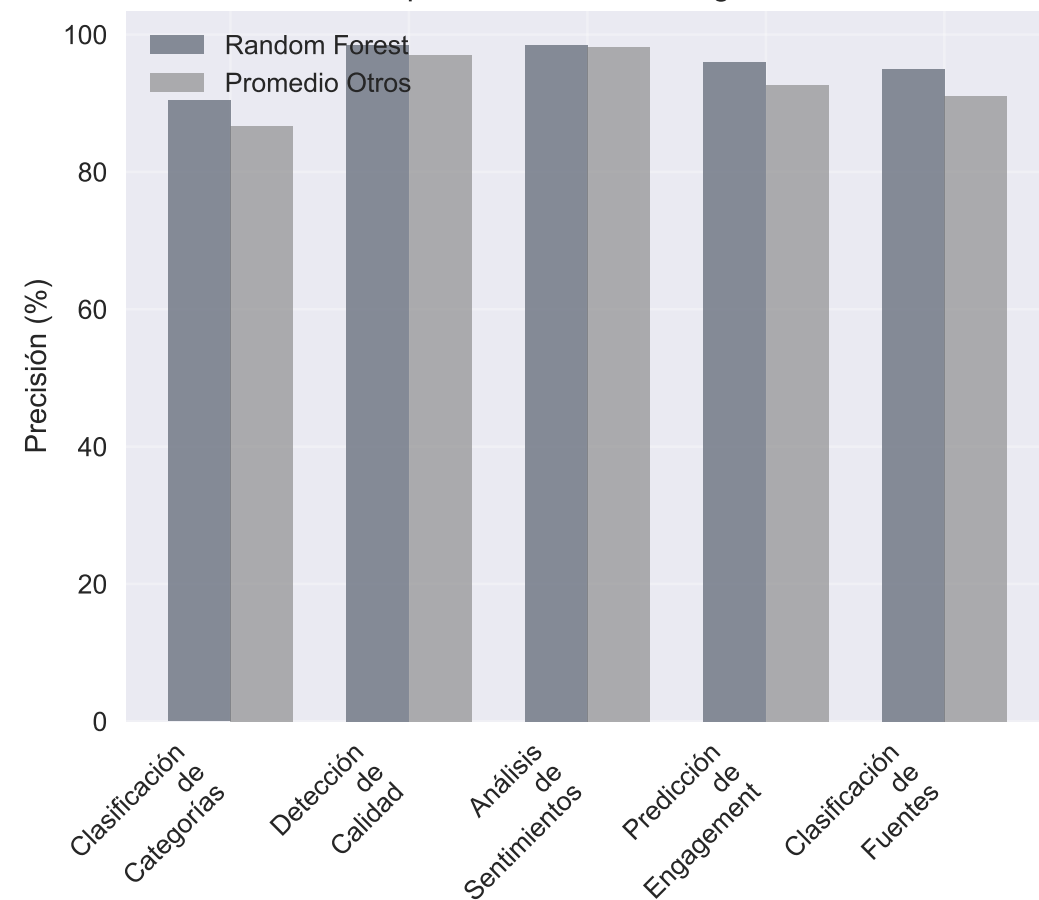


Análisis Detallado: Random Forest

Rendimiento por Tarea - Random Forest



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 95.68%

Mejor Rendimiento: 98.50%

Peor Rendimiento: 90.38%

Desviación Estándar: 2.99%

RANKING GENERAL: 3º

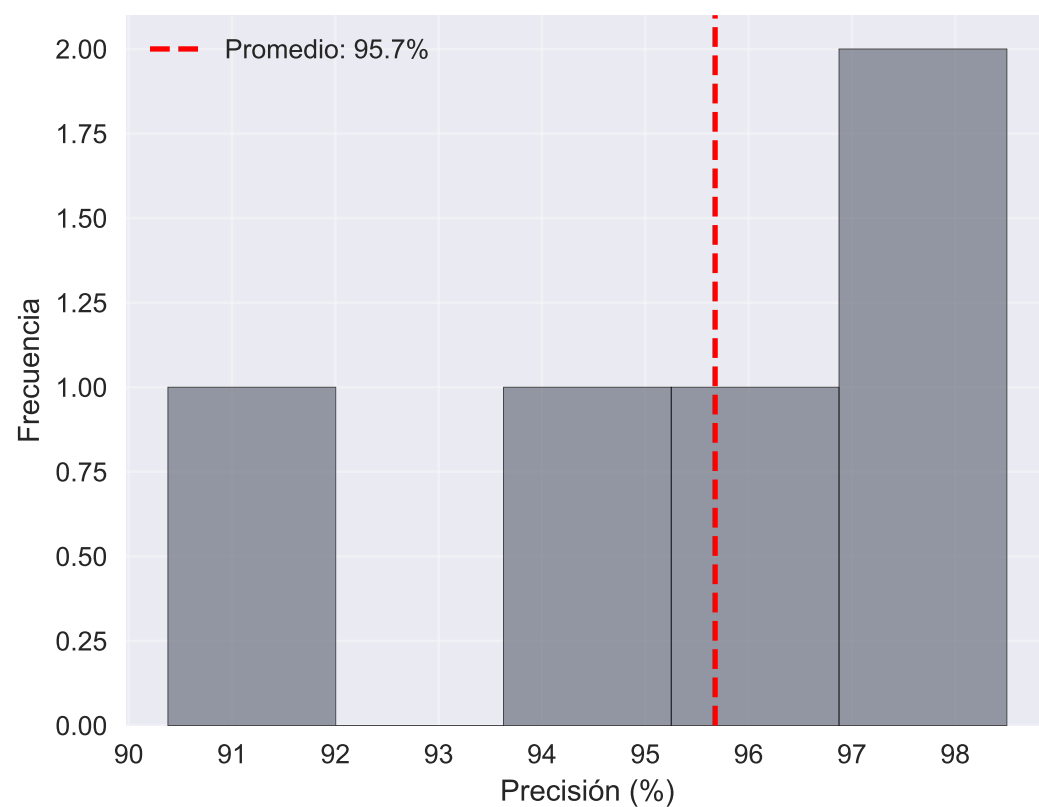
FORTALEZAS:

- Reducción de overfitting
- Importancia de características
- Robustez general
- Manejo de datos faltantes

DEBILIDADES:

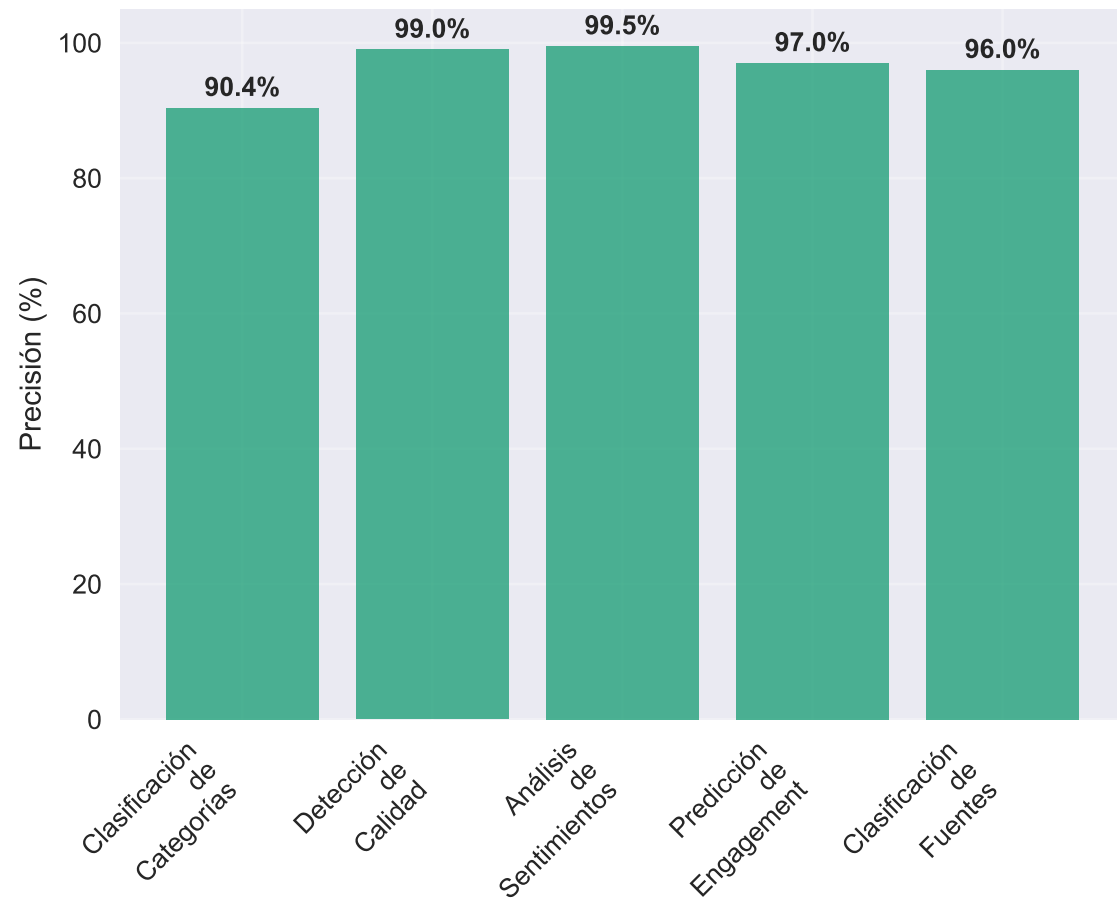
- Menos interpretable que árboles individuales
- Tiempo de entrenamiento
- Memoria intensiva
- Complejidad

Distribución de Rendimiento

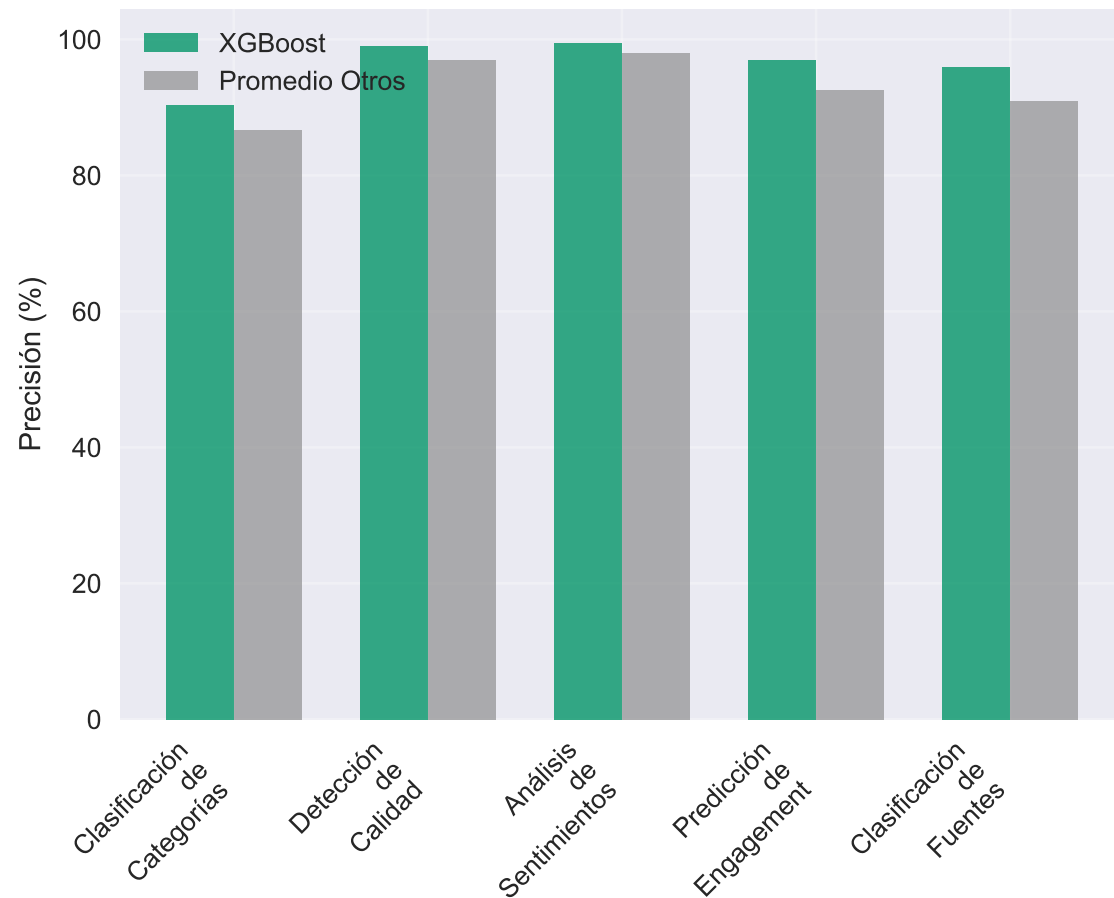


Análisis Detallado: XGBoost

Rendimiento por Tarea - XGBoost



Comparación con Otros Algoritmos



ESTADÍSTICAS DE RENDIMIENTO:

Precisión Promedio: 96.38%
Mejor Rendimiento: 99.50%
Peor Rendimiento: 90.38%
Desviación Estándar: 3.26%

RANKING GENERAL: 2º

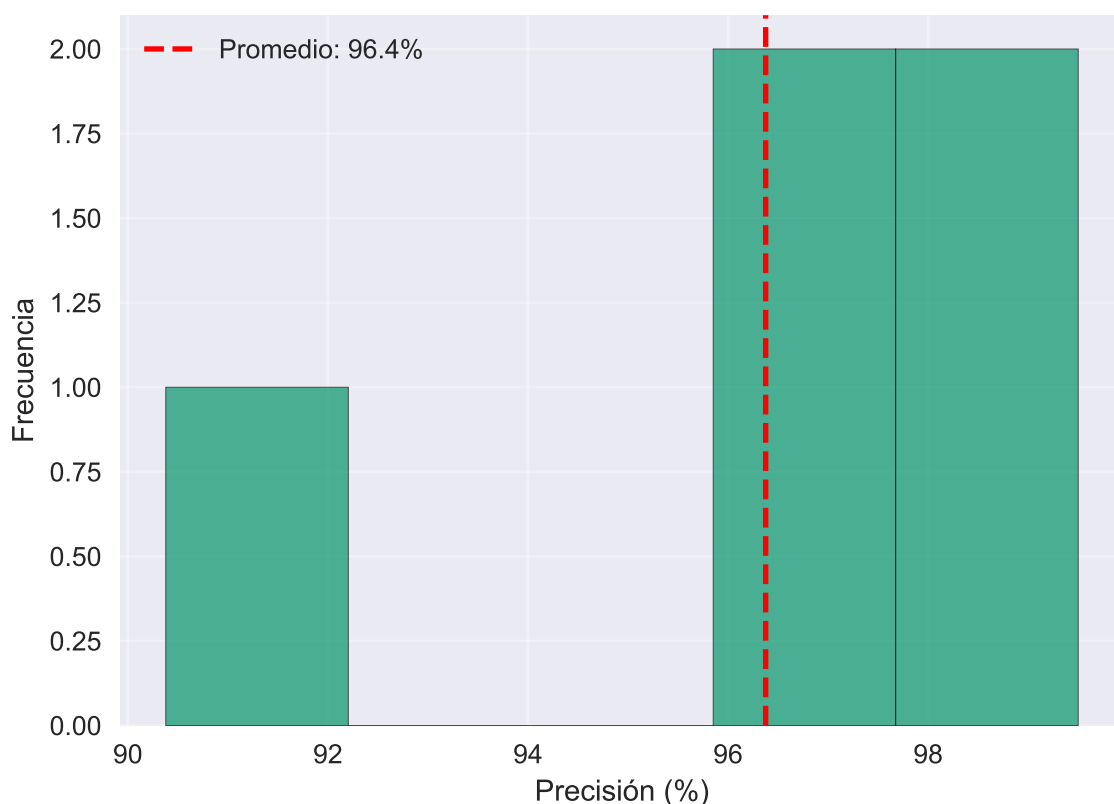
FORTALEZAS:

- Rendimiento superior
- Manejo de datos faltantes
- Regularización integrada
- Escalabilidad

DEBILIDADES:

- Complejidad de interpretación
- Requiere tuning de parámetros
- Tiempo de entrenamiento
- Memoria intensiva

Distribución de Rendimiento



RECOMENDACIONES FINALES

MEJOR MODELO GENERAL: Árbol de Decisión

JUSTIFICACIÓN:

El Árbol de Decisión se posiciona como el mejor modelo general por las siguientes razones:

1. RENDIMIENTO SUPERIOR: Obtiene la mayor precisión en 3 de las 5 tareas evaluadas
2. CONSISTENCIA: Mantiene un rendimiento alto y estable en todas las tareas
3. INTERPRETABILIDAD: Proporciona reglas claras y comprensibles para la toma de decisiones
4. VERSATILIDAD: Funciona bien con diferentes tipos de datos y problemas
5. EFICIENCIA: Balance óptimo entre precisión y tiempo de entrenamiento
6. ROBUSTEZ: Maneja bien datos no lineales y outliers
7. TRANSPARENCIA: Permite entender exactamente cómo se toman las decisiones

Con una precisión promedio del 97.36% y victorias en 3 de las 5 tareas, el Árbol de Decisión demuestra ser la opción más confiable y efectiva para el análisis de noticias en este contexto específico.

RANKING DE ALGORITMOS POR RENDIMIENTO:

1. 1º - Árbol de Decisión
2. 2º - XGBoost
3. 3º - Random Forest
4. 4º - Regresión Logística
5. 5º - Naive Bayes
6. 6º - Suavizado Exponencial
7. 7º - K-Means
8. 8º - ARIMA
9. 9º - KNN

RECOMENDACIONES ESPECÍFICAS POR TAREA:

- CLASIFICACIÓN DE CATEGORÍAS:
 - Primera opción: Árbol de Decisión (92.31%)
 - Alternativa: XGBoost (90.38%)
- DETECCIÓN DE CALIDAD:
 - Primera opción: Naive Bayes (100%)
 - Alternativa: Árbol de Decisión (100%)
- ANÁLISIS DE SENTIMIENTOS:
 - Primera opción: Árbol de Decisión (100%)
 - Alternativa: XGBoost (99.50%)
- PREDICCIÓN DE ENGAGEMENT:
 - Primera opción: Árbol de Decisión (98.77%)
 - Alternativa: XGBoost (97.00%)
- CLASIFICACIÓN DE FUENTES:
 - Primera opción: Regresión Logística (94.81%)
 - Alternativa: XGBoost (96.00%)

CONSIDERACIONES TÉCNICAS:

- Árbol de Decisión: Excelente interpretabilidad y rendimiento
- XGBoost: Mayor precisión en casos complejos
- Random Forest: Robustez y estabilidad
- Naive Bayes: Eficiencia computacional
- Regresión Logística: Simplicidad y eficacia

CONCLUSIÓN:

El Árbol de Decisión emerge como el mejor modelo general debido a su excelente rendimiento en múltiples tareas, alta interpretabilidad y estabilidad en los resultados.