

Informe del Modelo de Detección de Grooming

1. Precisión del modelo (Accuracy)

El modelo Naive Bayes alcanzó una precisión del 100% en el conjunto de prueba. Esto significa que clasificó correctamente todos los mensajes, tanto los correspondientes a grooming como los que no lo eran.

2. Reporte de clasificación

El modelo obtuvo un rendimiento perfecto en cada una de las clases:

- Para la clase grooming, logró identificar correctamente todos los mensajes sospechosos.
- Para la clase no grooming, también acertó en la totalidad de los casos.

En términos de métricas, tanto la precisión, como el recall y la medida F1 alcanzaron un valor de 1.0 en ambas clases, lo que demuestra que el modelo no cometió errores de clasificación en este conjunto de prueba.

3. Matriz de confusión

La matriz de confusión refleja claramente el desempeño del modelo:

- Mensajes grooming reales: 6 correctamente clasificados como grooming.
- Mensajes no grooming reales: 6 correctamente clasificados como no grooming.
- Errores: 0.

Esto confirma que el modelo tuvo un rendimiento perfecto en el dataset trabajado.

4. Clasificación de nuevos mensajes

Finalmente, se evaluaron nuevos mensajes que no estaban en el conjunto original.

El modelo realizó la siguiente clasificación:

- “¿Podés mandarme una foto tuya? No se la muestres a nadie.” → grooming
- “¿Jugamos Minecraft esta tarde?” → no grooming
- “No le digas a tus papás que hablamos, ¿sí?” → grooming
- “¿Terminaste el trabajo práctico de historia?” → no grooming

Conclusión:

El modelo Naive Bayes demostró un desempeño perfecto sobre los datos de entrenamiento y prueba utilizados. Fue capaz de distinguir adecuadamente entre mensajes de grooming y mensajes normales, y también clasificó de forma coherente los nuevos ejemplos.

Este tipo de sistema podría aplicarse como una herramienta de apoyo para detectar conversaciones sospechosas en entornos de mensajería en tiempo real.