

Detector de spam

Definición del problema

Objetivo

Desarrollar un sistema que analice correos electrónicos y determine automáticamente si son spam (correo no deseado) o correos legítimos, aplicando un conjunto de reglas predefinidas

Reglas de Clasificación

El sistema clasifica un correo como spam si cumple con al menos una de estas cinco reglas:

- Regla 1: Si el correo electrónico contiene la palabra clave "Whatsapp", entonces es Spam.
- Regla 2: Si el correo electrónico contiene un enlace a un sitio web conocido por distribuir malware, entonces es Spam.
- Regla 3: Si el correo electrónico tiene un remitente desconocido, entonces es Spam.
- Regla 4: Si el correo electrónico tiene un asunto que es demasiado bueno para ser verdad, entonces es Spam.
- Regla 5: Si el correo electrónico está mal escrito o tiene errores gramaticales, entonces es Spam.

El sistema recibe como entrada un archivo de texto (.txt) que contiene múltiples correos electrónicos en un formato específico:

- Cada correo comienza con una línea "De: [dirección de correo]"
- Seguido por una línea "Asunto: [texto del asunto]"
- Luego el cuerpo del mensaje
- Los correos están separados por una línea que contiene "---"