TEMA:

Análisis de Sistemas Colaborativos. Caso Stack OverFlow

HIPÓTESIS

Clasificación de preguntas y respuestas de Stack OverFlow.

INTRODUCCIÓN

El sistema StackOverFlow (SO) es uno de los sistemas colaborativos de mayor participación específicamente en Tecnologías de Software.

Algunos de los sistemas colaboratiovos QW son: Yahoo Questions, …

El trabajo propuesto consiste en el análisis y clasificación de las preguntas y respuestas posteadas, en base a algunos atributos obtenidos en la base de datos …..

La base de datos contenida en BigQuery de Google está compuesta por algunas tablas:

…

…

…

Los datos contenidos en estas tablas se describen a continuación.

…

….

….

….

Los datos analizados en su mayoría son datos ordinales discretos, intervalos (fechas), categóricos (tipo de post), y categórico nominal que corresponde al texto de preguntas, respuestas y comentarios.

Para hacer una exploración inicial obtuvimos algunos conceptos de la Wiki de SO:

Debido a la gran cantidad de información se seleccionaron los siguientes datos para hacer el análisis de las preguntas y respuestas:

….

….

…

PRE-PROCESAMIENTO DE DATOS

1) post publicados por minuto

3) tiempo mínimo de respuesta aceptada.

2) cantidad de usuarios por tema

3) preguntas y respuestas

4) comentarios

El estudio de minería de datos sobre Sistemas Colaborativos ha otorgado resultados significativos y aún existen casos de estudio como: Análisis de sentimientos sobre preguntas y respuestas para estudios de personalidad, predicción del tiempo que una pregunta tardará en ser respondida, predicción de respuestas aceptadas, entre otros.

…. se realizó una publicación denominada Mining ….. en este estudio se propone 4 enfoques para analizar si una repuesta será elegida como respuesta aceptada. Estos enfoques son: ……