**Author (s): Veronika Simoncicova1, Pavol Tanuska1, Hans-Christian Heidecke1, and Stefan Rydzi2**

**Títle of paper: Analysis of the Quality of the Painting Process Using Preprocessing Techniques of Text Mining**

**Journal: Advances in Intelligent Systems and Computing, 764.**

**Pag-: 30-38.**

**Year: 2019.**

**Problema que el autor desea resolver**

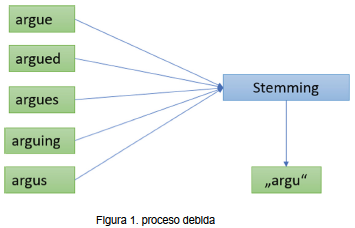
El problema para resolver abarca a la minería de texto, que es una nueva e interesante zona de investigación de la ciencia que trata de resolver la crisis de la sobrecarga de información mediante la combinación de técnicas de minería de datos, aprendizaje automático, procesamiento del lenguaje natural, recuperación de la información y gestión del conocimiento. Del mismo modo, la detección de enlace - un enfoque de rápida evolución para el análisis de texto que comparte y se basa en muchos de los elementos clave de la minería de texto – También proporciona nuevas herramientas para que las personas aprovechen mejor sus recursos de datos textuales crecientes. Las principales tareas de detección de enlace son para extraer, descubrir y unir escasa evidencia formar grandes cantidades de fuentes de datos, para representar y evaluar la significación de la evidencia relacionada, y para aprender patrones para guiar a la extracción, el descubrimiento y la vinculación de las entidades.

**Descripción del aporte del autor**

1. **Técnicas de minería de texto Preprocesing**

Los formatos de números, formatos de fecha y las palabras más comunes improbables para ayudar a la minería de texto, como preposiciones, artículos, y pro-sustantivos pueden ser eliminadas por las técnicas de pre-procesamiento.

* **Partiendo:** se utiliza este método para identificar la raíz / tallo de una palabra

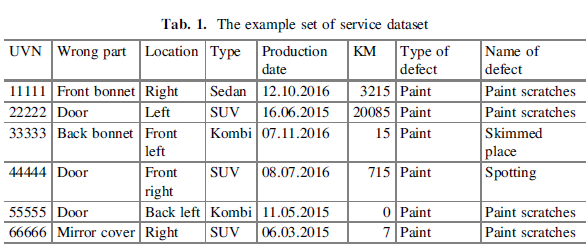


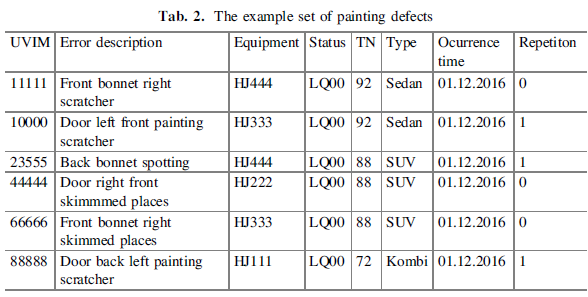
* **La eliminación / Detener la palabra:** Este proceso también reduce los datos de texto y mejora el rendimiento del sistema. Todos los documentos de texto que contienen estas palabras que no son necesarios para las aplicaciones de minería de texto, se desocupan.

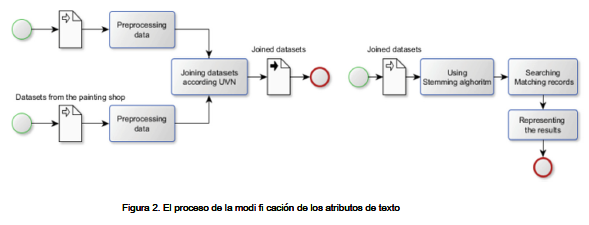
1. **Análisis los datos de la fabricación.**

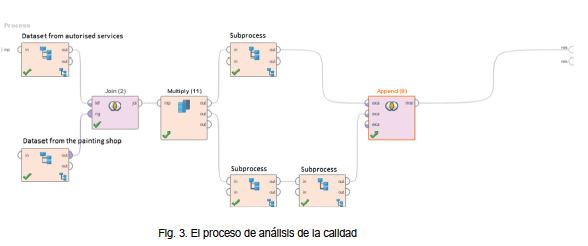
El objetivo es unir dos conjuntos de datos de las diferentes fuentes de datos y para adquirir información relativa al porcentaje de defectos transferidos desde la fabricación a los clientes finales.

* **RapidMiner.** Fuente abierta plataforma de análisis predictivos más utilizado para el análisis de extracción, preprocesamiento y visualización y análisis predictivo y modelos estadísticos, evaluación y despliegue
* **Conjunto de datos de Servicios autorizadas.** El sistema con fi datos de campo se utiliza para analizar la calidad de los productos producidos. En ese sistema, datos de servicio se almacena, en donde los diversos atributos detallen los defectos reclamados reportados que han sido reportados a los clientes en los centros de servicio autorizados. El objeto o producto único número (UVN), el nombre de defecto, la parte equivocada del producto, los defectos ubicaciones exactas en el producto, el tipo de producto, fecha de producción, tipo de defecto, el kilometraje y otros atributos que describen con precisión defectos en la reivindicada.



* **Conjunto de datos desde la Tienda de Pintura.** La base de datos del sistema de producción se utiliza para escribir los defectos de pintura en la producción, pero sólo una parte de los datos se pueden obtener a partir de esta base de datos.
* **El proceso de análisis conjuntos de datos.**





**Conclusiones**

Después de pre y procesar los datos obtenidos en ambas muestras, el objetivo principal era hacer eficiente y mejorar el análisis d la calidad de la pintura del producto. Sin embargo, no es posible comparar con una cantidad tan grande de datos de la tienda de pintura.

