

Estos dos últimos trabajos permiten tener una visión diferente de cómo se puede abordar el problema de clasificación de requisitos, debido a que no es impositivo que deba ser mediante el uso de algoritmos de aprendizaje supervisado, sino que, al hacer uso de algoritmos de aprendizaje no supervisado, se puede ampliar el espectro de aplicación al omitir la necesidad de etiquetar la base de datos de entrenamiento.

Vogelsang y Borg (2019) realizan un primer acercamiento en explorar cómo los científicos de datos realizan el proceso de ingeniería de requisitos en el diseño e implementación de sistemas basados en aprendizaje automático. Si bien, los autores indagan si en el proceso de creación de sistemas basados en aprendizaje automático se utilizan las mismas metodologías de ingeniería de requisitos para la definición de especificaciones del software del cliente, los autores encuentran que hay una gran variedad de nuevos conceptos como explicabilidad, libertad de discriminar decisiones de diseño, e incluso ámbitos legales, que actualmente no se utilizan para un proceso de ingeniería de requisitos en otro tipo de sistemas. Esta documentación es relevante debido a que, si bien la presente propuesta se basará en cómo clasificar los requisitos de desarrollo de software, se puede extender para futuras investigaciones la aplicación de dichas metodologías en el propio diseño de un sistema basado en aprendizaje automático ya que, como lo explican los autores, tiene incluso más parámetros que abordar.

Si bien, se ha observado cómo se tienen diferentes perspectivas para la clasificación de los requisitos de software, de manera general, se opta por utilizar modelos de aprendizaje supervisado para la resolución del mismo (Binkhonain y Zhao 2019). Esto, en conjunto con las correctas técnicas pre-procesamiento de texto, puede obtener una mejora significativa con respecto a los tiempos de clasificación de los requisitos de software (Iqbal, Elahidoost y Lucio 2018). No obstante, Gramajo, Ballejos y Ale (2019) resaltan que, si bien hay un amplio estudio en este campo, existe una escasez en conjuntos de datos de requisitos de software y de estudios enfocados a la clasificación de requisitos de software en el idioma español, lo cual propone una brecha de investigación que se intentará suplir con la presente propuesta.

Así, con un conjunto de datos de entrada pre-procesado, se intentará abordar diferentes técnicas de clasificación automática mediante el uso de modelos de lenguaje previamente entrenados Hey y col. (2020) y (Lee y Hsiang 2020).

## 6. Metodología

A continuación, se muestran los pasos para dar cumplimiento a cada uno de los objetivos específicos definidos en la Sección 3.

### 6.1. Generación de instancias de entrenamiento y validación

1. Se iniciará con el levantamiento de la información para generar el conjunto de datos con que se pretende entrenar los diferentes modelos. Para ello, se traducirá el conjunto de datos por (Dalpiaz y col. 2019) y se integrará con un conjunto de datos con información propia y etiquetada manualmente de requisitos de software utilizados en el ámbito educativo.
2. Una vez generado el conjunto de datos, se procederá con su división en instancias de prueba, entrenamiento y validación. Los dos primeros, se utilizarán para entrenar los diferentes modelos de clasificación y para afinar los parámetros de cada uno de los modelos de clasificación. El tercero y último, se utilizará para probar la mejor versión de cada modelo de clasificación.

### 6.2. Generación de modelos de clasificación

1. Teniendo el conjunto de datos, se procederá a seleccionar las técnicas de pre-procesamiento de texto que se aplicarán sobre el conjunto de datos.
2. Una vez definidas las técnicas de pre-procesamiento de texto, se seleccionará el conjunto de técnicas de aprendizaje automático con las que se generarán los modelos de clasificación de los requisitos de software.
3. Teniendo las técnicas de clasificación, se procederá a definir sus métricas de desempeño con las que se evaluará y seleccionará la mejor.
4. Luego, se seleccionarán las técnicas de búsqueda de parámetros para la clasificación que permitan mejorar las métricas de desempeño.
5. Finalmente, se generarán y entrenarán los modelos aplicando las técnicas definidas anteriormente. Así, se seleccionará el modelo con la combinación de parámetros que obtuvo los mejores resultados en las métricas de desempeño.

### 6.3. Validación de modelos de clasificación

Se evaluarán los modelos con la base de datos de validación, utilizando las mismas métricas de desempeño definidas anteriormente.

## 7. Cronograma

En la Figura 1 se muestran las actividades referentes al cumplimiento del primer objetivo específico.

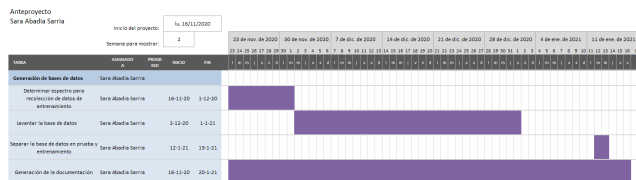


Figura 1: Actividades para el cumplimiento del primer objetivo. **Fuente:** Elaboración propia.

En la Figura 2 se muestran las actividades referentes al cumplimiento del segundo objetivo específico.

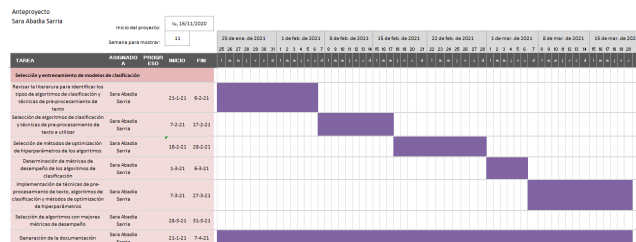


Figura 2: Actividades para el cumplimiento del segundo objetivo. **Fuente:** Elaboración propia.

En la Figura 3 se muestran las actividades referentes al cumplimiento del tercer objetivo específico.

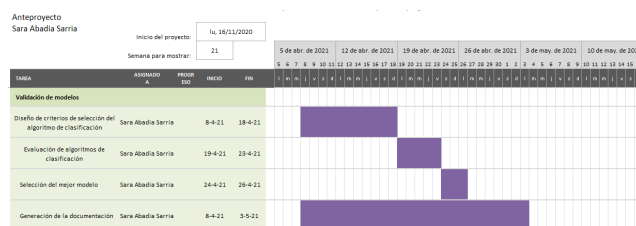


Figura 3: Actividades para el cumplimiento del tercer objetivo. **Fuente:** Elaboración propia.

En la Figura 4 se muestran las actividades referentes al la realización del documento del trabajo de grado.

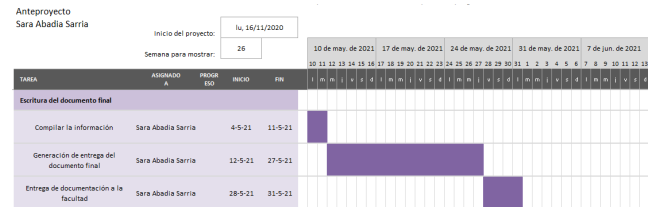


Figura 4: Actividades para la construcción del documento. **Fuente:** Elaboración propia.

## 8. Conclusiones

Como se pudo observar en las secciones anteriores, el proceso de clasificación de requisitos de software es una actividad fundamental para el éxito de proyectos de software y gracias a su optimización, se pueden mejorar los tiempos de implementación porque ahorra el análisis manual de requisitos a los analistas, arquitectos y desarrolladores, reduce el esfuerzo cognitivo requerido para dar sentido a las descripciones de los requisitos, aporta a que el proceso de desarrollo de software sea más eficiente, por consiguiente, aporta a disminuir el riesgo de extensión de los presupuestos en los proyectos de software. Retomando la Sección 5, recientemente han aumentado los estudios enfocados a la automatización de la clasificación de requisitos de software, lo que confirma la importancia de la temática y la existencia de un área con altas opciones de exploración y aportes.

La automatización de la clasificación de requisitos de software se realiza mediante el aprendizaje automático con metodologías de excelente sustento técnico que permiten automatizar el proceso con dos enfoques: el aprendizaje supervisado y el aprendizaje profundo. Son más numerosos los estudios basados en aprendizaje supervisado, por tanto, es el enfoque seleccionado en la propuesta.

Se identifica una oportunidad de futuros trabajos en la automatización de la clasificación de requisitos descritos en idioma español, pues los repositorios consultados en la Sección 5 utilizan en su mayoría, la información en el idioma inglés, de manera que, una brecha de investigación puede ser generar modelos de aprendizaje automático que permitan automatizar la clasificación de requisitos de software en idioma español u otros. Adicional, se identifica la oportunidad de generar un conjunto de datos en idioma español por la escasez