

## Declaración del Investigador Patrocinante:

Describe your role on this proposal development, including the benefit(s) of incorporating the applicant to your line of research and work group, as well as other aspects contributing to the relevance of its execution. The maximum length for this section is 1 page (Verdana font size 10, letter size is suggested).

La labor principal del investigador patrocinante será la de coordinar el trabajo de investigación y de aportar sus conocimientos en el área de Máquinas de Aprendizaje Distribuidas. El investigador es un académico jornada completa y tiene experiencia investigando en el área en cuestión, demostrable con publicaciones indexadas y conferencias en las áreas de regresión distribuida y métodos de aprendizaje automático.

De manera más específica el investigador patrocinante realizará las siguientes actividades durante la duración del proyecto:

1. Colaboración técnica y en infraestructura de procesamiento de datos para el proyecto, además de ayuda en redacción de reportes.
2. Coordinación de las etapas descritas en el Plan de Trabajo.
3. Supervisión de las metas establecidas en el proyecto.
4. Estudio y discusión de la literatura relevante.
5. Colaboración en el diseño del método propuesto.
6. Elaboración de artículos ISI en revistas indexadas.

En el núcleo del área de Máquinas de Aprendizaje existen primordialmente 2 problemas, el de clasificación y regresión. Estos 2 problemas comparten muchas similitudes, como la generación y selección de características. De hecho es posible ver de manera muy general el problema de clasificación como un subproblema del de regresión, pero con variable dependiente categórica. El problema de agrupamiento (Clustering) cae en la categoría de clasificación no supervisada. En trabajos anteriores del investigador patrocinante ha abordado la tarea de regresión en escenarios distribuidos y también de *BigData*. El tema de investigación propuesto, comparte sin duda las características de los problemas anteriormente investigados, por lo que resulta beneficioso poder explorar el mismo problema pero visto desde la tarea de agrupamiento.

Para finalizar, en escenarios de *BigData* se requiere de la participación de especialistas que ataquen problemas complejos desde distintas perspectivas, considerando como herramienta el computo distribuido y paralelo. Como resultado de este proyecto se espera satisfacer esta idea, incorporando nuevos métodos capaces de agrupar colecciones documentales con alta dimensionalidad en escenarios distribuidos, ya sea por el origen natural de los datos o como estrategia de procesamiento.