SPONSORING RESEARCHER STATEMENT:

Describe your role on this proposal development, including the benefits 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 (Must use letter size, Verdana size 10 or similar).

The sponsoring researcher collaborated in the elaboration of the work plan and he will coordinate the progress of the project. Additionally, this researcher will contribute with his expertise in the field of Distributed Machine Learning algorithms. The researcher is a full-time professor of the sponsoring institution and has a well known research experience supported by several publications in WOS journals and international conferences in distributed data regression and neural networks for classification and regression.

More specifically the researcher will perform te following activities during the execution of this project:

1. Collaboration in technical aspects for the design of the algorithms and in infrastructure for data processing.
2. Coordination of the stages described in the work plan.
3. Supervision of the goals contemplated in the project.
4. Discussion and study of the relevant literature to the project.
5. Contribution on the discussion and elaboration of the proposed methods and their experimental evaluation.
6. Participation in the elaboration of written documents for dissemination and publications in WOS articles.

At the core of Machine Learning there are two main tasks, namely Classification and Regression. Both tasks are tightly related since it is possible to formulate the classification task as a particular case of regression with a categorical target variable. The Clustering task is also considered as an unsupervised classification task. We expect that the experience and knowledge of the sponsoring researcher in Distributed Regression and Classification methods may be exploited in the context of a Distributed Clustering scenario. Undoubtedly, the proposed research topic shares the traits of the previously mentioned subjects thus this projects also offers the opportunity to explore the more general Classification problem from a different angle in a distributed data setting.

Finally, BigData problems require the integration of specific knowledge coming from experts that study the phenomenon from varied perspectives that also include techniques for parallel (at the CPU and/or GPU level) and distributed computing. We expect that this project satisfies this restriction by joining this partial efforts (sponsoring and principal researchers) and produces new methods capable of identify meaningful groups of documents from massive and distributed text collections.