Hi,

Fact Set has shared the below assignment for you to work on, as part of their evaluation process.

Here is the hackathon problem statement and data attached:  
  
**Problem statement**  
The discovery of private companies is a challenge as there are close to no regulation rules that can bring them under our radar. Amongst the primary sources of information are the company websites, the problem is to find those websites and to automate the discovery.  
  
You are given two lists of company websites currently known. The companies in the first list are all venture capital companies; the companies in the second tab are not.  
  
**Problem 1**  
  
Using this source material, we want you to build a model that, based on a company website, can predict if the company is a venture capital firm or not.  
  
**Problem 2**  
  
Based on this list of company's websites, determine a list of keywords that can be used in a web search engine to discover new similar companies. You can use the model from the first problem to measure your accuracy.

**Expected outputs**  
You can either provide your **code and research material in a zipped file or a github repository (preferred)**. A markdown file at the root must explain the structure of the folder, and more generally, outline your strategy and your results.  
  
More than the results you will achieve, we will consider your capability to address a problem, restate it if needed, break it down to individual pieces and manage the process of experimentation. We will also judge your ability to present your research in an intelligible way to both specialists and non-specialists.  
  
**Instructions to submit assignment:**

Please send in your work to below Mail IDs :  
  
**To:**  
V S Pavan Bachiraju - [vspavan.bachiraju@factset.com](mailto:vspavan.bachiraju@factset.com)  
Mickael Delattre - [MDelattre@factset.com](mailto:MDelattre@factset.com)

Satish Kumar Raju - [vkalidindi@factset.com](mailto:vkalidindi@factset.com)

**CC:**

Anupma Ranjan - [aranjan@factset.com](mailto:aranjan@factset.com)

T Satyendra Kumar - [stiruvuru@factset.com](mailto:stiruvuru@factset.com)

Sindhura Chilakapati - [sindhura.chilakapati@insofe.edu.in](mailto:sindhura.chilakapati@insofe.edu.in)

Subject Line: **YourName\_INSOFE\_DSHackathonSubmission**

In case of any queries regarding the Problem Statement or Data Set, you can write your query to V S Pavan Bachiraju - [vspavan.bachiraju@factset.com](mailto:vspavan.bachiraju@factset.com) and/ or Mickael Delattre - [MDelattre@factset.com](mailto:MDelattre@factset.com)

Deadline: **12:00 PM on 2nd March, 2020 (Monday)**

**I need you to revert to this mail to confirm that you understood the above instructions and that you will complete and share the task before the mentioned time on 2nd March. Failing to do so, your profile will not be considered for ANY further opportunities through INSOFE. (**Please read the Policy Document for more information**)**