|  |  |
| --- | --- |
| **Module:** | Big Data |
| **Assignment No:** | 1 |
| **Lecturer(s):** | Andrew Shields (andrew.shields@staff.ittralee.ie) |

**Assignment**

Working individually, you are to engage in the analysis of a data set using data analytics technologies or a combination of. For example Hadooop, Shiny and R Big Data tools. This should demonstrate your ability to apply the skills developed in the area of Big Data analysis technologies. Accompany this with a short paper that reflects on your work; and relates to your experiences and the available literature.

You should follow the KDD data analytics project life cycle discussed in class. The KDD process can vary in the number of steps. However, your assignment should cover as a minimum the steps below.



While candidates may **NOT** copy a tutorial or book chapter which they have not developed themselves you may reference existing tutorials and this is expected as part of you research into the topic once fully acknowledged and referenced. Candidates are required to demonstrate their ability to work under their own initiative.

This assessment is worth 50% of the module

* This assignment is much more about quality rather than quantity
* Research (find) and review several research papers at least five (preferably more).
* The formulation of your paper’s “theme” (title)
* Your approach MUST be “action” oriented (case study, applied etc) but may contain “theory” oriented (concept, investigation, State of the Art review)
* Your paper should be 6 pages in length

The following outlines what you should present in your answer in order to show your understanding of the chosen topic/area.

**Title**

* It’s important - put a little thought into it!
* Descriptive, interesting, novel
* Appropriate
* There may be a slight change to the proposed title – but nothing major!

**Abstract**

* Concise summary of paper
* 200 words maximum can be less (I will be strict about this)
* Your final abstract will have changed from the initial submission.

**Introduction**

* Give an overview of the area covered in the report, you should give a background to the topic demonstrating its importance to information systems. Giving a Context and reason why you have selected the topic e.g. what, why, how etc.

**Approach**

* Akin to “research method” chosen

**Discussion**

* This should provide the core of the report a description (summary) of the area – “State of the Art” so to
* speak. What are the major issues involved in the area / topic? How do / might these apply?

**Findings**

* You need to demonstrate what you have reviewed and discovered. Highlight where the area is “going”?
* Outstanding issues, unresolved questions, unknowns etc.

**Conclusion**

* Indicate what you have learned (key finding(s))? And what recommendation(s) you would make.

**References**

* Use a consistent style (Harvard)
* Include journal/conference proceeding references.

**Deliverables:**

1. **Practical Files:** Any practical input and outputs files from your data analysis should form part of the assignment. Code files, data files, and scripts result files etc. that demonstrate the analysis and technologies developed/used as part of the assignment. (theses should be submitted to blackboard or a CD-ROM/USB disk)
2. **Environment Setup:** you may use a standard machine/laptop, virtual machine or cloud service to run you system on. Your document should contain a brief section on how you achieved this.
3. **Word Document report:** with content describing the results of you analysis and the experiments that you ran.

Your report should have a section for each stage of the KDD data analytics life cycle.

1. Identify the problem
2. Design the data requirements
3. Pre-process the data
4. Perform analytics over the data
5. Visualising the data

You may also include any experiments and test that were unsuccessful and explain why they didn’t work.

Finally conclude your report by writing up your views and reflection on the overall experience referring to relevant literature where appropriate. In addition, provide some insights into the practical application of the technology(s).

**Grading Criteria:**

* Active and engaged participation in the development of the deliverables.
* Configuration and setup of a Data Analytics environment using Hadoop, R, Shiny etc.
* Clear evidence of work on each stage of the analytics life cycle stages.
* Application of technologies and insights into their use in a broader context.
* Evidence of logically structured and clear demonstration of data analysis with evidence of significant amount of outside reading which is integrated appropriately into the reflection and analysis content.
* Clear findings and conclusions from the data
* Clear analysis and interpretation of the process and experience with evidence of managing own learning through illustrated with examples where appropriate.
* Working data files to accompany analysis.
* Referencing as per standard college style (Harvard, Anglia Ruskin)