

# University of Glasgow

Peter Triantafillou

11/3/2016

## University Partner:

University of Glasgow

## Academic Supervisor - Name and Email Address:

Christos Anagnostopoulos; christos.anagnostopoulos@glasgow.ac.uk

## Suggested Project Title:

Predictive Modelling for Big Data Explorations: A Statistical Learning Approach

## Suggested Project Summary:

Description: Predictive modeling and the adoption of machine and statistical learning methods over large-scale data become necessity for gaining insight of complex processes to prove scientific theories and discoveries, to support decision making, and enhance strategic planning in different areas e.g., the economy, industry, healthcare, etc. Query-driven exploration over Big Data through discovering insights, patterns and trends in a timely fashion are becoming necessity in this field. This PhD research will focus on a novel problem of predictive analytics over distributed data in the sense that it can be deployed in environments in which data owners and sources restrict access to their data (e.g., due to security/confidentiality reasons, or cost reasons) and allow only certain statistical summaries or aggregation functions to be constructed/executed over the data.

Challenges: The challenge of this research relies on the idea of gaining insight knowledge only from interrogation/analytics queries and their results over distributed data. Moreover, it is challenging to develop novel query-driven supervised and unsupervised machine and statistical learning techniques over the analytics queries that can mine as much knowledge from restricted access data as possible. Such query-driven predictive analytics methods aim at multiple levels of knowledge abstraction and information fusion to allow the system to learn complex functions, mapping the input to the output indirectly, and provide evidence that on big data, sophisticated predictive analytics algorithms can achieve comparable or even better performance than those 'traditional' methods that explicitly require data access. In addition, the research will concentrate on problems associated with optimally scheduling of data access based on learning the query patterns over federated data nodes.

## Collaboration Sought for the Project:

The University of Glasgow and the Industry Sponsor will collaboratively explore a number of different issues such as: machine and statistical learning in high dimensional settings, information and knowledge fusion, scalable data access methods based on machine learning, complex analytics query processing and optimization, with direct applications on urban data, smart cities, and polyomics data. Real application domains

will enhance the applicability of the novel predictive modelling methodologies from this PhD research dealing with extracting and exploiting knowledge only from limited data access and cost/energy consuming predictive analytics tasks.

**Published or Private?:**

Yes