UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

PhD Course in Economics, Statistics and Data Science - ECOSTAT
XXXVIII cycle, a.y. 2022/2023

Scholarships

N. 1 linked to research project: "*Custom approaches for aerospace industry in Advanced Analytics*"

Company: Leonardo S.p.A.

Curriculum: Big Data and Analytics for Business

Abstract:

The research project will focus on developing complex decisional models applied to business improvement, cost reduction and change management, tailoring the solution to the aeronautic industry context. The candidate shall assess cutting edge models developed for the financial services industry, improving, and further developing these capabilities for the best performance, while leveraging on Big Data Analytics platforms. The PhD project shall also explore major techniques for enterprise decision management and behavioral economics, applied and enhanced to the specific scenario of an aerospace product supply chain. Development of innovative methods for clustering analysis and similarity measures shall be explored for customer behavioral identification applications. Also, the implementation of the models shall meet strict requirements in terms of computational scalability and parallel performance. The main envisaged outcome is the systematic definition of a set of decisional models, which might become the very foundation of a large-scale application of such techniques in aerospace industry. The outcome is a set of mathematical tools for assisting the different leadership roles of a common OEM company structure, with automated functionalities for model assisted decision making.

Intellectual property clauses agreed with the Company apply to this scholarship

PhD Executive Positions

N.1 linked to research project: *"Machine Learning for Improved Process Control"*

Company: MEMC Electronic Materials

Curriculum: Big Data and Analytics for Business

Abstract:

Process control, in the manufacturing world, is a fundamental tool for obtaining a competitive advantage as it allows to understand, analyse, and reduce the variability of the production processes. This means improving the cost structure (less waste, increased yield) and the quality of the finished product (fewer defects for end customers). The methodology mainly used in the company is the statistical process control (SPC), through which the critical parameters of the production process can be controlled. The elements that characterize the

SPC are a) the measurement system, including sampling frequency and sample size; b) anomaly alarm system; c) reaction mechanism (RFC) to bring the process back into control.
However, the SPC is a univariate control methodology that does not allow to control the interactions between the various critical parameters. Furthermore, the basic hypothesis of the SPC is the independence of the data. These two aspects undoubtedly represent a weak point because, in real production processes, many time series are non-stationary and moreover the interactions between the different variables are on the agenda.
To overcome the weaknesses of the traditional control methodology, the goal of the project is to build a multivariate anomaly recognition system using machine learning techniques. In the first phase, unsupervised learning algorithms will be used, for example auto-encoder, one-class Support Vector Machine, etc. to build an anomaly identification system. In the dynamics of signals or correlations the new system must be able to detect small deviations from standard conditions to prevent excursions. The algorithm uses historical data of the critical parameters selected to identify process anomalies.
In the second phase, the data collected on the anomalies will be classified into real / false problems to have a reference target that can be studied with supervised learning algorithms, for example lightgbm, xgboost, catboost, etc. This, based on experience, will make it possible to correct the predictions of the anomalies identified by the unsupervised algorithm.

N. 3 linked to research project:

1) *"Deep Learning for Business in Natural Language Processing"*
2) *"Model driven decision-making framework: an enterprise applications scope"*
3) *"Applications of Dynamic Network Models in Finance"*

Company: Intesa Sanpaolo

Curriculum: Big Data and Analytics for Business

Project 1:
Natural Language Processing (NLP) is a subfield of Artificial Intelligence which has close ties with Machine Learning. Basically, NLP works towards making the computer understand what we humans mean and intend. There are many pre-trained models that use Deep Learning for general text but only a few for text in finance domain. The financial text contains specialized words or words with a meaning different from the common one. The NLP models for finance are finBERT and, the newest, FinEAS (Financial Embedding Analyst of Sentiment). Furthermore, the source code released from these pre-trained models covers mainly Sentiment Analysis tasks, among which the most requested is Business Classification. Financial texts must be tagged with Economic Sector, Business Sector, Industry Group and Industry from for example TRBC Sector Classification. This research project aims to examine these models, find possible improvements, and explore how to implement Business Classification.

Project 2:
Artificial intelligence (AI) is continuously reshaping services concepts in the decision-making systems. However, the volume of data has exceeded the enterprise ability to manage the increasing number of silos of data, in order to take advantages from it. In the meantime, the adoption of an huge number of software engineering practices such as MLOps (Machine Learning Operations), in business operations, has allowed to catch value from data in a new and easiest way. The aim of this proposal is to focus on a framework that grants the use of AI from the business layer like a service, based on a model-based approach for the abstraction of the AI models.

Project 3:
The fundamental mathematic concept to model the interconnection pattern of a network is a graph. To explain it, we can start with a set of nodes representing the unit that constitute the network (e.g. people, economic assets, companies...). The nodes are connected by links: the presence of a link may have different interpretations depending on the application (if we consider the nodes as companies the links may be an economic relationship, for example the presence of a payment). Sometimes it is useful to associate a positive scalar to each link, to be referred to link's weight, with the aim of quantifying its activity level.
The network theory has many application sectors, and one of this is the financial, especially in risk management. For example, if the network is composed of companies and the links are the payments between

two companies in a specific time interval, we can study how the phenomenon of default flows through the nodes. This phenomenon has similarities with the epidemic contagion models, in fact a default of a company could generate a cascade of default of his neighbors. In addition to the dynamics of the network, we can also study its topology in order to include its characteristics as features in Risk models.