PhD in Computer and Control Engineering

Research Title: Knowledge Discovery from Financial Data

The effects of market globalization have produced an enormous amount of financial data coming from various contexts. Budgets, financial reports, business plans, share prices, and economic news articles represent a huge mass of information potentially relevant for private and public stakeholders.

Extracting value for these data collections entails coping with multi-modal, heterogeneous, and potentially Big data collections. For example, news about financial stock markets can be shared through videos, radio and TV newspapers, or social media; share price series can be analyzed at various data granularities according to the purpose of the domain experts; financial reports can be periodically crawled from Web-based interfaces. However, the content of financial resources is often very technical, partly redundant, written in different languages, and contextualized to specific domains.

Machine learning and data mining techniques have already found application in the financial domain to tackle specific use cases, among which (i) stock market prediction, (ii) currency and cryptocurrency trend discovery, (iv) fundamental share analysis, (v) portfolio management and optimization, (vi) competitive intelligence, and (vii) emerging market analysis.

The research activity will regard the study and application of innovative machine learning and data analytics approaches to addressing real financial use cases. The selected cases will be customized on real industrial and research challenges and will be addressed by means of cutting-edge data-driven solutions relying on the analysis of multi-modal, multi-lingual, and large-scale datasets.
The PhD student will address the study and application of smart Big data analytics solutions to solve relevant financial issues. The research activities involve multidisciplinary knowledge and are aimed at finding smart solutions that can be easily extended and adapted to different use cases. More specifically, the research will address the following key issues related to financial Big data analysis:

(i) *Lack of public data.* In many financial contexts, the raw data collections are often private or not easy to retrieve. For example, in many real cases public financial reports just provide a summarized, high-level view of the market trends. More detailed information is often kept undisclosed (at least at a finer granularity). Hence, there is a need for transferring domain knowledge and trained models from related domains or contexts in order to effectively support experts’ decisions.

(ii) *Risk exposure.* Properly managing financial resources is a crucial task in any business activity. Delegate choices regarding investments, asset management, and marketing strategies to machine learning-driven systems poses significant challenges in terms of capability of data-driven systems to properly balance risks and rewards.

(iii) *Temporal correlation between financial trends.* Most of the financial data continuously vary over time (e.g., share prices, currency values). However, detecting reliable temporal correlations among historical data can be challenging due to the non-stationarity of the analyzed series. Instead of modelling data as multivariate time series, an aggregated view of the analyzed data could provide a time-independent description of the underlying time series.

(iv) *Volume and variety of the data sources.* The analyzed data are usually acquired from various media (online newspapers, public Web platforms, social media, video communication infrastructures). Data are usually stored into Big collections of multimodal data. Furthermore, the retrieved textual and video contents are provided in various languages. Hence, there is a need for analytical processes able to integrate and analyze heterogenous data.

The PhD work plan can be summarized as follows:
(a) In the first year the student will study the state-of-the-art solutions to develop decision support systems based on financial data analysis. According to the preliminary overview of the existing literature, the open issues related to financial data integration, risk management, and financial series forecasting will be identified. Preliminary solutions to the aforesaid issues based on the extension of existing decision support systems will be proposed and evaluated on benchmark data.

(b) In the second year the student will investigate the portability of the existing solutions to different domains. Specifically, she/he will focus on transferring learning models and domain knowledge to related domains in order to overcome the limitations due to the lack of data for supervised and unsupervised analyses.

(c) In the last year the student will study new approaches to handling
multimodal and multilingual data sources. Furthermore, she/he will investigate the applicability of time-based solutions to specific use cases and compare their performance with that of machine learning models relying on time-independent features.

During the PhD the student will attend international conferences and workshops, and will participate to research challenges, e.g., the Financial Entity Identification and Information Integration (FEIII) Challenge.

<table>
<thead>
<tr>
<th>Skills and competencies for the development of the activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The candidate must have good knowledge of machine learning and data mining techniques, data preparation techniques, and business analytics tools.</td>
</tr>
<tr>
<td>The candidate should have good programming skills (proficiency in the Python language is recommended) and basic knowledge of Big Data frameworks (e.g., Hadoop, Spark).</td>
</tr>
</tbody>
</table>