In consideration of the determination of the Regione Piemonte – Direzione Coesione sociale No. 481 of August 24, 2021 which approved the following apprenticeship position for the PhD project proposal submitted by the Politecnico di Torino in the framework of a specific regional call for proposals (Apprendistato di Alta Formazione e Ricerca 2016-2018 – Avviso Pubblico per la realizzazione dei percorsi formativi di: Laurea triennale e magistrale, Diploma Accademico di primo e secondo livello, Master di primo e secondo livello Universitario, Dottorato di ricerca e Diploma accademico di formazione alla ricerca, Attività di ricerca approvato con Determinazione 537 del 3/8/2016 e s.m.i.):

**PhD in Computer and control engineering**

**Research project “Sviluppo e applicazione di tecniche di Natural Language Understanding per l’arricchimento di modelli predittivi”**

**Politecnico di Torino – Maize srl**

NLU per arricchimento di modelli predittivi

| Supervisor           | Prof. Luca Cagliero – Dipartimento di Automatica e Informatica Politecnico di Torino
                      | luca.cagliero@polito.it
                      | Francesco Tarasconi – Maize srl
                      | francesco.tarasconi@h-farm.com |

| Contact              | https://dbdmg.polito.it Research group: Database and Data Mining (DBDM) - Dipartimento di Automatica e Informatica - Politecnico di Torino
                      | https://www.celi.it/ (Company) |
## Context of the research activity

One of the most challenging issues addressed by modern organizations is to fruitfully exploit the available business- and customer-related data. To transform the raw data into actionable knowledge there is often the need to cope with heterogeneous data sources. Specifically, in many business scenarios the largest part of the available data is unstructured and consists, for instance, in large raw text documentary bases.

The research activity is mainly focused on the design, study, and development of innovative methods of Natural Language Understanding, which leverage unstructured data in textual form to solve complex business-oriented problems. The research will address both methodological aspects related to the study of state-of-the-art NLP techniques based on Deep Learning technologies and applicative aspects focused on developing effective and efficient solutions tailored to real case studies.

The research will focus on deeply understanding the capabilities of the developed models, to be adapted and integrated into innovative Decision Support Systems. The aim is to help domain experts to take crucial decisions.

Finally, the research will deepen the study of continuous learning and feedback methodologies with the goal of improving model performance over time and to make models reactive to sudden changes.

The Company Maize has planned for the winner of this position a collaboration within a contract of high apprenticeship according to the Italian Legislative Decree 81/2015, art. 45.

## Objectives

The research project has three main objectives:

1) Study of the state of the art of NLP technologies, with particular emphasis on data representations and language models. A particular attention will be also paid to the recently proposed Data Fusion techniques, whose aim is to integrate multimodal data sources (e.g., text, videos, audio) for the resolution of one or more specific business problems.
2) The apprenticeship will then focus on developing and testing innovative methodologies to create predictive models (e.g., sales forecasting) that exploit the large pool of heterogeneous data available to organizations. At this stage, a key aspect will be the maximization of the explanatory power of the phenomena of interest as it is crucial for effectively supporting business decisions. Pattern extraction techniques, clustering algorithms, and Explainable AI architectures are examples of established techniques that will be considered during the research activities.

3) The PhD candidate will also study how to integrate new data sources that can improve the robustness of predictive models. The goal is to make them capable of intercepting causal links (e.g., representing why certain customers orient themselves on certain types of products) rather than relying solely on variable proxies (e.g., aggregate historical series of sales).

The application of the developed solutions to practical use cases will allow the research to tightly collaborate with people with multidisciplinary expertise, integrate the solutions into real Decision Support Systems, and carry out qualitative and quantitative assessments.

Skills and competencies for the development of the activity

The candidate shall be **less than 30 years old** at the moment of the hiring from the company.

The PhD candidate is expected to carry out all the research activities in a proactive approach: during the whole apprenticeship she/he will have to profitably collaborate with other researchers, both in Maize srl and in the academia, to understand the best practices, to put into practice the theoretical knowledge, and to interpret the empirical results in a productive way.

From a scientific point of view, the PhD candidate will have to explore cutting-edge solutions to effectively and efficiently cope with heterogeneous data sources, perform rigorous empirical assessments, and provide clear evidence of the benefits of the achieved research findings. The most significant results will be published in international conferences and journal papers. Hence,
the researcher will become proficient in scientific paper writing and scientific knowledge dissemination (including oral presentation at conference and workshop venues).

From a business viewpoint, the PhD candidate will be asked to tailor the designed methodologies and systems to real case studies. To this aim, she/he will explore the feasibility of the data-driven methodologies from a variety of viewpoints, interact with business experts to find the necessary optimizations and tradeoffs, and appropriately synthesize the information in order to allow the exploitation of the new knowledge acquired.