FCT Fundação para a Ciência e a Tecnologia

Research in Data Science and Artificial Intelligence applied to Public Administration

Index

Index

Introduction

Knowledge is a pillar for the development of societies as well as the driver for many new and innovative applications. One example is given by the technical achievements in the last decades in the field of computing and communications that, among many fantastic developments, also allow the collection and processing of unprecedented large amounts of data, leading to deep changes in the methodologies used in diverse activities and fields of knowledge. Data is obtained routinely by the functioning of many services, namely by those provided by the public administration bodies, and represent a valuable information resource for research and knowledge creation. Data analysis may help to improve and optimize processes with examples ranging across societal activities so diverse as medicine, mobility, environment, security, education or employment.

Data science, which comprises the transformation, analysis, visualization and presentation of data, has the potential to create a more informative environment for public debate and political decision with the ultimate goal of achieving benefits for our quality of life and society. The continuous improvement of public services, the optimization of resources, the mitigation of fraud and error, the increase of economical returns are a few examples. Linking the available data pool to the existing scientific methodologies in advanced computing and artificial intelligence is a major task that requires targeted collaborative actions.

The Fundação para a Ciência e a Tecnologia (FCT), as the Portuguese agency for funding of research and technology, has taken this challenge of promoting scientific knowledge through the analysis of large amounts of data available in the public administration, and their application to societal relevant issues. The goal is to improve significantly the services provided to citizens and businesses, as well as to improve decision-making processes, so that they will increasingly be based on in-depth knowledge of reality and technically supported by evidence.

With this in mind, and as part of the National Digital Skills Initiative e.2030, Portugal INCoDe.2030 (Axis 5 - Research), in 2018, FCT launched a research Program in Data Science and Artificial Intelligence in Public Administration, to support new R&D projects with partnerships between public administration and scientific institutions. After a preliminary exploratory action, three competitive public tenders were programmed to open annually, with an overall allocation of €10 million.

In the 2018 and 2019 editions of the calls for Scientific Research and Technological Development Projects in Data Science and Artificial Intelligence in Public Administration, 32 research projects were selected for funding, and are now underway. They span across several fields from food safety to health care and on-line gambling, partnering and using data available in various public institutions, with research carried out by diverse teams from our scientific community. These projects are summarized here.

In complement to the public dissemination of this targeted research funding, and to the fostering of interaction within the scientific community, FCT wishes to increase awareness on the enormous potential of the available data pools in public administration institutions for science and innovation, and to foster new partnerships linking research and administration aiming at better services and processes to society.

Helena Pereira
President of Fundação para a Ciência e a Tecnologia

PROJECTS OF SCIENTIFIC RESEARCH AND TECHNOLOGICAL DEVELOPMENT IN DATA SCIENCE AND ARTIFICIAL INTELLIGENCE IN PUBLIC ADMINISTRATION

MOBILITY AND TRANSPORT

MOPREVIS DSAIPA/DS/0090/2018

MODELING AND PREDICTION OF ROAD TRAFFIC ACCIDENTS IN THE DISTRICT OF SETÚBAL

PRINCIPAL INVESTIGATOR	Paulo de Jesus Infante dos Santos
LEADING INSTITUTION	Universidade de Évora
PARTICIPATING INSTITUTIONS	Guarda Nacional Republicana
FUNDING	€ 299 986

In Portugal the number of crashes with victims has been increasing since 2011 and the district of Setúbal is one of the districts with the highest number of crash accidents. Big Data Analytics is the process of automatically examining large amounts of data to uncover hidden patterns, unknown correlations, and other information that can help organizations make more informed decisions. Based on the large amount of data available, the overall objective of this project is to identify factors that enhance the accidents in the district of Setúbal and to build predictive algorithms and models that can be used experimentally, in real time, by GNR of Setúbal, in order to take measures to reduce road accidents. For this project it was formed a multidisciplinary team that integrates computer engineers, specialists in geographic information systems (GIS) and mathematicians with specialization in probability and statistics, to: i) construct an information system about the accidents occurred in the district of Setúbal, combining several sources of information: digital platform (storage solution, large-scale data processing and management with distributed architecture), data acquisition module (module for extracting information from accident documents with automatic processing techniques of natural language and artificial intelligence and interface module for other systems, namely automatic data acquisition on traffic intensity in adjacent streets and climatic data) and module for information recover; ii) classify sites with high number of accidents (hotspots); iii) identify the determinants factors that potentiate the occurrence of accidents and their severity; iv) draw the profile of the individuals involved, comparing it by type (collision, trampling or scraping); v) construct predictive models for the number and severity of accidents, as well as for the most likely places for its occurrence, and also to obtain a model that can predict an accident given a road segment and given a time period. The final objective is to build a digital tool to support real-time decision making. The success of this project will certainly have a great social and economic impact, and the idea can be replicated to other districts.

iLU DSAIPA/DS/0111/2018

INTEGRATIVE LEARNING FROM URBAN DATA AND SITUATIONAL CONTEXT FOR CITY MOBILITY OPTIMIZATION

PRINCIPAL INVESTIGATOR	Rui Miguel Carrasqueiro Henriques
LEADING INSTITUTION	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC ID/INESC/IST/ULisboa)
PARTICIPATING INSTITUTIONS	Câmara Municipal de Lisboa and Laboratório Nacional de Engenharia Civil
FUNDING	€ 299 725

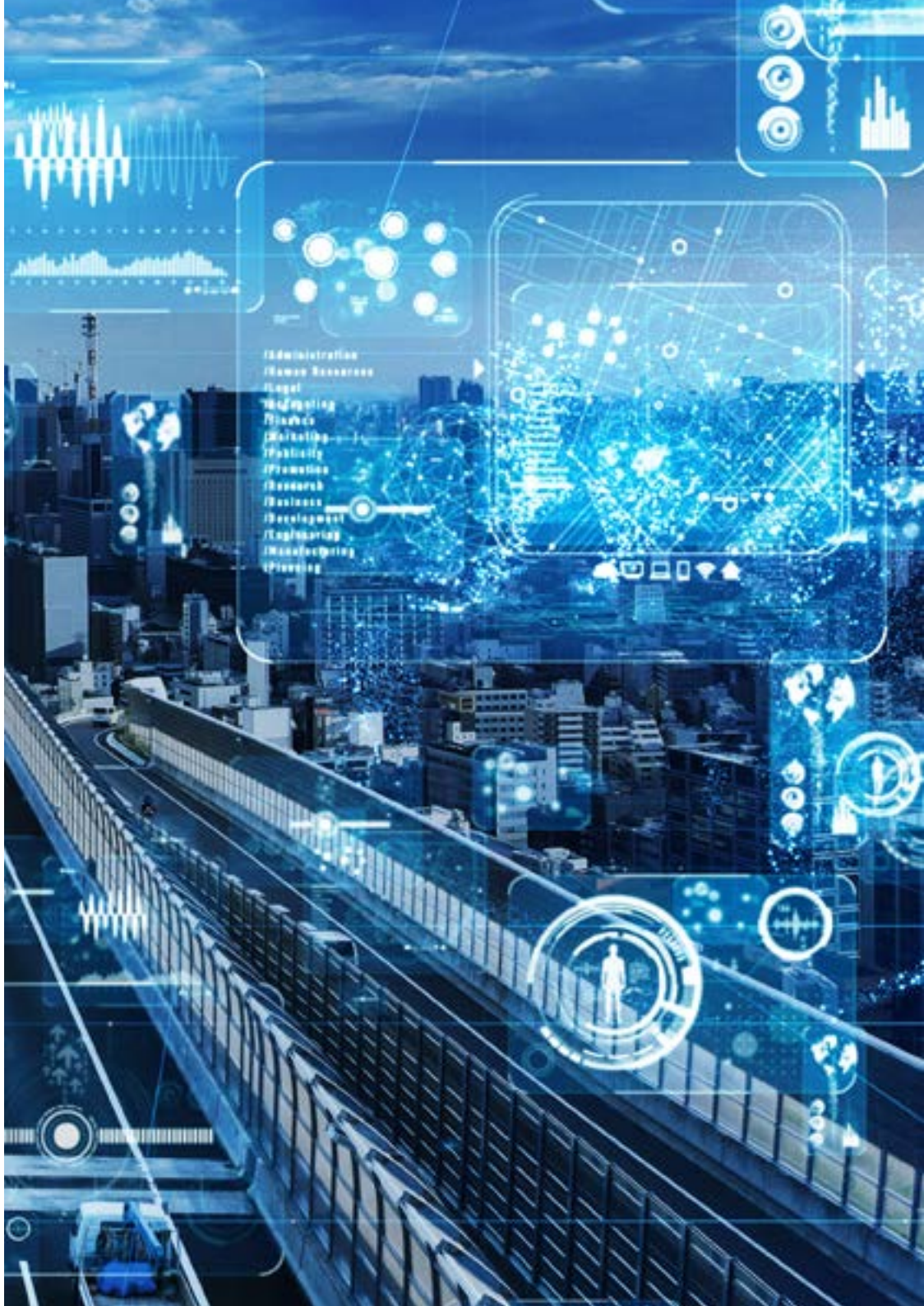
Mobility of passengers and freights in most European capital cities such as Lisbon is not yet sustainable. In addition, mobility data is dispersed through several entities/operators. Motivated by this reality, the Lisbon City Council (CML) established in the last years acute efforts to collect and access all available mobility data, including road traffic data from loop counters, geolocalized speed data from mobile applications, public transport data (bus, subway, train and bike modalities), and situational context data (public events, meteorological conditions, construction works, etc.). Despite these efforts, the potentiality of learning from such data diversity remains untackled. In this context, iLU project aims to: 1) consolidate the multiplicity of data sources on city mobility stored in the Plataforma de Gestão Inteligente de Lisboa (PGIL) and guarantee its real-time updatability; 2) discover actionable spatiotemporal patterns of mobility from such heterogeneous data sources, particularly non-trivial correlations between road traffic and situational context data; 3) anticipate traffic congestion using advanced and integrative predictive models; 4) real-time support of mobility decisions through the use of deep reinforcement learning to positively condition the city traffic by, for instance, controlling traffic lights and road message panels. As a result, iLU proposes to address four major challenges: 1) the lack of integrative views combining the input of multiple sensor modalities (such as mobile data and loop counters) and heterogeneous traffic modalities (such as private and public transport); 2) the absence of situational context data in traffic predictions; 3) the inability to accurately forecast traffic for time horizons above 15 minutes; 4) the untapped utility of data-driven control and simulation tools for urban mobility optimization. Expected scientific contributions include advanced algorithms to mine context-sensitive spatiotemporal traffic patterns; and deep reinforcement learning methods to optimize traffic in accordance with context-sensitive predictions. These contributions will be made available within an auditable decision support system, attemptively deployed over the National Infrastructure for Distributed Computing (INCD), to be promptly used by CML. In addition to the immediate utility of iLU for improving the mobility in Lisbon and its managing, the contributions will be designed to ensure its interoperability and scalability to other cities in Portugal and beyond.

FailStopper DSAIPA/DS/0086/2018

EARLY FAILURE DETECTION OF PUBLIC TRANSPORT VEHICLES IN OPERATIONAL CONTEXT

PRINCIPAL INVESTIGATOR	Rita Paula Almeida Ribeiro
LEADING INSTITUTION	INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
PARTICIPATING INSTITUTIONS	Metro do Porto, S.A.
FUNDING	€ 95 148

The occurrence of failures in public transport vehicles during its regular operation is a source of numerous negative impacts that affect not only the operator company but, especially, the clients. In this context, the early detection of such failures can avoid the cancellation of trips. This project focuses on the analysis of the failures of the Compressed-Air Production Unit (APU) in the Metro do Porto (MP) fleet, one of the equipment that most contributes to the cancellation of trips. The goal of the project is to develop a system based on real-time data analysis to notify the maintenance team of the existence of a failure in development, which is undetectable according to traditional maintenance criteria, avoiding its occurrence during the train operation. The use of automated data analysis systems for predicting failures is a cutting-edge technique in high-end industries such as space or aeronautics and is either inexistent or very incipient in less technological sectors. The central APU system installed on the roof of vehicles from MP feeds different units, which perform functions of different degrees of criticality. The maintenance regime currently applied to the APU system is based on pre-scheduled and corrective maintenance activities to respond to events where the failure has already manifested itself in its ultimate consequences. Our goal is to install a set of sensors in the APU system of a vehicle so that data is collected regularly. This data is pre-processed to form a data set based on which we can develop an early failure detection model, following an unsupervised or a semi-supervised approach. Ideally, this model should trigger an alarm to allow the intervention in a very embryonic phase of the failure, in which its symptoms and consequences are still imperceptible, thus avoiding the cancellation of trips. Once the failure detection model is implemented, we intend to study and infer what are the most probable causes of the failure. For this purpose, we will explore probabilistic and/or contextual approaches methods. An effective fault diagnosis can reduce the costs of the repairing plan. Moreover, it can be used to optimize the sensor equipment required, and thus, improve the conditions of application of this early failure detection system to the entire fleet of vehicles of the MP. As a long-term perspective, we believe that this study can be extended to other subsystems of the fleet such as traction converter or automatic doors.



HEALTH AND QUALITY OF LIFE

IPOscore DSAIPA/DS/0042/2018

PREDICTING THE RISK OF COMPLICATIONS OF SURGICAL TREATMENT AND DEFINE PROGNOSIS OF CANCER PATIENTS THROUGH CLINICAL AND BIOPATHOLOGICAL DATA INTEGRATION

PRINCIPAL INVESTIGATOR	Rafael Sousa Costa
LEADING INSTITUTION	Instituto de Engenharia Mecânica
PARTICIPATING INSTITUTIONS	Instituto Português de Oncologia do Porto Francisco Gentil, E.P.E.
FUNDING	€ 247 056

Biomedical research is primarily concerned with the understanding of biological systems and disruption mechanisms involved in disease. In recent years, technological advances in related areas of research have pushed high-throughput instrument precision, and increasingly accurate large-scale data are being collected at an outstanding pace. Considerable effort has been put into ensuring the scalability of tools for the storage and managing of these massive streams of spatio-temporal data inducing a rapid change in clinical and health care research, but analysis remains a challenge. In this context, there is a growing demand for computational methods with ability to extract and build meaningful knowledge from the deluge of clinical data, while coping with its noisiness and incompleteness. While these new possibilities open up tremendous potential, there are very significant challenges to making this vision a reality. Several authors emphasized the need for integrative approaches to provide a broader understanding of oncological diseases. These approaches should infer relationships between clinical and personal data of the risk. However, and although promising first attempts integration wise are arising, currently no effective integrative system in Portugal have been proposed to date using both personal and clinical data to deeply study several type of cancers and analyze the risk of postoperative complications in cancer patients. Under this scope, the present proposal aims the development of novel mathematical models and machine learning methods for personalized clinical data analysis, towards the P4 Medicine: Predictive, Preventive, Personalized and Participatory. The expected outcome is the analysis of large amounts of clinical data in order to extract and build sophisticated knowledge, and the creation of a web application to store the data and calculate a IPOscore that can provide added-value to surgeons, which will have the capability of generating new hypotheses in cancer patients. A knowledge discovery system will integrate powerful data mining algorithms to unravel potentially relevant links between patient profiles and clinical data, and predictive models for the risk of complications after surgery in cancer patients. Disease diagnostic and prognostic markers, disease progression rates, and patient profiles, will be tackled. Together with the challenging task of studying cancer diseases we will also embrace the challenging topic of developing efficient models and mining algorithms for clinical data integration.

NeuroBioAI DSAIPA/DS/0065/2018

ARTIFICIAL INTELLIGENCE-BASED NEUROIMAGING BIOMARKERS FOR THE DIAGNOSIS OF NEUROPSYCHIATRIC ILLNESSES

PRINCIPAL INVESTIGATOR	Diana Maria Pinto Prata
LEADING INSTITUTION	FCiências.ID - Associação para a Investigação e Desenvolvimento de Ciências
PARTICIPATING INSTITUTIONS	Centro Hospitalar de Lisboa Centro, E.P.E., Hospital Professor Doutor Fernando Fonseca, E.P.E., Serviços Partilhados do Ministério da Saúde, E.P.E., NEUROPSYAI - Serviços e Investigação em Neurociências, Lda. and Hospital Distrital de Santarém, E.P.E.
FUNDING	€ 299 925

Nowadays, diagnosis of neuropsychiatric disorders such as Alzheimer's and Parkinson's diseases rely on clinical interview, observation and neuropsychological testing, and, when in doubt, on blood and cerebral-spinal fluid exams, or imaging exams such as CT, MRI and PET. Diagnosis is still heavily dependent on the clinician's experience, and it is very hard at early stages of disease. There is a huge dependency on the clinicians' visual inspection of the images and its comparison with a memorized image of what the disease should look like. As a consequence, these diseases typically take on average 2 years to correctly diagnose. Diseases become more severe, they stay under- or mis-diagnosed, their treatment is delayed, and hence patients' and carer's distress is high. We are excited about combining artificial intelligence (AI) and neuroscience to revolutionize mental healthcare. We are currently working to capitalize on the information contained in brain scans acquired every day at hospitals and clinics. We do this by training AI algorithms to recognize disease signatures in those brain scans, using databases of diagnosed patient brain scans. We now have prototypes able to detect disease signatures at early stages of disease; which will be able to provide clinicians, not only a patient-personalized descriptive report of several brain measures but, most innovatively, a score probability of the patient presenting a particular neuropsychiatric illness. This enables a more sensitive, accurate, objective and quantitatively-based diagnosis and thus earlier and more cost- and time-effective disease management. This project aims at the statistical validation and the technical and clinical-administrative trialling of biomarker models for diagnostic classification for the most common neurodegenerative illnesses, Alzheimer's' and Parkinson's. Moreover, there is potential for easy scalability disease-wise: the successful access to retrospective data from hospitals will allow further neuropsychiatric illnesses data to explored with the same neuroimaging and AI-based statistical pipeline in our servers (for example, multiple sclerosis, migraines, epilepsy, schizophrenia and bipolar disorder). Our vision is that with tight collaboration with the Portuguese administration services, we will be able to use existing public health data which, via our neuroimaging AI-based approach, can then give back to the same national public health system a cost-effective disease management solution.

ICDS4IM DSAIPA/DS/0084/2018

INTELLIGENT CLINICAL DECISION SUPPORT FOR INTENSIVE MEDICINE

PRINCIPAL INVESTIGATOR	Manuel Filipe Vieira Torres dos Santos
LEADING INSTITUTION	Universidade do Minho
PARTICIPATING INSTITUTIONS	Centro Hospitalar Universitário do Porto (CHP/MS)
FUNDING	€ 264 889

Real-time, online-learning and intelligent decision support systems are of most importance to supply intensive care professionals with important information in useful time. Decision errors affect 10% of hospitalized patients. In the case of 30-day postoperative, 12.6% of deaths were associated to medical errors. An excellent decision-making is focused into 3 principal categories: (1) accurate data, (2) pertinent knowledge, and (3) appropriate problem-solving skills. Combining the results achieved in previous projects and making use of the data collected from the bedside monitors, ventilators, pharmacy, electronic health record, electronic nursing record, laboratory system, patient medical history, therapies and procedures, this project aims to extend the actual state of the art in order to support clinical decision-making by: 1 - Translating clinical notes based in natural language/narratives into useful data for analytics; 2 - Automatic problem (clinical findings) detection; 3 - Discover cause-effect relations among problems, semiologic notes, assessment notes and clinical decisions (planning); 4 - Improve intelligent decision support on therapies, orders and procedures; 5 - Extend to other intensive medicine areas. The expected results with greatest societal impact and in the ICU decision and management are: 1 - Augment the accuracy on the clinical decisions - the physicians will be supported in their decisions about the most adequate therapies/procedures. They will be able to assess what-if scenarios and decide better; 2- Promote evidence-based clinical decisions – clinical decisions on therapies, procedures and orders (e.g, lab exams) will follow open models based in a clinical ontology for intensive medicine; 3 - Acting proactively – the decision models obtained will be used in association with prediction models to anticipate good decision with non-return situations; 4 - Augmenting the quality of life – a more accurate decision combining patient medical history with decision scenarios can improve the patient life quality; 5 - Costs reduction without compromising the efficiency – cheaper decisions will be proposed without compromising patient outcome; 6 - Reduce the number of clinical errors – decreasing the number of bad decisions taken patients will recover faster and will have a better quality of life reducing costs; 7 - Improve quality of ICU– quality performance will increase.

LAIFEBLOOD DSAIPA/AI/0033/2019

ARTIFICIAL INTELLIGENCE FOR BLOOD MANAGEMENT

PRINCIPAL INVESTIGATOR	Pedro Tiago Gonçalves Monteiro
LEADING INSTITUTION	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC ID/INESC/IST/ULisboa)
PARTICIPATING INSTITUTIONS	Associação do Instituto Superior Técnico para a Investigação e o Desenvolvimento and FCiências.ID - Associação para a Investigação e Desenvolvimento de Ciências
FUNDING	€ 222 687

The Instituto Português do Sangue e da Transplantação (IPST) is a public institute of the Ministry of Health, and the recognized authority for the collection and regulation of blood donation and transplantation at the national level. Operational productivity of IPST is crucial for the blood supply chain in Portugal, since it monitors the quality and safety of donation, procurement, processing, storage, preservation and distribution of human blood, blood components, organs, tissues and human origin cells. The goal of LAIfeBlood project is to provide IPST with new tools to improve efficiency of blood supply operations in Portugal. LAIfeBlood project is particularly focused on: (1) extracting temporal and geographical patterns from historical collection and transfusion data; (2) developing predictive models for the expected collections and transfusions in different temporal and geographic areas; and (3) developing software tools to optimize the planning of collections at the national level according to the availability of IPST teams and resources and collection events. LAIfeBlood project tackles an important societal challenge of maximizing blood services effectiveness while minimizing wastage of an increasingly scarce resource and optimizing the efficiency of blood supply operations. We expect our results to be used to derive implementable policies at IPST to plan collections and manage the daily inventory rebalancing problem, as well as contributing to the development in the field of optimization and dynamic supply chain management.

PrevOcupAI DSAIPA/AI/0105/2019

PREVENTION OF OCCUPATIONAL DISORDERS IN PUBLIC ADMINISTRATIONS BASED ON ARTIFICIAL INTELLIGENCE

PRINCIPAL INVESTIGATOR	Hugo Filipe Silveira Gamboa
LEADING INSTITUTION	NOVA.ID.FCT - Associação para a Inovação e Desenvolvimento da FCT (UNL)
PARTICIPATING INSTITUTIONS	Universidade Nova de Lisboa
FUNDING	€ 234 864

Work-related disorders (WRD) have a major impact on the well being of individuals and their quality of life, productivity and absenteeism, resulting in a great impact on the global economy. The absence of recent or relevant national statistics concerning WRD, leads to a lack of drive to promote studies that characterize the working population, in terms of work-related morbidity and mortality and scientifically supported policies and strategies for occupational health(OH) improvement. This is particularly poignant when addressing the issue in Public Administration (PA). Hence, those diseases are underrepresented on public debates and global population awareness. In spite of the reported work on clinical occupational history, and studies evaluating the occupational exposure to hazards, based on the analysis of biosignals of repetitive activities, there are no studies that combines the historical OH and data from daily working activity, with biosignals. The project "Prevention of occupational disorders on PA based on Artificial Intelligence" will answer to the aforementioned question, by contributing to: the identification and characterization of profiles related with potential WRD, of the Portuguese population and their relationship with professional categories and risk situations at work; forecast the progress of occupational disorders and the associated risk factors, to evaluate the socio-economic impact in morbidity and mortality, with emphasis to the reduced life expectancies caused by disorders and incapacity; the characterization of daily activities, while profiling a sector of workers of PA; the identification of clusters of working activities prone to measure the exposure to risk factors; the definition of national indicators to help monitor and give alerts for occupational risk at PA. Three computational models, based on artificial intelligence (AI) techniques, will be developed: (1) global characterization of occupational disorder profiles; (2) design of daily working activities profiles of an individual; and (3) integration of the information resulting from both previous models, enabling to infer the level of occupational risk of a worker based on his/her occupational history and the data from daily monitorization. This project aims to identify occupational risk, by combining the power of historical records of the global population and the precision data of personalized occupational exposure records. The combination will be achieved through advanced machine learning and AI techniques.

AI4Life DSAIPA/DS/0054/2019

ARTIFICIAL INTELLIGENCE AS AN INSTRUMENT FOR THE REDUCTION OF INFANT AND YOUTH MORTALITY: UNDERSTANDING ITS DETERMINANTS AND PREDICTING OUTCOMES

PRINCIPAL INVESTIGATOR	Susana Margarida da Silva Vieira
LEADING INSTITUTION	Instituto de Engenharia Mecânica (IDMEC)
PARTICIPATING INSTITUTIONS	Faculdade de Medicina da Universidade de Lisboa, Faculdade de Ciências Médicas (UNL) and Direcção-Geral da Saúde
FUNDING	€ 222 804

One of the nine goals defined by the World Health Organization in 2025 focuses on the reduction of premature mortality from noncommunicable diseases by 25%. The plans to reduce premature mortality have been primarily focused on the age group between 35 and 69 years. However, despite the higher incidence of mortality in this age group, infant-juvenile mortality (<=18 years) also represents a significantly high social and economic burden to the country. The study of mortality and its determinants is a complex challenge for the age group 35-69, but the study of infant and juvenile mortality poses even greater challenges. At present, mortality monitoring at Direcção Geral de Saúde (DGS) is essentially based on a descriptive statistical analysis that reveals the risk of mortality by cause, age, sex and geography. The analysis is limited, since it is based on a small number of demographic characteristics. It lacks a holistic view of the life span of the citizen, of the varied events that characterize his health state. Machine learning techniques allow the discovery of important patterns and relationships in the data and may lead to the discovery of knowledge previously unknown to humans. In this way, it is possible not only to confirm and objectify previous notions about certain patterns, but also to obtain new knowledge. In this context, the goals of this project are: 1) to collect and integrate information from several data sources, which have never been combined before, in order to characterize the infantile-juvenile population in several dimensions; 2) analyze the data using traditional statistics and visualization techniques; 3) use advanced artificial intelligence techniques, which allow finding patterns in multidimensional data, not yet appreciated, in order to extract useful knowledge; 4) identify distinct profiles through unsupervised machine learning techniques; 5) use supervised machine learning techniques, combined with the patterns and knowledge obtained in the previous phase, to successfully predict which individuals, children and young people in particular, have a higher risk of mortality. The main objective of the whole project is to leverage existing information in public administration databases and others in order to support decision-makers regarding the best response to emerging diseases, better adaptation of public health intervention programs and improve the capacity of the health systems in the future.

DEMONSTRADOR

ANALYSIS OF ANTIBIOTIC PRESCRIPTION PATTERNS

PRINCIPAL INVESTIGATOR	Joana Gonçalves de Sá
LEADING INSTITUTION	Nova School of Business and Economics, Universidade Nova da Lisboa
PARTICIPATING INSTITUTIONS	Serviços Partilhados do Ministério da Saúde, E.P.E.
FUNDING	€ 123 860

This project is of particular relevance not only to the scientific community but also to policy makers, public health officials and the public in general. If successful, it might offer a very inexpensive system to improve AB prescription. Thus, and besides the normal scientific publications and communications, we will also help prepare a policy-oriented report, building on the conclusions and our experience, hoping that it can aid officials and policy makers in future projects of this nature.

This project has three main goals:

1. Develop a method to identify under and over prescription of antibiotics;
2. Identify a gold-standard for antibiotic prescription, following the WHO and local experts;
3. Propose an intervention to reduce inappropriate prescription and measure its impact.

The method will be specific to antibiotics but easy to generalize to other prescription medication.

MANAGEMENT OF NATURAL RESOURCES

IPSTERS DSAIPA/AI/0100/2018

IPSENTINEL TERRESTRIAL ENHANCED RECOGNITION SYSTEM

PRINCIPAL INVESTIGATOR	José Manuel Matos Ribeiro Fonseca
LEADING INSTITUTION	Instituto de Desenvolvimento de Novas Tecnologias (UNINOVA/FCTUNL/UNL)
PARTICIPATING INSTITUTIONS	NOVA Information Management School (UNL) and Direção-Geral do Território
FUNDING	€ 124 600

The Copernicus programme is the European Union Earth Observation (EO) programme, headed by the European Space Agency, and the developer of the Sentinels EO satellites. The IPSentinel is the Portuguese infrastructure developed by DGT and IPMA for storing and providing images of Sentinel satellites, covering the Portuguese territory and its search and rescue area. This free EO data has been used to inform environmental models, business strategies and political decisions. The main goal of this project is to explore the applications and limitations of artificial intelligence algorithms with accelerated processing hardware capabilities, as a unit of the IPSentinel for the digestion of large volumes of remotely sensed data, to produce level-3 products for land applications with the least amount of human intervention. We propose exploring two artificial intelligence approaches, one applying active learning techniques, and another based on fuzzy logic. Our research programme is based on two complementing research lines. The first line tackles the problem at the software level in two branches: 1) we propose to explore the active learning strategy for specific-class mapping in the context of remotely sensed big data to minimize the human intervention. Also, we propose to explore the use of past mapped regions and predictive models to support independent active learning process, in a "learning from the past" strategy. 2) we propose to apply fuzzy logic, coupled with multicriteria and normalisation techniques, to perform data fusion of EO images. This allows the fusion of different spectral bands of an EO image into a higher-level image that can be used to improve land cover classification. The second line tackles the problem of data processing acceleration at the hardware level. We propose to explore the use dedicated hardware (GPU and FPGA) for accelerated computing in the context of remotely sensed big data for the derivation of level-3 land cover products. The advantages of the proposed system are 3-fold: 1) as an IPSentinel unit, the learning machines have a fast and extensive access to the data catalogue. This integration allows the fast use of new data to produce updated level-3 products and ancillary information can be used to inform the learning process. 2) by using dedicated hardware and software for accelerated computing, learning tasks could be promptly performed. 3) by minimizing the human intervention in the learning process, the workload is greatly removed from the users.

WISDom DSAIPA/DS/0089/2018

WATER INTELLIGENCE SYSTEM DATA

PRINCIPAL INVESTIGATOR	Nelson Jorge Gaudêncio Carriço
LEADING INSTITUTION	Instituto Politécnico de Setúbal
PARTICIPATING INSTITUTIONS	Câmara Municipal do Barreiro, Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC ID/INESC/IST/ULisboa), Associação do Instituto Superior Técnico para a Investigação e o Desenvolvimento, Infraquinta - Empresa de Infraestruturas da Quinta do Lago, E.M. and EMAS - Empresa Municipal de Água e Saneamento, E.M.
FUNDING	€ 288 450

In Portugal, municipalities are responsible for providing water supply and sanitation services, either under direct management or indirectly through concessions. A sustainable management of these services needs the collection and the continuous updating of a broad set of data (e.g., pressure, flow, consumption) which need to be treated for generating valuable and usable information not only for the daily control, operation and management of the systems, but also for supporting the current and future planning and management of the urban water infrastructures. Depending on the degree of maturity of the water utility, these data can be collected manually by an operator in situ, stored in files or even on paper, or collected systematically and continuously, in real time, through telemetry systems. Collected data correspond to time series of pressure, water or volume records, tank levels, energy consumption and water quality parameters (e.g., chlorine, pH). Other type of available data includes infrastructure and consumer characteristics and meteorological data. The analysis of these data and their transformation into useful information for water utilities (WU) requires advanced tools. This research project aims at the development of algorithms and models that allow to extract knowledge from the data, supporting the WU in the decision-making and, thus, improving the management of its systems by reducing water losses and ensuring the provision of drinking water in quantity and quality. All algorithms and models developed are incorporated in the prototype allowing (1) to treat time series data and categorize/typify consumers, (2) to predict water consumption and quality parameters, (3) to detect and locate leaks, (4) to recognize anomalous events (e.g., illicit consumption, burst and meter deterioration). This tool allows to reduce water losses, which has a positive impact on environment, improving the use of water resources and reducing associated consumed energy. It will also allow to have systems more resilient to climate changes, given the developed better and smarter predictive models, that allow the WU to prepare contingency plans in time to face then increasingly more frequent extreme events, such as droughts. The project outcomes have the potential to be replicated to other 256 water utilities existing in Portugal, mainly managed by the public administration (i.e., municipalities).

PAMWater DSAIPA/AI/0099/2019

PREDICTIVE AND ANALYTIC MODELS FOR THE OPTIMIZATION OF MULTI-MUNICIPAL WATER SYSTEMS

PRINCIPAL INVESTIGATOR	Paulo Jorge Freitas de Oliveira Novais
LEADING INSTITUTION	Universidade do Minho
PARTICIPATING INSTITUTIONS	Águas do Norte, S.A.
FUNDING	€ 235 193

Water systems' companies face many challenges such as to optimize their energetic efficiency, analyse water characteristics or even the effect that weather has in water volumes. This is the challenge being now faced by Águas do Norte, a multi-municipal water systems' company with more than 1000 infrastructures regarding sanitation and water supply recording data every day, without stopping. Literature shows that main focus is given to the optimization of treatments applied to water rather than optimizing the entire process. Our approach to these problems focuses on improving the processes rather than the treatment. Machine learning models will be able to learn and understand the temporal dependence and contextual information from historical observations being then able to generate predictions for future timesteps, allowing one to adapt and optimize a system or a behaviour in function of something that has not yet happened. This project aims to conceive and develop deep learning models to create a predictive and analytic platform for multi-municipal water systems so that they may become efficient, resilient and sustainable. Such platform, entitled as PAMWater, shall be used as a decision support system, allowing decision makers to have access to a vast set of descriptive and predictive information. It will focus on 6 specific goals: (1) model the impact of the weather into volumetric flows, which may lead to hydraulic failure, (2) model and detect abnormal and illegal discharges based on historical data from industrial wastewater characteristics, (3) model and forecast systems' inflows based on water consumption, (4) model the urban water cycle to assess performance of the multi-municipal water systems, (5) model and forecast water quality variations in rivers and (6) optimize processes to improve energetic efficiency and reduce carbon footprint. These goals are achievable due to the high volumes of data, its historic and richness. Deep learning models, namely LSTMs and GRUs, due to their characteristics and ability to handle temporal dependences, are the perfect candidates for modelling the referred goals. Therefore, as soon as PAMWater is deployed, Multi-municipal Water Systems will have the ability to know, beforehand, predicted values of several parameters and adapt its processes in order to work efficiently and consequently provide a better service to the population and to the nature.

MATISSE DSAIPA/DS/0026/2019

A MACHINE LEARNING-BASED FORECASTING SYSTEM FOR SHELLFISH SAFETY

PRINCIPAL INVESTIGATOR	Marta Isabel Belchior Lopes
LEADING INSTITUTION	NOVA.ID.FCT - Associação para a Inovação e Desenvolvimento da FCT (UNL)
PARTICIPATING INSTITUTIONS	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC ID/INESC/IST/ULisboa), Instituto de Telecomunicações, Associação do Instituto Superior Técnico para a Investigação e o Desenvolvimento and Instituto Português do Mar e da Atmosfera, I.P.
FUNDING	€ 220 094

The marine environment provides a range of ecosystem services and benefits, including the provision of protein food sources. Shellfish cultivation and harvesting from natural seed banks respond to the increasing demand for seafood products and contributes to the economic sustainability of coastal regions. However, shellfish may act as vectors of contaminants to humans, and to safeguard public health, shellfish are routinely monitored for microbial quality, metal contaminants, and marine toxins derived from a natural phenomenon named Harmful Algal Blooms (HABs). A constant and statutory monitoring is in place to ensure that safe levels are not exceeded. The presence of marine toxins is the most critical environmental factor that affects shellfisheries, leading to recurrent closures to harvesting. The present strategy, following the EU legislation, is reactive, thus able to respond only after shellfish contamination, which frequently leads to severe economic losses and disruptions to the social fabric and cultural identity of coastal towns. The MATISSE project aims at developing proactive strategies to anticipate the environmental challenges posed to the shellfish industry. Based on a large data set collected from several sources, including remote sensing products and historical data from the routine environmental survey of the shellfish producing areas, useful forecasting models will be built to guide management actions. Artificial intelligence and machine learning tools will be developed to predict shellfish contamination based on the complex, high-dimensional, time-series data provided by the different data sources. The predictive ability of the models will be assessed through validation based on the historical in-situ measurements, acquired via routine environmental surveys. The project have two main objectives: 1) supporting the shellfish production sector through the prediction of toxins and faecal contamination and anticipating changes to harvesting permissions, and 2) supporting the Public Administration by improving the characterization of the shellfish production areas and providing tools for a more adequate harvesting licensing and sustainable use of the coastline. The project will provide a functional prototype to predict the risk of shellfish contamination, which can be a powerful tool to anticipate closures, and mitigate economic losses.

GrassData DSAIPA/DS/0074/2019

DEVELOPMENT OF ALGORITHMS FOR IDENTIFICATION, MONITORING, COMPLIANCE CHECKS AND QUANTIFICATION OF CARBON SEQUESTRATION IN PASTURES

PRINCIPAL INVESTIGATOR	Ricardo Filipe de Melo Teixeira
LEADING INSTITUTION	Associação do Instituto Superior Técnico para a Investigação e o Desenvolvimento
PARTICIPATING INSTITUTIONS	Instituto de Financiamento de Agricultura e Pescas, I.P.
FUNDING	€ 239 734

Grassland farmers in Portugal are supported by a range of financial instruments. Some of these instruments carry restrictions to ensure that the land is kept in good agronomic conditions. These restrictions require a mandatory maximum coverage of grasslands with shrub encroachment and their control using no-tillage or conservation tillage methods. These practices are also important systems for soil conservation and carbon sequestration. The IFAP is responsible for collecting data from supported farmers (declarative and third-party), performing conformity assessments and carrying out the payments. The procedure for conformity assessment is time and resource intensive. Project GrassData aims to obtain monitoring and compliance algorithms to automatically assess the performance of public policies (such as avoidance of shrub encroachment and no-tillage) designed to incentivize soil protection and carbon sequestration in pastures. The algorithms developed here will use farmer microdata held by IFAP for calibration and validation. Methodologically, this project will use an innovative and recently proposed approach of jointly using remote sensing, machine learning, and soil process-based modelling (PBM). The project comprises 6 work packages (WP). In WP1 a remote sensing-based algorithm will be developed to classify land used in the study region as pasture and assign a probability to the classification. In order to do so, super-resolution methods will be applied to produce high resolution images from the combination of satellite data (high temporal resolution) and aerial pictures collected with airplanes (high spatial resolution). In WP2 an algorithm will be developed to determine whether the parcels identified as pastures comply with good agronomic conditions restrictions, i.e. if shrub encroachment is over the limit. In WP3 a different algorithm will be developed to infer whether the parcels identified as pastures were recently tilled. In WP4, for plots identified as pastures that are below the limit for shrub encroachment, and for all pasture types in the study region, the PBM model RothC will be calibrated to estimate soil organic matter dynamics and carbon sequestration. WP5 then integrates those developments into a single product that will be tested by IFAP and for a different application region, in order to test the readiness for generalization of the product. WP6 ensures the organization of the project and dissemination of results.

RELIABLE DSAIPA/DS/0111/2019

REAL TIME DASHBOARD ON BUILDINGS OCCUPANTS RISK DURING EXTREME WEATHER EVENTS

PRINCIPAL INVESTIGATOR	Carlos Augusto Santos Silva
LEADING INSTITUTION	Associação do Instituto Superior Técnico para a Investigação e o Desenvolvimento
PARTICIPATING INSTITUTIONS	Instituto Nacional de Saúde Dr. Ricardo Jorge (INSARJ), Agência Regional para o Desenvolvimento da Investigação Tecnologia e Inovação – Associação (ARDITI) and ADENE - Agência para a Energia
FUNDING	€ 239 538

Portugal is the country in Europe where the death rate in winter has a higher correlation with outdoor temperatures, although it has one of the mildest winter climates. The same happens in summer during heat waves. This is explained by: the fact that most buildings in Portugal do not comply with the minimum requirements regarding thermal comfort conditions, occupant´s average low income and high energy prices, and by the aging of the population, since elderly people have less physiological mechanisms to handle large and sudden variations of temperature. The objective of the RELIABLE project is to develop a public geographic dashboard with real time information and high spatial resolution on the buildings occupants' health risk during extreme weather events (EWE) - heat and cold waves. This dashboard will extend the current surveillance capabilities in two dimensions: 1) improving the current forecasting models for risk by integrating new sources of public data using machine learning algorithms; 2) increasing the spatial and time resolution of the warnings, if possible, to the statistical subsection level (BGRI). This will be done through the collection, aggregation, processing and modelling of four sources of data: a) The historical information regarding healthcare indicators during EWE; b) Weather forecasts with high spatial resolution, where detailed information is calculated at the BRGI scale; c) the national census data of 2011, which needs to be updated, especially regarding the age and sociodemographic indicators. This will be done by upgrading the INE data with new sources of data like digital information from municipalities, satellite imagery and crowdsource information; d) The database of more than 1 million buildings energy performance certificates that were already issued by ADENE in order to estimate the indoor thermal comfort conditions during the EWE. This is the key source of information as it allows to estimate in detail the thermal comfort experience inside the buildings. The joint processing and analysis of all these data streams will enable the development of a model that calculates a new indicator to the resolution of the statistical subsection BGRI, which correspond in urban environments to a street or in rural areas to houses/farms. The model and the data will be available in a public dashboard that can be incorporated in the websites of different public authorities and the information will be updated every 6 hours. The webpage will have an API that will enable that any person may get direct warnings regarding a specific location.

ACCESSIBILITY AND QUALITY IN PUBLIC SERVICES

DERM.AI DSAIPA/AI/0031/2018

USAGE OF ARTIFICIAL INTELLIGENCE TO POWER TELEDERMATOLOGICAL SCREENING

PRINCIPAL INVESTIGATOR	Maria João Medeiros de Vasconcelos
LEADING INSTITUTION	Associação Fraunhofer Portugal Research
PARTICIPATING INSTITUTIONS	Serviços Partilhados do Ministério da Saúde, E.P.E.
FUNDING	€ 299 156

Skin cancer corresponds annually to about one-third of all cancers detected in Portugal, affecting one in seven people throughout their lifetime. This problem is aggravated by an absolute shortage of dermatologists currently working in the National Healthcare Service (NHS) which represent only 60% of the estimated required resources for the existing traditional model. This project aims to improve the existing Teledermatology processes between Primary Care Units (PCU) and Hospital Dermatology Departments (HDD) in the NHS for skin lesions diagnosis through the usage of Artificial Intelligence (AI). The framework will change processes by assisting both: general practitioners (GPs), in PCU, through a computer vision-based mobile application integrated with the eReferral system (SIGA-VAI); and dermatologists, in HDD, through an AI-powered Risk Prioritization and Decision Support (RPDS) platform, to be included in the eReferral system. On the one hand, the computer vision-based mobile application will guide the GPs in the acquisition process of macroscopic skin lesion images via an easy and intuitive workflow specifically adapted to this referral procedure. On the other hand, machine learning and computer vision approaches will be explored to create an AI module for automatic risk categorization of referral requests. Particularly, the DERM.AI project aims to achieve results that exceed the current state-of-the-art in skin cancer screening by merging dermatological imaging analysis (e.g. extraction of significant features based on the ABCD rule) with textual data (clinical structured information) received through the eReferral process. The introduction of m-Health solutions in the PCU scenario can play an important role by: 1) ensuring that the gold-standard guidelines are being followed; 2) enabling a quick and intuitive acquisition of dermatological data; 3) automatically assessing and ensuring the collected dermatological image quality. Adding to these advantages, the proposed AI-based framework also improves the HDD by: 1) providing quality information to specialists; 2) triage support through a machine learning module that will assist dermatologists in cases prioritization; 3) having a decision support system whose performance and robustness improves continuously over time with previously validated data.

DATA2HELP DSAIPA/AI/0044/2018

DATA SCIENCE FOR OPTIMIZATION OF EMERGENCY MEDICAL SERVICES

PRINCIPAL INVESTIGATOR	Vasco Miguel Gomes Nunes Manquinho
LEADING INSTITUTION	Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC ID/INESC/IST/ULisboa)
PARTICIPATING INSTITUTIONS	Associação do Instituto Superior Técnico para a Investigação e o Desenvolvimento and Instituto Nacional de Emergência Médica, I.P.
FUNDING	€ 294 036

Emergency medical services in mainland Portugal are coordinated by Instituto Nacional de Emergência Médica (INEM). The response time to a medical emergency may determine the life or death of a person. There are two important moments in INEM's operations: (1) the amount of time to answer each 112 call, and (2) the amount of time the emergency vehicle spends from dispatch until it reaches the emergency location. The Centro de Orientação de Doentes Urgentes (CODU) is the division of INEM responsible for answering to 112 calls (dispatch center). While CODU locations are centralized in four cities, the emergency vehicles to be dispatched are distributed throughout the country. Each vehicle is assigned to a base and it is assumed its dispatch is from the base location. In order to minimize the response time, CODU dispatches the closest vehicle to the emergency situation. However, due to population movements (e.g. summer months due to tourism) in certain areas, INEM also needs to schedule the number and location of the emergency vehicles accordingly in order to optimize the response time. The goal of the Data2Help project is to provide INEM with new tools to improve its operational results through the optimization of its resource assignment, resulting in an improved and faster response to medical emergencies in Portugal. In particular, the Data2Help project is focused on: (1) forecast the expected workload of CODU, (2) optimize the schedule of CODU staff to cope with the expected demand, (3) develop predictive models for the expected demand of emergency vehicles in different geographic areas, and (4) develop software tools to optimize the number of active emergency vehicles and staff across the country at each work shift. The Data2Help project will integrate INEM's information systems with other publicly available information such as weather reports, epidemic reports (e.g. flu reports), among other. As a result, an integrated information system with historical data of medical emergencies, response times and dispatched vehicles, as well as other information on the operational response of INEM will be built. Moreover, relevant external data (weather, epidemics, demographics, wildfires, etc.) that might be correlated with the number of emergency situations in certain locations will also be available. In the Data2Help project, new models and algorithms will be developed in order to provide INEM with optimized planning schedules for both INEM staff and emergency vehicle location.

IPPU DSAIPA/AI/0087/2018

IDENTIFYING AND PREDICTING EMERGENCY ADMISSIONS

PRINCIPAL INVESTIGATOR	Maria Joana Patrício Gonçalves de Sá
LEADING INSTITUTION	Fundação Calouste Gulbenkian
PARTICIPATING INSTITUTIONS	Associação do Instituto Superior Técnico para a Investigação e o Desenvolvimento and Serviços Partilhados do Ministério da Saúde, E.P.E.
FUNDING	€ 283 104

Emergency Care Units (ECUs) are medical facilities that deal with unplanned patient turnout, for a very large range of conditions, often urgent or acute, and frequently life-threatening. Therefore, ECUs need to find a difficult balance between having enough resources (human and others) to deal with an unexpected surge in patients, while reducing wasteful practices of sustaining more resources than required. Thus, timely information regarding possible variations in patient inflow is fundamental for proper planning and quality of service. But since a broad spectrum of reasons lead people to ECUs, hospital admissions vary significantly. From acute events, to lack of alternatives, or just out of concern, different reasons have different underlying dynamics, are guided by different factors, timings, and motivations. Thus, a combination of uncertainty and large variability, makes the problem of emergency forecasting a very complex challenge, with great impact on quality of care. Here, we propose focus on top drivers of ECU seeking behavior and use a Data Science and Machine Learning (ML) approach to study variations in emergency peaks and possible factors that might predict them. We will build on previous work from all members of the research team (RT) in health-related data mining and management, social media analysis, network science, text miming, and ML, and offer a simple prediction, that can be used by decision-makers and reduce uncertainty in ECU patient inflow. This proposal is particularly timely as it brings together an important challenge in disease monitoring, and novel ways of tackling it, mostly originating from the so-called digital revolution, while being particularly attentive to risks and future impact.

EPISA DSAIPA/DS/0023/2018

ENTITY AND PROPERTY INFERENCE FOR SEMANTIC ARCHIVES

PRINCIPAL INVESTIGATOR	Maria Cristina de Carvalho Alves Ribeiro
LEADING INSTITUTION	INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência
PARTICIPATING INSTITUTIONS	Universidade de Évora and Direção-Geral do Livro, dos Arquivos e das Bibliotecas
FUNDING	€ 299 238

The National Archive of Torre do Tombo (TT) is the backbone of the Portuguese institutional memory, managed by DGLAB. The vast amounts of archival description metadata help them find and contextualize the documents. Being at the forefront of the archival world, TT designed its online description system 20 years ago, according to the standards by the International Council of Archives (ICA). Metadata in TT is mainly composed by textual descriptions of the context and contents of the documents. Meanwhile, the archival assets evolved to encompass growing amounts of born-digital information and the interoperability requirements of cultural heritage repositories grew. A new generation of description tools is needed that includes libraries, archives and museums (LAM), and is more fine grained, more flexible and specially more machine-actionable. The huge step represented by such a paradigm shift raises many issues, some of which this project is devoted to solve. The first problem is the effective migration between the ICA and the CIDOC CRM standard, requiring both the use of existing crosswalks and the inference of the new relations with semi-automated methods. The second problem is the support to description, with tools that automate part of the generation of the more complex CIDOC CRM metadata records. The third has to do with interfaces for both human users and machines, improving user access to archives and promoting interoperability with both archives and global semantic networks. The role of TT as a large archival institution (it integrates the headquarters in Lisbon and the majority of the district archives) and also as a regulator for the state, municipal and private archives, ensures the impact of the project results in case the paradigm shift becomes a rule. Furthermore, the extensive record of innovation of TT makes it a respected voice in the ongoing debate on the archival description evolution. Three main impacts are expectable from the project. The proposed change in cultural heritage metadata will give users a better knowledge of the repository and an improved tool for more precise and richer retrieval. The second impact is a stronger presence in the aggregators, mainly in Europeana, that already uses a similar description approach. The third impact is the potential to deal with metadata assets in different platforms, from Excel files to archival description systems, and thus contribute to the integration in the Digital Archive of the Public Administration of diverse administrative as well as research assets.

SNS24 Scout.AI DSAIPA/AI/0040/2019

APPLICATION OF ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE PROCESSING METHODOLOGIES IN THE TRIAGE, COUNSELLING AND REFERRAL SERVICE OF SNS24

PRINCIPAL INVESTIGATOR	Paulo Miguel Torres Duarte Quaresma
LEADING INSTITUTION	Universidade de Évora
PARTICIPATING INSTITUTIONS	Serviços Partilhados do Ministério da Saúde, E.P.E.
FUNDING	€ 239 321

The Triage, Counselling and Referral Service (TAE) is a telephone service provided by the Contact Centre of the National Health Service - SNS24. Telephone service is provided by nurses and follows pre-defined clinical algorithms. Triage is based on a specific clinical algorithm (out of a set of 59), and the choice of the most appropriate algorithm is extremely important and relevant. The final referral indicated by the algorithm may be: Self-care (AC); Observation in Primary Health Care (CSP); Observation in Hospital Emergency (UH); Transference to the National Institute of Medical Emergency (INEM); Transference to the National Poison Information Center (CIAV). The selected clinical algorithm should ensure high safety and should have high discriminatory capability. In this context, the SNS24 Scout.AI will apply Artificial Intelligence (AI) methodologies, aiming the development of decision support tools with two main objectives: 1. Support the nurse in the selection of the most appropriate clinical algorithm; 2. Provide support to Directorate-General of Health (DGS) in the optimization process of the design of clinical algorithms and their referrals. The first objective will be achieved by identifying the most appropriate algorithms for a given set of symptoms, with adjustment for age and sex. The AI methodology to be applied will be based on a classifier built on automatic learning algorithms on an anonymised data set, obtained from contacts of the SNS24 in 2017 and 2018. This data represents accumulated experience in around 2 million cases. After the creation of the prediction model, it will be implemented in the SNS24 TAE Service as decision support, indicating in real time which algorithms are most likely to be used. The second goal is to create a support tool in the process of optimising the design of clinical algorithms and their referrals. In order to achieve these objectives, Natural Language Processing (NLP) and Machine Learning (ML) techniques developed specifically for the Portuguese language will be used, as well as Knowledge Representation and Reasoning (KRR). Therefore, the SNS24 Scout. AI project will automatically analyse the information recorded within the SNS24 TAE Service, using a built-in classifier using ML techniques to help nurses select the most appropriate clinical algorithm. It will also support the optimisation of the design in the SNS24 algorithms by analysing the results of the referral of each case, improving its discriminatory capacity and clinical safety.

FrailCare.AI DSAIPA/AI/0106/2019

INTELLIGENT FRAILTY PATHWAYS FOR THE PORTUGUESE ELDERLY POPULATION AND IMPACT ANALYSIS OF THE TELECARE SERVICE SNS24 PROXIMIDADE SÉNIOR

PRINCIPAL INVESTIGATOR	Helena Cristina de Matos Canhão
LEADING INSTITUTION	Universidade Nova de Lisboa
PARTICIPATING INSTITUTIONS	Serviços Partilhados do Ministério da Saúde, E.P.E. and V4H - Associação Para a Investigação em Valor e Inovação Tecnológica em Saúde (Value4Health.CoLab)
FUNDING	€ 236 949

This proposal aims at supporting the upscale of a telecare screening and intervention Public service (SNS24) for elderly with risk of frailty, to a national level. Artificial Intelligence (AI) experts will join clinical experts in Ageing and the telecare service professionals to develop intelligent tools that aim at improving the efficiency of intervention pathways and the cost-effectiveness of the national public health telecare service for senior citizens with frailty. Telecare digital solutions can support early risk detection of frailty, as also personalized intervention and interaction with senior citizens who live at home, in order to provide effective delivery of preventive healthcare. The SNS24 (www.sns24.gov.pt) is a telephone and online service of the Portuguese National Health Service. As the Contact Centre of the National Health Service, it is an integral part of the SPMS - Shared Services of the Ministry of Health, E.P.E. It supports the citizens in advice with acute illness as well as non-emergent health complaints, and offers a set of services to solve health-related issues avoiding the need to go to primary care unit or hospital. In 2018, SNS24 started a pilot program called Senior Proximity, with collaboration of primary healthcare institutions in two different regional areas (ACES Porto Oriental and ACES Oeste Sul), to target and support elderly population with frailty. The main objectives were to follow in a telecare service, a group of elderly citizens with frailty to: prevent health occurrences; early detect needs; promote integrated care in health, social and safety dimensions; and to contribute for a healthy and active ageing. As this pilot is running with successful measures, there is the need to evaluate the impact of this telecare intervention in such a societal burden related to frail elderly. Two challenges are addressed in this proposal: to develop an optimized model for personalised telecare interventions for the Elderly Portuguese population, in order to improve the effectiveness and scalability of the Senior Proximity program in the prevention of Health occurrences; to develop an automatic method to measure the impact and cost-effectiveness of the SNS24 Senior Proximity program through the use of artificial intelligence, in order to evaluate the feasibility of scaling-up this service to a national program level.

EDUCATIONAL SUCCESS

DS4AA DSAIPA/DS/0032/2018

UNDERSTANDING THE DRIVERS OF ACADEMIC ACHIEVEMENT: EVIDENCE FOR PORTUGAL'S HIGH SCHOOL SYSTEM

PRINCIPAL INVESTIGATOR	Tiago André Gonçalves Félix de Oliveira
LEADING INSTITUTION	Universidade Nova de Lisboa
PARTICIPATING INSTITUTIONS	Direção-Geral de Estatísticas da Educação e Ciência
FUNDING	€ 157 738

The Europe 2020 Strategy aim at tackling "the problem of early school leavers by reducing the dropout rate to 10% from the current 15%, whilst increasing the share of the population aged 30-34 having completed tertiary education from 31% to at least 40% in 2020". As the European Commission recently stated, "early school leaving is an obstacle to economic growth and employment. It hampers productivity and competitiveness, and fuels poverty and social exclusion". Presently, school dropout, which is one of the most common forms of school failure, is 14% in Portugal, well above the 10.7% of the European average. This project intends to explore the antecedents of academic achievement (AA) at national scale in Portugal, using public high-school students' micro data. We will focus on factors identified that affect AA (students', parents' and schools' characteristics), providing evidence on why some students achieve better AA than others. To do so, we will rely on supervised AI methods, using the vast amount of historical data provided by DGEEC, and build a model that highlights the most important factors leading to the AA. Moreover, logistic model trees (LMTs) will be used to analyse the data. To strengthen our results, sentiment analysis techniques will also be used to take advantage of the vast amount of textual data available. By using feature extraction techniques and transfer learning algorithms, we plan to discover new variables/factors that can improve the LMTs' predictive power. The main contributions of the proposal are the following: i) be among the first initiatives to use DS methods for large-scale AA study; ii) engender AA by providing to decision-makers, schools, and teachers a better understanding of its drivers, as well as individual (student-level) prediction of AA; iii) help Portugal and other member-states to achieve the Europe 2020 goals for AA; iv) for researchers, shed some light on AA's antecedents, using state-of-the-art methods that, to the best of our knowledge, have not yet been employed in this context; v) make use of data that is already collected and available but has not yet been used; vi) develop and make available a set of predictive models that, at the beginning of each year, give a complement to each teacher regarding the predicted AA score; vii) efficiently use LMTs to solve regression and classification tasks in a context characterized by a vast amount of real-world data, thus allowing for a more informed decision making process.

ModEst DSAIPA/DS/0039/2018

STUDENT FLOW MODELLING IN THE PORTUGUESE EDUCATIONAL SYSTEM

PRINCIPAL INVESTIGATOR	Luís Miguel Parreira e Correia
LEADING INSTITUTION	FCiências.ID - Associação para a Investigação e Desenvolvimento de Ciências
PARTICIPATING INSTITUTIONS	Direção-Geral de Estatísticas de Educação e Ciência
FUNDING	€ 246 950

The Direção-Geral de Estatísticas da Educação e Ciência (DGEEC) has a vast amount of data on the 2.0 M students yearly on the Portuguese school system, from pre-scholar to doctoral programmes, with annual in/out flow of around 90 K students. Models of the student flow along the whole system can be very important to plan for the educational infrastructure resources, to predict the distribution of competences of new human resources in the labour market, and to identify problems along the educational system and subsequently propose specific actions to overcome them. The outcome of such modelling can therefore have a large impact on the Portuguese educational system and on the Portuguese society in general. The goal of ModEst project is to mine the DGEEC data isolated and augmented with contextual variables of socio-economic data, available in INE (and Pordata), and to mine them to obtain sectorial models of segments of the educational system, as well as an overall model, that can be useful for defining policies as well as to take specific corrective measures.

HEALTI.CO DSAIPA/DS/0102/2019

FACTORS TO PROMOTE HEALTHY DIALOGUE AND BEHAVIOURS IN ONLINE SCHOOL COMMUNITIES

PRINCIPAL INVESTIGATOR	Manuel Marques-Pita
LEADING INSTITUTION	COFAC - Cooperativa de Formação e Animação Cultural, CRL
PARTICIPATING INSTITUTIONS	Fundação Calouste Gulbenkian and Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa (INESC ID/INESC/IST/ULisboa)
FUNDING	€ 221 795

The increasing possibility for people to connect to each other online is creating a hyper-connected information ecosystem. This hyper connectivity brings important opportunities for people to exchange knowledge. For example, by engaging in social dialogue, through which they can revise beliefs and misconceptions in light of sound arguments and evidence. The opposite is also true, and currently prevalent. Human vulnerabilities, from those determined at the neurological level to the socio-cultural, and technological, are being exploited in an unprecedented way. This has led to what is known as the post-truth era, a time in which people choose to believe alternative facts, even when presented with information that debunks that belief. As people become exposed to more and more information, the number of micro-level individual decisions they need to make on a daily basis increases, deliberation is replaced with fast, impression-based choices. People often resort to following the crowd too. In this context, the individual choices we make concerning the engagement in more deliberative and healthy dialogue have potentially a massive impact in social collective dynamics. Therefore, one of the first places in which we should be encouraging, and engaging in healthy dialogue, is the school. Indeed, the United Nations agenda for 2030 sets several goals for sustainable development, of which a focus on education that improves on quality, inclusiveness, equality and lifelong learning is one of the top five goals set for Portugal. This research proposal is concerned with using data to better understand the underlying dynamics of social conversation online within school communities – that is, secondary school students, and their teachers. This new understanding can be used in different ways, particularly in long term educational programmes, to foster more empathetic, inclusive, civil and healthy dialogues online. We propose to 1) convert student and teacher generated utterances to a representation that captures four important features with likely impact on conversation health: key phrases, sentiment, topic and arguments; 2) Working with conversations represented this way, we will use sophisticated data/network science algorithms to identify and fully characterise the key factors that drive different modular dynamic building blocks of online conversation; 3) the newly acquired knowledge will be encoded in a sophisticated AI-based system. This system will support the public administration in the ethical design of policies and programs that create novel, contextualised, learning scenarios for young people, as well as teachers to engage in dialogue.

SECURITY AND SUPPORT FOR CITIZENS AND BUSINESSES

GADgET DSAIPA/DS/0022/2018

ONLINE GAMBLING ADDICTION DETECTION

PRINCIPAL INVESTIGATOR	Mauro Castelli
LEADING INSTITUTION	Universidade Nova de Lisboa
PARTICIPATING INSTITUTIONS	Universidade de Coimbra, FCiências.ID - Associação para a Investigação e Desenvolvimento de Ciências and Turismo de Portugal, I.P. (MEI)
FUNDING	€ 295 291

The exploration and practice of online gaming and betting in Portugal began on May 25, 2016, having been issued 11 licenses and registering about 1800 million euros of revenue at the end of 2017. According to the European Gaming and Betting Association, Europe currently represents the largest international market for online gambling. Gambling should be regarded as a leisure and entertainment activity. When the player does not respect these values, gambling can generate detrimental effects that could result in an overall deterioration of social and familial relationships. In 2012, the European Commission released a statement highlighting the need for regulatory policies to aid in the detection of pathological gambling behaviours. In Portugal, the Gambling Inspection and Regulation Service is "responsible for the control and regulation of gambling activities in casinos and bingo halls, as well as online gambling and betting." To comply with operational objectives, this authority receives, on a daily basis, all data related to online gambling activities pursued by every user on every online platform with services that are accessible to Portuguese citizens. This authority acknowledges a profound scarcity of actionable data regarding the actual scope of gambling addiction, and a consequent lack of expertise about how best to deal with this problem. The same authority observes that "the human dimension and economical and social relevance of this issue (i.e., gambling addiction) demands scientific studies." To tackle this problem, we propose a system based on a version of Recurrent Neural Networks the architecture of which will be optimized by a neuroevolution algorithm. To effectively resolve the problem under consideration, this system must be able to render efficient comparisons of time series associated with different gamblers' behaviors, in a way that also takes the temporal dimension of the problem into account. The system, therefore, must be able to: 1) identify common behavioral patterns among gamblers within in an acceptable timeframe; 2) detect actions that are representative of a risky behavior in the context of gambling; and 3) run in real-time, to allow for continuous control of gambling activities. Successful implementation of the system and its integration with the system currently in use by the gambling control authority will enable efficient modeling and detection of online user behaviors associated with gambling addiction. The social impact of the project is enormous, given its inherent capacity to reduce the social costs associated with gambling addiction.

AI-4-MUFF DSAIPA/DS/0088/2019

ARTIFICIAL INTELLIGENCE ON THE MANAGEMENT OF THE DEGREE OF READINESS IN URBAN FIREFIGHTING

PRINCIPAL INVESTIGATOR	António Carlos Bárbara Grilo
LEADING INSTITUTION	NOVA.ID.FCT - Associação para a Inovação e Desenvolvimento da FCT (UNL)
PARTICIPATING INSTITUTIONS	Instituto de Investigação e Desenvolvimento Tecnológico em Ciências da Construção, Energia, Ambiente e Sustentabilidade (ITeCons) and Autoridade Nacional de Emergência e Protecção Civil
FUNDING	€ 236 482

In Portugal, the allocation of technical, human and financial resources to the fire brigades in charge of the urban fire response, despite considering the existing risk in an urban area and the rapid response, it is also motivated by the local population's willingness to have a satisfactory response in the case of an accident, and by the dynamics and culture of each one of the communities. These reasons have compelled to many fire brigades in regions with more population. However, the National Authority for Emergency and Civil Protection (ANEPC) considers that this distribution of resources does not guarantee an efficient and equitable intervention throughout the country or a timely response to those living in areas with lower demographics. While the public funding of fire brigades is only a part of the total corporate budget, prudent management of public and private financial resources should be made. This raises the research question that this AI-4-MUFF project intends to answer: How to ensure effective preparedness to fight urban fires by making efficient use of human resources (fire-fighters), fire-fighting equipment, and financial resources? The main objective of AI-4-MUFF is to develop a decision support tool to support ANEPC decision-makers and local fire brigades to make more technically and comprehensive decisions, but also to contemplate the policy component of the decision considering the communities and municipalities contexts. This project intends to develop a System for Supporting the Strategic Decision to Combat Urban Fires supported by empirical evidence, through the application of Machine Learning techniques, and the interconnection of multi-objective optimization models and agent-based simulation environment (Agent-based Model simulation). The expected results of the AI-4-MUFF are from the scientific point of view: 1) a conceptual model for the management of urban fires based on Theory-building using Machine Learning techniques; 2) a decision support model that integrates a multi-objective optimization model and an agent-based simulation model, in order to solve the trade-offs that may emerge from the results of the optimization model and simultaneously consider the uncertainties of the phenomenon of urban fire. From the point of view of benefit to the public entities, a functional prototype of the system of Support to the Strategic Decision to the Combat of the multi-level urban Fire will be developed parametrized for the ANEPC, pilot municipalities, and corporations of firemen, validating in a real environment.

AICE DSAIPA/DS/0113/2019

DATA SCIENCE AND OVER-INDEBTEDNESS: USE OF ARTIFICIAL INTELLIGENCE ALGORITHMS IN CREDIT CONSUMPTION AND INDEBTEDNESS CONCILIATION IN PORTUGAL

PRINCIPAL INVESTIGATOR	Leonardo Vanneschi
LEADING INSTITUTION	Universidade Nova de Lisboa
PARTICIPATING INSTITUTIONS	Faculdade de Psicologia da Universidade de Lisboa and Direção-Geral do Consumidor, Ministério da Economia
FUNDING	€ 239 244

In the last decades, credit expanded in Portugal, developing different debtors' classes and social strata. Despite the decline in unemployment, many families continue to face financial difficulties, and a representative part of the Portuguese population still cannot pay their debts. This caused an increase in the family effort rate to 73% in the first semester of 2018, while in 2017 the effort rate was 70.8%. By the end of 2014, in a population of around 10 million, 4.4 million were indebted. Previous research on over-indebtedness has concentrated on individuals' socio-economic, personal, and situational circumstances. Research indicates that vulnerability for over-indebtedness is mainly determined by socio-economic factors and financially relevant life events such as job loss. Given this scenario, this project proposes the use of Machine Learning (ML) for developing descriptive and predictive models, to understand the influencing factors of over-indebtedness on Portuguese consumers. Descriptive models will be obtained using Unsupervised ML algorithms like Self Organizing Maps and Agglomerative Hierarchical Clustering and will be used for establishing consumer clusters and guidelines for over-indebtedness regulation and consumer financial empowerment. Predictive models will be obtained using Supervised ML algorithms. A rigorous comparison between the different used algorithms is crucial and will be carried on by means of Automated ML, a set of techniques for automatizing the ML process, enabling the evaluation of thousands of models with multiple combinations of parametrization, and different types of feature selection methods. A challenge of this project is generating models that, besides extremely reliable and robust, are also easily interpretable. These models should help investigate and verify the influence of psychological factors such as attitudes towards debt, time preferences, tendency to decide based on improper heuristics, financial literacy, among other factors. Interpretation of the ML models is at the basis of the development of a set of interventions to assist in the Alternative Dispute Resolution (RAL) of consumer debt. The objectives of the project are: 1) characterize and describe over-indebtedness of Portuguese consumers using unsupervised ML; 2) create reliable supervised ML models to help to predict the factors that influence over-indebtedness; 3) develop interventions to assist in the RAL of consumer debt.

MAPINTEL DSAIPA/DS/0116/2019

INTERACTIVE VISUAL ANALYTICS PLATFORM FOR COMPETITIVE INTELLIGENCE

PRINCIPAL INVESTIGATOR	Fernando José Ferreira Lucas Bação
LEADING INSTITUTION	Universidade Nova de Lisboa
PARTICIPATING INSTITUTIONS	Agência para o Investimento e Comércio Externo de Portugal, E.P.E.
FUNDING	€ 211 253

Competitive Intelligence (CI) is concerned with gathering and analyzing information on any aspect of the business environment needed to support executives in strategic planning and decision-making. AICEP's mission is to promote Portuguese exports and to secure foreign direct investment, playing a major role in the economic development and job creation in Portugal. CI is a key component of AICEP's activity, which requires keeping track of current affairs and sift through the endless flow of news. The ability to quickly identify, process and understand the impact of political, economic and social events in the world arena constitutes a fundamental requirement to accomplish AICEP's mission. In an activity in which timing and assertive responses are of the essence, having analysts sorting through the Internet to find relevant news and updates is ineffective, time-consuming and expensive. In this project we propose a neural network-based document clustering system, which is a genuine content-addressable memory system, meaning it clusters and stores text documents so they can be retrieved based on their content. Using Self-Organizing Maps (SOM's), documents are mapped as points on a semantic map, in a topological order that describes the similarity of their contents. This map is then used as a metaphor to interact with the text corpus, creating an interactive visual analytics tool, allowing for its exploration through different, but correlated views and iteratively selecting and examining documents. This means the user can browse through point-and-click, brushing and linking between different views and graphical features or by using prototype examples (e.g. "all the documents similar to this"). The possibility of coupling improved semantic properties with clustering and visual analytics will allow the creation of effective interactive visual analytics systems, improving access, while supporting serendipity in retrieval, browsing, and exploration of text documents. The interactive visual analytics system proposed here, MapIntel, will represent a significant improvement over the present situation at AICEP and a relevant scientific contribution in the domain of text mining and visual analytics. The ability to cluster and make sense of very large amounts of text documents will not only improve the quality of the strategic planning and decision-making process, but it will also optimize resources by freeing analysts to focus on analysis and adequately briefing policymakers.

DEMONSTRADOR

SMART SUPPORT TO ENTREPRENEURS (CHATBOTS)

PRINCIPAL INVESTIGATOR	Hugo Gonçalo Oliveira
LEADING INSTITUTION	Universidade de Coimbra
PARTICIPATING INSTITUTIONS	Agência para a Modernização Administrativa, I.P. and Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento (INESC-ID)
FUNDING	€ 106 568

In order to increase the efficiency of a service, such as the Entrepreneur's Desk (in Portuguese, Balcão do Empreendedor, BDE), it is important to decrease the number of questions made directly to human employees. A solution to minimize this problem is to develop an artificial agent, available 24 hours a day, that can help entrepreneurs. Such agent must have a communication interface, through which the entrepreneur can ask questions, using the Portuguese language, and see their doubts answered automatically. To answer the questions, the agent must use the contents of the BDE or redirect the user to a BDE area where the doubts can be clarified.

This project thus aims to model intelligent agents, capable of helping entrepreneurs by providing automatic responses, in Portuguese, to clarify their doubts related to the BDE and associated services. Agents may use different data sources and different approaches, with different levels of complexity, and will be integrated in a platform that will manage the available agents and try to make the best out of each one to arrive at the best possible answers.

Moreover, the following objectives are also included: (1) Given the dispersion and variety of data, carry out a survey of the BDE data, respective formats, "clean" and restructure them, not only to be used by the different agents, but also having in mind their public availability for exploitation in future research; (2) Study the best Natural Language Processing techniques for computing the semantic similarity of texts written in Portuguese and their application in question-answering agents.

IA.SAE DEMONSTRADOR

ARTIFICIAL INTELLIGENCE IN FOOD AND ECONOMIC SECURITY

PRINCIPAL INVESTIGATOR	Luís Paulo Reis
LEADING INSTITUTION	Faculdade de Engenharia da Universidade do Porto
PARTICIPATING INSTITUTIONS	Autoridade de Segurança Alimentar e Económica
FUNDING	€ 125 000

This project aims to develop a fully functional system for complaints analysis and classification, and selection of economic agents to be inspected and the corresponding inspection routes, for the Portuguese Economic and Food Safety Authority (ASAE). The latest techniques of Artificial Intelligence (AI) and Machine Learning (ML) will be used, in order to develop this system, together with all information available at the ASAE databases and information gathered from external sources. The project intends to use all the information available in the ASAE databases and also external sources in order to improve prevention in the areas of food security and economic surveillance. The risk analysis, complaint analysis, inspection and route generation models to be created include the simulation of the behaviour of diversified economic agents and the use of information fusion and extraction techniques, georeferencing methods, machine learning, optimization, data and text mining, natural language processing, among others.

The project will analyse in an intelligent and semi-automatic way the complaints and denouncements, received electronically, and carry out their separation and prioritization by using natural language processing and machine learning techniques. Global risk matrices based on the volume of consumption, default rate, product/service, and other factors will also be analyzed and optimized (using machine learning). Through the fusion of information related to each economic agent and the georeferencing of all Portuguese Economical agents, their history and information from multiple sources, an intelligent selection will be carried out and a system for the generation of flexible georeferenced inspection plans will be developed. In order to allow full control and effective use of the system, a configuration module and intelligent information display will also be developed. The system will have feedback, with information/corrections from operators and specialists and the results of each inspection, in order to allow its continuous learning throughout its operation.

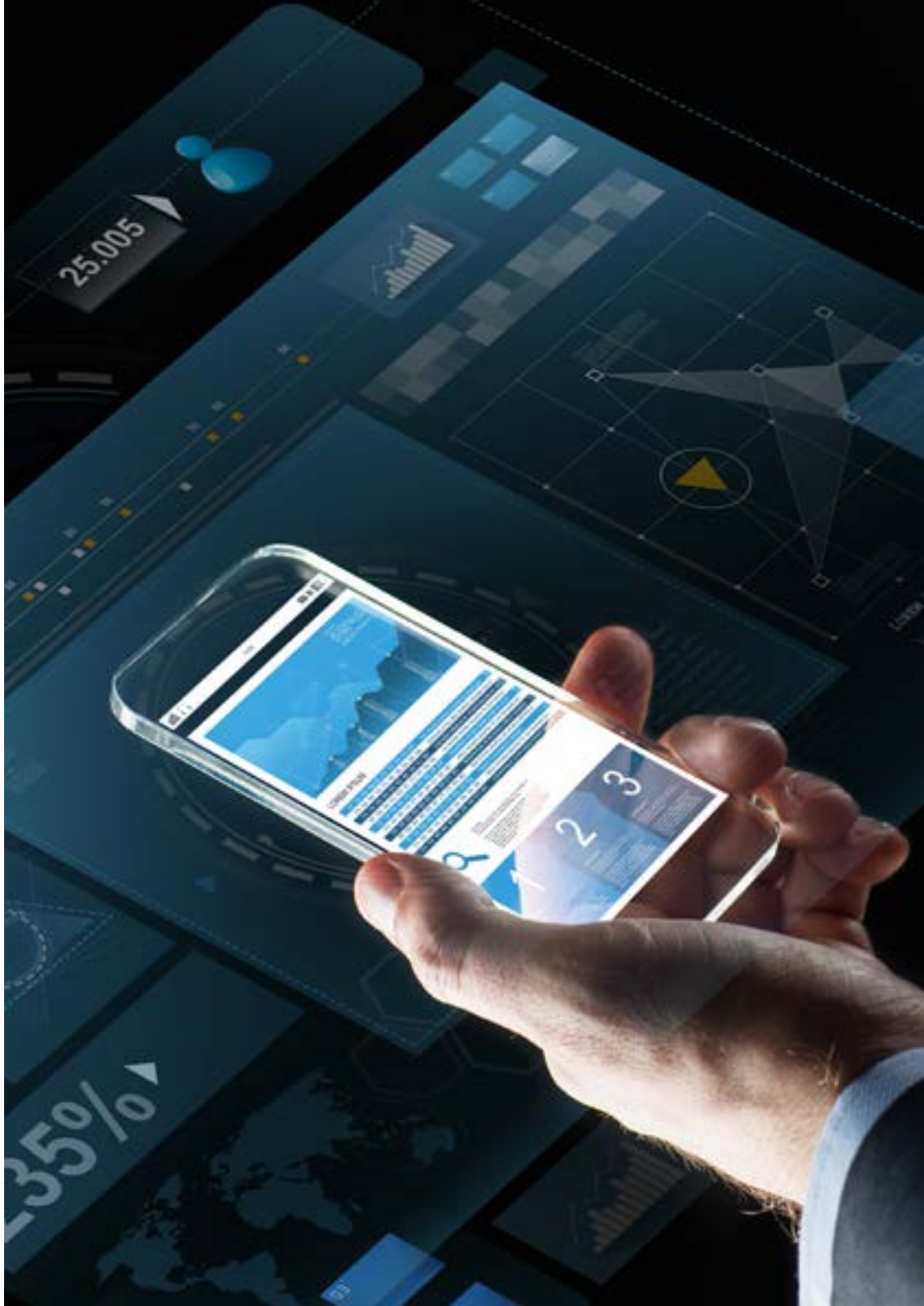
The global aim of this project, rather then increasing the number of detected administrative offenses and criminal cases, is to achieve, by a well-developed decision support process, based on AI techniques, to ASAE specialists, a real and clear awareness of economic agents that the infraction will be more easily identified, thus allowing an increase in national food and economic security.

DEMONSTRADOR

RISK ASSESSMENT OF LONG-TERM UNEMPLOYMENT

PRINCIPAL INVESTIGATOR	Leid Zejnilović
LEADING INSTITUTION	Universidade Nova de Lisboa
PARTICIPATING INSTITUTIONS	Instituto do Emprego e da Formação Profissional, I.P., Universidade de Chicago, Universidade do Porto and Rede Europeia Anti-Pobreza
FUNDING	€ 122 556

The project aims to develop (and partially implement) a roadmap for intelligent unemployment management at the Institute for Employment and Vocational Training (Instituto do Emprego e Formação Profissional – IEFP) in Portugal, with a prediction system that identifies individuals at a higher risk of becoming Long-term Unemployed (LTU) in its core. The objective of the project is to develop a system that allows: i) IEFP and its local offices to better identify individuals at a high risk of long-term unemployment, using machine learning algorithms ii) effectively distribute resources to help as many individuals who need the most support, and iii) track the impact of the support and use this information for future interventions.



PUBLISHED BY
Fundação para a Ciência e a Tecnologia
www.fct.pt

PUBLISHING DATE
2020