July 17 - August 6, 2022

WebValley 2022

Data Science and AI tools for Air Quality Monitoring

webvalley.fbk.eu

22nd edition

WebValley 2022

WebValley is the international FBK summer school for data science and AI-based interdisciplinary research. The school runs in a tech lab, set up at the Artigianelli high school in downtown Trento. The WebValley Lab provides computing resources and devices to test new ways of exploring the principles of applied data science and predictive models.

Students joining the school work in a lively and interactive environment together with a group of selected experts, also interacting through teleconference with other labs. Each year, the team includes students from Trentino, nationals and internationals. More than 400 students (17-19y old) have attended the WebValley camps since its first edition in 2001, as true protagonists of a challenging research project. Fellowships are sponsored by FBK and partner organizations, covering tuition and accommodation, as well as computing and scientific resources. Families can be asked for a contribution for sports and social activities organized on weekends.

The requirements to participate are:
- High School student (for Italy: 4th year completed)
- Good knowledge of English
- Enthusiasm in science and new technologies
- Above-average school records
- 1 Student Motivation letter
- 1 Teacher's Recommendation letter
- 1 recorded video interview with the candidate answering 5 written questions

Applicants need to demonstrate their inquisitiveness, their programming skills (if present) and their interest in STEM domains or in the topic of the year. In addition, aptitude to teamwork is crucial requirements to participate.

Candidates are scored both for background knowledge and for motivation to contribute to the project.

The 2022 Challenge

In 2022 the team of about 20 students, supported by FBK researchers and other tutors of international level will delve into a project of AI for a widely distributed air quality assessment, exploring possible relation patterns between the pollutant agents distribution detected by diverse sensors based on different technologies, in collaboration with FBK centers Sensors and Devices and Digital Health & Wellbeing (Research Unit Data Science for Health).

Air quality assessment based on IoT sensors solutions can become a key approach for smart cities development: thus, AI-powered data science is a critical tool for the comprehensive understanding of the whole IoT sensor network. Furthermore, the use of state-of-the art micro and nanoelectronics allows the development of low cost sensors, thus paving the way to a near future with diffused gas sensing platforms helping the decision-making process of the environmental agencies. In this framework, a growing number of tools are developed and tested to support diverse scenarios where a distributed air quality monitoring is needed with a particular interest in predictive systems raising red flags when signals of worsening environmental conditions are detected.

In particular, the WebValley 2022 Team will be involved in the development, implementation and validation of AI algorithms aimed at the monitoring of pollutant agents through the elaboration of datasets acquired from innovative sensors and consolidated technologies. An evaluation of the data collected by the Environmental Protection Agency on pollutants in recent years is also envisaged, to investigate the impact of the COVID-19 restrictions on air quality.

In details, the Team will delve into the computational tools needed to analyze and make sense of the data, i.e. data science and machine/deep learning solutions and high-quality software collaboratively produced by the participants, after having been provided with the essential domain knowledge and effective operative, communicative, and organizational tools.

Throughout the project evolution the participants will develop technical skills in data science, acquiring working experience on machine learning and life science methodologies, including reproducibility and interpretability for AI solutions in environmental fields, and the basics of deploying models on the cloud. Participants will acquire basic knowledge of gas sensing as well and learn how a sensor becomes a detector, experiencing the critical role of microelectronic components, whose future availability in Europe is one of the targets of the IPCEI european program.

The Format

In the first week, introductory courses in data science, visualization and AI (e.g. Python and machine learning) software are provided to the whole team, with an emphasis on the specific domain of the project (e.g. biomedicine, digital agriculture, etc.). Such initial concentration efforts provide a large spectrum of tools among which the participants can choose the most proper ones for developing the research project, including programming languages and AI frameworks such as Keras/TensorFlow and PyTorch for deep learning.

The second stage of the experience (2nd and 3rd week) outlines a learning environment which is *intentionally shaped*, where the participants have the chance to work independently on the research project, typically divided in smaller groups that are formed on the basis of the students' personal interests and the specific tasks required to tackle the challenge. The teamwork sessions will be marked by interactive experiences designed with a specific methodology that aims to develop fundamental problem solving skills while setting the goals of the challenge, and to increase the quality of the cooperation among the teams.

Lab is open all day, but group activities and leisure time are also part of the three weeks camp.

The Goals

1. Encourage smart students to be entrepreneurs in science
2. Interdisciplinarity
3. Develop teamwork, collaboration, fast-prototyping attitudes
4. Expose to challenging research themes of strong ethical interest
5. Using high quality data from scientific or statistical institutions
6. Gain experience about the hardware and data it generates
7. Promote the adoption of standard formats and share data policies
8. Deduce innovative, efficient, and effective education and communication models to be reproduced within the Italian and, potentially, the European school system

The Program

- Data Science & Tools
- Unix + GitHub
- Python intro + clinic
- Numpy & Scipy & Pandas & Pyplot
- Data Visualization
- Machine learning for the life sciences
- Basics of XAI: interpretability in artificial intelligence and data science
- Data science & privacy
- Data integration for Health
- DL theory, apps & implementations
- Widely distributed data and time series analysis and forecast
- Environmental data and pollutants
- Gas sensing technology
- Sensor calibration and gas detector integration
- Data collection and analysis in gas sensing
- Project Data
- Meetings and brainstorming sessions

SPECIAL EVENT
- Friday, Aug 5th - Final presentation of the results / of the project