LINKSFOUNDATION.COM

FONDAZIONE links
PASSION FOR INNOVATION

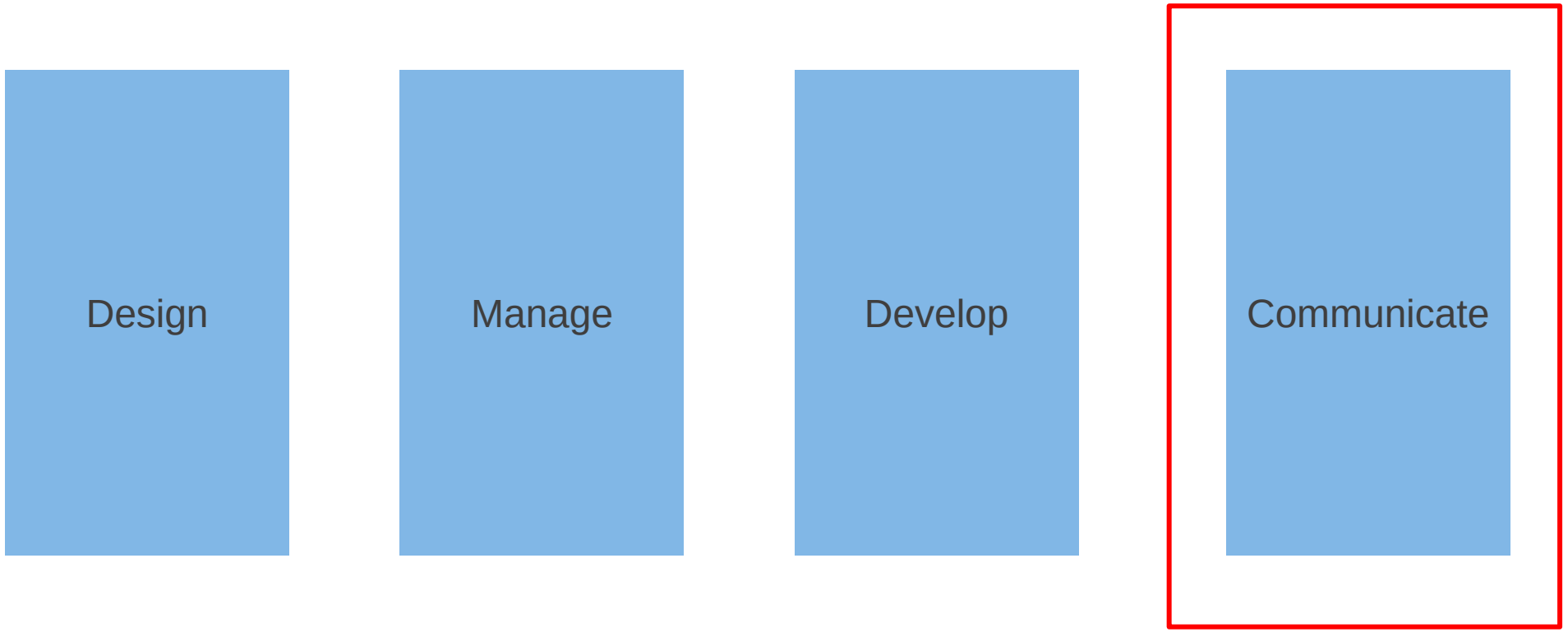
Applied Data Science Project

L15 – Report: Paper and Deliverable

Politecnico di Torino

ellis
European Laboratory for Learning and Intelligent Systems

Pillars



Report type

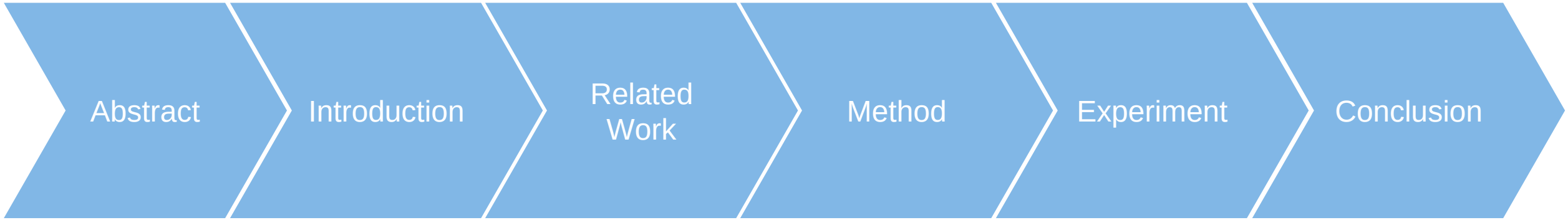
- Paper
  - concise reporting
  - usually, the reader of a paper is a researcher

- Deliverable
  - it is a longer report than a paper
  - usually, the reader of a deliverable is an innovator with knowledge about the topic of the deliverable

Report type

- Paper
  - concise reporting
  - usually, the reader of a paper is a researcher
- Deliverable
  - it is a longer report than a paper
  - usually, the reader of a deliverable is an innovator with knowledge about the topic of the deliverable

Paper outline



Abstract

The whole paper condensed in a few hundred words

1 sentence for each of the following:

- What are the objectives
- How does the work address the objectives
- What about the experiment
- What are the results
- Which conclusions can be derived

Introduction

The introduction sets the scene by:

- stating the objective(s) and the background
- introducing the methodology
- providing the remainder of the whole paper

Related Work

This section lists the relevant research activities that have been used as inspiration sources

While listing each:

- A few sentences to summarize the whole work taken as inspiration. This is expected accompanied by the references

- 1 sentence to underline one of the concepts and/or methodologies further worked out in your work

The take home message of this section is to have a fair and factual comparison with the state of the art meaning what others did before you

Method

This section contains:

- a problem statement described with words and with formulas
- a diagram to summarize the method
- a detailed description and formulas
- a complete list of technical tools and configurations

Experiment

A detailed description of the dataset
- why it is relevant for the work
- how it has been collected/found
- statistics of the dataset (no. of records, no. of features, …)

Experiment configuration
- whether there are assumptions to run the experiments
- how the experiment is performed

Evaluation
- how the evaluation is performed
- tables summarizing the performance
- take home message from the evaluation: what was successful, what not and why

Conclusion

This section

- restates the objective
- recaps the method and how it addresses the objective
- provides a short summary of the evaluation and whether it has been successful or not
- shares some lesson learned to be ported to other studies
- outlines future additions to this work

Length

Concise report

Length depends on the venue where the paper is presented:

- less than 10 pages if a conference
- from 10 to 25 pages for a journal (though there are exceptional cases of longer papers)

Hands on

UAQA: Urban Air Quality Assesment

Luca Catalano
Politecnico di Torino
Turin, Italy

Giacomo Rosso
Politecnico di Torino
Turin, Italy

Bruno Spaccavento
Politecnico di Torino
Turin, Italy

ABSTRACT

This paper extends the PRESTO architecture [13], a transformer-based model initially designed to analyze multimodal pixel time series composed from multi-source satellite imagery. The final goal is to forecast pollutant concentrations using these pixel time series which convey different types of measurements (weather, temperature, other pollutant concentrations, land characteristics, etc.). Through self-supervised learning, mimicking Masked-Language pretraining strategies, PRESTO efficiently captures spatio-temporal patterns, generating compact embedding for pixel time series. Leveraging these embedding a downstream task of pixel time series forecasting is developed. With a regression head built upon an MLP regressor, this model achieves accurate predictions. PRESTO's adaptability to multi-source and multimodal learning makes it a promising solution for efficient and versatile feature extraction on pixel time series.

ACM Reference Format:
Luca Catalano, Giacomo Rosso, and Bruno Spaccavento. 2018. UAQA: Urban Air Quality Assesment. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The utilization of machine learning alongside extensive remote sensing datasets has led to substantial societal benefits in various domains. Applications may range from monitoring progress on sustainable development goals to enhancing weather or pollutants forecasting, from bolstering disaster management capabilities to improving smart city design.

Machine learning models tailored for remote sensing applications face challenges in handling multimodal and multi-source data, particularly in scenarios where labelled datasets are scarce. The main challenges in this field are:

- **Highly multimodal and multi-source data.**
- **A highly informative temporal dimension.**
- **Missing data handling.**
- **Unavailability of labelled datasets.**

These scenarios have brought the research towards self-supervised techniques and the models that are best suited for this kind of task are considered to be transformer-like architectures[14].

This paper introduces a novel approach that combines a customized transformer architecture named PRESTO[13] for feature extraction and an MLP regressor for forecasting. The proposed method is designed to address the unique characteristics of remote sensing data, including temporal dynamics and information collected from diverse sensors. The main contribution may be summarized as follows:

- **PRESTO extension** to different data sources (multisource model) and to a specific finer time granularity (daily concerning the PRESTO monthly granularity).
- **Preprocess pipeline** to align both spatially and temporally different sources.
- **Generalization of PRESTO pretrain strategy** to encompass different sources and strategies.
- **Forecasting pipeline** using the PRESTO encoder and an MLP regressor.
- **Label weighting** to effectively train the model using both golden and synthetic labels.

Specifically, our work focused on the Urban area of Milan and we used both different satellites as time-dynamic data sources and some time-static measurements for the pretraining of PRESTO[13]. For the downstream forecasting task, certain measurements obtained from monitoring stations in Milan served as the ground truth, while additional synthetic labels were created to augment the training dataset.

2 RELATED WORKS

One key aspect in the design of machine learning algorithms for remote sensing is the consideration of the characteristics inherent to remote sensing data. These characteristics encompass the highly multimodal nature of the data, stemming from a multitude of Earth-observing satellites equipped with diverse sensors. Additionally, the temporal dimension holds paramount importance due to the dynamic nature of Earth's landscapes, necessitating approaches that emphasize pixel time series modelling.

Furthermore, the unique metadata associated with remote sensing data, specifically location data and timestamps, has proven instrumental in augmenting machine learning algorithms [15]. Chal-

Project concluded in this course at the edition of 2023-2024

App for editing

Overleaf
https://www.overleaf.com

Google Docs

Word

These 3 applications allow to edit collaboratively

Overleaf offers an editing environment optimized for writing scientific articles, just check the constraints of the free access

Consider to export your paper when complete to PDF to maximize the portability

Paper and ADSP

The whole table of contents reported in this presentation:

- Abstract
- Introduction
- Related Work
- Method
- Experiment
- Conclusion

Expected lenght 4-6 pages (up to) with references

You will receive the template

Report type

- Paper
  - concise reporting
  - usually, the reader of a paper is a researcher

- Deliverable
  - it is a longer report than a paper
  - usually, the reader of a deliverable is an innovator with knowledge about the topic of the deliverable

Deliverable outline

It is optional and it depends on the deliverable type whether it is only about the solution or if it is intended to give an overview of what already exists (output of a desktop analysis)

Revision history

It is a table, and it summarizes the relevant changes made on the deliverable

What is a relevant change worth to be reported in the Revision history table?

- table of contents
- edit of sections
- internal revision of sections
- external review of sections
- address comment from internal/external reviewers

Any change has an owner, contributor(s), and date

Executive summary

- Objective of the work
- Challenges being addressed
- Remainder of the deliverable with a list of the arguments reported in the document
- Contributions of the different partners/contributors

Introduction

The introduction sets the scene by:

- stating the objective
- defining why the presented work is relevant for a sustainable development
- introducing the methodology
- providing the remainder of the whole deliverable

Proposed solution

This is the bulk of the work

It contains:

- at least 1 functional diagram that summarizes the whole solution, and additional diagrams to focus on technical aspects of the solution
- detailed description of the solution
- screenshots of the final output either being an application with a user interface or an application with an API
- experiments and evaluation to support the statement this solution is proposed because it resulted to be effective, more useful, …
- implementation details

Background/State of the art

It is optional and it depends on the nature of the work

It is present when the work has involved a desktop analysis

It is not present if the work focused on implementing an application

If it is present, it proposes a comparison of how other projects are addressing the task or research activity outputs. This is provided along with a set of references

Conclusion

This section

- restates the objectives
- provides a summary of the whole work
- shows how the objectives have been addressed by the solution
- details future activities

Annex

In a data science project, the annex generally contains snippets of code with descriptions and tests

Length

Usually, it is around 50 pages without counting the Annex

Annex may be lengthy though it should be of a reasonable number of pages to be glanced by a reader and look for a function or an example

Hands on

DELIVERABLE

D1.2 Technological State of the Art and Mockup solutions

Project Acronym:	easyRights	
Project title:	Enabling immigrants to easily know and exercise their rights	
Grant Agreement No.	870980	
Website:	www.easyrights.eu	
Deliverable Type:	Report	
Version:	1	
Date:	31/03/2020	
Responsible Partner:	LINKS	
Contributing Partners:	21C, UTH, NTNU, IED, CAP	
Reviewers:	Angela Errore (PMO), Fareeda Akbar (BCC), Bill Mitsios (LAR), Francesco Molinari (POLIMI)	
Dissemination Level:	Public	X
	Confidential – only consortium members and European Commission Services	

D1.2 Technological State of the Art and Mockup solutions, easyRights an H2020 project

App for editing

Google Docs

Word

Both applications allow to collaborative edit the making of a presentation

Consider to export your paper when complete to PDF to maximize the portability

FONDAZIONE links

PASSION FOR INNOVATION

Thank you for your attention.

Questions?

CONTACTS

Giuseppe Rizzo

Program Manager (LINKS Foundation) and
Adjunct Professor (Politecnico di Torino)

giuseppe.rizzo@polito.it

FONDAZIONE LINKS
Via Pier Carlo Boggio 61 | 10138 Torino
P. +39 011 22 76 150
LINKSFOUNDATION.COM