Guiding Principles and Concepts for the Use of AI in DTE

AI4DTE – Digital Platform Demonstrator

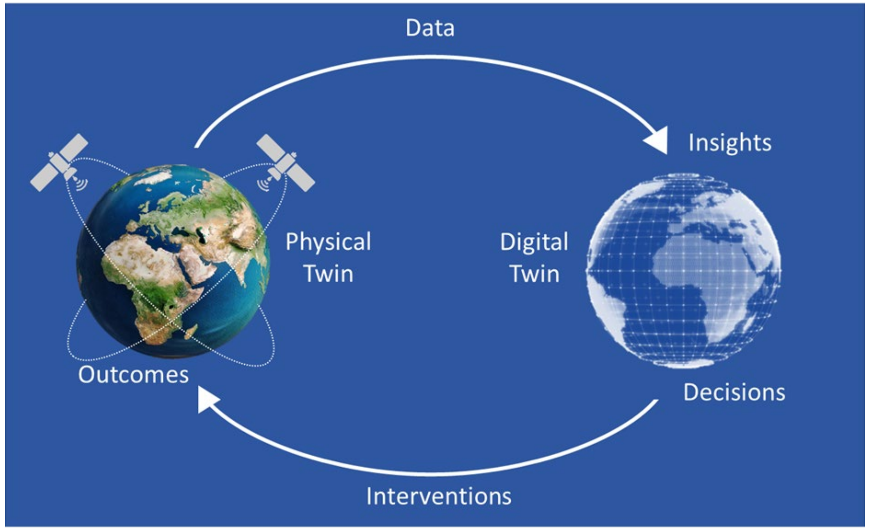
Jonny Langstone – Software Engineer
(Telespazio UK)

Téo Bloch – Data Scientist
(Telespazio UK)

Introduction

AI4DTE Project Overview

- 20-month project on behalf of the ESA Φ-lab to develop and demonstrate a prototype cloud-based platform for AI-enabled applications in the context of the DTE initiative.



AI4DTE Project Objectives

Develop Core AI Infrastructure – Create a robust, scalable AI software stack to handle various DTE domains

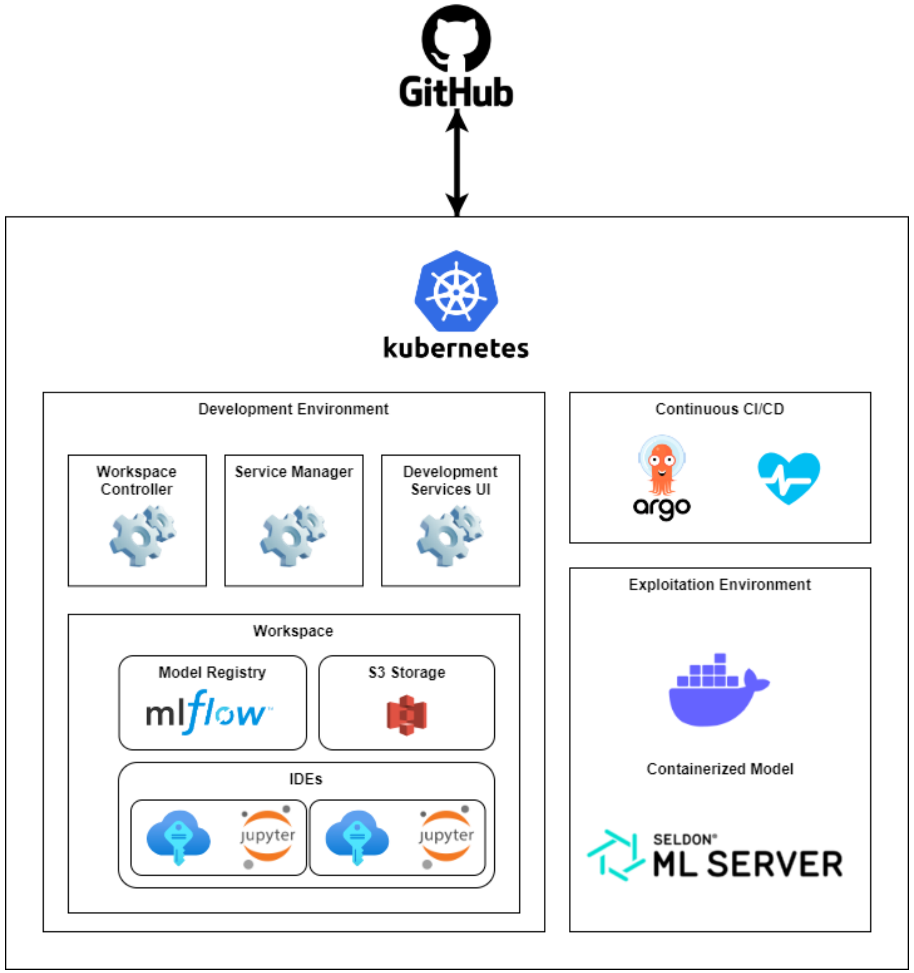
DTE Use Cases - Validate the platforms effectiveness with real-world scenarios with help from consortium partners

Refine – Understand the constraints and requirements guiding the use of AI in DTE- Identify limitations, risks and opportunities.

Roadmap - Create a strategic plan for future operationalization of the platform informed by insights gained during the project – growing set of tools and capabilities

Collaboration – Build and foster a community of AI4DTE solution providers and establish partnerships with DT user communities

Architecture Overview



GitHub Integration
Utilizes GitHub for comprehensive platform governance, managing memberships through GitHub organization and team structures, and repository management to track all development changes.

Kubernetes for Orchestration
Employed as the backbone of the platform, ensuring scalability and robust management of containerized workloads across development and production environments.

CI/CD With Argo
Automates the deployment of resources and updates the platform –Declarative GitOps approach.

Development Environment
Core elements of the software stack – reconciles with GitHub and creates resources – e.g. shared workspaces tailored to GitHub team memberships

Workspace & IDE
S3 Storage (training data), Model Registry and user specific IDE's can be spawned – Development as a Service (DaaS)

Exploitation Environment
Independent component that allows consumer users to make use of published models – Model as a Service (MaaS) – API Endpoint.

DTE Use Case Implementation – Road Condition

Summary

- The Road Pavement Condition use case is divided in two blocks:
  - The ***Current Road Pavement Condition*** model, which uses Synthetic Aperture Radar (SAR) satellite data, survey data from TRACS vehicles and rainfall records to retrieve updates on the current condition of the road pavement.
  - The ***Future Road Pavement Condition*** model, which can use the outputs from the current condition model and rainfall statistics to estimate the road pavement condition with a 1-year lead-time.

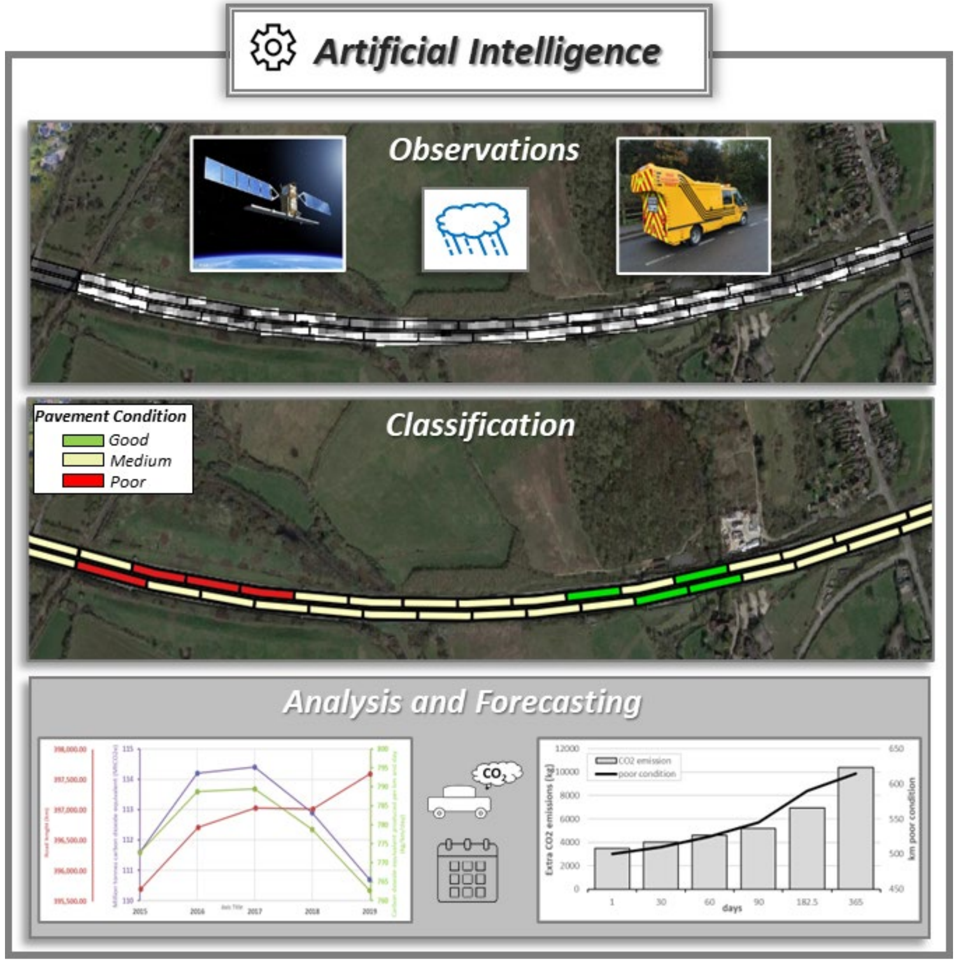
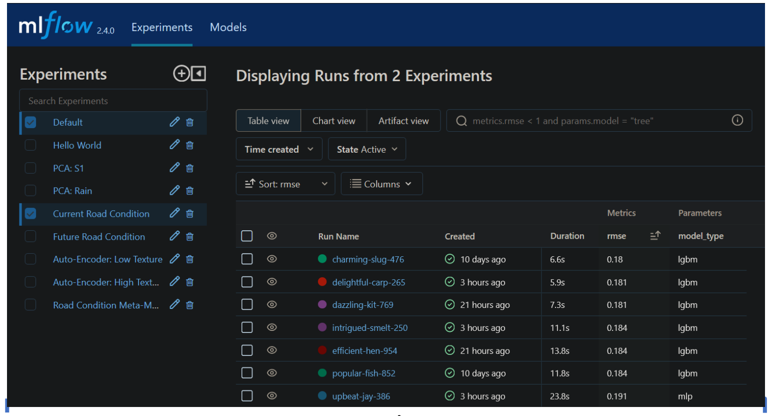


Image credit: Telespazio UK Ltd
Project: Artificial Intelligence for Digital Twin Earth (AI4DTE) Software Stack

DTE Use Case Implementation – Road Condition

Data Preparation

- **Pull and prepare data from object store.**
- **Pre-process data.**
- **Generate synthetic data for class imbalance.**



MLFlow used within the IDE/workspace to log performance and models, and create API access to models.

Current Road Condition

Train inference models

We explore both a neural network (RMSE: 0.191) and a gradient boosted tree ensemble (RMSE: 0.178).

Quantify impact of synthetic data

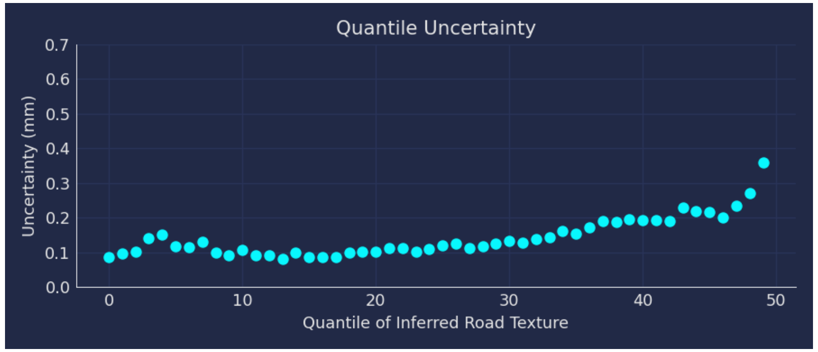
Synthetic data allowed a ~4% decrease in RMSE for the tree ensemble, and ~10% decrease in neural network RMSE.

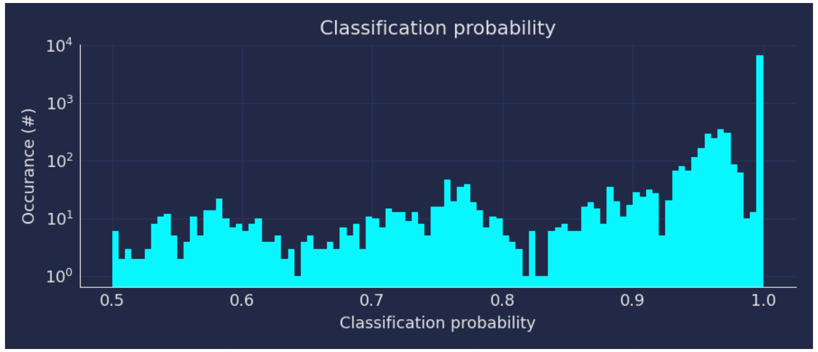
Apply best model to whole dataset

Useful for analysing performance and uncertainty calculations.

Quantify uncertainty

Parametric uncertainty estimated using distribution of ground truth in each of 50 quantiles of inferred road surface condition. Classification probability can then be derived from road condition inference and uncertainty.





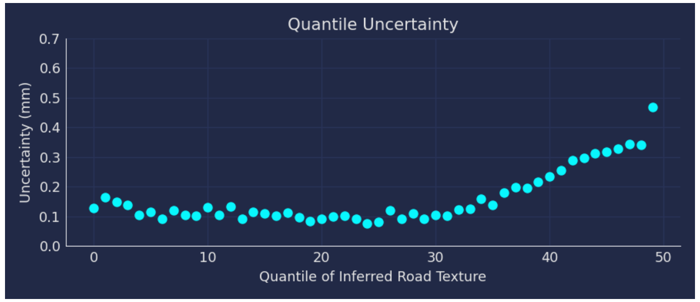
DTE Use Case Implementation – Road Condition

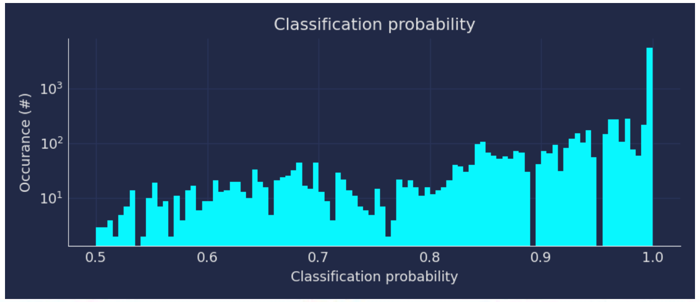
Data Preparation

- **Pull and prepare data from object store.**
- **Pre-process data.**
- **Generate synthetic data for class imbalance.**

Current Road Condition

- **Train inference models.**
- **Quantify impact of synthetic data.**
- **Apply best model to whole dataset.**
- **Quantify uncertainty.**





Future Road Condition and CO_2

Train prediction models

Neural network RMSE: 0.274.

gradient boosted trees RMSE: 0.200.

Apply best model to whole dataset

Useful for analysing performance and uncertainty calculations.

Quantify uncertainty

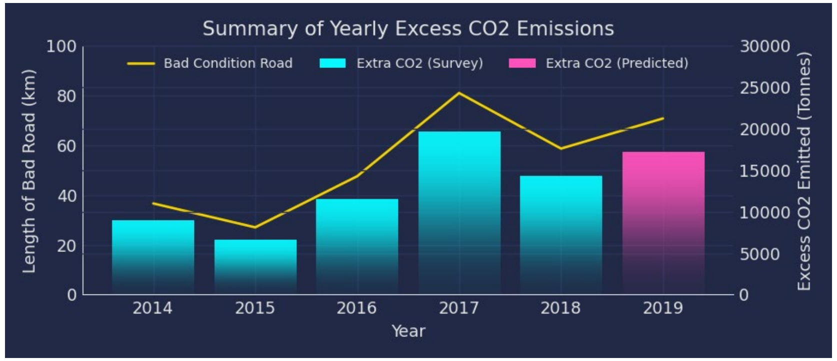
Quantile-based estimation of parametric uncertainty.

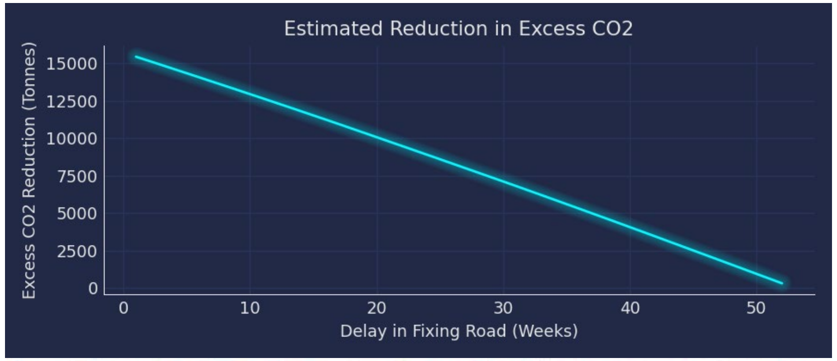
Classification probability can then be derived from road condition prediction and uncertainty.

Calculate Excess CO2 emissions

Based on academic research (Setyawan et al., 2015), bad condition roads lead to a 2.49% increase in CO_2 emissions.

CO_2 emissions per kilometre per day can be Estimated from government reports (DofT: Transport and Environment Statistics and DofT: Road Lengths in Great Britain).





DTE Use Case Implementation – Road Condition

Data Preparation

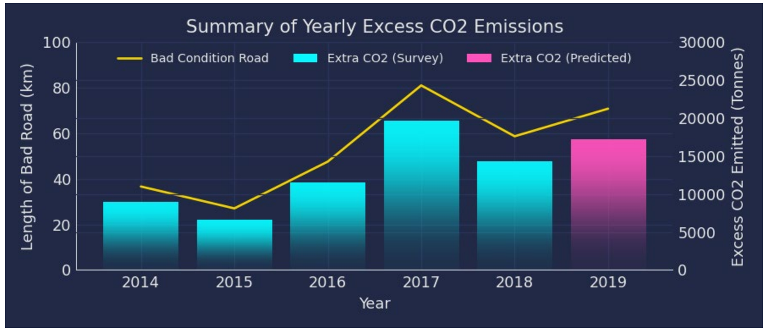
- Pull and prepare data from object store.
- Pre-process data.
- Generate synthetic data for class imbalance.

Current Road Condition

- Train inference models.
- Quantify impact of synthetic data.
- Apply best model to whole dataset.
- Quantify uncertainty.

Future Road Condition and CO_2

- Train prediction models.
- Apply best model to whole dataset.
- Quantify uncertainty.
- Calculate Excess CO2 emissions.



Decision Support App

Built and hosted within the platform

Guiding Principles and Looking Ahead

Modularity
- Decoupling – pluggable components

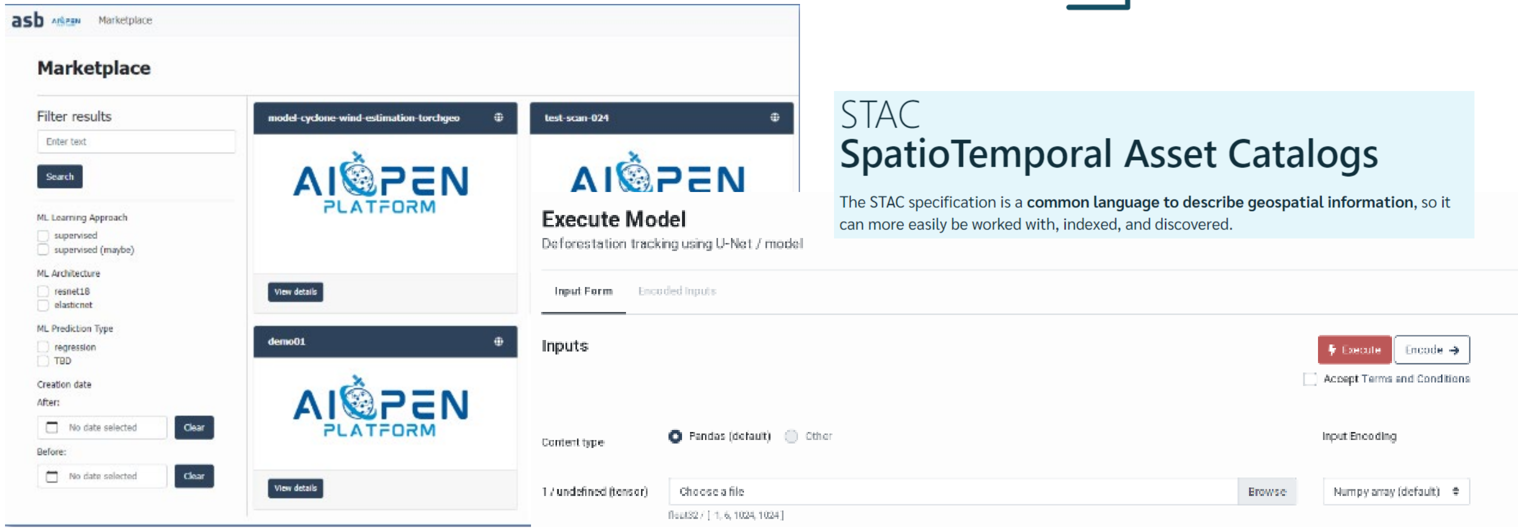
Evolvability
- Cloud agnostic infrastructure
- Additional GPU to scale workloads
- Integration initiatives – e.g. Destination Earth / Sentinel Hub API

Trustworthiness and Reliability
- Strong model and data governance to maintain system integrity and trust.
- Communicating uncertainties fosters trust and credibility in the platform.
- Presentation the models – data story telling techniques

User Enablement
- Minimize technical complexities, enabling users to focus on domain-specific tasks – Interactive UI

Discoverability
- Cataloguing of Models and Training Data - Marketplace

STAC



STAC
SpatioTemporal Asset Catalogs

The STAC specification is a **common language to describe geospatial information**, so it can more easily be worked with, indexed, and discovered.

Thank You

Jonny.Langstone@telespazio.com
Teo.bloch@telespazio.com