## Instruction & Guidelines

- You are allowed to use any library, built-in function or code base for this assignment.
- You are only allowed to use classical ML models or their ensemble, and your pipeline must not contain Deep Learning (DL) models. For example, use of the AutoGrad library from Pytorch is allowed but not a CNN/MLP. Similarly, generating code from GenAI/code-LLMs is allowed but generating directly predictions from these is not allowed.
- This is a competition, so each is competing against the others in the class. The higher your rank on the leaderboard, the higher the marks. Currently, there is no lower bar on performance, but after a week we will reassess if students are pushing the performance or not.
- You are given Train and Test splits. You are supposed to use only the Train set for model training. Groundtruth for Test split is not available. Sample submission is available in the competition link.

## Dataset Description

The dataset used for predictive modeling was generated by the Wild Blueberry Pollination Simulation Model, which is an open-source, spatially-explicit computer simulation program that enables exploration of how various factors such as plant spatial arrangement, outcrossing and self-pollination, bee species compositions and weather conditions either in isolation and combination,
affect pollination efficiency and yield of the wild blueberry agroecosystem.

The simulation model has been validated by the field observation and experimental data collected in Maine USA and Canadian Maritimes during the last 30 years and now is a useful tool for hypothesis testing and theory development for wild blueberry pollination researches.

Table 1: Dataset Features Description

| Features | Unit | Description |
|---|---|---|
| Clonesize | m2 | Average blueberry clone size |
| Honeybee | bees/m2/ | Honeybee density |
| Bumbles | min | Bumblebee density |
| Andrena | bees/m2/ | Andrena bee density |
| Osmia | min | Osmia bee density |
| MaxOfUpperTRa | bees/m2/ | Maximum air temperature (Blooming Season) |
| nge | min | Minimum air temperature (Blooming Season) |
| MinOfUpperTRan | bees/m2/ | Average of the maximum daily air temperature The |
| ge | min ℃ | highest record of the lower band daily air temperature |
| AverageOfUpperTRa | ℃ | The lowest record of the lower band daily air |
| nge | ℃ | temperature The average of the lower band daily air |
| MaxOfLowerTRange | ℃ | temperature Rainy days during the bloom season. |
| MinOfLowerTRange | ℃ | The average of raining days for entire bloom season |
| AverageOfLowerTRa | ℃ | |
| nge RainingDays | Day | |
| AverageRainingDays | Day | |

## Task

Your task is to use any kind of regression model to predict the yield of wild blueberries.