Submitted to the Syngenta AI Challenge

FOR INTERNAL USE ONLY (leave blank) Submission Number:

Team Number:

**INSTRUCTIONS TO AUTHORS**

**• Use this template in either the .doc or text with style file format.**

**• Do not remove the INTERNAL USE ONLY line above**

**• Please do not change the font size margins or headers**

**• Unless otherwise requested, do not include personal names or information in your write-up**

**• Failing to do any of the above may result in a delay in evaluating your submission**

Title of Your Submission

Corresponding Author

Address, Town, State, Zip code, Country, CorrespondingAuthor@email.com

You should replace this paragraph with your abstract. As we know, the challenge is to develop a model that predicts the 2016 yield of the seed varieties from the provided class of 2014 seeds. This model could be used to help scientists analyze large amounts of seed data more efficiently and effectively. Each submission should also provide the criteria used to select the varieties that were truly “elite,” a clear description of the methodology and theory used, results from your model that justify the selection, and appropriate references. This paper should provide estimates of Type I errors in the 2014 data and recommendations to reduce them and identify patterns in the genetic data that support how you arrived at your conclusion. In the abstract, please briefly summarize your variety recommendation and key points. Here is one possible example. “The following seed varieties are “elite” and will perform the best in farmers’ fields in 2016. Our key points would be stated here.”

*Key words:* Key words summarizing the main techniques used in your approach, for example: tabu search, stochastic optimization, neural networks

1. Introduction

In the introduction, you should include an overview of the approach. Throughout the document, appropriate citations, such as this imaginary website of the American Mathematics Institute (2005) and figures, such as Figure 1, should be used. For pagination reasons, figures and tables may appear on different pages as does Figure 1. You may also refer to additional Codalab.org pipelines and external models submitted through the Codalab.org page.

 Remember, the entries will be evaluated based on the following criteria: (1) the rigor of the solution, which will be assessed by the alignment with historically observed variety responses at evaluation (which



Figure 1: Production Possibilities Frontier.

are not part of the data distributed to researchers), (2) Effectiveness of the approach in identifying the best varieties for the test years, and (3) Transparency, clear interpretation, and self-documentation of models.

2. Criteria used to select the seed varieties

In this section, you should list the criteria you used to select the ‘truly elite’ seed varieties. As noted in the Challenge explanation, each seed variety of any plant has a unique genetic composition and must pass through a series of “stage gates” in order to be selected by scientists to breed. Each year, after the data from yield tests are analyzed, breeders decide whether to continue testing the variety or discard it. At the final stage gate is the decision to offer the seed variety to growers – the varieties that have been chosen as “elite.”

3. Estimates of Type I Errors

In this section, you should estimate Type I errors in the previous years’ seed data, explain how you arrived at this conclusion, and provide recommendations to reduce these errors. Varieties that are not successful (non-elite) after they become commercial are considered Type I errors.

4. Methodology

Discuss your methodology for identifying patterns in genetic information that predict whether a variety is “elite.” Charts, diagrams, flowcharts or other visualizations may be useful to communicate your thinking. Including discussion and considerations, such as the assumptions that were made, the scope and early considerations, may provide a useful framing of your solution. Given the multi-disciplinary aspect of the problem, background information may be useful to include or reference. It is vital that you document your methodology in sufficient detail and clarity that it can be understood and evaluated.

4. Quantitative results

In accordance with your methodology, present quantitative results that justify your seed variety selection.

5. Team members

List all the team members associated with the submission, including the Corresponding Author. This section is required if there is more than one person on a team. A team’s solution should be submitted once (as opposed to each member of the team submitting the same solution individually).

• Team Member One (aka Corresponding Author), Address of Corresponding Author, CorrespondingAuthor@gmail.com

• Team Member Two, Address of Team Member Two, TeamMemerTwo@company.com

• Team Member Three, Address of Team Member Three, TeamMemberThree@nonprofit.org

6. Exogenous data sets (optional)

Since the data sets provided in the Challenge have geographic coordinates, researchers have the option to use additional geo-referenced data sources (i.e. ISRIC, VegScape, and Drought Monitor among others). Any additional datasets used must be available for public use and properly cited. In this section, please list any exogenous data used, along with a reference.

• Exogenous Data Set One: reference or publically available website to access the data set

• Exogenous Data Set Two: reference or publically available website to access the data set

• Exogenous Data Set Three: reference or publically available website to access the data set

7. Supplementary materials (optional)

The description of the entry in the submission template should be self-explanatory. If you upload supplementary files along with your submission in Codalab, please provide a description of the files here.

• Supplementary File One Name: file one description

 • Supplementary File Two Name: file two description

Acknowledgments

Necessary acknowledgements to funding agencies or others should be made in this section. This section is optional.

References

American Mathematical Institute (2005) Better predictors of geospatial variability