Beck's Hybrids – Seed Quality and Germination

Team Members - Ropan Datta, Avery Frye , Joseph Hsin, Zarian Iqbal, Abigail Jones, Swastik Patel, Karthik Rajan, Akshath Ravikiran, Basant Sharma, Audrey Ward, Bruce Ward, Ruiming Xu

Methods (Fall 2024)

In the end, we shared our models and decided as a team on the Boosted Decision Trees

TA: Advaith Hari

Results



BECK'S

Model Predictions

This displays the predicted versus true germination rates of the saturated germination test. These specific results show the following results:

MSE: 90.7; R²-Score: 0.51

Despite being the model that fits worst, a linear trend is still visible.

Fig. 11: Map Feature for predictions across fields

Leaflet Libraries

Fig. 10: Predicted vs. Actual Germination Rates

and built weather-based models for each.

along with a 95% confidence interval

Germination due to underlying variance

Web Application

Model Scores

The interactive interface is used to make these predictive models easily user-friendly.

- A key feature is the mapping tool
- Data is sorted by location
- Each marker is a predicted germination
- score

Conclusion and Future Work

In Review

- Explored crop seed germination processes
- Developed several easy-to-use models that would give accurate predictions
- · Integrated this tool with a user-friendly interactive webpage

Future Goals

- Implementing real-time dashboards based on current weather conditions
- Collecting an increased amount of field data to generate additional insights
- · Make improvements to and ultimately finalize our model

Acknowledgments

- Thank you to our mentors at Beck's Hybrids: Will Hirschfeld, Sean Hostetler, and Emily Tuttle
- Thank you to our mentors on The Data Mine staff: Ashley Arroyo, Cai Chen, and Mia Sartain

References

https://www.rcc-acis.org/ scikit-learn Libraries

The Data Mine Corporate Partners Symposium 2025



Beck's Hybrids is the third largest seed company in the US, dedicated to delivering the best hybrid seeds to Midwest farmers.

Goal/Motivation

Predict germination rates and factors affecting germination Streamline prediction process with visuals using machine learning

Problem

- Beck's faces challenges monitoring and predicting corn germination rates
- Need to identify key factors to improve seed qualit and germination rates



- · Provide Beck's employees and farmers with an effective tool to improve crop and germination rate quality
- Utilize easy-to-understand predictive models able to accurately predict germination results

Our Approach

- Worked with planting, harvesting and lab germination test results datasets
- · Developed several predictive models using datasets, (allows options for Beck's) Random Forest Classifier: Strong validation
- Boosted Decision Tree: Ideal model outputs
- 0 Linear Regression: Easy to work with
- Created interactive tool to streamline prediction process

Methods (Fall 2024)

Random Forest Classifier

Combines predicted scores of multiple Decision Trees to minimize overfitting and ensure high accuracy.



Scatterplot for Prediction Fig. 2: Linear Regression prediction chart



This model utilizes relationships between data points to find a line of best fit to determine final . vields.

- Easy to understand
- · Helps reveal relationships and trends among variables
- Key variables: Total Gradeout Units. % Conversion, and Average Actual Harvested Moisture
- · Assessed method by comparing with models and evaluating R^2 score of 0.76

Fig 6: File upload space Fig 5: Base page for model Back End

Fig. 3: Visualization of Decision Trees correcting errors

Fig. 4: Features identified as relevant by the model

Prediction

Training Score

Fig. 8: Learning Curves for Weather Model

We built a back-end in SQL. This was used to connect the data to the front-end.

- Takes data, and stores it (and the employee's account
- Works with the front-end team to connect the code

model accordingly

0.75 -

0.70

0.65

0.55 -

This team worked to continued development with the Boosted Decision Trees model to improve its training scores to as close as possible to the validation scores.

- Improved performance of previous models
- Created and validated Weather Based Models · Integrated models between the Back End and
- Front End

model as the best machine learning method to continue work with for our end-product Methods (Spring 2025) Front End Using JavaScript in React, we built an interactive interface for the model.

Kev features include:

Boosted Decision Tree

previous trees

Many decision trees working together

to predict germination scores create

Each tree corrects errors from

Optimized Hyperparameters:

Number of Trees: 600

Maximum Depth: 6

Learning Rate: 0.05

o MSE: 2.5

o R²-Score: 0.65

 This allows for a much better fit. compared to only one tree

Maximum Number of Leaves: 50

Results (For Warm Germination):

our Boosted Decision Tree model.

- Login for secure data storage Upload files – to add prediction data Map feature - users can view
- results on an interactive map





information) in a database

Accounts for information and displays the results from the

Fig. 7: Languages used: React, Python, Postgres, Postman

Data Team