

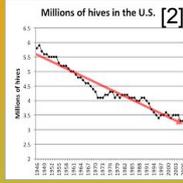
Year 5: Developing a Novel Multiple Linear Regression Model to Optimize Honey Bee Gut Immunity Using a Lactic Acid Bacteria Probiotic Mixture

Varun Madan, Lake Highland Preparatory School, Orlando, FL, USA

Background

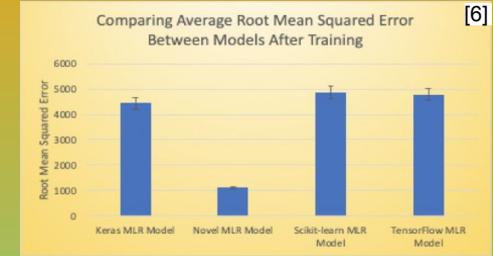
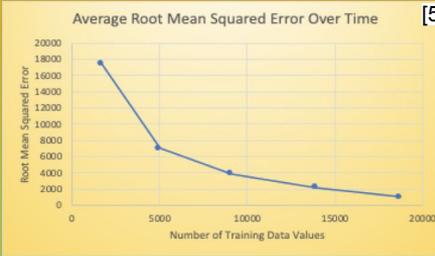
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad [1]$$

- One of the most significant contributors to the Honey Bee decline^[2] is *Nosema ceranae*,^[3] a single-celled fungal parasite that decreases the overall gut immunity.
- The first three years of the project involved successfully developing a Lactic Acid Bacteria probiotic supplement for hives.
- In Year 4, an existing multiple linear regression (MLR) model^[1] helped find ideal dosages that maximized *Nosema* reduction.
- However, the previously-used machine learning (ML) libraries only minimized the residual sum of squares when the explicit training command was given, which greatly compromised predicting accuracy.



Results

The novel MLR model's root mean squared error (RMSE) when compared with the experimental values significantly decreased over time and was significantly less than that of its ML alternatives



An unpaired t-test also revealed that the differences between the final batch of data predictions (18,000 trials) and experimental results at various input values were statistically insignificant, reaching a peak accuracy of **92.8%** on input values within 3 standard deviations of the mean.

Project Design

The purpose for this year is to create a MLR model with novel continual learning (CL) algorithms to help farmers predict the optimal amount of treatment for their hives

Use the existing bee trials from different sources to develop and train the novel multiple linear regression model



At selected doses, determine experimental *Nosema* concentration in honeybee gut



Compare model's^[4] predictions to experimental results as well as existing machine learning alternatives

After the CL algorithms were developed along with a outlier-resistant function, the model was continuously trained with over 18,000 values from external studies & previous years. The model's predictions were then statistically compared against experimental values & existing machine learning algorithms (Keras, TensorFlow, & Scikit-learn) that were trained with the same data.

Conclusions

- The results showed an extremely significant improvement in output accuracy, as the margin of error in predicted *Nosema* concentration significantly decreased throughout the entire training period.
- The model's novel continual-learning algorithms were also significantly more accurate than existing machine learning alternatives, without sacrificing valuable processing time in the process.
- This multiple linear regression model is currently being integrated into an iOS application with the intention that farmers will potentially be able to determine the ideal dosage of the bacterial treatment moments after recording their *Nosema* counts.