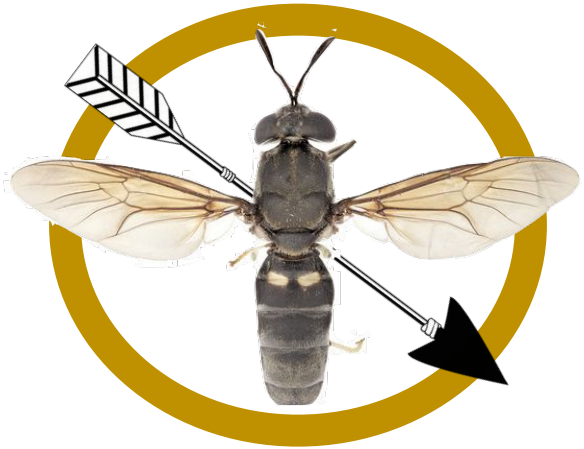
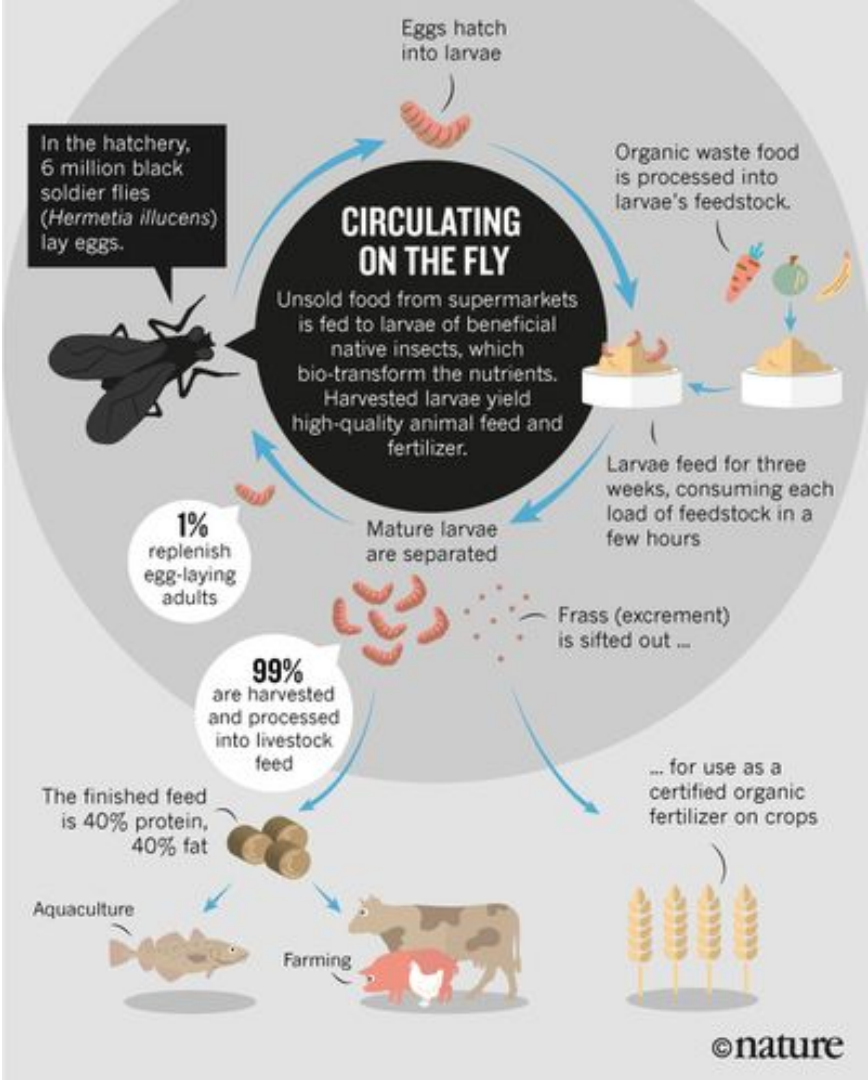
The Soldier Fly Hunger Games

Hayley & Kasia



Entomics

- Entomics are developing technology solutions for optimising waste conversion.
- They use black soldier flies (BSF) to convert organic waste into
  - Fertiliser
  - Animal feed
  - Biofuels



Project Outline

- Entomics currently inspect the reared larvae daily and use human intuition and a basic rules-based system to feed
- This project will predict how much waste to feed each tray of BSF and be able to predict as well or better than the current system

Tray lifecycle

Day 0 → ~ Day 24

Daily Inspection (e.g. 06/10/2017, Tray 331)

Temperature (Categorical)

Previous 3 feeds (Categorical)

Age (Discrete)

Moisture (Categorical)

Food left (Categorical)

Black larvae (Boolean)

Dead Larvae (Boolean)

Overpopulated (Boolean)

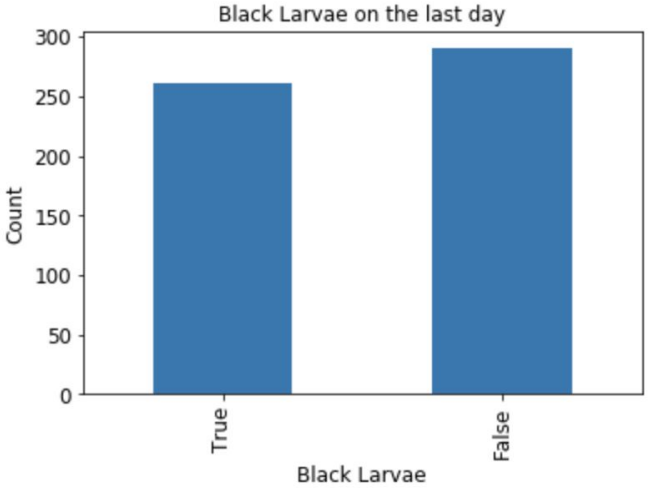
Historical data

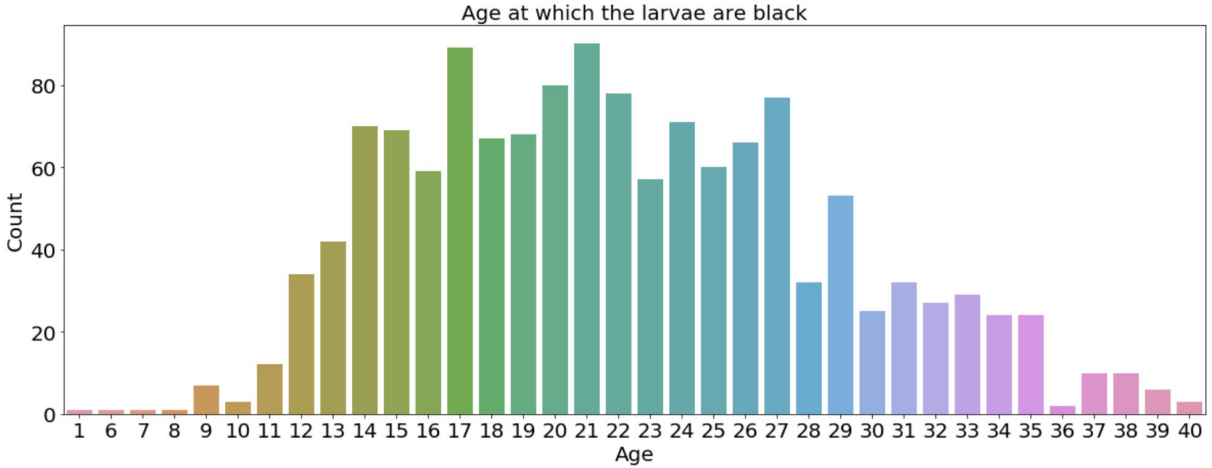
- Data collected 06/10/2017 to 31/05/2018
- 13 features
- 11,491 data points
- Trays usually harvested once larvae turn black (about to pupate)
- Fed between 0 to 7L of food daily

	Date	Tray	Age	Temperature	Moisture	Food Left	Black Larvae	Dead Larvae	Overpopulated	Previous Feed 3	Previous Feed 2	Previous Feed 1	Feed
0	06/10/2017	331	24	0	-3	None	True	False	False	0.0	3.0	3.0	3
1	06/10/2017	335	24	-1	-3	None	True	False	False	0.0	1.0	0.0	1
2	06/10/2017	338	24	-1	-3	None	True	False	False	0.0	1.0	0.0	1
3	06/10/2017	357	21	-1	-3	None	True	False	False	0.0	1.0	0.0	0
4	06/10/2017	358	21	-1	-3	None	True	False	False	0.0	1.0	0.0	0

Data Quality

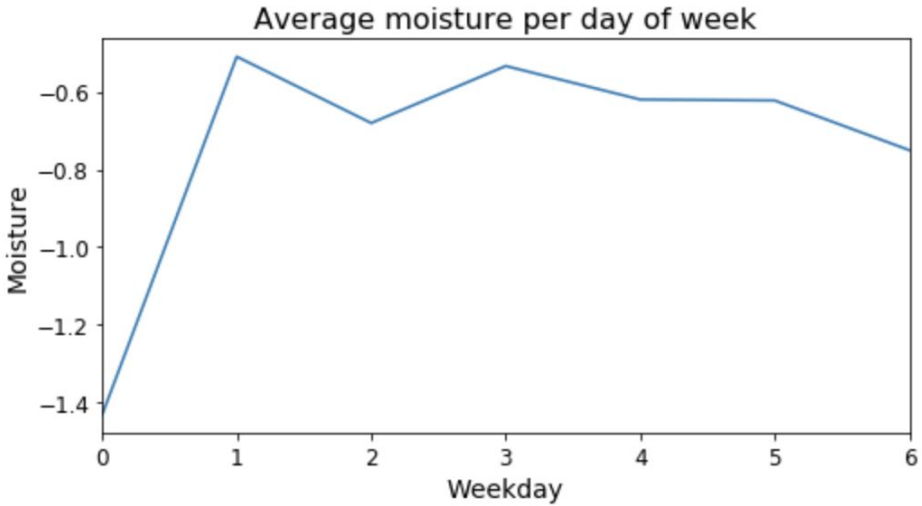
- **Dead Larvae**, **Overpopulated** and **Black Larvae** changed from True to False on consecutive days
- **Black Larvae** not always True on the final day
- Missing observations

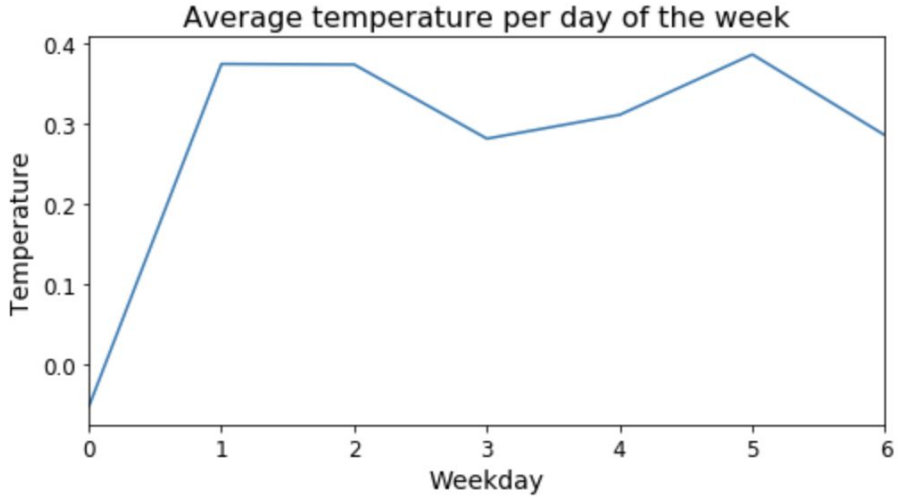




Data Quality

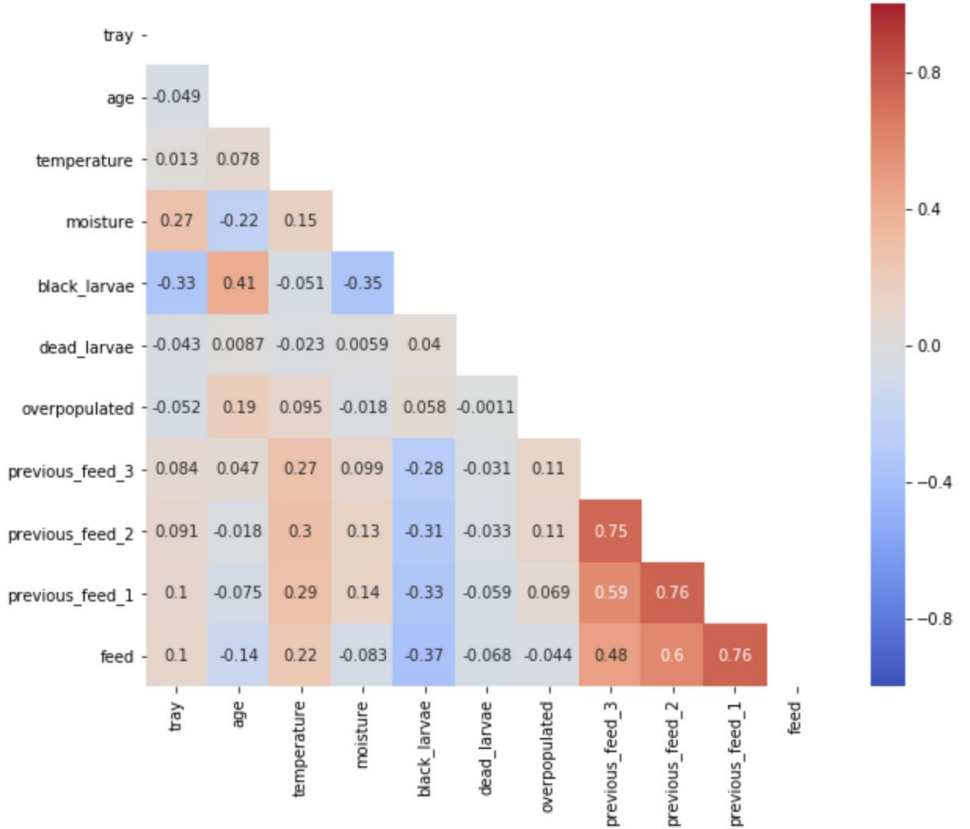
- **Previous feed 1** did not always match **Feed** variable from day before
- Hidden variation: Monday had the lowest temperature and moisture
- Duplicates





Preprocessing: Removing Previous Feed 2 & 3

- **Previous feeds** were highly correlated
- Kept the most recent feed (**Previous Feed 1**)



Preprocessing

- Train & test sets were grouped by **Tray**
- One hot encoding **Food Left**
- **Previous Feed 1**: NaN changed to 0

X_train

	tray	age	temperature	moisture	food_left	black_larvae	overpopulated	previous_feed_1	weekday
8085	828	21	1	0	0	False	False	5.0	1
2416	512	8	1	-2	0	False	False	5.0	2
7678	817	17	1	0	0	False	False	4.0	4
4398	591	12	1	-1	0	False	False	6.0	1
8716	876	15	1	0	0	False	False	3.0	4

y_train

8085	0
2416	5
7678	4
4398	6
8716	4

Models

Model description	Accuracy (%)	MAE	MSE	Mislabelled (out of a total of 3322)
Logistic regression	48.71	0.7	1.3	1704
Linear regression	N/A	0.67	0.91	N/A
Support vector machine	56.95	0.58	1.1	1430
K-nearest neighbours	58.19	0.62	1.2	1389
Decision tree classifier	73.99	0.39	0.78	864
Random forest	66.68	0.46	0.89	1107
XGBoost	65.56	0.48	0.92	1144

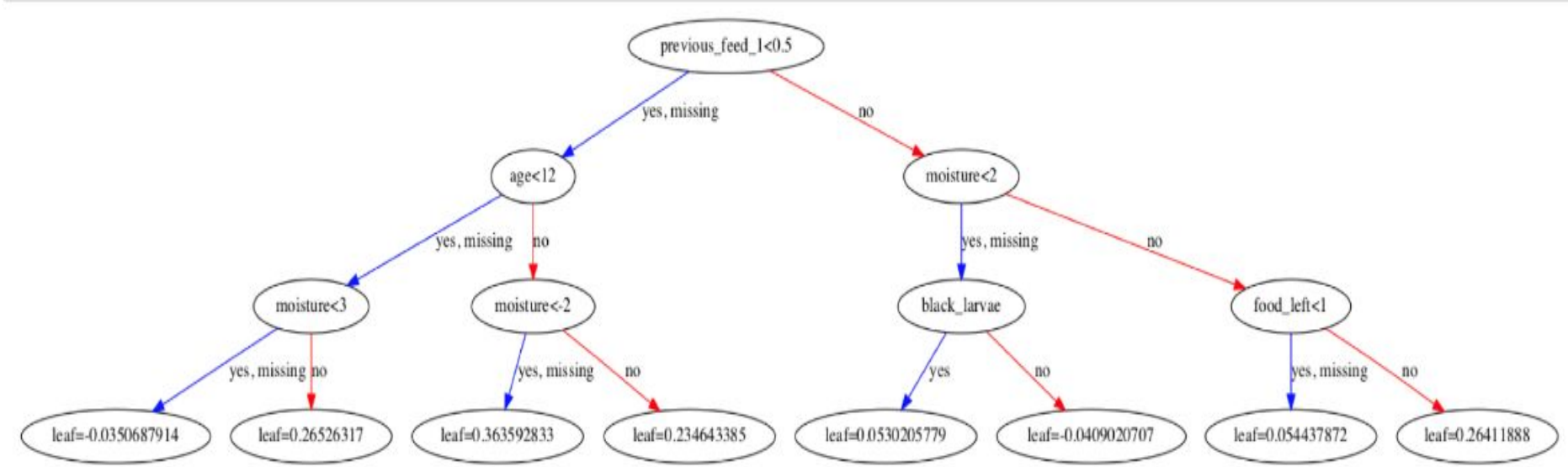
DTC: highest accuracy, but overfitting the data and trees were too complex

Best trade-off between interpretability and MAE: **XGBoost**

```
xgb_optimised = xgb.XGBClassifier(colsample_bytree=0.6, gamma=1.5, learning_rate=0.02, max_depth=4, min_child_weight=10,
       n_estimators=1000, n_jobs=-1, objective='multi:softprob',
       random_state=42, subsample=0.8)
```

Models

Sample Tree (XGBoost)

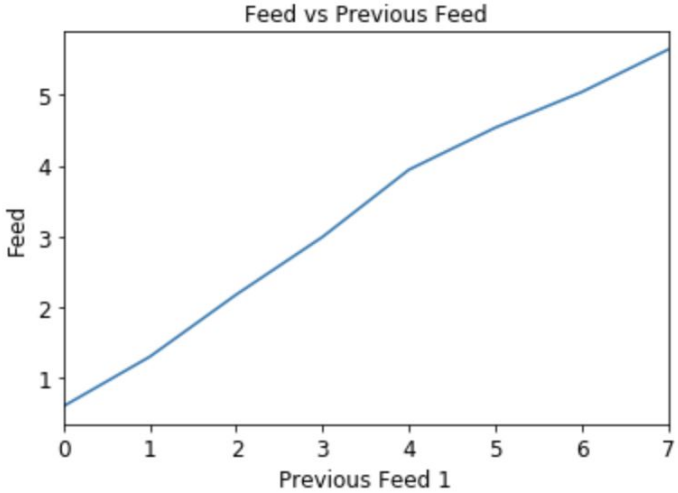


Models

Previous Feed 1 was the most significant feature

If larvae were fed more in the past, it is likely that they will continue to feed on a large amount of food in the future

Weight	Feature
0.3104	previous_feed_1
0.1720	black_larvae
0.1180	food_left
0.0951	age
0.0903	moisture
0.0663	overpopulated
0.0531	temperature
0.0500	tray
0.0447	weekday

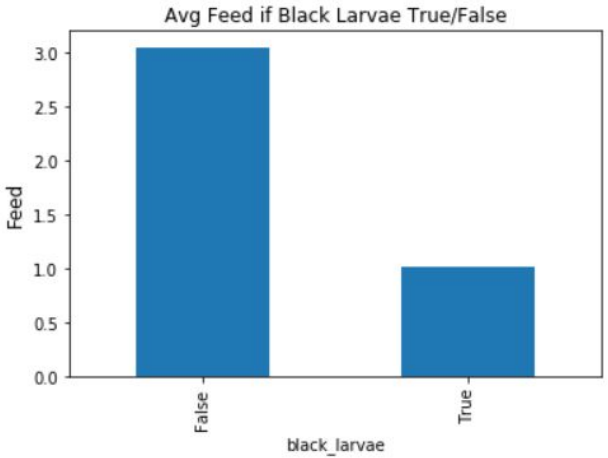


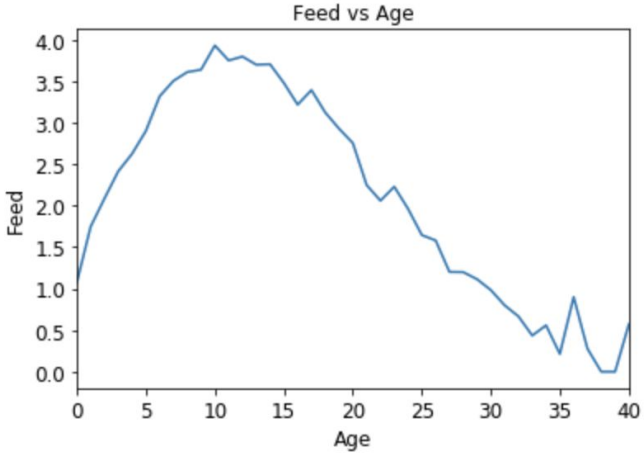
Models

The presence of **Black Larvae** and tray's **Age** were both important factors

Larvae tend to consume less food when they are very young and just before they are about to pupate

Weight	Feature
0.3104	previous_feed_1
0.1720	black_larvae
0.1180	food_left
0.0951	age
0.0903	moisture
0.0663	overpopulated
0.0531	temperature
0.0500	tray
0.0447	weekday





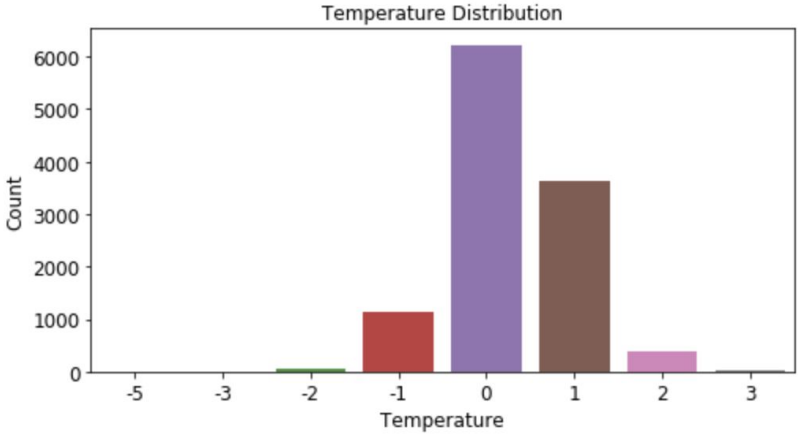
Models

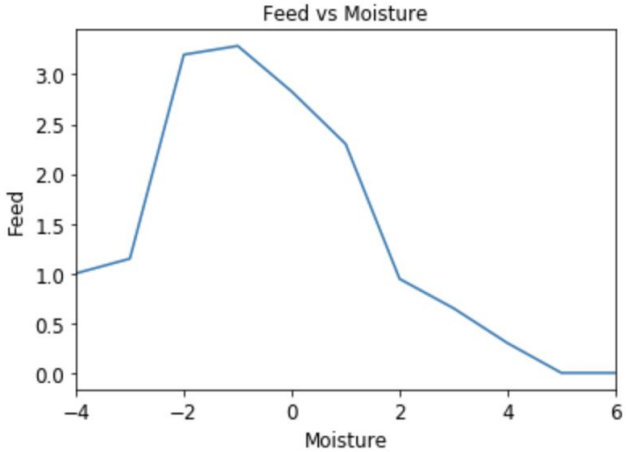
Moisture also reasonably important

Higher levels of humidity contribute to lower feed

Weight	Feature
0.3104	previous_feed_1
0.1720	black_larvae
0.1180	food_left
0.0951	age
0.0903	moisture
0.0663	overpopulated
0.0531	temperature
0.0500	tray
0.0447	weekday

Temperature - smaller effect than anticipated





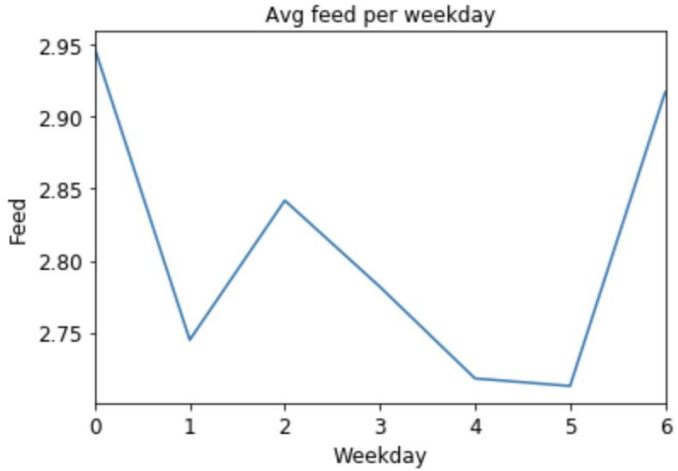
Models

Tray and **Weekday** variables not insignificant

Some of the variance in the ‘feed’ values cannot be explained by the recorded tray conditions

Hidden variables?

Weight	Feature
0.3104	previous_feed_1
0.1720	black_larvae
0.1180	food_left
0.0951	age
0.0903	moisture
0.0663	overpopulated
0.0531	temperature
0.0500	tray
0.0447	weekday



Conclusion

Further work to improve predictions:

- Further optimise the model
- Predict missing observations
- Group by similar trays

Future work

Moisture & **Temperature**
as continuous variables

Human intuition - who recorded the data each day?

More data points (e.g. intraday readings)

Which trays produced the highest quality of insects?

Thank you!