Multiple Task Learning for Cantaloupe Melon Harvest Forecasting

Daniel L. Silver
Acadia University
27 University Ave, Carnegie Hall 310
Wolfveille, Nova Scotia B4P2R6
danny.silver@acadiau.ca

Geoffrey Mason
Acadia University
27 University Ave., Carnegie Hall 310
Wolfveille, Nova Scotia B4P2R6
geoffreypi@gmail.com

Mac MacLeod
Scotia Weather Services Inc.
11 Thornhill Dr., Suite 240
Dartmouth, Nova Scotia B3B1R9
mmacleod@scotiaweather.com

David Colville
Nova Scotia Community College
Annapolis Valley Campus
Middleton, Nova Scotia B0S 1P0
david.colville@nscc.ca

Andy Vermeulen
Vermeulen Farms Limited
48 Rabbit Square Road
Canning, Nova Scotia B0P 1H0
a.vermeulen@xcountry.tv

ABSTRACT

This project takes a data mining with multiple task learning approach to develop a forecast model that predicts cantaloupe melon harvest (number of melons) each day for up to 14 days for fields located in the Annapolis Valley of Nova Scotia. The final predictive model forecasts tomorrows expected cumulative yield as well as the expected growing degree days (GDD) given the current day of the year, observed and forecasted GDD and recent harvest data. The model iteratively predicts these outcomes one day into the future for 14 days using the current weather forecast. The model's predicted GDD is used in those rare cases where the forecasted weather is not available from the weather service. The model was deployed in the summer of 2014, with a harvest forecast being mailed out twice a week to the farmer starting in late July. The results show that the model was particularly helpful in accurately forecasting yield during the start of the season, allowing the farmer greater time to plan logistics for harvest and delivery to market. During the season, the model significantly overestimated the number of melons picked because of (1) an anomaly in melon planting times and (2) marketing factors that were not factored into the model. A number of useful recommendations were developed by the team for future studies.

CCS CONCEPTS

•Computing methodologies →Supervised learning by regression; Neural networks; •Applied computing →Agriculture;

KEYWORDS

data mining, predictive modeling, harvest forecasting, multiple task learning, neural networks.

ACM Reference format:
Daniel L. Silver, Geoffrey Mason, Mac MacLeod, David Colville, and Andy Vermeulen. 2017. Multiple Task Learning for Cantaloupe Melon Harvest Forecasting. In *Proceedings of KDD 2017 Halifax conference, Halifax, Nova Scotia, Canada, August 2017 (KDD-2017),* 8 pages.
DOI: 10.475/123_4

1 INTRODUCTION

Success in the harvest of crops and their delivery to market depends on knowing when the produce is ready to be picked and the planning of human and material resources to maximize crop yield [10]. Cantaloupe melons are particularly challenging, as the harvest period is short and the required human and transportation resources demanding [9]. This requires careful planning and logistics based on the best possible prediction of crop yield. Scotia Weather Services Inc (SWS) wishes to extend their abilities in weather forecasting to provide forecasts of crop yield to the agri-food industry - from farmers to markets [2, 4, 5, 7].

This project takes a data mining approach and uses mutilple task machine learning technology to develop forecast models that predict melon harvest (number of melons) per day up to two weeks in advance. The predictive models forecast daily harvest given the current day of the year, observed and forecasted weather data and field information [1, 6, 8]. Vermeulen Farms (VF) of Canning, NS provided domain expertise and melon harvest data from 2007 to 2013. In addition, we obtained (1) observed weather data from Environment Canada as well as an existing Nova Scotia Community College (NSCC) weather station proximal to Vermeulen Farms, and (2) high quality weather forecast data from the Weather Network and Scotia Weather Services for two weeks in advance. The major challenge anticipated in the project was making accurate harvest predictions out over the two week period so that the advance information would be useful for farm and market planning. We also installed and tested automated weather and photography equipment during the 2014 harvest season so as to more accurately record field data.

Based on the results of this project, we anticipate collaborating with SWS to develop the requirements for a fully automated forecasting system. The long-term goal is to develop a cloud-based harvest forecasting system for a variety of crops. This project is a significant step toward SWS being able to develop a unique and innovative crop forecasting service in Canada ultimately benefiting SWS and the agricultural industry as a whole.

1.1 Objectives

The project had four objectives: (1) To determine the best way to formulate the data mining problem and the most important predictor variables. (2) To build more accurate predictive models and test them live during the 2014 growing season using a semi-automated delivery method that sends forecasts to Vermeulen Farms by email. The predictive model will forecast daily harvest up to 14 days in advance given the day of the year, observed and forecasted weather data as inputs. (3)To install and test automated weather and photography equipment in the field so as to more accurately record field data during the 2014 season. (4) To develop recommendations for future work in this applied area.

1.2 Approach

A standard data mining approach was used to develop and test the predictive models. We enriched the observed weather data we had with additional soil temperature and moisture and solar radiation data from a NSCC weather station as well as high quality daily weather forecast data from SWS. Historic melon yield data from Vermeulen Farms was also reviewed, cleaned and consolidated. This data was used to develop and test artificial neural network models that predict the crop yield for up to 14 days in advance. A confidence interval was estimated for each daily prediction.

Beginning in mid-July the best model was deployed to make forecasts for the 2014 growing season. A semi-automated approach was taken, whereby daily observed weather and new weather forecasts, obtained from the weather station and SWS, respectfully, were used to make yield predictions. These daily forecasts were delivered to Vermeulen Farms by email for their decision making. The machine learning methods we employed are tried and proven in the literature [6]. Innovation occurred in the areas of data representation, the handling of missing weather values and model development for predictions up to two weeks in advance. Following the final harvests in September, the results were analyzed and a report was prepared by the team.

Raw weather data was collected from the weather station and prepared for model development. Because of missing data in the historic records, work had to be done to impute values based on best estimates from nearby weather stations. Two new cameras (with temperature loggers) were set up in the melon field. This new equipment provided a means of measuring the temperature in the field and the growth of the melons over the 2014 growing season. The camera images were analyzed to monitor melon growth over the season and capture temperature changes and field activities.

Notes were made during the project and in November the team completed a requirements analysis that resulted in a series of recommendations for future work.

The remainder of this paper will follow the CRISP Data Mining methodology, reporting on the six major steps undertaken by the project: business understanding, data understanding, data preparation and analysis, model development, model evaluation, deployment, summary and recommendations.

2 BUSINESS UNDERSTANDING

The ability to accurately forecast melon harvest data is directly dependent on access to weather forecast data. Current Numerical Weather Prediction (NWP) models are able to generate weather forecasts up to 14 days into the future. This was determined to be the practical limit for our melon forecasts. Vermeulen Farms felt that this would be more than enough to meet their farm planning needs, although their markets would prefer even more advanced forecasts.

The energy provided by the sun to plants is well understood to be the dominant factor in their growth and production of fruit. Specifically, the growing degree days (GDDs) metric is considered the best indicator of plant growth in the agricultural field. GDD is defined as [(Maximum temperature of day) + (Minimum temperature of day)]/2 - (Base temperature); where negative values are treated as zero, and Base temperature for cantaloupes is 10 °C (varies for each crop). Most crops become ripe for harvesting once the GDDs accumulated from the time of planting exceeds a certain value. Other factors that we assumed would play a major role based on our literature survey were solar radiation, total precipitation and relative humidity.

We came to realize that observed and forecasted weather data, and harvest data may be unavailable at times. This meant that our system would have to be robust enough to handle missing data. Our solution was to create a multiple task predictive model [3] that estimated all required input values for the next melon forecast interval. These estimates could be used to replace any missing input values required by the model. More specifically, as shown in Figure 3, we would develop a model that accepted the previous 14 days weather information and melon harvest, and predicted the current days weather and melon harvest. This is referred to as an auto-regressive model. By using this model iteratively, we could predict as far out as desired, interchangeably using our own previously predicted values of weather or melon yield as replacements, if the actual data was unavailable.

3 DATA UNDERSTANDING

The harvest information obtained from Vermeulen Farms was from 3 fields making up a total of about 30 acres. These fields produce a large quantity of the cantaloupe melons sold in Nova Scotia. The data contained information about the number of melons harvested at each size, however an analysis of the records indicate a number of uncertainties in the breakdown of the total melon count. Vermeulen Farms recommended that for this initial project that we focus on predicting the total number of melons independent of size.

It was our understanding going into the project that the count of melons harvested was driven largely by the plant yield in the field, and that the major problem for the project team was predicting this yield. This is a factor of considerable importance that will be revisited later.

4 DATA PREPARATION AND ANALYSIS

The data used in the project can be broadly divided into melon harvest data and weather data. The initial preparation and analysis of these data sets was done independently, before combining them.

Figure 1: Image of one of the cantaloupe melon fields studied.

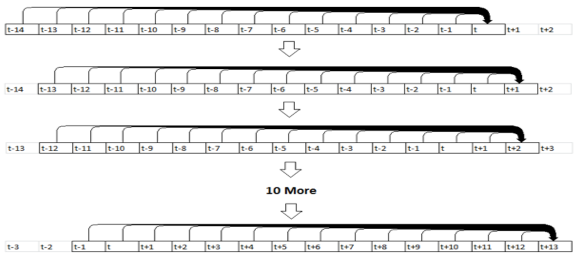

Figure 2: Characterization of the auto-regressive approach taken to predict weather and melon harvest values.

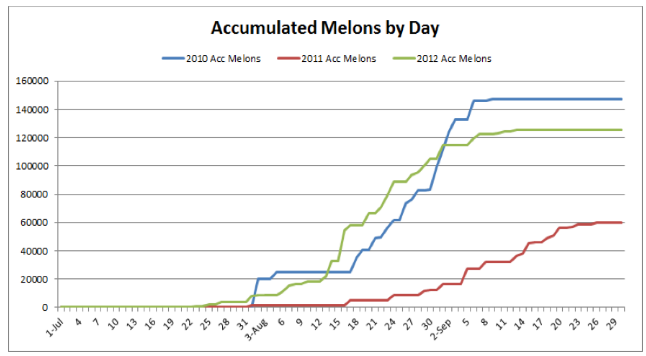

Figure 3: Accumulated melons for 2010, 2011, and 2012 from July 1 to September 30.

4.1 Melon Harvest Data

The harvest records for the melons spanned from 2007 2013. The data for all years were provided on hand-written sheets of paper, except for 2013 which was transferred digitally as a spreadsheet. All harvest data was entered and combined into a single MS Excel spreadsheet file. An initial Meta-Data Report (MDR) was produced following emails and phone calls between the Acadia team and Vermeulen Farms during which the harvest data was cleaned and made complete as possible.

It was decided that the data from 2007 - 2009 was not sufficiently complete to be used for model development. It was also decided that 2013 was an anomalous year of extremely poor yield that should not be used. This analysis confirmed our decision to not break predictions down by melon size. A final MDR was produced after this data was combined with the weather data.

The melon count data is not normally captured daily by VF. Several days of harvest may pass before the total count is taken. For this reason we moved to considering the target variable as being the accumulated melon harvest for each day of the season. The melon harvest for each day was accumulated as a running total, which equaled the total melons picked to date. This value is monotonically increasing and serves to decrease the noise in the data. The difference between the predictions of any two consecutive days can then be calculated as the increase in melons to be harvested. Figure 4 presents the three years of harvest data, 2010-2012, used to develop and test the forecast models. It shows a wide variation in the harvest, with 2010 being an excellent year (147,000 melons) and 2011 being a poor year (60,000 melons).

4.2 Weather Data

There were two sets of observed weather data; one was obtained from the publicly available Environment Canada records, and the other one was provided by NSCC from their Station 60 weather station.

The Environment Canada (EC) data was retrieved from their website and inspected for missing data and errors. Any problems with the data were fixed using data the Acadia team had previously acquired from the Kentville Agricultural Centre for other projects. A total of 1096 records spanning 2010-2012 were used in the analysis to prepare for building the statistical models (see Appendix A).

The NSCC data was received, and was inspected for missing data and errors. An initial MDR showed that there was enough missing NSCC weather data to warrant using the more complete EC data, when needed, for the construction and testing of the models.

4.3 Data Analysis

The weather data that was available with high quality included: average air temperature, maximum air temperature, minimum air temperature, heating degree days, cooling degree days, dew point temperature, relative humidity, wind direction, wind speed, and total precipitation.

Our initial analysis of the data focused on understanding which variables were most important to predicting the accumulated melon harvested per day with a focus on GDD, solar radiation, precipitation, and humidity.

Solar radiation appeared to be correlated, but the solar radiation data was missing for 2010 and would not be available as an input during deployment, so it was not used for building the predictive models. It was suspected that soil moisture and soil temperature would correlated strongly with the melon harvest as well, but that soil data was not available with sufficient quantity or quality to determine is this was true. Precipitation and humidity showed a good correlation with melon count, but GDD showed the strongest relationship.

In the end, it was recommended that the predictive models should use the previous 14 days weather data and melon counts, along with the forecast for the current days weather. Further details will be discussed in the next section.

5 MODEL DEVELOPMENT

An auto-regressive method was used to develop models to predict the accumulated melons. To predict the number of melons on day t, the autoregressive model accepted the day of the year, the previous 14 days (days t-1 to t-14) of weather data and melon data, and the weather forecast data for day t. To produce a 14 day forecast (days t to t+13), the predicted melon harvest for day t was used as an input to predict day t+1s melon harvest, along with the weather forecast data for day t+1. This process was repeated until all 14 days into the future had a prediction. In this way, any source of weather containing the appropriate data could be used to produce a 14 day melon harvest forecast.

A multiple task learning neural network approach [3] would also predict the weather data of day t. So, even in the case where no weather data (observed or forecasted) was available, and no melon harvest data was available, the model could use the last recorded values to predict the next days values. Note that this technique has the advantage of being able to predict beyond the 14 days of weather forecast, but with an increasingly lower confidence.

All of the models built were multiple task learning back-propagation neural networks. A standard approach was taken to determine the best network architecture and configuration and learning parameters. In the end, we settled on networks with 28 inputs, 50 hidden nodes, and 2 outputs - one for predicting the accumulated melon count, and the other for predicting the GDD. We used a learning rate of 0.001 and a momentum value of 0.9.

A serious consideration during this project was the lack of data; the harvest season spanned a maximum of 90 days (July September) of each year, and only the data from 2010, 2011, and 2012 was judged to be complete and accurate. This gave a maximum of 270 examples with which to train and evaluate the models. To deal with this small data set, we used cross-validation to build and test our models on the available data. Cross-validation is a technique where multiple models are built, each time using n-1 set of blocks of data to train a model and the remaining block of data to test the model. We used a 5-fold cross-validation, which means that 5 models were built and tested using 5 blocks of data each containing 20% of the data. We trained the models on three of those blocks (60% of the data), used one block (20%) as a a tuning set to prevent over-fit of the model to the training data, and tested it on the remaining block (20%). If the results are displayed by year, the values from the 5 test blocks are combined and then broken up by year so as to compute the error or each year.

The final data that was used to develop the model deployed during the 2014 season consisted of all of the melon harvest data from 2010-2012 and matching observed weather data from Environment Canada.

6 MODEL EVALUATION

To determine which of the weather variables were most important to the prediction of accumulated melon count, models were built that used each of the weather variables independently. The resulting models were compared to a persistence model that predicts the next value to be the same as the current value. Models were also developed using the GDD for each day (10 Deg Days) and the accumulated GDDs (Deg Days Summed from May 5).

The results are presented in Figure 5. Using only the previous accumulated melon harvest gave the lowest error. The most important weather variable combinations were found to be in order of importance: (1) the accumulated GDD (summed from May 5th, the average planting date), (2) all variables together, and (3) the maximum temperature. The persistence model appears to be the best because these results are only for a prediction one day in advance. It is well known that persistence model does poorly as one uses it to predict further into the future.

We examined numerous models build from combinations of the above variables looking for significant second order affects that would produce more accurate results. No superior models were found.

The final multiple task model used the previous 14 days accumulated melon harvest and accumulated GDD to predict the next days accumulated melon harvest and accumulated GDDs. This model was then used 14 times to produce a 14 day melon harvest forecast as shown in Figure autoregres. Figure 6 shows the mean and standard error results from a 5-fold cross-validation series of models presented by year. Note that each model has relatively low RMSE variation around the mean.

The models were tested to see how their accuracy varied over the 14 day prediction period, and were compared to a basic persistence model that always predicted the previous days value. Figure 7 show the forecast results of a persistence model and regressive models using alternatively the observed or predicted weather data.

A final production model was built from all of the 2010-2012 data using 80% of the data as the training set, and the remaining 20% as a tuning set to prevent over-fitting the model to the training data.

7 2014 DEPLOYMENT

In order to produce a melon harvest forecast for day t, we required the past 14 days of weather data (t-1 to t-14), a 14 day forecast of the weather (t to t+13), and real or estimated values of the past 14 days of melon harvest (t-1 to t-14). The NSCC weather station 60 data was used whenever possible as this was the closest station to the field. At times this data had to be replaced by Environment Canada data. Scotia Weather Services weather forecast data was used for all predictions except for the first two reports.

A forecast of the expected number of melons in the field was produced approximately twice a week from July 18 to September 12.

We received information about the melon harvest at VF twice during the 2014 harvest season. The first melon harvest information was received on August 1, so we were able to make predictions for

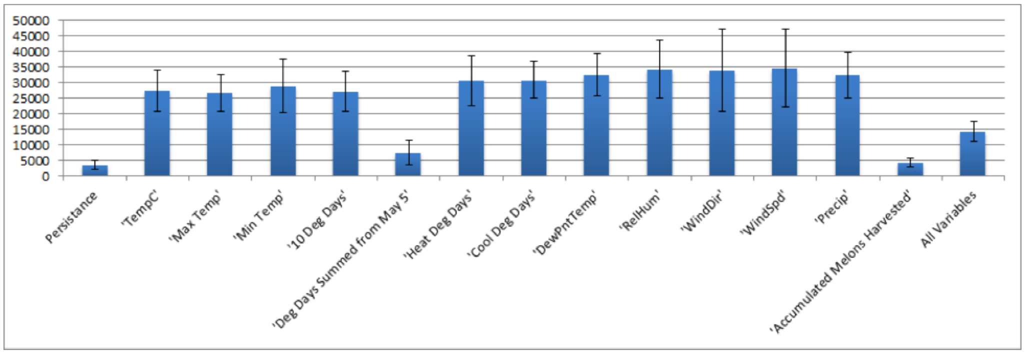

Figure 4: RMSE for models built using various weather variables.

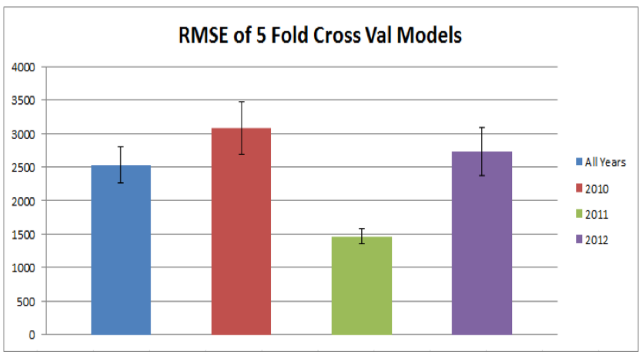

Figure 5: RMSE of 5 fold cross validation models and the standard deviation for all years combined and each year individually.

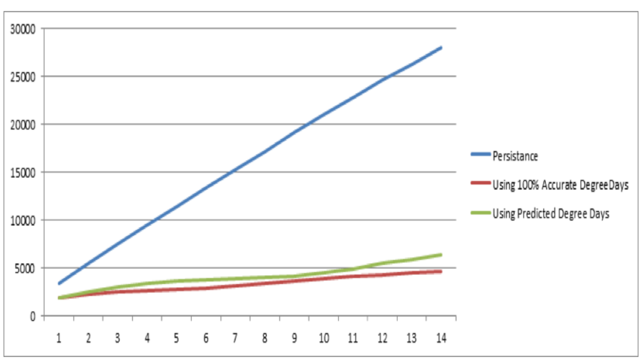

Figure 6: RMSE of a persistence model and regressive models using observed or predicted weather data.

August 2 using that information. The last set of harvest records arrived on September 7. Unfortunately, this meant that for most of the season predictions of accumulated melons was being based on prior predictions of accumulated melons.

7.1 Acquisition of Observed Weather Data

The weather information can be divided into two sets of data; the observed GDD from day t-1 to t-14, and the predicted GDD calculated from the weather forecasts for day t to t+13:

The observed weather data came from one of two sources: (1) NSCC Weather Station 60, which was located near the VF fields where the melons were growing. This data was originally manually emailed to us roughly once a week, but later the process was automated so we received updates daily. (2) Environment Canadas records from Kentville, which was located further from the field. We obtained this data by downloading it directly from the EC website.

These different sources were compared for accuracy and we chose to use the NSCC data whenever posisble, followed by the EC records when the NSCC values were not available. The GDDs were calculated from the maximum and minimum temperatures available from the weather data.

7.2 Acquisition of Forecasted Weather Data

The weather forecast data came from one of two sources: (1) Scotia Weather Services, which provided us a forecast localized specifically to the VF fields. We received these predictions daily as an automated email starting on July 29, and continuing daily until we stopped producing melon forecasts. (2) The Weather Networks weather predictions for Kentville. Originally, we manually obtained this data from their website only when needed, but in mid-August we began to save this information each day.

The important data from the weather forecasts was the maximum and minimum temperatures, which were used to calculate each days GDD. We preferred to use the SWS forecast to calculate the GDD, but used the WN forecast on the rare occassion when the SWS forecast was not available. We could have also used our own predicted GDD value, if neither of the forecasted data was available.

7.3 Acquisition of Observed Melon Data

We have assumed that the observed melon data would be provided to us on a frequent basis, preferably biweekly with minimal delay, but that turned out to be very challenging for Vermeulen Farms. We received melon harvest data twice during the entire harvest period; on August 1 and September 7. This meant that we had to use our models previous predictions of the melon harvest as input for future predictions for most forecasts that were issued (except those shortly

Estimated Error	2194.071	1551.517	1246.253	1237.393	1704.684	2322.274	1898.943	2231.22	2334.019	2549.624	2600.844	2502.721	2758.121	3000.869
Days Ahead	1	2	3	4	5	6	7	8	9	10	11	12	13	14
August	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Upper Error Bound	37543.58	43081.64	46059.17	46050.31	46517.6	47135.19	46711.86	47044.14	49716.68	53764.86	53816.08	53779.81	54670.99	55732.95
Predicted Melons in Field	**35349.51**	**41530.13**	**44812.92**	**44812.92**	**44812.92**	**44812.92**	**44812.92**	**44812.92**	**47382.66**	**51215.24**	**51215.24**	**51277.09**	**51912.87**	**52732.08**
Lower Error Bound	33155.44	39978.61	43566.67	43575.53	43108.24	42490.65	42913.98	42581.7	45048.64	48665.62	48614.4	48774.37	49154.75	49731.21

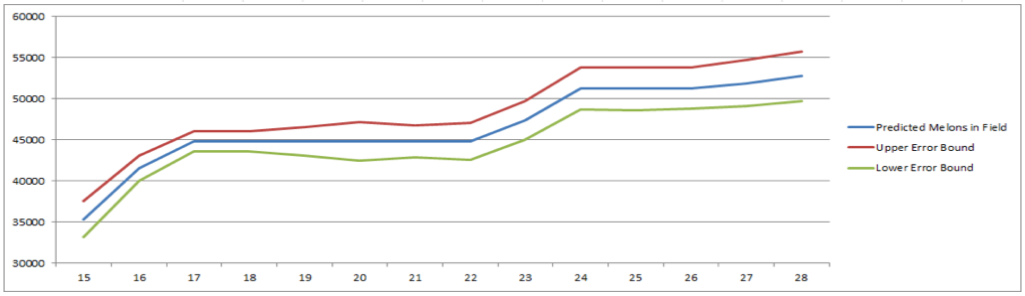

Figure 7: RMSE of the model, used to make 2014 predictions, for each of the 14 days evaluated over all days from 2010, 2011, and 2012.

after August 1). The automated or semi-automated collection of harvest data is subsequently a major area of improvement for any future projects in this area.

7.4 Communication of Melon Harvest Forecasts

Each 14 day accumulated melon harvest forecast was placed in a spreadsheet showing the actual values and an easy to read graph that included estimated error bars. This spreadsheet was then emailed to everyone involved in the project, most importantly to the owner and operator of VF. Figure 8 shows an example of one of the melon forecast reports sent out on August 15.

7.5 Evaluation of 2014 Predictions

After we had received the final results of the melon harvest, we were able to compare the actual melons harvested to our predictions of accumulated melons. Figure 9 shows the actual harvest as well as the predicted harvest (and 95% prediction interval) over the entire harvest season. Figure 10 compares the actual 2014 harvest to the previous three years used to build the predictive model.

The team examined the predicted and actual harvest graphs and discussed the reasons for the discrepancies observed during the second week of August and from September 6 onward with the farm owner. The following are our conclusions:

(1) The flat section of the actual harvest graph from August 7 to 18 is because of the late maturity of a different variety of melon. This same phenomenon can be observed from August 5 to 17 in the 2010 harvest data, also shown in Figure 10 . This is not observed in the 2011 or 2012 data that was used to build the deployed model. During this period Vermeullen Farms was waiting for melons to mature. If melons had been available, they would have been picked as predicted by the model. This suggests that planting times and variations in plant maturity dates are important inputs to consider in future work.

(2) The flat line of the actual harvest graph after September 7 as compared to rise in predicted harvest is most concerning. The team estimates that 50% of the melons yielded by the plants remained in the field at the end of the season. This was due to a market saturation that occurred early in the season. As can be seen in Figure 14, this does not occur in either 2010 or 2012. Based on Figure 11 we conclude that our model continued to predict values close to the true yield of the melon field (140,000) rather than the number of melons that the market would accommodate (70,000). This suggest that market factors are import input variables for predicting the harvest of a crop versus its yield on the plant.

Another factor in the accuracy of our model's predictions is the error introduced by the forecasted weather versus the observed weather. The models were built properly using observed weather, however it is the forecasted temperature and therefore GDD that drives the predictions. Errors in weather forecasts are unavoidable and so we must expect errors in harvest predictions. Figure 15 compares the results of predicting the melon harvest using the actual weather, SWSs forecast, and WNs forecast. Unsurprisingly, the actual weather had the lowest error. The difference between the accumulated melon predictions given the actual versus forecasted weather is insignificant, particularly beyond six days.

8 SUMMARY AND RECOMMENDATIONS

8.1 Summary

The goal of this project was to develop a model that can predict the number of cantaloupe melons available to harvest. The Acadia team worked with Scotia Weather Services (SWS), Vermeullen Farms (VF), and the Nova Scotia Community College (NSCC) on the project. The long-term vision was to develop a cloud-based harvest forecasting service benefiting SWS and the agricultural industry as a whole.

Multiple neural network models were developed in an effort to find the best possible combination of training and test data, network

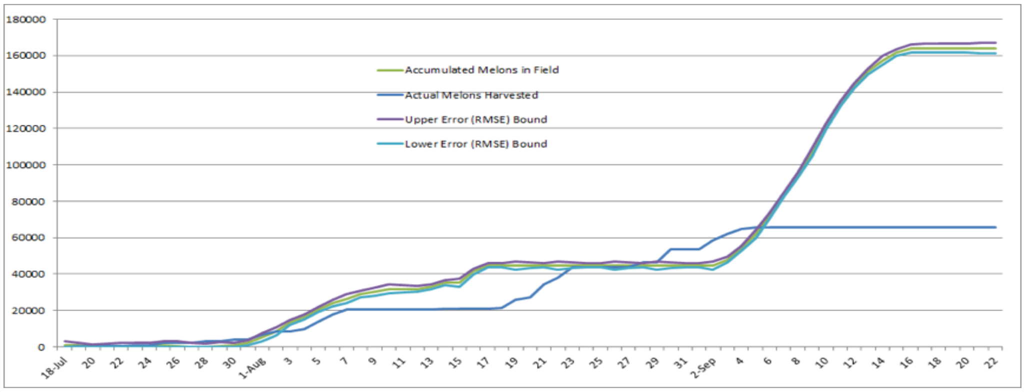

Figure 8: Comparison of the actual number of melons harvested versus the predictions (Accumulated Melons in Field) of the number available, from July 18 to September 23.

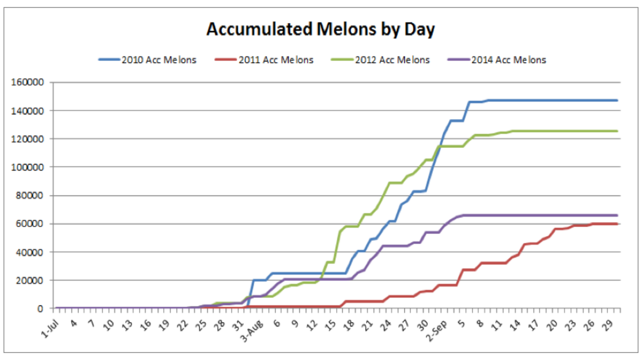

Figure 9: Accumulated melons for 2010, 2011, 2012, and 2014 from July 1 to September 30.

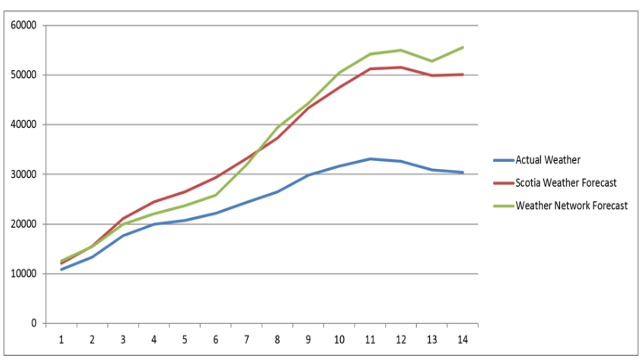

Figure 10: RMSE of the model evaluated over the 2014 growing season, using actual weather data and predicted weather data from Scotia Weather Services and the Weather Network.

configuration and parameter settings. A series of cross validation studies using 60% of the data for training, 20% for validation and 20% for test allowed us to selected the best model combination. Results were record and analyzed to ensure the best neural network model configuration and parameter settings. A final model was built using 80% of all of the available data from 2010-2102 for a training set and 20% as a validation set. This model was used to generate all predictions for 2014. The estimated error of this model was based on the average error of the cross validation models using the prior years data.

Using the final model, we use our neural network model and weather predictions from SWS (or WN) to produced melon forecasts from mid-July until the end of September. Each forecast used the previous 14 days of GDDs and melon harvests to predict the melon harvest for the current day and another 13 days into the future. We produced 15 forecasts from July 18 to September 12, which covered the 2014 growing season. The forecasts were emailed out to all participants in the project. We had expected to receive actual harvest data each week. However, due to the demands on the farm crew, harvest data was received only twice during the season, and once more at the end of the season. Therefore, most of our predictions used our previous predictions as inputs.

All observed and predicted data for the 2014 season was recorded. This included the melon forecasts and the weather predictions that were received. The evaluation of the melon harvest predictions during the season were hampered by the lack of weekly harvest data, so the analysis did not yield results until after these results were obtained at the end of the season.

The melon harvest evaluation showed that the model did well on forecasts in the early part of the season. However after August 7, the results show that the model overestimated the number of melons picked. This is for two reasons determined in a post-harvest analysis. First, there was a period of about 10 days in August during which a different variety of melon had to mature before it was ready to be picked. Unfortunately, a similar delay was only observed in one of the years used to build the predictive model. Second, market factors played a significant factor in our overestimation of melon yield. Beyond the first week of September, the market demand was low enough (either in general for melons, or for certain sizes of melons) that few melons were picked even though they were available in the field. Thus, the actual number of melons picked each week during the 2014 harvest was a function of both the yield in the field and the market demand. Since market factors were not inputs to the model, the model did not take into consideration their affect.

We found that our yield forecasts generally became less accurate and had less confidence as we predicted further days into the future. However, there was an interesting stabilization of accuracy beyond 10 days (see Figure 8). The cause for this is uncertain, but it may be related to a similar stabilization of the temperature forecast (and therefore the GDD) that were used as input to our model.

8.2 Recommendations

The following are the recommendations from this project effort as determined by the project partners:

(1) When harvesting a crop by hand on a busy farm there are challenges to accurately recording and communicating yield on a daily basis. Automated methods of capturing the yield at the field or grading warehouse are needed. There needs to be a cultural shift toward valuing automated or semi-automated methods of data capture.
(2) If market demand is a significant factor then it should be considered in future harvest estimation studies. One has to be sure what the training data actually represents - a decline in yield or a decline in produce demand. One would have to have access to historical produce market factors, in order to develop models that took such factors into consideration.
(3) The use of automated cameras is a tremendous resource for capturing agricultural activities and interventions over extended periods of time. The images can be reviewed for undocumented activity, animal intrusions, severe weather impacts, and vandalism.
(4) Having a basic weather station or weather logged in each field that can accurate record air and soil temperature (min/max over each day), humidity, precipitation, and solar radiation are important. Air temperature is the most important factor.
(5) A study is needed to determine the true market for harvest predictions and the dollar value of this information. VF valued the information and would consider paying for it in the future given improvements.
(6) The time of delivery of harvest yield predictions, its content and its format is very important to the success of the forecast information being used by the recipient. VF reviewed and used our predictions only once or twice during the season. This had mostly to do with the time of its arrival.
(7) At the outset we thought that precipitation and humidity would have played a significant role in the harvest forecast. These variables were dropped based on them making no significant contribution to accumulated melon predictions for the next day. We suspect we should have looked more carefully at their contribution to predictions for days 7-14 into the future. This should be considered in future studies.

Acknowledgement: We would like to acknowlege funding provided by the National Research Council through their IRAP Program and by NSERC through their Engage Grant Program.

Figure 11: The project team discussing options in one of the cantaloupe fields under study.

REFERENCES

[1] J.T. Baker and V. R. Reddy. 2001. Temperature Effects on Phenological Development and Yield of Muskmelon. *Annals of Botany 87* 87 (2001), 605–613.
[2] Phil Kenkel Byoug-Hoon Lee and B.Wade Brorsen. 2013. Pre-harvest Forecasting of County Wheat Yield and Wheat Quality using Weather Information. *J. Agriculture and Forest Meteorology* 2013, 168 (2013), 26–35.
[3] Richard A. Caruana. 1997. Multitask learning. *Machine Learning* 28 (1997), 41–75.
[4] Moh. Hajiyan. 2012. Early Prediction of Crop Yield. *Technical Report, School of Engineering, University of Guelph* (2012).
[5] Valdimir Konecny Jiri Statstny and Oldrich Trenz. 2011. Agricultural Data Prediction by Means of Neural Network. *Agric. Econ. Czech* 57, 7 (2011), 365–361.
[6] V.R. Reddy J.T. Baker, D.I. Leskovar and F.J. Dainello. 2001. A Simple Phenological Model of Muskmelon Development. *Annals of Botany 87* 87 (2001), 615–621.
[7] Raju Paaswan and Shahin Begum. 2013. Regression and Neural Networks Models for Prediction of Crop Production. *Int. J. of Science and Engineering Research* 14, 9 (2013), 98–110.
[8] G. Bourgeois S. Jenni, D. Cloutier and K. Stewart. 1996. A Heat Unit Model to Predict Growth and Development of Muskmelon to Anthesis of Perfect Flowers. *J. Amer. Soc. Hort. Sci.* 121, 2 (1996), 274280.
[9] G. Bourgeois S. Jenni, K. Stewart and D. Cloutier. 1998. Predicting Yield and Time to Maturity of Muskmelons from Weather and Crop Observations. *J. Amer. Soc. Hort. Sci.* 123, 2 (1998), 195201.
[10] Bharat Misra S. Veeenadhari and C.D. Singh. 2011. Data Mining Techniques for Predicting Crop Productivity A Review Article. *IJCST* 2, 1 (2011), 90–100.