Micro-climate prediction system using IoT data and AutoML

Dash M, Sharma A, Jalem S. R

Department of Artificial Intelligence, Wolkus Technology Solutions Private Limited, Bengaluru, India

A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

Abstract.

Microclimate variables like temperature, humidity are sensitive to land surface properties and land-atmosphere connections. They can vary over short distances and even between sections of the farm. Getting the accurate microclimate around the crop canopy allows farmers to effectively manage crop growth. However, most of the weather forecast services available to farmers globally, either by the meteorological department or universities or some weather app, provide weather forecasts for larger areas. To address this issue we developed a ~100 m spatial resolution AutoML framework that predicts hourly temperature and humidity over a period of 24 hours. The system uses one year of historical data from both IoT sensors and local weather forecasts for training and predicts temperature and humidity using individual models. The models were developed using a gradient boosting machine (GBM) approach. To account for model drift and data drift, an autoML framework was developed to automate model training on a monthly basis. The autoML framework uses i) a Bayesian optimization-based Hyperopt library to automate hyperparameter selection for the GBM models and ii) MLflow for model training, logging, and deployment purposes. For continuous deployment of the models, the autoML framework was integrated with Kubeflow for production-level serving. The models were developed using historical data from 6 different districts of Maharashtra, India and the accuracy was tested for over 50 grape farms (~50 ha) from that region via live deployment. Compared to local weather forecasts, the models showed a 15% and 30% decrease in mean error for temperature and humidity prediction, respectively. The developed microclimate framework also outperformed in predicting extreme temperature and humidity conditions by ~30%. Timely prediction of extreme weather conditions would be helpful in effective crop protection and crop management. Though the AutoML Microclimate framework was developed and tested for grape farms in the Maharashtra region of India, it can be easily extended to other regions and crops as well.

Keywords.

IoT, Smart climate, Micro-climate, AutoML, Bayesian Optimization

Introduction

Agricultural productivity is at high risk due to global warming and climate change especially in the tropical and subtropical regions. Crops are sensitive to climate change, including changes in temperature and humidity (Rukhsana *et al.,* 2021). Among the changes, temperature increase has the most likely negative impact on crop yields as rising temperatures causes shortened crop duration and increased respiratory losses (Zhao *et al.*, 2017). On the other hand, extreme humidity increases the likelihood of pests and diseases, thereby, affecting both crop productivity and quality (Rabbi *et al.*, 2019). One way to overcome challenges faced due to these extreme weather conditions is to accurately forecast the weather in advance. Since climatic parameters can vary even between sections of the farm, local weather forecasts require microclimate consideration to improve forecast accuracy and for better crop management. Most of the local weather forecasts available to farmers are at a regional level. Without high resolution hyperlocal microclimate predictions, effective crop management becomes a challenge for the farmers. The objective of this study is to develop a ~100 m spatial resolution AutoML framework that predicts hourly temperature and humidity over a period of 24 hours with higher accuracy that can help with effective crop management.

Material and Methods

Data collection

The study was conducted for 50 grape farms located across 6 different districts of Maharashtra, India. Each of these farms were installed with the Fasal IoT (https://fasal.co/) device on or before January 2020. The microclimate data was collected using the Fasal IoT device (Fig. 1) that is equipped with 12 sensors. The devices measure several microclimate parameters like temperature, humidity, air pressure, soil moisture (primary and secondary root zone), soil temperature, leaf wetness, solar intensity (lux), rainfall, wind speed and wind direction. The data is recorded in hourly frequency and uploaded to a cloud database using 4G cellular service. The BME sensors used to measure temperature and humidity were placed at plant canopy level (~2 meter from ground). Along with the IoT data, local forecast data from several weather service providers were also collected. Forecast data was collected using their paid API service. The forecast data was updated four times in a day and stored in our cloud database. All the data collection, data preprocessing and model development were done using Python 3.7 and supported libraries.

Data preprocessing

For developing the data we used 1.3 years (December 2019 to March 2021) of historical temperature and humidity forecast data as well as the IoT sensor data from the region. For forecast data, we used the data from 2 services based on their accuracy and service reliability (very less or no API failures). Due to errors in data collection, we performed data cleaning by removing outliers using the Interquartile Range (IQR) approach (Wanga and Yong, 2021). Post data filling, missing values were imputed using the k-Nearest Neighbor (KNN) algorithm. The IoT data and forecast data were normalized with respect to datetime as well as units. accelerate the training speed of the model, and make the model more accurate than before, min-max normalization was used to transform all numerical features to zero and one.

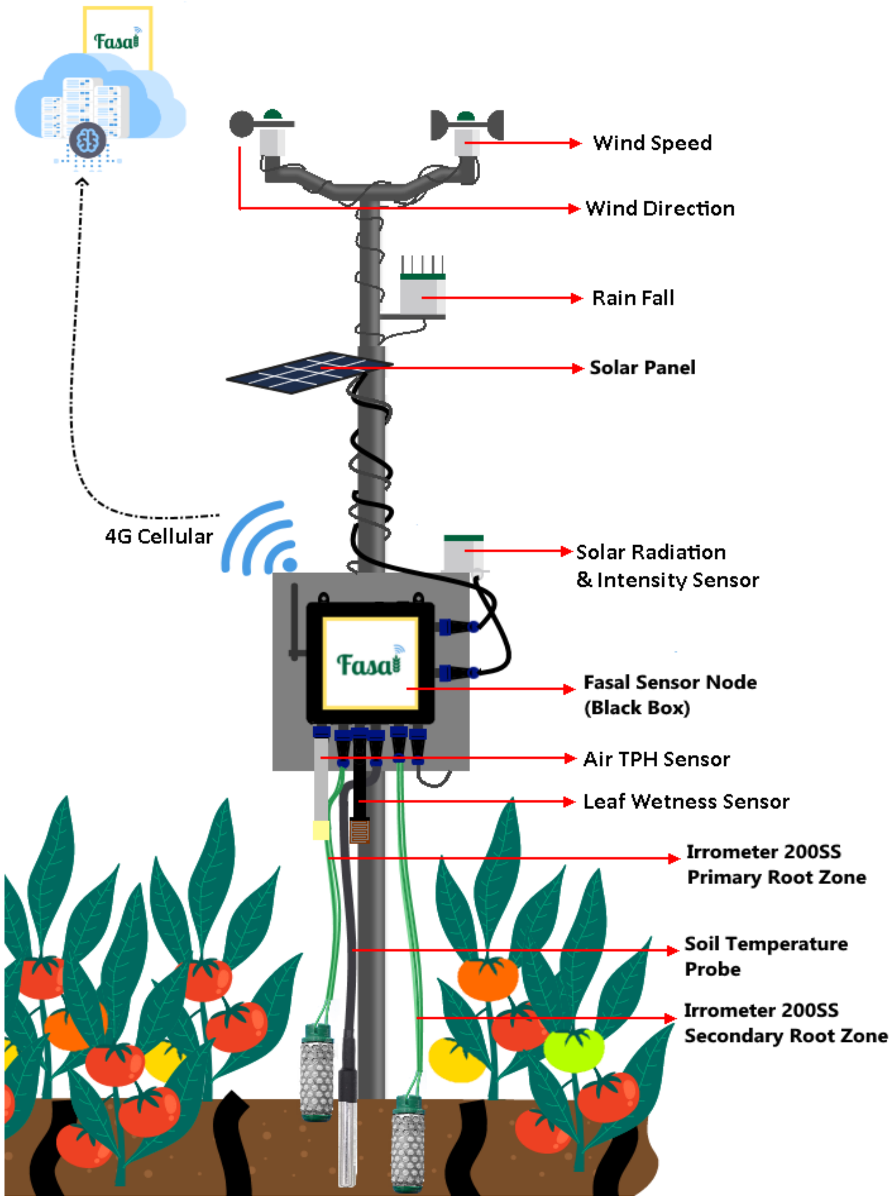

Fig 1. Fasal IoT device used for data collection. The device has 12 sensors and data from the device is uploaded to Fasal Cloud every hour.

Model development

Microclimate parameters like temperature and humidity are associated with some time features. Therefore, we created some time features, such as the hour of day and month. In addition to these, we also used time lag features like the temperature and humidity 24 hour or 48 hour before. We used the Pearson correlation coefficient for feature selection followed by feature importance to further filter features. We used one year of historical data (December 2019 to December 2020) for training and 2 months of data (January 2021 to February 2021) for testing separate models for temperature and humidity. Both training and testing data went through the same preprocessing process as described above. We used XGBoost algorithm for training the models. Ten-fold cross-validation was used to test algorithm accuracy and adjust the parameters for higher accuracy. Hyperparameter optimization was performed using the Hyperopt library that uses a Bayesian optimization algorithm to select the best set of hyperparameters. We used both MAE (mean absolute error) and RMSE (root mean squared error) as metrics for model performance analysis.

Model Productization

Post model development, Flask was used to develop an API to use the model for prediction. The Flask API was containerized using Docker and then deployed using Google Cloud Run. We used Airflow to schedule data collection and feature engineering that was cached using Redis. This data was the input for the API request. MLFlow was used to store models and run the training pipeline. The forecasted temperature and humidity from the developed API was stored in BigQuery. BigQuery was also used to store training data.

Results and Discussion

We developed two separate models for temperature and humidity. Table 1 compares the accuracy of the developed model post deployment from March 2021 and April 2021 in comparison to the forecasted from two weather service apps that were also used in developing the model. We used several metrics to compare the model performance. MAE (mean absolute error) and RMSE (root mean squared error) to get an understanding of average error distribution and any outliers in prediction. Both MAE and RMSE were lower for the model forecast in comparison to the two sources. However, both the MAE and RMSE values were very similar in case of the model prediction suggesting fewer outliers in the prediction error for the model output. Therefore, to further understand the forecast improvement, we looked into what percentage of unfavorable conditions were correctly predicted by the model vs the two sources. During the fruit ripening stages of grapes (March 2021 to April 2021), high temperature >32 °C and humidity <50% or >85% have been shown to be unfavorable and affect harvest quality (Mori *et al.*, 2007). For temperature prediction, prediction error <1°C was considered acceptable whereas for humidity it was <2%. Compared to the forecast data from external sources, the developed AI models showed almost 30% improvement in predicting unfavorable weather temperature and humidity conditions.

Table 1. Comparison between the temperature and humidity forecast from the developed model (Model) vs two different commercial weather forecast service providers (Forecast source 1 and Forecast source 2).

Parameter	Forecast source	MAE	RMSE	% of extreme conditions captured with threshold error
Temperature	Model	3.08	3.96	86%
	Forecast source 1	4.24	6.44	54%
	Forecast source 2	3.96	6.18	57%
Humidity	Model	5.65	5.88	79%
	Forecast source 1	7.26	12.63	42%
	Forecast source 2	7.88	13.27	39%

After May 2021 we observed model drift or a degradation in the temperature model's accuracy or an increase in RMSE (Fig. 2). When we re-trained the model and incorporated data from March and April 2021, we observed the model performance to be improved as shown in Fig. 2. Similar pattern was also observed for humidity model prediction.

Model drift is a common phenomenon and occurs due to alteration in the environment. To address this we have developed an AutoML framework (Fig. 3) where anytime a degradation in a model's performance occurs, the system will trigger re-training of the model. We used RMSE as a parameter to track model performance since it captures any outlier in the prediction accuracy as shown in Table 1. The training pipeline (including model testing) was developed using the same approach described earlier. MLFLow with Google Cloud Platforms's (GCP) Cloud Storage and data backend was used to store all the models and artifacts. Kubeflow was used to deploy the training pipeline for automating training processes or for long running training.

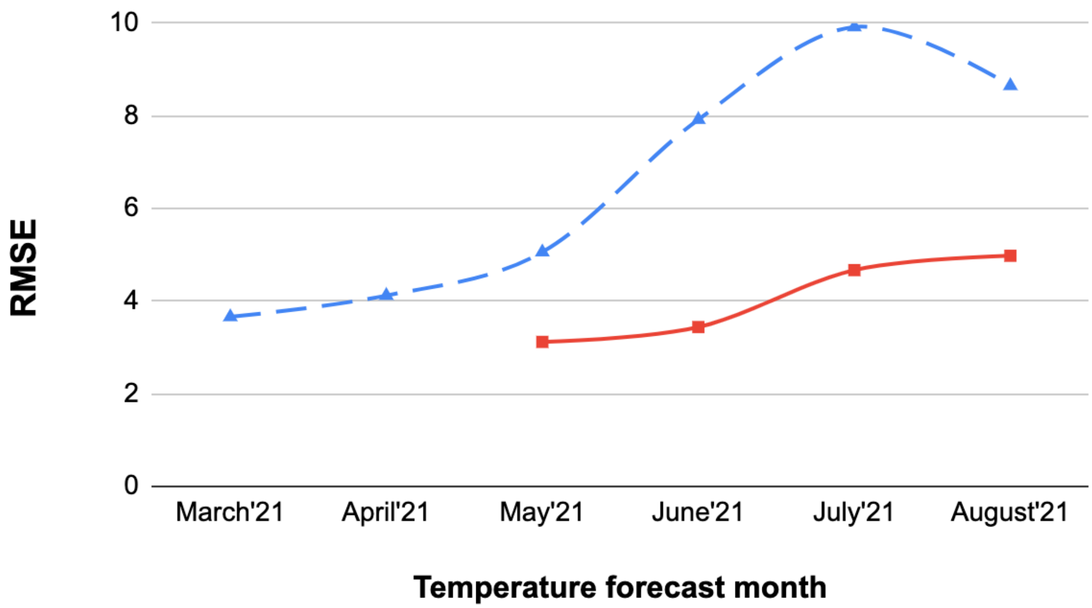

Fig 2. Model drift in temperature model v1 prediction (dashed line) as displayed with an increase in RMSE value from May 2021 onwards. When the model was re-trained the RMSE values were again decreased (line with square points).

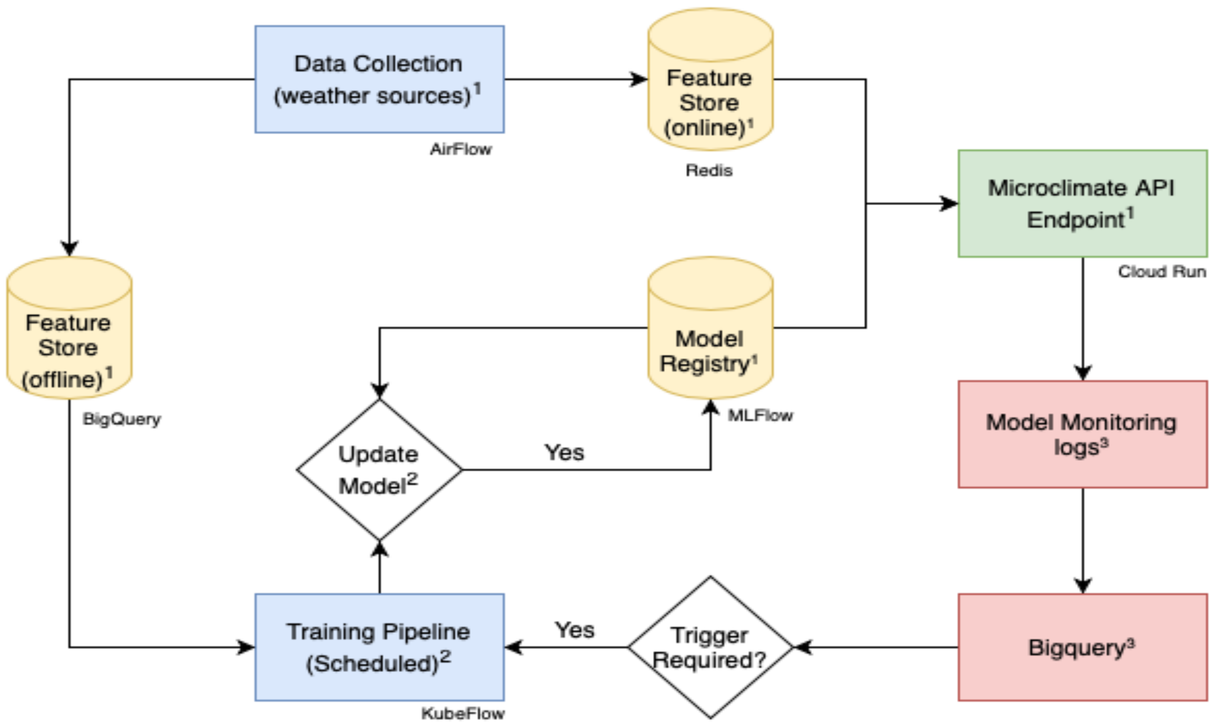

Fig 3. AutoML framework to automate training of the models whenever the model's accuracy degrades below a certain threshold. We used RMSE as a metric to trigger the training pipeline.

Conclusions

This study shows that microclimate forecasts can be significantly improved by combining IoT and machine learning technology. The degradation observed in most machine learning's models over time due to model drift, can be corrected using an AutoML framework. The AutoML framework should encompass a training pipeline for continuous model training and development as well as CI/CD pipeline to continuously deploy the model to production servers.

References

Mori K., Goto-Yamamoto N., Kitayama M., Hashizume K. (2007). Loss of anthocyanins in red-wine grapes under high temperature, Journal of Experimental Botany, Volume 58, Issue 8, Pages 1935–1945, https://doi.org/10.1093/jxb/erm055.

Rabbi B., Chen Z.-H., Sethuvenkatraman S. (2019). Protected Cropping in Warm Climates: A Review of Humidity Control and Cooling Methods. Energies, 12, 2737. https://doi.org/10.3390/en12142737

Rukhsana, Alam A., Mandal, I. (2021). Impact of Microclimate on Agriculture in India: Transformation and Adaptation. In: Rukhsana, Alam, A. (eds) Agriculture, Food and Nutrition Security. Springer, Cham. https://doi.org/10.1007/978-3-030-69333-6_3.

Wanga J., Yong S. (2021). An AETA Electromagnetic Disturbance Anomaly Extraction Method based on Sample Entropy. IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). Vol. 5. IEEE, 2021.

Zhao C., Liu B., Piao S., Wang X., Lobell D.B., Huang Y., Huang M., Yao Y., Bassu S., Ciais P., Durand, J.L. (2017). Temperature increase reduces global yields of major crops in four independent estimates. Proceedings of the National Academy of Sciences, 114(35), pp.9326-9331.