Sensor Data Processing-Agricultural Crop Yield Prediction

Prashant Kumar 202018005 | Prashun 202018006 | Sonal Nathwani 202018034 | Abhijeet Kumar 202018042 | Jay Shah 202018051 | Shubham Sharma 202018052 | Smit Gandhi 202018057

I. Abstract

— This is a report for a project that attempted at crop yield prediction using sensor data of weather as well as historic crop production data. For this, data has been collected from data.gov.in and worldweatheronline. The Facebook's prophet model has been used as the main forecasting algorithm. Later, a dashboard has been created for better representation of the experiment which is hosted on the local system.

II. Introduction

The focus of this experiment is on sensor data processing and is more inclined towards agricultural data captured through various sensors or recorded in history. Agricultural market has a huge potential in a nation like India. Farmer's in India play a major role in feeding the growing population of the country. This makes crop analysis and prediction as important as crop production. Farmers can use crop yield prediction data to make their decisions about crops.

Agriculture yields' prediction is one of the major challenges in machine learning. There are many factors affecting crop yield such as crop genotype, environmental factors like soil conditions, farmers efforts such as proper irrigation, timely plantation, etc. This forecasting is specifically relied on climate features to predict crop yield. Based on the considered datasets, we have taken into consideration the climate features specific to the agricultural seasons in India such as Rabi, Kharif and whole year and then have tried to make the crop yield predictions for the upcoming year.

III. Methodology

A. Dataset

To perform our predictions, there were two main types of datasets that were basically needed for us. They are: Historic data related to crop yield and data related to the climatic conditions. For crop yield production data we used a dataset that had crop data for all districts of all states of India for all crops grown and their seasons from the year 1998 to 2014. The dataset was gathered from data.gov.in[1]. This dataset can be viewed using the dashboard that we created for demonstration purposes. For climate data, we used data available via worldweatheronline [2] for eight cities of India namely Bengaluru, Bombay, Delhi, Hyderabad, jaipur, Kanpur, Nagpur, Pune from the year 2009 to 2020. The climate features which we have incorporated are temperature, humidity and precipitation. The granularity of this data was hourly.

Now, our main concern was to make use of this available data so as to make the predictions. Also the timeframe for which the data was available was different and we had to find an intersecting time so as to perform the experiment. The data from 2009 to 2014 was finally considered keeping this constraint in mind. Also, one form of data that we had was of an hourly time format whereas the other was one value for a particular season of the year. Hence we had to bring both the data to the same level to run a prediction model on it. We did perform some mathematical operations to get a single value for the weather parameters such as averaging for the temperature and humidity and adding up for precipitation. All these operations were performed for particular seasons such as July-October timeline for kharif crops and November-March timeline for the Rabi crops.

B. Facebook's Prophet Model

The model that was chosen as the main forecasting model is Facebook's Prophet algorithm. The reason for choosing this was that this package offers precise borders that are easy to tune. A person who isn't trained to work with time series forecasting can also easily adapt to this model with minimal knowledge of time series and only specific domain or business knowledge. The prophet is an open-source package released by the Facebook team of Core Data Scientist which is available for download in CRAN and PyPI as it is a forecasting tool that can be implemented in both R and Python.

One of the biggest advantages of using this model is that it is more inclined towards curve-fitting rather than explicitly accounting for temporal dependence, unlike other time series models. This definitely helps in capturing the seasonalities and trends in a better way which may not be easily taken into account by the typical time series models.

As this model was mainly designed for the purpose of forecasting at scale at Facebook, it incorporates all the features affecting their performance. It uses a decomposable time series model with three main model components: trend, seasonality, and holidays. An important benefit of the decomposable model is that it allows us to look at each component of the forecast separately. They are included in the following equation:

$$y = g(t) + s(t) + h(t) + \epsilon(t) \quad (1)$$

g(t) is the trend function that models non-periodic changes in the value of the time series. s(t) represents the periodic changes (e.g., it is in weekly and yearly seasonality). h(t) is the effects of holidays that occur on potentially irregular schedules that may be available for one or more days.epsilon t stands for any missing idiosyncratic changes that aren't accommodated in the model. It is assumed that ε (t) is normally distributed. This is what the actual model is made up of. However, the kind of predictions that we want to make have no effect on holidays, and hence in our algorithm that parameter is omitted.

1) **Saturated, NonLinear Growth:** The trend is modeled by fitting a piecewise linear curve over the trend or the non-periodic part of the time series.There are two trend models that have been implemented: a nonlinear saturation growth model and a piecewise linear model. The majority of the time the growth that has to be taken into consideration is a non-linear type of growth that saturates at a carrying capacity. For Facebook, the carrying capacity could be the number of people having access to the internet facilities whereas in our case that could be the availability of agricultural land. This type ort of growth is usually presented using the logistic growth model, which in its most basic form is :

$$g(t) = \frac{C}{1 + exp(-k(t-m))} \quad (2)$$

2) **Seasonality:** Any business time series model, usually exhibits a periodic seasonality which is depicted by the behaviours of the results. For example, it could be a 5-day work week or some vacation period that affects the facebook business whereas in our case it would be the seasonal weathers such as kharif and rabi which exhibit a typical repeating periodic seasonality. The prophet relies on a fourier series to provide a flexible model. Seasonal effects s(t) are estimated with the following function:

$$s(t) = \sum_{n=1}^{N} \left(a_n cos(\frac{2\pi nt}{P})\right) + \left(b_n sin(\frac{2\pi nt}{P})\right) \quad (3)$$

Here P is the regular period that anyone expects the time series to have (e.g.P = 365.25 annual data or P = 7 weekly data, when we measure our variability in days).Parameters [a1, b1, ….., aN, bN] need to be estimated for a given N to model seasonality.For time series, if anyone believes the high frequency components are just noise and should not be considered for modelling, anyone could set the values of N from to a lower value. If not, N can be tuned to a higher value. For yearly and weekly seasonality, researchers at facebook have found N = 10 and N = 3 respectively to work well for most problems.

3) **Holidays and Events:** For a given forecasting problem, prophet uses the union of the global set of holidays and the country-specific ones. The country specific holidays are supposed to be given as input by the user in a specified format. Also a window period around the holiday can be set as a parameter by the analyst so as to consider the effect of those days around the holiday also a holiday itself. However, as mentioned earlier this component is omitted in our case of forecasting.

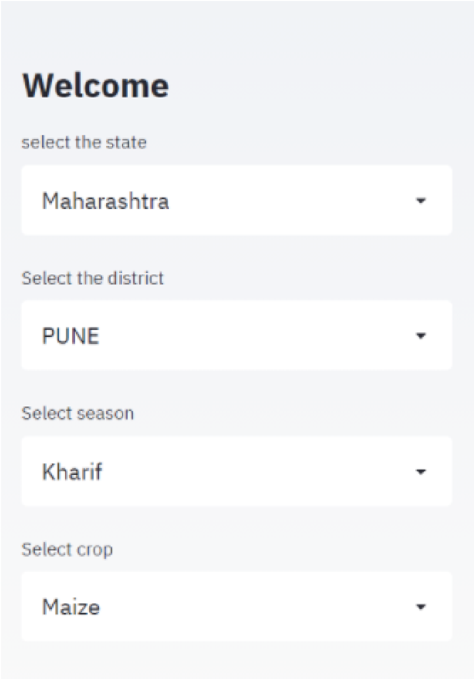
IV. Implementation

For the implementation, we have basically made use of the following packages and the entire code is in python: fbprophet, streamlit, numpy, pandas, seasons. The use of fbprophet has been done for forecasting purposes as mentioned earlier. This model basically requires a timeline parameter and then a prediction parameter. In our case the timeline parameter is the year and the prediction parameter is the yield(crop production for rabi/kharif for that year). Along with this multiple regressors could be added to this model to predict based on other external parameters that the analyst thinks that affects the prediction value. Here, we have added four other regressors to perform the predictions. These regressors include area, temperature, humidity and precipitation. Hence the final predictions that are made are based on all these factors.

Now, once the model has been implemented, there is a simple dashboard that has been created using streamlit which consists of a few drop down menus to select from based on the availability of the data with us at our end. Based on those choices there are a few visualizations that come up on the screen. These visualizations depict the historic crop yield for the particular selection, the historic weather parameters data for the particular selection and the graph with the historic as well as predicted data. The following are the images that would give a clear look and feel of the dashboard.

A. Dashboard

Users can select state, district, season, crop. Based on selection of season, The values of climate features are calculated. After the user selects all options, crop yield data is shown. Year, Area and production for that crop, in the season.



After the user selects all options, crop yield data is shown. Year, Area and production for that crop, in the season.

Crop Production Prediction

state: Maharashtra district:PUNE season:Kharif crop:Maize

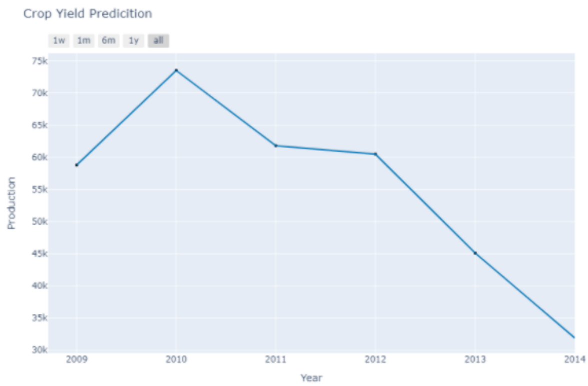
	Crop_Year	Area	Production
134203	2003	5500	12400
134237	2004	7000	13600
134293	2006	8500	19600
134320	2007	74	207
134347	2008	5700	13400
134371	2009	4900	10100
134394	2010	10500	24500
134416	2011	10500	31400
134442	2012	9500	23300
134469	2013	11800	33000
134496	2014	13300	54500

After climate features are added. A dataframe is shown to the user like this.

Regressors: Area, Temp, humidity, Precip

	ds	Area	y	Temp	humidity	Precip
134371	2009	4900	10100	22.9682	86.7588	516.2000
134394	2010	10500	24500	23.2432	87.0664	338
134416	2011	10500	31400	23.7093	80.0488	290.9000
134442	2012	9500	23300	23.9380	78.1619	365.6000
134469	2013	11800	33000	23.5501	82.0779	307.3000
134496	2014	13300	54500	24.2029	75.4648	324.6000

So if the user selects the Kharif season, for each year, the average temperature is calculated for the months which fall in the Kharif season. Based on the regressors selected by the user. A prophet model is trained on 5 years data and production for 6th year is predicted i.e. 2014. A graph is shown with values of production for training period and prediction period. Due to lack of climate data, the prediction period is only for 1 year.
We have also added an option to select a regressor from temperature, humidity and precipitation.



The black dots show the actual value for 5 year. And the predicted value is without black dot.

V. Evaluation

To evaluate the performance of the algorithm that we have implemented we checked the directional accuracy of our predictions and it gave a score of 1.00. This says that our implementation would always give the current prediction in terms of positive or negative directions i.e. for the chosen parameters, would the yield increase or decrease. However to evaluate the quantification of the increase or decrease values we did not calculate any specific metric as due to lack of data we had the availability of only one observation to perform the prediction on and hence it was not of much insight. So we have displayed the absolute error value for each prediction based on the options that a user selects.

VI. Conclusion

Through this approach we tried our hands on time series forecasting sensor data for agricultural crop yield predictions by using facebook's prophet algorithm. We achieved results that gave us great directional accuracy however there can be further improvisation in tuning the parameters and regressors to get close to the actual predictions even after correct direction of predictions.

References

[1] Data.gov.in,https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics.

[2] Worldweatheronline, https://www.worldweatheronline.com/

[3] Taylor, Sean J., and Benjamin Letham. "Forecasting at scale." The American Statistician 72, no. 1 (2018): 37-45.

[4] Ramesh, D., and B. Vishnu Vardhan. "Analysis of crop yield prediction using data mining techniques." International Journal of research in engineering and technology 4, no. 1 (2015): 47-473.

[5] Directional Accuracy,https://en.wikipedia.org/wiki/Mean_directional_accuracy.