Dulux/Yates sales forecast model using weather and live chat data.

Recommendations to improve the current model:

* Train the model on more recent sales and weather data. Currently we have weather data from 2009 until July 2016, sales data from 2012 until June 2017 and live chat data from 2015 until June 2017. Therefore the model is trained and tested on the overlapping data only from 2012 to July 2016 for sales and weather purposes. Including live chat data limits the time period used to around one year from 2015 to July 2016.
* The current model only uses features with information from 5 weeks ago in order to predict sales volume a month ahead (i.e. weather data, chat and sales volume from five weeks ago). Historical features using sales and weather data from more than 5 weeks ago should be used to improve the timing and amplitude of the sales forecast. This would improve the accuracy of the model.
* Using historical sales, weather and live chat data from more than 3 months ago would enable the model to forecast 3 months in the future thereby allowing the business to plan ahead accordingly.
* In addition, features could be generated per Bunnings store trade area
* The current model is run per product/product category per area. This should be extended to a model per climate zone. Three different climate areas for Australia appear to be sufficient. However, we have downloaded data from 470 weather stations in Australia at the moment. The data from all stations in one climate area/zone would need to be averaged.
* Depending on real time availability of sales data across all products and stores/areas more recent sales data could be used. For example if sales data is updated every 2 weeks, the model could use the most recent data in addition to historical sales data. In combination with current weather data, or even using forecasted weather, the model becomes nearly real-time.
* The model uses a combination of weather features but could also account for interactions and non-linear relations between variables to improve the sales forecast. We are using multiple linear regression at the moment, i.e. the model does not account for non-linear behaviour.
* Creating monthly, quarterly and yearly model features for weather data, sales and chat volume. For example "rain/temperature/sales volume in the last month/quarter/year" would account for seasonality.
* Mathematical model tuning, e.g. scaling the data and/or put it in a range of values between zero and one to make it easier for the model to pick up trends are common practise.
* A model run on sales data from all products across all stores using weather and sales features yield good results. Depending on the forecast ability of the model one model might be enough to answer several business questions.
* The model currently uses data from three states and 37 products only. Using data from all states for all products would most likely improve the models predictive capabilities. The more data the better.