**TCCC India Process Documentation**

Step1: Extracting the data.

Temperature and precipitation data are extracted from “Weather Data” file. After extraction make sure precipitation value is converted to daily value by dividing by the total no. of days.

Core sparkling (Independent variable) and CCI core sparkling (Dependent variable) data are taken from “Long Term Category Value Volume” file.

Ensure that one has selected the right Market, States and products.

Step2: Once extraction is done; we need to Forecast the results for Total Category (core sparkling) and which is an independent variable using “Auto Arima” method.

Unemployment (after extracting from the link divide it by 100) is also Forecasted using auto-arima.

Economy is calculated based on the Previous year’s growth.

(R code is written to run the Auto Arima and forecast the results for the same).

Step3:

Once required data is ready, we can start building the model. Variables used for weather data are as follows:

VolSales(Dependent Variable):

No transformation used in any of the model.

Independent Variables:

Tmax:

Overall transformation used in the model:

D\_Tmax(First difference),Ln\_Tmax(log of tmax),Lag\_tmax

Avg\_precipitation:

Overall transformation used in the model:

lag1\_avg\_prcp

Total Spk:(Category Sales)

Overall transformation used in the model: Split the Spk variable based on trend, taken log and lag for the same.

Unemployment:

Overall transformation used in the model:

Lag1(Unemployment), Log(Unemployment)

Economy:

Overall transformation used in the model:

Log(Economy)

Dummies – Have used a couple of dummies in few models

Seas1(Seasonality): Used in few models to capture the seasonal trend of sales across the given time frame

As of now we have used above independent variables, not necessarily that these must be used, we can change / add variables and/or any transformations used.

The ultimate intent is to use independent variables to provide us w best fit hence forecast.

We would then consider only the variables and transformed variables which give a good fit and proceed further.

Step4:

Data set contains 80% of Training Data and 20% of Testing data which is represented by the variables x,y z in the python code .

X:Y=range of training data set

Y:Z=range of testing data set

The remaining rows will be used for forecasting the results.

Approaches used for building the model are:

* OLS (Ordinary Least Squares): It’s an additive model and is a method for estimating the unknown parameters in a linear regression model. It produces the best possible coefficient estimates when your model satisfies the OLS assumptions for linear regression.

Priors used in this technique is Flat priors; Models have been built using modelling toolkit of R

* Elasticity calculations :-

If Log of temp is used – (Coefficient/Avg temp +1)/Avg Sales

If Temp is used directly – Coefficient/Avg Sales

* Elasticity range :

Temp : 2% to10%

Precipitation : -0.5% to -1.0%

* The purpose of OLS is to get elasticities of Temp and Precip which is then used for dynamic coefficients calculation
* RF(Random Forest) :Package used in this approach is Sklearn RandomForestRegressor()

None of the parameters are assigned in this model as we have got good fit without tuning the parameters

* ANN(Artificial Neural Network): Package used is Sklearn neural network MLPRegressor().

Parameters tuning is required in this algorithm to get a good fit.

Below are the few parameters used in ANN Algorithm:

**hidden\_layer\_sizes, max\_iter, batch\_size.**

**Eg:**

**hidden\_layer\_sizes = (256)** [The number of neurons in the hidden layer]

**max\_iter=50**, Maximum number of iterations

**batch\_size=1**, The size of a batch **(**size of neurons sent in batch**)** must be more than or equal to one and less than or equal to the number of samples in the training dataset

Above parameters can be altered accordingly until we get better MAPE for the same.

Step 5:

The objective of this exercise is to get the predicted results for training data set, testing data set and to get the forecasted results using Ensemble Model.

Ensemble methods use multiple learning algorithms (Here it’s OLS, RF, ANN) to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone.  It takes a weighted average of individual predictions to form a final prediction.

The package used is Sklearn VotingRegressor()

[VotingRegressor](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html#sklearn.ensemble.VotingRegressor) is to combine conceptually different machine learning regressors and return the weighted average of predicted values.

Step6: Calculation of Due To’s:

|  |  |
| --- | --- |
| Temp Diff | Tmax(current year)- Tmax(previous year) |
| Precip Diff | precip(current year)- precip(previous year) |
| DT Temp | Beta \* x(current year) -Beta\*x(previous year)  x : tmax or ln(tmax) |
| DT Precip | Beta \* x(current year) -Beta\*x(previous year)  x : precipitation |
| % DT Temp | DT Temp / raw or transformed tmax |
| % DT Precip | DT Precip / raw or transformed precip |
| DT Temp using Elasticity | Temp Diff \* Temp elasticity \* previous year VolSales |
| Sales Index | Average sales of the particular month **/** Average of sales for entire time period |
| ***DT Temp using Sales Index*** | Sales Index \* DT Temp |
| ***DT Precip using Sales Index*** | Sales Index \* DT Precip |
| ***DT Weather Sales Index*** | DT Temp using Sales Index + DT Precip using sales Index |
| ***DT % Temp using Sales Index*** | DT Temp using Sales Index / previous year VolSales for the particular month |
| ***DT % Precip using Sales Index*** | DT Precip using sales Index / previous year VolSales for the particular month |
| ***DT % Weather using Sales Index*** | DT % Temp using Sales Index + DT % Precip using Sales Index |
| ***DT Temp MoM/YoY*** | Month over month difference of DT Temp using Sales Index |
| ***DT Precip MoM/YoY*** | Month over month difference of DT Precip using sales Index |
| ***DT Weather MoM/YoY*** | Month over month difference of DT Weather Sales Index |

Same has been calculated for the Economy and unemployment variables.

Step6: Adding all the required variables to the Power BI.