CHAPTER - 5

TESTING

Testing plays a major role, it helps in discovering the defects and also the main objective of testing is, problems being detected by the end user must be avoided. The proposed system contains three modules with various functionalities. The four modules are crime type and crime location, forecasting, arrest rate and the crime rate and count week, month and year wise for crime scene in jewel theft and murder. They work one after another, result of first module induces the module 2 i.e., sequential workflow can be observed in the proposed system. Testing each and every functionality of the modules is very important in this case. Crime prediction in jewel theft murder saves the time and makes easy process for the police investigation.

5.1 INTRODUCTION

The test cases are passed for different modules. The proposed system architecture contains three modules. Each module is depended on the before modules, which is the sequential work flow. If our proposed system architecture fails, then all our objective and functional modules of the system is entirely lost. Therefore, testing each cases and every functions of the modules is important in our work and the sequential work flow can be observed from the process flow. Jewel theft murder prediction gives effective method for the investigation process and also the modules gives us the forecasted crime scene rate. Thus testing is done with many possible test cases to ensure that our system functions are per the proposed system architecture and as well as by adding the new features and data for testing.

5.2 TESTING THE MODEL

Prophet includes functionality for time series cross validation to measure forecast error using historical data. This is done by selecting cut off points in the

history, and for each of them fitting the model using data only up to that cut off point. We can then compare the forecasted values to the actual values. This cross validation procedure can be done automatically for a range of historical cut offs using the cross validation function. We specify the forecast horizon (horizon), and then optionally the size of the initial training period (initial) and the spacing between cut off dates (period). By default, the initial training period is set to three times the horizon, and cut offs are made every half a horizon.

The output of cross validation table 5.1 is a data frame with the true values y and the out-of-sample forecast values yhat, at each simulated forecast date and for each cutoff date. In particular, a forecast is made for every observed point between cut off and cut off + horizon. This data frame can then be used to compute error measures of yhat vs. y. Here we do cross-validation to assess prediction performance on a horizon of 365 days, starting with 730 days of training data in the first cutoff and then making predictions every 180 days.

	horizon	**RMSE**	**MAE**	**MAPE**	**MDAPE**	**coverage**
0	51 days	7.007761	2647.2175	1851.73	0.390	0.0
1	55 days	3.120698	1766.5498	1109.87	0.229	0.0
2	56 days	2.980418	1726.3885	1012.91	0.214	0.0
3	60 days	2.975610	1724.9956	1007.99	0.209	0.0
4	61 days	2.954189	1718.7729	969.95	0.203	0.0

Table 5.1 Cross validation performance metrics

The performance metrics utility can be used to compute some useful statistics of the prediction performance (yhat, yhat_lower, and yhat_upper compared to y), as a function of the distance from the cutoff (how far into the future the prediction was). The statistics computed are mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percent error (MAPE), and coverage of the yhat_lower and yhat_upper estimates.

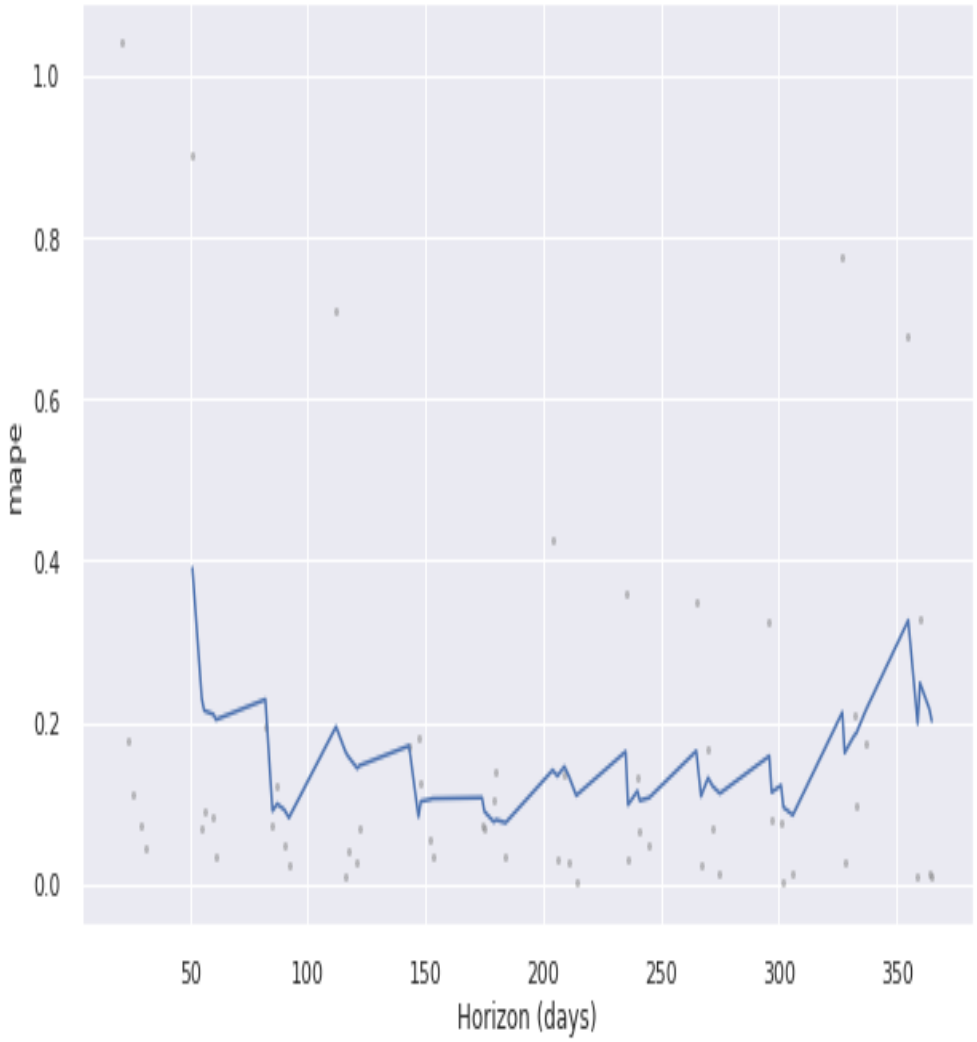

Figure 5.1 MAPE

Cross validation performance metrics can be visualized with plot_cross_validation_metric, as shown for MAPE in figure 5.1. Dots show the absolute percent error for each prediction in forecast. The blue line shows the MAPE, where the mean is taken over a rolling window of the dots.

5.3 TEST CASES

Testing our works is split into three main modules. The three modules are Crime type and location, jewel theft murder crime rate analysis using various factors and forecasting and error metrics. For the above modules and sub modules 16 test cases are passed and verified. For testing our work the test case id, test case scenario, test case secondary considerations, expected outputs, status for the test case and the remarks for the test case are taken into consideration. Test cases are verified and the results, status and remarks for the particular scenarios are processed.

5.3.1 MODULE-1 CRIME TYPE AND LOCATION

For module 1 crime type and location four test cases are passed and tested. In crime type and location test case scenario, data sets are first loaded and all set to pre process for the segregating the crime types and crime location wise data for the jewel theft murder. These are the four scenarios for the first module. The secondary considerations, the expected output, status and the remarks for the respective first module scenarios are for loading huge amount of crime data which forms the crime data collection. And in the expected output all the data sets are loaded and the status is passed, the crime data are collected from various sources. And next is the pre processing where the missing values andnoisy values are checked, in the expected output the unwanted columns are removed and missing values are filled with the zero, then the test case is passed after the removal of all the unwanted datasets and now the data is suitable for analyzing and for the crime rate forecasting.

TC ID	SCENARIO	SECONDARY CONSIDERATIONS	EXPECTED OUTPUT	STATUS	REMARKS
TC01	Load dataset	Huge amount of crime data are loaded for the crime data collection.	All datasets are loaded	Pass	Crime data are collected from various sources.
TC02	Pre-processing	Check for missing values and noisy values	Remove unwanted columns and fill missing values with zero	Pass	Now the data is suitable for analysis and forecasting.
TC03	Crime type	Check for Jewel theft murder Crimes	The dataset contains only jewel theft murder crimes.	Pass	Visualize the crime types in the form of bar graph.
TC04	Location description	Check for Jewel theft murder crime locations	Jewel theft murder Crime locations	Pass	Visualize the location in the form of bar graph.

Table 5.2 test case for module-1 crime type and crime location

The next is the crime type for which the jewel theft murder crimes are checked and the data sets contains only jewel theft murder crimes, in the expected output only the data which contains the jewel theft murder crimes type are verified. The test case is passed now and the crime type which contains all the crime data are visualized in the form of bar graph representation. The last test case is the location description for which the jewel theft murder crime attempted location are checked and from which only the particular locations

aremarked and verified for jewel theft murder crimes andthe status is passed, now the location is visualized in the form of graph. The above table explains the testing procedure for the module 1 crime type and crime location.

5.3.2 MODULE-2 JEWEL THEFT MURDER CRIME RATE ANALYSIS USING VARIOUS FACTORS

For the module 2 jewel theft and murder crime rate analysis using various factors seven test cases are passed and tested. The jewel theft murder crime rate is analyzed using various factors they are year, month of the year, day of the week, week of the month, district, arrest rate and last is the arrest rate by district. The seven scenarios are jewel theft murder crime count by year, the next is the crime count by month of the year, crime count by day of the week, crime count by week of the month, crime count by district, jewel theft murder arrest rate and the last test case scenario is the jewel theft murder arrest rate by district. The secondary considerations, the expected output, status and the remarks for the respective second module scenarios are, the first scenario jewel theft murder count by year for which the module 1 becomes true and the expected output analyses the jewel theft murder crime count per year and the test case is passed and the output is visualized in the form of graph by the crime count by year. The next is the jewel theft murder crime count by month of the year where the secondary considerations are checking for the crime count by month of the year alone. And the expected output analyses the jewel theft murder crime count by the month of the year only and the test case is passed and it is visualized in the form of graph. The next is the crime count by day of the week where the jewel theft murder crime count by day of the week is verified and visualized in the form of graph. And in the expected output only the crime count of the jewel theft murder is analyzed for day of the week. The next is the jewel theft murder crime count by week of the month where the secondary consideration checks for the jewel theft murder crime rate and the

crime count by week of the month and they are represented in the form of bar graph.

TC ID	SCENARIO	SECONDARY CONSIDERATIONS	EXPECTED OUTPUT	STATUS	REMARKS
TC05	Jewel theft murder crime count by year	After Module 1 becoming True	Analyzing the jewel theft murder crime count by year	Pass	Visualizing the crime count by year in the form of graph.
TC06	Jewel theft murder crime count by month of the year	Check for Jewel theft murder crime count by month of the year	Analyzing the jewel theft murder crime count by month of the year	Pass	Visualizing the crime count by month of the year in the form of graph.
TC07	Jewel theft murder crime count by day of the week	Check for Jewel theft murder crime count by day of the week	Analyzing the jewel theft murder crime count by day of the week	Pass	Visualizing the crime count by day of the week in the form of graph.
TC08	Jewel theft murder crime count by week of the month	Check for Jewel theft murder crime count by week of the month	Analyzing the jewel theft murder crime count by week of the month	Pass	Visualizing the crime count by week of the month in the form of graph.

TC ID	SCENARIO	SECONDARY CONSIDERATIONS	EXPECTED OUTPUT	STATUS	REMARKS
TC09	Jewel theft murder crime count by districts	Check for Jewel theft murder crime count by districts	Analyzing the jewel theft murder crime count by districts	Pass	Visualizing the crime count by districts in the form of graph.
TC10	Jewel theft murder arrest rate	Check for Jewel theft murder arrest rate	Analyzing the jewel theft murder arrest rate	Pass	Visualizing the arrest rate in the form of graph.
TC11	Jewel theft murder arrest rate by districts	Check for Jewel theft murder arrest rate by districts	Analyzing the jewel theft murder arrest rate by districts	Pass	Visualizing the arrest rate by districts in the form of graph.

Table 5.3 test case for module-2 jewel theft murder crime rate analysis

The expected output analyses the jewel theft murder crime count by week of the month alone and the test case is passed for this scenario. And now the graph is formed for the visualization of the crime count by day of the week. The table 5.3 represents the module analysis. The next scenario is the jewel theft murder crime count by district where the secondary consideration checks for the crime count by district alone. The test case is passed and the expected output for this scenario is analysis the jewel theft murder crime count by district alone. Further the results are made into a form of graph for the visualization purpose. The next scenario is the jewel theft murder arrest rate, for which the secondary

consideration checks for the arrest rate in jewel theft murder case alone and they are analyzed then the test case is passed and now the arrest rate is visualized in the form of graph. The last scenario is the jewel theft murder arrest rate by district where the secondary consideration checks for the arrest rate only in the particular district where the crime has occurred. Then the test case is passed and then for the visualization purpose the graph is represented only for the arrest rate by district.

5.3.3 MODULE-3 FORECASTING AND ERROR METRICS

TC ID	SCENARIO	SECONDARY CONSIDERATIONS	EXPECTED OUTPUT	STATUS	REMARKS
TC12	Build model	Checking for the crime rate in jewel theft murder forecasting model	Model is built	Pass	Model is ready for forecasting
TC13	Splitting data	80% training data and 20% testing data	Train test split	Pass	Fit the train data into model
TC14	Forecasting	Actual values	Predicted values	Pass	Forecasted value with graph
TC15	Trend change points	Check for trend changes	Trend change	Pass	Trend change points in the form of graph
TC16	Error metrics	Check for errors in the model	RMSE,MSE, MAV values	Pass	Model built is a good fit

Table 5.4 test case for module-3 forecasting and error metrics

The above table 5.4 represents the test cases for the module 3 forecasting and error metrics. Same as module 2 module 3 also has many factors for the forecasting and error metrics, they are Forecasting the crime rate for jewel theft murder, forecasting components, trend change points for jewel theft murder and last the error metrics. There are five scenarios for these factors those are to build the model and next is the splitting the data and next is forecasting the model and next is the trend change points and last scenario is the error metrics.

The secondary considerations and the expected output and status and the remarks for the respective second module scenarios are, for the first scenario building the model and the secondary consideration checks the crime rates in the jewel theft murder forecasting model. The expected output forms a building of the model for the crime in jewel theft murder. Then the test case is passed then at last the model is build and ready for forecasting. The second scenario is the splitting the data where the secondary consideration splits the 80% for the training data and rest of the 20% for the testing data. The expected output trains and tests the data and then splits. The test case is passed and the model is fit and the data is trained into the model.

The next scenario is the forecasting for which the actual values of data are tested and trained. And the secondary consideration is the actual values, from which the data are tested and trained. The output becomes the predictive values and the test case is passed then the graph is visualized for the forecasted values. The next scenario is the trend change points where the secondary consideration data is checked and the trends points are plotted in the form of graph for the visualization purpose. The last scenario is the error metrics where the secondary consideration checks for the errors in the model and the expected output values are the RMSE, MSE and MAV values.Then the test case is passed and the model is built which proves to be a good fit model after all the test cases are verified.

5.3.4 MODULES TESTED

1. Crime type and location
2. Jewel theft murder crime rate analysis using various factors
3. Forecasting and error metrics

Each modules includes the sub module function wherethe first module represents crime type and location description. Where the test case are checked for the crime scene in jewel theft murder. The second module shows the analysis of jewel theft murder crime rate using various factors like year, day of the week, month of the year, week of the month, districts and sub module includes arrest rate and arrest rate by districts. The third module represents the forecasting model, changes in trend and error metrics.

Module 1 - Crime type and location

Module 2 - Jewel theft murder crime rate analysis

- Crime count by year
- Crime count by month of the year
- Crime count by day of the week
- Crime count by week of the month
- Crime count by district
- Arrest rate
- Arrest rate by district

Module 3 - Forecasting

- Forecasting the crime rate for jewel theft murder
- Forecasting components
- Trend change points for jewel theft murder
- Error metrics

5.4 SUMMARY

The system is tested with various test cases. There are total 16 test cases used in the system. Each modules represents the process involved to achieve the forecasting model. Module 1 shows the test cases involved in the preparation of datasets. Module 2 shows the test cases involved in the analysis process and module 3 shows the test cases involved in the building of model and forecasting. The last test case shows the error metrics which is used to evaluate the model performance. The development process in this context refers to the workflow framing, coding and integrating the code into the modules considered. Errors are liable to happen but, there must always be a proper check to encounter these errors. The test cases that were considered also give a brief idea on how the system behaves. If any deviation occurs in the normal behavior, it marks the inefficiency of the system.The overall motive of testing is to build a forecasting model which proves as a good fit and produce high quality results.