

Contents

Problem Description.....	2
Main Tasks:.....	2
Methodologies:.....	2
The datasets are provided as cited below for the Retail Sales Forecasting:.....	3
Submission 1 (18th Nov 2017 EOD) : Time Series Model Building for all product categories.....	4
Time Series Model.....	4
Documents, to be submitted to Grader tool:.....	4
Submission 2 (19th Nov 2017 EOD) : Regression Model Building for the ProductCategory “WomenClothing” :.....	4
Documents, to be submitted to Grader tool:.....	5
Error Metrics:.....	5
Hints to Proceed:.....	5
Submission 3 (24th Nov 2017 EOD) :.....	6
Documents, to be submitted to Grader tool:.....	6
Final documentation with additional efforts and improvements made during the week using the prescribed template. The document should consist of the following aspects also:.....	6
Final commented code.....	6
Final Prediction Results.....	6
Final presentation for viva with the name “viva.ppt”	6

Sales Forecasting of Retail Clothing Product Categories

Problem Description

Forecasting is an important approach to plan the future effectively and efficiently. A time series is a sequence of data points, typically consisting of successive measurements made over a uniform time interval. Time series forecasting is the use of a model to predict future values based on previously observed values.

A leading retailer in USA, wants to forecast sales for their product categories in their store based on the sales history of each category. Sales forecast has very high influence on the performance of the company's business and hence these sales forecasts can be used to estimate company's success or performance in the coming year. Accurate forecasts may lead to better decisions in business.

Sales or revenues forecasting is very important for retail operations. Forecasting of retail sales helps retailer to take necessary measures to plan their budgets or investments in a period (monthly, yearly) among different product categories like women clothing, men clothing and other clothing and at the same time they can plan to minimize revenue loss from unavailability of products by investing accordingly.

Main Tasks:

1. **Target attribute is "Sales(In ThousandDollars)"**
2. You are expected to build a framework that provides monthly forecasts of the next 12 months **for each product category** (3 product categories) with past sales history using time series model.
3. You are expected to build a framework that provides monthly forecasts of the next 12 months **for "WomenClothing" product category alone with other causal attributes** using regression model.
 - a. Compare the results obtained from time series model and regression model for **"WomenClothing" product category and analyse the impact of causal variables.**

Methodologies:

The forecasting is expected to be made for next 12 months ie., from Jan 2016 to Dec 2106 using Time series model and also regression model as cited below:

1. Based on sales history using time series model for each product category
2. Based on other various influencing factors of the sales such as Holiday events, weather changes, macro economic factors etc. using the regression model for **“WomenClothing” product category alone.**

The datasets are provided as cited below for the Retail Sales Forecasting:

1. Sales data: “Train.csv”

1. This table contains the Temporal data like Year, Month, and Product category and Sales (In ThousandDollars), provided for the period 2009 to 2015

2. Weather data: “WeatherData.xlsx”

2. This is obtained for the period from 2009 to 2016 from different sources and has the information about the weather conditions like temperature, wind speed, humidity, visibility, precipitation etc.

3. Event/Holidays data: “Events_HolidaysData.xlsx”

3. It's obvious that we shop more during the holidays or during festivals etc. Hence, the event data for the period from 2009 to 2016 has the information like Year, Month-Date, Event, Day category(federal holiday or event)

4. Macro Economic Data: “MacroEconomicData.xlsx”

4. This data is provided for the period from 2009 to 2016 with the details like CPI, GDP, Cotton production, mill usage, unemployment rate etc.

5. Attributes Details : “AttributesDescription.xlsx”

5. This has the details of attributes for the datasets cited above (1 to 4)

Using the data given, you must come up with the model/models that can forecast the monthly sales for 2016

Note: Missing values are denoted as “”, “NA”, “?” in the datasets
In WeatherData.xlsx file, the blanks in the attribute “WeatherEvent”
should be considered as not applicable but not the missing values

Submission 1 (18th Nov 2017 EOD) : Time Series Model Building for all product categories

Time Series Model

- o The initial sales data given is monthly sales data by product category. You have to make month level forecasts for the year 2016 (Jan 2016 to Dec 2016) using time series models for each product category with the past historical data provided for the period from 2009 to 2015.

Use Sales history (“Train.csv”) and Report on

- pre-processing,
- problem understanding,
- Exploratory Data Analysis,
- building the time series model using the past sales history and
- sales forecasting results

using the time series model for each product category in R Notebook or Jupiter notebook format.

Documents, to be submitted to Grader tool:

- a. Commented code of Time series Model with the name **“submission.R”** or **“submission.ipynb”** as the case may be.
- b. Upload the sales forecasting obtained for Jan 2016 to Dec 2016 from time series model to Grader tool with the name **“prediction.csv”**.

Submission 2 (19th Nov 2017 EOD) : Regression Model Building for the ProductCategory “WomenClothing”:

The time series data (“Train.csv”) and causal attributes datasets along with Train.csv (“WeatherData.xlsx”, “Events_HolidaysData.xlsx”, “MacroEconomicData.xlsx”) -----should be considered for regression analysis and forecasting the sales for next 12 months ie., for monthly forecasting of 2016

Regression Model Approach

- **Regression Model:**
 - o **Details of Data:**
 - Consider the data related to the period from 2009 to 2015 for building the required framework.

- ❖ Train Data : Consider the data provided for the period from 2009 to 2015

- ❖ Test Data : Test set would be of Jan 2016 to Dec 2016 and the predictions to be made for next 12 months., ie., for Jan to Dec for 2016

Note: The datasets of weather, events & macroeconomic data are provided for the period from 2009 to 2016. You have to use the data for 2009 to 2015 for train data and 2016 data for test data.

- The data, for other causal attributes, is extracted from different sources and it is raw data (mostly daily observations) and hence to be converted into monthly aggregated data to make month level forecasts using regression models.

- o **Data Aggregation:**

- The weather data is provided day-wise and Event/Holidays data is provided with the details of individual dates. Of course, the macro economic data is by month. Hence, these datasets should be aggregated and to be converted as month-wise data so that forecasting can be done for next 12 months.

- o **Feature engineering & Feature selection :**

- It is very important aspect with which the model accuracy can be improved well. It is recommended that emphasis be given on feature engineering and feature selection during the preprocessing to obtain the right set of features to develop the models.

- **Think about how to split your data into train and validation.**

Documents, to be submitted to Grader tool:

- a) Commented code of Regression Model with the name **“submission.R” or “submission.ipynb”** as the case may be.
- b) Upload the sales forecasting obtained for Jan 2016 to Dec 2016 to Grader tool with the name **“prediction.csv”**.

Error Metrics:

- o Consider MAPE as error metric and tune the model accordingly.

Hints to Proceed:

1. Do the preliminary and exploratory analysis of time series data

2. Understand the following aspects:
 - Is there a significant trend?
 - Is there a significant seasonality and is the seasonality important?
 - Is there evidence of the presence of business cycles?
 - Are there any outliers in the data that need to be explained?
 - How strong are the relationships among the variables available for analysis?
3. Build the time series model.
4. Estimate the parameters of the chosen time series model
5. Present your analysis for deciding on the appropriate model
6. Elaborate on difficulties and alternative approaches
7. Whenever necessary, transform the data into a time series that appears to be a realization of a stationary time series model.

Note: The hints provided are just high level only, you please use your knowledge and explore further to come-up with proper approach and conclude the models and results.

Submission 3 (24th Nov 2017 EOD) :

Documents, to be submitted to Grader tool:

Final documentation with additional efforts and improvements made during the week using the prescribed template. The document should consist of the following aspects also:

- a. Explain why you have chosen your model as the best model for Time series
- b. Explain why you have chosen your model as the best model for Regression
- c. Comparative study, of the results obtained from time series model and regression model, for the ProductCategory "WomenClothing" and also give a report the impact of causal variables on the results.

Final commented code

Final Prediction Results

Final presentation for viva with the name "viva.ppt"

Note: Please follow the naming convention for the submission files as cited above only.