SALES FORECAST AND KPI PREDICTION

Final AI project from SW800 – MAY/2018 batch

PROJECT DESIGN PRESENTATION

Course Coordinator: Vimala Mathew, Scientist/Engineer 'D'
Team Members: Aiswarya , Aslam Bughari, Chandru, Civaprasad, Hani Esmayil, Salman Faris
Batch: Advanced Diploma in Artificial Intelligence (SW800 – MAY/2018)

Introduction

- ***Connect*** is an IT infrastructure, networking, hardware and software supply chain organization with **5 stores** within GCC. We have collected **7 years** of their sales data and trained ML with one year data.
- The first **challenge** is **predicting the sales** for Connect
- An ensemble learning technique known as **Random Forest Regression** is used to predict the sales data
- The manager of each store has been tasked to **predict the sales** of the store for **up to six weeks in advance**.
- The actual sales history and the complete sales data from the stores are used **to predict the future sales of each store.**

Tools

DATA STRUCTURE

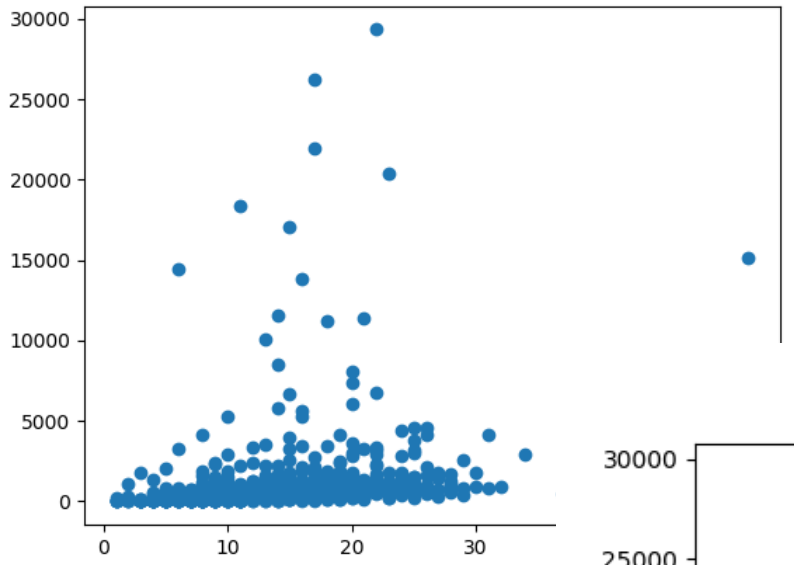
COLUMN NAME	TYPE	DESCRIPTION
TYear	Categorical	The year of the transactions
TMonth	Categorical	The month of transaction
SALESMAN_CODE	Continous	Unique ID assigned to each Salesman
CATEGORY_CODE	Continuous	Unique ID assigned to each product category
ActualSales	Numerical	The historical sales till date
ActualReturn	Numerical	The product returned to store on expiry, damaged for a period
TargetSales	Numerical	The targeted sales of the company to a salesman
TargetReturnPercent	Numerical	The percentage on sold product that was returned
TargetCustVisit	Numerical	The targeted visits by customer
ActInvoiceCustVist	Numerical	Historical data of customer visits
ActualCashSales	Numerical	Historical Sales By cash
TargetCashSales	Numerical	Targeted sales by cash
ActualCreditCollection	Numerical	Historical sales by credit
TargetCreditCollection	Numerical	Targeted sales by credit
ActualTotalCollection	Numerical	Historical total sales
TargetCollection	Numerical	Targeted total sales
KPI	Numerical/Dependent	The commission for the salesman
Branch	Categorical	The branch number
TDayofweek	Numerical	The day of the week in numbers in 1 to 7
StateHoliday	Numerical 1 or 0	State holiday or not in 1 or 0
Customer	Numerical	The customer that has made a purchase
Topen	Numerical 1 or 0	Opened days in 1 or 0
Dates	Numerical-Date	The date the sales occurred in a numerical DATE data

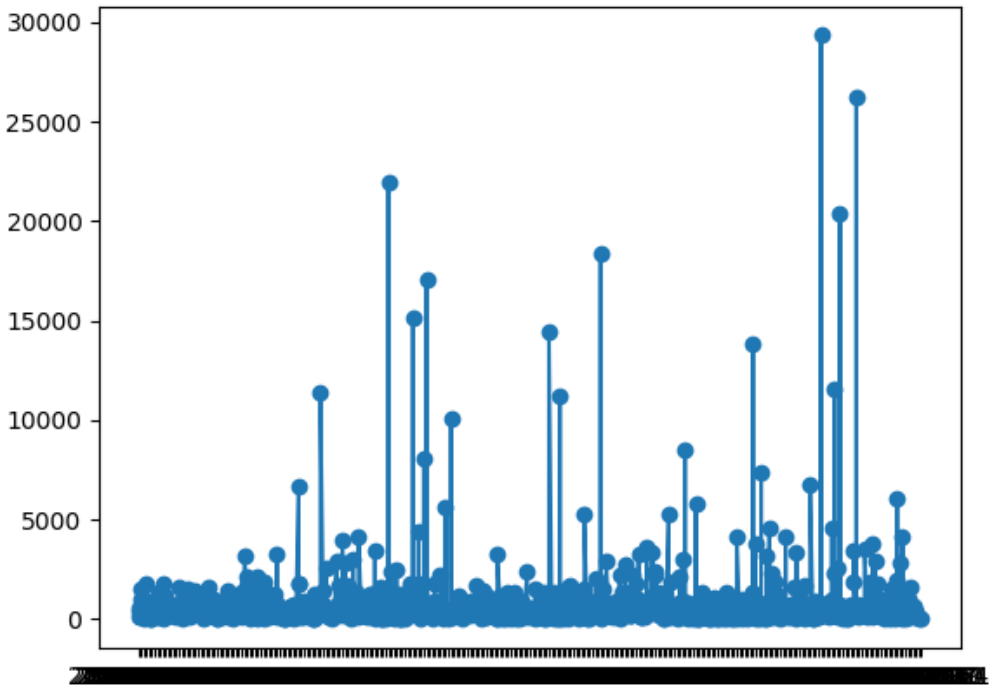
TRAIN DATA

	A	B	C	D	E	F	G	H	I
1	Branch	TDayOfWe	TDate	Sales	Customer	TOpen	Promo	StateHolic	TWeekend
2	720	7	25-Jun-17	0	0	0	0	1	0
3	720	7	26-Jun-17	0	0	0	0	1	0
4	720	7	27-Jun-17	0	0	0	0	1	0
5	720	7	28-Jun-17	0	0	0	0	1	0
6	720	7	2-Jul-17	139	8	0	1	1	0
7	720	3	2-Aug-17	1080	11	0	1	1	0
8	720	3	2-Aug-17	0	0	0	1	1	0
9	720	3	2-Sep-17	0	0	0	1	1	0
10	720	3	3-Sep-17	0	0	0	1	1	0
11	720	3	4-Sep-17	0	0	0	1	1	0
12	720	3	5-Sep-17	0	0	0	1	1	0
13	720	7	########	11	9	0	1	1	0
14	KHS	7	25-Jun-17	0	0	0	0	1	0
15	KHS	7	26-Jun-17	0	0	0	0	1	0
16	KHS	7	27-Jun-17	0	0	0	0	1	0
17	KHS	7	28-Jun-17	0	0	0	0	1	0
18	KHS	7	2-Jul-17	139	8	0	1	1	0
19	KHS	3	2-Aug-17	0	0	0	1	1	0
20	KHS	3	2-Aug-17	0	0	0	1	1	0
21	KHS	3	2-Sep-17	0	0	0	1	1	0
22	KHS	3	3-Sep-17	0	0	0	1	1	0
23	KHS	3	4-Sep-17	0	0	0	1	1	0
24	KHS	3	5-Sep-17	0	0	0	1	1	0
25	KHS	7	########	213	17	0	1	1	0

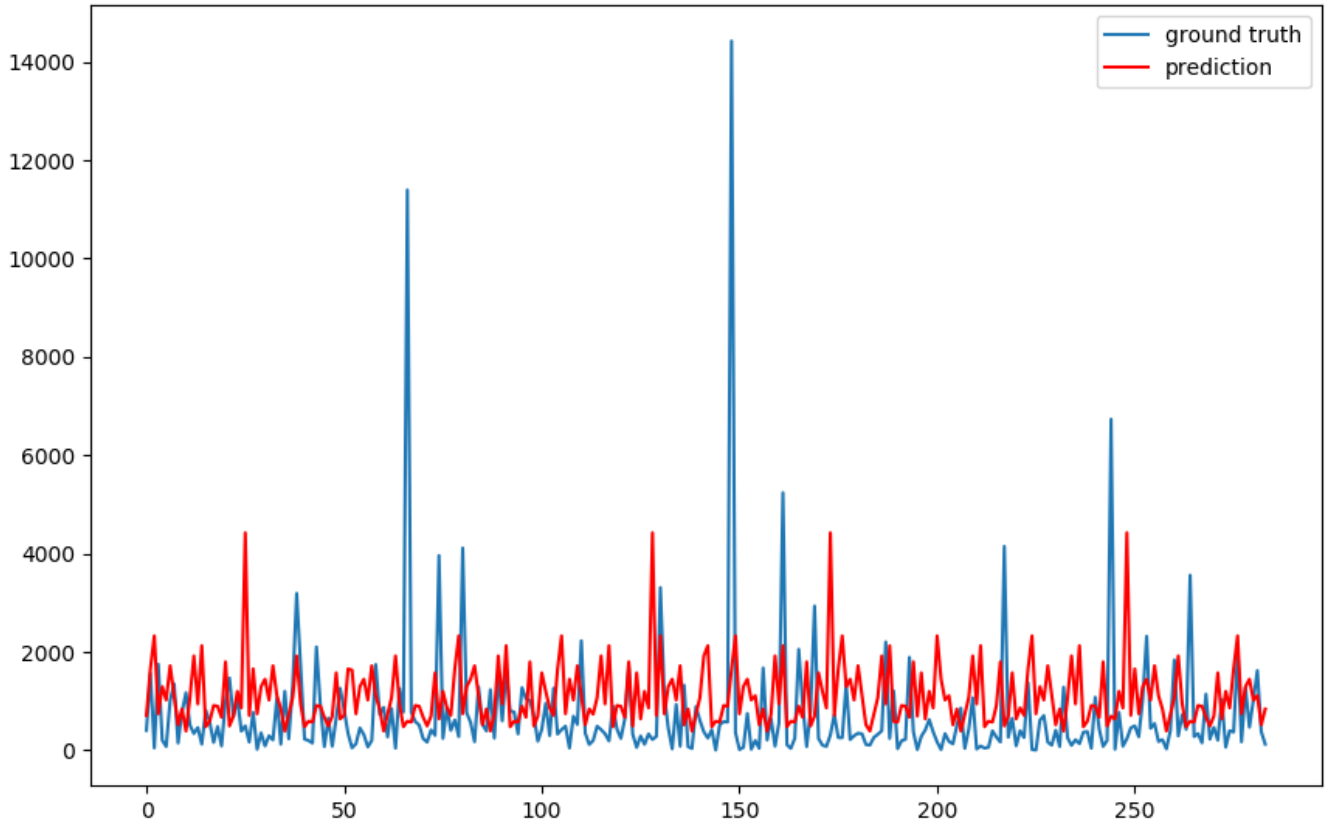
Customer Vs Sales

Sales Vs Date





Actual Vs Prediction on Train Data



Random Forest

- Now with the training data we created in the last step, we can finally **apply the Random Forest algorithm**.
- Random Forest is an **ensemble learning** technique that constructs multiple decision trees via randomization.
- The '-Attrs' flag defines what variables are categorical and which are continuous, with variables with the option C being categorical and those with an option Q being continuous. Only the column competition distance is a continuous variable.

ARIMA

(Autoregressive Integrated Moving Average Model)

- An ARIMA model **is a class of statistical models** for analyzing and forecasting **time series** data. It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful **time series forecasts.**
- It is a generalization of the simpler **Auto Regressive Moving Average** and adds the notion of integration. This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:
  - **AR**: Auto regression. A model that uses the dependent relationship between an observation and some number of lagged observations.
  - **I:** Integrated. The use of differencing of raw observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.
  - **MA**: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Prediction

- Next, we'll make a prediction using the model we created. The response variable at the time of learning is **LN(1 + t1.sales)** after converting the scale, so the reverse conversion would be **EXP(predicted-1).**
- The system **predicts** the sales for **upcoming 20 days** taking the testing data as the data for the first 10 days of sale.

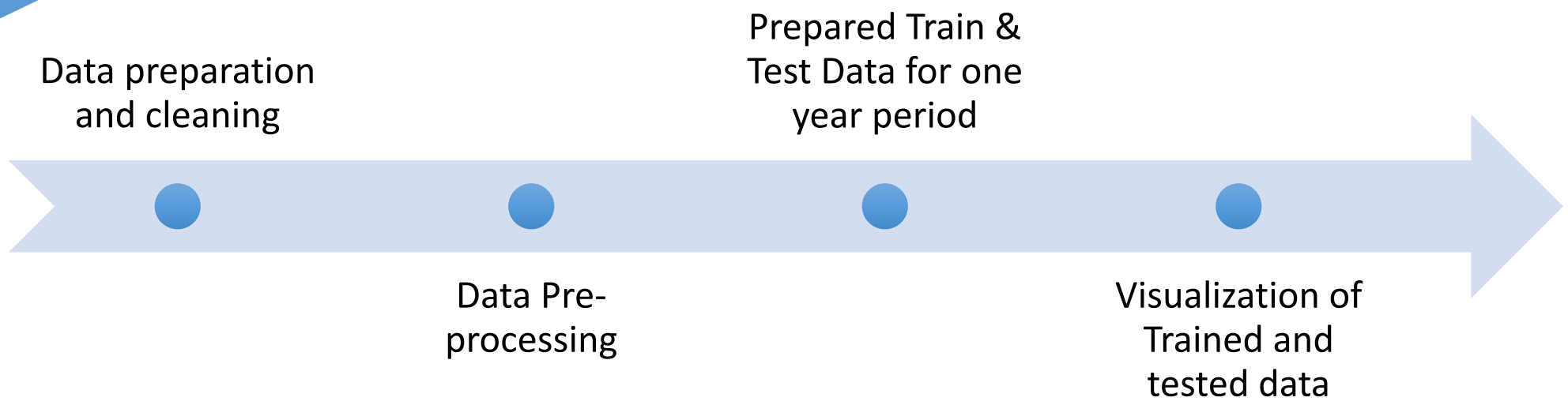
Evaluation

To evaluate the strength of our predictions, we will use the following equation:

Where y_i is the sales of that the *i*th day for that store, and $\hat{y}_i$ is the predicted value.

$$RootMeanSquarePercentageError = \sqrt{\frac{1}{n}\sum_{I=1}^{n}\left(\frac{Y_i - \hat{Y}_i}{y_i}\right)^2}$$

Where are we now



Expected Deliverables

Correlation between data

Forecasting of Sales against category

Forecasting of Sales against Salesman

Prediction of Sales for a salesman for a specific period

KPI Analysis

KPI Prediction

Thank you