Leveraging Machine Learning for Competitive Advantage

Dr. Rachna Dhand

BHP
Senior Data Scientist Cybernetics

ASX Illuminate, 2017

If you can’t explain it simply, you don’t understand it well enough.
Outline

1. What is Machine Learning?
2. Case Study 1. Time Series Forecasting: Sales Data (2016)
What is Machine Learning?

Machine Learning (ML) is the collection of statistical algorithms that have the ability to learn from the data and make future predictions.

- **Statistical Methods**
  - Statistical Algorithms that bind data & response into some form of Likelihood Function that can be used to predict the response.

- **Algorithms**
  - Target of prediction is known and Model is trained on historical data.

- **Supervised Learning**
  - Target depicts unknown behavior. Used in Pattern Detection.

- **Unsupervised Learning**

"Artificial Intelligence is not a Man versus Machine saga; it’s in fact, Man with Machine synergy."
Sudipto Gosh
Supervised Learning – Process Flow

1. Raw Data
2. Transformations
3. Partioning
4. Model Building
5. Validation on Test Data
6. Predictions

- Scaling, Outlier Detection and Target Definition
- Sampling & Segregation into Training & Testing Partitions
- Modelling the Training Dataset: Classification or Regression
- Validation Metrics: RMSE, R Square or ROC, Confusion Matrix

Machine Learning is all about Problem Solving!
Machine Learning - Layers

- Theoretical Distributions
- Classification & Regression
- Clustering & Segmentation
- AI & Web Service for Knowledge Consumption

**Machine Learning Algorithm**

- Extracting Insights
- Identifying Patterns
- Making Predictions
- Visualization & Intervention

Leveraging Machine Learning for Competitive Advantage
Case Study 2 – Optimizing Returns on Newspaper Publishing

Problem
A problem that most newspaper companies encounter daily is how to predict the right number of newspapers to print and distribute among distinct sales points. The aim is to predict newspaper demand as accurately as possible to meet customer need with minimum number of returns, missed sales and oversupply. The Newspaper Team seeks to optimise the publication quantities in order to maximise opportunities for sales and minimise returns. They are determined by Routes or retail outlets.

Objective
- Forecast the Sales per Route to minimise the return/Sells-Out per Route on 14-Day cycle.
- Optimise Returns to on Forecasted Sales for each Route as per suggested bands.

Methodology
To achieve this objective the project was designed in two phases.
Phase 1
Forecast Net Sales for each Route (3200 Routes) based on the Draw total and external factors like week number, week day, public holidays, school holidays and interventions of high predictive value.

Phase 2
Generate Predicted Returns with objective function of minimising returns per route, looking at the forecasted Sales for Routes a fortnight in advance.

“Statistics are ubiquitous in life, and so should be statistical reasoning.”
Alan Blinder
Methodology – Optimizing Returns

The publishing house had approximately 3200 Routes that need to be forecasted for sales and optimized for Returns. The first step was Clustering them into appropriate groups for Modelling (Newsagents, Hotels, Schools, Deli, or Supermarkets).

The second step was to identify the co-variates that impacted individual clusters & hence define them within the data (Public Holidays, School Holidays, Weekdays, Weekends, Australia Day, Election Day).

The third step was to model the clusters for forecasting Sales for all outlets. Seasonal ARIMA Modelling was applied to all outlets within each cluster to forecast the fortnight ahead sales.

Finally, the forecasted sales were optimized using power function to predict returns for each outlet, given the return bands.
ARIMA models are, in theory, the most general class of models for forecasting a time series which can be made to be “stationary” by differencing (if necessary), perhaps in conjunction with nonlinear transformations such as lagging or deflating (if necessary).

A random variable that is a time series is stationary if its statistical properties are all constant over time. A stationary series has no trend, its variations around its mean have a constant amplitude, and it wiggles in a consistent fashion, i.e., its short-term random time patterns always look the same in a statistical sense. The latter condition means that its autocorrelations (correlations with its own prior deviations from the mean) remain constant over time, or equivalently, that its power spectrum remains constant over time.

ARIMA means Auto-Regressive Integrated Moving Average. Lags of the stationary series in the forecasting equation are called "autoregressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

1. p is the number of autoregressive terms,
2. d is the number of nonseasonal differences needed for stationarity, and
3. q is the number of lagged forecast errors in the prediction equation.

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.
Idea Concept – Data Streamlining

- Sales Data (2013 – 2015) excluding Sundays
- Clustering the Routes (Newsagents, Hotels, Petrol,...)
- Re-Structuring the Data (Adding Public Holiday, School Holiday, Special Events Days, Week Number Indexes)

- Modelling the Clusters for each Route with it’s ARIMA Forecast (Parallelizing the ARIMA building for 9 Clusters, 3200 Routes)
- Formulating the Medium Expected Returns Optimisation Formula
  \[ Y = 1.05x^{0.5} \]

Dynamic variation & Seasonal Sales Trends captured with Fourier Terms (On Public Holidays, School Holidays, Special Event Days in a year)
Net Sales Forecasting: Phase 1
Initial Modelling Description
ARIMA Seasonal (ACF and PACF)

Leveraging Machine Learning for Competitive Advantage
Net Sales Forecasting: Phase 1
Modelling Description

Execution Flow: Every Data Science problem encounters a unique flow of computation while Modelling and Scoring.

For the Publishing house, the challenge was

a. to incorporate the dynamic impact of all the covariates like Public Holidays, School Holidays, Weekends and Week days & special Event days.
b. model each Route with a separate ARIMA Model and total number of Routes were 3200 within 9 Clusters.
c. optimize the computation time for Models for each Route as each ARIMA cluster computation took 6 hours to 19 hours depending on the number of Fourier terms used in each cluster.

Parallelized the computation of ARIMA equation for Each Cluster over 20 cores

Execution Time for Computation was reduced to 20 minutes to 30 minutes for each cluster

“Errors using inadequate data are much less than those using no data at all.” Charles Babbage
Net Sales Forecasting: Phase 1
Initial Modelling Results
14-Day Forecasts

Forecasts from ARIMA(3,1,1)(1,0,1)[6]

Forecasts from ARIMA(1,1,1)(2,0,1)[6]

Leveraging Machine Learning for Competitive Advantage
Net Sales Forecasting: Phase 1
Initial Modelling Description
ARIMA Seasonal (Super Market Routes)
Net Sales Forecasting: Phase 1
Modelling Results (Super Markets)
Routes: 00014, 00072

Forecasts from ARIMA(5,1,0)(2,0,0)[5]

Forecasts from ARIMA(5,1,0)(2,0,0)[5]

Leveraging Machine Learning for Competitive Advantage
Net Sales Forecasting: Phase 1
Modelling Results
For AIR North West routes

<table>
<thead>
<tr>
<th>Routes</th>
<th>ME</th>
<th>RMSE</th>
<th>MAE</th>
<th>MASE</th>
<th>ACF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Sal_07432R01</td>
<td>-0.2535</td>
<td>8.670248</td>
<td>4.968498</td>
<td>0.8631931</td>
<td>-0.004641654</td>
</tr>
<tr>
<td>2 Sal_07436R01</td>
<td>-1.563</td>
<td>34.09363</td>
<td>24.36629</td>
<td>0.7568571</td>
<td>-0.001620553</td>
</tr>
<tr>
<td>3 Sal_07437R01</td>
<td>-0.0209</td>
<td>8.902918</td>
<td>5.424503</td>
<td>1.109197</td>
<td>-0.04309073</td>
</tr>
<tr>
<td>4 Sal_07438R01</td>
<td>-0.0589</td>
<td>8.401815</td>
<td>5.17907</td>
<td>1.073647</td>
<td>-0.09545113</td>
</tr>
<tr>
<td>5 Sal_07439R01</td>
<td>0.0128</td>
<td>3.677939</td>
<td>2.467908</td>
<td>0.9032865</td>
<td>0.06744089</td>
</tr>
<tr>
<td>6 Sal_07440R01</td>
<td>-0.0216</td>
<td>11.64424</td>
<td>7.573748</td>
<td>1.143413</td>
<td>-0.0001654223</td>
</tr>
</tbody>
</table>

MAE is the Mean of Residuals for the fit (Mean = \( |e_i| \))

RMSE is the root of square of the Mean of the Residuals of the fit (RMSE = \( \sqrt{\text{mean}(e^2)} \))

MASE The mean absolute scaled error is simply

\[ \text{MASE} = \text{mean}\left(|q_i|\right). \]

Similarly, the mean squared scaled error (MSSE) can be defined where the errors (on the training data and test data) are squared instead of using absolute values.

“If you can’t explain it simply, you don’t understand it well enough.” Albert Einstein
After forecasting the net sales, the firm determines expected return for the outlet according to its policy. It has three types of expected return quantities for forecasted net sales intervals. These are best, medium and worse return quantities. Before determining the circulation, one of the types is selected according to its policy.

Finally the Circulation report holds the expected Circulation quantities for each Route a fortnight advance. The formula is

\[
\text{Circulation} : \text{Forecasted Net Sales} - \text{Expected Medium Return}
\]

"It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers", Alan Turing in 1951
Do you have any questions?