Implementation of a Mobile Alert System for Sales Forecasting Based on a Tree Augmented Naïve Bayes Classifier

Claire Gallagher

M.Sc. Software Design and Development, College of Engineering & Informatics, NUI Galway
Enterprise Services IT Business Intelligence, HP Galway
Introduction

Background

Objectives and Goals

Approach

Methodology

Results

Future Work
Background

HP Enterprise Services - Big Data, Cloud, Mobility, Security

Project Aim:
Create a sales analytics and alerting solution presented through a mobile app

1) Identify deals with potential to Win that existing model predicted as a Loss
2) Identify deals at risk of Loss that existing model predicted as a Win

Goals:
• Apply predictive analytics to HP sales data using a Bayesian classification algorithm to boost current sales predictions
• Present opportunity predictions through a mobile app and provide key alerts to the sales force based on sales opportunities that may be incorrectly classified

More efficient sales pipeline
Accelerate sales force productivity
Increased sales
Maximise business returns
Employee specific alerts
Approach Diagram

**INPUTS**

- Qualitative Sales Data
- Quantitative Sales Data
- Historical Sales Data + Actual Outcome (Win/Loss)

**ANALYTICS**

- Weighting Function
- Classifier
- Classification Algorithm

**OUTPUTS**

- Win
- Borderline Loss
- Loss

Difference in predicted outcome = ALERT!
Methodology

- Literature review and research
- Determine the most appropriate and accurate classification algorithm
- Analyse sales data and significant attributes
- Train, test, and validate Tree Augmented Naïve Bayes classifier
- Evaluate results and accuracy

Technologies Used

- R programming
- ‘bnlearn’ package in R
- SQL Server
- QlikView
- HP Vertica
- D3, JavaScript, HTML
Tree Augmented Naïve Bayes Classifier

Bayesian Classifiers
- Probabilistic classifier
- Accounts for previous evidence
- Outputs probability of success

Naïve Bayes
- Strong independence assumptions between features
- Simple fixed structure

Tree Augmented Naïve Bayes
- Stems from Naïve Bayes
- Interdependencies allowed between factors

TAN
- Predicts "Status" of opportunity based on other influencing factors
- Each attribute depends on the class attribute and one other (allows for dependencies between attributes)
- Win Probability for each Opportunity
- Categorised into:
  - Win: > 80%
  - Borderline Loss: 40-80%
  - Loss: < 40%

M. G. Madden (2009)
Sales Data

<table>
<thead>
<tr>
<th>TRAINING</th>
<th>TESTING</th>
<th>VALIDATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical Sales Data with Actual outcome of opportunity – Won/Lost Jan 2010 – July 2014 Deals with existing model prediction included</td>
<td>• Test on training data • Training: testing – 70:30 • 10 fold cross validation</td>
<td>• Open opportunities • August 2014 – January 2015 • Not used for training model in any way • Prediction compared to Actual outcome for evaluation of model • Compared to existing model predictions</td>
</tr>
</tbody>
</table>
Learning Curve Analysis

Repeated cross validation with increasing training dataset size (beginning with most recent data)

Depicts the evolution of learning performance (classification accuracy) as a function of training set size
ROC Curve Analysis

True positive rate (Sensitivity) vs. False negative rate (1-Specificity)

High Area Under Cure (AUC) = High Accuracy

Used for analysis of binary categorical threshold
Confusion Matrix

True Positive Rate (TPR): % of deals that were correctly predicted as a win.

False Positive Rate (FPR): % of deals that were incorrectly predicted as a win.

True Negative Rate (TNR): % of deals that were correctly predicted as a loss.

False Negative Rate (FNR): % of deals that were incorrectly predicted as a loss.

<table>
<thead>
<tr>
<th>Prediction Positive</th>
<th>Actual Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE POSITIVES</td>
<td>WIN 1127</td>
</tr>
<tr>
<td>FALSE POSITIVES</td>
<td>LOSS 117</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction Borderline Loss</th>
<th>Actual Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>BORDERLINE LOSS WINS</td>
<td>88</td>
</tr>
<tr>
<td>BORDERLINE LOSS LOSSES</td>
<td>98</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction Negative</th>
<th>Actual Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>FALSE NEGATIVES</td>
<td>LOSS 122</td>
</tr>
<tr>
<td>TRUE NEGATIVES</td>
<td>1198</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TAN Classifier Predictions</th>
<th>Total Population 2540</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR</td>
<td>81.37%</td>
</tr>
<tr>
<td>TNR</td>
<td>84.78%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>83.27%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Naïve Bayes</th>
<th>Total Population 2540</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR</td>
<td>89.44%</td>
</tr>
<tr>
<td>TNR</td>
<td>77.64%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>82.87%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Existing Predictive Model</th>
<th>Total Population 2540</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR</td>
<td>87.93%</td>
</tr>
<tr>
<td>TNR</td>
<td>29.44%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>55.39%</td>
</tr>
</tbody>
</table>

Accuracy: TP + TN / Total Population
Mobile App Alerting System

Bar chart: Win /Borderline Loss / Loss comparison of both prediction models

Heatmap: Opportunities predicted differently by Bayesian model are highlighted

Opportunity ID: 232425
Name: Acme Ltd
Deal at Risk of Loss due to New Deal Type and High Contract Value
Future Work

Development of mobile app page

Variable influence analysis

SQL Server $\rightarrow$ HP Vertica

Automate data feeds and process

Integrate into existing ES Sales dashboards and mobile app
Summary

Developed a Tree Augmented Naïve Bayes model for Sales Forecasting

TAN Accuracy significantly improves upon HP’s existing qualitative model

Combination of qualitative and quantitative sales data

Model is applied to drive a mobile app alerting system
Key References


Thank you!

Any Questions?