Predicting On-time Delivery in the Trucking Industry

Authors: Rafael Duarte Alcoba, Kenneth W. Ohlund
Advisor: Matthias Winkenbach
Agenda

- Motivation
- Methodology
- Results
- Conclusion
The US trucking industry...

Dominates the commercial transportation industry with 83.7% of the revenue

- 1.4% WATER
- 1.5% RAIL INTERMODAL
- 3.2% AIR
- 4.6% PIPELINE
- 5.6% RAIL
- 83.7% TRUCK

Connects the entire US territory

Is expected to grow 21% over the next 10 years

1.4% WATER
1.5% RAIL INTERMODAL
3.2% AIR
4.6% PIPELINE
5.6% RAIL
83.7% TRUCK
Research Questions

- How can companies engaged in logistics optimize resources while improving customer service levels?
- Can on-time delivery in trucking be predicted?
- Can a predictive analytics model indicate which combinations of variables lead to delays?
Gathering Data

- Loads within the United States (more than 6,000 locations)
- Restricted to FTL (full truckload)
- Data from October 1, 2014 to September 30, 2016
- Binary decision variable for on-time delivery (0 = delayed; 1 = on-time)
Fishbone Diagram

Variables Potentially Affecting On-time Delivery

LOAD
- Commodity
- Weight
- Mode/Equipment
- Contract vs Spot
- Industry (dairy, paper…)
- Team vs Ind. Driver
- High risk / High Value

LANE
- Geography
- Distance
- Weather
- # of stops

CARRIER
- Size
- History w/ Coyote
- Safety rating
- CSA Score
- Tracking Method

PROCESS
- Tender Lead Time
- Appt Scheduling
- Bounces
- Incidents
- Pickup Time (buckets)
- Hours of Operation
- Origin Facility

OPERATIONS
- Arrived Time Pickup
- Departed Time Pickup
- # of stops
- Ops Team
- Tenure

FACILITY
- Late dispatch
- Ontime Pickup
- Days of the week
- Season
- Hours of Operation
- Origin Facility
- Destination Facility

On Time Delivery
Sampling & Partitioning

On-Time Delivery

Data

- Imbalanced
  - 95% on-time
  - 5% delayed

Model

- Overfitting
  - Rule of Thumb
    - 50% on-time
    - 50% delayed

- 75% Training
- 25% Validation

Undersampling
Model Selection

- **Response**
  - Categorical (0 or 1)
  - Continuous

- **Predictors**
  - Categorical

- **Logistic Regression**
- **Neural Network**
- **Bootstrap Forest**

- **Goal**: find an explanatory model with high interpretability
- **Main model**: LR
- **Assess Performance**: NN and BF
Variable Selection

Multi-Collinearity

Correlation Matrix

<table>
<thead>
<tr>
<th>Correlations</th>
<th>Contract-Spot</th>
<th>Duration at StartSegment</th>
<th>Historical Volume</th>
<th>Incidents per Volume</th>
<th>OnTime of StartSegment</th>
<th>Facility Type Appt of EndSegment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contract-Spot</td>
<td>1.00</td>
<td>-0.07</td>
<td>0.13</td>
<td>0.28</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>Duration at StartSegment</td>
<td>-0.07</td>
<td>1.00</td>
<td>-0.06</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Historical Volume</td>
<td>0.13</td>
<td>-0.06</td>
<td>1.00</td>
<td>-0.23</td>
<td>-0.34</td>
<td>-0.09</td>
</tr>
<tr>
<td>Incidents per Volume</td>
<td>0.26</td>
<td>-0.01</td>
<td>-0.23</td>
<td>1.00</td>
<td>0.09</td>
<td>0.02</td>
</tr>
<tr>
<td>OnTime of StartSegment</td>
<td>0.05</td>
<td>0.02</td>
<td>-0.34</td>
<td>0.09</td>
<td>1.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Facility Type Appt of EndSegment</td>
<td>0.07</td>
<td>0.03</td>
<td>-0.09</td>
<td>0.02</td>
<td>0.03</td>
<td>1.00</td>
</tr>
</tbody>
</table>

PCA / MCA

Stepwise Regression Output

- Standard forward search
- Starts from an empty model
- At each step the model selects a variable that increases maximum likelihood fit.

\[
\text{LogWorth} = -\log_{10}(p - \text{value})
\]
Performance Evaluation

Build models using six explanatory variables with statistical significance

Confusion Matrix to assess the **predictive** power of the models

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
<th>err</th>
<th>missed delays</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C₀</td>
<td>C₁</td>
<td></td>
</tr>
<tr>
<td>C₀</td>
<td>(n_{0,0}) = number of C₀ cases classified correctly</td>
<td>(n_{0,1}) = number of C₀ cases classified incorrectly as C₁</td>
<td>(\frac{n_{0,1} + n_{1,0}}{n})</td>
</tr>
<tr>
<td>C₁</td>
<td>(n_{1,0}) = number of C₁ cases classified incorrectly as C₀</td>
<td>(n_{1,1}) = number of C₁ cases classified correctly</td>
<td></td>
</tr>
</tbody>
</table>
## Predictive Performance (Validation dataset)

### Main model: LR

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>0</th>
<th>1</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>219</td>
<td></td>
<td>209</td>
<td>429</td>
</tr>
<tr>
<td>1</td>
<td>1,805</td>
<td>6,337</td>
<td>8,142</td>
<td></td>
</tr>
<tr>
<td>Σ</td>
<td>2024</td>
<td>6546</td>
<td>8570</td>
<td></td>
</tr>
</tbody>
</table>

err = (n₀₁ + n₁₀)/n  
missed delays = n₀₁/n  

- Model interpretations vs “Black Box” approach
- High visibility of the predictors
- Robust results

### Assess Performance: NN and BF

#### NEURAL NETWORK

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>0</th>
<th>1</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>254</td>
<td></td>
<td>175</td>
<td>429</td>
</tr>
<tr>
<td>1</td>
<td>2,075</td>
<td>6,067</td>
<td>8,142</td>
<td></td>
</tr>
<tr>
<td>Σ</td>
<td>2329</td>
<td>6241</td>
<td>8570</td>
<td></td>
</tr>
</tbody>
</table>

err = (n₀₁ + n₁₀)/n  
missed delays = n₀₁/n  

- 26.25% error rate
- 2.04% missed delays

#### BOOTSTRAP FOREST

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>0</th>
<th>1</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>243</td>
<td></td>
<td>186</td>
<td>429</td>
</tr>
<tr>
<td>1</td>
<td>2,058</td>
<td>6,084</td>
<td>8,142</td>
<td></td>
</tr>
<tr>
<td>Σ</td>
<td>2301</td>
<td>6270</td>
<td>8570</td>
<td></td>
</tr>
</tbody>
</table>

err = (n₀₁ + n₁₀)/n  
missed delays = n₀₁/n  

- 26.18% error rate
- 2.17% missed delays
Predictive Performance (Testing dataset)

New dataset to gauge model’s accuracy and robustness

- **Validation**

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>(\Sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>219</td>
<td>209</td>
</tr>
<tr>
<td>1</td>
<td>1,805</td>
<td>6,337</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td>2024</td>
<td>6546</td>
</tr>
</tbody>
</table>

\[
err = \frac{(n_{0,1} + n_{1,0})}{n} = 23.50\%
\]

- **Test**

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>(\Sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>23</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>452</td>
<td>1,448</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td>475</td>
<td>1,498</td>
</tr>
</tbody>
</table>

\[
err = \frac{(n_{0,1} + n_{1,0})}{n} = 25.44\%
\]

- **Validation**

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>(\Sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2.6%</td>
<td>2.4%</td>
</tr>
<tr>
<td>1</td>
<td>21.1%</td>
<td>73.9%</td>
</tr>
</tbody>
</table>

- **Test**

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>(\Sigma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.2%</td>
<td>2.5%</td>
</tr>
<tr>
<td>1</td>
<td>22.9%</td>
<td>73.4%</td>
</tr>
</tbody>
</table>
Application - Results

Using model results to prioritize loads requiring attention

<table>
<thead>
<tr>
<th>LoadStopID of StartSegment</th>
<th>LoadStopID of EndSegment</th>
<th>Contract-Spot</th>
<th>Duration at StartSegment</th>
<th>Historical Volume</th>
<th>Incidents per Volume</th>
<th>OnTime of StartSegment</th>
<th>Facility Type Appt of EndSegment</th>
<th>Prob [On-time]</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXX1</td>
<td>YYY1</td>
<td>1</td>
<td>14:39</td>
<td>111</td>
<td>0.12</td>
<td>0</td>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>XXX2</td>
<td>YYY2</td>
<td>1</td>
<td>0:08</td>
<td>2011</td>
<td>0.07</td>
<td>0</td>
<td>1</td>
<td>20%</td>
</tr>
<tr>
<td>XXX3</td>
<td>YYY3</td>
<td>1</td>
<td>16:55</td>
<td>1010</td>
<td>0.08</td>
<td>1</td>
<td>1</td>
<td>42%</td>
</tr>
<tr>
<td>XXX4</td>
<td>YYY4</td>
<td>1</td>
<td>5:30</td>
<td>1349</td>
<td>0.07</td>
<td>1</td>
<td>1</td>
<td>57%</td>
</tr>
<tr>
<td>XXX5</td>
<td>YYY5</td>
<td>1</td>
<td>1:30</td>
<td>654</td>
<td>0.03</td>
<td>1</td>
<td>1</td>
<td>66%</td>
</tr>
<tr>
<td>XXX6</td>
<td>YYY6</td>
<td>1</td>
<td>2:30</td>
<td>1077</td>
<td>0.06</td>
<td>1</td>
<td>0</td>
<td>74%</td>
</tr>
<tr>
<td>XXX7</td>
<td>YYY7</td>
<td>1</td>
<td>1:40</td>
<td>6</td>
<td>0.00</td>
<td>1</td>
<td>0</td>
<td>80%</td>
</tr>
<tr>
<td>XXX8</td>
<td>YYY8</td>
<td>1</td>
<td>0:15</td>
<td>4</td>
<td>0.00</td>
<td>1</td>
<td>0</td>
<td>81%</td>
</tr>
<tr>
<td>XXX9</td>
<td>YYY9</td>
<td>1</td>
<td>0:01</td>
<td>85</td>
<td>0.00</td>
<td>1</td>
<td>0</td>
<td>82%</td>
</tr>
</tbody>
</table>
Application - Results
Using model results to drive actions

![Prediction Profiler diagram]

- **Inc. per Volume**: known when the load is tendered
- **Historical Volume**: known when the load is tendered
- **Cont. (0) vs Spot (1)**: known after pick-up
- **Facility Type Appt**: known after pick-up
- **Duration Start Seg.**: known after pick-up
- **On-time Start Seg.**: known after pick-up
Conclusion

1. Resources can be optimized using the Logistic Regression Model
2. On-time delivery can be predicted
3. Using a combination of six variables with high statistical significance can deliver predictive power

<table>
<thead>
<tr>
<th>Actual</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.1%</td>
<td>2.6%</td>
</tr>
<tr>
<td>73.9%</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

- Tracking 23.7% of the loads
- Missing only 2.4% of loads that will be late
Conclusion

Trade-off: Resource Reduction vs Missing Error

<table>
<thead>
<tr>
<th>Cut Off</th>
<th>Prediction</th>
<th>Actual</th>
<th></th>
<th>Tracking</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>2.6%</td>
<td>21.1%</td>
<td>2.4%</td>
<td>73.9%</td>
<td>2.4%</td>
</tr>
<tr>
<td>0.6</td>
<td>4.1%</td>
<td>53.8%</td>
<td>0.9%</td>
<td>41.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>0.7</td>
<td>4.9%</td>
<td>86.6%</td>
<td>0.1%</td>
<td>8.4%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>
Suggestion for Future Research

- Increased availability of online information through new technologies
- Readiness to store records on remote servers using (cloud servers)
- Predictive model able to capture information from online records could bring new insights and complement the analysis presented in this study
backup slides
## Variables

Build models using six explanatory variables with statistical significance

<table>
<thead>
<tr>
<th></th>
<th>Inc. per Volume</th>
<th>Historical Volume</th>
<th>Cont. (0) vs Spot (1)</th>
<th>Facility Type Appt</th>
<th>Duration Start Seg.</th>
<th>On-time Start Seg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob. [0]</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>1</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Prob. [1]</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>0</td>
<td>-</td>
<td>1</td>
</tr>
</tbody>
</table>
Reweighted Confusion Matrix

**Original Confusion Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>Σ</td>
</tr>
<tr>
<td>Actual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>2,192</td>
<td>2,093</td>
<td>4,285</td>
</tr>
<tr>
<td>1</td>
<td>950</td>
<td>3,335</td>
<td>4,285</td>
</tr>
<tr>
<td>Σ</td>
<td>3,142</td>
<td>5,428</td>
<td>8,570</td>
</tr>
</tbody>
</table>

\[
\text{err} = \frac{n_{0,1} + n_{1,0}}{n}
\]

\[
\text{Err for predicting 1 and actual } = 0
\]

35.51%

24.42%

**Reweighted Confusion Matrix**

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>Σ</td>
</tr>
<tr>
<td>Actual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>219</td>
<td>209</td>
<td>429</td>
</tr>
<tr>
<td>1</td>
<td>1,805</td>
<td>6,337</td>
<td>8,142</td>
</tr>
<tr>
<td>Σ</td>
<td>2,024</td>
<td>6,546</td>
<td>8,570</td>
</tr>
</tbody>
</table>

\[
\text{err} = \frac{n_{0,1} + n_{1,0}}{n}
\]

\[
\text{Err for predicting 1 and actual } = 0
\]

23.50%

2.44%
# of observations

Total Sample: 522,920
Excl. Outliers or missing values: 342,800
Undersampling: 34,280
Validation: 8570

## Logistic Regression

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
<th>0</th>
<th>1</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>219</td>
<td>209</td>
<td>429</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>1,805</td>
<td>6,337</td>
<td>8,142</td>
</tr>
<tr>
<td>Σ</td>
<td></td>
<td>2024</td>
<td>6546</td>
<td>8570</td>
</tr>
</tbody>
</table>

$$\text{err} = \frac{n_{0,1} + n_{1,0}}{n}$$

Missed delays = 2.44%
Predictive Models

- **Logistic Regression**
  - Simply saying, it works with the same ideas as linear regression, but for a categorical output.
  - Relies on mathematical equation relating predictors with the outcome.

- **Neural Network**
  - Machine Learning technique. It mimics the activity in the brain, where neurons are interconnected and learn from experience.

- **Bootstrap Forest**
  - Variation of Random Forests. It combines results from multiple trees to improve predictive power.