Hunter Douglas Quality Project

Freddie | Griffin | Hathaway | Meg | Yuru

Company Overview

- Employs over 17,000 people worldwide
- Operates in more than 100 countries
- Composed of 168 companies with 68 manufacturing and 100 assembly plants and marketing organizations
- Over $2.5 billion in sales
- Worldwide manufacturer of architectural products (acoustic ceilings, rain screens, building facades) and window coverings
- Headquarters in Rotterdam, the Netherlands
- North American Headquarters –Pearl River, NY, USA

Goal!

Decrease Customer Rejection Rate (CRR)

The company will cover all the cost from shipping products back to remake and all other costs which will decrease the company's profit.

Our goal is to decrease the CRR so that good products rate will be increased. We can increase sales to improve the profit.

Another goal is to improve our customer loyalty and retention to increase revenue.

Time is money, but so is reputation

Sending out flawed blinds harms both

- How can we accurately and effectively double check blinds that are likely to be flawed to avoid extra costs, both immediate and from a damage to HDs reputation?
- **We aim to create a predictive model to help HD identify the probability of failure before sending out.**

Problem Description

Ho

Hypothesis

Our model is able to predict failure based on some factors.

By accurately predicting which blinds are likely to be flawed BEFORE sending them out, our model can save Hunter Douglas time, money, and reputation damages that impact the bottom line.

Architecture & Approach

Target

Warehouse A

Roller Shades

Preprocessing

Data Cleansing

Remove duplicates and incomplete records

Only keep factory-fault records within 90 days.

Predictors

Continuous: NET_SALES, REMAKE_UNITS, WIDTH, HEIGHT

Categorical: SALES_ORDER_LINE, ALLIANCE_LEVEL, REGIONAL, 'REGION_STATE, FABRIC, 'COLOR, MATERIAL, SYSTEM, day_of_week, OPERATING_SYS_OPT_ID,

Target Variable

Remade (binary, where '1' indicates remade original order)

Architecture & Approach

The goal of a good machine learning model is to get the right balance of Precision and Recall, by trying to maximize the number of True Positives while minimizing the number of False Negatives and False Positives.

Random Forest Regressor

Analytic Approach 1

Random Forest Regressor

- New features: a measure of the relative uniqueness of the feature
  - `vc_color, vc_fabric, vc_materials`
  - **calculated as** `prepare_file['vc_color'] = prepare_file.groupby('COLOR_ID_x')['COLOR_ID_x'].transform('count')/len(prepare_file)`
- Features
  - `FLAWED`(target), `WIDTH_x, HEIGHT_x, NET_SALES_UNITS_x, COLOR_ID_x, ORIGINAL_MATERIAL_ID_x, FABRIC_ID_x, OPERATING_SYSTEM_ID_x, OPERATING_SYS_OPT_ID_x, vc_color,vc_fabric, vc_materials`

Random Forest Regressor performs better when data is categorical and unbalanced. As Decision Tree is prone to overfitting, Random Forests are used in practice to better generalize the fitment. RF provide a good balance between precision and overfitting. *However, logistic regression is easier to explain and much faster to train and execute.*

Random Forest Regressor Result

threshold = 0.15 10 trees

Precision	0.31807
Accuracy	0.9459
F1	0.5219
Recall	0.5537

Predicted Failure Rate	6.023%
Actual Failure Rate	5.746%
Mean Absolute Error	6.895%

Feature		Predicted	
		0	1
Observed	0	49336	1534
	1	1384	1717

- Set our threshold at 0.37 to achieve an equal amount of predicted failures to actual failures. Then set the threshold to 0.15 according to the standard of Hunter Douglas.
- Chose to use 0.37 because 0.15 gave us a prediction rate of 11 %, which is too high to check all those.

Feature Importance Top 10

Feature	Importance	Rank
Width	**0.302243**	**1**
Height	**0.272038**	**2**
vc_color	0.026975	3
vc_materials	0.023783	4
vc_fabrics	0.021360	5
Net_Sales_Unites	0.020071	6
Color_ID_801	0.009274	7
Color_ID_201	0.009228	8
Color_ID_101	0.006873	9
Color_ID_301	0.006791	10

Neural Network

Analytic Approach 2

Neural Network

- New feature
  - `volume = width x height`
  - `vc_color/fabric/materials`
- One-hot encoding / dummies
- Validation set (0.2)
- Improve Performance
  - Normalization
  - Adding noise
  - Resampling

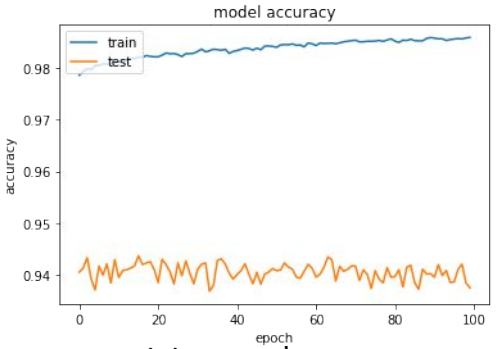
Neural Network can learn and discover inherent/generic features in objects that will be helpful in classification. Usually worry less about the feature engineering part. *However, it requires "relatively" large datasets to work well and needs the infrastructure to train them in a reasonable time. Not a easy algorithm to start with and hard to interpret predictors because of "black box."*

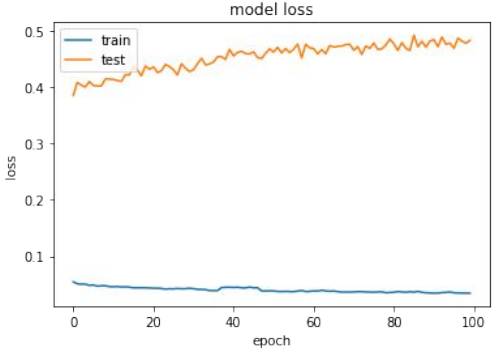
Neural Network Result

threshold = 0.15

Precision	0.4034 / 0.4864	Predicted Failure Rate	7.059%
Accuracy	0.9435 / 0.9394	Actual Failure Rate	5.778%
F1	0.4010 / 0.4034	Mean Absolute Error	7.66%
Recall	0.4481 / 0.4481		

Feature		Predicted	
		0	1
Observed	0	56677	2823
	1	2014	1635





- Have to make decision between precision and accuracy.
- Set threshold at 0.15 according to the standard of Hunter Douglas.
- Can't provide feature importance because of mechanism.

Economic Impact Model

Software | Hardware | Personnel | IT | Consulting

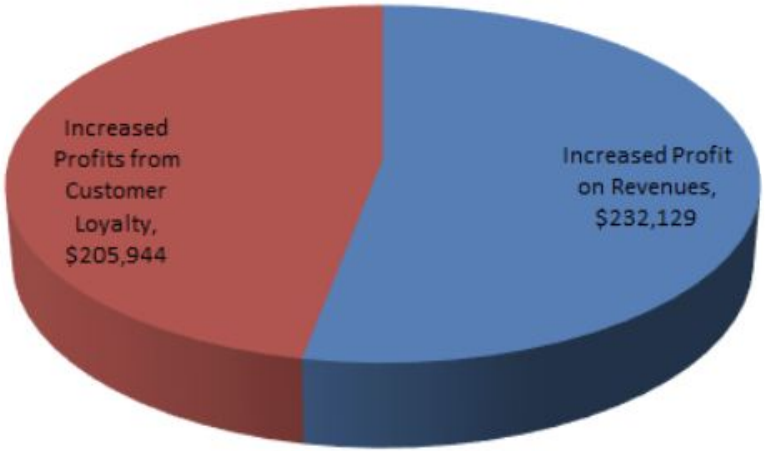
- Increased predictive power from our models leads to more than $45,000 in saving for Hunter Douglas at this factory alone
- Limited expected costs make implementation of our recommendations cost-effective
- Continuous improvement of model going forward would continue to create tangible value for Hunter Douglas

Table 1. Summary Results

Return on investment (ROI)		248%
Payback period (years)		0.41
Net present value	$	105,423
Annual total cost of ownership	$	26,907
Total 5-year benefits	$	438,073

Economic Impact Model

Benefits



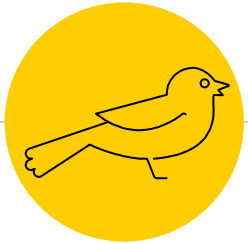
- Direct
  - Reduction in CRR rate directly leads to increased profits due to cost savings
- Indirect
  - 0.05% increase per year in profits stemming from reduced CRR rate which can increase customer loyalty

Lessons Learned

- Data cleansing and model training are time-intensive.
- We could pick the optimal algorithm based on the key metrics (time efficiency, precision) we chose.
- While feature engineering is cool, and could potentially lead to better results, not to spend too much time on creating new features.
- Neural Network doesn't have too much limitation on feature engineering because large data helps the performance of prediction; however, the model will be overfitted if the sample is too small.
- Modeling economic impact for a business is inherently uncertain.
- Savings of a few percentage points can have a dramatic impact on profitability.
- The more we learn, the less we know.

Conclusion and next steps

- The Random Forest model performed better than the Logistic Regression models (LassoLarsCV & GBRT) and Neural Network in our occasion.
- Our models only forecast the probability of returning for roller shades in warehouse A. In order to obtain a more comprehensive model, HD should expand the choice of products and adjust model parameters accordingly.
- Our models are able to calculate the returning probability if the user inputs proper 2019 data into our models.
- Width, height, and color are the most influential predictors. Hunter Douglas should pay extra attention to these factors before sending out their products.



Thanks!

Any questions ?