Using Python Libraries to Predict, Optimize, and Provide End Users Decision Confidence

Paul Chen, Soyeon Baik, Matthew A. Lanham
Purdue University, Krannert School of Management
chen3876@purdue.edu; baik6@purdue.edu; lanhamm@purdue.edu

Krannert School of Management

ABSTRACT

This study is designed to predict and minimize the window breakage rate by controlling the parameters in the manufacturing process. This will help window manufacturers to understand the important factors that affect the window breakage, control them, and minimize the risk and cost for breakage. Pycaret, the opensource and low-code library that automates the machine learning workflow is used for prediction. Pyomo, another open-source python library is used for optimization. This study suggests an adaptable business solution by using simple and intuitive python packages.

INTRODUCTION

Current average window breakage rate is 6% with an opportunity cost of $100 per 1$m^2$ size. Reducing the breakage rate will lead to a potential increase in profit for window manufacturers.

Various factors are involved in the window manufacturing process, which makes it challenging for manufacturers to understand and recognize what are the important factors and how they can adjust them to reduce window breakage rate in the real manufacturing process.

Table 1 contains the parameters in the window manufacturing process. We aim to find the coefficient and optimized options for each parameter that can be adapted right into the manufacturing process.

Type	Factors	Type	Factors
Customer spec	Window Size	Process	Edge Delegation Rate
	Window Type		Cut speed
Supplier	Window Color		Spacer Distance
	Glass Thickness		Silicon Viscosity
	Location		

Table 1. Parameters for Window Manufacturing

RESEARCH OBJECTIVES

- Which factors impact the window breakage rate?
- What is the feasible mix of parameters that can be applied to real window manufacturing process settings to minimize the breakage rate?

LITERATURE REVIEW

Following academic pieces of literature are referred to conduct this research. The main difference from the former research is that we mainly focus on interpretability and adaptability for decision-makers in the real manufacturing operation processes.

Study	Prediction	Optimization
Building A Classfication Model With Pycaret	Classification	
Pyomo — Optimization Modeling in Python		Optimization
Smart “predict, then optimize”	Prediction	Optimization
Our Study, 2022	**Prediction**	**Optimization**

Table 2. Literature Review Summary and Comparison

METHODOLOGY

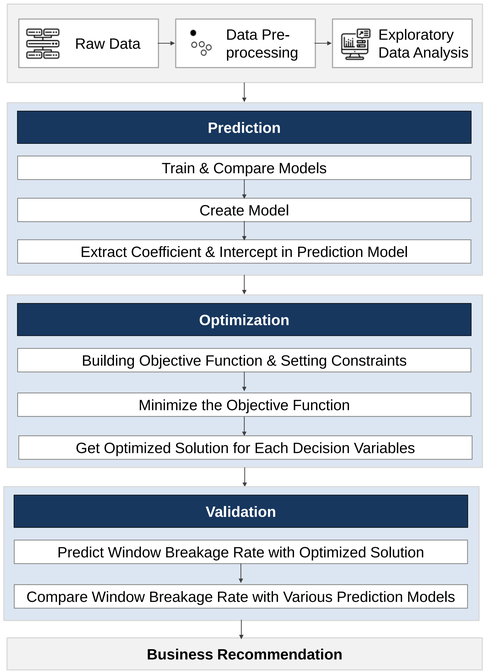

Fig 1. Detailed Methodology

STATISTICAL RESULTS

Prediction Model

Pycaret library automatically generates various machine learning models to predict window breakage rates.

Model	MSE	RMSE	R2
Extra Tree Regressor	7.12	2.65	78.17%
Light GBM	7.86	2.79	76.0%
Huber Regressor	12.90	3.58	60.8%
Linear Regression	14.22	3.75	56.9%

Fig 2 . Prediction Model Performance Comparison (MSE, RMSE, R2)

Optimization Modeling

Optimal values for decision variables that manufacturers can actually control to reduce the window breakage rate.

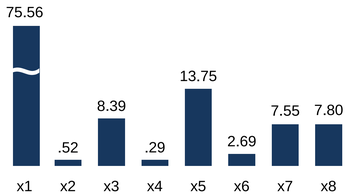

Fig 3. Optimized Settings for Parameters

Variables	Description
X1	Window size
X2	Glass thickness
X3	Ambient temp
X4	Cut speed
X5	Edge deletion rate
X6	Spacer distance
X7	Window color
X8	Silicon viscosity

Table 3. Parameters

Prediction Based On Optimization

Predicted breakage rate based on the optimized settings is reduced by **99.85%**

Model	Predicted Breakage Rate
Extra Tree Regressor	1.82%
Light GBM	2.41%
Huber Regressor	0.010%
Linear Regression	0.009%
Average (Historical Data)	6.27%

Fig 4 . Window Breakage Rate Prediction with Optimized Parameter Options

EXPECTED IMPACT

With the optimal settings for each parameter, the window breakage rate can be reduced to almost zero. The simple linear regression makes it easy for window manufacturers to understand what factors affect the window breakage rate and adjust settings to minimize it.

Based on the historical data, 6% is the average window breakage rate in the industry. The potential savings from this decision-making in this industry are **$6,300** per 1,000 batches.

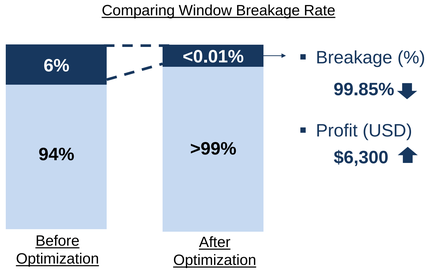

Fig 5. Total Savings ($) based on predicted breakage rate

CONCLUSIONS

Developing and choosing models with the highest accuracy is important. However, we should always keep in mind what is the purpose of the analysis. We tried to apply data analytics to help decision-makers in the real business problem. In this case, models that can suggest actionable solutions are more crucial than the accuracy itself. Also, lowering the barrier for the non-analytical decision-makers to make an informed decision is important. With free, simple, and user-friendly python packages, this research can be stretched to many business operational fields for better decision-making.

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham and our industry partner for this opportunity, their guidance, and support on this project.