Predicting Appointment Conversion with Probability Scoring ML Models

Juily Vasandani, Saumya Bharti, Deepankar Singh, Shreeansh Priyadarshi, Matthew A. Lanham

Purdue University Krannert School of Management

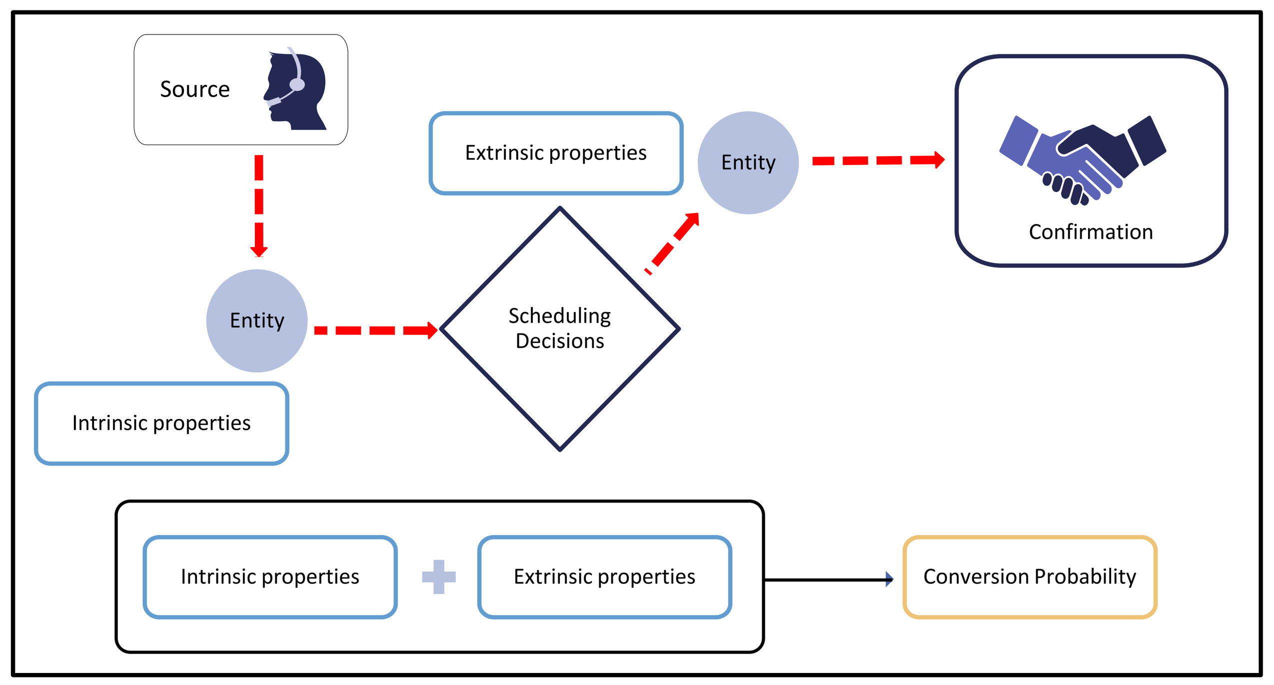
jvasanda@purdue.edu; bhartis@purdue.edu; singh681@purdue.edu; spriyad@purdue.edu; lanhamm@purdue.edu

ABSTRACT

This study examines and tests several supervised learning methodologies to predict *and* maximize appointment conversions. Our research is motivated by costs incurred by the industry-agnostic problem of employee idle time, one of most significant consequences of canceled appointments. Using proprietary data from a national partner, we develop and test machine learning algorithms to build an ensemble model that predicts conversion probability. Our work is novel because we implement probability scoring to evaluate our model and improve its overall predictability. In production, our algorithm can lead to a **boost of nearly 6% in project revenue**.

INTRODUCTION

When a customer cancels an appointment, firms must deal with the impact of employee idle time, costing companies **over $100 billion dollars** in annual wages.



Introducing predictability into the sales process allows resources to be directed towards clients with the highest probability of conversion, minimizing idle time.

This study aims to answer the following **research questions**:

- What are the drivers that impact the likelihood of an appointment conversion?
- Can predictability in the conversion process reduce employee idle time costs?

LITERATURE REVIEW

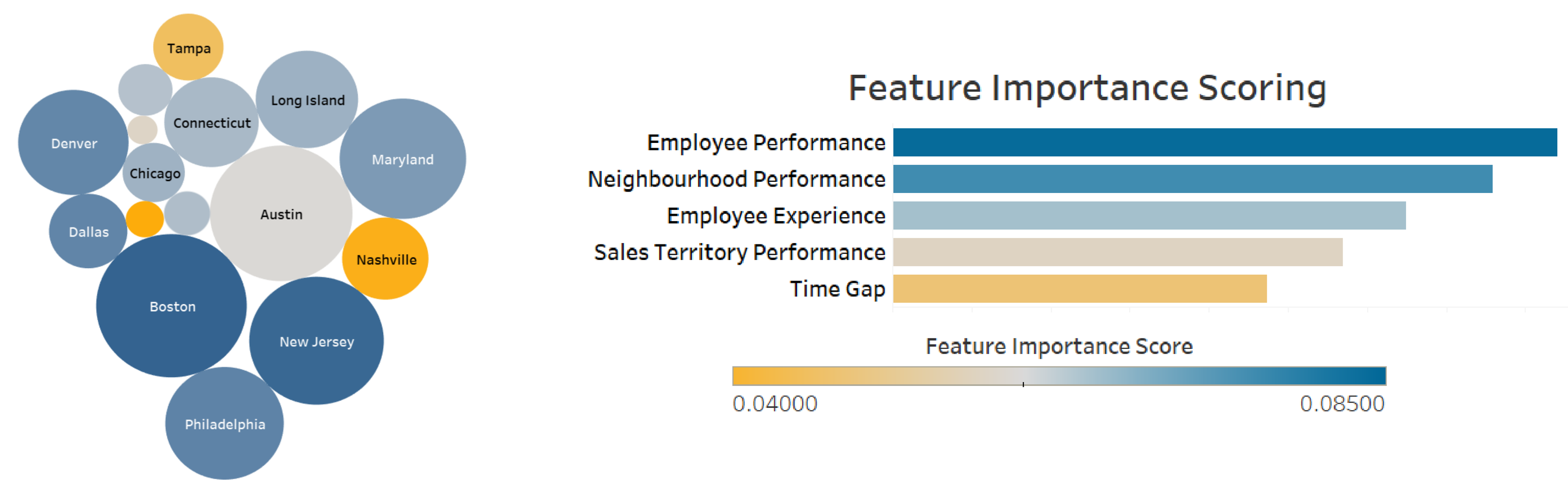
Although not much data exists in the appointment management space, similar binary classification problems like attrition are more common. However, no research exists that utilizes probability scoring to evaluate and improve model predictability.

Study	Algorithm(s) Used	Probability Scoring
Mohammadi et al. (2019)	Logistic regression; Multilayer perceptron; Naïve Bayes	✗
Ahman et al. (2019)	Tree-based models (DT & RF); Boosting models (Gradient Boost & XGBoost)	✗
Pamina et al. (2019)	K-nearest Neighbors (KNN); Tree-based models (DT & RF); Neural network	✗
Vafeiadis et al. (2015)	Artificial neural network; Support Vector Machine; Logistic regression	✗
Vasandani et al. (2020)	***Ensemble model:* ANN and Boosting algorithms**	✓

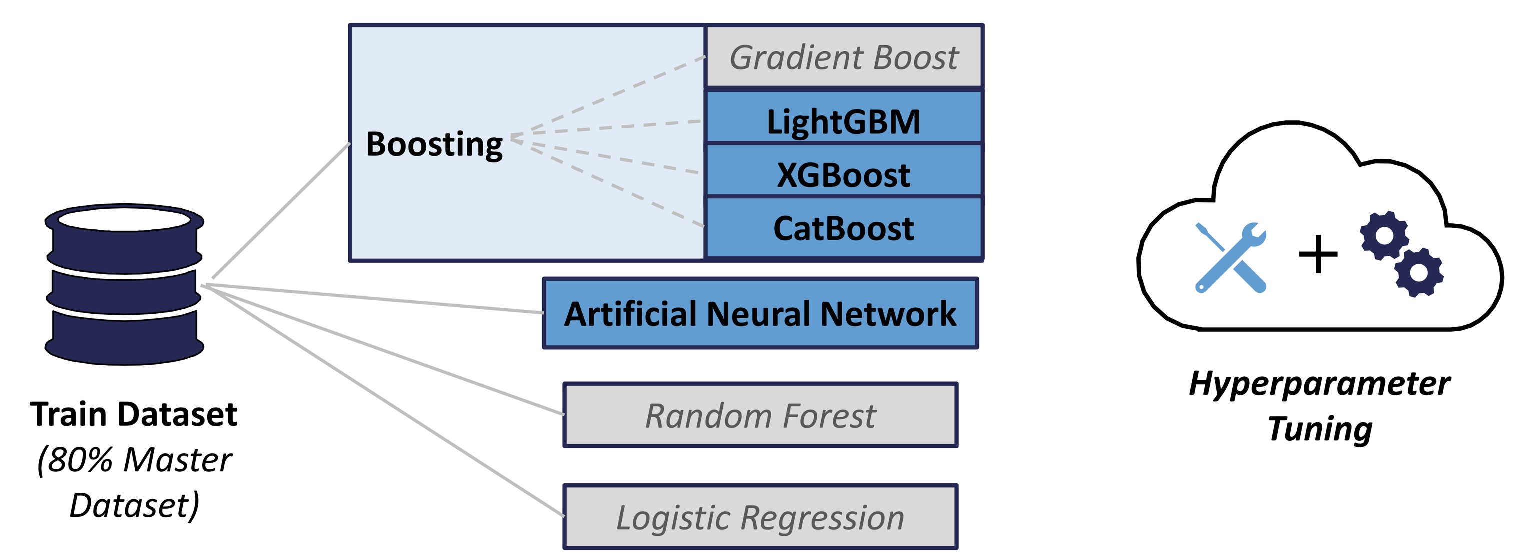
METHODOLOGY

1 *Data Understanding, Cleaning, & Feature Engineering*

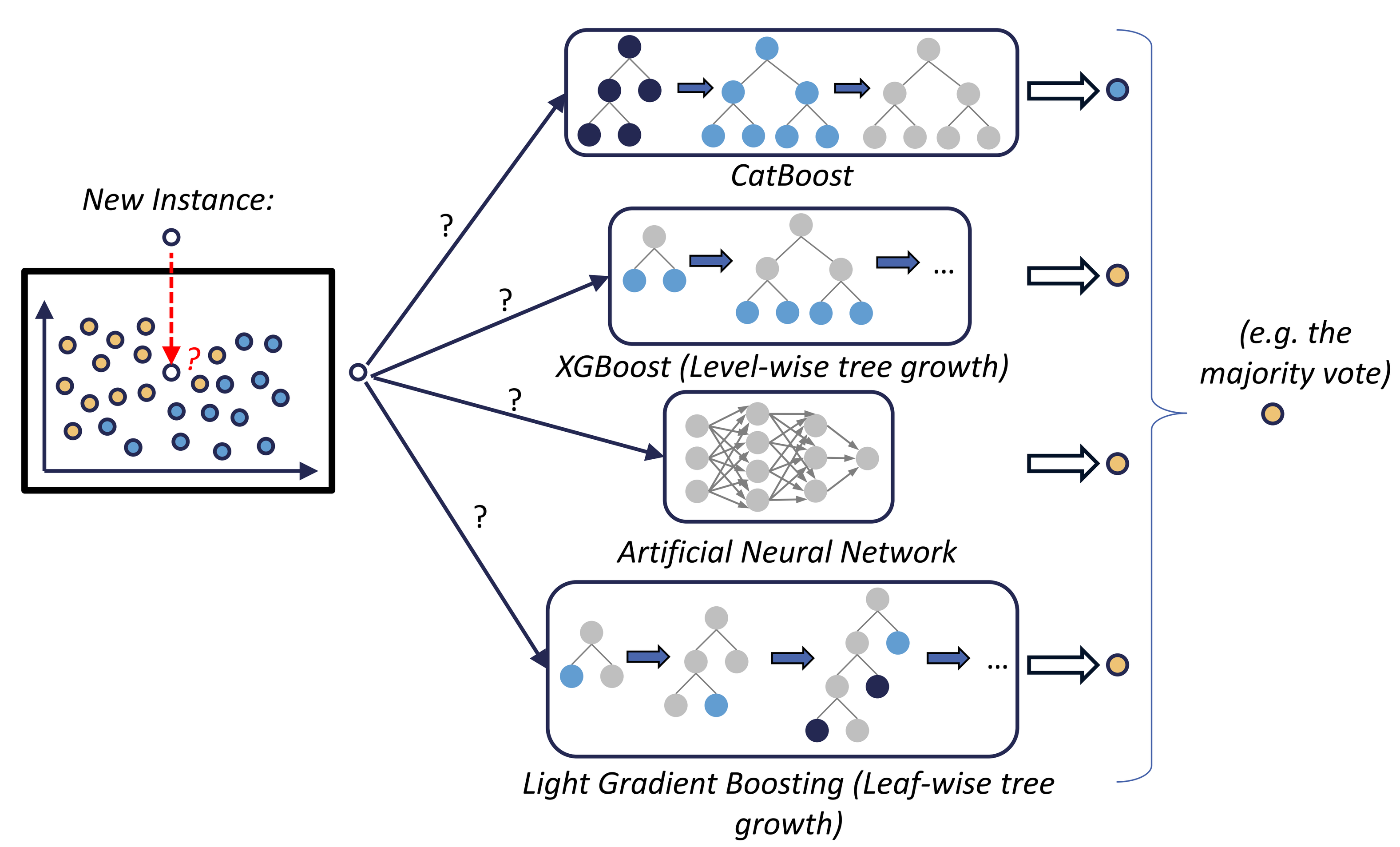
2 *Feature Testing & Importance*



3 *Model Selection & Algorithm Tuning*



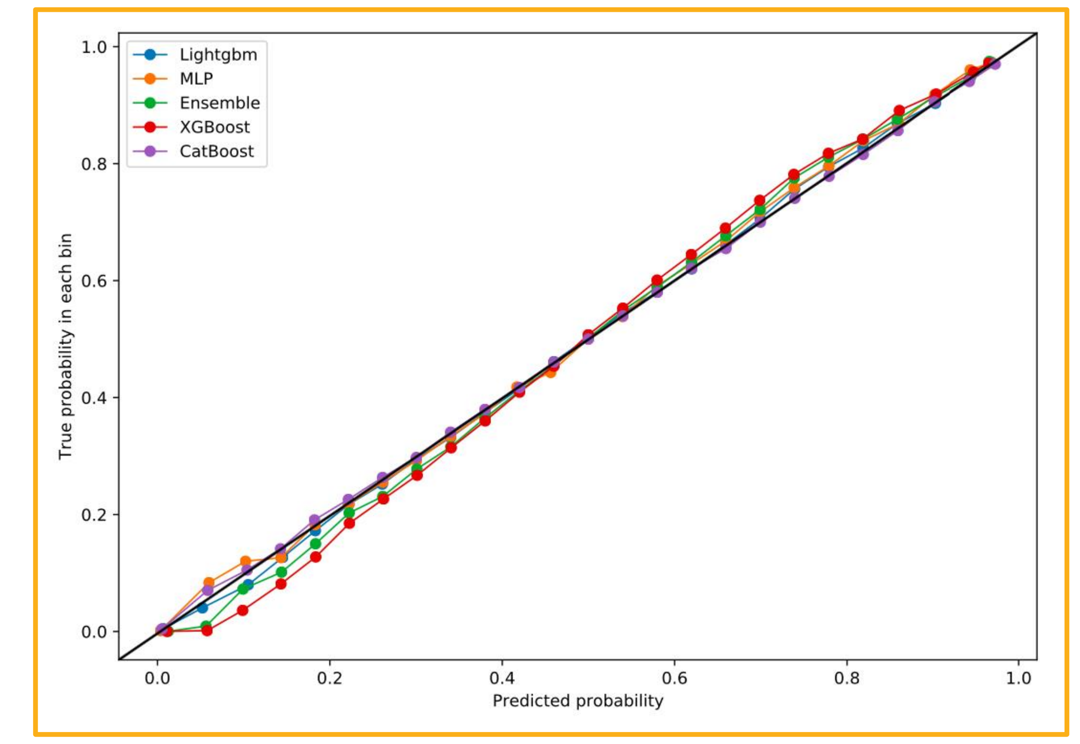
4 *Ensemble Creation*



RESULTS

1 *Model Evaluation*

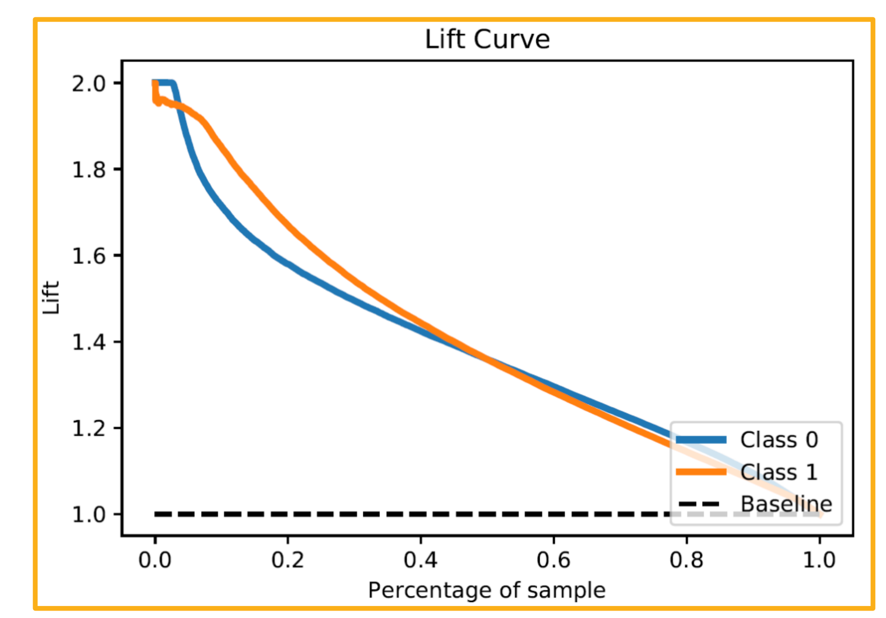
A calibration plot provides a diagnostic to determine whether the predicted values are reliable, providing a visual representation of the final performances.

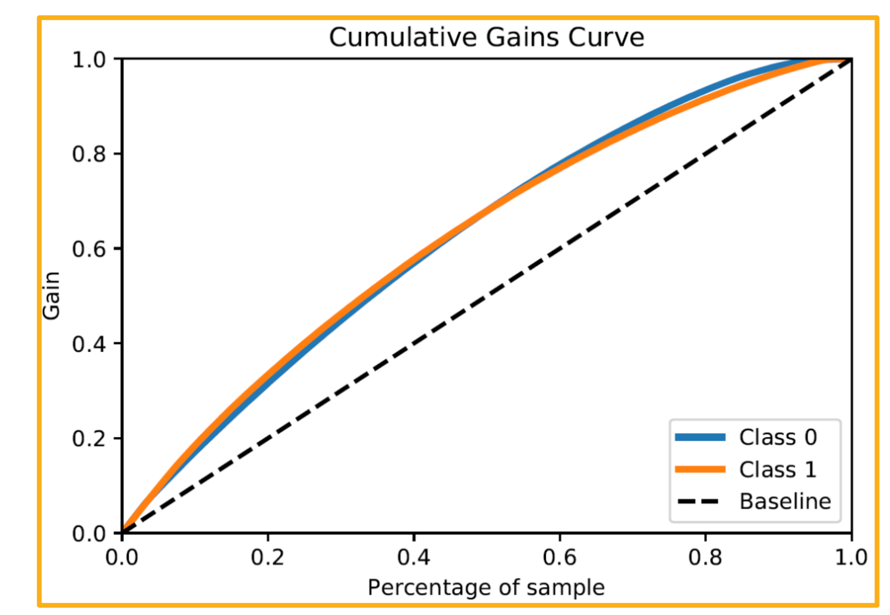


Models	Log Loss
XGBoost	0.587610044
ANN	0.579266692
CatBoost	0.579286963
LightGBM	0.579380672
Ensemble	***0.578761449***

Log loss measures classifier performance, increasing as the predicted probability diverges from the actual label.

2 *Lift/Gain:*





Lift and gain charts allow us to gauge the relative improvement in the conversion rate when targeting high-quality leads compared to selecting a lead at random.

CONCLUSIONS

On average, our model is estimated to produce a **1.4x lift** in the conversion rate, boosting it to approximately 37%. Using conservative project conversion estimates, this **increases the number of confirmed projects by 30%**.

Assuming average project value of $13,000, implementation of our ensemble model in production leads to a **5.8% boost in revenue** – nearing approximately **$18,000,000** in value for our partner.

ACKNOWLEDGEMENTS

We thank Professor Matthew Lanham and our company sponsor for their constant guidance and support on this project.