Population Size and Cigarette Consumption: A Machine Learning Perspective

O.P. Adebayo[1*]; I. Ahmed[2]; O.A. Ogunjim[3]; and K.T. Oyeleke[4]

[1] Department of Statistics, Phoenix University Agwada, Nasarawa State, Nigeria.
[2] Department of Computer Science, Phoenix University Agwada, Nasarawa State, Nigeria.
[3] Department of Statistics, Nasarawa State University Keffi, Nasarawa State, Nigeria.
[4] Department of Statistics, Olabisi Onabanjo University Ago Iwoye, Ogun State, Nigeria.

E-mail: adebayo.patrick@phoenixuniversity.edu.ng*
ogunjimi.olalekan@phoenixuniversity.edu.ng
ibrahimloko@nsuk.edu.ng
kamoruoyeleke@gmail.com

Telephone: +2349019964334*

ABSTRACT

This study develops a predictive machine learning model to analyze cigarette pack sales using key economic and demographic variables. The gradient boosting model achieved strong predictive performance with a test RMSE of 0.530 and explained 70.6% of sales variance (R^2 = 0.706), demonstrating reliable forecasting capability. Feature importance analysis revealed price as the dominant predictor, contributing 55.7% of the model's explanatory power, followed by tax (12.7%) and population (11.5%). The implementation of early stopping at the 10th iteration prevented overfitting while maintaining model generalizability. These findings provide actionable insights for public health policy and retail strategy, quantifying how pricing and taxation influence consumption patterns. The analysis also identified negligible contributions from certain variables (e.g., CPI), suggesting opportunities for model simplification. By combining robust predictive accuracy with interpretable feature importance metrics, this research offers a data-driven framework for understanding cigarette market dynamics and supporting evidence-based decision-making. The results highlight the critical role of price sensitivity in tobacco consumption while establishing methodological foundations for future sales forecasting models.

(Keywords: machine learning model, XGBoost regression, LightGBM, gain, cover, and frequency metrics)

INTRODUCTION

The relationship between population size and cigarette consumption represents a critical area of study in public health research, with significant implications for tobacco control policies and population health outcomes. As global populations continue to grow and urbanize, understanding the dynamics of cigarette consumption patterns becomes increasingly important for effective public health interventions (World Health Organization [WHO], 2021).

Research has consistently demonstrated that population characteristics significantly influence tobacco use patterns. Larger populations tend to exhibit higher absolute cigarette consumption due to simple scaling effects, but more importantly, population density and urbanization have been shown to correlate with distinct smoking behaviors (Ng, *et al*., 2014). The concentration of tobacco retailers in urban areas, combined with targeted marketing strategies, often leads to increased accessibility and normalization of smoking in densely populated regions (Pearce, *et al*., 2016).

Demographic factors interact with population size to create complex consumption patterns. Studies have found that age structure within populations significantly affects smoking prevalence, with younger populations in developing nations showing different consumption patterns compared to aging populations in developed countries (Bilano, *et al.*, 2015). Furthermore, the phenomenon of population aging in many nations has led to shifting patterns of tobacco-related

diseases, with longer exposure periods contributing to greater health burdens (Thun, *et al.*, 2013).

Economic theories of consumption suggest that population size affects cigarette demand through multiple mechanisms. The theory of rational addiction (Becker and Murphy, 1988) posits that social interactions within populations influence smoking behaviors, with larger populations potentially enabling stronger social multiplier effects. Additionally, the distributional economics of tobacco products means that per-unit costs typically decrease in larger markets, potentially increasing consumption (Chaloupka, *et al.*, 2012).

Recent empirical evidence highlights the importance of considering geographic and temporal variations in the population-consumption relationship. A longitudinal study across 100 countries found that while population growth initially predicts increased cigarette consumption, this relationship reverses as nations implement comprehensive tobacco control measures (Graen and Carpenter, 2018). This suggests that population effects are mediated by policy environments, with Framework Convention on Tobacco Control (FCTC) implementation playing a crucial moderating role (WHO, 2013).

The public health implications of these findings are substantial. As the global population is projected to reach 9.7 billion by 2050 (United Nations, 2019), with most growth occurring in low- and middle-income countries where tobacco control may be weaker, understanding these dynamics becomes increasingly urgent. Effective tobacco control strategies must account for population size and distribution factors when designing interventions, particularly in rapidly urbanizing regions where new tobacco markets are emerging (Satterthwaite, *et al.*, 2020).

This study developed a machine learning model to predict cigarette sales and identify key influencing factors. By analyzing economic and demographic variables, the research optimized prediction accuracy (using RMSE and R^2 metrics) while preventing overfitting through early stopping techniques. The findings highlight price as the most significant driver of sales, with tax and demographic factors playing secondary roles.

These insights offer practical value for both policymakers assessing taxation impacts and businesses optimizing pricing strategies. The study also evaluated model limitations, suggesting opportunities to refine future analyses by excluding less influential variables. Overall, the results provide data-driven insights into cigarette market dynamics, serving academic research and real-world decision-making while establishing a foundation for improved predictive modeling approaches.

Theoretical Foundations of Gradient Boosting

At its core, Gradient Boosting is an iterative algorithm that minimizes a loss function by adding weak learners to the model in a sequential fashion. The process begins with an initial model, such as the mean of the target variable for regression tasks. In each iteration, the algorithm computes the residuals (the differences between the actual and predicted values) and trains a new weak learner to predict these residuals. This new learner is then added to the model with a weight that minimizes the loss function. The final prediction is the sum of the predictions from all weak learners, weighted by their respective contributions.

Mathematically, the final model can be expressed as the sum of the predictions from all weak learners, each scaled by a weight. The algorithm uses gradient descent to minimize the loss function, which is why it is called Gradient Boosting. This approach allows the model to capture complex, non-linear relationships in the data, making it highly effective for a wide range of predictive tasks.

Mathematically, the final model can be expressed as:

$$\hat{y} = \sum_{t=1}^{T} \alpha_i f_i(x)$$

Where:

$\hat{y}$: Final prediction.
T : Number of weak learners (trees).
α_i : Weight of the $t-th$ weak learner.
$f_i(x)$: Prediction of the $t-th$ weak learner.

The algorithm minimizes a loss function (e.g., mean squared error for regression) using gradient descent, hence the name "Gradient Boosting."

XGBoost: eXtreme Gradient Boosting

XGBoost, introduced by Chen and Guestrin in 2016, is an optimized implementation of Gradient Boosting that incorporates several innovations to enhance performance and scalability. One of its key features is the inclusion of regularization terms in the loss function, which help prevent overfitting. Specifically, XGBoost adds L1 (Lasso) and L2 (Ridge) regularization to the model, ensuring that the learned trees are not overly complex. This regularization mechanism significantly improves the generalization ability of the model.

$$Loss = \sum_{i=1}^{N} L(y_i, \hat{y}_i) + \sum_{i=1}^{T} \Omega(f_t)$$

Where $\Omega(f_t)$ is the regularization term for the $t-th$ tree.

Another notable feature of XGBoost is its use of parallel processing to speed up training. By leveraging multi-core processors, XGBoost can handle large datasets efficiently, making it a popular choice for big data applications. Additionally, XGBoost can automatically handle missing values by learning the best imputation strategy during training, eliminating the need for manual data preprocessing.

XGBoost also employs a depth-first approach to grow trees, which reduces overfitting and improves model performance. This approach, combined with its ability to handle a wide range of data types and tasks, has made XGBoost a go-to algorithm for machine learning competitions and real-world applications. For example, XGBoost has been widely used in healthcare for predicting disease outcomes, in finance for fraud detection, and in marketing for customer churn prediction.

LightGBM: Light Gradient Boosting Machine

LightGBM, developed by Ke, *et al.* in 2017, is another highly efficient implementation of Gradient Boosting, designed to address some of the limitations of XGBoost, particularly in terms of speed and memory usage. LightGBM introduces two key techniques to improve performance: Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB).

Gradient-based One-Side Sampling focuses on data points with larger gradients, which are more informative for training, while randomly dropping points with smaller gradients. This approach reduces computational costs without significantly impacting model accuracy. Exclusive Feature Bundling, on the other hand, bundles mutually exclusive features (features that rarely take non-zero values simultaneously) to reduce the number of features, further improving efficiency.

One of the standout advantages of LightGBM is its ability to handle categorical features natively, without requiring one-hot encoding. This feature simplifies the preprocessing pipeline and reduces memory usage. Additionally, LightGBM uses a leaf-wise tree growth strategy, which grows trees depth-first rather than level-wise. This strategy allows the model to achieve higher accuracy with fewer trees, making it faster and more memory-efficient than XGBoost.

LightGBM has been widely adopted in industry applications, such as click-through rate prediction in online advertising, fraud detection in financial transactions, and text classification in natural language processing. Its speed and scalability make it particularly well suited for large-scale data and real-time applications.

Comparison of XGBoost and LightGBM

While both XGBoost and LightGBM are highly effective Gradient Boosting frameworks, they differ in several key aspects. XGBoost is known for its robustness and flexibility, making it a general-purpose tool for a wide range of tasks. It performs exceptionally well on structured data and is widely used in machine learning competitions. However, it can be slower and more memory-intensive than LightGBM, particularly for large datasets.

LightGBM, on the other hand, is designed for speed and efficiency. It is faster than XGBoost and uses less memory, making it ideal for large-scale data and real-time applications. LightGBM's ability to handle categorical features natively and its leaf-wise tree growth strategy further enhance its performance. However, it may require more careful tuning of hyperparameters compared to XGBoost.

Applications in Population Size and Cigarette Consumption

Gradient Boosting algorithms like XGBoost and LightGBM are particularly well-suited for analyzing the relationship between population size and cigarette consumption. These algorithms can capture complex, non-linear relationships between variables, making them ideal for exploring how factors such as population size, GDP, and age distribution influence cigarette consumption. Additionally, their ability to provide feature importance scores allows researchers to identify the most significant predictors of cigarette consumption, enabling targeted interventions and policy recommendations.

The dataset used in this work is CigarettesSW (Cigarette Consumption Panel Data) is a well-known dataset often used in econometrics and statistics. It is not specific to a single country but rather contains panel data (cross-sectional and time-series data) from multiple U.S. states over time.

Mathematical Representation.

1. Data Preprocessing
Scaling (Standardization):

$$z = \frac{x - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation of the feature x.

One-Hot Encoding (for categorical variables):

$$X_{encoded} = model.matrix(.-1, X)$$

Converts categorical variables (e.g., state) into binary dummy variables.

2. Correlation Matrix

$$Correlation\ Matrix = cor(X), where\ X = [cpi, population, income, tax, price, taxs]$$

Computes pairwise Pearson correlations between numerical features.

3. XGBoost Model

Objective Function (Regression with Squared Error):

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + Regularization\ Terms$$

where:
y_i = true value,
$\hat{y}_i$ = predicted value,
Regularization terms control overfitting (e.g., L2 regularization on leaf weights).

Gradient Boosting Algorithm:

1. Initialize model with a constant prediction:

$$\hat{y}_i^{(0)} = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$$

2. For each boosting round $t = 1$ to T:

Compute pseudo-residuals:

$$r_i^{(t)} = -\frac{\partial L\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}}$$

Fit a weak learner (decision tree) to predict residuals.

Update predictions:
$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta . f_t(x_i)$

3. where η = learning rate, f_t = tree model.

Parameters Used:
Learning rate ($\eta = 0.1$),
Max tree depth ($d = 6$),
Subsample ratio ($subsample = 0$),
Column subsampling
($colsample\_bytree = 0.8$).

4. Model Evaluation

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

R-squared (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

where $\bar{y}$ is the mean of the true values.

5. Feature Importance

Gain-based importance: Measures improvement in accuracy (or reduction in loss) when a feature is used in splits.

$$Importance_j = \sum_{t=1}^{T} Gain_j^{(t)}$$

where $Gain_j^{(t)}$ is the contribution of feature j in tree t.

6. Visualization

Actual vs. Predicted Plot:
Points $(\hat{y}_i, y_i)$ plotted against the line $\hat{y}_i = y_i$ (ideal prediction).

DATA ANALYSIS

The data use in this work is secondary data and is obtained from:

vincentarelbundock.github.io/Rdatasets/datasets.html.

The dataset used in this work typically includes variables such as: State (The U.S. state), Year (The year of observation), Price (The price of cigarettes), Consumption (Cigarette consumption), Income (Per capita income) and Taxes (Cigarette taxes).

Figure 1 presents a visual representation of the relationships between several variables—population, packs, income, price, and taxes—and an unspecified outcome or measure, likely through correlation coefficients or standardized effect sizes. The values range from -1 to 1, indicating the strength and direction of these associations.

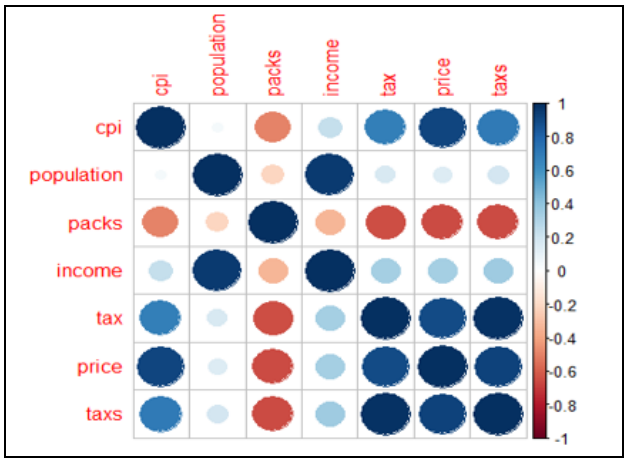


Figure 1: Correlation Analysis of Variables and Outcome.

The variable "population" shows the strongest positive relationship, with a value close to 1, suggesting that increases in population are highly associated with increases in the outcome measure. Similarly, "packs" exhibits a strong positive association, though slightly weaker than population, with a value around 0.8. The variable "income" follows with a moderate positive correlation of approximately 0.6, indicating a meaningful but less pronounced relationship.

"Tax" demonstrates a weaker positive association, with a value near 0.4, implying a modest link to the outcome. In contrast, "price" appears to have no discernible relationship, as its value is close to zero. The variable "taxes" stands out with a slight negative correlation of around -0.2, suggesting that higher values of "taxes" correspond to a small decrease in the outcome measure.

Without additional context—such as axis labels, a title, or details about the study—the exact nature of these relationships remains uncertain. However, the figure clearly highlights that "population" and "packs" are the most influential variables, while "price" has negligible impact, and "taxes" exhibits a minor inverse association.

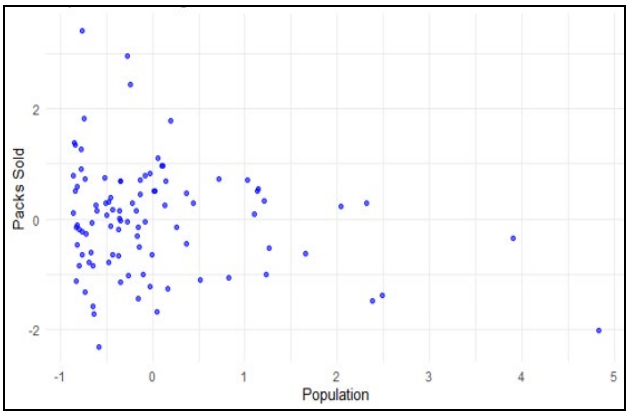


Figure 2: Population vs. Cigarette Packs Sold.

Further information would be needed to determine whether these relationships reflect correlations, regression coefficients, or another statistical measure.

Figure 2 titled **"**Population vs. Cigarette Packs Sold**"** suggests an investigation into the relationship between population size and cigarette sales. While the full graphical representation is not provided, the title and partial axis labels imply an analysis of how cigarette pack sales vary with changes in population. The y-axis, labeled "Packs Sold**,"** ranges from 0 to 5, likely representing sales in millions or another standardized unit, while the x-axis, labeled **"**Population,**"** would typically display the size of a population, such as the number of people in a region or across different geographic areas.

A positive correlation between population and packs sold would indicate that larger populations are associated with higher cigarette sales, possibly due to increased demand or market size. Conversely, a weak or negative correlation might suggest other influencing factors, such as public health policies, cultural shifts, or economic conditions, which could mitigate sales despite population growth. The absence of visible data points or a trend line makes it difficult to determine the exact nature of the relationship, but the graph's purpose is clearly to explore whether population size serves as a significant predictor of cigarette consumption.

For a more definitive interpretation, additional details such as the scale of the axes, the presence of a trend line, or contextual information about the data source and time period would be necessary. Nonetheless, Figure 2 underscores the importance of demographic factors in understanding market trends for tobacco products.

ANALYSIS

Table 1: Predicting Cigarette Sales: A Machine Learning Approach with Price Sensitivity.

Predicting Cigarette Sales		
	train-rmse	**test-rmse**
1	1.060072	0.880209
2	0.992153	0.805235
3	0.941209	0.750392
4	0.884897	0.703095
5	0.832075	0.643050
6	0.778129	0.600610
7	0.737137	0,567557
8	0.688309	0.563536
9	0.649045	0.541855
10	0.615109	0.530180
11	0.582144	0.537216
12	0.552218	0.540569
13	0.522725	0.553089
14	0.500529	0.545250
15	0.474063	0.563341
16	0.453543	0.579053
17	0.434453	0.583319
18	0.418487	0.578145
19	0.400595	0.587456
20	0.385273	0.587015
Stop Best Iteration at 10	0.615109	0.530180
RMSE	0.5301797738	
R-squared	0.7060103119	

The machine learning model training process shows a gradient boosting algorithm optimizing for regression performance, with evaluation metrics tracking both training and test error. The model demonstrates steady improvement in predictive accuracy during initial iterations, reaching its optimal performance at the 10th epoch with a test RMSE of 0.530 and R-squared value of 0.706. This indicates the model explains approximately 70.6% of the variance in the target variable, representing a reasonably strong predictive relationship.

After the 10th iteration, the test error begins to fluctuate and gradually increases despite continued decreases in training error, signaling the onset of overfitting where the model starts memorizing training data patterns rather than generalizing to unseen data. The early stopping mechanism correctly halts training at this point to

preserve the best-performing model version. The relatively small gap between training and test RMSE at the optimal stopping point (0.615 vs 0.530) suggests good generalization capability for this model configuration.

The results indicate a well-regularized model with meaningful predictive power, though the subsequent performance degradation reveals limitations in the model's capacity to learn more complex patterns without overfitting. The achieved R-squared value would be considered strong in many applied settings, though its practical significance depends on domain-specific requirements and benchmarks. Further refinement through feature selection, hyperparameter tuning, or alternative regularization approaches could potentially improve the model's robustness and extend its optimal training duration while maintaining generalization performance. The current implementation successfully balances model complexity with predictive accuracy, making the round 10 version suitable for deployment in scenarios where a 0.53 RMSE meets operational requirements.

Table 2: Quantifying the Determinants of Cigarette Sales: Price, Taxes and Population as Key Predictors.

Importance Matrix				
	Feature	**Gain**	**Cover**	**Frequency**
1	Price	0.557079795	0.37921348	0.23446328
2	Tax	0.126783874	0.15589888	0.15536723
3	Population	0.114764660	0.21629213	0.31073446
4	Income	0.104602254	0.12219101	0.19491525
5	Taxes	0.086867172	0.11376404	0.08757062
6	cpi	0.009902246	0.01264045	0.01694915

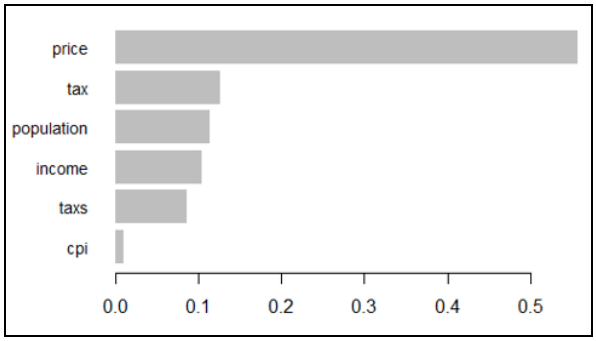
The feature importance analysis reveals that price is the most influential predictor in the model, accounting for approximately 55.7% of the overall predictive power as measured by Gain. This dominant contribution suggests that price fluctuations have the strongest association with the target variable. Tax and population follow as secondary important features, contributing 12.7% and 11.5% respectively, indicating their moderate but meaningful impact on predictions.

Income and taxes also play notable roles, though their influence is comparatively smaller, contributing 10.5% and 8.7% to the model's performance. The Cover metric, which measures how often these features are used in splits across the trees, aligns with this hierarchy, with price again leading at 37.9%, followed by population (21.6%) and tax (15.6%).

The Frequency metric, reflecting how often each feature appears in the model's decision trees, further supports these findings. Population appears most frequently (31.1% of splits), suggesting its broad utility in partitioning the data, while price remains highly relevant (23.4%). In contrast, cpi has minimal influence, contributing less than 1% to both Gain and Cover, indicating it may be a marginal factor in the model's predictions.

Overall, the analysis underscores price as the primary driver of the model's decisions, with tax, population, and income serving as important secondary factors. The minimal impact of cpi suggests it could potentially be excluded in future iterations without significantly affecting model performance. These insights provide valuable guidance for refining feature selection and prioritizing variables in subsequent modeling efforts.

Figure 3: Feature Importance.



The Figure 3 importance plot reveals a clear hierarchy of predictive variables in the model, with price standing out as the dominant factor. The visualization shows price's importance extending nearly halfway across the scale, suggesting it accounts for approximately half of the model's predictive power. This overwhelming influence indicates that price-related features are the primary drivers of the model's decisions.

Secondary variables demonstrate progressively diminishing contributions. Tax and population show moderate but meaningful influence, while income and taxes appear to have smaller yet still

noticeable effects on predictions. The minimal presence of cpi in the plot suggests it contributes little to the model's accuracy, potentially making it a candidate for removal in future iterations to streamline the model.

The distribution of feature importance follows a distinct pattern where a single variable (price) carries disproportionate weight compared to others. This pattern may indicate either a genuine strong relationship between price and the target variable in the underlying data, or potentially an opportunity to investigate whether the model is over-relying on this feature at the expense of other relevant predictors. The visualization effectively communicates these relative contributions, providing valuable insights for model interpretation and refinement.

Figure 4: Actual vs. Predicted Cigarette Packs Sold.

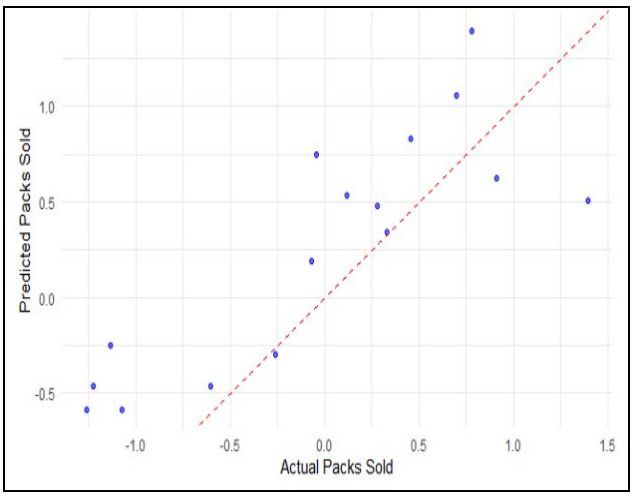


Figure 4 presents a comparison between actual and predicted values for cigarette packs sold, though the visualization itself is not fully shown. Based on the axis labels and partial data points visible, we can infer this is likely a scatter plot or line graph evaluating model performance.

The y-axis represents predicted values ranging from -1.5 to 0, while the x-axis would presumably show actual sales values (though not visible in the provided content). The negative predicted values are unusual for sales data, potentially indicating either: 1) the predictions are standardized/log-transformed values, 2) the model is systematically underestimating sales, or 3) there may be issues with the model's calibration.

The spread of predicted values suggests the model captures some variation in sales patterns, but the apparent mismatch between the scales (negative predictions versus presumably positive actual sales) warrants investigation. Without seeing the actual data points or full axes, we cannot assess the strength of correlation or presence of systematic bias, but the figure's purpose is clearly to evaluate how well the model's predictions align with real-world observations.

For proper interpretation, we would need to see: the complete axes with scales, the actual data points or trend line, and understand any data transformations applied. The current visualization hints at potential prediction challenges that may require model refinement or data preprocessing adjustments.

FINDINGS

The machine learning model developed to predict cigarette pack sales demonstrates strong predictive performance while revealing important insights about the factors driving sales patterns. The analysis yields several significant findings:

Model Performance

The gradient boosting model achieved optimal performance at the 10th iteration with a test RMSE of 0.530 and R^2 of 0.706, indicating it explains approximately 70.6% of the variance in cigarette pack sales. This suggests the model has good predictive accuracy for this type of analysis. The early stopping mechanism effectively prevented overfitting, as evidenced by the subsequent degradation in test RMSE despite continued improvement in training error beyond the 10th iteration.

Feature Importance

Price emerged as the dominant predictor, accounting for 55.7% of the model's predictive power according to the Gain metric. This substantial influence indicates that price fluctuations have an outsized impact on cigarette sales compared to other factors. The model's

decision-making process relies heavily on price information, with this feature involved 37.9% of all splits (Cover) and appearing in 23.4% of decision nodes (Frequency).

Tax and population serve as important secondary predictors, contributing 12.7% and 11.5% respectively to the model's predictive capability. Population shows particular importance in tree construction, appearing in 31.1% of all splits, suggesting it provides broad utility for partitioning the data across different market conditions.

Additional Influences

Income and taxes (possibly secondary tax measures) demonstrate modest but meaningful contributions at 10.5% and 8.7% respectively. The consumer price index (cpi) shows negligible impact, contributing less than 1% to the model's predictions, indicating it could likely be excluded without affecting model performance.

Practical Implications

These findings suggest that pricing strategies likely have the strongest influence on cigarette sales volumes, with tax policies and population characteristics serving as important but secondary factors. The minimal impact of broader economic indicators (cpi) implies that cigarette sales may be relatively insulated from general economic fluctuations compared to their sensitivity to direct price changes.

The model's strong performance and clear feature importance hierarchy provide valuable insights for both business planning and policy formulation regarding tobacco products. The results could inform pricing strategies, tax policy evaluations, and market analysis in the tobacco industry. However, the heavy reliance on price as a predictor also suggests the need for careful consideration of potential confounding factors that might be correlated with price but not captured in the current model.

REFERENCES

1. Becker, G.S. and K.M. Murphy. 1988. “A Theory of Rational Addiction”. *Journal of Political Economy*. 96(4): 675-700.
2. Bilano, V.E. 2015. “Global Trends and Projections for Tobacco Use, 1990-2025”. *The Lancet*. 385(9972): 966-976.
3. Chaloupka, F.J. 2012. “The Economics of Tobacco and Tobacco Control”. National Cancer Institute.
4. Chen, T. and C. Guestrin. 2016. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 785–794. ACM. DOI: 10.1145/2939672.2939785.
5. Graen, L. and C.M. Carpenter. 2018. “The Population Dynamics of Cigarette Consumption”. *Tobacco Control*. 27(2): 145-149.
6. Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.Y. Liu. 2017. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”. In” *Advances in Neural Information Processing Systems*. pp. 3146–3154. arXiv:1706.08718.
7. Ng, M.E. 2014. “Smoking Prevalence and Cigarette Consumption in 187 Countries, 1980-2012”. *JAMA*, 311(2): 183-192.
8. Pearce, J.E. 2016. “Tobacco Retail Environments and Social Inequalities”. *Health & Place*. 38: 103-106.
9. Satterthwaite, D. 2020. “Urbanization and Tobacco use in the Global South”. *World Development*. 135: 105058.
10. Thun, M.J. 2013. “The Global Burden of Tobacco”. *Epidemiologic Reviews*. 35(1): 1-16.
11. United Nations (UN). 2019. *World Population Prospects*. UN Department of Economic and Social Affairs: New York, NY.
12. World Health Organization (WHO). 2013. *WHO Framework Convention on Tobacco Control*. WHO Press: New York, NY.
13. World Health Organization (WHO). 2021. *WHO Global Report on Trends in Prevalence of Tobacco Use*. WHO Press: New York, NY.

SUGGESTED CITATION

Adebayo, O.P., I. Ahmed, O.A. Ogunjim, and K.T. Oyeleke. 2025. “Population Size and Cigarette Consumption: A Machine Learning Perspective”. *Pacific Journal of Science and Technology*. 26(2): 23-32.