**OPIM 5604 – Section 713**

**Predictive Modeling**

**Professor: Pankaj Prakash**

**Pranaydeep Khare, Vikram Krishnareddy, Bharath Kumar Reddy Pahti & Eva Wiggins**

**Project White Paper**

**Bike Sharing**

**Table of Contents:**

**Executive Summary 3**

**Business Understanding**

**Methodology**

**Evaluation**

**Deployment**

**References**

**Appendix**

**Executive Summary:**

This project analyzes how to develop and use various predictive models for the forecasting of use and demand of bike rentals in a bike sharing business venture. The need for an accurate forecast is necessary for maximum business optimization using the historical data available. Using data points on variables such as season, temperature, humidity, weather pattern, day of week, membership and occasional use allows the business to optimize inventory needs, while reducing cost and improve user satisfaction.

For successful business outcomes the need to understand customer behavior as well as likes and dislikes is essential. These datapoints will then allow the business to tailor the user experience to a personal level.

Initially we looked at basic carts to determine if the historical data could point to usage trends. These charts showed that the data points to trends based on weather patterns, casual vs registered users and time of day. For more precise predictions to tailor personalized service, a tracking system that show patterns in individual usage as well as preferred routes and duration will improve ROI.

To find our best model in predicting customer need, we used different models to predict demand. Some of the models included linear regression, decision tree, boosted tree, bootstrap forest and neural networks.

The bootstrap forest model emerged as our best model, showing an R2 score of 0.854 on the training and 0.781 on the validation data. The business can use the model prediction a daily operations forecast that determines bike distribution and allows for cost savings while boosting profitability and achieving better health and wellness objectives for the general population.

Our model can be deployed into a daily business operations matrix that informs business decisions.

**Business Understanding:**

Bike sharing programs important in the ability to improve health outcomes for users, but they are also a great predictor of the mobility of people, more so than other transit systems. Transit systems account for 21% of all CO2 emissions in 2022.

A diagram of a bike sharing system

AI-generated content may be incorrect. (Clockston, November, 2021)

To combat traffic congestion and improve climate change related issues of CO2 pollution where transportation accounts for 20.7% of global fossil CO2 emmissions (Worldometer, n.d.), introducing bike share programs will be a valuable tool. It also will deliver positive health benefits. The company currently has inconsistent availability of bikes due to changing demand as well as other variables. The goal is to reduce operational inefficiency by deploying our predictive model into our business operations. Once the model has been integrated into the daily operational matrix, the forecast can be automated. The goal is to reduce customer dissatisfaction due to lack of bike availability, pinpoint areas of opportunity and increase profitability.

Based on usage data, we can predict high volume usage and direct alternate transport as needed or even add more bikes during high volume times/dates.

A major operational challenge is fleet rebalancing or knowing how many bikes are needed, when and where.

Predict bike use demand with a higher accuracy to:

* Improve availability: ensure supply meets demand
* Optimize planning by reducing redundancies and unnecessary labor expenses
* Elevate the client user experience by maximizing availability and offering personalized rewards programs
* Inform strategic decision making

To improve the client experience we strive to prevent:

* Empty docking stations when needing to rent
* No available docks when returning bikes
* Reduce issues with broken bikes or bikes in need of repair
* High costs due to inefficiencies in the distribution model

**Methodology:**

Our target variable is the number (count) of bikes rented per hour. The data has 15 columns and 17,379 rows of hourly data. (Bike Sharing Dataset, 2013)

* **Timestamp features**: hour, weekday, month, year.
* **Weather features**: temperature, humidity, wind speed, weather condition.
* **Categorical variables**: holiday, working day.
* **Target variables**: casual, registered, and total count (sum of casual and registered).

Having looked at the initial data, we determined based on creating charts looking at variables such as registered vs. casual users and weather as well as seasonal data that we could find correlations that are usable when running prediction models.

A screenshot of a graph

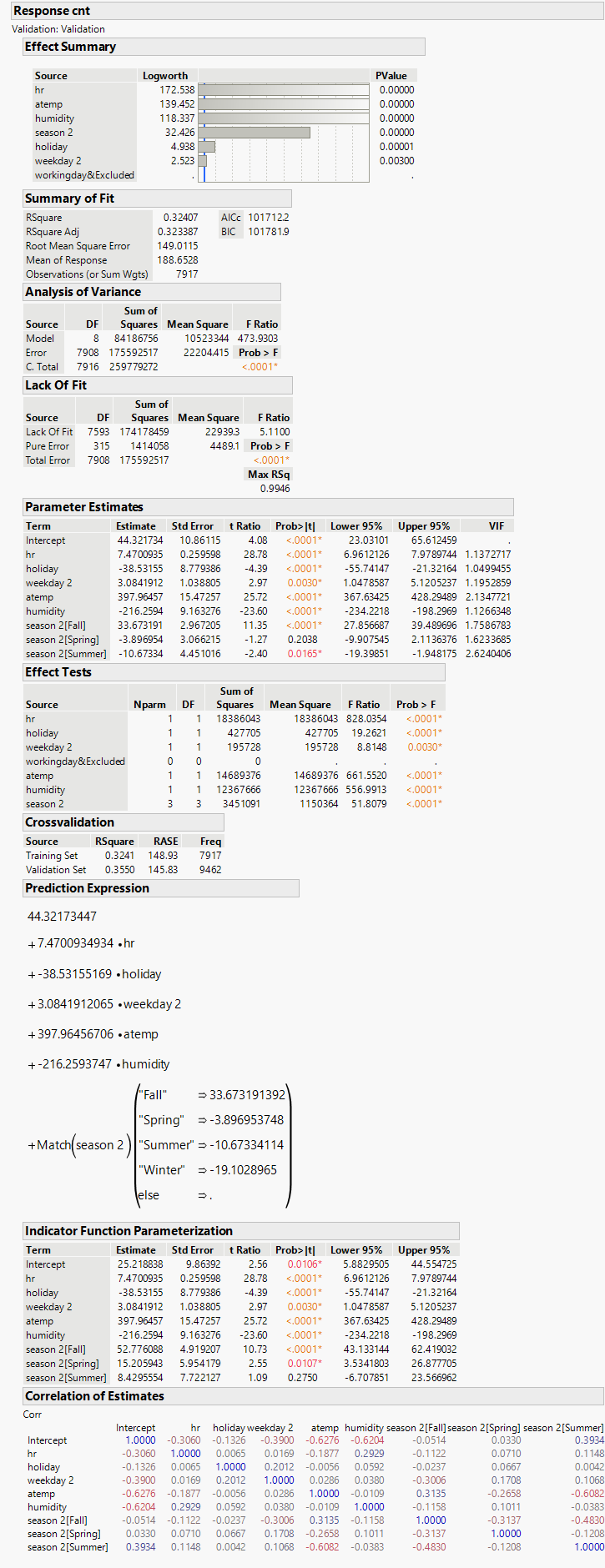
AI-generated content may be incorrect.  A graph on a white background

AI-generated content may be incorrect.

A graph on a white background

AI-generated content may be incorrect.  A graph on a white background

AI-generated content may be incorrect.



**Linear Regression**

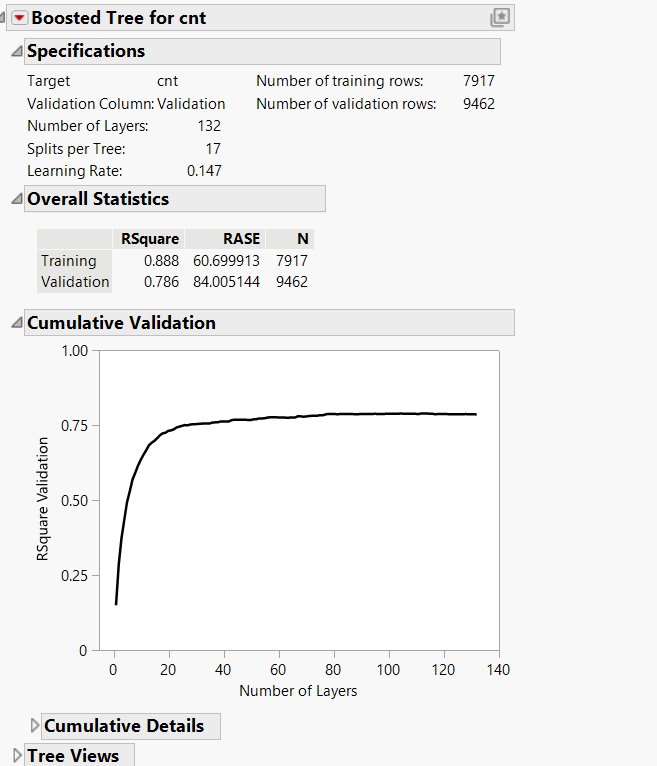
Model Performance

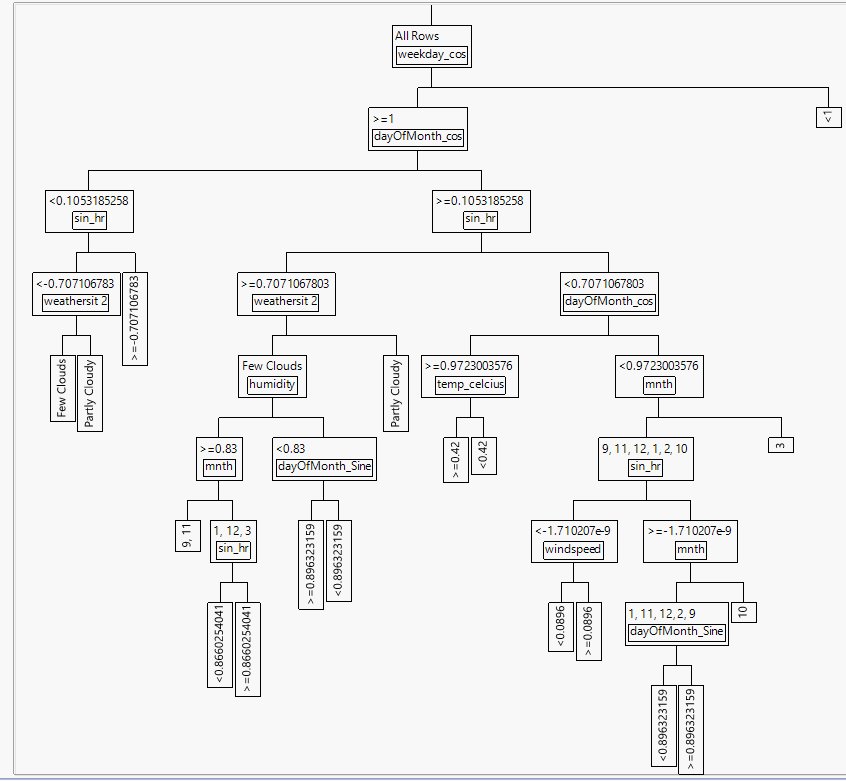
* Method: Simple baseline model​
* Performance:​
* Training R²: 0.3241​
* Validation R²: 0.3550​
* Validation RMSE: 149.01​
* Mean of Response: 188.6​

​

Key Insights:​

* Captures general demand patterns​
* Works well as a baseline model​
* Limited in handling non-linear relationships





**Boosted Tree:**

Model Performance​

* Training R²: 0.888​
* Validation R²: 0.786​
* Training RASE: 60.70​
* Validation RASE: 84.01​
* ​

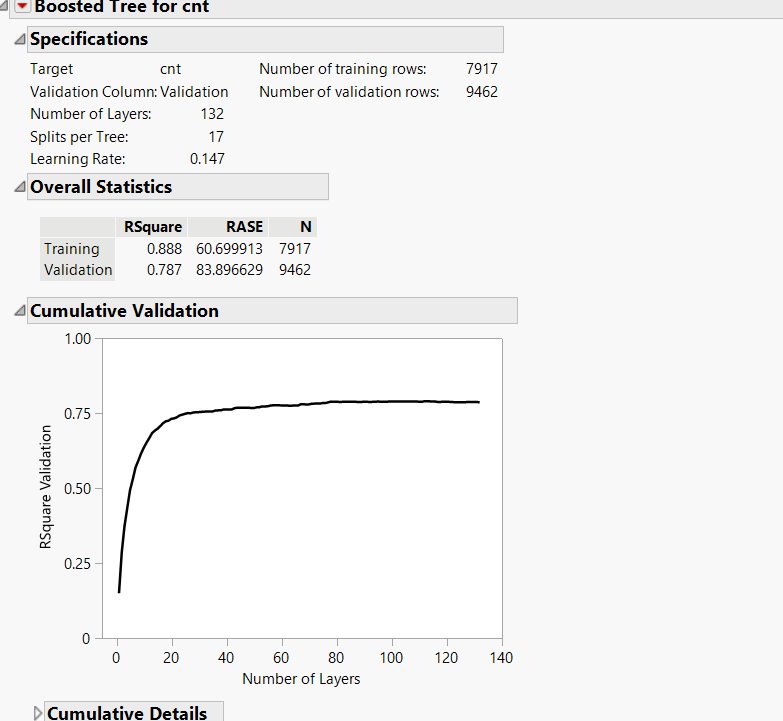
Key Parameters:​

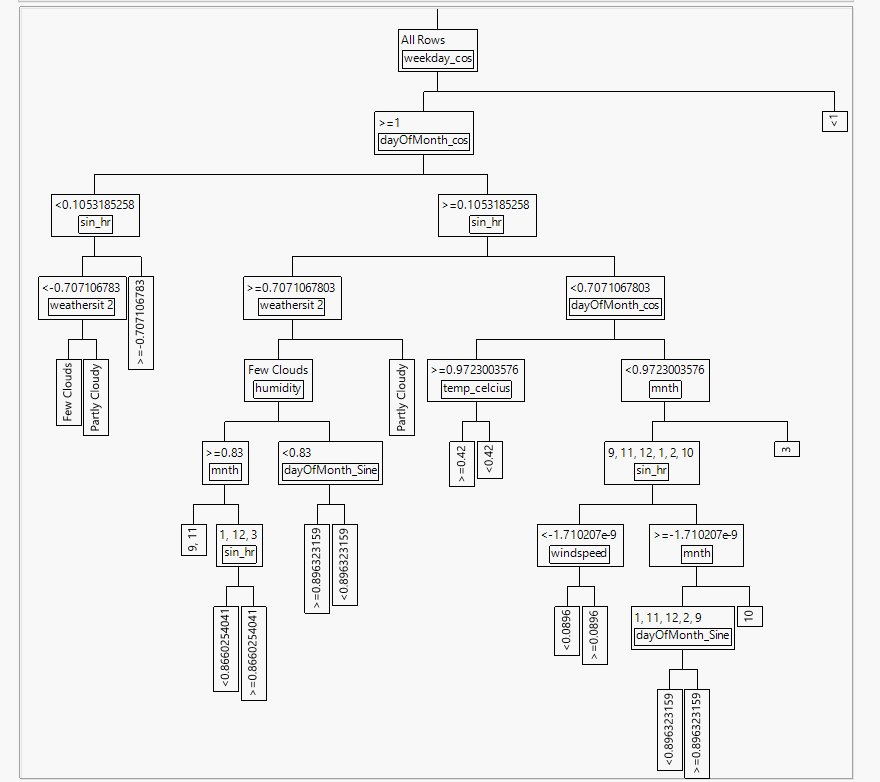
* Number of Layers (Trees): 132​
* Splits per Tree: 17​

​

Key Insights:​

* Good model fit with high R² on training (0.888) and decent validation R² (0.786).​
* Slight overfitting observed due to the performance gap between training and validation.​
* Majority of performance improvement occurs within the first 30–50 layers.​
* Validation R² curve flattens after ~60 layers, suggesting additional layers offer limited benefit.​
* Model is reasonably accurate but could be further optimized to reduce validation error.



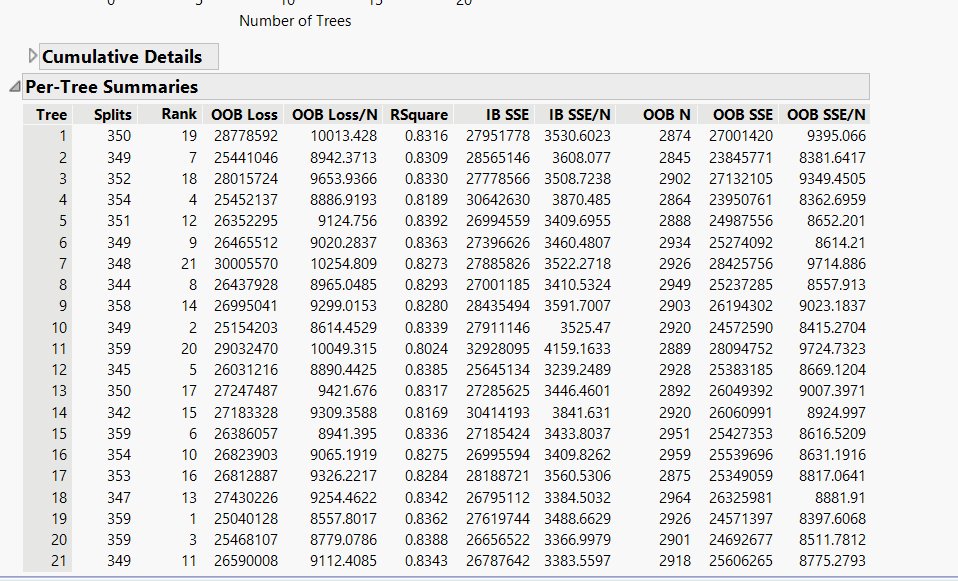


**Decision Tree:**

**Model Performance:**​

* Training R-Square: 0.888 (Good Fit)​
* Validation R-Square: 0.787 (Strong Generalization)​
* Training RASE: 60.70​
* Validation RASE: 83.90​
* Performance consistent with typical model behavior​
* Slight performance drop indicates minimal overfitting​
* **Key Parameters:**​
* Number of Layers: 132​
* Splits per Tree: 17​
* Learning Rate: 0.147​
* Key Insights:​
* High Predictive Power: Explains ~89% of training variance, ~79% of validation​
* Slight Overfitting: Acceptable given model complexity​
* Learning Rate Impact: High-rate speeds learning, may affect stability​
* Well-balanced model for accuracy vs. complexity





Bootstrap Forest:

Model Performance: ​

* Training R-Square: 0.854​
* Validation R-Square: 0.781​
* Training RASE: 69.13​
* Validation RASE: 85.04​

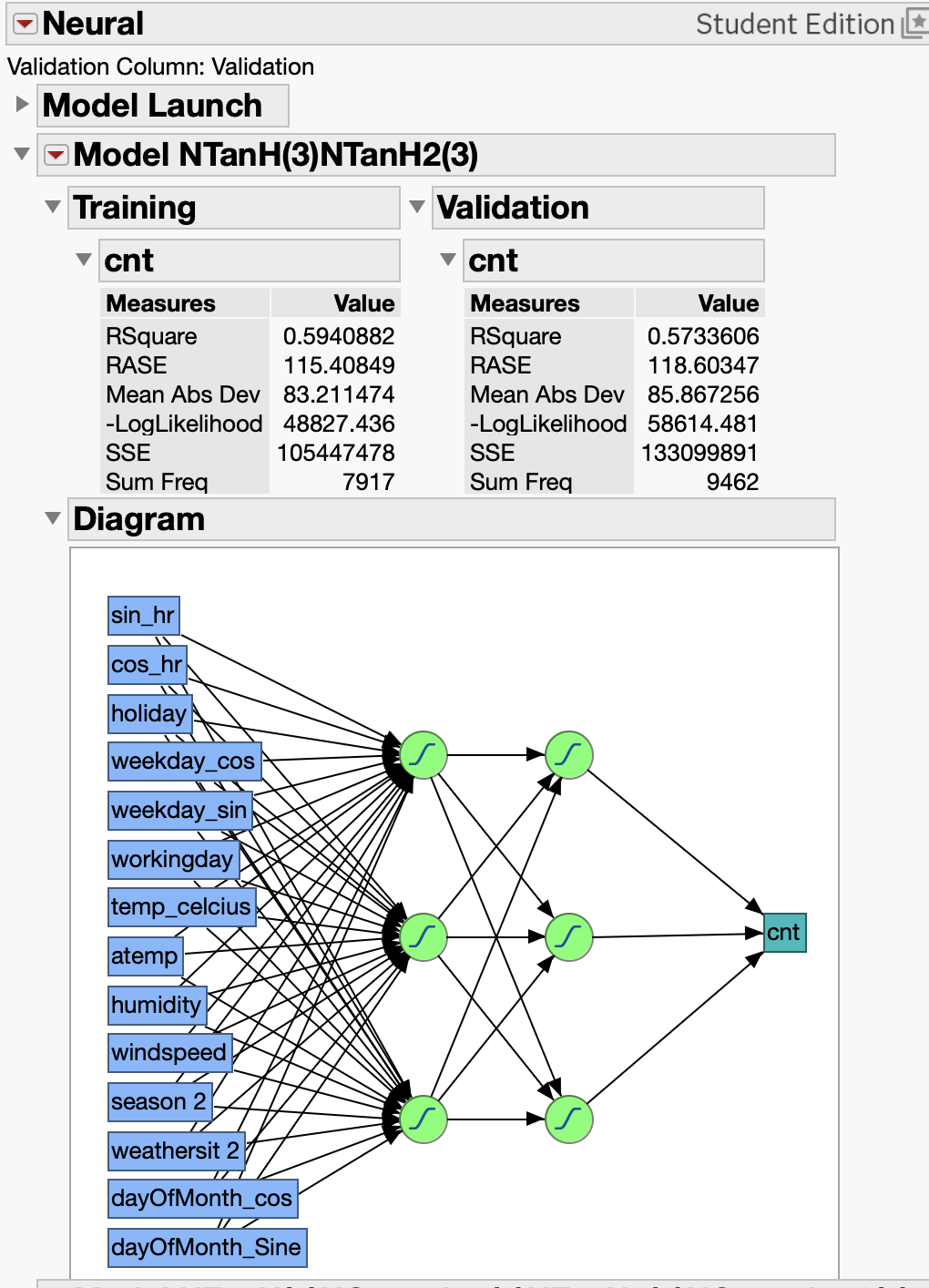
**Key parameters:** ​

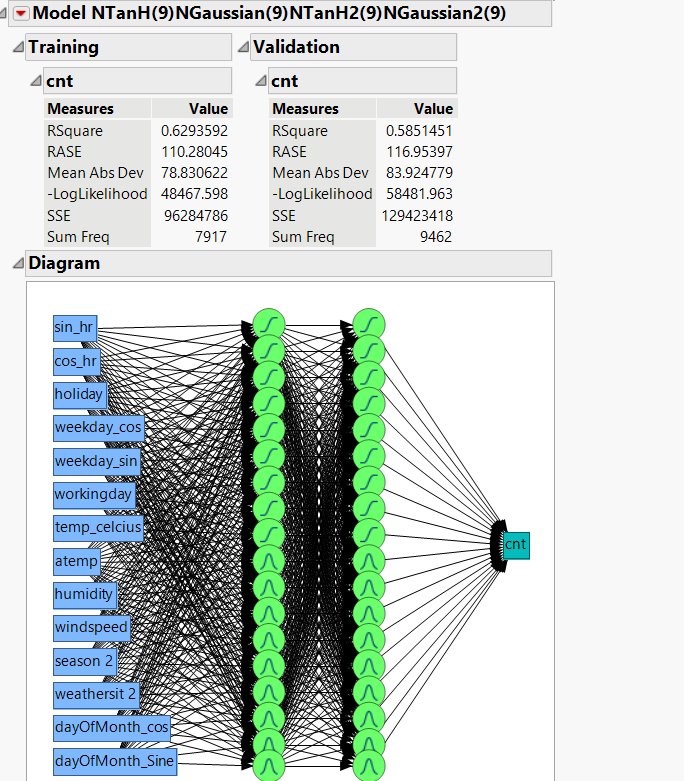
* Out-of-Bag RASE: 59.42​
* In-Bag RASE: 93.97​
* Validation stabilizes by ~21 trees​

​

**Key Insights**:​

* Strong performance with fewer trees (54)​
* Ensemble reduces variance vs. single trees​
* Out-of-bag validation helps reduce need for test data​
* Good generalization: validation R-Square ~78%RASE suggests modest error in predictions





**Evaluation:**

**Model R2 Score RMSE/RASE**

Linear 0.32407 145.83

Decision Tree 0.888 83.89

Boosted Tree 0.888 85.03

Bootstrap Forest 0.854 84.01

Neural (TanH) 0.58549 121.88

Neural (Gaussian) 0.629359 116.59

Best model for our business operations:

**Business Case Development**

Forecast demand:

* Cost savings due to supply management
* Customer Retention and engagement – offer different incentives to casual users and regular users
* Maximize Revenue

ROI Considerations:

* Cost of model development tools and integration into business model
* Benefits of reduced costs and increased revenue
* Reassess model periodically

**Deployment:**

A graph of bicycle sharing

AI-generated content may be incorrect.(Richter, 2015)

Bike sharing businesses have changed transportation in many parts of the world, especially in urban areas as traffic congestion, sustainability and health trends have become a focus not only for legislators but people in general. By capitalizing on the capabilities of predictive modeling, our business is able to adjust the operational objectives to become proactive. The model when implemented will allow the business to reduce cost and improve the client experience. Continued development and data insights will allow the business to offer a personalized service to customers to develop loyalty and improved ROI.

Our best model enables real time data driven solutions, reduce inefficiencies and improve client satisfaction.

**Implementation:**

* Integrate predictive model into business operational matrix
* Update predictions with local event calendar and weather forecasts
* Notifications for out of model trends

**Challenges:**

* Privacy concerns especially in user data
* Performance monitoring
* Oversight and monitoring depend on human interpretation
* Vandalism and theft

‘…Bike sharing has not only altered the fabric of urban transit but also sparked a cultural shift towards shared economies and sustainable living. The transformation from rudimentary beginnings to a tech-savvy, user-centric model exemplifies the potential of collaborative consumption and the power of innovation in shaping the future of urban transportation.’ (Faster Capital, 2025)

 (Singh & Katare, 2023)

**References:**

# Bibliography

Bike Sharing Dataset. (2013, December 19). UC Irvine Machine Learning Repository.

Clockston, R. L.-R. (November, 2021). Health Impacts of Bike-Sharing Systems in the U.S. *Environmental Research Vol 202*.

Faster Capital. (2025, April 12). *Bike Sharing Economy: Riding the Wave: The Business Models Behind Bike Sharing*. Retrieved from fastercapital.com: https://fastercapital.com/content/Bike-Sharing-Economy--Riding-the-Wave--The-Business-Models-Behind-Bike-Sharing.html

Richter, F. (2015). Bike-Sharing is Taking Off Around the World. *Statista*.

Singh, O., & Katare, L. J. (2023, June). Retrieved from Allied Market Research: https://www.alliedmarketresearch.com/bike-sharing-market-A11309

Worldometer. (n.d.). *worldometers.info*. Retrieved from worldometer: https://www.worldometers.info/co2-emissions/