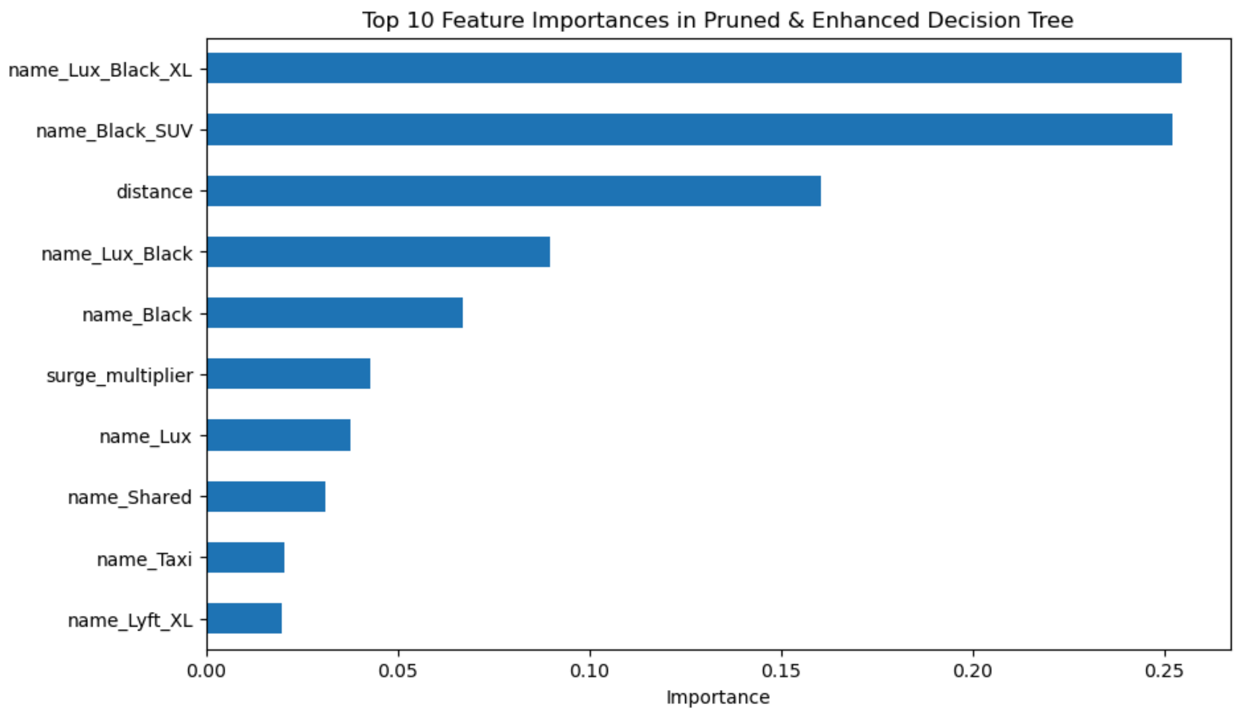
Holly Cohan, Yesh Onipede, Yu Xia

Week 11 Report: Risk, Bias and Ethical Considerations

The objective of this week's assignment is to build transparency into the prediction process and analyze potential risks and biases in the dataset and model output. This will be achieved by identifying important features, explaining individual predictions, evaluating bias in the input dataset and model outputs, and discussing potential risks of using the model.

Top 10 Important Features



The analysis of the top 10 features reveals interesting insights about the decision tree's key predictors for the target variable. The `name` features (e.g., `name_Lux_Black_XL`, `name_Black_SUV`) account for a significant portion of the predictive power in the decision tree. These variables represent the type of car used for the ride, which inherently reflects the pricing structure of the ride. For instance:

- Premium Car Types: Features such as `name_Lux_Black_XL`, `name_Black_SUV`, and `name_Lux_Black` suggest that luxury vehicles are often associated with higher fares due to superior service or exclusivity.
- Economy or Shared Options`: Features like `name_Shared` and `name_Taxi` highlight the importance of lower-cost options in determining the fare.

The `distance` variable is another critical factor and a logical driver for ride pricing. Longer distances typically translate to higher fares, which is likely why this variable ranks highly among the predictors. Similarly, the `surge_multiplier` feature indicates the influence of real-time demand-supply dynamics on pricing. While it ranks lower compared to `distance` and car types, it still plays a substantial role, particularly during peak hours or high-demand periods.

The combination of `name` features and `distance` likely captures how ride type influences pricing per unit distance. For instance, a `Lux_Black_XL` ride covering 10 miles will cost significantly more than a `Shared` ride for the same distance. The heavy reliance on `name` features could make the model sensitive to changes in ride type categorizations (e.g., if a new ride type is introduced). `distance` remains a stable and robust predictor, ensuring that predictions align with the intuitive understanding of ride costs.

Select 5 predictions at random

Sample	Actual Price	Predicted Price	Absolute Error	Percentage Error
1	9.5	10.58	1.08	11.41
2	36.0	20.50	15.50	43.06
3	22.5	19.60	2.903	12.90
4	13.5	16.84	3.34	24.76
5	45.0	48.18	3.18	7.07

For **Sample 2** (Actual Price: 36.0, Predicted Price: 20.50):

- The model predicted **20.50** instead of the actual value **36.0**, resulting in a significant error.
- **Feature Contributions:**
  - **Distance:** The distance contributed heavily due to its linear relationship with the price. For sample 2, the distance is 3.64 which is considered as average.
  - **Surge Multiplier:** Played a secondary role, amplifying or reducing the base rate. The surge multiplier for sample 2 is 1.0 which is the baseline level.
  - **name_Uber_XL:** This feature ranked as the 11th most important factor in the model's predictions.

Examining Sample 2, which has a relatively high error, we found that although all the key features are within a relatively average range, the model still fails to predict the actual price accurately. This discrepancy may be attributed to external factors, such as supply levels, which are not accounted for by the model but could significantly influence pricing.

Protected Categories and Biases Analysis
The points in our dataset represent rideshare instances rather than individuals, meaning there are no explicitly protected categories such as gender, race, ethnicity, or religion. However, certain variables, such as source and destination, which indicate the pickup and drop-off locations, could indirectly correlate with demographic or socio-economic characteristics based on geographic trends. For instance, some neighborhoods might predominantly consist of specific income brackets, racial groups, or other demographic profiles. Additionally, the level of car requested can serve as a potential indicator of economic status. For Example, features like `name_Lux_Black_XL` and `name_Black_SUV` dominate the list, reflecting that the model prioritizes distinguishing between high-end ride types. Variables like

`name_Shared` and `name_Taxi` are also significant but have comparatively lower feature importance. This might suggest a smaller variance in fare pricing within this category.

Features such as `name_Lux_Black_XL` and `name_Black_SUV` are heavily prioritized by the model, indicating the model may focus on high-end ride types. The higher MSE for these features' specific groups (e.g., name_Lux_Black_XL = 1) suggests the model struggles more with predicting prices for high-end rides, possibly due to greater variance or fewer training examples for these categories.

Bias Reduction Strategies
To mitigate bias, it is important to address the trade-off between fairness and accuracy through potentially rebalancing the dataset or penalizing disparities may slightly reduce the overall model performance but ensure more equitable predictions.We addressed potential biases in the model by removing outliers in the price variable, defined as values beyond three standard deviations from the mean. The intent was to minimize the undue influence of extreme, atypical fare values on the model's training process, which could skew predictions and reduce generalizability. After removing these outliers, we retrained the decision tree model on the modified dataset and re-evaluated its performance across the training, validation, and test sets. The results indicated a decline in the model's ability to generalize, with overfitting observed as the model became too tailored to the training data.

Performance Metrics for Original Data and Enhanced Data

Model Type	**Metrics**	**Training Set Values**	**Validation Set Values**	**Testing Set Values**
Pruned Decision Tree Model on Enhanced Data (with updated coordinates and eta_minutes_rh) – Original model	**MSE**	2.6845	3.0536	**3.0240**
	RMSE	1.6385	1.7474	**1.7390**
	R^2	0.9675	0.9627	**0.9639**
Pruned Decision Tree Model on Enhanced Data (with updated coordinates and eta_minutes_rh) with price outliers removed	**MSE**	2.4186	3.9584	**3.8990**
	RMSE	1.5552	1.9896	**1.9746**
	R^2	0.9668	0.9517	**0.9534**

In conclusion, simply removing outliers is not the best idea for generalization. To enhance the model's ability to capture higher price predictions while addressing bias, there is another strategy we can consider. First, instead of removing all outliers, we could use a more nuanced approach, such as applying robust scaling or transforming the price variable (e.g., log transformation) to reduce the impact of extreme values while retaining important information about higher fares. Another approach, which is likely more effective, is regularizing the model and fine-tuning hyperparameters to strike a balance between fitting higher prices and maintaining generalizability. This ensures the model performs equitably across all fare ranges. The current model uses hyperparameters inherited from the original model built in Week 7. While these parameters were effective for the original dataset, they may not be optimal for the pruned dataset after outlier removal. Updating these hyperparameters through techniques like grid search or random search could improve the model's ability to capture the complexity of higher price predictions while reducing overfitting to the training data. This adjustment would allow the model to better reflect the variability in fare prices, particularly at the higher end, without compromising its ability to generalize to unseen data.