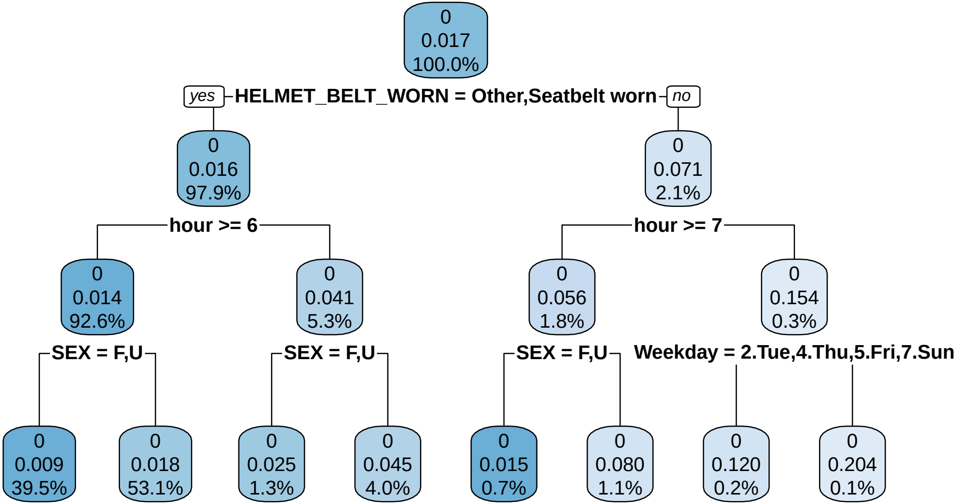
# Foundations of Statistical and Machine Learning for Actuaries
# (Tree-based) Ensemble methods and Interpretability

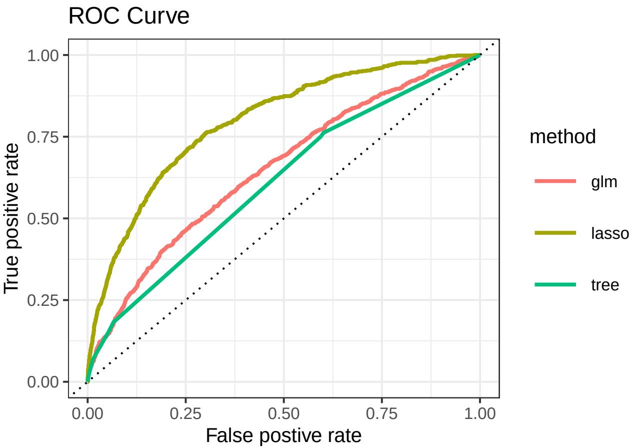
Edward (Jed) Frees, University of Wisconsin - Madison
Andres M. Villegas, University of New South Wales

July 2025

# Tree: VicRoads Crash Data



# Tree: VicRoads Crash Data (ROC)




**AUC**

| Method | Train | Test |
|---|---|---|
| glm | 0.663 | 0.650 |
| lasso | 0.809 | 0.797 |
| tree | 0.629 | 0.610 |

# Advantages and disadvantages of Trees

**Advantages**

- Easy to explain
- (Mirror human decision making)
- Graphical display
- Easily handle qualitative predictors

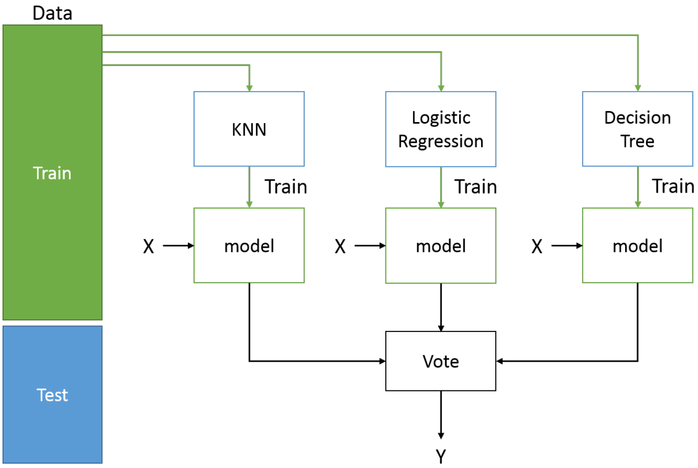
**Disadvantages**

- Low predictive accuracy compared to other regression and classification approaches
- Can be very non-robust

Is there a way to improve the predictive performance of trees?

- Ensemble methods
- Bagging, random forest, boosting

# Ensamble methods




Ensembles tend to have lower error and produce less overfitting

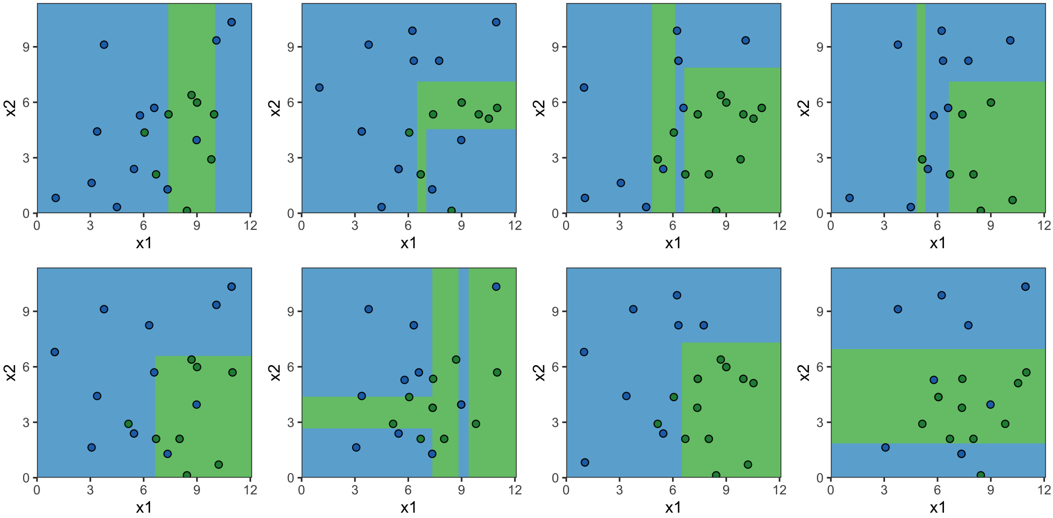
# Bootstrap Aggregation (Bagging)

- A **general-purpose** procedure to reduce the variance of a statistical learning method
  - particularly useful and frequently used in the context of decision trees

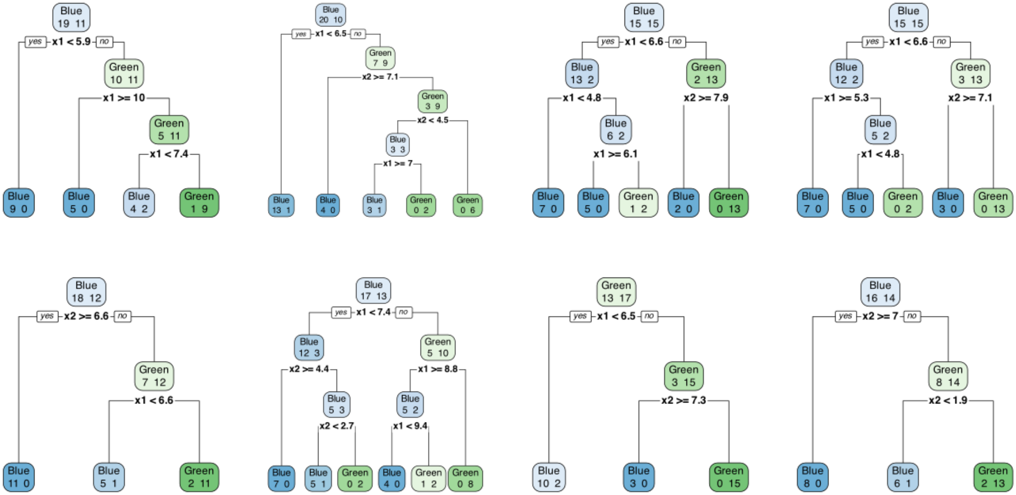
Bagging procedure

- Bootstrap
  - sample with replacement repeatedly
  - generate $B$ different bootstrapped training data sets
- Train
  - train on the $b$th bootstrapped training set to get $\hat{f}^{*b}(x)$
- Aggregate
  - Take a majority vote of all of the trained models

# Bagging: Illustration



# Bagging: Illustration



# Random forest

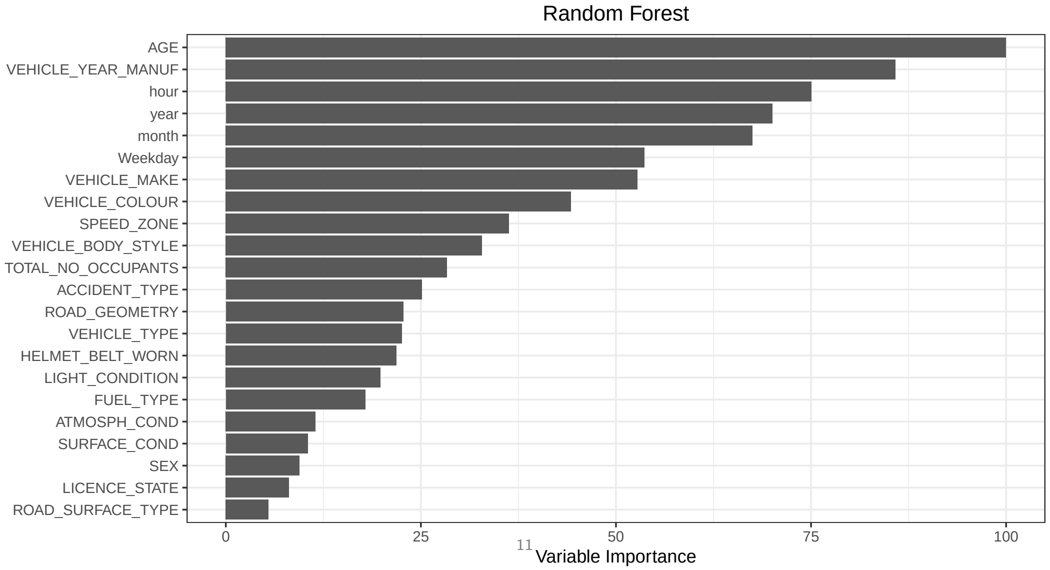
Random forests decorrelates the bagged trees

- At each split of the tree, a fresh random sample of $m$ predictors is chosen as split candidates from the full set of $p$ predictors
- Strong predictors are used in (far) fewer models, so the effect of other predictors can be properly measured.
  - Reduces the variance of the resulting trees
- Typically choose $m \approx \sqrt{p}$
- Bagging is a special case of a random forest with $m = p$

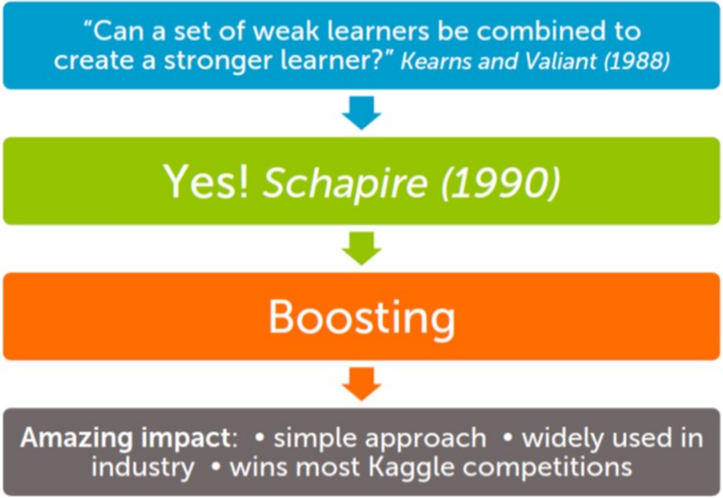
# Bagging/Random Forest: Variable Selection and Importance

- Bagging and Random Forest can lead to difficult-to-interpret results, since, on average, no predictor is excluded
- Variable importance measures can be used
  - Bagging classification trees: Gini index reduction for each split (measure of node purity)
- Pick the ones with the highest variable importance measure

# Random forest: Variable Importance VicRoads Data



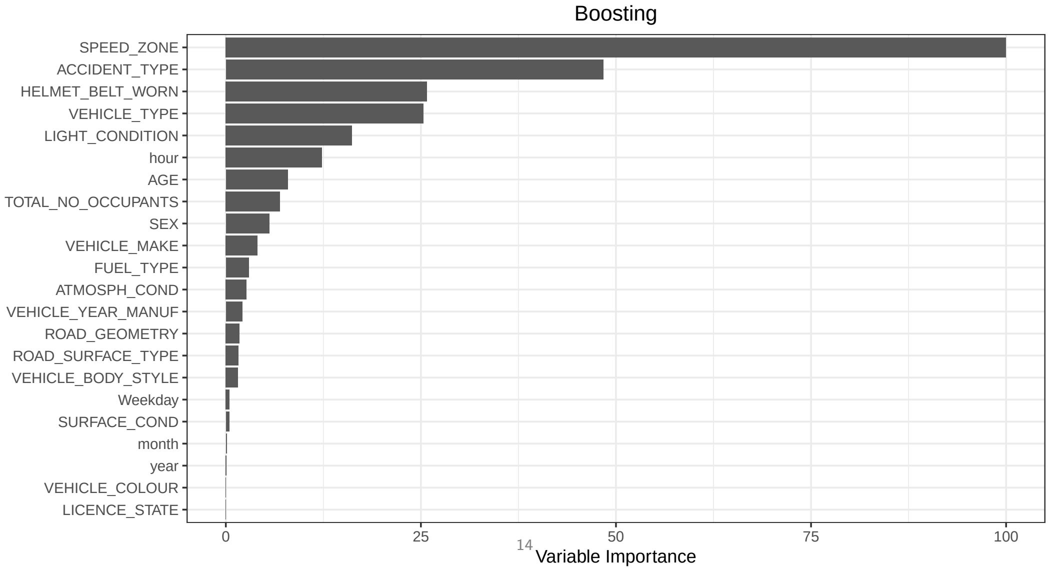
# Boosting motivation



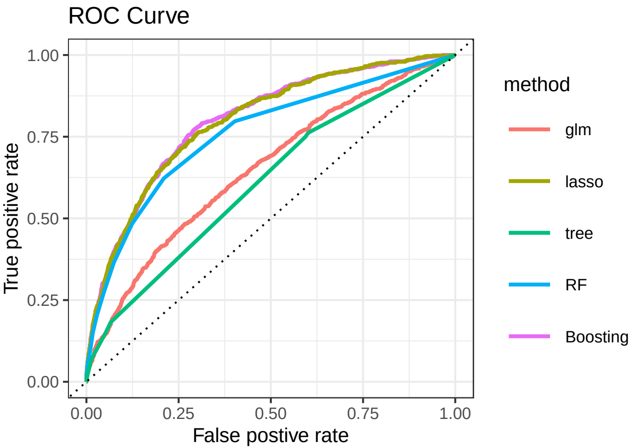
# Boosting procedure

- A general approach that can be applied to many statistical learning methods for regression or classification
- Involves combining a large number of decision trees
  - trees are grown sequentially
  - using the information from previously grown trees
  - no bootstrap - instead each tree is fitted on a modified version of the original data (sequentially)
- Unlike standard trees, boosting learns slowly - by focusing on the residuals and hence focusing on areas the previous tree did not perform well.

# Boosting: Variable Importance VicRoads Data



# Comparison of methods: VicRoads Crash Data




**AUC**

| Method | Train | Test |
|---|---|---|
| glm | 0.663 | 0.650 |
| lasso | 0.809 | 0.797 |
| tree | 0.629 | 0.610 |
| RF | 0.676 | 0.759 |
| Boosting | 0.813 | 0.800 |

# Summary of key concepts in (tree-based) ensemble methods and interpretability

- Bagging
    - Bootstrap aggregation
    - Reduces variance of a statistical learning method
    - Particularly useful for decision trees
- Random Forest
    - Decorrelates the bagged trees
    - Randomly selects predictors at each split
- Boosting
    - Sequentially builds trees
    - Focuses on residuals of previous trees
- Interpretability
    - Variable importance measures
    - Partial dependence plots
    - LIME
    - SHAP values