Washington D.C. Traffic Accidents Dashboard - Machine Learning Edition

Sam Clark and Joanne Choi
7/28/2022

Models

Random Forest

- Independently builds trees then uses bootstrap aggregation (bagging) and provides final output based on average or majority voting after combining results of all models

XGBoost

- Trees are built sequentially and weak learners are improved. Each new tree is boosted from the previous tree to improve the overall model

Predicting Number of Accidents

Modeling Methods

- Random Forest Regression
- Linear Regression
- XGBoost Regression

Features

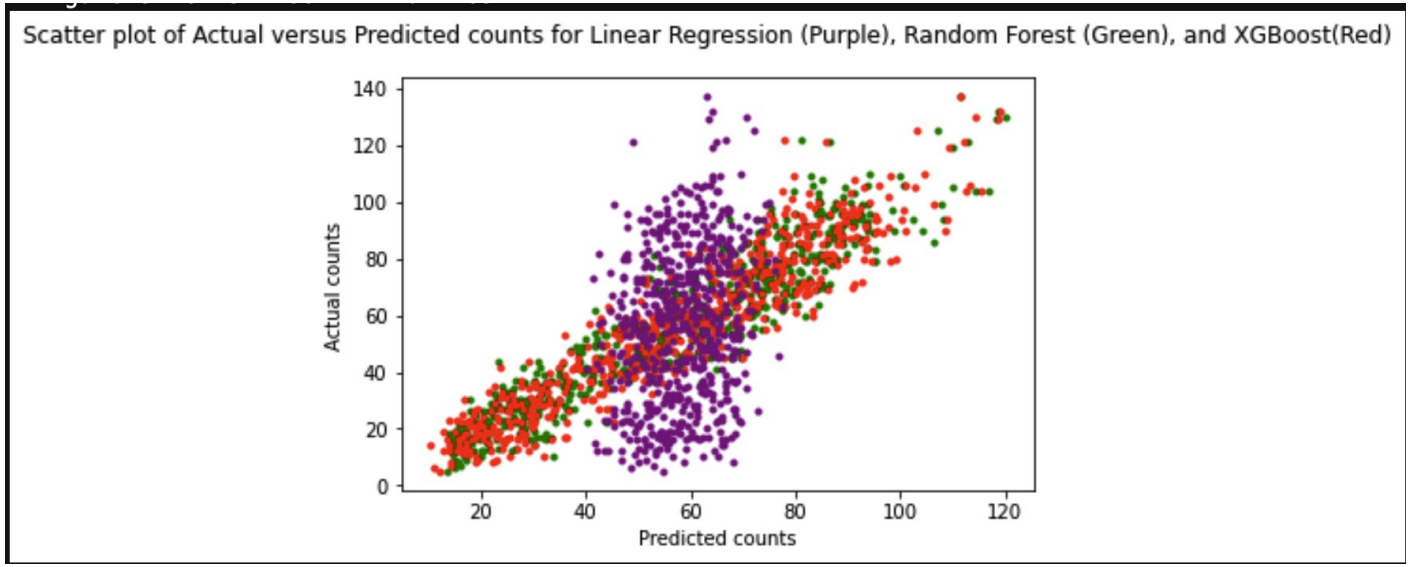
- Hour [0-23]
- Month [Jan-Dec]
- Weekday [Mon-Sun]
- Temperature
- Visibility
- Cloud Cover
- WindSpeed
- Precipitation
- Weather Conditions [Clear, Overcast, Partially Cloudy, Rain, Rain & Overcast, Rain & Partially Cloudy, Snow, Snow & Overcast, Snow, Partially Cloudy]

Regression results and metrics

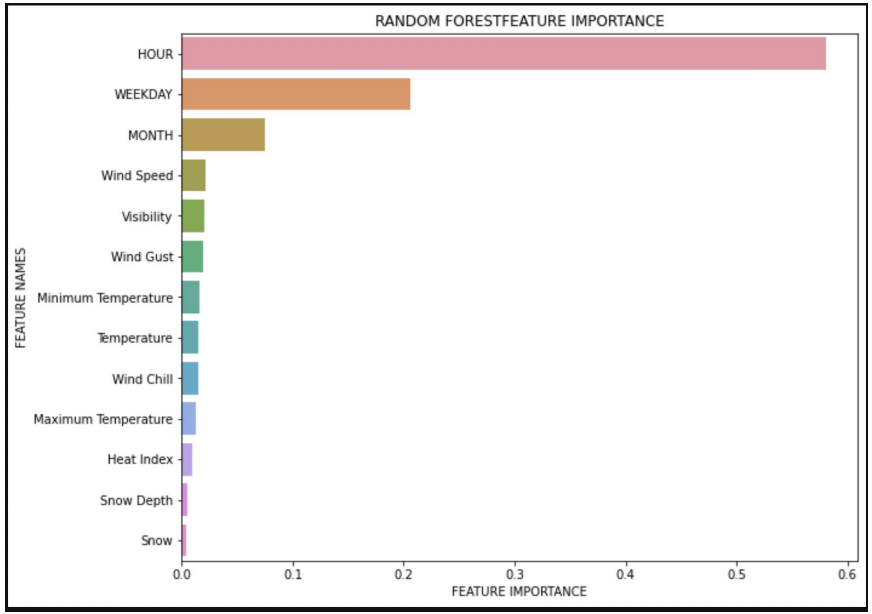
	Accuracy	RMSE	Minimum	Mximum
Linear Regression	29.60%	26.5	40.3	77.4
Random Forest	82.12%	9.6	12.5	120.2
XGBoost	81.22%	9.7	10.2	118.9

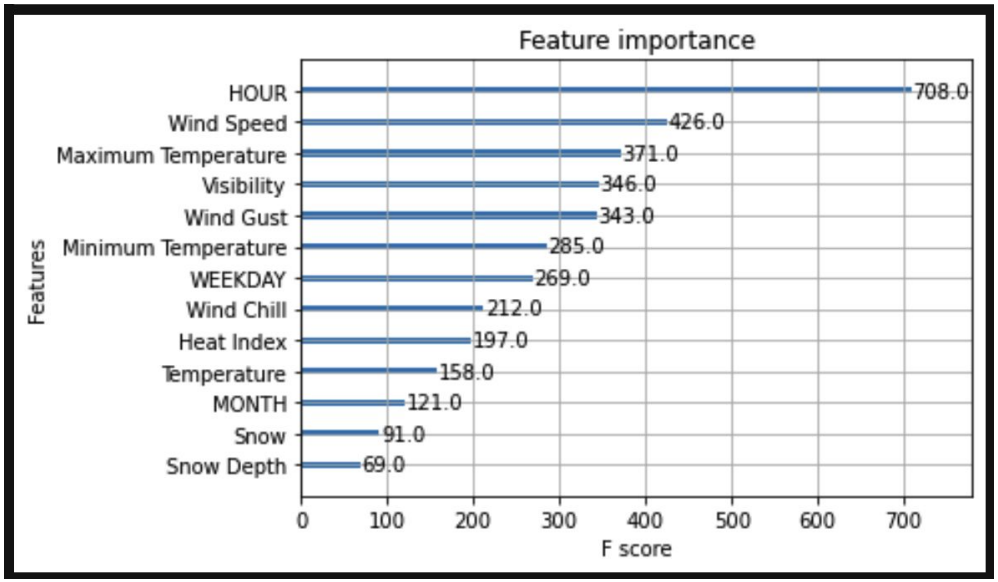
- Linear regression predictably performed poorly
- Random Forest and XGBoost had similar Performance

Comparison of the 3 Regressions



Feature Importance in Random Forest and XGBoost





Injury Severity Classification

Classes	Features
• Minor • Major • Fatal	• Hour [0-23] • Month [Jan-Dec] • Weekday [Mon-Sun] • Temperature • Visibility • Cloud Cover • WindSpeed • Precipitation • Weather Conditions [Clear, Overcast, Partially Cloudy, Rain, Rain & Overcast, Rain & Partially Cloudy, Snow, Snow & Overcast, Snow, Partially Cloudy]

Injury Severity Classification

Imbalanced Data Distribution

Severity Class	Count
Minor Injuries	16363
Major Injuries	864
Fatal Injuries	64

Injury Severity Classification

Random Forest (unbalanced)

Accuracy: 93%

	Recall	Precision	F1
Minor	0.99	0.95	**0.97**
Major	0.03	0.10	**0.05**
Fatal	0.00	0.00	**0.00**

Confusion Matrix

	Minor	Major	Fatal
Minor	**4797**	72	1
Major	243	**8**	0
Fatal	25	0	**0**

XGBoost (unbalanced)

Accuracy: 95%

	Recall	Precision	F1
Minor	1.00	0.95	**0.97**
Major	0.00	0.00	**0.00**
Fatal	0.00	0.00	**0.00**

Confusion Matrix

	Minor	Major	Fatal
Minor	**4870**	0	0
Major	251	**0**	0
Fatal	25	0	**0**

Injury Severity Classification

- Used SMOTE to balance the data

Class	Count
Minor Injuries	16224
Major Injuries	864
Fatal Injuries	64

Class	Count
Minor Injuries	16224
Major Injuries	16224
Fatal Injuries	16224

Injury Severity Classification

Random Forest (balanced & optimized)

Accuracy: 74%

	Recall	Precision	F1
Minor	0.99	0.79	**0.88**
Major	0.36	0.76	0.48
Fatal	0.86	0.68	**0.76**

Confusion Matrix

	Minor	Major	Fatal
Minor	**4804**	2	93
Major	1119	**1360**	2345
Fatal	138	441	**4300**

XGBoost (balanced & optimized)

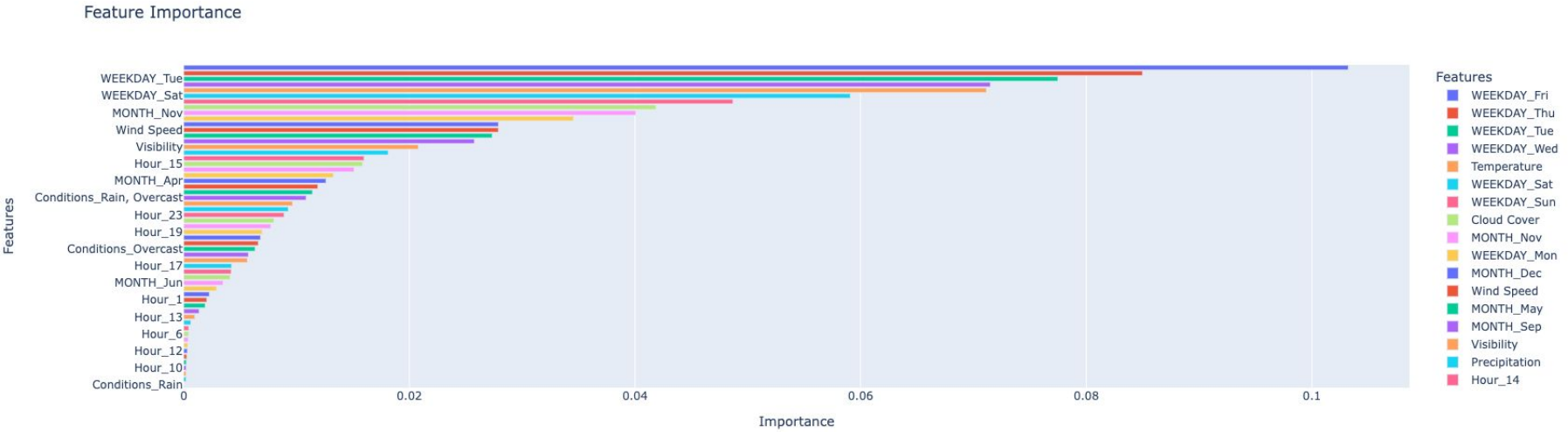
Accuracy: 97%

	Recall	Precision	F1
Minor	0.95	0.98	**0.96**
Major	0.98	0.94	**0.96**
Fatal	0.98	0.99	**0.99**

Confusion Matrix

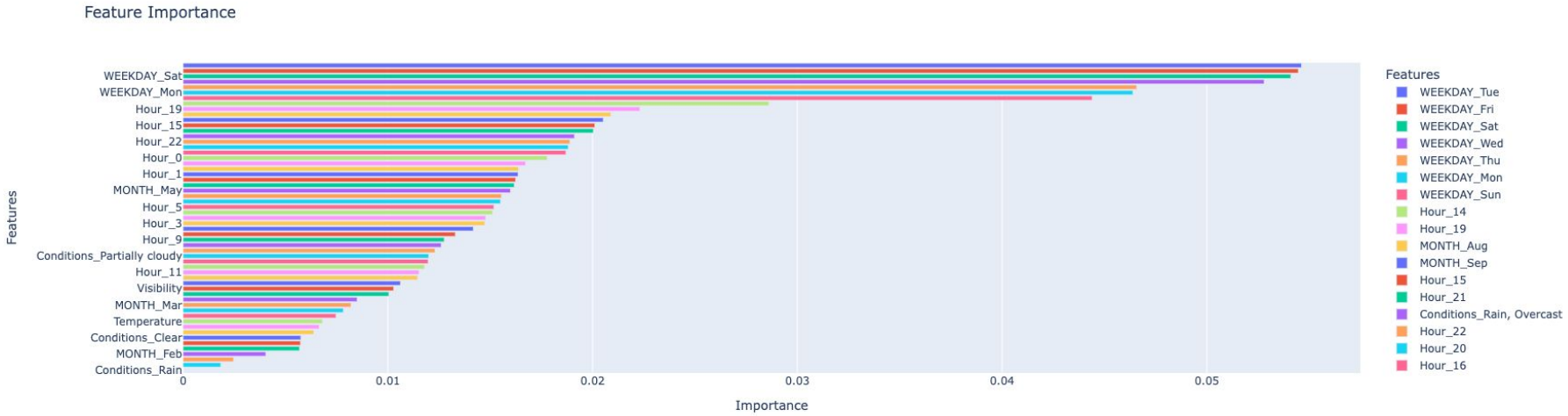
	Minor	Major	Fatal
Minor	**4795**	91	13
Major	225	**4518**	81
Fatal	23	7	**4849**

Injury Severity Classification



Random Forest Feature Importance

Injury Severity Classification



XGBoost Feature Importance

Next Steps

- Finish tuning hyperparameters for models
- Import models into the dashboard
- Continue working on the paper

UMBC

Thank you!