PREDICTING THE SEVERITY OF ROAD VEHICLE ACCIDENTS

How to improve rescue services in road vehicle accidents by implementing a machine learning classifier

28.08.2020, Germany

Content

Business Understanding

In the US there are:

- 280 million vehicles
- 6 million accidents per year
- 35.600 fatalities per year (2018)

How to reduce fatality rate?

- ML classifier to predict severity of accidents
- Enable rescue services to send appropriate equipment and number of paramedics

Data Understanding

Data set

- SDOT Traffic Management Division
- Traffic Records Group
- 2004 to present date
- 194.673 data points
- 38 columns
- 2 severity codes (prop damage / injury)

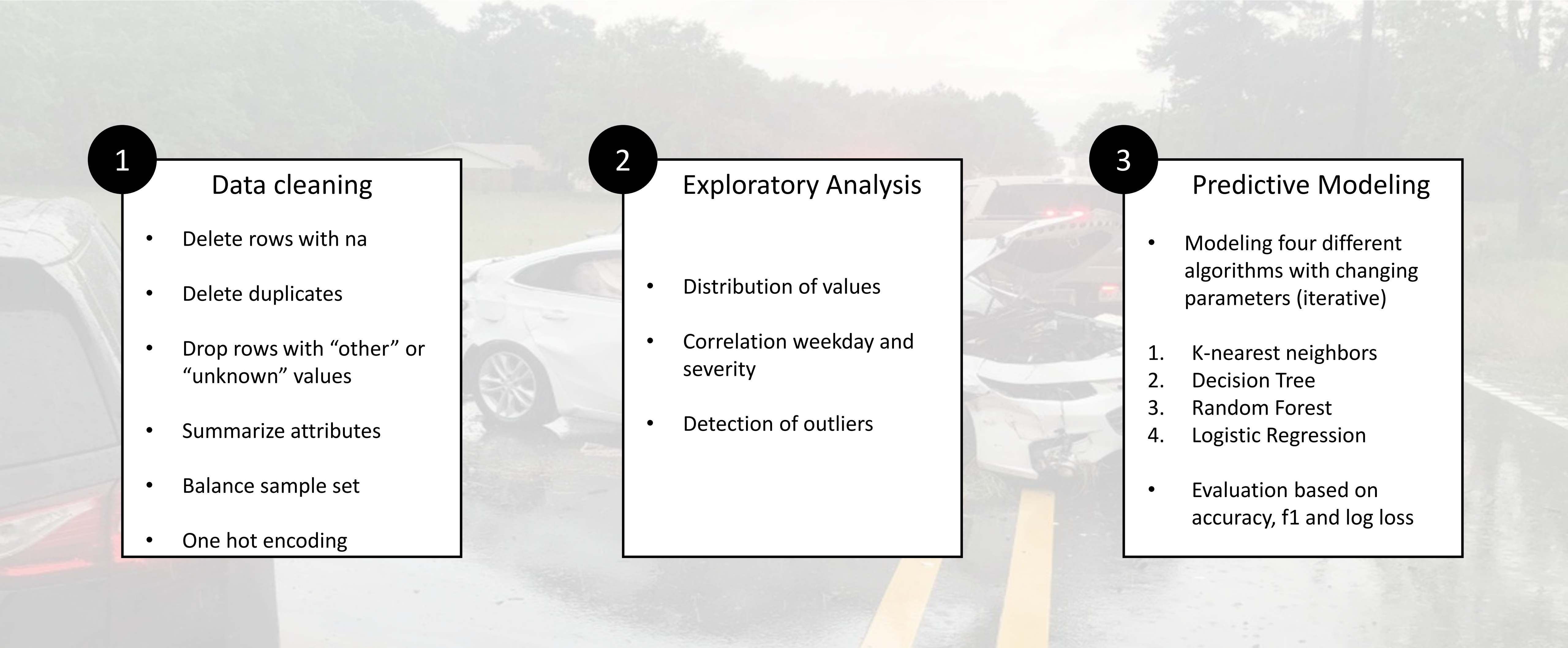
Feature selection

- External/environmental features only
- Must be possible to visually/directly determine feature by anyone
- 5 features selected for modeling:
  - Person count
  - Vehicle count
  - Junction type
  - Weather
  - Road conditions
  - Light conditions

First five rows of resulting data frame

	SEVERITYCODE	PERSONCOUNT	VEHCOUNT	JUNCTIONTYPE	WEATHER	ROADCOND	LIGHTCOND
0	2	2	2	At Intersection (intersection related)	Overcast	Bad_Conditions	Daylight
1	1	2	2	Mid-Block (not related to intersection)	precipitation	Bad_Conditions	Dark - Street Lights On
2	1	4	3	Mid-Block (not related to intersection)	Overcast	Good_Conditions	Daylight
3	1	3	3	Mid-Block (not related to intersection)	Clear	Good_Conditions	Daylight
4	2	2	2	At Intersection (intersection related)	precipitation	Bad_Conditions	Daylight

Methodology



1 Data cleaning

- Delete rows with na
- Delete duplicates
- Drop rows with “other” or “unknown” values
- Summarize attributes
- Balance sample set
- One hot encoding

2 Exploratory Analysis

- Distribution of values
- Correlation weekday and severity
- Detection of outliers

3 Predictive Modeling

- Modeling four different algorithms with changing parameters (iterative)

1. K-nearest neighbors
2. Decision Tree
3. Random Forest
4. Logistic Regression

- Evaluation based on accuracy, f1 and log loss

Results

Knn

	k	Accuracy_Score	F1_Score
4	5.0	0.61225	0.611384
9	10.0	0.60975	0.606648
6	7.0	0.60875	0.605882
8	9.0	0.60500	0.604183
3	4.0	0.60350	0.594858

Decision Tree

	Max_Depth	Accuracy_Score	F1_Score
6	7.0	0.64850	0.647881
5	6.0	0.64750	0.646833
7	8.0	0.64425	0.644000
4	5.0	0.64375	0.643603
3	4.0	0.64275	0.641532

Highest accuracy with decision tree
with max_depth = 7

```
              precision    recall  f1-score   support

           0       0.66      0.61      0.63      1979
           1       0.64      0.69      0.66      2021

   micro avg       0.65      0.65      0.65      4000
   macro avg       0.65      0.65      0.65      4000
weighted avg       0.65      0.65      0.65      4000
```

Random Forest

	N_estimators	Accuracy_Score	F1_Score
54	55.0	0.64125	0.641044
4	5.0	0.64075	0.640637
97	98.0	0.64000	0.639960
30	31.0	0.64000	0.639937
66	67.0	0.63950	0.639420

Logistic Regression

	C	Accuracy_Score	F1_Score	Log_loss
2	0.001	0.62300	0.622618	0.656204
0	0.100	0.62125	0.620931	0.654979
1	0.010	0.61950	0.619010	0.654798

Discussion and Conclusion

Discussion

- **Max accuracy of 0.6485**
- **Further data might be needed**
- **Choice of features must be optimized**
- **Classifier is not biased towards a certain result**
- **Future research might include:**
    - Possibility to predict probability of accident
    - Include further severity grades

Conclusion

- **Machine learning classifier was developed in order to predict the severity of an accident**
- **Based on data set of accidents in Seattle (2004 to today)**
- **Best model was found by using Decision Tree with max_depth = 7**
- **Model can be of great help for rescue services**