Predicting the relationships between a car accident and weather conditions

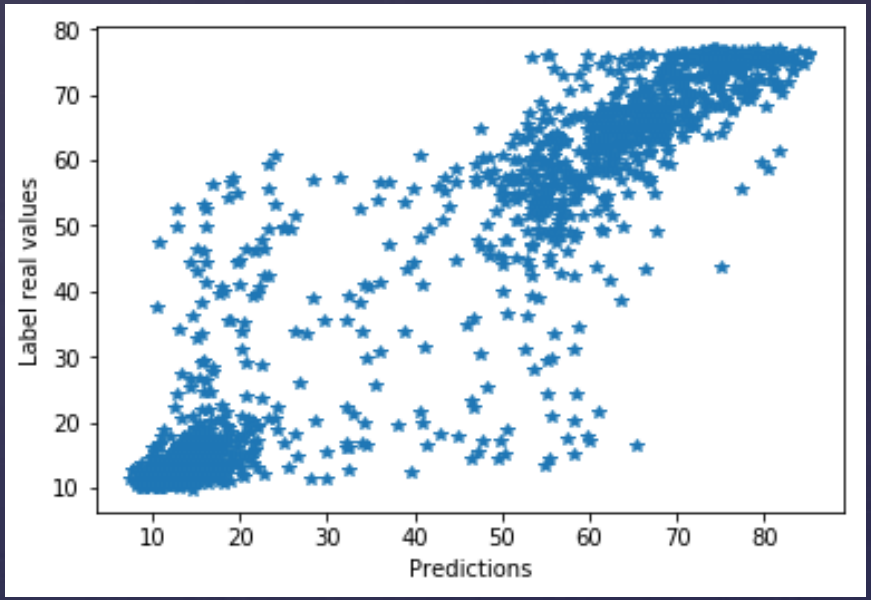
Predicting the relationships between a car accident and weather conditions

- Car drivers;
- Information services;
- Road department (improvement, correcting road signs)

Data acquisition and cleaning

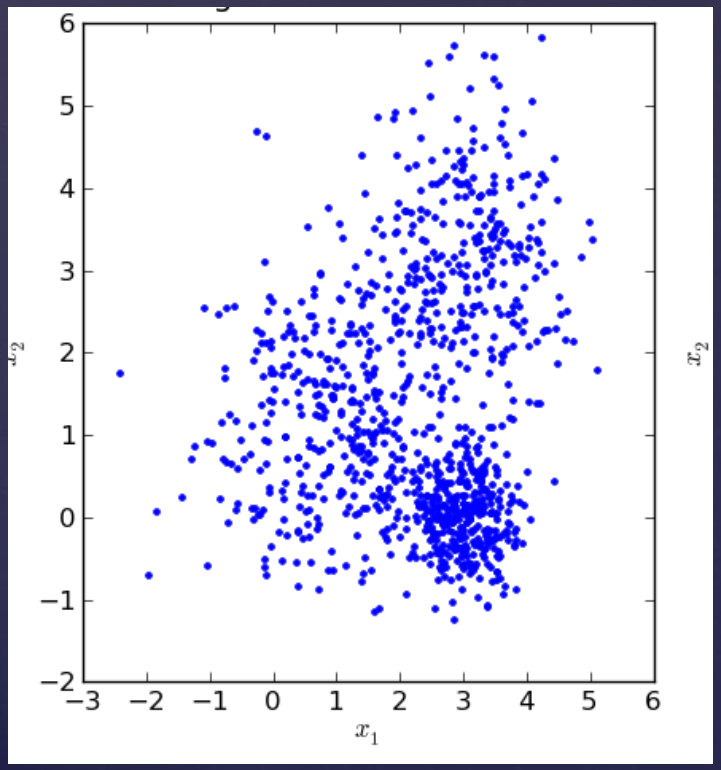
- ●Dataset (you can download here)
- In total, 1890 rows and 38 features in the raw dataset. ●
- Duplicate, highly similar or highly correlated features were dropped.
- Cleaned data contains 5 features.
- NaN values.

Regression models: dealing with unbalanced dataset



- Unweighted model prioritize the error of those players.
- Resulting in narrower predicted range.
- Assigning more weights to underrepresented players help with this problem.

Regression models performance



Log loss: 0.598-0.601

- Accuracy: ○ 0.532-0.576
- SVM performed best among single algorithms, but the differences were small.

Conclusion and future directions

- Built useful models to predict the relationships between a car accident and weather conditions.
- Accuracy of the models has room for improvement.