Hack the Crash

Presented at the HackZurich on the 18 to 20 September

Thales Solutions

Today's Presentation

1 Data processing

2 Final model

3 Specifications and recommendations

1 Data processing

- Merging, cleaning and transforming the data:
  - Merging the datasets
  - NaN values are replaced by special values
  - Categorical strings are encoded
  - All data is represented by space and compute efficient data structures
  - Scale the data by feature


```
Public  →  0x01
Private →  0x02
```

```
DataFrames
Tensors
```

1 Our feature representations

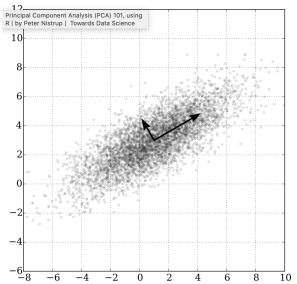
- Each data point is encoded into a feature vector

```
roof_type
foundation_type
age_building
has_secondary_use
```

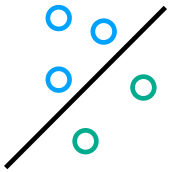
→

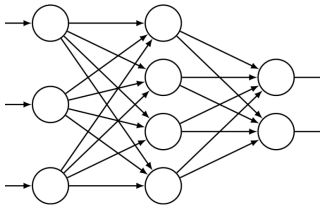
```
2
12
54
0
```

- The feature vector is reduced to a subset of features via principal component analysis (PCA)



- Can be used by all sorts of models, from SVM's, through Tree Classifiers, up to Neural Networks





2 Final model

- RandomForestClassifier
  - 250 estimators
  - Class weight balancing
  - Used with a subset of PCA features

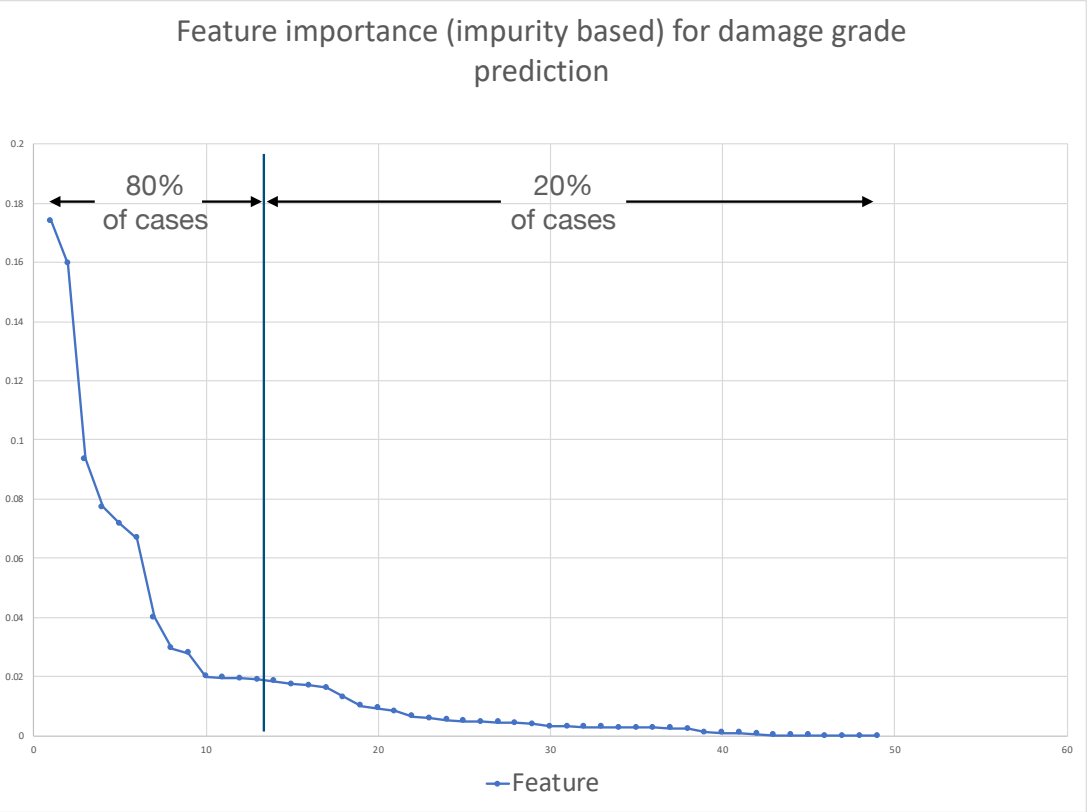
- Benefits
  - Intuitive model that allows for nice insights and interpretations
  - Competitive with other alternatives

3 Key problems and possible solutions

Problems		Solutions
Prediction of future damages on buildings	→	Model trained on 49 specific features
Find recommendations to reduce damage grades	→	Extract key features that predict (high) damage grade

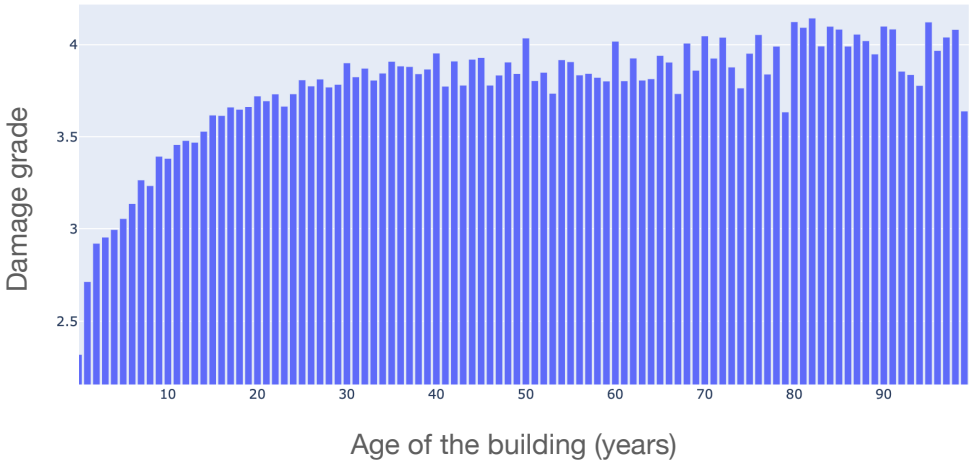
3 We got those results

- The Top 10 features in our list are indicators for 80% of the cases where damage was reported
- Building attributes like age, area (in square feet), height make around 34% of the feature importance for damage grade prediction
- The bottom 20% consist of 39 features
- 80/20 Rule



3 Recommendations

- The age of the building is the strongest indicator. The damage grade almost doubles over the 25 first years of a building.

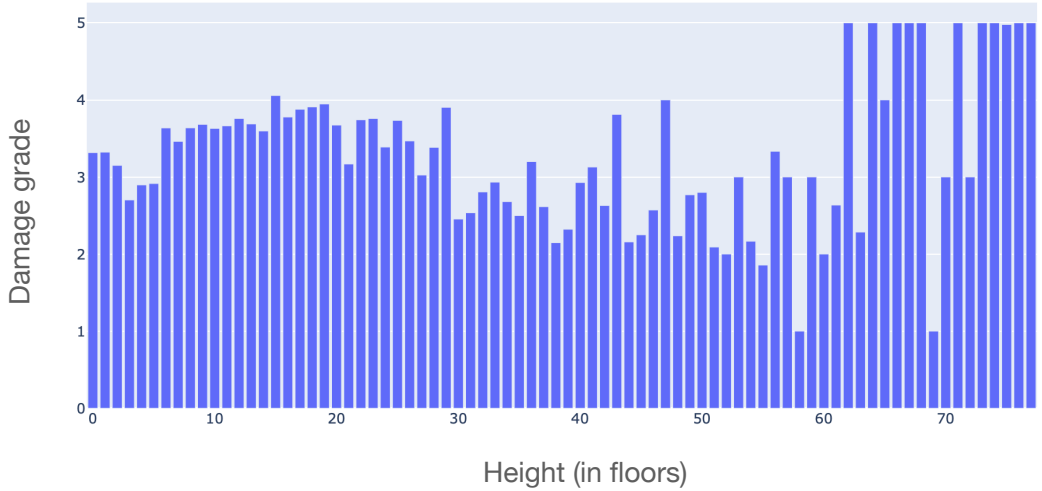


→ We recommend careful maintenance

3 Recommendations

- The height of the building is also among the top-5 strongest indicators. From the 60th floor on, the risk of buildings being highly damaged (grade 5) is extremely high.

→ Build in width not in height



Recap of Today's Presentation

1 Data processing

2 Final model

3 Specifications and recommendations