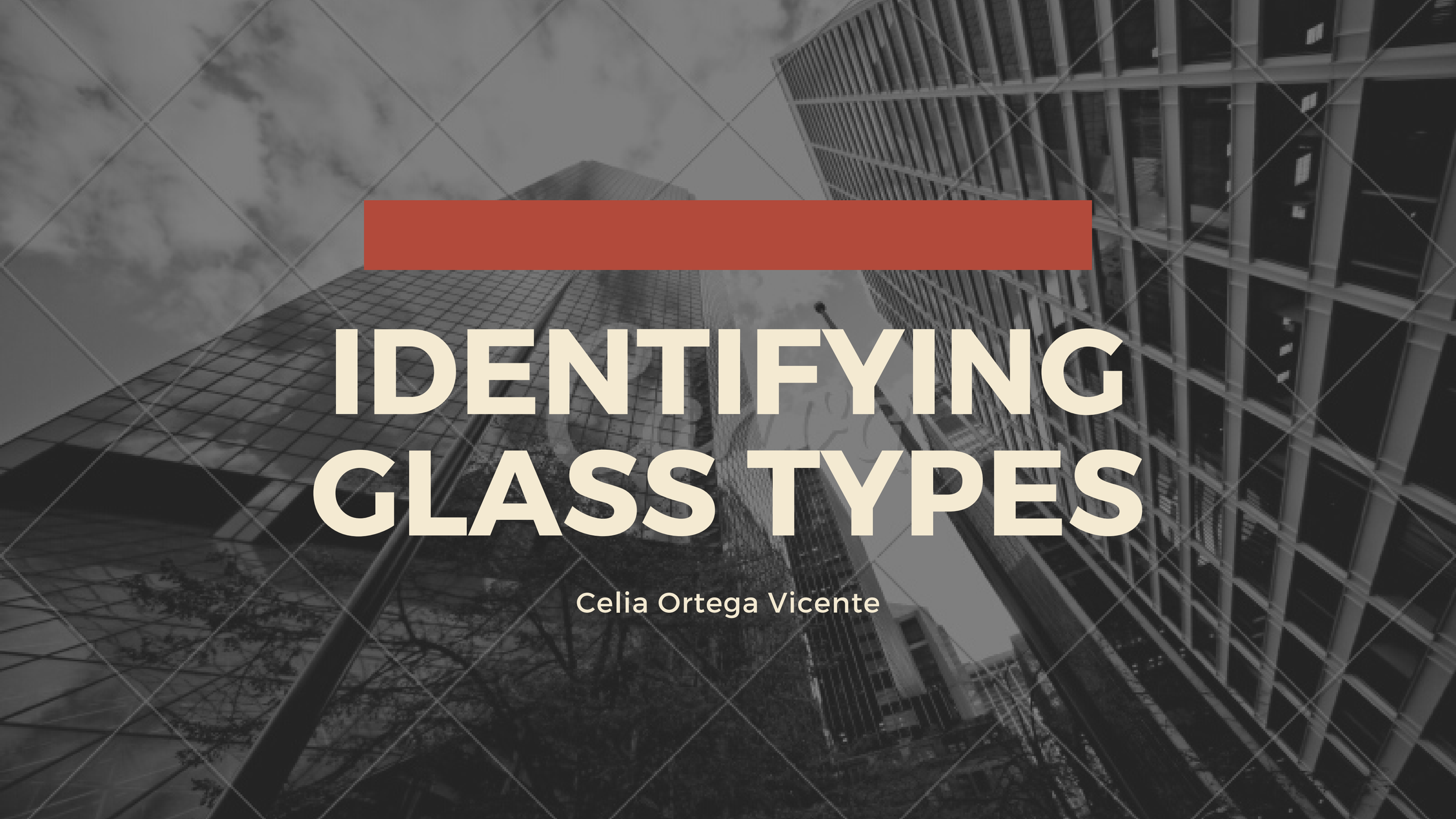


IDENTIFYING GLASS TYPES

Celia Ortega Vicente

WORKFLOW

DATA COLLECTION

We collected glass samples from UCI Machine Learning Repository

DATA PREPROCESSING

Clean the data to ensure its accuracy and ease of use.

MODEL TRAINING

Test three models and to see which one is the most effective.

What for?

- Understanding the composition of glass samples.
- Could be crucial for criminal investigations.
- Investigating new methods for glass production to reduce cost and increase efficiency.

FEATURES

- **Refractive Index**
- **Sodium (Na)**
- **Magnesium (Mg)**
- **Aluminum (Al)**
- **Silicon (Si)**
- **Potassium (K)**
- **Calcium (Ca)**
- **Barium (Ba)**
- **Iron (Fe)**

Glass Types

BUILDING WINDOWS - FLOAT PROCESSED

BUILDING WINDOWS - NON-FLOAT PROCESSED

VEHICLE WINDOWS - FLOAT PROCESSED

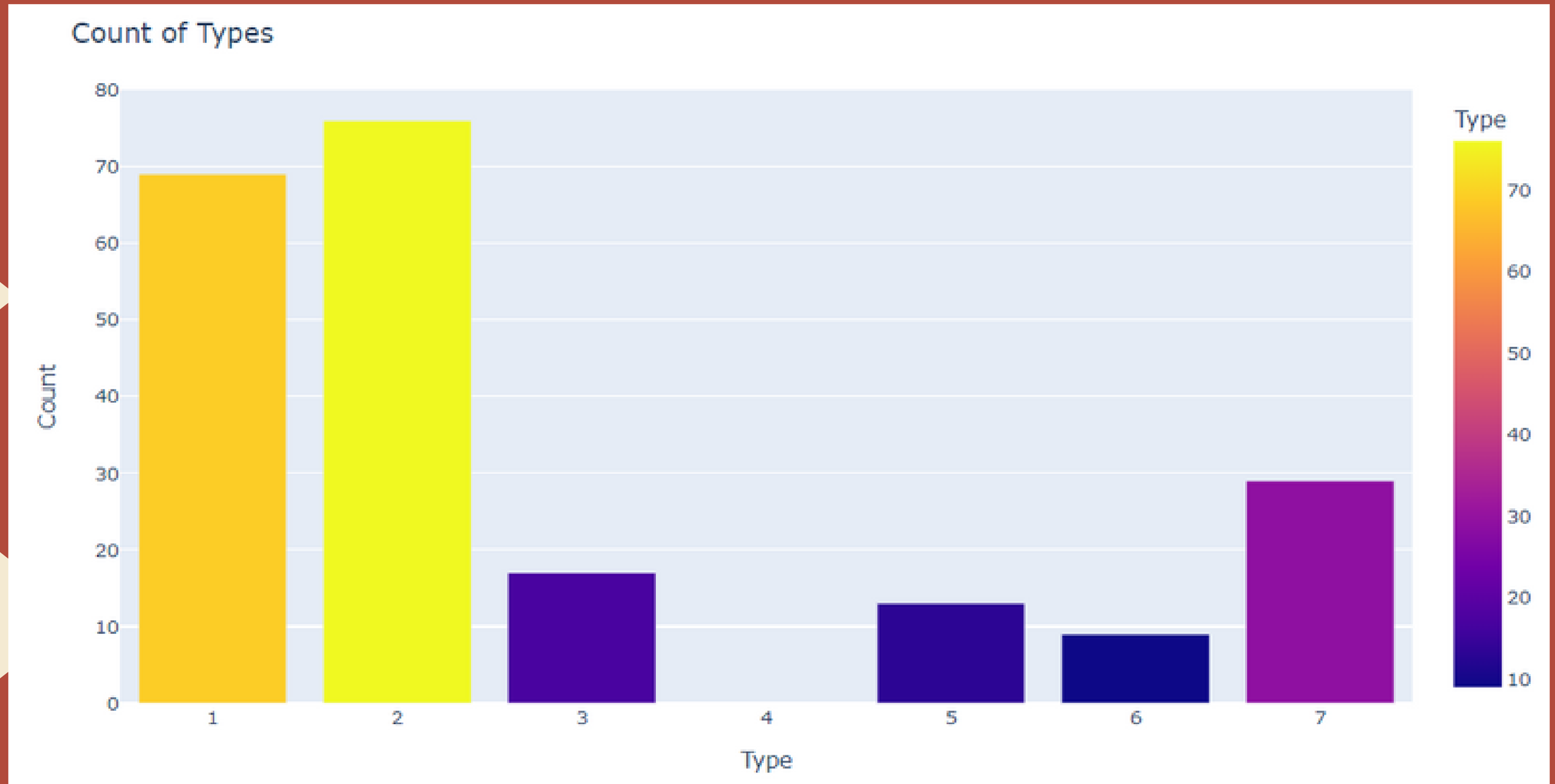
CONTAINERS

TABLEWARE

HEADLAMPS

DATA IMBALANCE PROBLEMS

- Missing one type of glass in data (type 4)
- Large difference in sample sizes between classes



MODELS

1 Random Forest

Kappa score

- With imbalanced data: **0.785**
- With balanced data: 0.756

2 Gradient Boosting Classifier

Kappa score

- With imbalanced data: 0.714
- With balanced data: 0.658

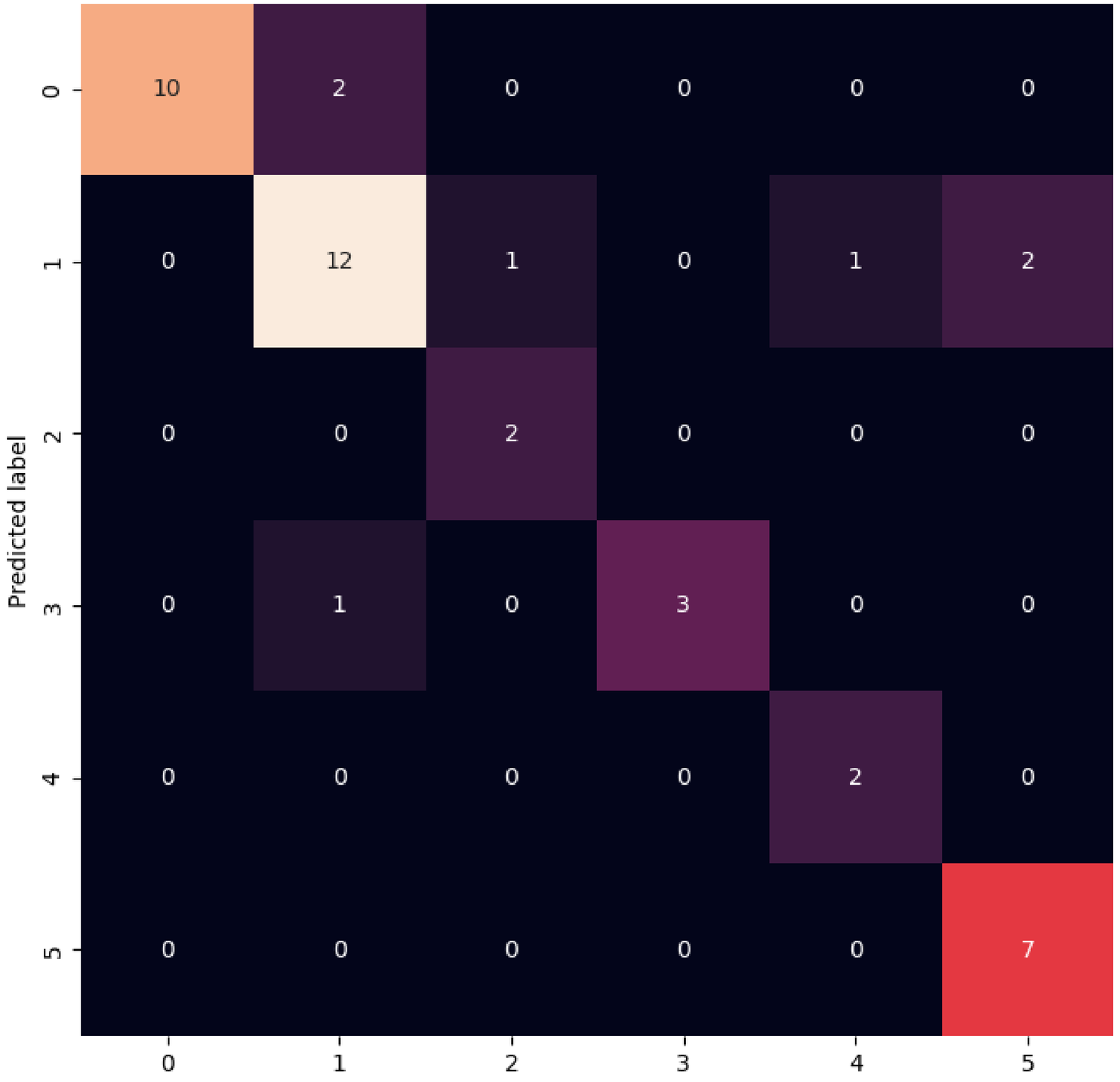
3 K-Nearest Neighbours

Kappa score

- With imbalanced data: 0.534
- With balanced data: 0.664

CONFUSSION MATRIX

ONLY THE 16% OF THE DATA WON'T BE GOOD PREDICTED



TRY IT YOURSELF