Data science and machine learning presentation 2

- Presented by: By Aly Elhadad(ID: 270067), Finnian Logan-Riley (ID: 269947) and Arran Logan-Riley (ID: 269945).

- Date: 20/06/2022
- University: Hochschule Furtwangen
- Course: Data Science and Machine Learning in Business

Question Definitions

- Q1 - Implement a prediction model for "Qualität"
  - Predict quality of products
- Q2 – Implement a prediction model for "Warpfaktor"
  - Predict the Warp Factor of the products
- Q3 - Implement a prediction model for "Fehler"
  - Predict mistakes/errors in products
- Q4 - Implement a prediction model for "Gammawert"
  - Gamma value
- Q5- Implement a prediction model for "LScore"
  - Predict LScore
- Q6 – implement a prediction model for "XKlasse"
  - Predict XKlasse

Take home messages

- It is possible to predict LScore with at 93%
  Quality at a 95% and WarpFaktor at 99%
- It is not possible to predict XKlasse
  - Opportunity to improve business
- It is possible to predict Fehler at 88%
- It is possible to predict Gammawert
- Opportunity to get more data during the production process
- Implement other inputs from suppliers
- All results currently based on input values
  - Durchmesser, Höhe und Gewicht

- **Full code can be found at https://github.com/FinnianHBLR/data-science-prediction**

Reasoning for analysis

- The goal is to reduce waste and improve the production quality
- Prediction of product grading to save time
- Less reliance for manual grading
- We have chosen 6 product properties to achieve this
- Most importantly, to optimise the SmartBuild production process

Data applied

- Data from SmartBuild production line
- The properties Diameter (Durchmesser), Height (Höhe) and Weight (Gewicht) are measured at the
start of the production line.
- The quality, mistakes and other output values are measured at the end of the production line
- Number of data entries: 10,000
- Data split 0.3/0.7 - 0.2/0.8
- Data age is unknown
- No data quality issues were found

Approach to findings

- Started with downloading and reviewing the SmartBuild data
- Checked the data quality
- Agreed on which tasks should be done by who
- We selected the appropriate models to be used
  - Only common data science methods such as DescionTrees, XGBoost were applied.
- Programmed the first models
- Full Git branching organising with code reviewers assigned
- We reviewed code through and made edits
- Evaluation is to be carried out
- Specific focus on evaluation rather than the model itself
- If a solution cannot be found, verification will be carried out
- Conclusions can then be finally drawn bringing all predictions together

Overview

- Q1 Quality
  - XGBRegressor
  - Accuracy
  - MAE
  - MSE
- Q2 Warpfaktor
  - Linear Regression
  - Accuracy
  - MAE
  - MSE
- Q3 Fehler
  - XGBClassifier (eXtreme Gradient Boosting)
  - ROC Curve evaluation
- Q4 Gammawert
  - XGBRegressor
  - MAE
- Q5 LScore
  - DecisionTreeClassifier
  - Confusion Matrix
- Q6 XKlasse
  - RandomForestClassifer
  - Logistic Regression
  - Class prediction error
  - Confusion Matrix

Q1 - Quality

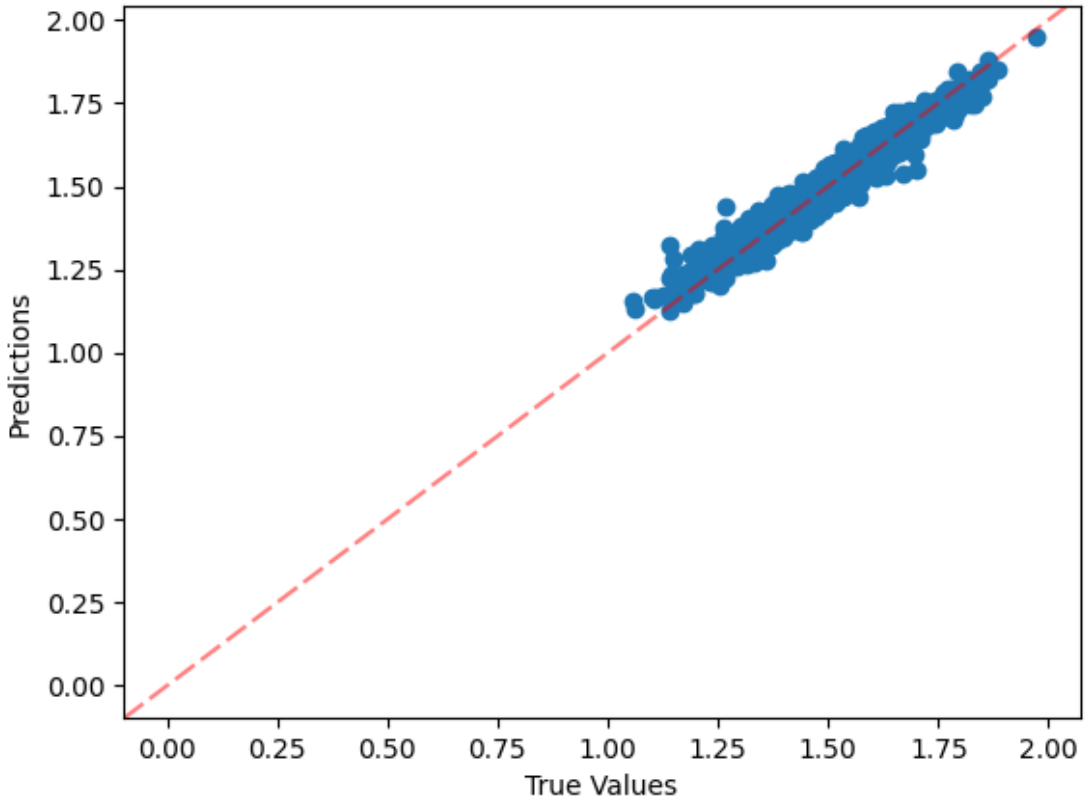
Model details:

1. XGBRegressor, Train size = 0.8
2. MAE = 0.022
   MSE = 0.00083
3. Data inputs -> 10,000 entries of Durchmesser, Hoehe and Gewicht

- Observations:
- The model can predict the Quality based on the

  Three inputs with up to 0.953 accuracy

  Therefore, The Quality can be predicted with a good accuracy.



Q2 -Warpfaktor

Model details:

1. Linear regression, Train size = 0.8
2. MAE = 1.65

   MSE = 4.23
3. Data inputs -> 10,000 entries of Durchmesser, Hoehe and Gewicht

- Observations:
- The model can predict the Warpfaktor based on the

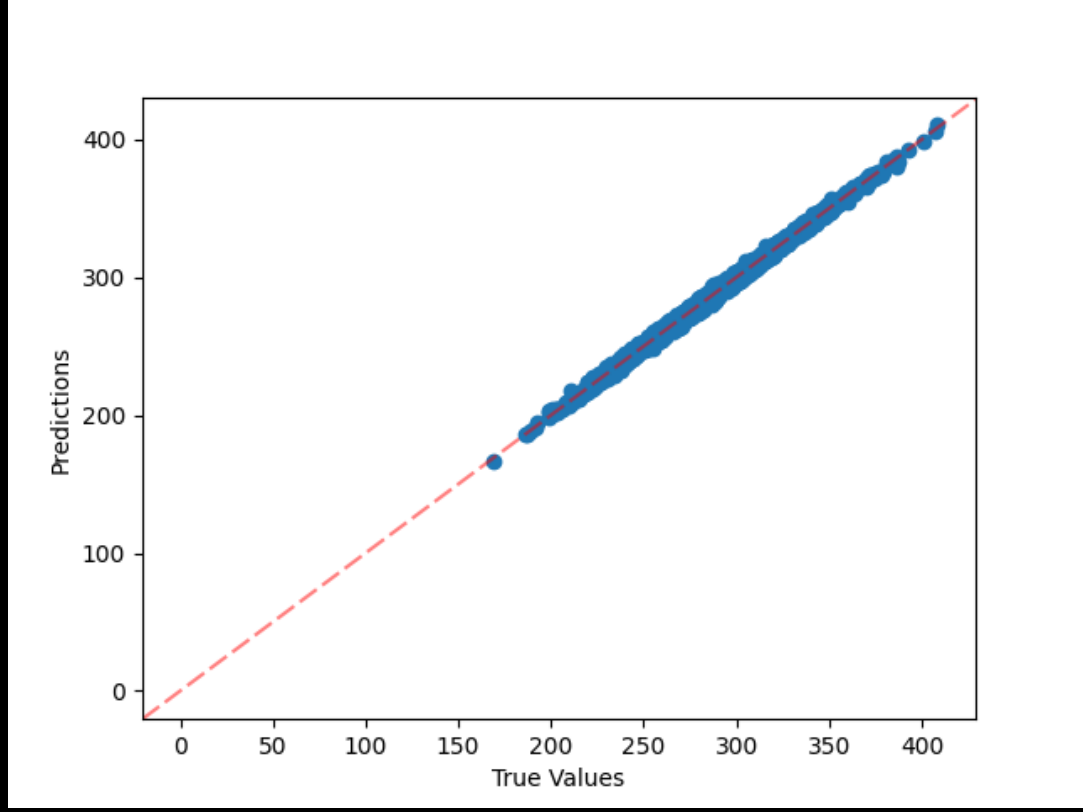
  Three inputs with up to 0.997 accuracy

  Therefore, The Warp faktor can be predicted with a good accuracy.
  The Coefficients for different Variables are:

  Durchmesser 1.1

  Hoehe 1.05

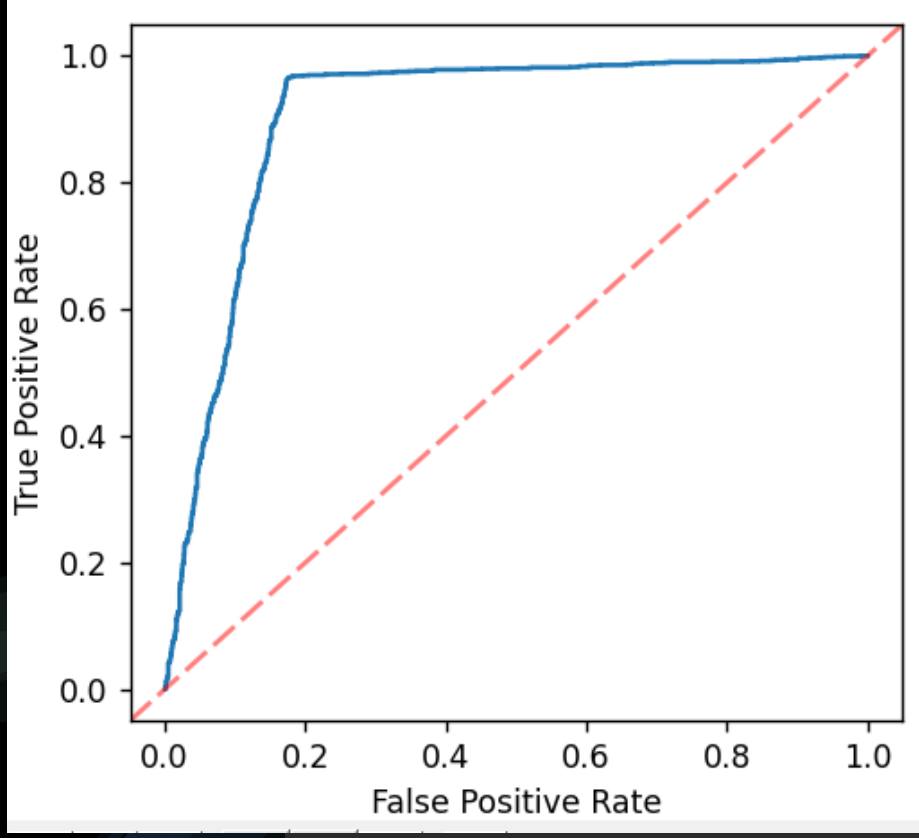
  Gewicht 1.0



Q3 Fehler

Model Details:

- Prediction model used - XGBClassifier
- Data inputs – 10,000 entries of Durchmesser Hoehe Gewicht
- Good overall performance
- **1434** entries were successfully predicted as nein and **1233** entries were successfully predicated as Ausschuss
- 2,667 out of the 3,000 entries had a correct prediction. This means it has an 88% accuracy of predicting the correct values

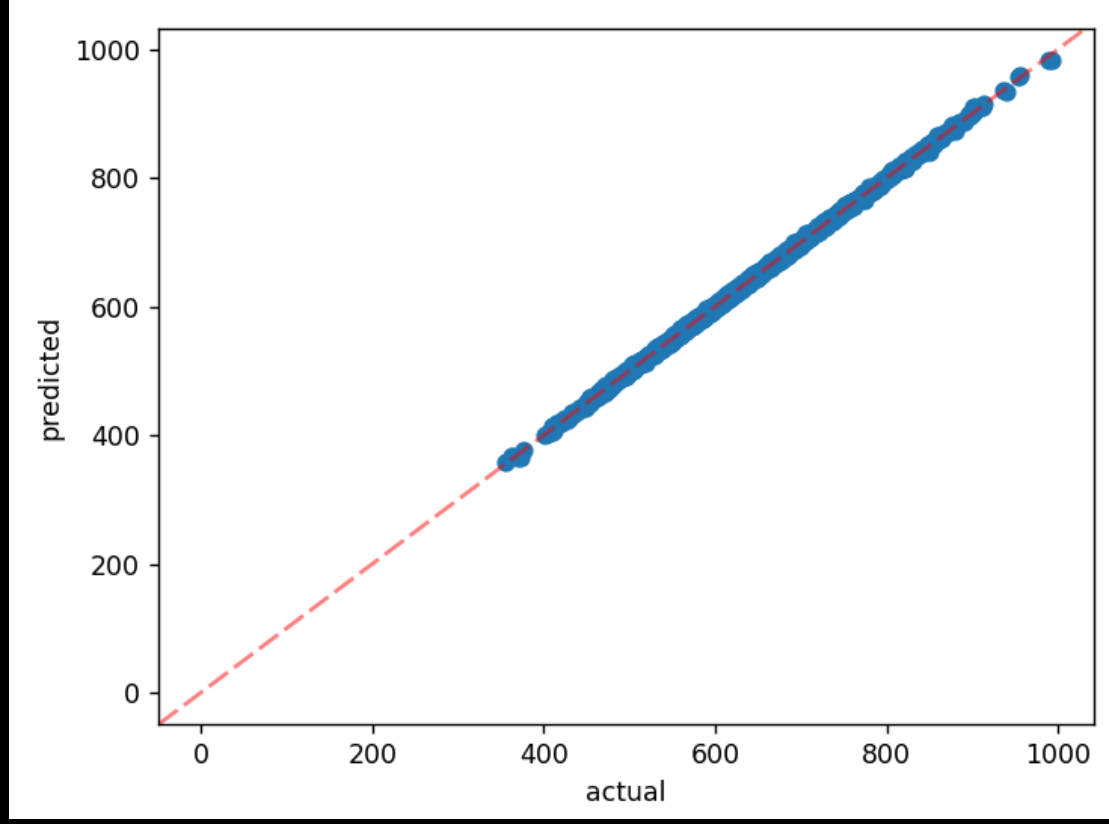



```
[[1434   78]
 [ 255 1233]]
```

Q4 Gammawert

Model Details:

- Prediction model used - XGBRegressor to predict the Gammawert value
- Data inputs – 10,000 entries of Durchmesser Hoehe Gewicht
- Good overall performance
- MAE of 1.79



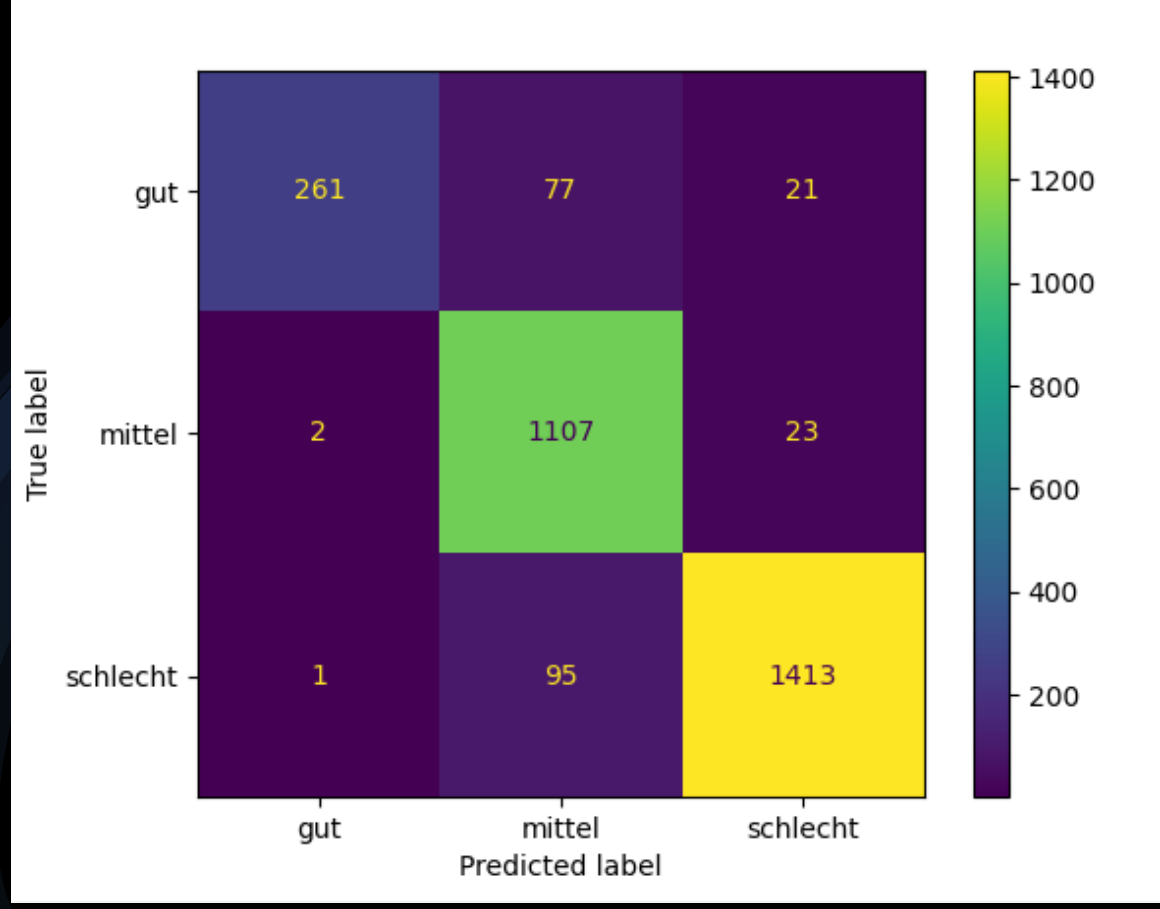
Q5 LScore

Model details:

- Decision Tree Classifier
- Depth 4
- Confusion Matrix
- 2,781 correct, 219 wrong predictions

- Observations
  - Accuracy – correct classifications/total
    - 0.93
  - Precision – Ability to identify positive classification
    - Total: 0.93
      - Gut - 0.99
      - Mittel 0.87
      - Schlecht 0.97
  - Further data could improve the results

- From the model applied it is possible to predict the LScore of a product with the input data given.



Q6 XKlasse

Details of models:

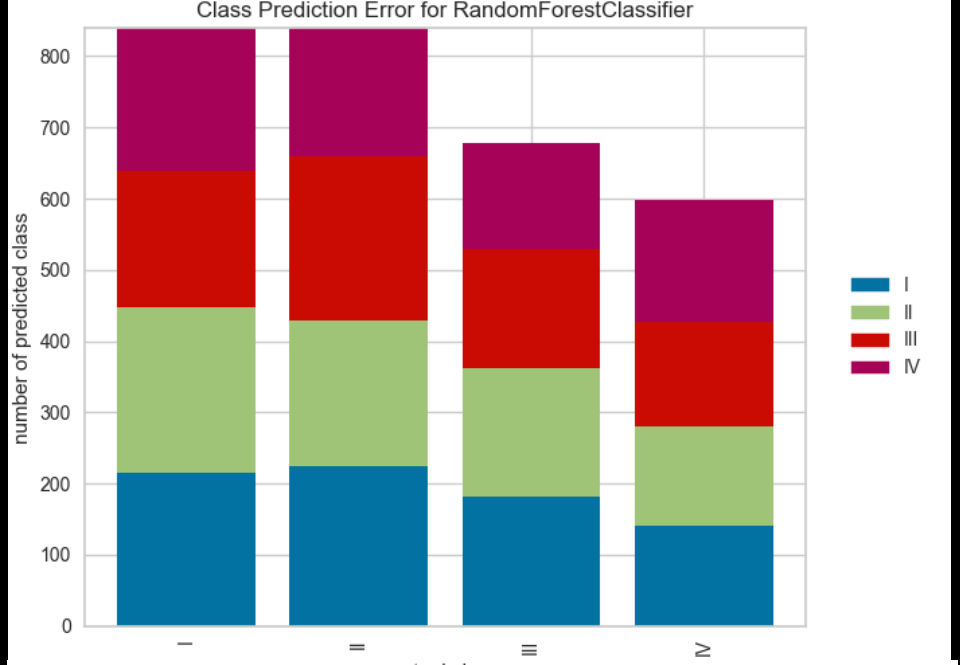
Two models with two different types of evaluation

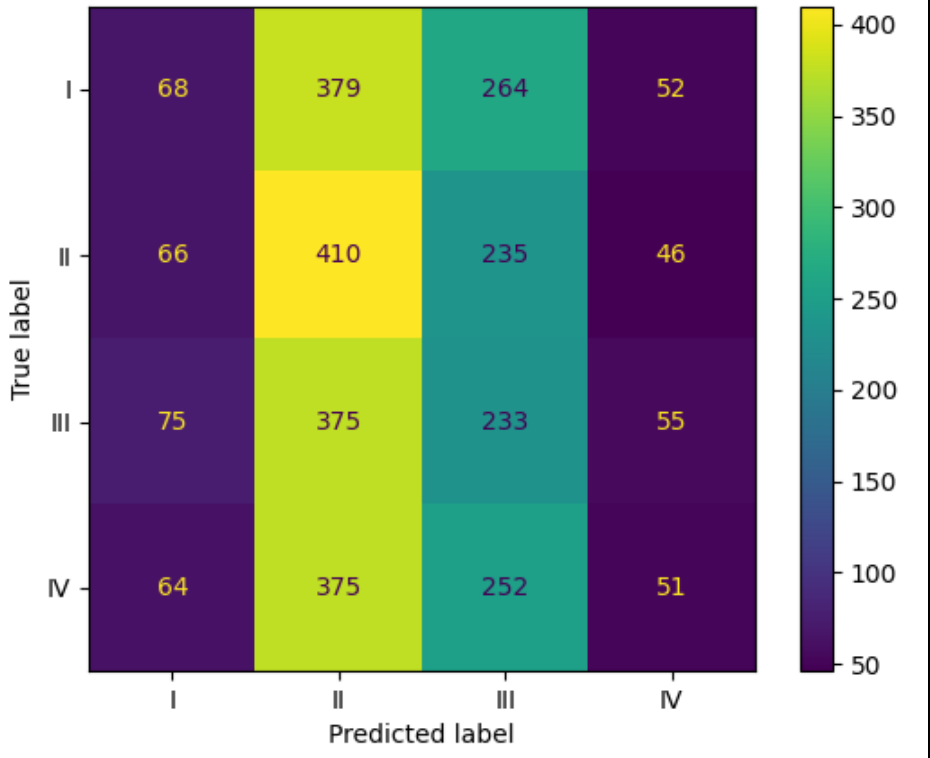
RandomForestClassifer

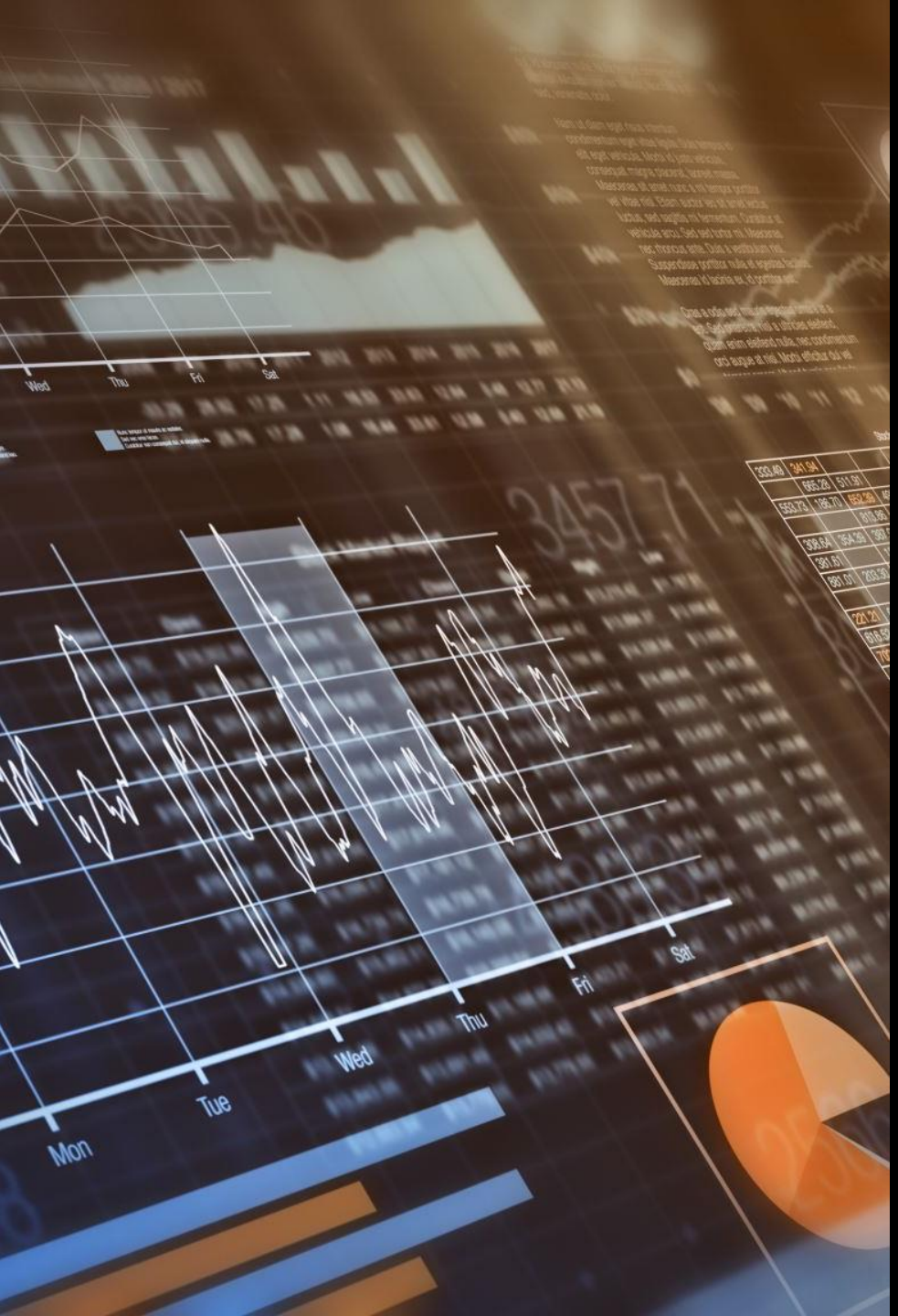
- Random distribution of predictions
- Class prediction error evaluation

Further investigation with Logistic Regression

- Confusion matrix
- Observations
  - Precision
    - I 0.25
    - II 0.27
    - III 0.24
    - IV 0.25
  - This could be for multiple reasons, but to the best of our abilities this value is not possible to predict. This may need further insight.
- **Currently** it is not possible to effectively predict XKlasse.







Reflection - Implications for SmartBuild

- There could be missing input data that could be added
  - Automatic measurements during production, paint time, worker, process time, machine metrics.
  - External inputs such as material suppliers or paint suppliers
- Another source is based on the outcome of other values could be predicted (Not just inputs)
  - Especially with automatic Fehler and Qualiteat grading this can work
- Effective predictions to be implemented
  - LScore
  - Gammawert
  - Fehler
  - Qualiteat
  - Warpfaktor
- However, for legal/critical grading
  - Must be cautions
  - Need to reach human level
  - Implement a review schedule for regular check-ups on systems or in the case of new machinery
- Further data investigation/opportunities for improvement
  - XKlasse
  - What is needed for correct prediction e.g., new sensor data
- Questions