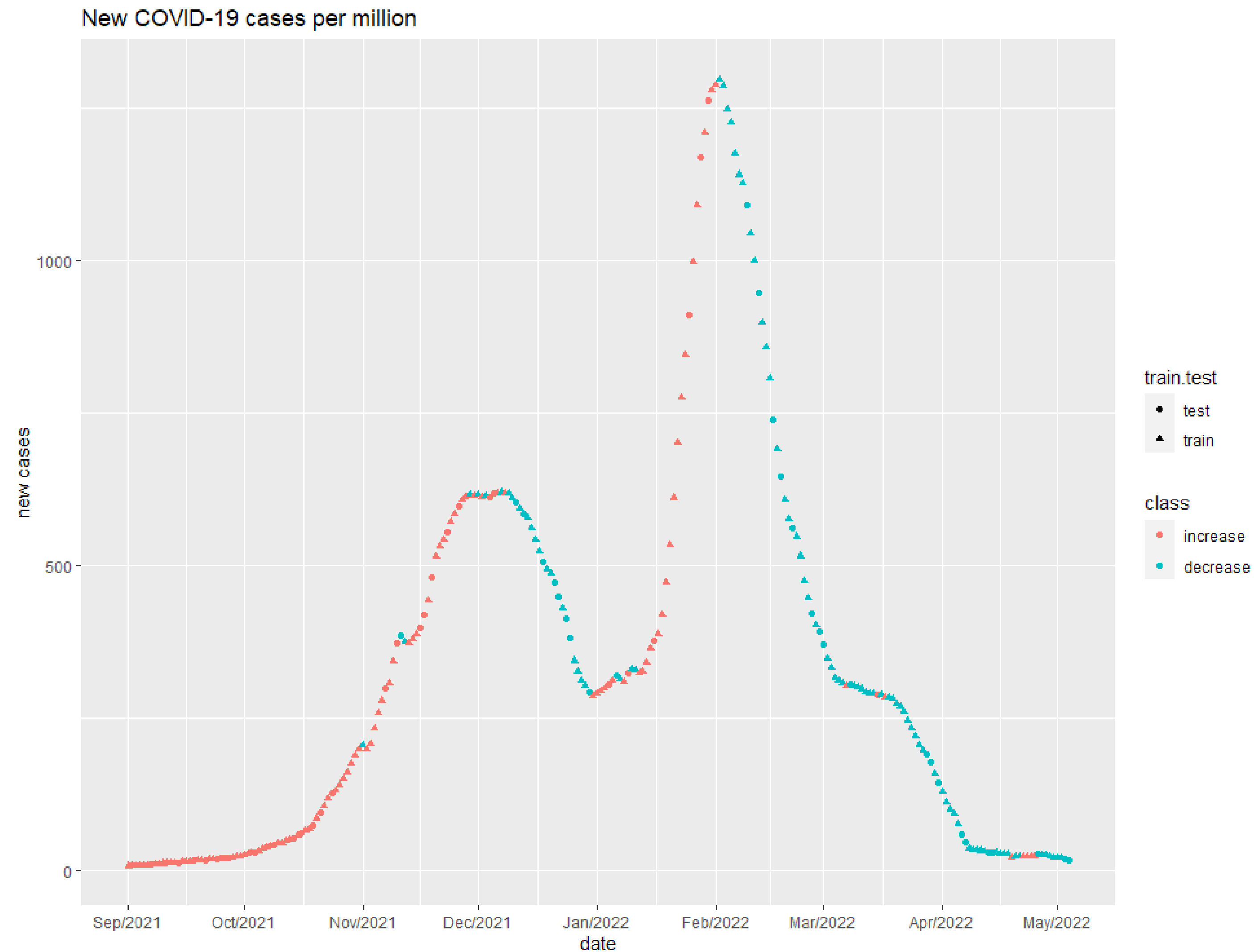# Warsaw University of Technology

COVID-19 new cases classifier
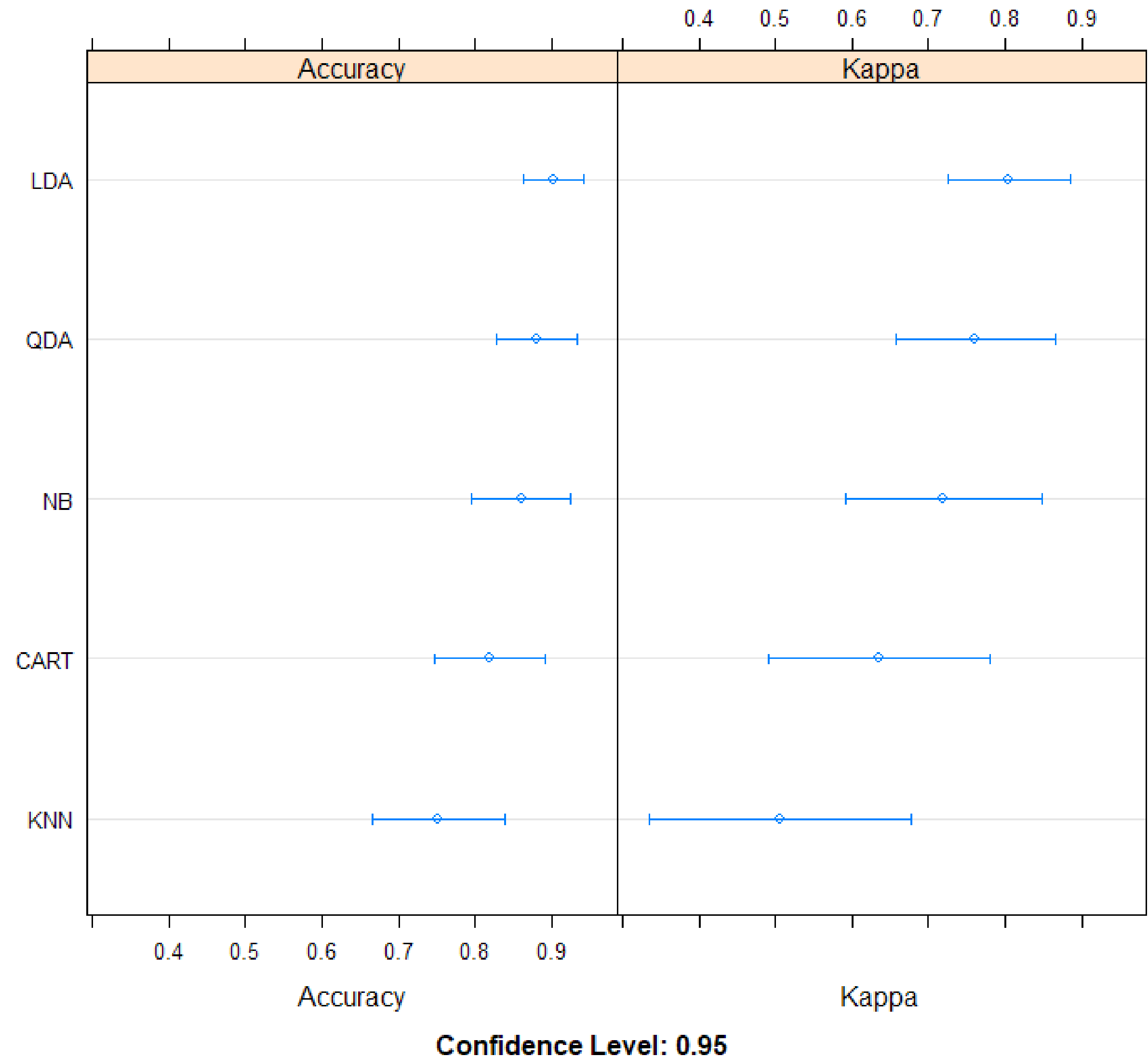
Bogumił Wierzchowski

Data

- Data gathered by the **Center for Systems Science and Engineering** (CSSE) at **Johns Hopkins University**
- From **September 2021** to **May 2022**
- 75%/25% **train/test** split
- All data **relative to population**
- Smoothed by a **moving average** with a length of 7

New COVID-19 cases per million

Warsaw University of Technology
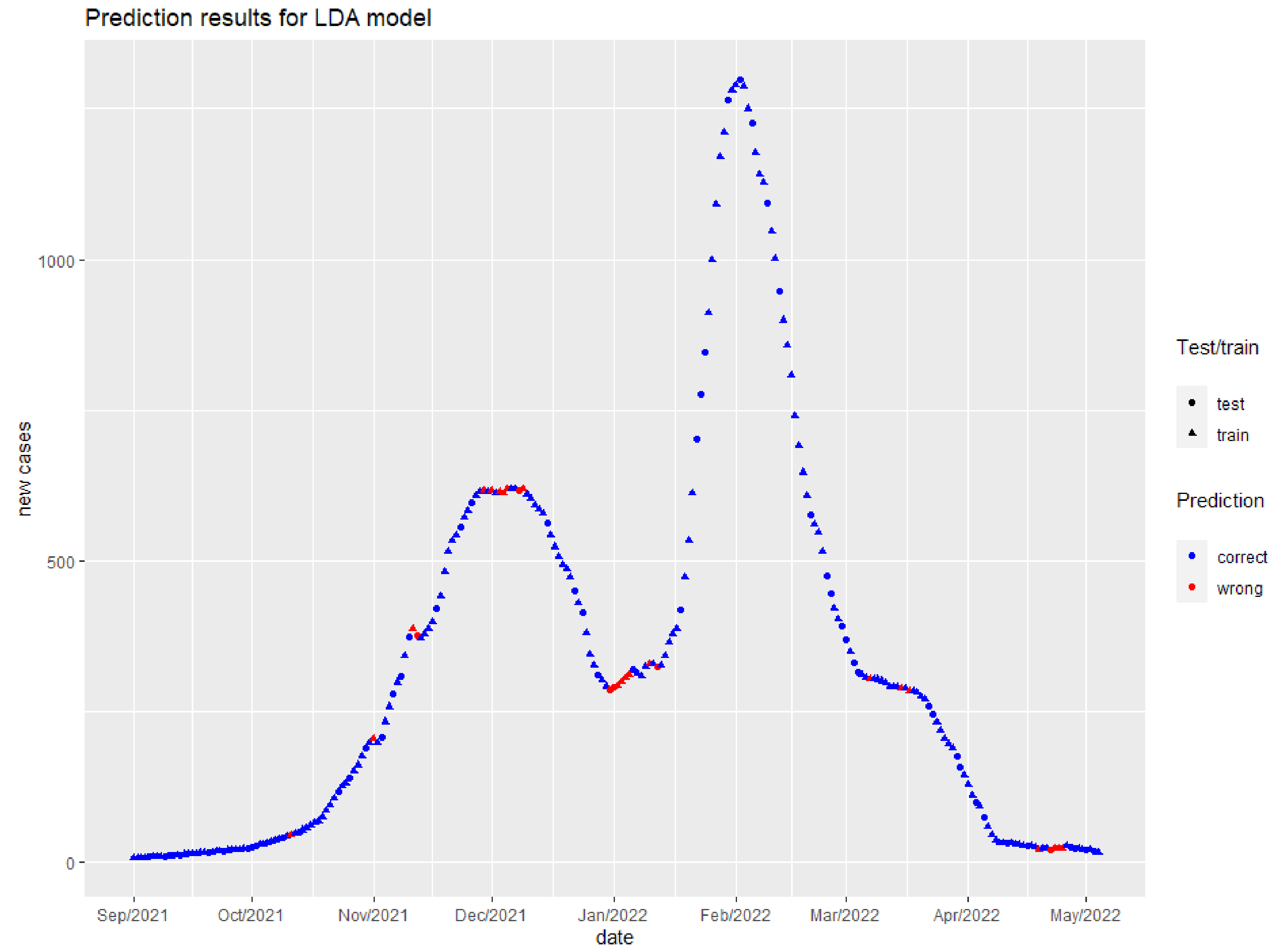
Classifiers validation

- 10-fold **cross validation** on train set
- KNN validated on **grid** from 2 to 20 NN
- Factors used for classification: **new deaths**, **new vaccinations**, **new tests**, **tests per case** + daily increases for all of them

Warsaw University of Technology

LDA model results

- **Train** accuracy: 0,886
- **Test** accuracy: 0,902
- **Total** accuracy: 0,890

Prediction results for LDA model

Warsaw University
of Technology

Sources

1. Data on COVID-19 (coronavirus) by Our World in Data (https://github.com/owid/covid-19-data/tree/master/public/data)
2. Used R libraries: Hmisc, ggplot2, caret

Warsaw University
of Technology