

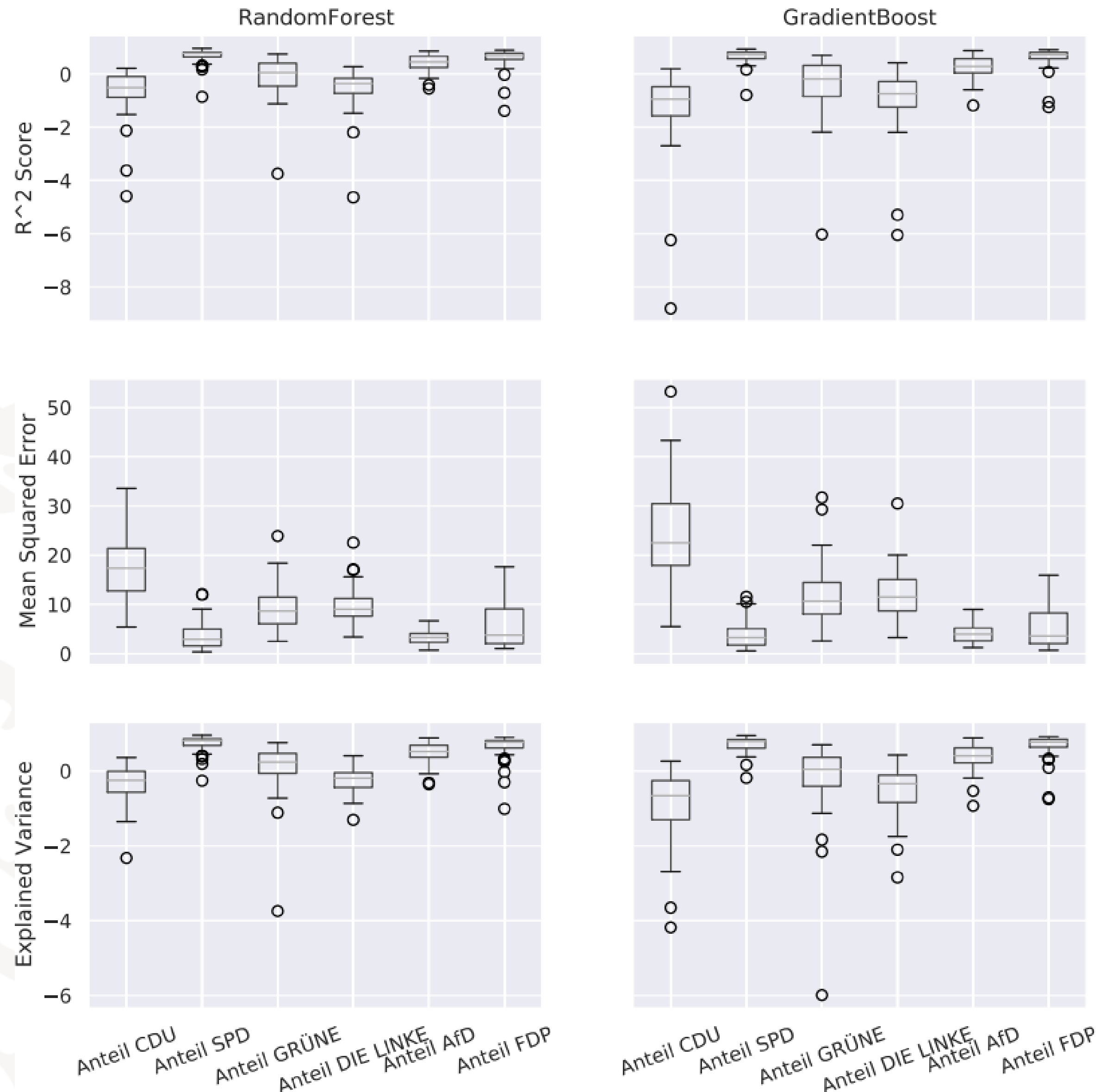


Idee

- Vorhersagen der Wahlergebnisse auf Basis der Daten zur Einkommensverteilung
- Bzw. Erkennen von Zusammenhängen zwischen Wählerschaft und Einkommen
- Input: mehrere kontinuierliche Daten zur Einkommensverteilung
- Output: kontinuierliche Daten zum Anteil der Parteien am Wahlergebnis
- → Multi Target Regressionsproblem

- Untersuchte Algorithmen:
 - SKLearn GradientBoostingRegressor
 - SKLearn RandomForestRegression
- SKLearn MultiOutputRegressor
- Wiederholtes Training für unterschiedlich Zwecke mit Analyse des Ergebnis:
 - Einfaches Training mit 1 zu 4 Test- /Trainingsplit
 - Wiederholten Training auf gleichen Test- /Trainingsdaten
 - Cross-Validation
 - Repeated 5-Fold-Cross-Validation aufgeschlüsselt nach Parteien

Auszug aus den Ergebnissen



- Analyse der Modelle in Abhängigkeit zu gewählten Test- Und Trainingsdaten
- Boxplot verschiedener Metriken zur Vorhersage der beiden Algorithmen Random Forest und Gradient Boost
- Ausgabe beschränkt auf CDU, SPD, Grüne, Linke, AfD und FDP
- Explained Variance, Mean Squared Error, R²
- 10 Wiederholungen 5-Fold Cross Validation
- Datenpunkt bezieht auf eine Instanz eines Modells mit eigenen Trainings- und Testdaten