EuroSciPy
Szczecin 2024

Conformal Prediction with MAPIE: A Journey into Reliable Uncertainty Quantification

EuroSciPy, Szczecin

28th August, 2024

Claudio G. Giancaterino

TOWIN LAB
Towards Innovation Lab

Summary

- ❑Conformal Prediction overview
- ❑MAPIE introduction
- ❑Regression use case

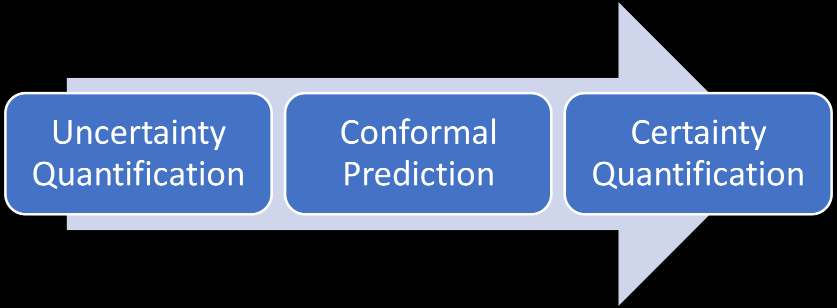
Introduction to Uncertainty Quantification

Uncertainty quantification is a crucial aspect of predictive modeling. Traditional predictions typically yield a single-point output for a given input, relying on models that provide the most likely outcome. These models are unable to capture the full complexity of the real world or are based on incomplete data.

Uncertainty can be broken down into:
-aleatoric uncertainty;
-epistemic uncertainty.

https://deepai.org/machine-learning-model/text2img

Why Conformal Prediction ?

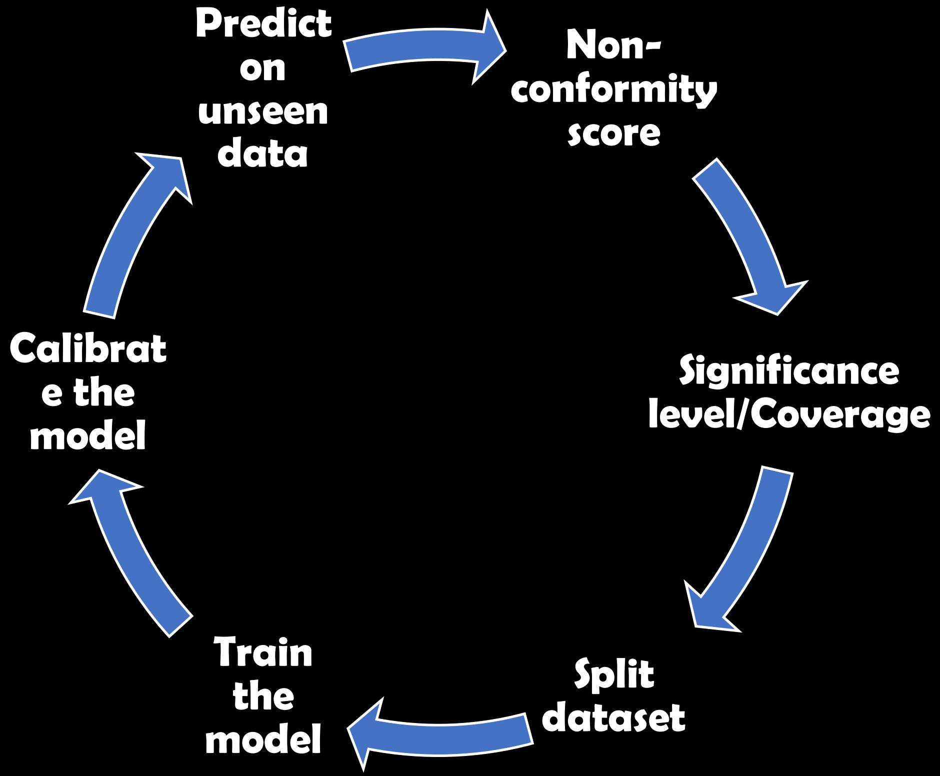


Conformal prediction is a powerful machine learning framework used to evaluate the uncertainty of predictions. It turns point predictions into prediction regions, it provides prediction sets for classification task or prediction intervals for regression task.

Some advantages of using conformal prediction for uncertainty quantification:
-distribution-free;
-guaranteed coverage;
-model-agnostic
-easy implementation

EuroSciPy
Szczecin 2024

How does Conformal Prediction work in a nutshell ?



Towards Innovation Lab

Type of Conformal Predictors

-Transductive Conformal Predictors (TCP): TCP leverages the entire dataset for training and requires model retraining for each new prediction. This approach ensures that each prediction is made with the most up-to-date information.

-Inductive Conformal Predictors (ICP): ICP splits the data into a training set and a calibration set. The model is trained once on the training set, and the calibration set is used to adjust the prediction intervals.

MAPIE (Model Agnostic Prediction Interval Estimator)

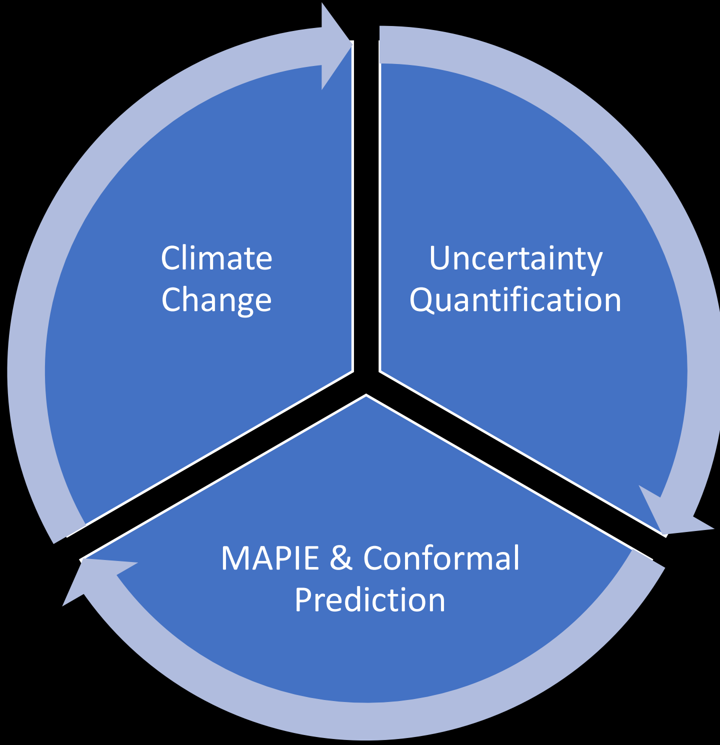
MAPIE is based on two types of predictor families:

-Split Conformal Prediction

-Cross-Conformal Prediction

EuroSciPy Szczecin 2024

Use Case



TOWIN LAB
Towards Innovation Lab

EuroSciPy Szczecin 2024

Data Collection

[Co2emissions byvehicles]

Data are retrieved from Kaggle about CO2 emissions released by Canadian government's open data website. There are 7385 rows and 12 columns. The outcome CO2 emissions are for combined city and highway driving.

Pixabay: CO2 emissions by vehicles

LAB
Towards Innovation Lab

Conformalized Quantile Regression (CQR)

Conformalized quantile regression extends traditional quantile regression by incorporating conformal prediction principles.

Python

```python
# fit MAPIE conformal quantile regressor using LightGBM estimator
np.random.seed(0)
LGBM_cqr = MapieQuantileRegressor(estimator=LGBM, cv="split", \
alpha=alpha, method= "quantile")
LGBM_cqr.fit(X_train, y_train, X_calib=X_cal, y_calib=y_cal, random_state=0)
# predictions and scores
LGBM_cqr_results, LGBM_cqr_predictions_df = \
calculate_predictions_and_scores(LGBM_cqr, X_test,"QRegressor", alpha)
# fit MAPIE conformal quantile regressor using QuantileRegressor estimator
np.random.seed(0)
QR_cqr = MapieQuantileRegressor(estimator=QR, cv="split", \
alpha=alpha, method= "quantile")
QR_cqr.fit(X_train, y_train, X_calib=X_cal, y_calib=y_cal, random_state=0)
# predictions and scores
QR_cqr_results, QR_cqr_predictions_df = \
calculate_predictions_and_scores(QR_cqr,X_test,"QRegressor", alpha)
```

Naive regression

The naive solution is the simplest method of conformal prediction.

Python

```python
# fit MAPIE naive regressor using LightGBM estimator
np.random.seed(0)
LGBM_naive = MapieRegressor(estimator=LGBM, method= "naive")
LGBM_naive.fit(X_train, y_train)
# predictions and scores
LGBM_naive_results, LGBM_naive_predictions_df = \
calculate_predictions_and_scores(LGBM_naive,X_test,"Regressor",alpha)
# fit MAPIE naive regressor using QuantileRegressor estimator
np.random.seed(0)
QR_naive = MapieRegressor(estimator=QR,method= "naive")
QR_naive.fit(X_train, y_train)
# predictions and scores
QR_naive_results, QR_naive_predictions_df = \
calculate_predictions_and_scores(QR_naive,X_test,"Regressor",alpha)
```

Jacknife regression

The Jacknife method is a resampling method and enhances the naive method by using a leave-one-out approach.

Python

```
# fit MAPIE jackknife regressor using LigthGBM estimator
np.random.seed(0)
LGBM_jacknife = MapieRegressor(estimator=LGBM, method= "base", cv=5)
LGBM_jacknife.fit(X_train, y_train)
# predictions and scores
LGBM_jacknife_results, LGBM_jacknife_predictions_df = \
calculate_predictions_and_scores(LGBM_jacknife,X_test,"Regressor",alpha)
# fit MAPIE jackknife regressor using QuantileRegressor estimator
np.random.seed(0)
QR_jacknife = MapieRegressor(estimator=QR,method= "base", cv=5)
QR_jacknife.fit(X_train, y_train)
# predictions and scores
QR_jacknife_results, QR_jacknife_predictions_df = \
calculate_predictions_and_scores(QR_jacknife,X_test,"Regressor",alpha)
```

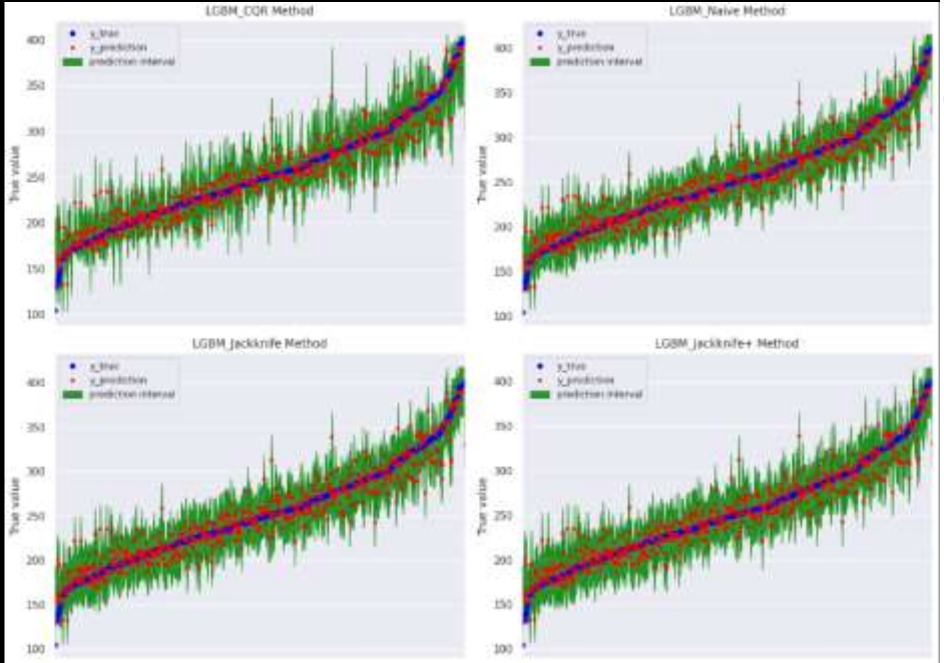
Jacknife+ regression

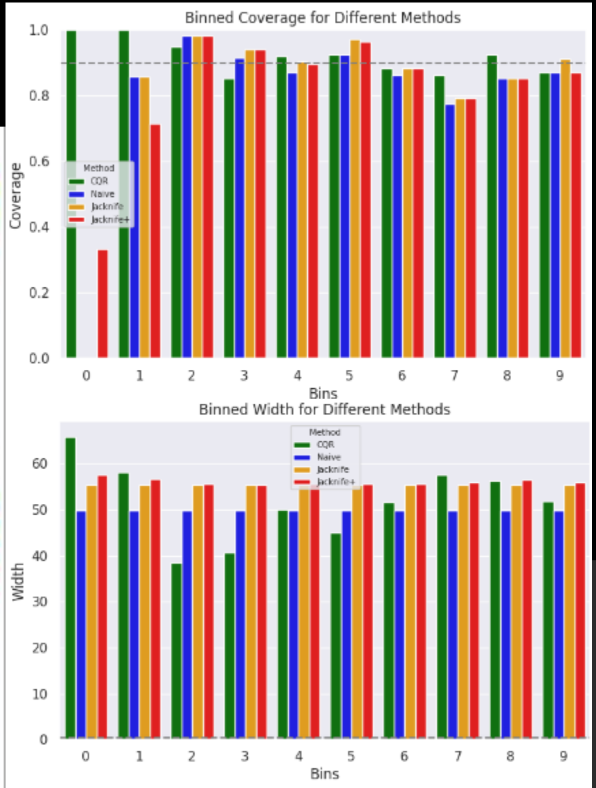
The Jackknife plus method follows the Jackknife method by incorporating additional information from the training data to refine the nonconformity scores further.

```python
# fit MAPIE jackknife+ regressor using LigthGBM estimator
np.random.seed(0)
LGBM_jacknife_plus = MapieRegressor(estimator=LGBM, method= "plus", cv=5)
LGBM_jacknife_plus.fit(X_train, y_train)
# predictions and scores
LGBM_jacknife_plus_results, LGBM_jacknife_plus_predictions_df = \
calculate_predictions_and_scores(LGBM_jacknife_plus,X_test,"Regressor",alpha)
# fit MAPIE jackknife+ regressor using QuantileRegressor estimator
np.random.seed(0)
QR_jacknife_plus = MapieRegressor(estimator=QR,method= "plus", cv=5)
QR_jacknife_plus.fit(X_train, y_train)
# predictions and scores
QR_jacknife_plus_results, QR_jacknife_plus_predictions_df = \
calculate_predictions_and_scores(QR_jacknife_plus,X_test,"Regressor",alpha)
```

LightGBM Results

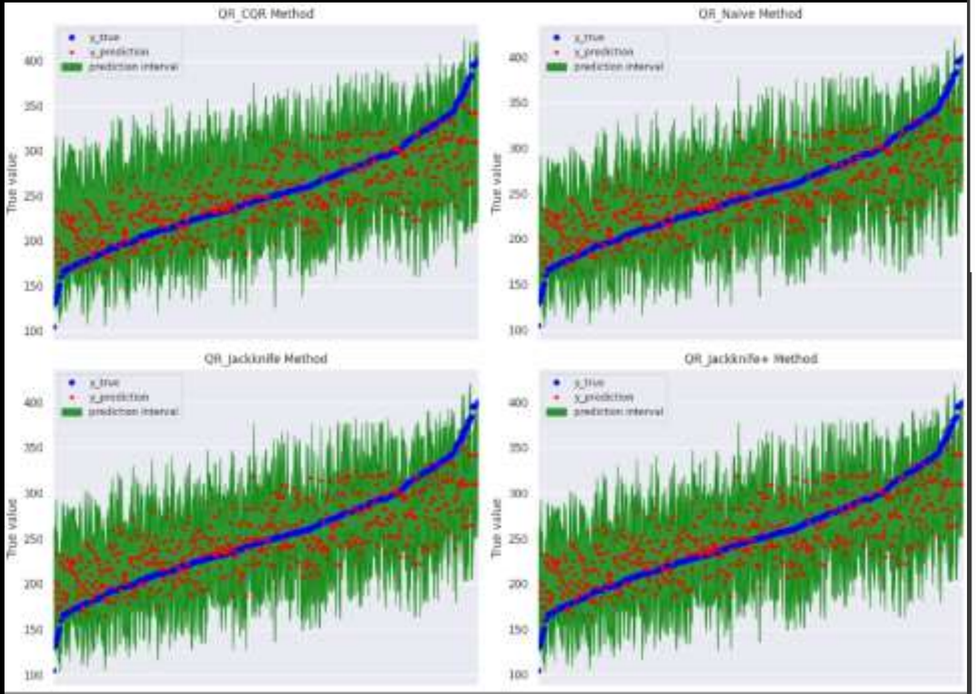
The conformalized quantile regression can satisfy all requirements for prediction intervals: efficient, adaptive, and valid.

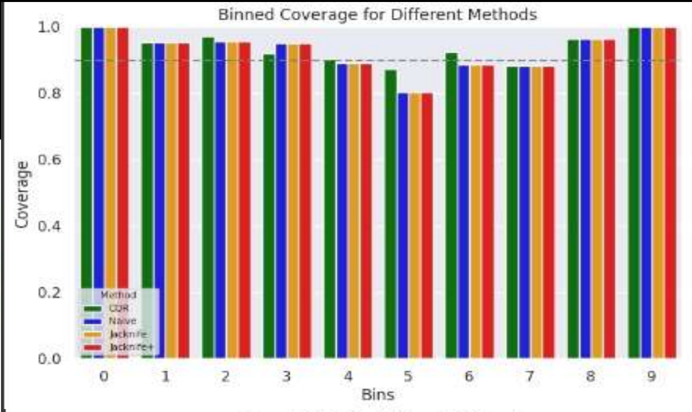


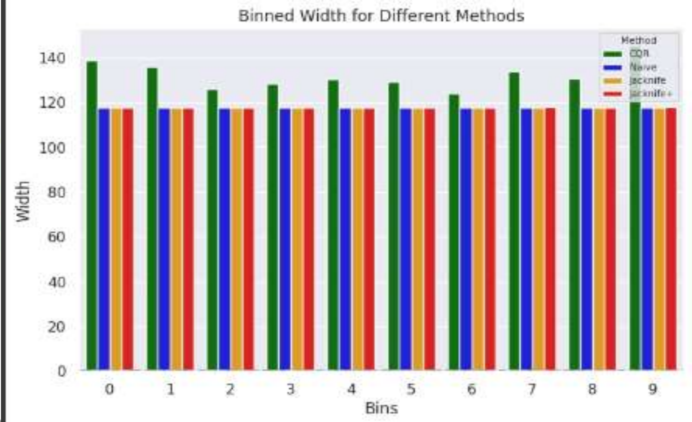


Quantile Regression Results

Quantile regression compared with LightGBM is poor in the prediction







Conclusions

MAPIE is the right tool for conformal prediction:

-MAPIE is designed to provide prediction intervals that can be applied to any predictive model without requiring modifications to the model itself.

-MAPIE ensures that the prediction intervals meet the desired coverage probability empirically by using a portion of the data to calibrate these intervals, providing a reliable measure of uncertainty for predictions.

-MAPIE it's straightforward to implement.

References

[MediumArticle]
[Notebook]
[Dataset]
[StreamlitApp]
[MAPIE_documentation]
[PracticalGuideAppliedConformalPrediction]
[ConformalPrediction]

EuroSciPy Szczecin 2024

Thank you

Keep in touch:

- Linkedin
- Newsletter
- Medium
- Website

Towards Innovation Lab