TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Jan Kokla 206588

APPLYING CONFORMAL PREDICTIONS TO THE REGRESSION PROBLEM: A CASE STUDY OF A FINTECH

Bachelor's Thesis

Supervisor: Gert Kanter
PhD

Tallinn 2023

Summary

In this thesis, the application of *conformal predictions* (CP) to a regression problem is explored to transform point predictions into probabilistic ones. Unlike deterministic predictions that yield a single forecast, probabilistic predictions offer a range of potential outcomes along with their corresponding probabilities. They enable decision-makers to better assess risk and the potential outcomes of various scenarios. Recently, CP has gained attention as it provides prediction intervals with guaranteed coverage relying on minimal assumptions and acting as a wrapper around any model.

Different methods of conformal predictions were compared and benchmarked using a Python package called MAPIE. These methods were evaluated against an alternative solution involving the use of *quantile regression* (QR), which employs a tailored loss function to produce prediction intervals. The models were evaluated by the correctness, ensuring the actual coverage aligns closely with expectations, and the adaptivity of prediction intervals, allowing for small intervals on easy inputs and larger intervals on more challenging inputs with some degree variability in the data.

While QR yielded the smallest prediction intervals, it fell short in achieving the desired coverage, rendering it unreliable and unsuitable for decision-making purposes. Alternatively, the *inductive conformal predictor* exhibited correct coverage but fixed interval widths, limiting its practicality in real-life applications. Finally, *jackknife+-after-bootstrap* with gamma conformity score, managed to offer the best trade-off between correctness and adaptivity. The thesis successfully demonstrated the feasibility of conformal prediction as a viable solution for uncertainty quantification, given its guaranteed coverage and adaptive nature of prediction intervals.