

Forum:[Hack Forums](#)

Thread: [How to Teach Machine Learning Models a Sense of Uncertainty](#)

Stanley - October 6, 2025, 11:35 pm

How to Teach Machine Learning Models a Sense of Uncertainty

Conformal prediction is a machine learning framework designed to quantify uncertainty in model outputs. Instead of making a single prediction, it provides a range of likely outcomes along with measures of confidence and credibility. Its agnostic nature means it can work with any algorithm and applies to both classification and regression tasks. Mondrian conformal prediction extends this framework to handle imbalanced datasets by ensuring equal error rates across classes. Together, these methods make AI predictions more interpretable, transparent, and reliable for real-world, high-risk applications.

<https://hackernoon.com/how-to-teach-machine-learning-models-a-sense-of-uncertainty?source=rss>

DAL-SHA - October 7, 2025, 12:56 am

Teaching models uncertainty is smart—helps trust AI more by showing confidence levels, especially in tricky or imbalanced data situations.

Vectura - October 7, 2025, 1:42 am

Really solid piece, conformal prediction is such an underrated concept. Teaching models to express how sure they are, not just what they think, is a big step toward safer, more trustworthy AI systems.

Fish - October 7, 2025, 1:42 am

I feel like this is really smart because right know they have too much false confidence

Ieste - October 7, 2025, 1:47 pm

Mondrian conformal prediction could be a game-changer for imbalanced data tasks.

JoshAllen - October 7, 2025, 3:53 pm

I really like this. AI puts out a lot of answers with confidence, even if it is not confident. Sometimes putting out false information.

Being able to communication that the answer is not confident, will let people try again or research deeper in a specific area to get the answer they are looking for.

A lot of these companies really don't want to do this because it could be seen as a weakness to others.

Hackers - October 16, 2025, 9:39 am

This is super important... if you train AI on producing the "most accurate" or the "best answer", it actually degrades.

Part of the artform of AI is teaching it to be wrong enough so it can understand differences in perspectives, answers, and solutions. From there, it can derive some of the best answers, and most accurate info... but it must arrive at this on its own.

