Deloitte.

Shedding Light on the Inner Workings of AI Models

Deloitte's Lucid [ML] Explainable AI Offering.

The Need

Machine Learning [ML] models are raising the bar in terms of prediction accuracy. Every day, Data Scientists around the world actively leverage the possibilities of ever broader and deeper sources of data.

The advantages are clear: Greater accuracy improves quality, reduces waste, and heightens efficiency. Yes, all this comes at the cost of model complexity. Their ability to deftly harvest information from vast quantities of data renders the models themselves complex. This is particularly true for deep neural networks (DNNs), arguably the greatest contributor to economic value from artificial intelligence technologies to date, and the underlying technology behind now famous large language models and other forms of Generative AI. The power of DNNs is entwined in its intricate web of nodes, organized into numerous hidden layers, each specialized in recognizing features and sub-patterns at varying levels of abstraction.

The downside: These high-performance models are "black boxes" – so opaque that not even their creators can know how they arrive at their conclusions or predictions. Providers and users place their faith in black box models because they work, often extremely well. Experience and back-testing provide a compass to refine their capabilities, a process of informed trial and error.

This poses a variety of challenges to companies seeking to exploit the power of AI, especially in regulated industries. The lack of transparency hinders adoption for banks and insurers, who are left to gaze enviously at clever innovations in other industries, such as retail commerce.

MAKING AN IMPACT THAT MATTERS since 1845

Our Solution: Lucid [ML]

Deloitte's explainable AI solution Lucid [ML] brings transparency to Machine Learning models. It achieves this through a variety of methods that track a feature's path throughout the model's learning process. The user may select the most suitable approach and degree of investigation.

The drivers of model behavior can be explained either at a global or local level. The global level is of particular interest to Data Scientists. The local level is of particular interest to Audit or Compliance. Another useful feature is the inclusion of contrastive explanations, known as "counterfactuals", which illustrates how far an individual data point was from the model's decision boundary.

Experts in explainable AI will note that methods such as LIME (local interpretable model-agnostic explanations) and SHAP (Shapley additive explanations) are freely available as open-source. Lucid makes use of these open-source packages for the "quick scan" functionality. Yet they are limited and can even be misleading. Lucid resolves this through a proprietary explainability method "LucidSHAP" that also functions where model features are not perfectly orthogonal – a characteristic not frequently observed in real life data.

LucidSHAP resolves this by feeding both feature combinations and their dependence information, simulating more realistic datasets. These simulated datasets are then passed to the black box model, resulting in a more accurate ranking of model drivers.

This is a computationally intensive process, not practical for low-power laptop CPUs. The latest release of Lucid has been optimized in direct collaboration between Deloitte and NVIDIA to detect GPUs and, if present, to opt for a GPU-optimized computation process. The result is a 70x acceleration in the calculation over the original CPU-limited release on large datasets.

Advantages/Benefits of Lucid [ML]

- Companies may apply ML models they understand more confidently to tackle their most complex problems.
- Advanced methods need not be hindered by opacity and comply with regulations.
- Model developers may communicate model functionality more easily to individual stakeholders for both models in operation or models in development.
- Gaining insights into business process inputs, such as identifying key variables, or prioritizing data quality efforts.

Example Use Cases

- Interpreting credit decisioning models.
- Fulfilling profiling transparency obligations (GDPR Article 12).
- Providing explanations to individually rejected loan applications (GDPR Article 20).
- Understanding drivers of positive or negative sentiment toward their company or produces on social media.
- Identifying blind-spots in computer vision – which characteristics require more training to distinguish objects

Contact

David Thogmartin

Director | Risk Advisory
AI & Data Analytics | aiStudio
dthogmartin@deloitte.de

Video demonstration

Interested in seeing Deloitte's Lucid [ML] for yourself?

Click here to watch a demonstration.

Deloitte.

Issue 03/2024