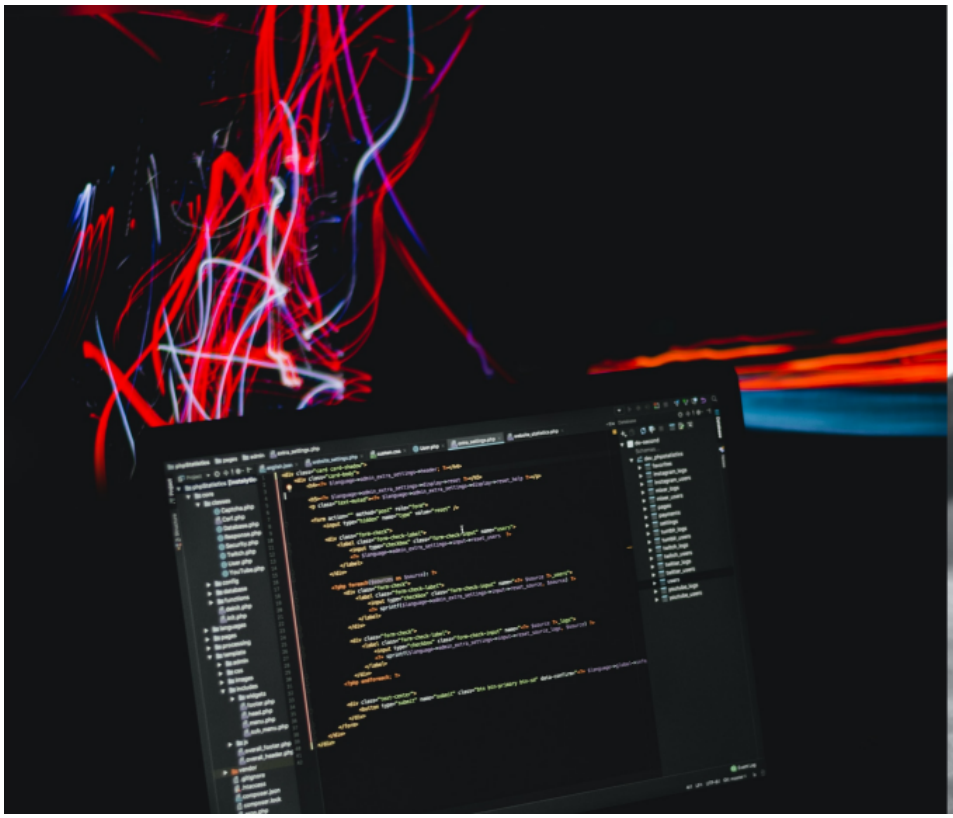
The Clever Hans Effect in Machine Learning: an overview by Bhusan Chettri

The Clever Hans effect occurs when a machine learning model performs well but relies on irrelevant cues or confounding factors in the training data. In this article Dr Chettri introduces the effect and emphasizes its importance in understanding.



Bengaluru, Karnataka Jul 24, 2023 (Issuewire.com**)** - Dr. Bhusan Chettri, a PhD holder in applied AI and machine learning with a strong research background, introduces the Clever Hans Effect in this article. Exploring its relevance to machine learning, the article examines common areas where the effect can occur and emphasizes the importance of awareness when developing data-driven machine learning systems. By understanding the Clever Hans Effect, one can enhance the development of accurate and robust machine learning algorithms that effectively capture relevant patterns in the data. This article serves as an introductory version of Dr. Chettri's recently published paper in the esteemed IEEE Spoken Language Technology 2022 conference held in Doha, Qatar.

Machine learning (ML) is a field of study that involves using algorithms to analyze and learn from data, with the goal of making predictions or decisions based on that data. ML has become an increasingly popular field of study in recent years, with numerous applications across a wide range of industries such as retail, healthcare, finance, marketing, to name a few. However, as with any powerful tool, it is important to use it with care and consideration. One issue that can arise in machine learning is the Clever Hans Effect, which refers to the phenomenon where an algorithm appears to have learned a solution to a problem but is just picking up on irrelevant cues in the data.

The term "Clever Hans" refers to a horse that gained fame in Germany in the early 1900s for appearing

to be able to solve arithmetic math problems and answer questions by tapping his hoof. The horse was owned by a math teacher named Wilhelm von Osten, who trained Hans to respond to various cues, such as changes in facial expressions or body language from his trainer. Von Osten believed that Hans was actually capable of understanding and solving math problems, but others suspected that the horse was simply picking up on cues from his trainer. As Hans gained more attention and became a popular attraction, scientists and psychologists began to investigate his abilities. A number of experiments were conducted to test whether Hans was actually solving math problems or just responding to cues from his trainer. These experiments eventually revealed that Hans was indeed responding to cues, often very subtle ones that were unintentional on the part of his trainer. This phenomenon became known as the "*Clever Hans Effect*" and has since been recognized as an important lesson in understanding how animals/humans (and machines) can inadvertently pick up on cues from their environment.

In machine learning, the ***Clever Hans Effect*** can occur when an algorithm is trained on a biased or incomplete dataset. For example, an algorithm may be trained to recognize cats based on a dataset that only includes images of cats on a green background. The algorithm may learn to associate green backgrounds with cats, rather than actually recognizing the features that make up a cat. This can lead to poor performance when the algorithm is applied to new, unseen data (where there is no green background, for example). The Clever Hans Effect is relevant to machine learning because it can cause algorithms to appear to be performing well on a task, when in fact they are just picking up on irrelevant cues in the data. This can occur when the algorithm is overfitting the data, meaning that it is learning to recognize patterns that are specific to the training data but may not generalize well to new data. In this way, machine learning algorithms can suffer from a similar phenomenon to Clever Hans, where they appear to be "smart" but are actually just picking up on patterns in the data that are not relevant to the problem they are trying to solve.

Dr Chettri explains that it is important to be aware of the Clever Hans Effect in machine learning because it can lead to inaccurate or biased predictions or decisions. For example, if a credit scoring algorithm is trained on historical data that contains biases against certain groups, such as minorities or low-income individuals, the algorithm may learn to discriminate against those groups without even realizing it. Similarly, a medical diagnosis algorithm that learns to recognize certain features in images of tumors may not actually be recognizing the cancer itself, but rather the particular imaging technique used to produce the images.

By understanding the Clever Hans Effect, data scientists and ML developers can develop more robust and accurate machine learning systems that truly learn the patterns in the data that are relevant to the problem at hand. This involves careful feature selection and regularization to avoid overfitting, as well as rigorous testing and evaluation to ensure that the algorithm is truly learning the underlying patterns in the data. Ultimately, being aware of the Clever Hans Effect can help build more trustworthy and effective machine learning systems that are grounded in a deeper understanding of the data they are analyzing. It should also be noted that the Clever Hans Effect can have serious implications for machine learning systems, particularly in high-stakes applications such as healthcare or finance. If an algorithm is trained on biased or incomplete data, it may make incorrect decisions that could have real-world consequences. It is therefore crucial to be aware of the Clever Hans Effect and take steps to minimize its impact when building machine learning systems.

References:

- https://theses.eurasip.org/media/theses/documents/bhusan-chettri-voice-biometric-system-security-design-and-analysis-of-countermeasures-for-replay-attacks.pdf
- https://ieeexplore.ieee.org/document/10022624

- https://www.issuewire.com/automatic-speaker-recognition-technology-and-spoofing-attacks-an-overview-by-bhusan-chettri-1737133109911273

Media Contact

Borac Solutions

info@boracsolutions.com

Source : Borac Solutions (https://ieeexplore.ieee.org/document/10022624)