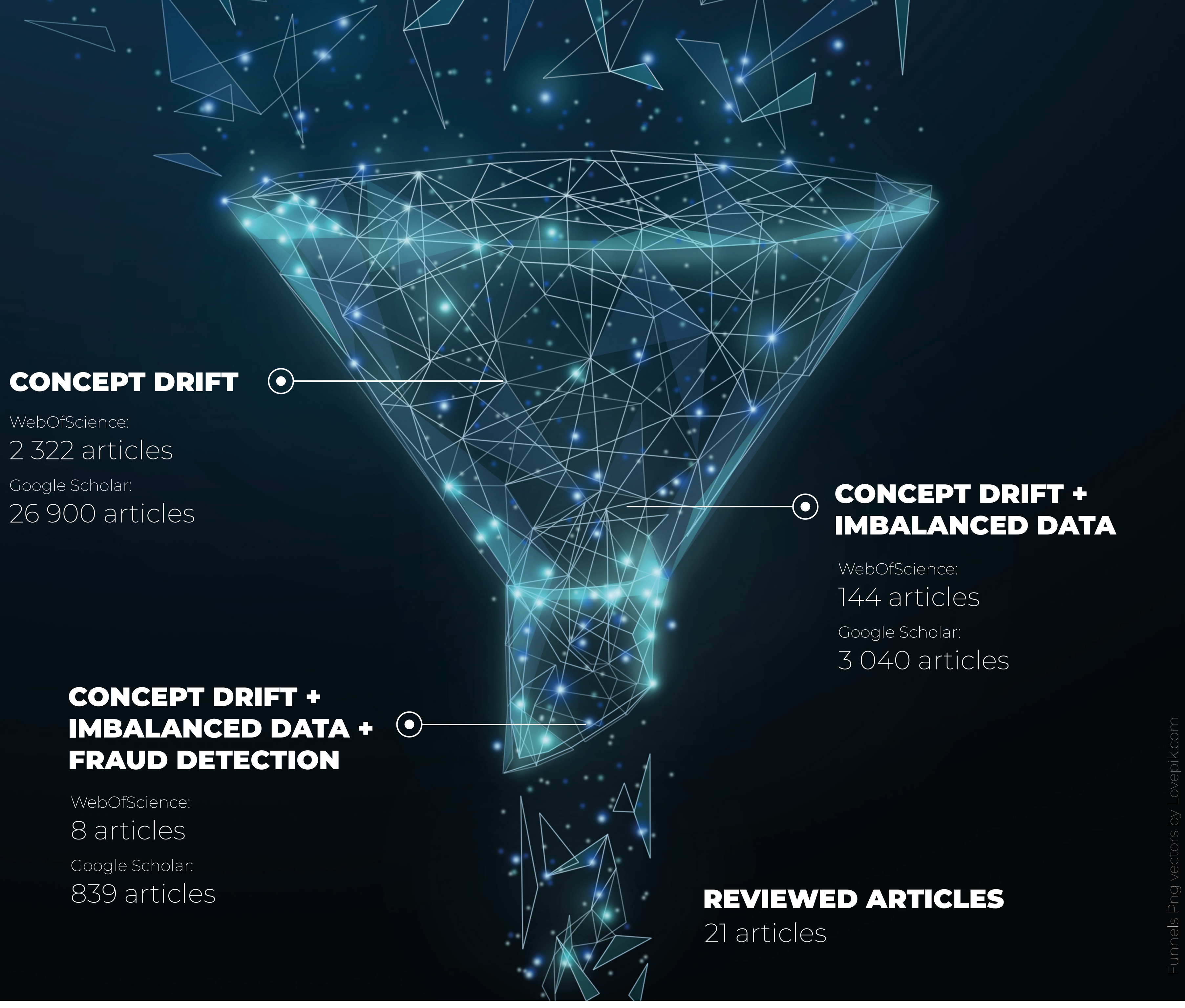
WHAT IS A CONCEPT DRIFT?

Does it affect machine learning performance?

Dalia Breskuviene and Gintautas Dzemyda,
Data Science and Digital Technologies Institute
Vilnius University

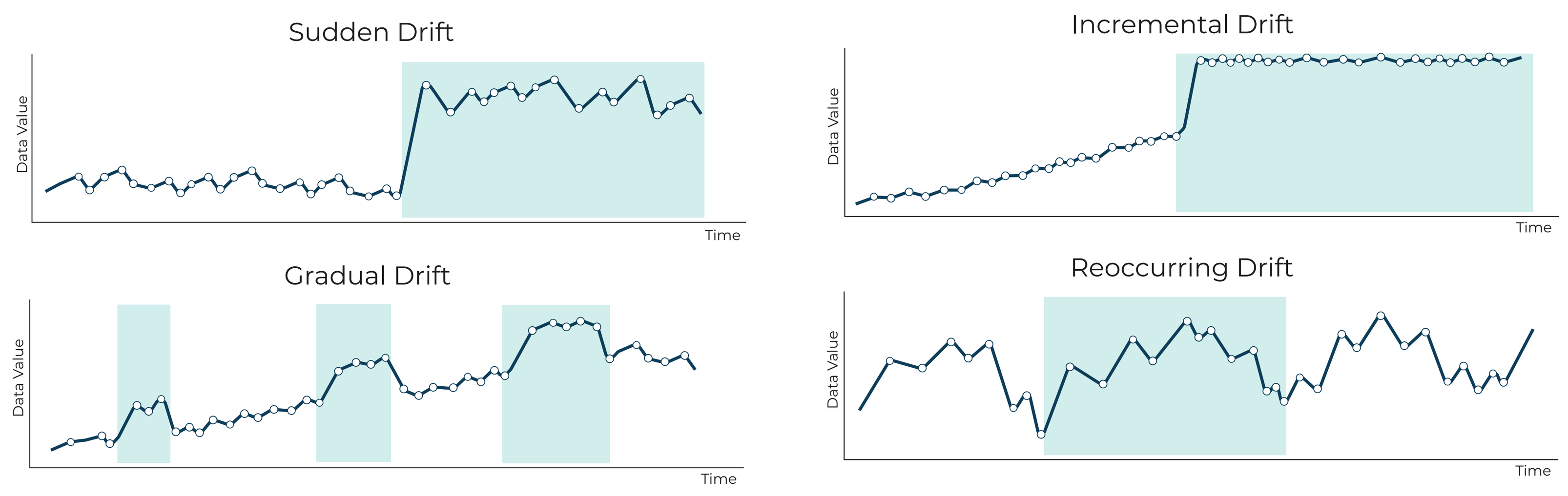


ABSTRACT

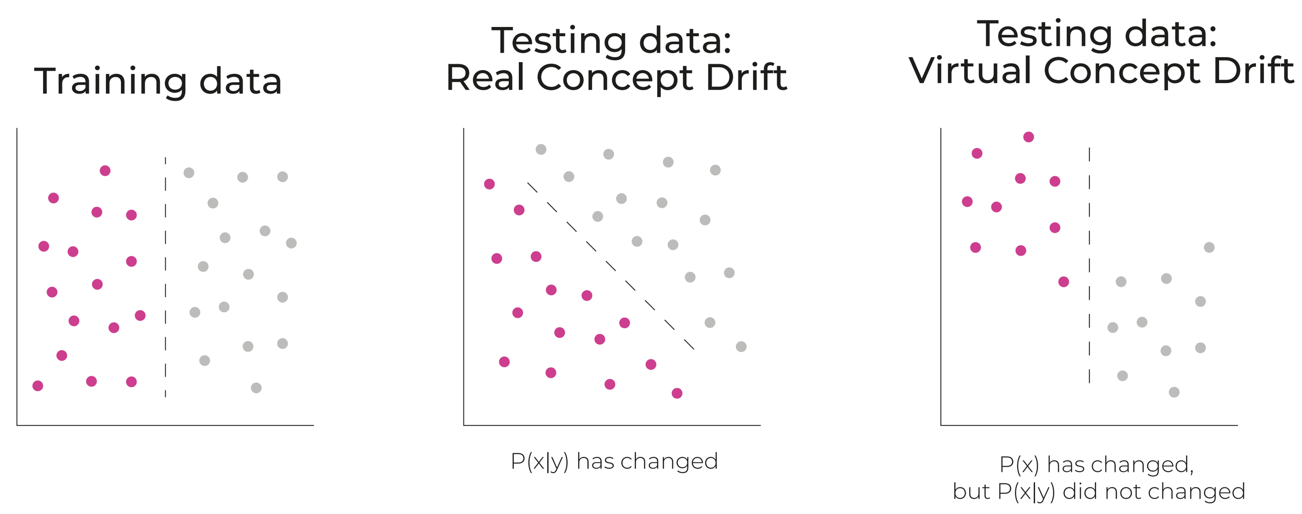
Traditional machine learning assumes consistent class proportions and data distribution over time. However, this is not the case in real-world scenarios like fraud detection, medical diagnosis, natural disaster prediction, or manufacturing quality control. **Concept Drift** is a phenomenon where the statistical properties of the target variable or data distribution change over time, meaning that models trained on historical data become less effective. This study explores concept drift, its manifestations, and its impact on machine learning, especially in **fraud detection**.

CONCEPT DRIFT DEFINITION

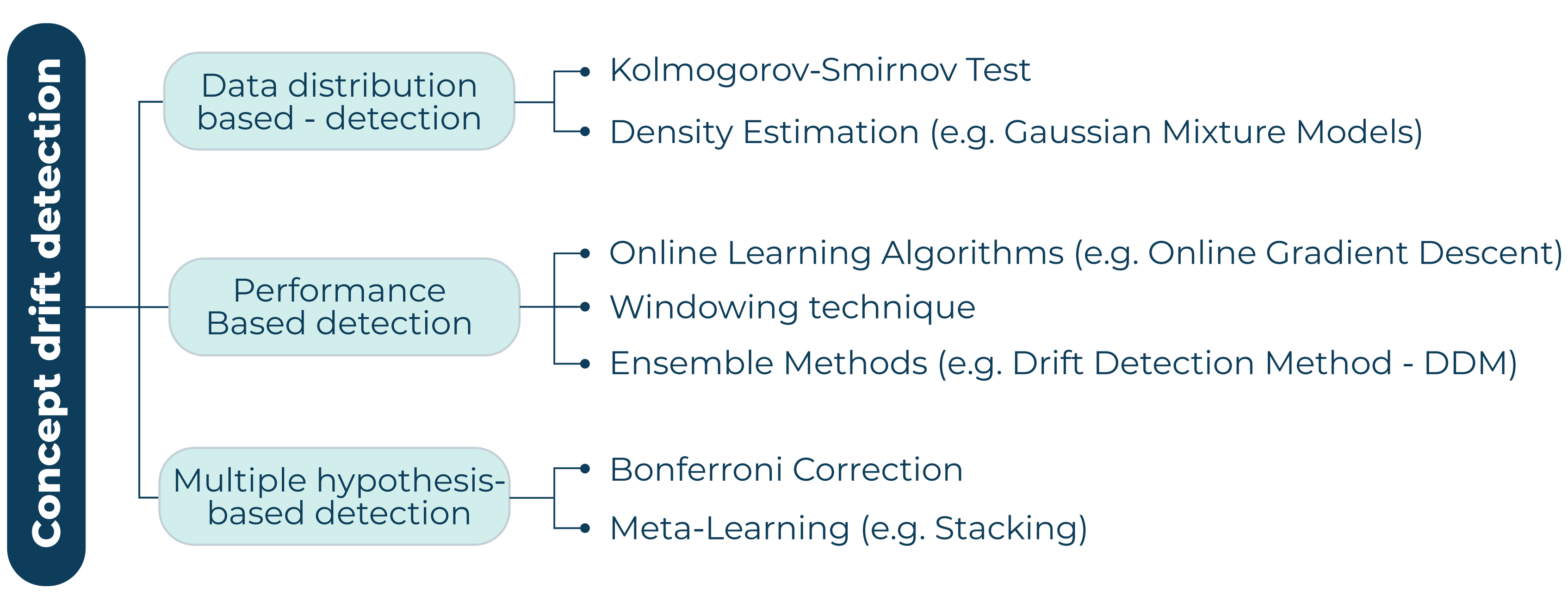
Concept drift is a fundamental idea in machine learning, encapsulating the phenomenon where the statistical properties of the target variable or data distribution change over time. These changes can be gradual or abrupt, challenging the conventional machine learning assumption that the data used for training and testing remains static.



Concept drift can be systematically categorized not only by the temporal dynamics but also by determining its effects, which can be attributed to the shifting nature of data distribution.



CONCEPT DRIFT DETECTION

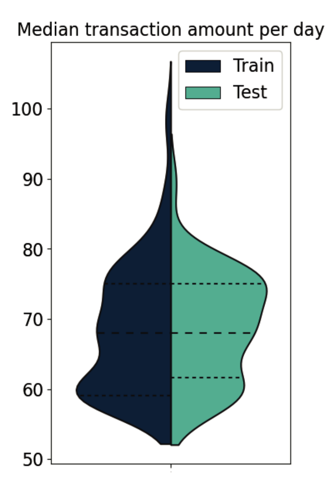


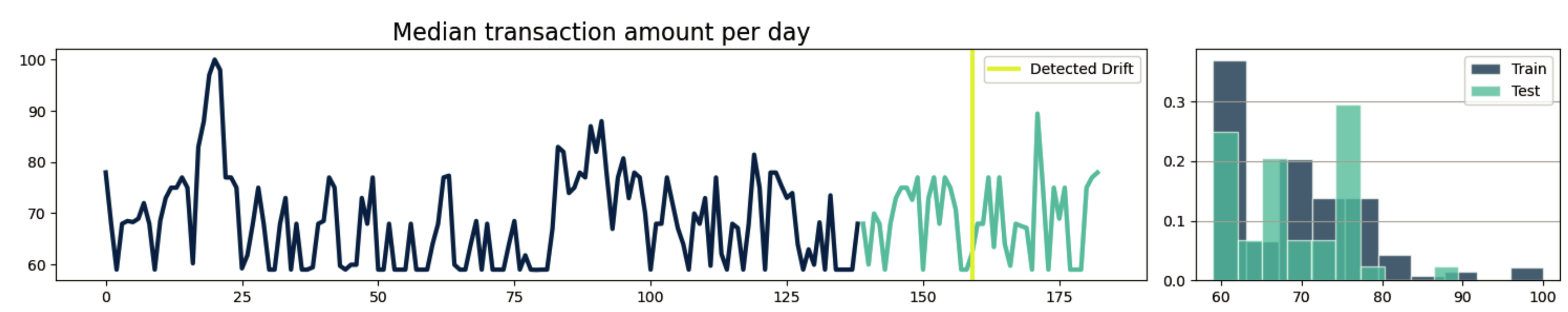
ML STATE-OF-ART TECHNIQUES

State-of-the-art techniques in ML for handling concept drift have moved beyond traditional approaches, such as periodic model retraining. Online learning algorithms, such as Online Passive-Aggressive Algorithms, enable continuous model updates with each new instance, ensuring adaptability to evolving data patterns. Ensemble methods, presented by Drift Detection Ensemble, leverage the power of combining multiple models to enhance robustness in the face of concept drift. Incremental learning techniques can efficiently update models without the need for full retraining. Deep learning approaches involving recurrent neural networks and long short-term memory networks have shown promise in adapting to dynamic environments. As these techniques evolve, addressing challenges and exploring future research directions, they play a crucial role in ensuring the reliability and adaptability of machine learning models in the presence of concept drift.

FRAUD DETECTION CASE STUDY

Picture yourself as a data scientist maintaining a promising fraud detection model. This model consistently processes predictions for an expanding user base. As the landscape evolves, changes in customer behavior due to the increasing number of customers and economic changes like raising interest rates further complicate the dynamic nature of the data. The challenge of addressing concept drift extends beyond merely maintaining prediction flow. It is a crucial aspect of upholding user trust, mitigating unforeseen costs, and guaranteeing the reliability of machine-learning models in a continually evolving environment.





HIGHLIGHTS

- Concept drift introduces variability in temporal data, necessitating adaptive models for sustained accuracy;
- Imbalanced data is much more sensitive to real concept drift as the scarcity of instances in the minority class amplifies the challenges associated with adapting to changes in the underlying data distribution;
- Fraud detection uses streaming data, where recognizing concept drift becomes particularly challenging, given the dynamic nature of the data flow.

REFERENCES

J. Read and I. Žliobaitė
A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy," IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3784–3797, Aug. 2018, doi: 10.1109/TNNLS.2017.2736643.
J. Gama, P. Medas, G. Castillo, and P. Rodrigues, Learning with Drift Detection, vol. 8. 2004, p. 295. doi: 10.1007/978-3-540-28645-5_29.
G. M. Paldino et al., "The role of diversity and ensemble learning in credit card fraud detection," Adv Data Anal Classif, Sep. 2022, doi: 10.1007/s11634-022-00515-5.

CONTACT ME