



Black Box vs. White Box Modelle in der Betrugsbekämpfung

Abstract: Am 19. November fand unser Online-Kolloquium zum Thema **Black Box vs. White Box Modelle in der Betrugsbekämpfung** statt. Grund genug an dieser Stelle die wichtigsten Erkenntnisse zusammenzufassen.

Den Beginn der Veranstaltung machte Dr. Gagl mit Einblicken, wie in den kognitiven Neurowissenschaften „Black“ Box und „White“ Box Modelle genutzt werden, um die Funktionen des menschlichen Gehirns zu verstehen. Konkret hat Dr. Gagl die unterschiedlichen Ansätze anhand von Beispielen aus der Forschung zur Objekt- und Worterkennung erläutert.

Sein Fazit zu den Vor- und Nachteilen zu den beiden Herangehensweisen lassen sich in vielen Aspekten auch auf den Einsatz in der Bekämpfung von Finanzkriminalität übertragen.

	Black Box Modelle	White Box Modelle
Struktur	Komplex	Relativ einfach
Vorhersage	Schnell relativ gut	Inkrementeller Fortschritt
Generalisierung	Gut	Schwer möglich Expertenwissen wichtig
Ursachenforschung und Intervention	Sehr schwer möglich	Sehr gut möglich

Quelle: Dr. Benjamin Gagl, Vortrag spotixx Online Kolloquium 19.11.2020

Wenn man mit dem Begriff „White Box“ die Vorstellung von **Interpretierbarkeit und Nachvollziehbarkeit** verbindet, so wird schnell klar, dass hier auch die **Zielgruppe** eine ganz wesentliche Rolle spielt. Was für einen Data Scientist transparent und nachvollziehbar ist, mag für den Fraud-Verantwortlichen ein großes Mysterium darstellen. Insofern spielt der Adressat und Nutzer eine große

Bedeutung bei der Einschätzung von White Box Modellen.

Klar wurde dies am Beispiel verschiedener Use Cases mit unterschiedlichen Beteiligten: Während z.B. Banken und Aufsichtsbehörden ein klares Bild von Transparenz für interne Rating Modelle haben, tun sich die gleichen Player im Bereich Geldwäsche ungleich schwerer mit quantitativen Modellen. Die Akzeptanz für den Einsatz von Machine Learning in der Geldwäsche Bekämpfung und False-Positive-Reduzierung fehlt an vielen Stellen noch!

In der Folge haben wir neben dem Faktor Mensch auch den **Faktor Maschine** beleuchtet – ein Aspekt, der oft nicht betrachtet wird. Genauso wie es eine Maschine, d.h. einen Algorithmus, gibt, die ein Modell erstellt, gibt es – analog zum Menschen – auch eine Maschine, die das Modell interpretieren muss! Viele proprietäre Fraud Monitoring Systeme in Finanzinstituten kennen jedoch nur einfache Regel-Interpreter. Also auch IT-Systeme tun sich mit White Box Modellen durchaus leichter.

Dies bildete den Brückenschlag zum Fraud Analytics Service von spotixx, der in der Folge live präsentiert wurde. Ein grundlegendes Designprinzip der Lösung ist es, dass sowohl Mensch und Maschine die erstellten Fraud Prevention Modelle in Form einfacher Regeln verstehen und anwenden können (https://spotixx.com/factsheet_ruleservice_de).

So konnten wir in unserem Kolloquium den weiten Bogen von der akademischen Forschung mit White Box Modellen hin zu praktischer Anwendung in Finanzinstituten schlagen.

Über Rückfragen zu unserem Kolloquium und zu unserer Lösung freuen wir uns sehr:

Ruppert Jaeschke – rj@spotixx.com oder +49 (0) 160 644 1933