

Data-Mining aus Freitextinformationen von Tierarztpraxen im Rahmen des Forschungsprojektes zur Erfassung des Antibiotikaeinsatzes bei Hunden, Katzen und Pferden

in Informatik / Bioinformatik / Statistik / Data Science / Mathematik / Physik

Der Einsatz der Antibiotika in der Tiermedizin kann die Entstehung von Antibiotikaresistenzen begünstigen, die die öffentliche Gesundheit im zunehmenden Maße gefährden. Im Rahmen einer am Institut für Veterinär-Epidemiologie und Biometrie ausgeschriebenen Masterarbeit soll ein Konzept zur Erfassung des Antibiotikaeinsatzes bei Hunden, Katzen und Pferden entwickelt werden. Eine Vielzahl von Daten bzgl. Patienten, Diagnose, Behandlung, Antibiotikaeinsatz und Resistenzen sollen aus den kooperierenden Laboren und Tierarztpraxen automatisiert über eine Schnittstelle in eine Datenbank abgelegt werden. Benötigte Informationen können an unterschiedlichen Stellen und auch untereinander kombiniert abgelegt sein (z. B. in großen Freitextfeldern). Dies könnte zur Folge haben, dass die Datenstruktur nicht mit der Projektdatenbank kompatibel ist. Mithilfe von Methoden der künstlichen Intelligenz und des maschinellen Lernens wird ein Modell trainiert, das in Freitextfeldern bestimmte Begriffe (z. B. Arzneimittelnamen und deren Abkürzungen, aber auch Dosierung, Dauer der Anwendung etc.) identifiziert und in die gewünschte Struktur der Projektdatenbank transformiert.

Voraussetzungen sind:

- Solide Kenntnisse und praktische Erfahrung mit den Methoden des Maschinellen Lernens für Data und Text Mining (entity recognition) in Python 3, Tensorflow, spaCy, BERT
- Teamfähigkeit, Verlässlichkeit, Kommunikationsfähigkeit

Für weitere Fragen stehen PD Dr. Roswitha Merle (roswitha.merle@fu-berlin.de) und Prof. Dr. Vitaly Belik (vitaly.belik@fu-berlin.de) zur Verfügung.