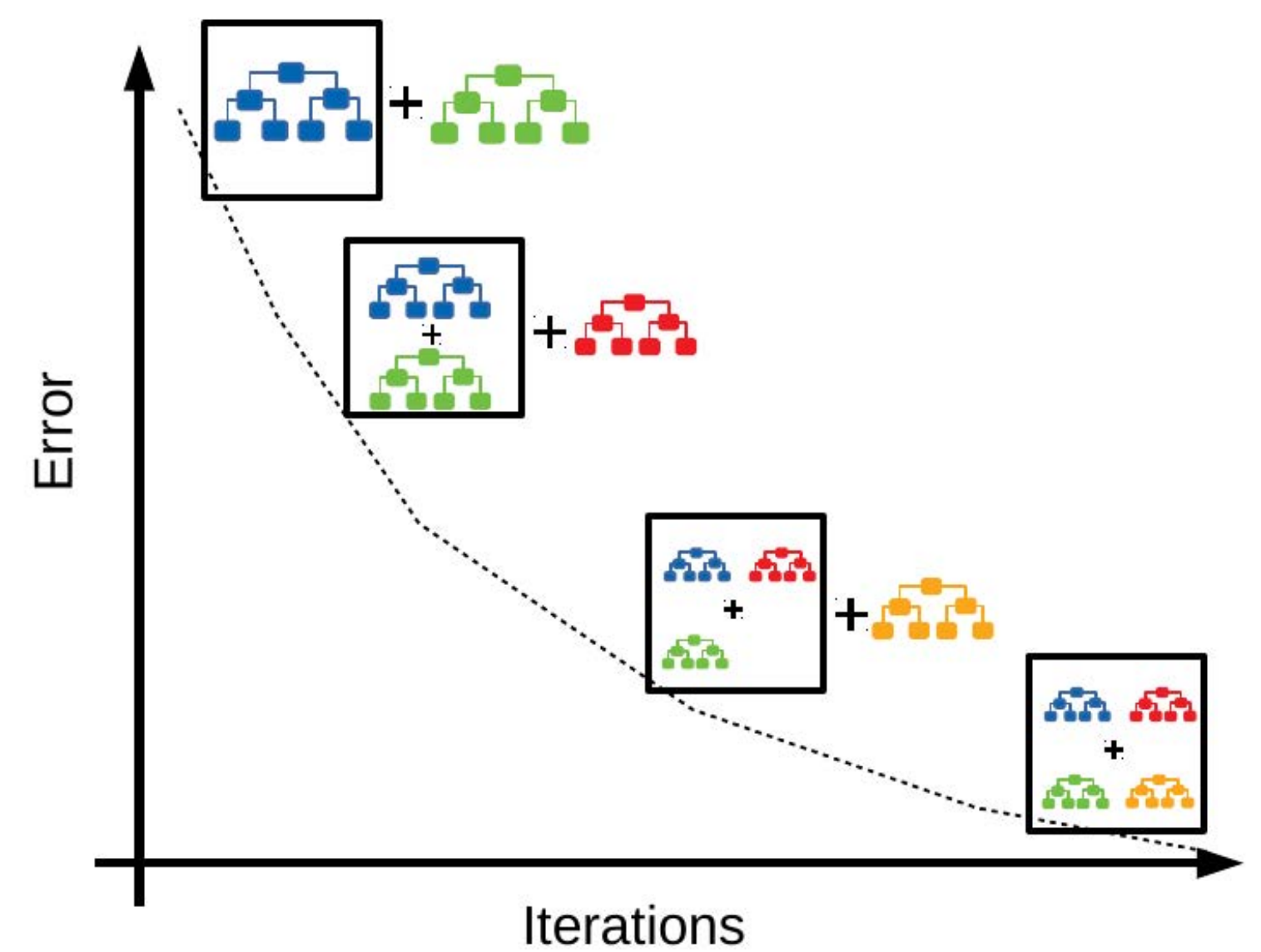
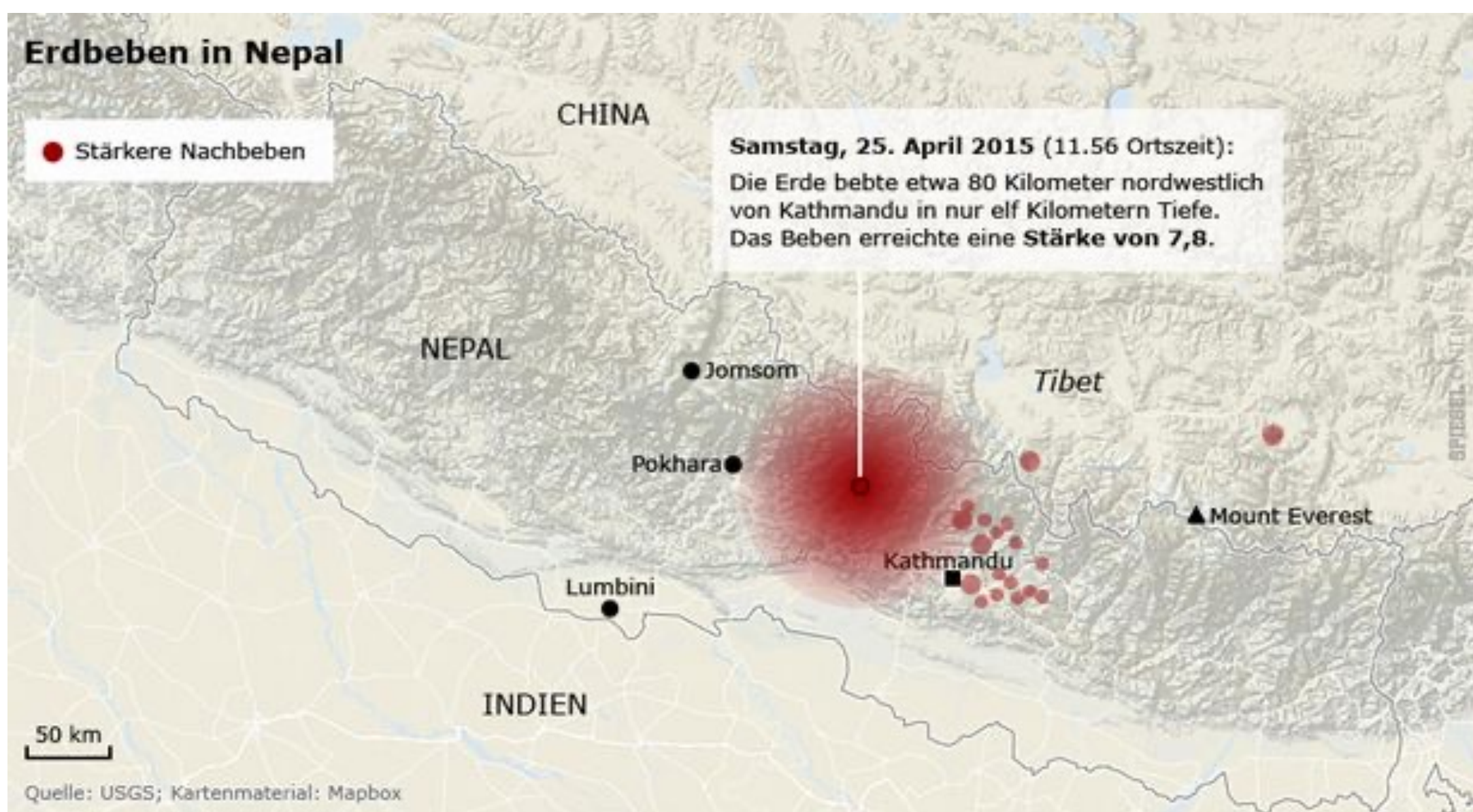


Data Science Methoden für die Vorhersage von Katastrophenfolgen am Beispiel des Erdbebens in Gorkha – Nepal in 2015



Fragestellungen:

Wie können Machine Learning Modelle verwendet werden, um Gebäudeschadensklassen vorherzusagen am Beispiel des Erdbebens in Gorkha?

Die Abschlussarbeit widmet sich der Frage wie Data Science bei der *Prävention, während der akuten Katastrophenlage* und der *Nachsorge im Katastrophenmanagement* und in der Versicherungsindustrie helfen. Es soll das Prinzip Data Science im Allgemeinen erklärt werden, um dann spezifischer auf die einzelnen Machine Learning Methoden einzugehen. Hierbei soll erklärt werden was ein Entscheidungsbaum, ein Random Forest und ein Boosting Tree ist und welches dieser Modelle das präzisere Ergebnis erzielt. Dies soll anhand des Beispiels des Erdbebens von Gorkha in Nepal im Jahre 2015 validiert werden. Das Erdbeben von Gorkha in Nepal ereignete sich am 25. April 2015 um 11:56 mit einer Stärke von 7,8 – 8,1 Magnitude auf der Richterskala, bei dem 8.891 Menschen umgekommen sind und 21.952 Menschen verletzt wurden. Nach dem Erdbeben wurde der Schaden statistisch erfasst, hierbei wurde für jedes Haus eine Schadensklasse zugewiesen, die zeigen sollen, ob ein Haus nicht beschädigt wurde, bis zur vollständigen Zerstörung des Hauses. Diese Schadensklassen sollen durch die oben genannten Modelle vorhergesagt werden. Die Daten wurden von DrivenData bereitgestellt. Diese müssen aufbereitet werden und mit der Programmiersprache R verarbeitet werden.

Das Erdbeben von Gorkha dient als Beispiel, um zu zeigen, wie eine Vorhersage mit verschiedenen Machine Learning Algorithmen durchgeführt und wie das Ergebnis bemessen werden kann. Der benutzte Datensatz besteht aus 260.000 die Gebäude mit 39 Merkmalen. Die Zielklassen beschreiben den Beschädigungsgrad der Gebäude nach dem Erdbeben - nicht beschädigt, beschädigt oder vollständig zerstört - Durchgeführt wurde die Vorhersage mit der integrierten Entwicklungsumgebung RStudio und der Bibliothek „caret“, dass eine Reihe von vorhandenen R-Paketen benutzt. Es wurden die Machine Learning Algorithmen Random Forest, GBM und XGBoost verwendet. Bei diesen Modellen wurde ersichtlich das die wichtigsten Variablen nahezu identisch sind. Vor allem die Merkmale für die Lokation der Gebäude, Art des Fundamentes und des Hochbaus sowie das Alter spielten eine wichtige Rolle. Das beste und präziseste Ergebnis mit 73,9% erreichte der Ensemble Learning Algorithmus XGBoost der eine verbesserte Variante des GBM ist.

