

## Einführung: Machine Learning

Beim Machine Learning werden die in einem Datensatz enthaltenen Informationen als Grundlage für eine automatisierte Verallgemeinerung herangezogen. Dadurch kann Machine Learning Forscherinnen und Forscher beispielsweise bei sehr großen und dadurch oftmals unübersichtlichen Datensätzen erste weiterführende Hinweise auf die in einem Datensatz enthaltenen Informationen liefern und den Prozess des Beschreibens, Erklärens und Vorhersagens vereinfachen. Hierzu wird beim Machine Learning in der Regel auf erlernbare Beispiele zurückgegriffen, die sich ebenfalls aus den in einem Datensatz enthaltenen Informationen ergeben. Über in *Algorithmen* eingebettete statistische Formeln soll dabei ein einschlägiges Modell gebildet werden, welches die zugrundeliegenden *Muster und Gesetzmäßigkeiten* über die erlernbaren Beispiele hinaus aufzuschlüsseln vermag. Als Algorithmus gelten dabei alle von Forscherinnen und Forschern vordefinierten Handlungsvorschriften zur Lösung eines oder mehrerer Probleme beim Beschreiben, Erklären und Vorhersagen auf der Grundlage eines Datensatzes. Demnach handelt es sich bei der Verkettung mehrerer statistischer Formeln zur Lösung eines oder mehrerer Probleme um einen *Machine Learning Algorithmus*, wenn die Verkettung entsprechend auf die erlernbaren Beispiele zurückzuführen ist. Mit dem insbesondere beim Beschreiben und Erklären des Datensatzes stattfindenden Lerntransfer soll Machine Learning dazu beitragen, auch unbekannte Daten beurteilen und darauf aufbauend einschlägige Vorhersagen treffen zu können.

Prinzipiell wird dabei in *Supervised Machine Learning*, also von Forscherinnen und Forschern überwachte Machine Learning Algorithmen, sowie dem automatischen Gegenstück des *Unsupervised Machine Learnings* unterschieden. Beim Supervised Machine Learning definieren Forscherinnen und Forscher sowohl *Input* als auch *Output* in einem Datensatz und ermitteln über die in den Algorithmen eingebetteten statistischen Formeln das zugrundeliegende Modell zur Vorhersage des Outputs – dieses Vorgehen wird auch *Classification* genannt. Entsprechend versuchen die Algorithmen des Unsupervised Machine Learnings Input und Output zunächst über die zugrundeliegenden Muster und Gesetzmäßigkeiten zu erfassen und zu systematisieren – beispielsweise mittels des sogenannten *Clusterings*. Sukzessive kann daraufhin automatisch ein entsprechendes Modell auf der Grundlage geeigneter statistischer Formeln gebildet werden. Dabei kommen insbesondere die bereits aus der deskriptiven, bivariaten und multivariaten Analyse bekannten statistischen Formeln zur Anwendung. Die nachfolgenden Videos verdeutlichen das Vorgehen an anschaulichen Beispielen...