Text von der Anfangspräsentation

# Folie 1 Einführung

Hallo, ich heiße Atanas Dimitrov und ich werde euch ganz grob meine Bachelorarbeit vorstellen. Der Titel ist „Automatische Auswahl für von maschinellen Lernverfahren für kausale Inferenz“. Erstmal kläre ich, was kausale Inferenz ist.

# Folie 2 Kausale Inferenz

Die kausale Inferenz gibt uns Kenntnisse darüber, inwiefern eine Handlung, die wir durchführen, ein Objekt bewirkt. Diese Handlung kann alles sein – geben von einem Medikament, Werbung, politische Entscheidung. Stellt euch vor, wir möchten ein Experiment mit zwei Gruppen – Behandlungsgruppe und Kontrollgruppe. Dabei haben wir Rahmen. Das Objekt kann als ein Vektor von Variablen dargestellt werden. Diese Variablen haben potenziell Einfluss auf das Ergebnis nach der Handlung. Dann haben wir die Handlung, die in unserem Fall als eine binäre Variable kodiert wird. Eins bedeutet, dass die Handlung auf das Objekt durchgeführt wurde, 0 – das nicht durchgeführt wurde. Und wir haben ein Ergebnisvektor. Die Aufgabe wäre festzustellen, welchen Unterschied oder welche Differenz die Handlung durchschnittlich verursacht.

# Folie 3 Synth-Validation

Die Wissenschaftler haben sich mit diesem Problem schon seit langer Zeit beschäftigt. Es wurden unterschiedliche Lösungen und Variationen von Lösungssätzen entwickelt. Eine von der aktuellsten heißt Synth-Validation. Das, was sie macht, ist datenbasiert eine von vielen Methoden automatisch zu wählen. Grundsätzlich werde ich mich in meiner Bachelorarbeit damit beschäftigen, die Synth-Validation zu verbessern und erweitern.

# Folie 4 Beitrag

Nachdem das Paper über Synth-Validation veröffentlich wurde, sind einige neuen Machine Learning Methoden für kausale Inferenz entstanden. Diese neuen Methoden werde ich neben den alten implementieren und zur Synth-Validation hinzufügen. Die Methode wurde bis jetzt nur mit generierten Daten getestet. Für meine Evaluation werde ich daneben auch echte benutzen. Ein anderer Beitrag wäre, dass Synth-Validation auf Julia geschrieben wurde und ich werde sie auf R implementieren.

# Folie 5 Evaluation

Am Ende der Arbeit werde ich eine Evaluation von den Ergebnissen machen. Die unterschiedlichen Methoden werden mit der automatischen Auswahl von der Synth-Validation verglichen werden. Da der echte Treatmenteffekt von den Daten verfügbar ist, wird als Vergleichsmetrik die true treatment error, also die Differenz zwischen den echten und den geschätzten Treatmenteffekt benutzt. Wie schon angesprochen, die Daten die für die Evaluation benutzt werden, sind echt und nicht generiert. Sie kommen aus dem medizinischen Bereich.