

# Tamara Bigler

## (Preisträgerin in Betriebswirtschaftslehre)

- > Alter: 25
- > Nationalität: Schweizerin

### Ausbildung

- > 2016 – 18: Master of Science in Business Administration, Universität Bern – *summa cum laude*
- > 2012 – 16: Bachelor of Science in Business Administration (Nebenfach Mathematik), Universität Bern – *insigni cum laude*
- > 2007 – 11: Matura mit Schwerpunkt Psychologie, Pädagogik und Philosophie, Seeland Gymnasium Biel



# Tamara Bigler (Preisträgerin in Betriebswirtschaftslehre)

## Berufliche Tätigkeiten

- > Seit 2018: Doktorandin/Assistentin, Ordinariat für Quantitative Methoden, Universität Bern
- > 2016 – 18: Hilfsassistentin, Ordinariat für Quantitative Methoden, Universität Bern
- > 2016: Praktikum im Riskmanagement, Deutsche Bank, Frankfurt
- > 2015: Praktikum als Quantitative Research Analyst, Straits Invest, Singapur
- > 2014: Hilfsassistentin, Institut für mathematische Statistik und Versicherungslehre, Universität Bern

# Tamara Bigler (Preisträgerin in Betriebswirtschaftslehre)

## Angaben zur Masterarbeit

- > Titel: **Eine Matheuristik für das Feature Selection Problem in der multiplen linearen Regression**
  
- > Betreuer: Prof. Dr. Norbert Trautmann, Dr. Oliver Strub  
Departement Betriebswirtschaftslehre der Universität Bern

# Tamara Bigler (Preisträgerin in Betriebswirtschaftslehre)

## Motivation

- > Durch die stetige Digitalisierung gibt es eine immer bessere Datenverfügbarkeit und immer grössere Datenmengen, welche potenziell verarbeitet werden können.
- > Die Daten können genutzt werden, um Regressionsmodelle für unterschiedliche Anwendungen zu trainieren. Zum Beispiel können Häuserpreise geschätzt werden aufgrund verschiedener Eigenschaften wie Fläche, Anzahl Badezimmer, Vorhandensein eines Balkons, etc.
- > Die Modelle sollen trotzdem interpretierbar bleiben; es sollen also nicht alle unabhängigen Variablen, sondern nur eine begrenzte Anzahl miteinbezogen werden.
- > Durch die Begrenzung der Anzahl an unabhängigen Variablen im Regressionsmodell kann weiter dem Overfitting entgegengewirkt werden, bei dem gegebene Beobachtungen der Ausprägungen einer abhängigen Variable sehr gut, aber zusätzliche Beobachtungen nur sehr schlecht erklärt werden.

# Tamara Bigler (Preisträgerin in Betriebswirtschaftslehre)

## Thema

- > Feature Selection in der multiplen linearen Regression:
  - Gegeben:
    - Eine Menge von unabhängigen Variablen (UV)
    - Eine Anzahl von Beobachtungen einer abhängigen Variable
  - Gesucht:
    - Eine Teilmenge der UV sowie deren Beta-Koeffizienten, so dass die gegebenen Beobachtungen der abhängigen Variable bestmöglichst erklärt werden können
    - Um eine hohe Interpretierbarkeit zu gewährleisten und Overfitting zu reduzieren, soll die Teilmenge der UV eine gegebene maximale Kardinalität nicht überschreiten
  - Zielkriterium:
    - Summe der quadrierten Fehler zwischen den Beobachtungen der abhängigen Variablen und der geschätzten Beobachtungen der abhängigen Variablen durch die lineare Regression (quadrierte Residuen)
- > Ziele der Masterarbeit:
  - Entwicklung einer Heuristik zur Feature Selection in der linearen Regression, welche in kürzerer Rechenzeit als ein exaktes Modell eine gute Auswahl an UV findet.

# Tamara Bigler (Preisträgerin in Betriebswirtschaftslehre)

## Methodik

- > Entwicklung der Matheuristik basiert auf
  - einem exakten Modell aus der Literatur (vgl. Bertsimas und King 2016)
  - Nachbarschaftssuche mittels Local Branching (vgl. Fischetti und Lodi 2003)
- > Die Matheuristik verbessert iterativ die Auswahl der unabhängigen Variablen in der Regression, indem in jeder Iteration eine neue zufallsbasiert ausgewählte Teilmenge der unabhängigen Variablen potentiell in die lineare Regression aufgenommen werden kann
- > Es können in jeder Iteration nur wenige unabhängige Variablen aus der Regression ausgeschlossen und wieder in die Regression aufgenommen werden
- > Vergleich der erklärten Varianz (anhand des korrigierten  $R^2$ ) der abhängigen Variable mittels der linearen Regression der Matheuristik mit der des exakten Modells
- > Vergleich der beiden Ansätze erfolgt unter einer Zeitlimite
- > Vergleich erfolgt anhand von vier Datensets mit weniger als 100 unabhängigen Variablen sowie vier Datensets mit mehr als 1'000 unabhängigen Variablen

# Tamara Bigler (Preisträgerin in Betriebswirtschaftslehre)

## Resultate

- > Für Datensets mit weniger als 100 unabhängigen Variablen findet die Matheuristik die bestmögliche Auswahl an unabhängigen Variablen in vergleichbarer Rechenzeit zum exakten Modell
- > Für Datensets mit mehr als 1'000 unabhängigen Variablen findet die Matheuristik bei einer begrenzten Rechenzeit von 10'800 Sekunden (= 3 Stunden) eine bessere Auswahl an unabhängigen Variablen bezüglich des korrigierten  $R^2$  als das exakte Modell
- > Bei einer Veränderung der Anzahl auszuwählender unabhängiger Variablen findet die Matheuristik unverändert bessere Teilmengen bezüglich des korrigierten  $R^2$  als das exakte Modell während einer begrenzten Rechenzeit von 10'800 Sekunden

# Tamara Bigler (Preisträgerin in Betriebswirtschaftslehre)

## Kommentar (Prof. Dr. Norbert Trautmann)

- > Sehr innovative Anwendung des Operations Research im Bereich Data Science
- > Erste Matheuristik (Kombination von heuristischen Suchstrategien mit Modellen der mathematischen Programmierung) zur Lösung des Feature-Selection-Problems mit grossen Datenmengen