

**Predictive Analytics.  
Mehr erfahren aus Daten.  
Intelligent analysieren.  
Machen Sie Wahrscheinlichkeit  
berechenbar.**

# **Agglomerative Hierarchical Clustering - Use case in life insurance under Solvency II**

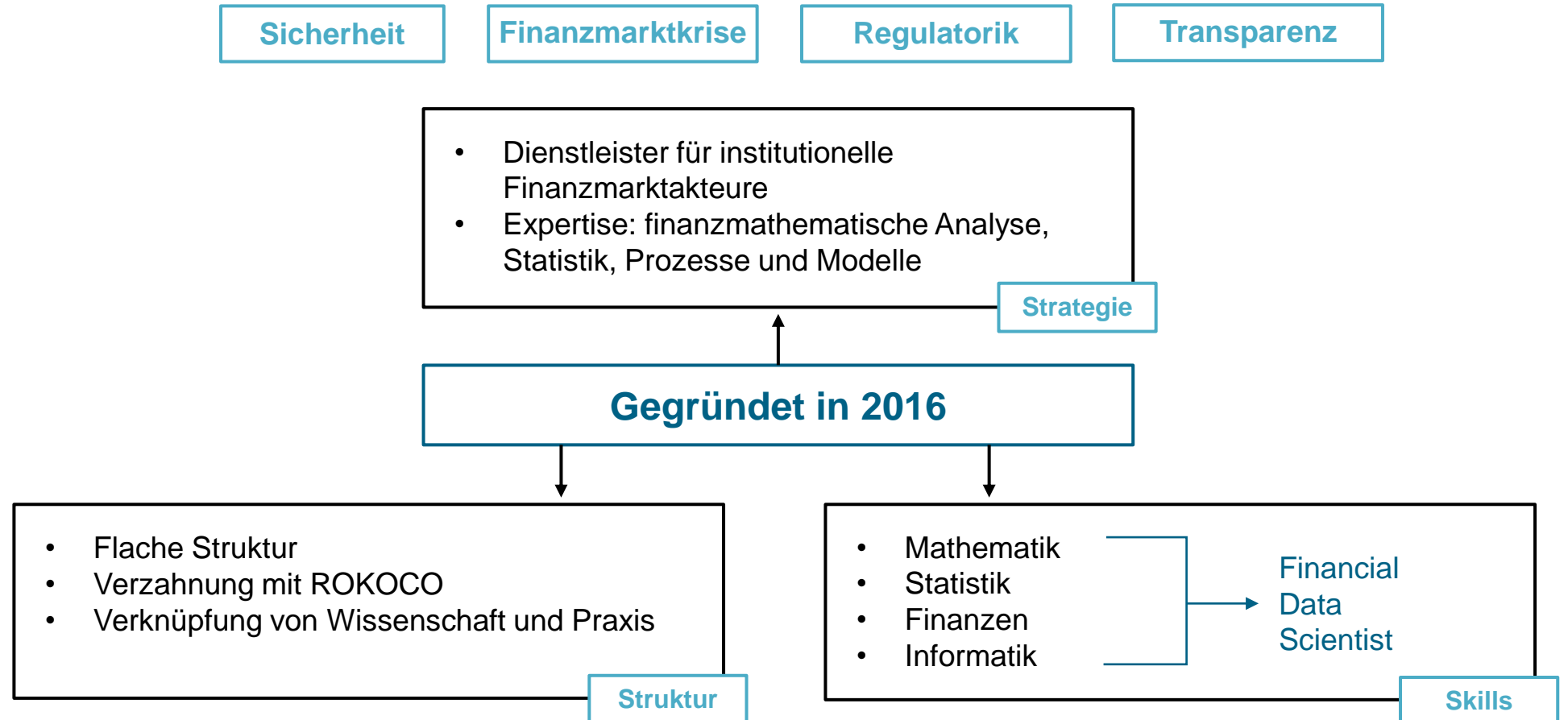
German Data Science Days 2022

14. Juli 2022

Sebastian Hantsch

# Wer wir sind und was wir tun ...

Tochtergesellschaft der ROKOCO GmbH (Aktuarielle Beratung)



# Wer wir sind und was wir tun ...

- Strategie
  - Dienstleister für institutionelle Finanzmarktakteure
  - Expertise: finanzmathematische Analyse, Statistik, Prozesse und Modelle
- Produkte & Services
  - Ökonomische Szenarien für regulatorische Zwecke
  - Real-World-Einschätzungen für Kapitalmarktsimulation
  - Künstliche Intelligenz & Machine Learning
  - Bilanzanalysen und -auswertungen
- Kunden & Mandanten
  - Fondsanbieter, Asset Manager, Versicherungen, Banken, Pensionseinrichtungen, Versorgungswerke
  - Deutschland, Österreich, Schweiz, Luxemburg

## Motivation (1)

- Im Zuge des europäischen Aufsichtsregimes Solvency II sind die Versicherer verpflichtet, eine Bewertung der Verpflichtungen gemäß Standardmodell oder internem Modell zu erstellen
- Hierzu wird in der Lebensversicherung typischerweise eine Hochrechnung der zukünftigen Unternehmensentwicklung vorgenommen
- Diese Hochrechnung berechnet für alle Versicherungsverträge im Bestand die in der Zukunft erwarteten Cashflows (Beiträge, Kosten, Versicherungsleistungen)
- Aufgrund der engen Kopplung der Aktivseite (Kapitalanlage) mit der Passivseite (Verpflichtungen aus Überschussbeteiligung) in der deutschen Lebensversicherung muss diese Projektion für jedes Kapitalmarktszenario erneut durchgeführt werden

## Motivation (2)

Größenabschätzung:

- Je nach Größe des Versicherungsunternehmens können im Bestand mehrere Millionen Einzelverträge vorliegen
- Typischerweise werden 1.000 – 10.000 verschiedene Kapitalmarktszenarien durchgerechnet

→ sehr lange Rechenzeiten!

## Was ist Bestandsverdichtung?

- Statt den Originalversicherungsbestand jeweils komplett in der Hochrechnung zu berechnen, kann stattdessen ein repräsentativer Teilbestand bestimmt werden, der bei entsprechender Gewichtung der gewählten Repräsentanten eine gute Näherung der erwarteten Cashflows liefert
- Hierzu wird eine Metrik definiert, die den numerischen Abstand zwischen zwei Versicherungsverträgen angibt, beispielsweise:
  - Numerische Attribute (z. B. Höhe des Beitrags, Höhe der Versicherungssumme) mittels euklidischer Metrik
  - Kategorische Attribute (z. B. Vertragsart, Geschlecht der versicherten Person) mittels 0/1-Metrik
- In der Regel werden numerische Attribute vorab auf den Werte-Bereich  $[0, 1]$  skaliert, um einer ungewünschten Gewichtung durch unterschiedliche Größenordnung von Attributen zu begegnen

# Was muss ein Verdichtungsalgorithmus leisten?

- Grundsätzlich ist die Anzahl der Arten, wie man  $k$  Repräsentanten aus einem Versicherungsbestand mit  $n$  Verträgen ziehen kann extrem groß („ $n$  über  $k$ “):

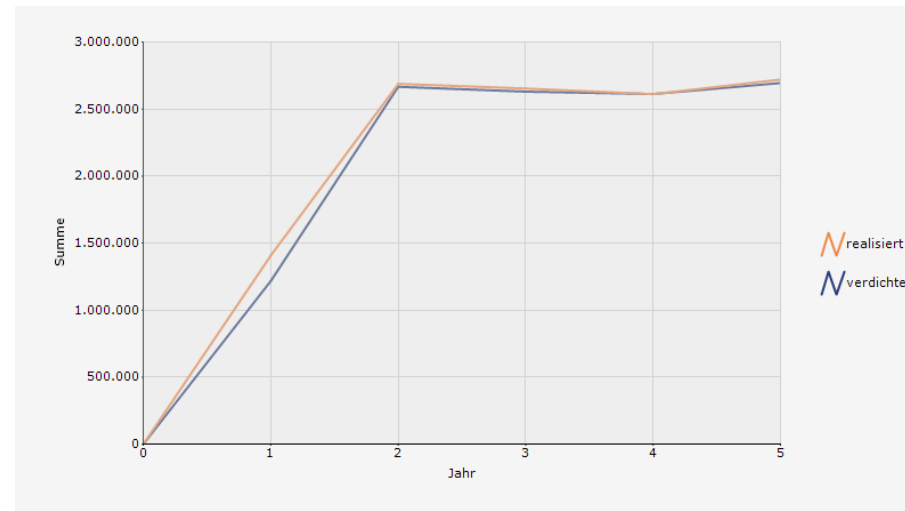
$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- Es gibt also bei einem Bestand mit 1.000.000 Verträgen  $\binom{1000000}{1000} \approx 1.5 \times 10^{3432}$  Möglichkeiten, um 1000 Repräsentanten zu ziehen
- Der Bestandsverdichtungsalgorithmus hat deshalb die Aufgabe, eine gute Auswahl der gewünschte Anzahl von Repräsentanten zu ermitteln, so dass bei entsprechender Gewichtung der Originalbestand bestmöglich getroffen wird, ohne den kompletten Lösungsraum zu durchsuchen



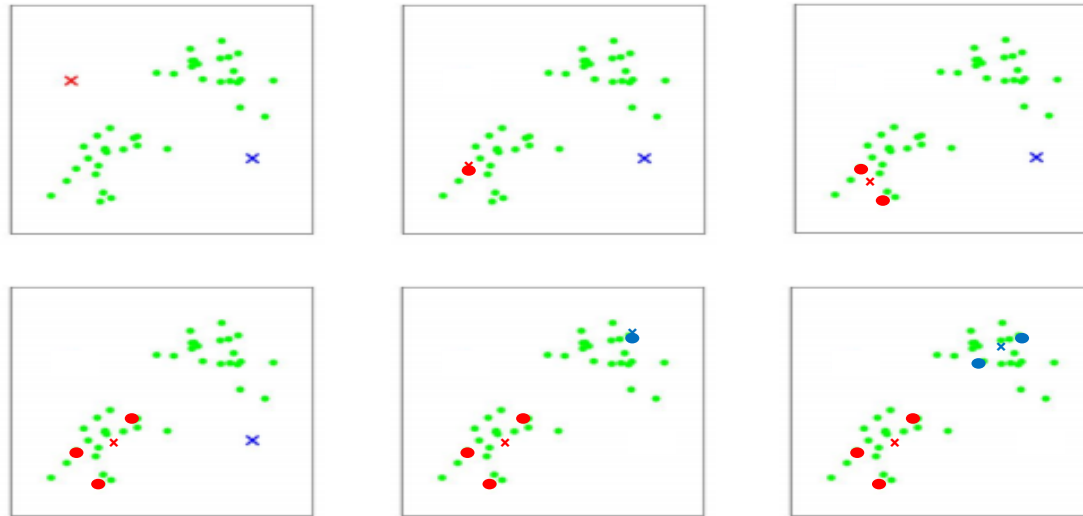
# ROKOCO Bestandsverdichtung

- Für den Kundenkreis der ROKOCO entwickelte Implementierung diverser Verdichtungsalgorithmen
- Einfache Definition der Metrik sowohl für numerische als auch kategorische Attribute
- Implementierung bewährter Algorithmen-Familien
- Ergebnisvisualisierung sowie Unterstützung zur Ermittlung der geeigneten Anzahl an Repräsentanten



# Clustering-Algorithmen

- kPrototypes
  - Wähle zufällig oder mit speziellem Verfahren k Objekte als Prototypen-Anfangslösung (Zentren der Cluster)
  - n Durchläufe oder bis sich die Zentren der Cluster nicht ändern:
    - Für jeden Datenpunkt:
      - Finde das nächste Zentrum
      - Weise den Datenpunkt dem Cluster mit dem nahen Zentrum zu
      - Update das Zentrum als Mittelwert der Datenpunkte im Cluster



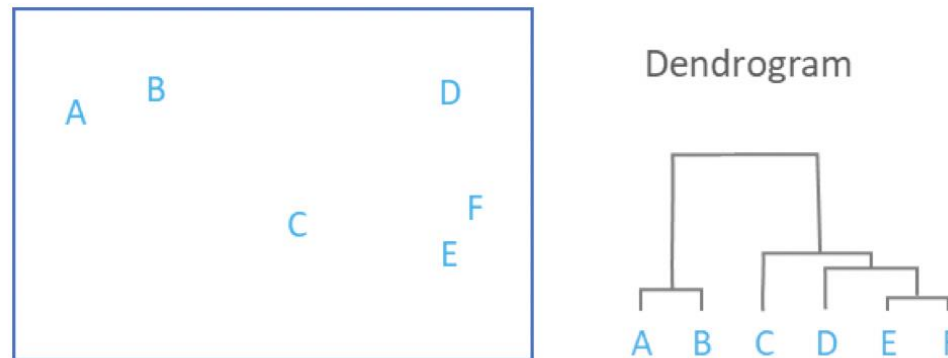
- Agglomeratives hierarchisches Clustering

# Agglomeratives Clusterverfahren

Algorithmus:

1. Zu Beginn werden für alle Verträge des Originalbestands jeweils ein separater Cluster erzeugt
2. Aus den vorhandenen Clustern werden jeweils zwei Cluster zu einem neuen Cluster verschmolzen, welche ein Verknüpfungskriterium minimieren
3. Fortsetzen bei Schritt 2, bis die gewünschte Anzahl an Clustern = Repräsentanten erreicht wurde

Grafische Darstellung im einfachen zweidimensionalen Fall mittels Dendrogramm:



## Verknüpfungskriterium: Ward`s method

- Minimierung der Varianzerhöhung bei Verschmelzung der Cluster:

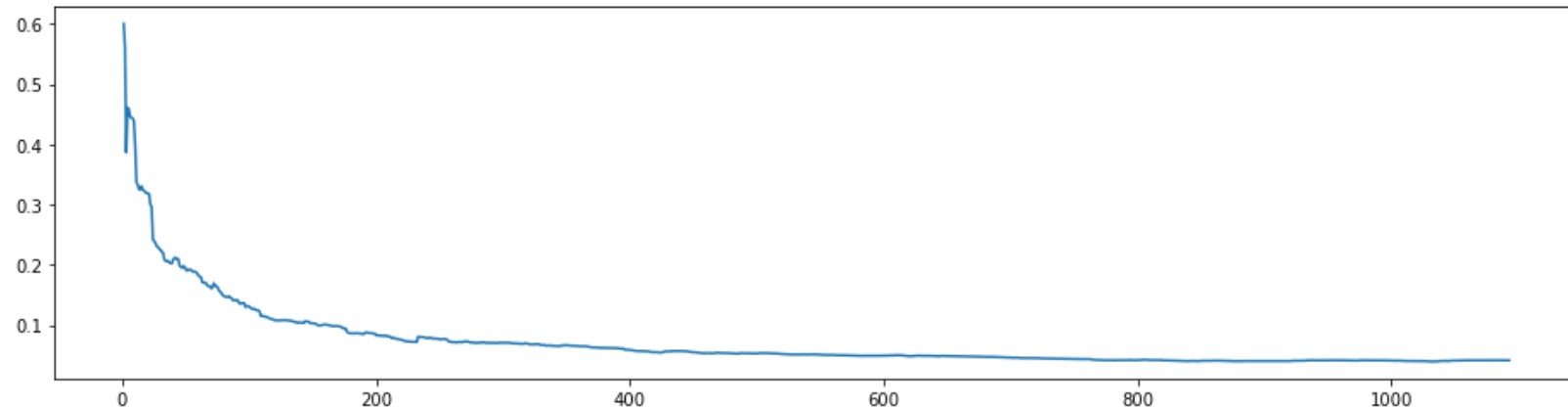
$$\begin{aligned}\Delta(A, B) &= \sum_{i \in A \cup B} \|\vec{x}_i - \vec{m}_{A \cup B}\|^2 - \sum_{i \in A} \|\vec{x}_i - \vec{m}_A\|^2 - \sum_{i \in B} \|\vec{x}_i - \vec{m}_B\|^2 \\ &= \frac{n_A n_B}{n_A + n_B} \|\vec{m}_A - \vec{m}_B\|^2\end{aligned}$$

- $\Delta(A, B)$  Varianzerhöhung durch Verschmelzung von Cluster A und B
- $\vec{m}_j$  Clusterzentrum von Cluster j
- $n_j$  Anzahl Verträge im Cluster

# Vor- und Nachteile

Vorteile:

- Deterministischer Algorithmus
- Anzahl Repräsentanten muss nicht vorab festgelegt werden, da die Zwischenschritte des Algorithmus einfach ausgewertet werden können



Nachteile:

- Hoher Arbeitsspeicherbedarf: Speicherkomplexität  $O(n^2)$

# Kontakt

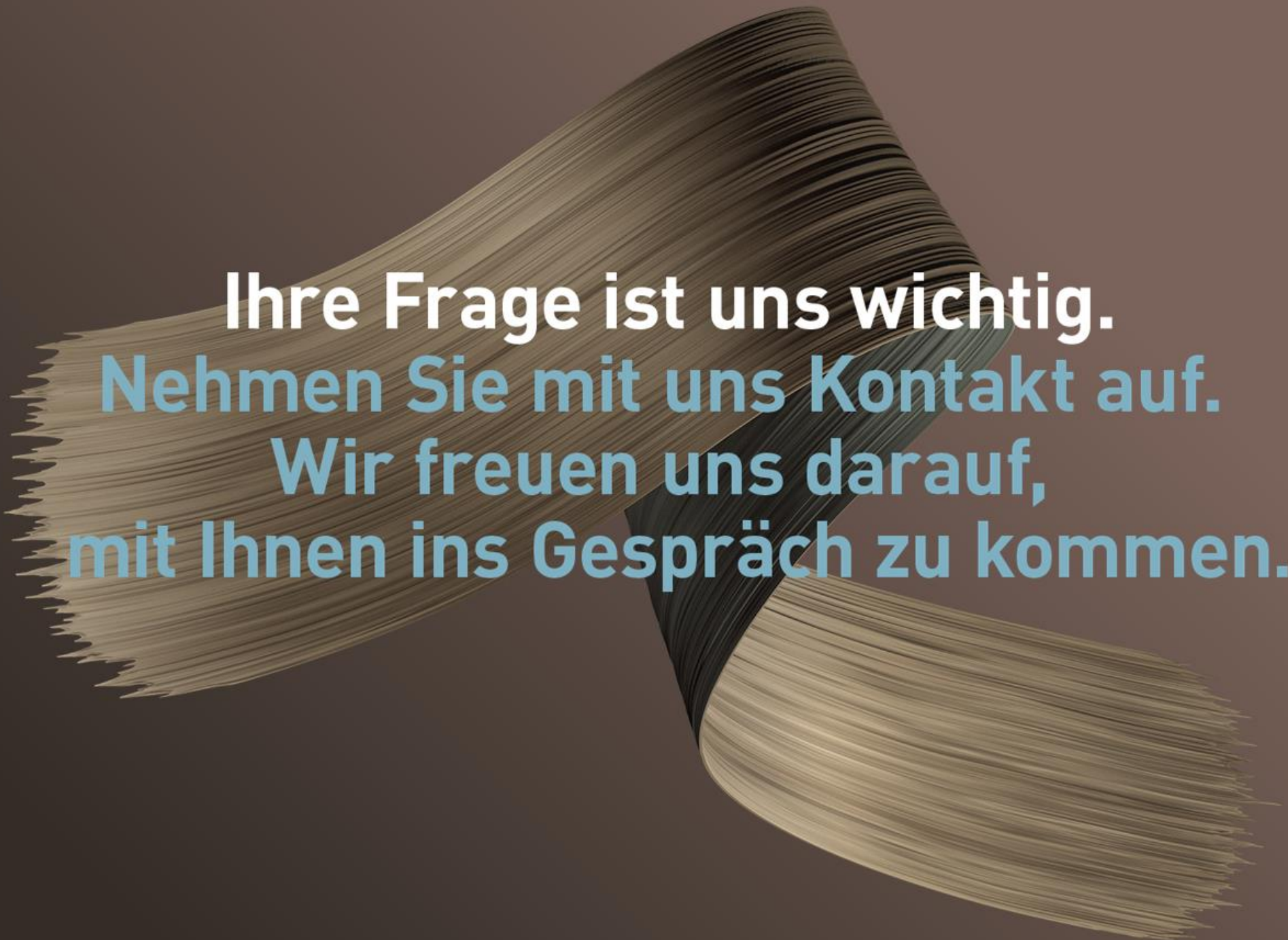
Aktuelle Stellenausschreibungen  
– auch für Praktikanten und Werkstudenten –  
unter **[www.rokoco.com](http://www.rokoco.com)**

**SEBASTIAN HANTSCH**  
Geschäftsführer

ROKOCO Predictive Analytics GmbH  
Ludwig-Ganghofer-Straße 6  
82031 Grünwald  
T. +49 89 6200 9357 25  
[anton.s.wittl@rokoco.com](mailto:anton.s.wittl@rokoco.com) | [www.rokoco.com](http://www.rokoco.com)

**DR. ANTON S. WITTL**  
Geschäftsführer

ROKOCO Predictive Analytics GmbH  
Ludwig-Ganghofer-Straße 6  
82031 Grünwald  
T. +49 89 6200 9357 25  
[anton.s.wittl@rokoco.com](mailto:anton.s.wittl@rokoco.com) | [www.rokoco.com](http://www.rokoco.com)



**Ihre Frage ist uns wichtig.  
Nehmen Sie mit uns Kontakt auf.  
Wir freuen uns darauf,  
mit Ihnen ins Gespräch zu kommen.**