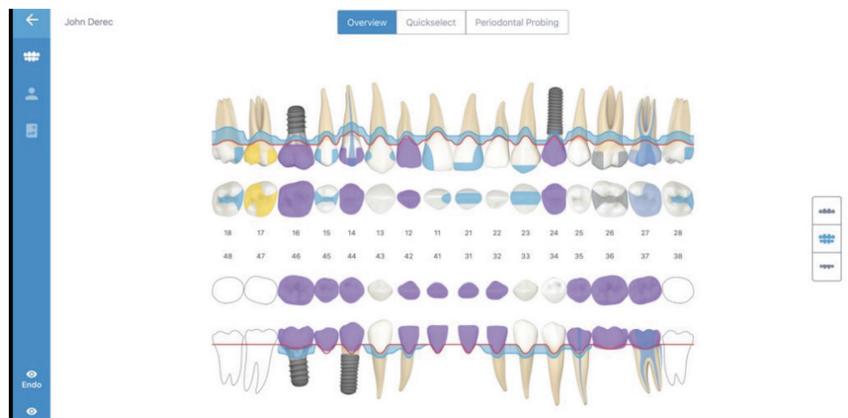


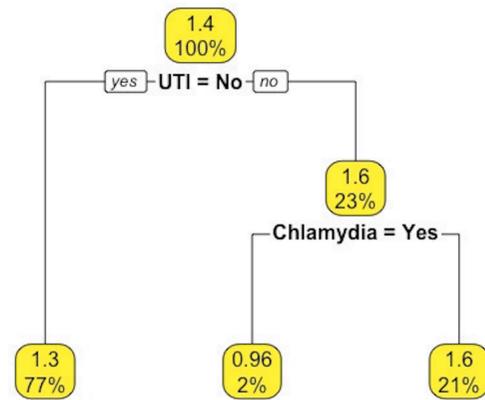
**Bachelor-Thesis Medizintechnik**

# Klinische Machine Learning Classifier für Big Data Pipe Line

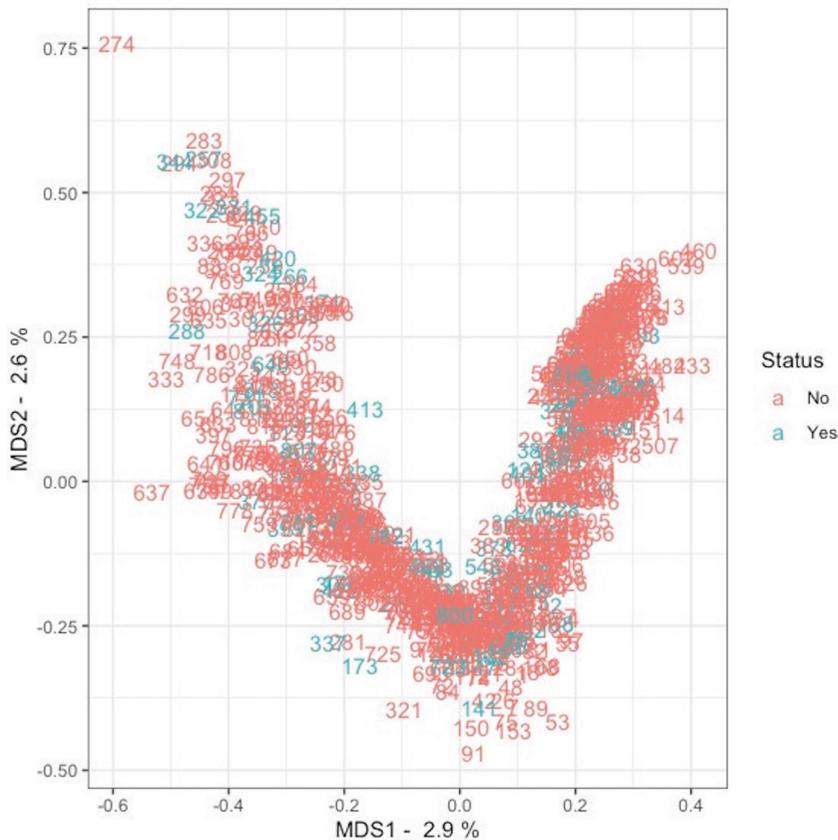
© Dental Charting App DERE



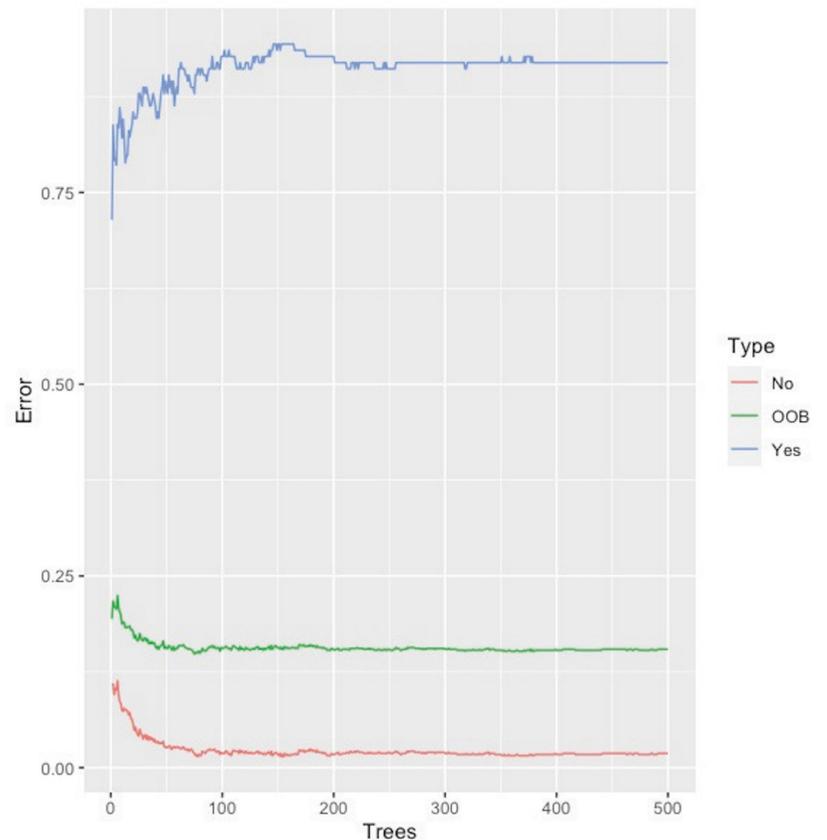
© Decision Tree



MDS plot using (1 - Random Forest Proximities)



© MDS Plot



© Plot der Fehler pro Anzahl Decision Trees

**Problemstellung**

Die Firma Nordend Global hat die Software DERE entwickelt, welche weltweit eine der einzigen Dental Charting Software ist, die es gibt. Dental Charting ist ein Prozess, bei dem Zahnärzte die Gesundheit der Zähne händisch auflisten und beschreiben. Für diese soll einen Machine Learning Classifier in Form von Decision Trees und Random Forests programmiert werden, damit in Zukunft basierend auf Patientendaten Therapievorschläge errechnet werden können.

Machine Learning ist ein Teilbereich der künstlichen Intelligenz, der Systeme in die Lage versetzt, automatisch aus Erfahrungen zu lernen und sich zu verbessern, ohne explizit programmiert zu sein. Dadurch kann das System vollautomatisch Wissen generieren, Algorithmen trainieren, identifiziert Zusammenhänge und erkennt unbekannte Muster.

**Lösungskonzept**

Die Daten einer Studie über „Treatment of Periodontal Disease and the Risk of Preterm Birth“ wurden verwendet, um einen Algorithmus zu

programmieren. Dafür mussten die Daten zunächst bereinigt werden, damit anschliessend Decision Trees und Random Forests programmiert werden konnten.

Ein Decision Tree ist ein mathematisches Diagramm, das baumartig und gerichtet ist. Es besteht aus einer Wurzel, Knoten, Ästen und Blättern. Die Knoten bilden Verzweigungspunkte, welche von einer Entscheidung abhängig sind. Sie führen über die Äste in die nächste Entscheidungsebene. Mit der Hilfe von Decision Trees lassen sich Entscheidungen finden.

Ein Random Forest generiert eine Vielzahl verschiedener Decision Trees und kombiniert sie anschliessend, um ein Gesamtergebnis zu erzielen. Diese werden dazu verwendet, um bestmögliche Entscheidungen oder Vorhersagen zu treffen. Decision Trees und Random Forests sind beides Verfahren von Nonsupervised Learning, welches ein Teilbereich von Machine Learning darstellt.

**Ergebnisse**

Während dem Programmieren der Decision Trees wurde schnell vermutet, dass die Daten im verwendeten Datensatz nur wenig korrelieren und deswegen ungeeignet für diese Art von Anwendung waren. Mit der Programmierung der Random Forests wurde diese Vermutung bestätigt.

**Dalia Küchler**

Betreuer:  
Fabian Ille

Experte:  
Philipp Stämpfli

Kooperationspartner:  
Daniel Bättig  
NordendGlobal AG, Feldmeilen