

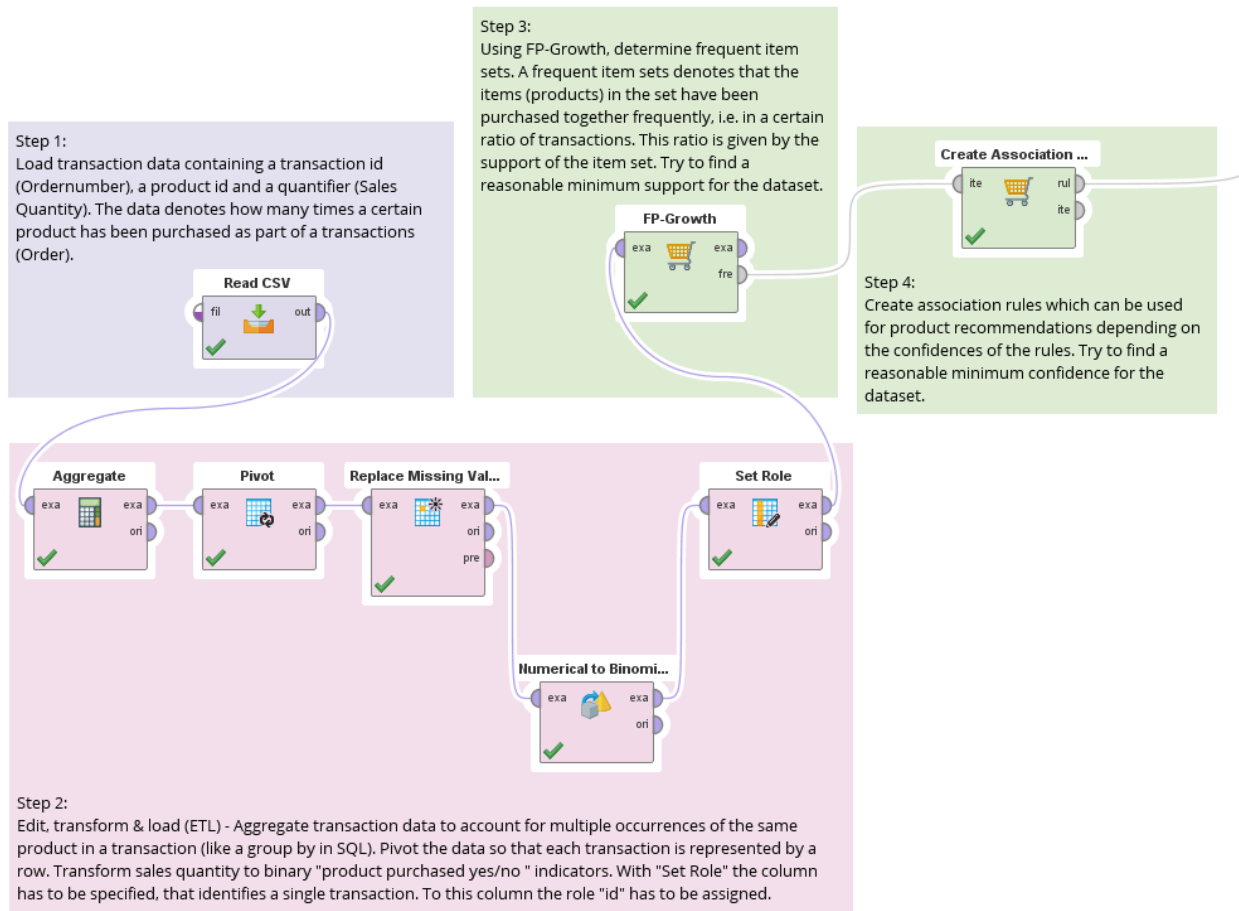
Übungsblatt 6
Stand: 06.12.2018

Bearbeiten Sie alle Aufgaben aller Übungsblätter – soweit nicht anders angegeben – in festen **Zweiergruppen**. Beide Gruppenmitglieder müssen die Aufgabe gemeinsam bearbeiten und die Lösung alleine erläutern können. **Die Lösungen dürfen nicht mit anderen Gruppen gemeinsam erarbeitet werden.**

Aufgabe 9 (Data Mining – Dirt Bikes)

Nach der erfolgreichen Erstellung eines Dashboards (Aufgabe 8) ist das Dirt-Bikes-Management Fan von Datenanalysen geworden. Der CEO hat mal was von Data Mining gehört. Daher bittet er Sie, ihm anhand eines einfachen Beispiels das Potenzial von Data Mining zu demonstrieren. Da Sie nicht so viel Zeit investieren wollen, haben Sie die schlaue Idee, einfach das Beispiel aus der Versicherungsbranche aus Ihrer Data Mining-Vorlesung nachzubauen.

- a. Suchen Sie sich eine Ihnen passend erscheinende Data Mining-Software/ Skriptsprache etc. aus. (Vielleicht ist ja der Magic Quadrant von Gartner hilfreich (<https://www.kdnuggets.com/2017/02/gartner-2017-mq-data-science-platforms-gainers-losers.html>)).
Hinweis: eine Software für Data Mining steht im Unternehmen (so wie auch in unseren Pools) nicht zur Verfügung.
- b. Verwenden Sie den Beispieldatensatz (*01_CarInsurance_example.xlsx*) und erstellen Sie einen Entscheidungsbaum, der das Attribut „Damage“ anhand der anderen Attribute beschreibt. Interpretieren Sie kurz textuell Ihren Entscheidungsbaum, und beschreiben Sie, wie ein Versicherungsunternehmen das Ergebnis in Ihren Prozessen verwenden könnte. Erstellen Sie eine Confusion Matrix um die Performance Ihres Klassifikators anhand der Accuracy zu bewerten.
- c. Der CEO ist (vermutlich) begeistert von Ihrem Entscheidungsbaum. Daher bittet er Sie, die Daten aus dem Vertriebsbereich nach interessanten Mustern zu untersuchen. Er ist hier insbesondere an einer Warenkorbanalyse interessiert für die Dirt Bikes-Produkte (ohne Zubehör/ Accessories). Die notwendigen Daten wurden schon aus dem OLTP-System für Sie extrahiert. (*warenkorb_data_ohne_acc.csv*). Führen Sie mit dem Datensatz eine Warenkorbanalyse durch, in dem Sie Assoziationsregeln ableiten, die beschreiben, welche Produkte häufig zusammen gekauft werden. Glücklicherweise hat ein Kollege schon mal hierfür eine Schulung gehabt. Er stellt Ihnen einen Musterprozess aus der Software RapidMiner zur Verfügung (siehe nächste Seite). Interpretieren Sie kurz textuell Ihr Ergebnis, und beschreiben Sie, wie Dirt Bikes das Ergebnis in Ihren Verkaufsbereich verwenden könnte.



- d. Sie wollen die Karriereleiter bei Dirt Bikes weiter hinaufklettern und erkennen Ihre Chance. Daher zeigen Sie Eigeninitiative und stellen dem CEO zusätzlich noch eine spannende Clusteranalyse auf Basis der Daten aus Ihrem Snowflake- bzw. Star-Schema oder aus den OLTP-Daten aus Übung 4.

Abzugeben Neben der Demonstration der Analysen im Rahmen des Praktikumstermins sind abzugeben in EMIL:

- Screenshots und/oder Konfigurationsdatei/ Skript für die Analysen aus b., c. und d.
- die Texte zur Interpretation der Ergebnisse aus b., c. und d.

Allgemeine Hinweise (gelten für alle Übungsblätter)

Ich akzeptiere nur Lösungen, auf denen **die Namen beider Bearbeiter, die Veranstaltungsbezeichnung, die Nummer des Übungsblattes und die Nummer der Praktikumsgruppe (1/2/3/4) notiert ist**. Außerdem müssen sie eine **Versionsnummer** auf dem Lösungsblatt notieren und die Versionen bitte bei erneuter Abgabe hochzählen.

Je Zweier-Gruppe ist in EMIL nur von einem Gruppenmitglied das Lösungsdokument hochzuladen. Die Abgabe der Lösungen muss spätestens am **Tag der Übung in EMIL** erfolgen (**Deadline: Ende des jeweiligen Praktikumstermins**).

Sind Lösungen von zwei/ mehreren Gruppen zu ähnlich, bekommen alle beteiligte Gruppen einen Fehlversuch und alle Gruppen müssen eine überarbeitete Version einreichen.