Database Semester Project

# Data Use Case

## Explain context & motivation

Wir wollen basierend auf Resultaten von ATP Turnieren die Wahrscheinlichkeit von Tennisspielen errechnen und darauf basierend auch von Turnieren eine Finalpartie mit einem Gewinner und der Wahrscheinlichkeit. Das Thema zum Tennis entstand bei der Suche nach einem Use Case zuerst wollten wir Vorhersagen treffen auf Sportresultate. Sport ist beliebt und es sollte genug Datensätze geben, die ersten Ideen waren Fussball und Formel 1. Diese Ideen wurden dann verworfen, weil es zu viele verschiedene Faktoren gibt welche Einfluss haben auf das Resultat. Beim Fussball hat es mindestens elf verschiedene Spieler pro Mannschaft und verschieden Aufstellungen und Umstellungen in einer Partie, welche nicht immer gleich sein muss. Bei der Formel 1 war es für uns nicht ganz klar wie wir die Leistung von einem Fahrer klar unterscheiden wollen zu der Leistung der Autos, weil man Daten brauchen wird von einem Fahrer als er noch unter einem anderen Rennstall fuhr. Also brauchten wir einen Sport, bei dem die Leistung von nur einem Spieler und seinem Gegner im Resultat widergespiegelt wird. Die Einflussfaktoren sollten gut verständlich und trennbar sein.

Tennis ist ein beliebter internationaler Sport welcher viele Datensammelt und diese sind auch zugänglich. In den Einzel-Partien sieht man die Leistung von dem Gewinner und von Verlierer, die Einflüsse sind die Oberfläche des Spielfeldes und die Form des Spielers. Es gibt natürlich noch weitere Faktoren wie Links-Rechtshänder, die Grösse des Spielers, die Stärke des Spielers in Form seiner Spielweise, diese verschiedenen Faktoren haben wir grob beschrieben mit der Historischen Leistung des Spielers gegen den Gegenspieler.

## Define use case & data-based decision support

Man kann mit der Wahrscheinlichkeit von einer Tennispartie erhalten durch das Eingeben von zwei Spielern und der Unterlage, zusätzlich kann mit einer Datums Eingabe auch die Wahrscheinlichkeit einer nicht existenten Partie in der Vergangenheit ausgegeben werden. Basierend auf diesem Prinzip können auch Turniere und auch nicht existente Turniere «Simuliert» werden, am Ende soll die Finalpartie und deren Gewinner ausgegeben werden.

Die Nutzer einer solchen Funktion sind Wettbüros und Fantasy League, um Vorhersagen zu treffen oder um nicht existente Turniere zu gestalten. Wettbüros können durch die Wahrscheinlichkeiten eine geeignete Preispolitik gestalten.

## Select & justify data sources

Die Datenquelle, welche verwendet wird, ist tennis-data.co.uk und die Datasets wurden vom User Tyler Woekenberg erstellt. Das Dataset von den Spielern hat einen ID, die Namen und auch das Geburtsdatum. Die Match Daten sind gut haben viele Informationen wie die Art des Spielfeldes aber auch die Punkte des Gewinners und des Verlierers. Diese Punkte Angabe können wir verwenden als einen Faktor für die Form Berechnung.

## Select & justify database technology

Wir werden als Technologie MySQL Workbench verwenden, es wird im Unterricht unteranderem behandelt. Die Entscheidung für SQL entgegen NoSQL lag in dem Dataset, welches wir brauchen für dieses Projekt. Die Matchdaten müssen noch aufgeteilt werden in verschiedene Tabellen, aber die Daten sind strukturiert und auch neue Datensätze folgen dem gleichen Schema.

Die Daten wurden in MySQL Workbench migriert durch den Insert Befehl, dabei wurden die Matchdaten verwenden um die Tabellen von der Serie, Pitch, Tournament aufzufüllen. Die Spielerdaten kommen von der Datenquelle atp\_player und man musste Sonderzeichen aussortieren.

## Build the system based on a clear system architecture