**Introduction:**
Malfunction or rejection of transplanted kidneys cause a high financial burden for the German healthcare system. A key reason for transplant rejection is therapeutic non-adherence, other reasons are not fully elucidated yet. Big data analytics and machine learning are technologies with a tremendous potential to reveal (unexpected) patterns and risk factors that are relevant for long-term transplant survival.

**System Architecture:**

**Costs caused by patients after KTx (Charité)**

<table>
<thead>
<tr>
<th>Year</th>
<th>Costs per Year</th>
<th>Number of Cases</th>
<th>Costs per Case per Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>1.913.955€</td>
<td>1.835</td>
<td>1.043,03€</td>
</tr>
<tr>
<td>2016</td>
<td>1.863.104€</td>
<td>1.822</td>
<td>1.022,56€</td>
</tr>
<tr>
<td>2017</td>
<td>3.447.812€</td>
<td>1.766</td>
<td>1.952,33€</td>
</tr>
</tbody>
</table>

**Cleansing:**

Python library to ease data wrangling, cleansing, and EDA

Support for handling missing values:
- Listwise deletion;
- Mean/ median/ mode imputation
- Linear regression

**Goals:** Reduction of re-hospitalizations as well as reduction of graft loss for kidney transplant patients

**Adherence:**
- The degree to which a patient correctly follows therapies and medical advices
- **Non-adherence** is one of the main reasons for unwanted re-hospitalizations and loss of kidney function
- Monitoring adherence has a high impact for the treatment of patients in terms of outcomes and thus cost reductions

**Real-Time Monitoring:**

**Data Source:**
- Producer/consumer software design
- Updates of data can trigger generation of messages (events) in real time
- All important events are available in the moment when needed

**Risk Prediction:**
- **Task:** Predicting the probability that an endpoint occurs in near future
- **Targets:** Re-hospitalization, mortality, infections, graft loss, ...
- **Data (retrospective):** 15 years, 8k patients
- **Features:** ~300 (demographic, vital, lab,...)
- **Method:** Gradient Boosted Regression Trees
- **Train/Dev/Test:** 80/10/10, 20-fold x-val
- **Preliminary Results (mortality):**
  - ROC: 91.10 (stddev: 2.57)
  - Prec/Rec F1: 86.72, 56.06, 68.10

**Dashboard:**
- Visualization of patients according to adherence
- **Data:** Real-time APP data, TBase
- **Method:** Rule-based adherence ranking based on data from MACSS platform

**Complex Event Processing:**
- Real-time stream processing for detection of sequence of events
- Matching of patterns, generated by experts
- Irrelevant data is discarded, allowing the process potentially infinite data streams

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780185.