Outcome Prediction

Group E

Team Members:

Haseebullah Shaikh (Team Lead)

Humza Asim

Hammad Irshad

Humza Moeen

Huzaifa Waseem

Outcome Predictions (ETL life cycle)

- Data of each table will be extracted from API by specifying particular path for each table.
    - Six data sources (six tables) can be used in this service for outcome predictions which includes appointment, patient councilor, price log, availability, councilor, and rating.
- Extracted data will be transformed in Apache Spark.
    - For successful rate four tables will be merged appointment, patient councilor, councilor and ratings, unnecessary columns will be dropped such as description, notes, user id etc.
    - Few queries will be implemented on merged table
        - Confirmed treatments will be retrieved
        - New column will be added which will specify outcome of treatment (successful or not successful) based on ratings.
        - Series of successful treatments will be counted for each patient.
        - Successful rate will be measured for each councilor.
    - For Time spent merged table will be added with a new column which will be the multiplication of appointment duration and number of appointments group by each patient id.
    - For money spent active payment records of price log table will be merged with previous merged table for calculating average amount spent on each councilor.
    - Final transformed data (prediction results table) will include four columns councilor id, successful rate, time spent and money spent.
- Transformed data (Predictions) will be loaded (stored) in database.
- Outcome Prediction API that will be exposed for predicting successful rate, time and money spent, API request must include councilor id the prediction response will be returned based on that id.