Data Science Methodology to determine and predict Customer Satisfaction in a Hospital

I have chosen Hospitals and customers satisfaction to apply data science methodology.

The revenue of a Hospital has been recently highly decreased. Decrease in Hospital satisfaction gives us the sign on the reasons of the decrease in the number of customers(patients) that formerly getting health care service from this hospital.   
core question is;  What are the reasons of the customer dissatisfaction in our hospital and how can we improve the satisfaction levels and improve our number of customers and overall benefit?  
In this level determination of our goal and objectives require close coordination and direct communication with the directors of the hospital and related subject matter experts to define the problem in a right manner. Involve the key sponsors into the understanding of the problem.  
I will gather the necessary data and execute analytic methodology, use necessary tools and create a model to answer the question in a data driven way. Finally we will reach better care, and customer satisfaction that will results in the rise of our revenue also.

1. Analytic Approach:

I will use descriptive models to determine the current status of the level of customer satisfaction, reasons in the background and direct reasons and relationships. Quantitative and qualitative methods will be in parallel used.

To determine the reasons of dissatisfaction and the decrease in the number of customers i will use statistical analysis on the data. Diagnostics to express what happened and why will be applied.

To make forecasting on the future of this trend I will conduct predictive tools and try to explain why will happen next.

I hope to find a prescriptive solution to increase the overall customer satisfaction and the revenue of the hospital as a follow on. Machine learning for relationships and classification will be one step and Decision Tree will be applied to predict future potential of the customers.

2.Data Requirements

I need the historical and current behaviors of the patience and their feedbacks and their reasons and comments showing their satisfaction (partially related with their expectations) and the causes of dissatisfaction and their assessment on the hospital service quality etc. I determine the ingredients at this level. Sources will be social media, Internet comments on hospital services and other open feedback sources, besides mainly from questionnaires and surveys done to patients, caregiver notes etc.

In this level i select cohort, to represent the patients in general as the patients who are customers of the hospital for the last five years and who live in the district. Dissatisfied customers will be in the cohort.

I will determine the content, formats, representations suitable for decision tree in this step. There will be one record per patient with columns representing variables. My content will try to cover all aspects of dissatisfaction reasons and hospital service history and feedbacks.

3. Data Collection

In this step I will gather data from external (social media, GOV Sites/rankings), avatar surveys, call center notes, patient complaints, caregiver notes (mostly un-structured), also questionnaires and cohort group surveys and interviews. Internal; data including structured information on patients,inpatient record system, billing info and service feedbacks will be in the examined data.

Descriptive statistics and visualization will help me to have initial insight over the content and the quality of my data. Gaps will be identified. I will do substitutions to defer redundant and inconsistent data. I will get rid of redundant data and merge some necessary meaningful data in this step.

4. Data Understanding and Preparation

I will encompass all activities related to constructing my data set in this step. Understanding the relationships, distributions I will use pairwise correlations and histograms to the data. In this manner I will decide what sort of data preparation may be needed to make variables more useful in my future model. Univariate statistics and other means also will help me to assess data quality. Missing values, misleading values will be determined. If needed I will redefine the satisfaction term and scales and also loop back to data collection to add new parameters.

In the sense of data preparation, I will deal with missing/invalid values, remove duplicates and ensure proper format.

In this phase I will aggregate transactional records.

5. Modeling and Evaluation

To determine relationships and current status of satisfaction and realted issues I will use Descriptive analytics model.

For the next steps that this hospital will take care of to improve the quality and predict customer/patient satisfaction predictive model and analytics will be in use. Training set that is formed from historical data will be used like a calibration gauge.

Decision Tree Model and examining the Type I errors( false positive) and Type II errors ( false negative) will be expressed and determined for our models.

In the evaluation level I will use statistical significance to ensure the data properly handled and interpreted within the model.

To evaluate decision tree models I will use ROC curve, True Positive (Sensitivity) rates vs False Positive Rates, try to balance the miscalculation costs. I will choose the optimal model.