



**C . A . R . E**

Cost • Analysis • Regressive • Evaluation

## **IS483 Project Proposal**

Optimizing Value of cancer care using Artificial  
Intelligence to predict individualized Cost-Outcomes

**Team CARE**  
**Version Number 1.0.1**  
**As of 15 Nov 2019**

## Team Members:

1. **Choy Yu Min Justin** ([justin.choy.2017@sis.smu.edu.sg](mailto:justin.choy.2017@sis.smu.edu.sg)):  
Project Manager / Solution Designer
2. **Chua Ming Yu** ([mingyu.chua.2017@sis.smu.edu.sg](mailto:mingyu.chua.2017@sis.smu.edu.sg)):  
Full Stack Developer / Solution Designer
3. **Guo Lingxing** ([lxguo.2017@sis.smu.edu.sg](mailto:lxguo.2017@sis.smu.edu.sg)):  
Full Stack Developer / Solution Designer
4. **Tanny Lai** ([tanny.lai.2017@sis.smu.edu.sg](mailto:tanny.lai.2017@sis.smu.edu.sg)):  
Graphic & UiUx Designer / Frontend Developer
5. **Tan Jia Yun** ([jiayun.tan.2017@sis.smu.edu.sg](mailto:jiayun.tan.2017@sis.smu.edu.sg)):  
Secretary / UiUx Designer / Frontend Developer
6. **Chang Jesslyn** ([jesslync.2017@sis.smu.edu.sg](mailto:jesslync.2017@sis.smu.edu.sg)):  
Secretary / UiUx Designer / Frontend Developer

## Faculty Supervisor:

- Name<sup>1</sup>

## Sponsor:

Organization: National Cancer Centre Singapore (NCCS)

Dr Wong Fuh Yong ([wong.fuh.yong@singhealth.com.sg](mailto:wong.fuh.yong@singhealth.com.sg))

- Senior Consultant Radiation Oncologist, Division of Radiation Oncology, NCCS
- Head, Department of Cancer Informatics, NCCS
- Chief Medical Informatics Officer, NCCS

---

<sup>1</sup> To be assigned by course coordinator

## Project Overview

### 1.1 Project Description:

**Summary of the goal: What are you building {brief, about 3 sentences}? You may include the X factor. For example, building an e-commerce site selling books using a variant of K-nearest neighbor data mining technique to recommend books. Be SPECIFIC. Don't say you are going to write a game! What game? It is good to do product comparison but put the details in the IS480 wiki.**

The goal of this project will be to develop an application to predict direct healthcare expenditure for cancer patients. Using historic clinical and financial data, we will build a model using machine learning to perform personalized predictions for future healthcare expenses for individual cancer patients. This model will be used to make predictions of the cumulative costs of sequential treatment over time and choice of treatment methods. Following this, an application will be built to visualize these predictions in an easily understandable way for financial planners to be able to better explain expected treatment costs of cancer to patients. This Visualizer also functions as a communication tool to aid patient engagement in their shared care.

### 1.2 Motivation:

**Why? What is the reason for doing this? This section could be merged with project description. Do not repeat the description; do not use vague adjectives (best, user friendly, commercial quality, scalable, interactive, one stop, efficient, improve productivity, etc.). Each of these adjectives must be quantified, otherwise, you will lose credibility. Do not claim everything, it does not solve all problems.**

National healthcare spending in Singapore has almost tripled from \$3.74 billion in 2010 to \$9.8 billion in 2016 (2.2% of GDP). This is projected to reach \$13 billion in 2020 (Basu, 2017). This unsustainable rate of increase is especially evident in cancer care. Major advancements in cancer diagnostics and treatment have led to significant improvements in clinical outcomes. However, the cost of therapy, including chemotherapy, targeted agents, and more recently, immunotherapy and proton therapy, is substantial and have contributed to some of the largest and most rapid increase in healthcare spending.

In seeking to control these unsustainable increases in healthcare costs, it is imperative that healthcare organizations can predict the likely future costs of individuals, so that management resources can be efficiently targeted to those individuals at highest risk of incurring significant costs. For the patient and their family, having advance knowledge of expected healthcare expenses allows them to prepare financially and emotionally and to make rational healthcare decisions to maximize value.

### 1.3 Stakeholders:

Stakeholder	Description
<b>Sponsor</b>	<p><b>Who initiated the project? Be specific about any relationship between the sponsor and the team. What is your contact person's role in the organization? Is it a faculty member, CEO, manager, liaison, etc.</b></p> <p>This project was initiated by <b>National Cancer Center Singapore (NCCS)</b>. NCCS is Singapore's foremost centre for the prevention, diagnosis, and treatment of cancer as well as cancer research. Our main person of contact is Dr Wong Fuh Yong, Senior Consultant Radiation Oncologist of NCCS Oncology.</p>
<b>User (Financial counsellors)</b>	<p><b>Is the user different from the sponsor? Who is your user? If it is not a person you know, describe the user persona for each user role (e.g. admin, buyer, seller). Projects with no sponsors require a listing of the targeted users for beta testing</b></p> <p>The primary group of users will be the financial counsellors, who will be using our application in order to better explain the immediate and long-term cost of cancer treatment to patients with meaningful visualisation &amp; predictive information.</p>
<b>User (Doctors &amp; Patients)</b>	<p>The secondary group of users will be oncologists and their patients who will use the applications in clinical consultations as an information tool to enhance patients' knowledge and to engage patients in shared care and medical decision making.</p> <p>Decision making is motivated by Value Drive Care where the choice of treatment strategies serves both to maximize health outcomes that matters to the patient at a cost that is rational and sustainable from the societal perspective.</p>

### 1.4 Deliverables:

**Outcomes: What is delivered (deployed) to the sponsor? Is it a proof of concept? Or release for live usage? Value Statement: What does the sponsor get out of this?**

An application that is able to make individualized cancer treatment costs based on a set of characteristics of the patient. These predicted costs and relevant clinical outcomes (e.g. survival) will then be displayed in a lay-person friendly form. The value to be gained by NCCS, are improvements in patient satisfaction with the communication process regarding diagnosis and treatment selection. This application will be deployed for financial counselling for specific procedures, and short and long term overall expenditure. It will also be deployed in the clinics for medical consultations with doctors, allied health personnel and medical social workers.

## 1.5 Scope:

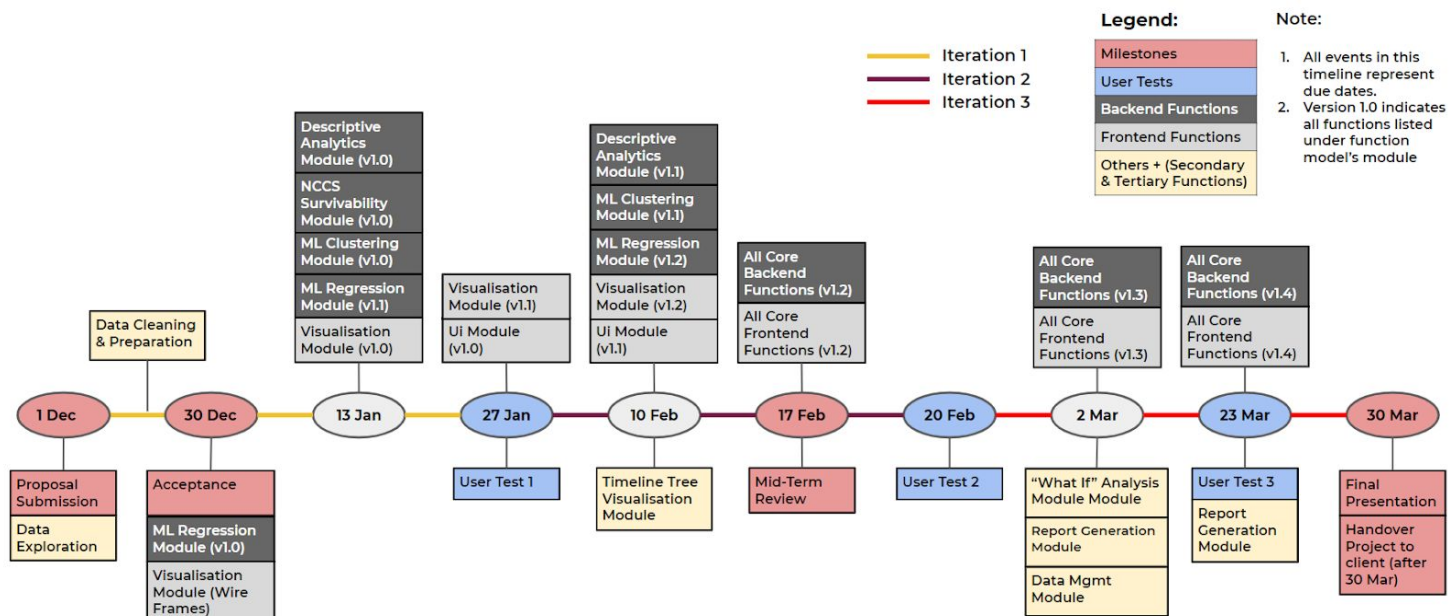
The main objective of this project is to empower the financial decision making for both doctors and patients using Machine Learning. The core functions of this project focus on achieving this, as well as providing the end users a lay-person friendly visualisation to understand the prediction results. Secondary and Tertiary functions are those that are of value, but will only be developed if time allows. For this scope, we assume that the cost of healthcare for the sample data will be reasonably close to that of the next 3-5 years into the future that we are predicting. However, we also acknowledge that the features given in the current data set might not be enough to give a reasonable regression prediction of the cost and other clinical outcomes - having compared the data set to similar studies. Nonetheless, in this study we will also assume first that the features available are sufficient for doing the prediction, and we will build our models based on these key features. The other functions listed here are only support functions that allow the program to be used more easily by the intended stakeholders.

*Note: Our Clustering module as listed in our functional model (below) seeks to use unsupervised learning to cluster models before feeding these clusters to be trained individually by the main Machine Learning Regression Model. It is not a clustering based predictor of its own. However, if the clustering results turn out to be meaningful, we will consider using the clusters to perform descriptive analytics per cluster. Even so, this model will not directly be used to do predictive analytics.*

<b>Core (Primary) functions</b>	
<b>Machine Learning Regression Module</b>	<ul style="list-style-type: none"> <li>• Trainable model based off suitable Machine Learning/AI models.</li> <li>• Predict patient treatment cost (1,2,5,10 years)</li> <li>• Predict clinical outcomes (5,10 years)</li> </ul>
<b>Machine Learning Clustering module</b>	<ul style="list-style-type: none"> <li>• Clustering/bagging module</li> <li>• Cluster patients to improve regression accuracy</li> </ul>
<b>Descriptive Analytics Module</b>	<ul style="list-style-type: none"> <li>• Generate summary statistics of selected patient groups</li> <li>• Generate summary statistics of general population</li> </ul>
<b>NCCS Survival module</b>	<ul style="list-style-type: none"> <li>• Calculate <b>survival</b> of patients with/without treatment</li> </ul>
<b>Visualisation module</b>	<ul style="list-style-type: none"> <li>• Generate visualisations based on Regression, survival, &amp; Descriptive Analytics module output</li> </ul>
<b>User Interface (Ui) module</b>	<ul style="list-style-type: none"> <li>• Home page</li> <li>• Patient cost &amp; Clinical outcome predictor page               <ul style="list-style-type: none"> <li>◦ 1, 2, 5, 10 year views (cost).</li> <li>◦ 5, 10 year views(clinical)</li> </ul> </li> </ul>
<b>Secondary functions</b>	
<b>“What-if” Analysis Module</b>	<ul style="list-style-type: none"> <li>• Predict costs &amp; clinical outcomes based on “what-if” patient conditions</li> </ul>
<b>Timeline Tree Visualisation Module</b>	<ul style="list-style-type: none"> <li>• Generate a timeline node based descriptive analytics based on each timeframe, represented by a node each.</li> <li>• Generate visualisation of this descriptive node model</li> <li>• Integrated with Ui Module</li> </ul>
<b>Report Generation Module</b>	<ul style="list-style-type: none"> <li>• Generate viewable report for saving/printing</li> </ul>
<b>NCCS value calculator module</b>	<ul style="list-style-type: none"> <li>• Calculate the relevant non-cost value indicators that are of interest to NCCS.</li> </ul>
<b>Tertiary functions</b>	
<b>Data Management Module</b>	<ul style="list-style-type: none"> <li>• CSV import new patient data for model training</li> <li>• Basic Create, Read, Update, Delete (CRUD) functions</li> </ul>

## Project Plan

We adopted a **staged waterfall** approach for this project. Modules have been allocated to be completed in respective iterations and there will be additional User Testings with NCCS on top of IS480's milestones. There are a total of 3 iterations: Iteration 1 is set to cover the data exploration, and building of some of the core functions of the solution; Iteration 2 is set to cover the Frontend modules and deliver the full minimum viable solution with several improvements and bug fixes from iteration 1; Lastly, Iteration 3 is set to cover enhancements to both front and backend parts of the solution, and most of the secondary functions if the project goes smoothly and the team possesses the spare capacity to do so. The first two user tests are set as feedback points, to gather the inputs from NCCS staff to better improve the solution in later iterations. User Test 3 however, will be structured more for the purpose of getting the overall feedback on the effectiveness of the project and solution. The general project timeline and its key dates are as shown below.



**Risks:**

No.	Risk	Risk Level	Mitigation Plan
1	Ensuring security of sensitive data	High	<p>1. Store data on encrypted SSD &amp; all work will be done locally on our work stations</p> <p>2. No usage of external programmes that requires of our data</p> <p>3. All workstations to be access controlled.</p>
2	<p>Data Risk - As our data is complex and pulled across a large time frame these are some of the concerns we have identified:</p> <p><b>1. Billing Data is likely to have high dimensions</b> as many records have fields of custom short text typed by staff</p> <p><b>2. Possibility of missing records</b> for cancer care expenditures before 2005 as SingHealth might not have fully went digital in their patient records yet.</p> <p><b>3. Possibility of missing records</b> for cancer care expenditures in private institution or other public institutions (non-SingHealth cluster) are not available for analysis.*</p> <p><b>4. Difficulties with attributing cost expenditure appropriately to direct cancer care.</b></p>	High	<p>1. We will work closely with NCCS data scientists and clinical specialists to do careful feature engineering and data cleaning</p> <p>2. Avoid using data from time spans with limited/inconsistent records. If necessary, we will work closely with NCCS to draft oversampling/undersampling strategies.</p> <p>3. If dimensionality of medication/treatment is too high due to similar medication/treatment i.e. (PANADOL EXTRA 20S vs PANADOL TAB 500MG) we consider working with NCCS clinical specialists to create mappings that compress dimensionality</p>
3	Due to data sensitivity and NDA, we are not be able to leverage on GPU servers for model training. Our hardware might lack the computational power to handle the large dataset.	High	Split dataset into smaller sample for training. Go through only 1 sample at every training to reduce the chance of GPU crashing.
4	Difficulty in explainability of models used. For some of our Machine Learning (e.g. unsupervised clustering), the resulting model may not be easily interpretable, thus limiting the usefulness of our model	Low	We will explore the use of several industry methods such as Feature importance, DeepLIFT, and Layer wise propagation, to reduce this risk.



**Resources:**

Resource	Description	
<b>Programming languages, frameworks &amp; Libraries</b>	<ol style="list-style-type: none"> <li>1. Python 3.7</li> <li>2. Node.js</li> <li>3. React</li> <li>4. Angular</li> <li>5. SQL</li> </ol>	<ol style="list-style-type: none"> <li>6. Python d3.js</li> <li>7. HTML</li> <li>8. CSS</li> <li>9. PHP</li> <li>10. Scikit Learn</li> </ol>
<b>OS:</b>	Windows	
<b>Hardware</b>	<ol style="list-style-type: none"> <li>1. Individual computers, GPU</li> <li>2. Encrypted SSD as main repository for NCCS data</li> </ol>	
<b>Platform</b>	Jupyter Notebook Visual Studio Code	
<b>NCCS Key Contributors</b>	Dr Wong Fuh Yong, Senior Consultant, Ms Justine Tan Siew Wee, Asst Director Clinic Operations, Ms Gladys Liao Chunxia, Senior Executive, Clinic Operations, Ms Esther Yoh Peck San, Assistant Manager, Finance, Ms Grace Lim Li Ping, Manager Clinic Operations	

## References:

- Azuaje, F. Artificial intelligence for precision oncology: beyond patient stratification. *npj Precision Onc* 3, 6 (2019) doi:10.1038/s41698-019-0078-1
- Ferroni, P., Zanzotto, F., Riondino, S., Scarpato, N., Guadagni, F., & Roselli, M. (2019). Breast Cancer Prognosis Using a Machine Learning Approach. *Cancers*, 11(3), 328. MDPI AG. Retrieved from <http://dx.doi.org/10.3390/cancers11030328>
- Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, Dimitrios I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Computational and Structural Biotechnology Journal, Volume 13, 2015, Pages 8-17, ISSN 2001-0370, <https://doi.org/10.1016/j.csbj.2014.11.005>.
- Zhao, M., Tang, Y., Kim, H., & Hasegawa, K. (2018). Machine Learning With K-Means Dimensional Reduction for Predicting Survival Outcomes in Patients With Breast Cancer. *Cancer Informatics*. <https://doi.org/10.1177/1176935118810215>
- Suk Lee, Eunbin Ju, Suk Woo Choi, Hyungju Lee, Jang Bo Shim, Kyung Hwan Chang, Kwang Hyeon Kim and Chul Yong Kim (November 27th 2018). Prediction of Cancer Patient Outcomes Based on Artificial Intelligence, Artificial Intelligence - Scope and Limitations, Dinesh G. Harkut, IntechOpen, DOI: 10.5772/intechopen.81872. Available from: <https://www.intechopen.com/books/artificial-intelligence-scope-and-limitations/prediction-of-cancer-patient-outcomes-based-on-artificial-intelligence>