Machine Reasoning Project Report

Smart Secondary School Proposer

Team Members

SUDALAIANDI RAJA SUDALAIMUTHU

JAYARAMAN REVATHI

JAYASRI RAGHUNATHAN

SUNIL VARGHESE

1 EXECUTIVE SUMMARY

Singapore has a strong education system. Singapore students aim high and they achieve very good results. This is recognised around the world. Singapore has good schools, with capable school leaders and teachers, and the facilities that are amongst the best in the world.

After their primary education, Singapore students face heavy competition to get admission to the top secondary schools. To manage the secondary school admission process, Ministry of Education has a process wherein every year students are admitted in schools primarily based on their PSLE (Primary School Leaving Examination) score. The other popular criteria used by students to select secondary schools are the Co-Curricular Activities (CCA) offered by the schools and the distance of the school from their home.

Our team of 4 Singapore residents have also gone through the selection process of selecting secondary schools for our children and the biggest challenge we found was to collect the disparate information from different sources and most importantly how to analyse this data to arrive at the most appropriate schools for our children.

We believe that many other Singapore parents would also be finding similar difficulties in analysing the schools available, so to alleviate their pain we decided to embark on this project to provide a smart secondary school proposer which would analyse and propose the best schools for their children based on their choices. For this project we have decided to use the school data available in public domain like MOE website and Kiasu Parents to name a few.

By using the knowledge and techniques imparted in our machine reasoning course, we first built a database of schools and their attributes from the MOE website. We used KIE tools JBPM and Drools to build the business processes and the decision rules. We also used Java and Google APIs to get real time information on the distances from home.

We had a tough but interesting time working on this project. It also allowed each one of us to understand the intricacies involved in choosing the right school for our children based on well-reasoned factors. We also felt that the scale and scope of the project was much bigger than what we initially thought and more time was required to explore all options.

2 BUSINESS PROBLEM BACKGROUND:

In Singapore selection of a secondary school for a child after PSLE is considered an important decision, as this would be having a major impact on the child's future for the next 4 to 6 years or beyond. Most of the parents would want to select a school which is a perfect fit for their child. However, there are a number of factors to be considered, for instance the PSLE score, distance of the school from home, CCAs offered by the school to name a few. With so many factors it becomes increasingly difficult for parents to choose a school for their child especially if the child has a high score in PSLE and has more options.

We believe that most parents take the option of choosing a school with the highest COP (Cut Off point) disregarding other factors like distance from school, which could really affect the child later on while studying in the school.

3 PROJECT OBJECTIVE

Having defined the business problem our group's aim was:

1. To come up with a school proposer which can identify the most suitable schools based on various factors such as the school's cut off point, the distance between home and school, the CCA interest of the child etc.

2. The school proposer would guide the parents to spell out their preferences and based on that our algorithm would recommend for them the schools which are the best fit for their child. The recommended schools would be shown as a ranked list.

4 PROJECT SOLUTION

The first step in solutioning was to come up with a knowledge model. Knowledge modelling is a two-step process:

1. Knowledge Acquisition

2. Knowledge Representation

4.1 KNOWLEDGE ACQUISITION

Knowledge acquisition sets up the sources from where we would get the relevant data to build our intelligent system. The sources are enumerated in the below table:

S/N	Source of Information	Information elicited	Acquisition Technique used
1	Public Domain – MOE website, Kiasu Parents Portal	The bulk of information is provided from these sources e.g.- - The secondary schools - Cut-off scores of each school - Location - CCAs available in school etc	- Web Scraping to get the data from MOE website - Google API to calculate distance between school and home
2	Expert – Parents of Secondary school students	From Parents we collected - Important considerations used when selecting a school	Focused discussions to elicit the tacit knowledge

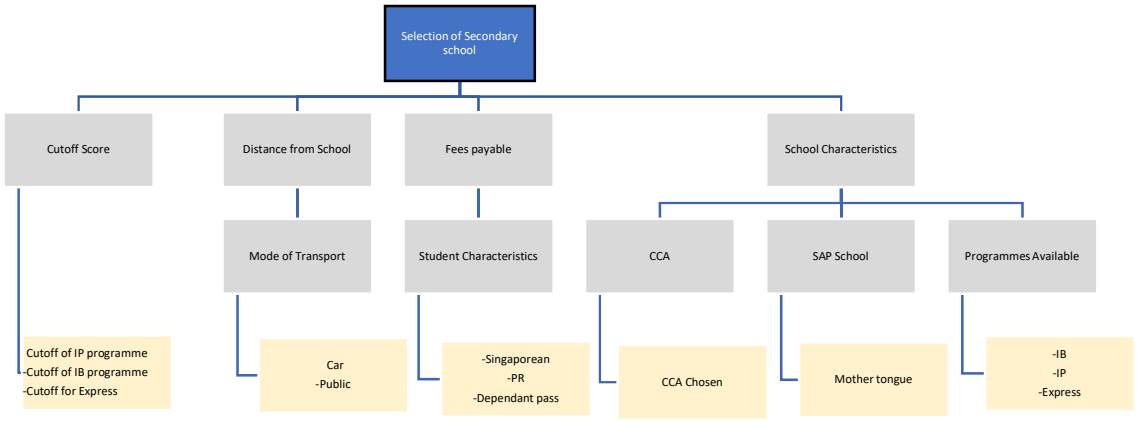
4.2 KNOWLEDGE REPRESENTATION

To formalize the knowledge that is collected during the acquisition stage, information needs to be adequately represented in a defined structure. Since we were going to recommend a ranked list of schools, we needed to associate certainty factors to each of the factors so that we can do scoring for them. Certainty factors are the measures of belief or the confidence that we associate with a particular factor. A list of scoring factors, and the associated certainty factors are provided in Annex 1.

The most appropriate representation we found was the inference diagram for our data.

4.3 INFERENCE DIAGRAM

The Inference diagram arranges the factors affecting student's chances of securing a place in a Secondary school in a hierarchical tree structure. The top most level node represents the decision of the proposed system, which in this case, recommends a group of schools to the user. This decision can be broken down into multiple layers of inferable sub-goals or subfactors before arriving at a list of "observables". These "observables" are gathered from users of the proposed system and they represent their inherent preferences. The table below illustrates an example using the inference diagram in Figure 2. The inferable sub-goals together with the "observables" are in fact, derived from the experiences and insights gained as a parent.

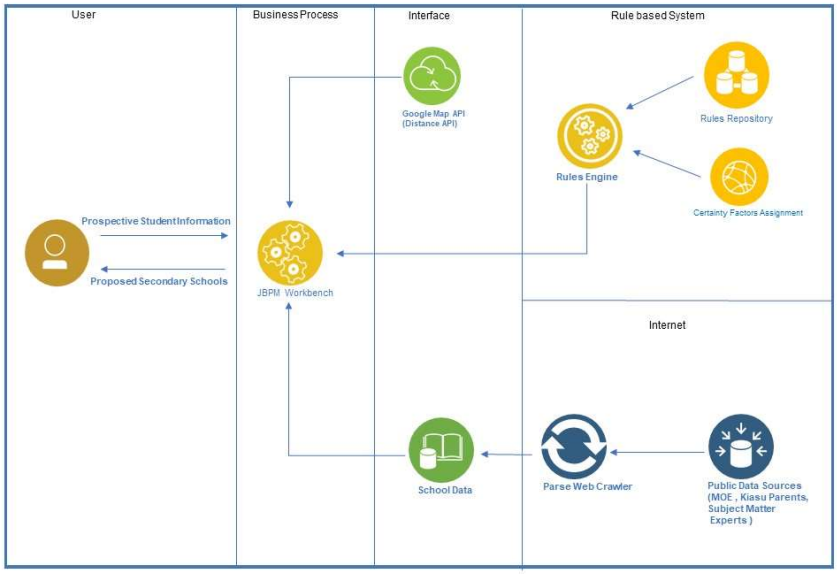


Legend: Top level Inference | Sub goals | Observables

S/N	Category	Information
1	Goal/top level inference.	A student looking for a secondary school
2	Sub-level goal	The student described in [1] would likely be interested in a school which certain school characteristics
3	Sub-level goal	SAP school is one of the characteristics which a school will have.
4	Observables	Mother Tongue will be the criteria via which the student can select a SAP school.

4.4 SYSTEM ARCHITECTURE

The system architecture diagram, illustrates how the application in the front-end has been interfaced with the back-end rule-based system.



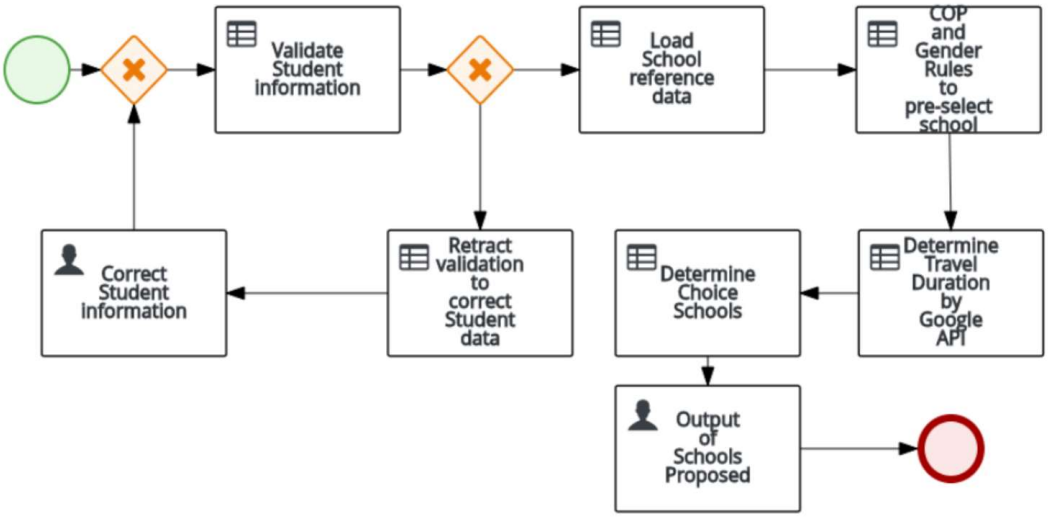
5 PROJECT SCOPE

While data mining can be performed continuously, in the context of this project, its scope is limited by the (i) School information as in March 2019 (ii) date and time of data extraction as well as (iii) the amount of data that is available at the time of extraction. The system provides a snapshot of the MOE schools information at the time of extraction. However, the team believes that the list does not change often so the results would be pretty accurate.

6 BUSINESS PROCESS DESIGN

The Business Process is shown in the diagram below.

School Data was crawled from MOE and Kisau Parents' websites using parseHub and Java class to Load all the school details was integrated with the KIE workbench. The distance between the school Postal code and the student's postal code is calculated by calling the Google distance Matrix API. The BPMN business process together with drools rules are used in KIE Workbench tool.



7 SYSTEM'S FEATURES

Despite the limited scope and assumptions, the team went through an in-depth thought process to implement significant features in the Proposer system which can substantially add value to potential students.

SYSTEM'S INTELLIGENCE & ROBUSTNESS

The School Proposer system is intelligent as it infers the users' preferences based on their input answer to the questions. They cover a range of factors such as availability of CCA, Mother Tongue, Cut off Points, Gender and Travel Time. The system uses Google APIs to calculate the distance dynamically based on the routes so that the travel time calculated are very accurate.

SCALABILITY

The system is developed to benefit users of the present as well as the future. The system is built with the capacity to scale to include any new schools as long as data is publicly available. This is because the team has adopted a web scrapping technique to retrieve information from the MOE website. All it needs to do is to increase the frequency of web scrapping.

8 LIMITATIONS

The entire decision process in selecting a school is complex as there are still umpteen number of factors to be considered. Despite the experience we had as parents in selecting schools for our children, it is inevitable that the system will fail to address the needs of each and every student. Due to shortage of time certainty factors were also not fully integrated into the system and may need more time to get this done. Also, only express students were considered in this round. It is certainly possible to consider them during further enhancement of the system to provide a more comprehensive result.

9 CONCLUSION

Our team had a great time working on this project, and we definitely picked up some useful skills along the way. Knowledge gathering was a crucial part of the entire process. Without a sound knowledge base taught in the lectures, we wouldn't have been able to build on system based on all the different rules. Building the system itself presented a whole new set of learning points. We got to apply practical knowledge of the KIE Server, as well as tap on our team's existing expertise in Java. Working on the exercise together allowed everyone to learn some or the other skills from one another. Overall, it was a tough initiation into the MTech programme but for which we are definitely not complaining.

10 IMPROVEMENTS

If we had a longer time frame to work on this project, we would have worked upon the following points of improvement:

1) Web Based GUI : We would have loved to include a Web based GUI, so that the look and feel could have been better.

2) Certainty factors: Though we wanted to include them from the very start however the time frame did not allow us to do that for all the factors. To improve the scoring model, we would like to include them.

3) Validations: Since the timeframe was short, some of the validations might have been loose. We feel that some of the validations can be tighter and could be done better in the next iteration.

Annex1

Scoring Model.xlsx