FINAL PROJECT REPORT

INTELLIGENT CHILD CARE MATCHER AND WAITLIST OPTIMIZER

TEAM MEMBERS

ANG PAU HUANG, EDWIN
CHEOK MEI LI (MAVE)
PADMAPRIYA MATHIVANAN
PRIYANSH MISHRA

TABLE OF CONTENTS

1 INTRODUCTION

1.1 EXECUTIVE SUMMARY

Choosing who to entrust your little one is an important decision. If you are putting your child into childcare for the first time, it may be a stressful experience since you want what is best for your child, but you don't know where to start. Your child's early experiences will have a lasting impact on their future habits and resilience. This is why choosing a quality childcare centre is essential.

There are various aspects in choosing the best option. This includes determining the type of childcare service i.e., full day or half day, finding out the childcare's philosophy and values, their teaching methods, their location, what food are fed to the children, and last but certainly not least, how they ensure the child's safety. We parents spend endless hours scouring the Internet for information, fearful of making the wrong decision.

Our team comprises of four members, amongst which two are new parents. We had faced these challenges when choosing the best childcare for our little one. Hence, to simplify this decision process, we created a recommendation system for parents to key in their preferences and get recommendations in one click without the need to surf through different childcare websites and various forums to get the information that we need.

During our childcare search process, we noticed that things aren't easy for childcare centres as well. They must call us to check on our interest after we last placed our child on wait list, typically multiple times. Next, they would arrange a time for us to go down to the childcare and meet the principal and teacher(s). Lastly, they call again to check if we would like to take up the vacancy. A childcare principal informed us that the situation had worsen since COVID-19, as they need to find a time to ensure the prospective parents do not meet the children in the childcare, limiting the time the parents can come down to the childcare.

In view of these difficulties, we came up with a system that will benefit both the parents and childcare centres. A free parent-facing recommendation system will help the parents view a list of suitable childcares within minutes (complete with reviews!). When parents select suitable childcares, their options will flow to the childcare-facing childcare matching system, into the childcare wait list. As a service to the childcares, the waitlist will include predictions on how likely the parents will take up a vacancy so that the childcare centres can optimise their resources spent on each parent. In return, the childcare will pay us a nominal fee that will allow us to provide free services to the parents.

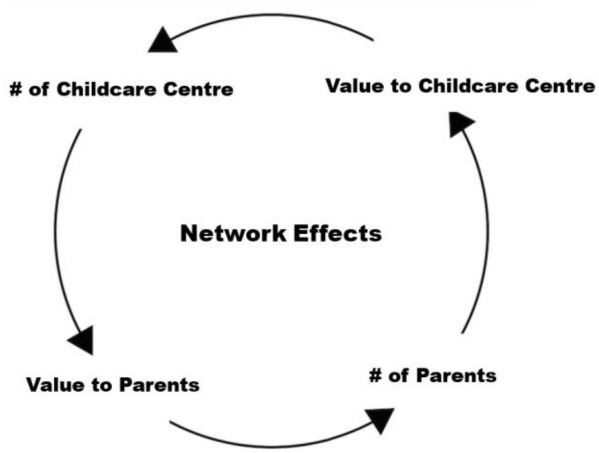


Figure 1: More System usage lead to network effects

As parents ourselves, we do not want to short-change parents and would not recommend childcares to skip parents who are deems less likely to take up a vacancy. On the other hand, childcares may wish to spend less resources on parents who are less likely to accept a vacancy (e.g., if parents do not pick up calls, how many more calls before continuing to the next parent). This in turn benefits parents who are more likely to accept vacancies.

As we build this system, we contacted a childcare centre on whether they will take up our product, to which they had given a positive response. We applied knowledge from lectures and hands-on exercises to build this system. For instance, we incorporated KIE Drools for building the parent-facing recommendation system and used rules, genetic algorithm and decision trees to provide predictions for the childcare-facing childcare matcher.

This COVID-19 pandemic has not made it easy for us to collaborate as we did not see each other in person during project development period. However, we had regular online meetings and helped each other through the challenges of working on this project. Overall, we are very grateful for this project experience, and this report is a log of our learnings.

1.2 BUSINESS PROBLEM BACKGROUND

Problems faced by parents

During the search for a suitable childcare centre for their child, parents in Singapore will often identify a list of potential childcares based on factors like distance and fees. Parents do also often conduct a lot of online research on past parents' reviews on potential childcare centres. It can be challenging for a parent to be able to identify a childcare centre that meet his/her preferences.

Figure 2: Parents asking reviews on Facebook[1]

[1] https://www.facebook.com/groups/541425542723577/

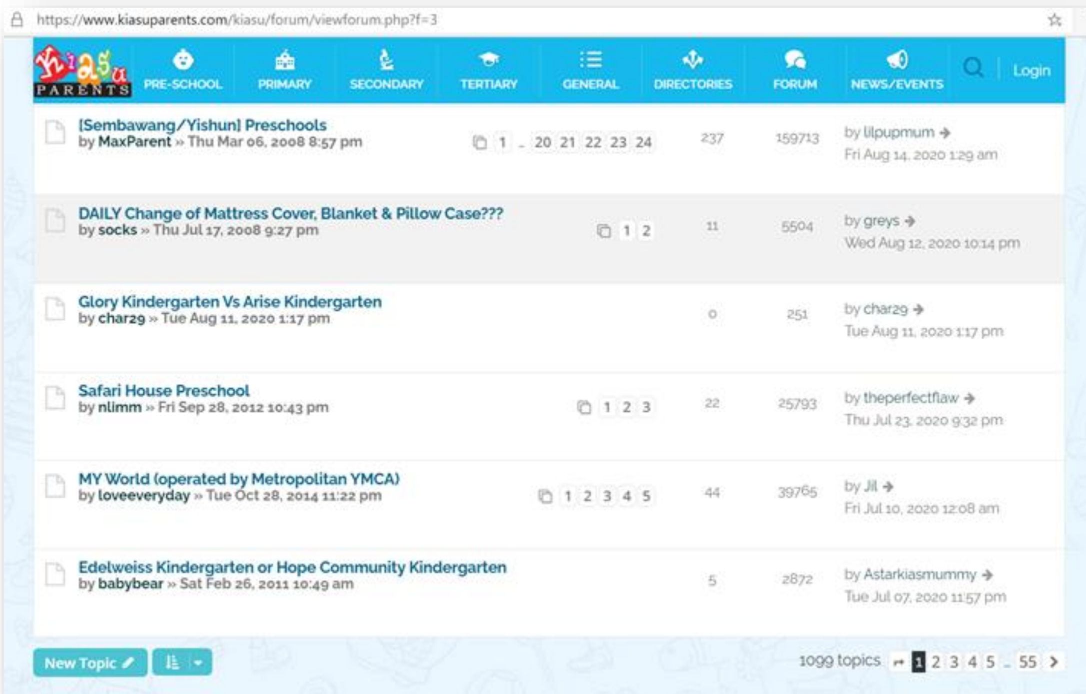


Figure 3: Parents asking for reviews on KiasuParents[2]

<u>Problems faced by childcare centres</u>

Separately, a recurring problem among childcare centre operators today is the need to manage an inflated waitlist of prospective parents which does not reflect true demand[3]. One reason is parents who had secured childcare places in a centre would normally not remove their names on waitlist of other centres. Even if Early Childhood Development Agency (ECDA), the regulatory and developmental agency for the early childhood sector, is able to coordinate the waitlists to reduce inflation issues, the problem remains that some parents may already have found alternative childcare methods, like domestic helpers and did not bother to withdraw their applications. The resources spent in managing waitlists are high, which includes contacting and planning for parent's visits to childcare centres.

This is evident from a feedback from a pre-school principal. She mentioned that parents take time getting back to the preschools on the parents' interest, even when they are interested.

---

[2] https://www.kiasuparents.com/kiasu/forum/viewforum.php?f=3

[3] https://www.asiaone.com/singapore/why-childcare-supply-and-demand-dont-add

Furthermore, challenges of managing a waitlist has worsened since the COVID pandemic, as pre-schools need to either conduct virtual tours or can only allow parents to enter the pre-school when the children are not around, such as in the evenings or Saturdays.

Existing solutions

It is also important for us to understand what are the existing solutions are for parents to register interest (in childcare centre). The following solutions do not help parents quickly glean reviews on the childcares nor help childcares differentiate parents' interests (besides by interest registration date).

1. Registering through ECDA website[4]. This website allows parents to register up to 10 childcares, without allowing parents to rank the childcares. It provides limited information on childcare centres such as levels offered by the childcare, the fees etc.

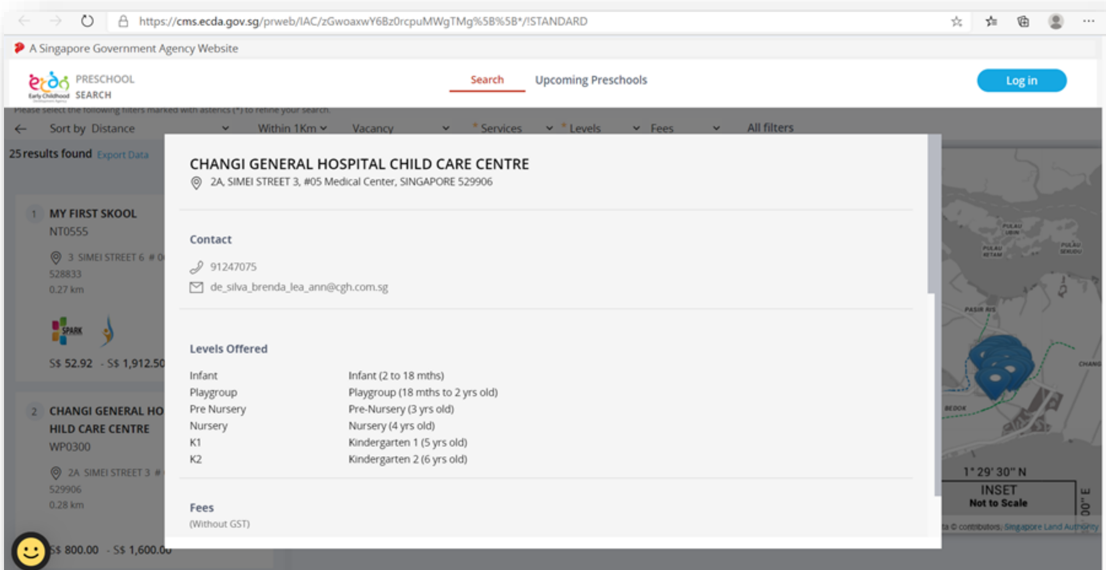


Figure 4: Different levels offered by childcare centre

[4] https://www.ecda.gov.sg/pages/aboutus.aspx

2. Alternatively, parents can go straight to the childcare centres to register interest.

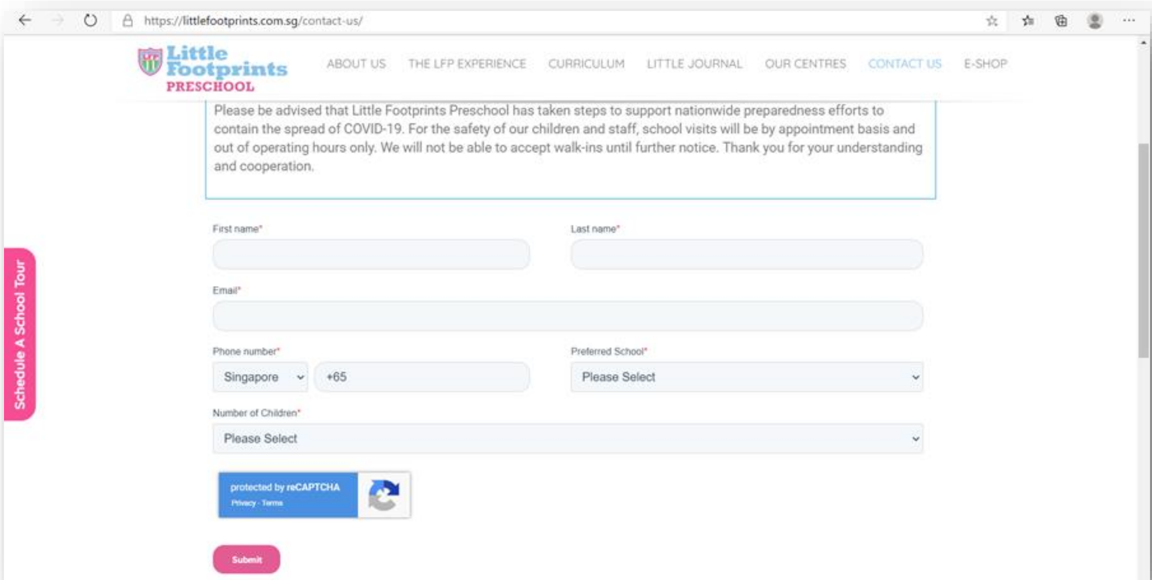


Figure 5: Registering interest at childcare centre's website

<u>Growing problem</u>

As the number of childcare centres and enrolments in Singapore increase, our system offers solutions toward these problems.

Preschool Enrolment Statistics Singapore 2019

Year	2012	2013	2014	2015	2016	2017	2018	April 2019
Total no. of child care centres	1,016	1,083	1,143	1,256	1,342	1,419	1,495	1,517
Total no. of child care centre places	92,779	101,597	109,694	123,327	137,278	149,803	167,421	171,660
Total Enrolment	75,530	73,852	83,928	95,414	103,221	110,826	119,195	119,333
Enrolment in full-day programme	65,826	65,650	75,518	86,898	95,357	103,404	111,838	111,725

Source: ECDA[5]

Table 1: Increasing number of childcare centres

[5] M890651 - Number, Capacity And Enrolment In Child Care Centers

1.3 PROJECT OBJECTIVES

This project aims to deliver a system that provides:

1. Parents with childcare recommendations, along with sentiment analysis based on actual parents' reviews from our system. Parents can then rank and submit their interest on the childcares through our system.

2. Childcare centres with an optimised waitlist of parent interests. which include information like likelihood of prospective parent taking up placements. Childcare centres can then better optimise their resources on managing waitlists. In view of PDPA, only essential information on parents will be included in waitlists.

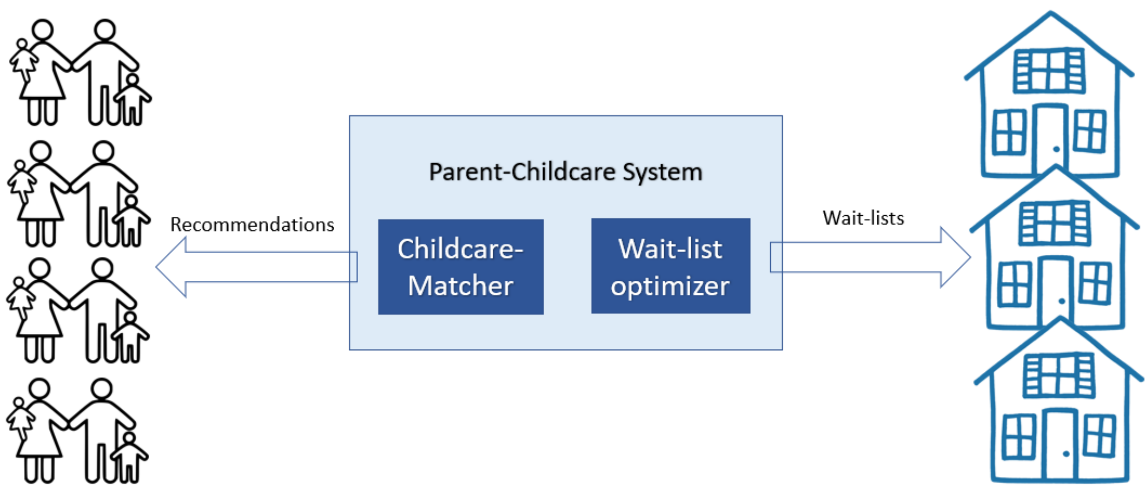


Figure 6: Childcare-matcher and waitlist sub-systems

1.4 KNOWLEDGE ACQUISITION:

To learn about challenges faced by parents and childcare centres, we performed knowledge acquisition. We gained this knowledge from documented sources like forums and undocumented sources like consulting human experts like the principal of childcare centre.

Source of information	Information explored	Acquisition methods
Project Model I – Childcare-Matcher		
ECDA	Childcare centre data	Extracted data based on certain features
KiasuParents portal / Forum	Childcare reviews and comments/discussions	Web scrapping to extract the comments grouped by school names
Distance database	Postal sector information of Singapore from Wikipedia. Distance of postal section using google API with constraints 1, 2 and 3 km and postal code of all childcare centres by web scrapping addresses.	Web scrapping and extraction of information
Project Model II – Waitlist Optimizer		
Discussion with Childcare centres	Engaged childcare centre to understand about their challenges with waitlist management and whether our approach can solve their problems	Through email

Table 2: Knowledge acquired in project

2 SYSTEM DESIGN

2.1 USER WORKFLOW: PARENTS

Parents are able to access our website on the internet. User friendly forms will be presented to parents to gather requirements. Our system then uses the inputs to produce recommendations for the parents. If parents find a suitable childcare centre among the recommendation, they may rank the childcare to indicate their preference level and submit their interest to our system on the website.

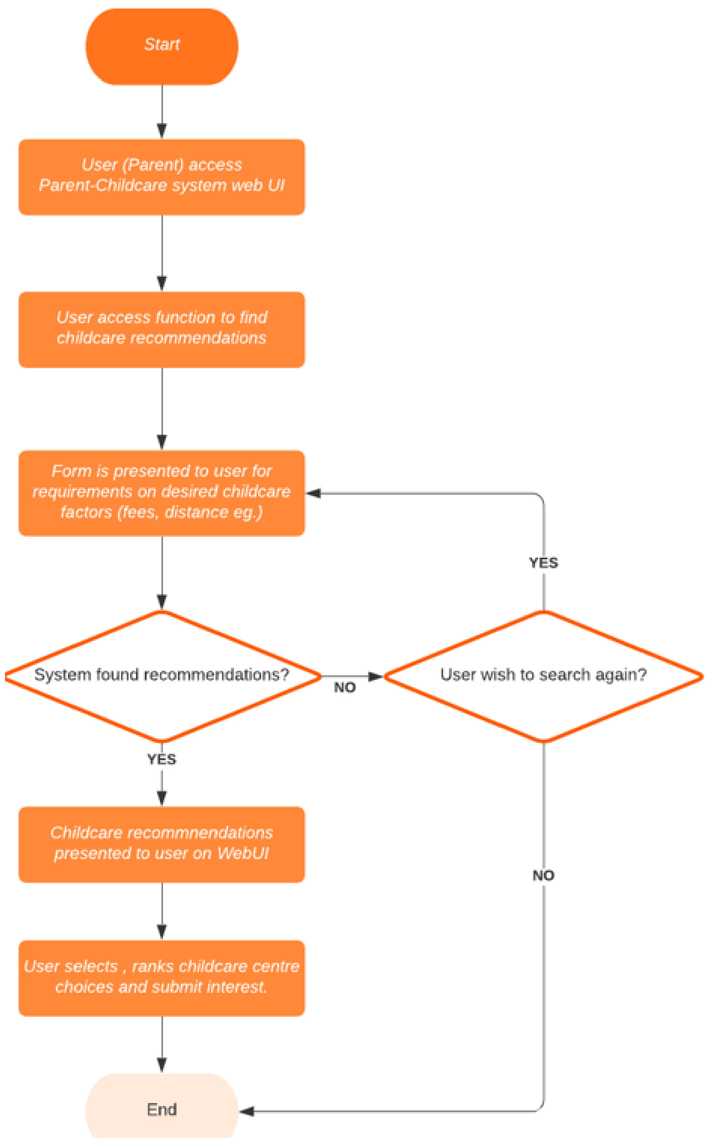


Figure 7: Childcare-matcher user workflow

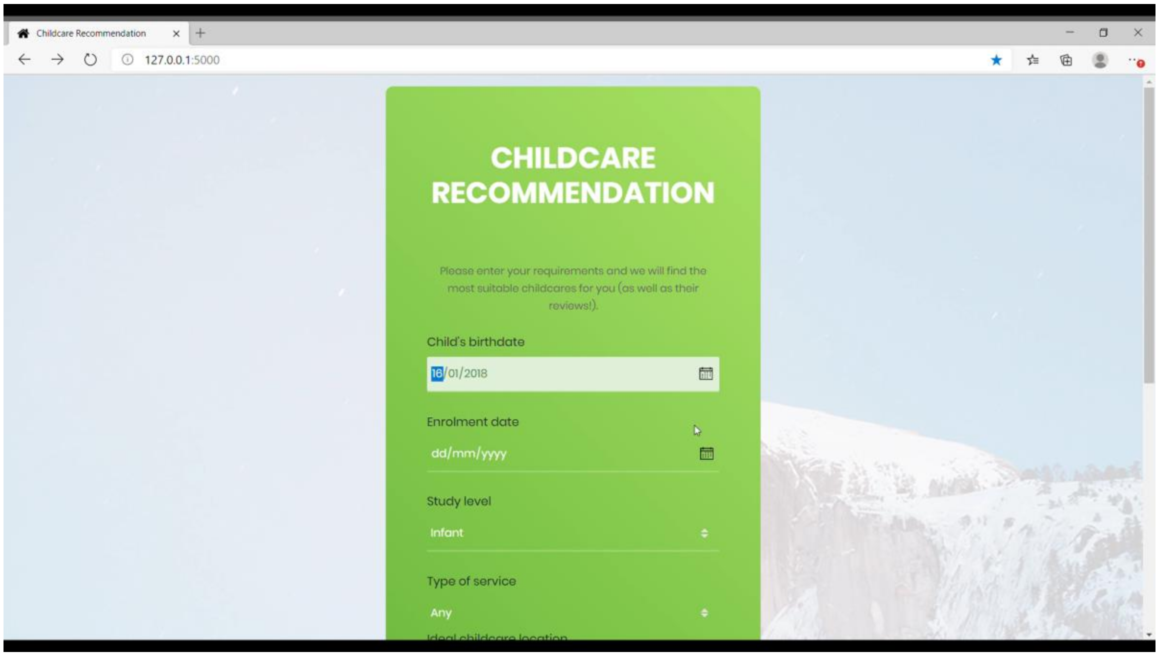


Figure 8: Form for parents to fill up

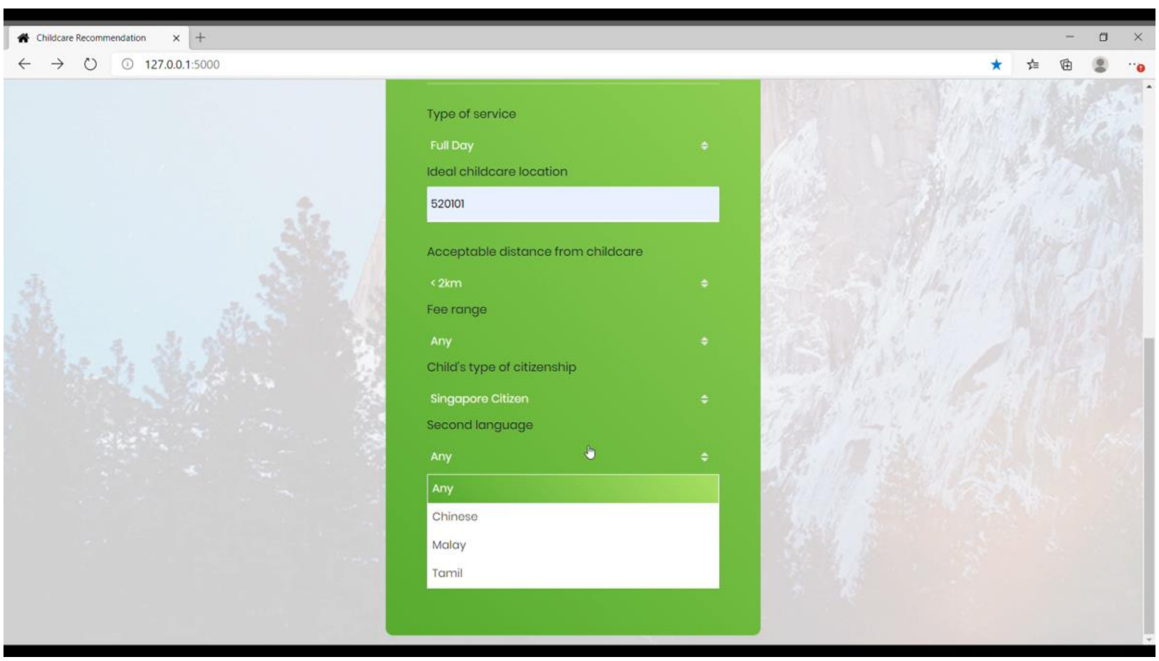


Figure 9: Form for parents to fill up

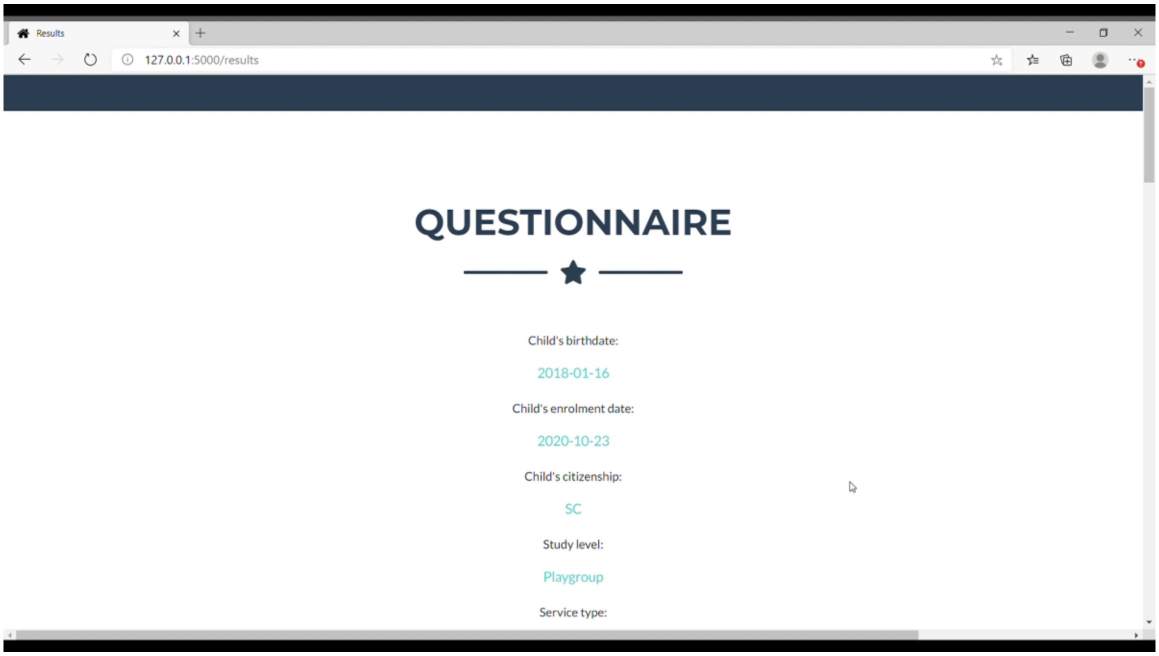


Figure 10: Returned recommendations – first part with answers from parents' forms

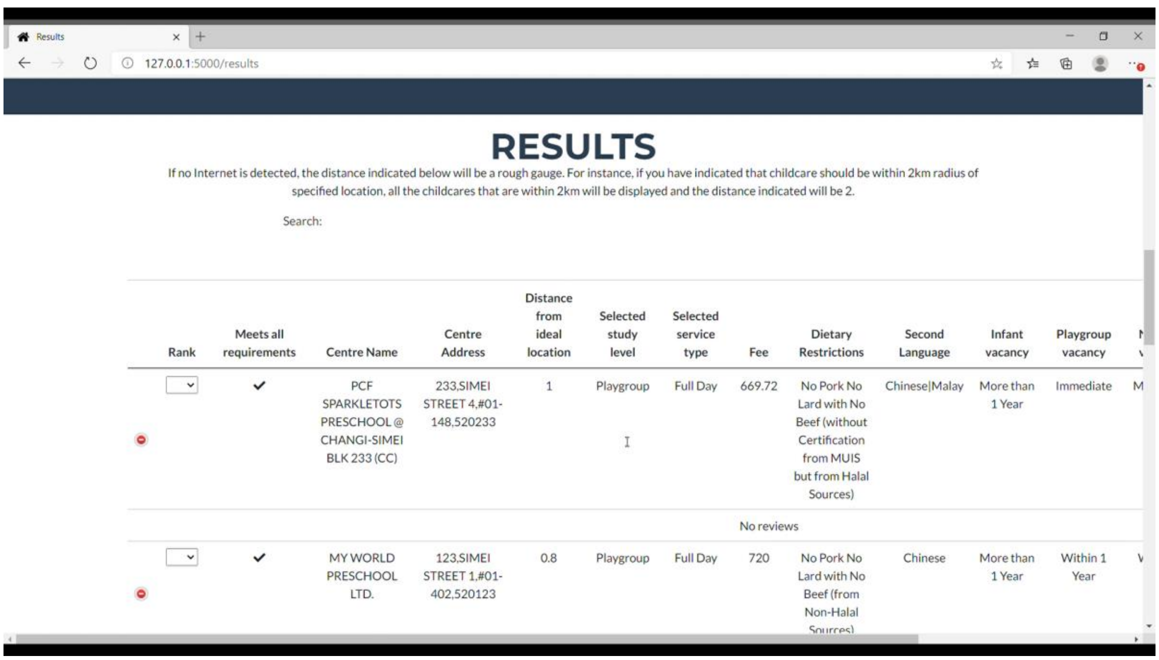


Figure 11: Returned recommendations – second part with childcare details

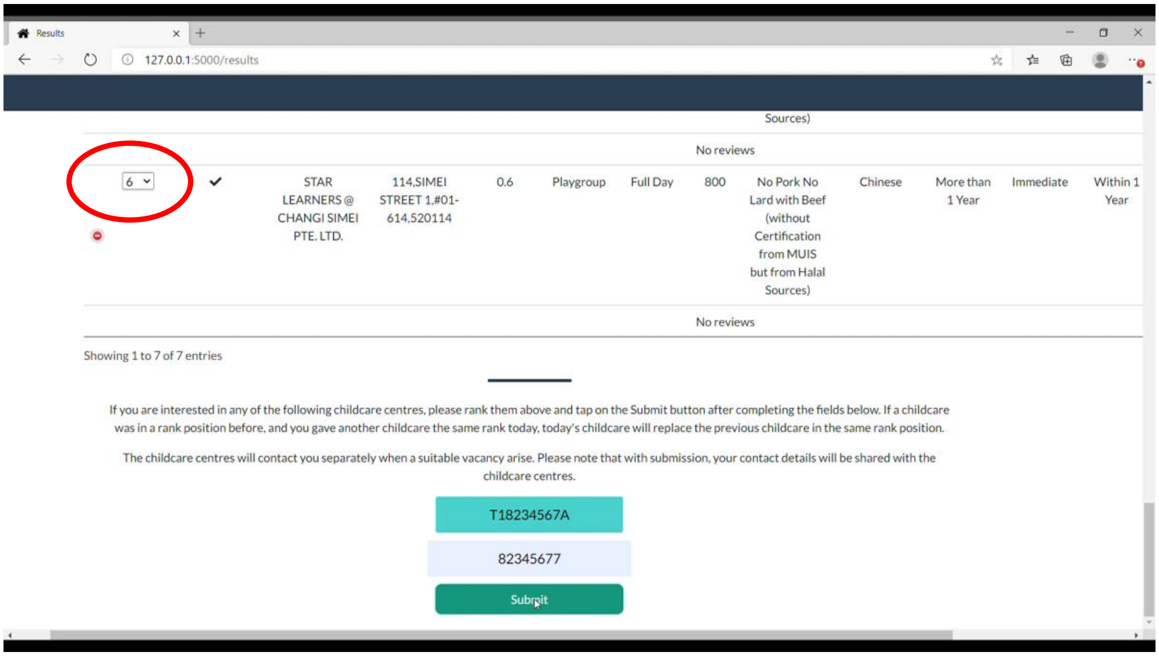


Figure 12: Parents submit their ranking of childcare (red circle) and details of their children

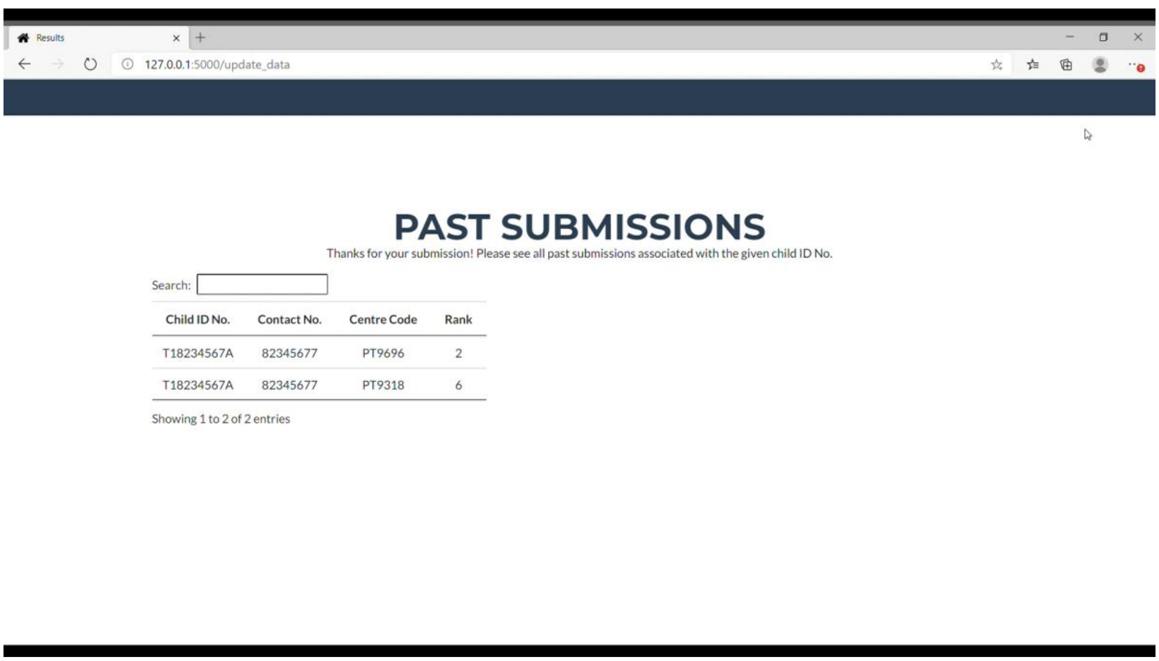


Figure 13: Upon submission, parents will see any of their past submissions

2.2 USER WORKFLOW: CHILDCARE CENTRES

Childcare centre operators are able to access a website on the internet. The operators may search (based on dates eg.) and retrieve waitlists of prospective parents. When waitlist of parent interests is presented to the system users, it will also include information like likelihood of prospective parent taking up placements.

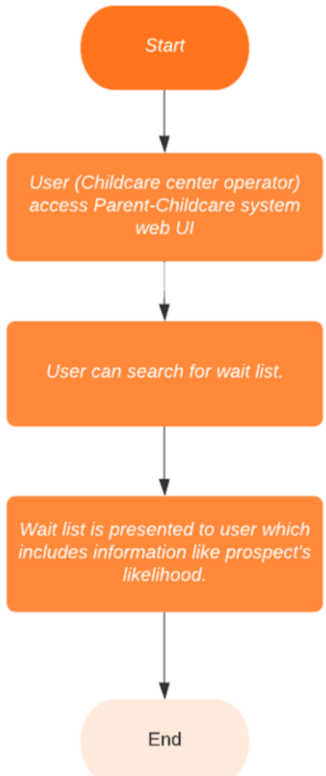


Figure 14: Waitlist optimizer user workflow

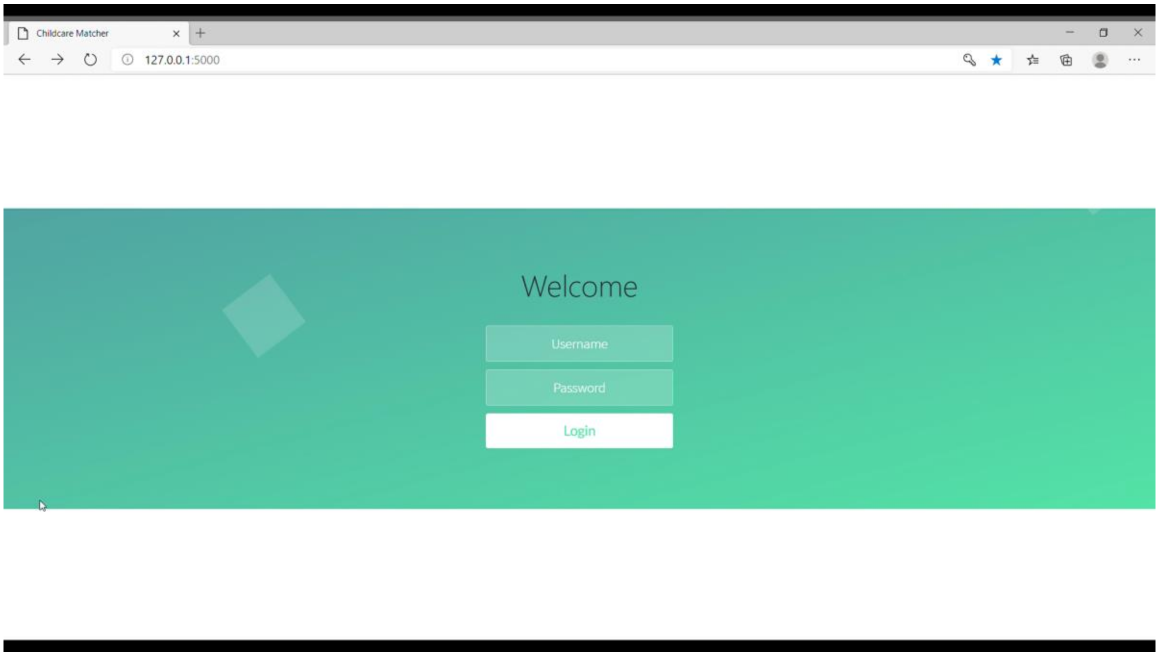


Figure 15: Login for childcare

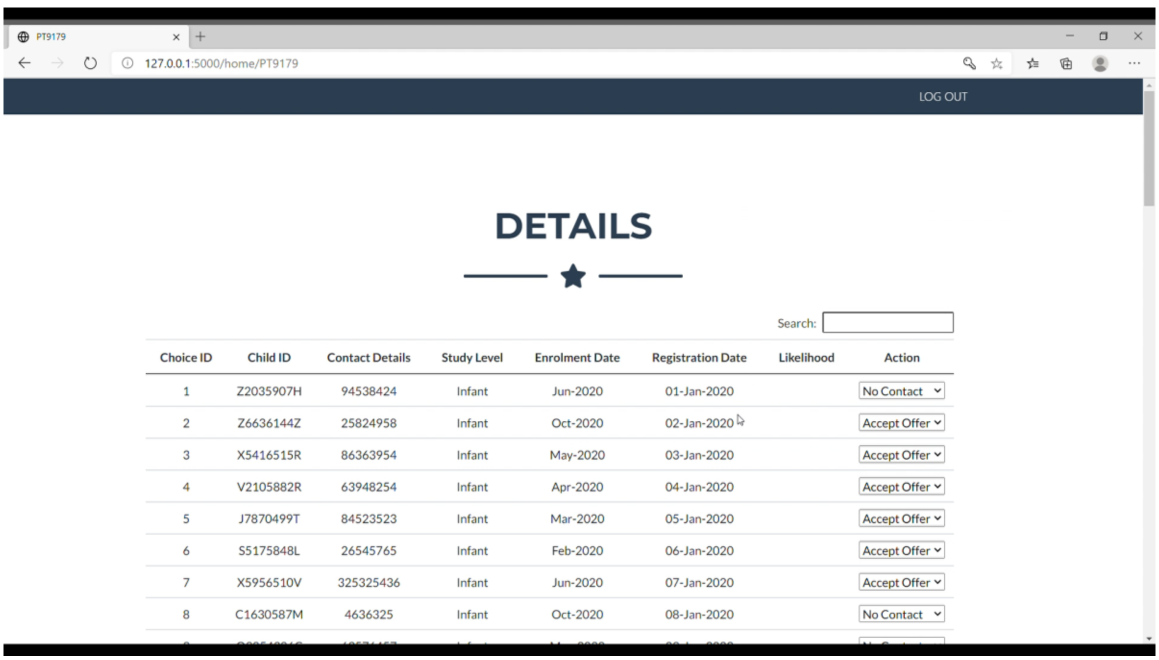


Figure 16: Waiting list of parents' interest in a particular childcare

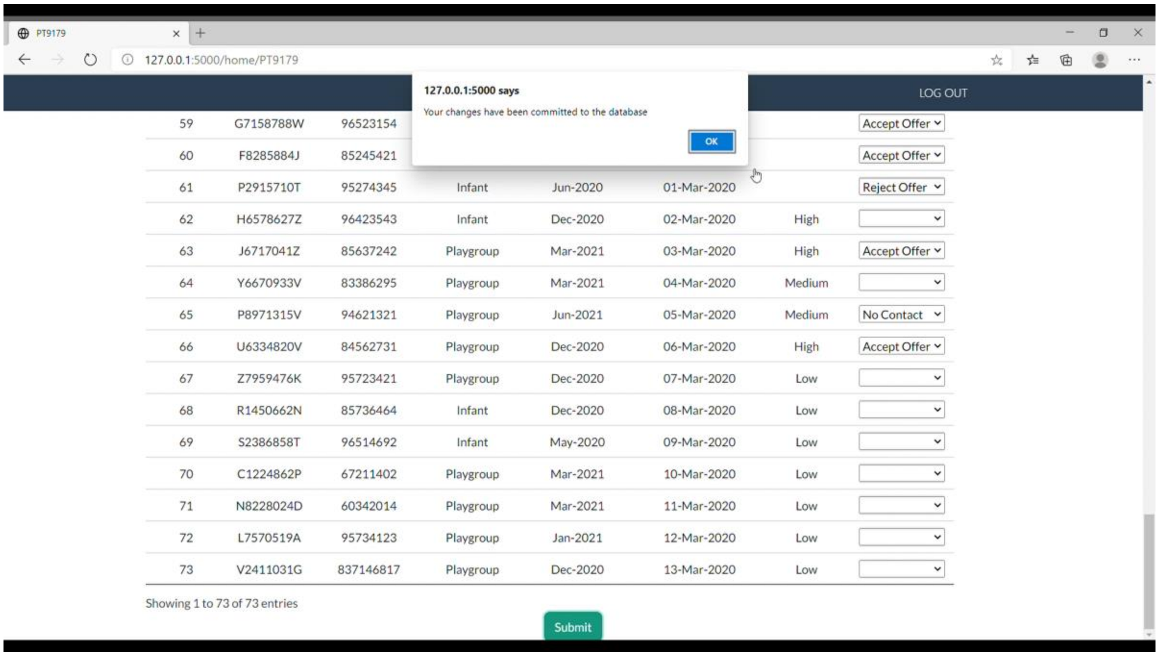


Figure 17: Childcare update whether parent accepted offer/rejected offer/no contact. This serves as future reference for our decision tree model when predicting likelihood of parents accepting offer.

2.3 SYSTEM ARCHITECTURE

Figure 9 shows the overall architecture of the system:

1. **Childcare-Matcher**: It is composed of two separate processes, one for the parent to receive recommendations, another for parent to submit their interest (of childcare centre) to our system

2. **Waitlist optimizer**: Childcare centre operator receive waitlist information from our system

3. **Interface** between front end and back end systems.

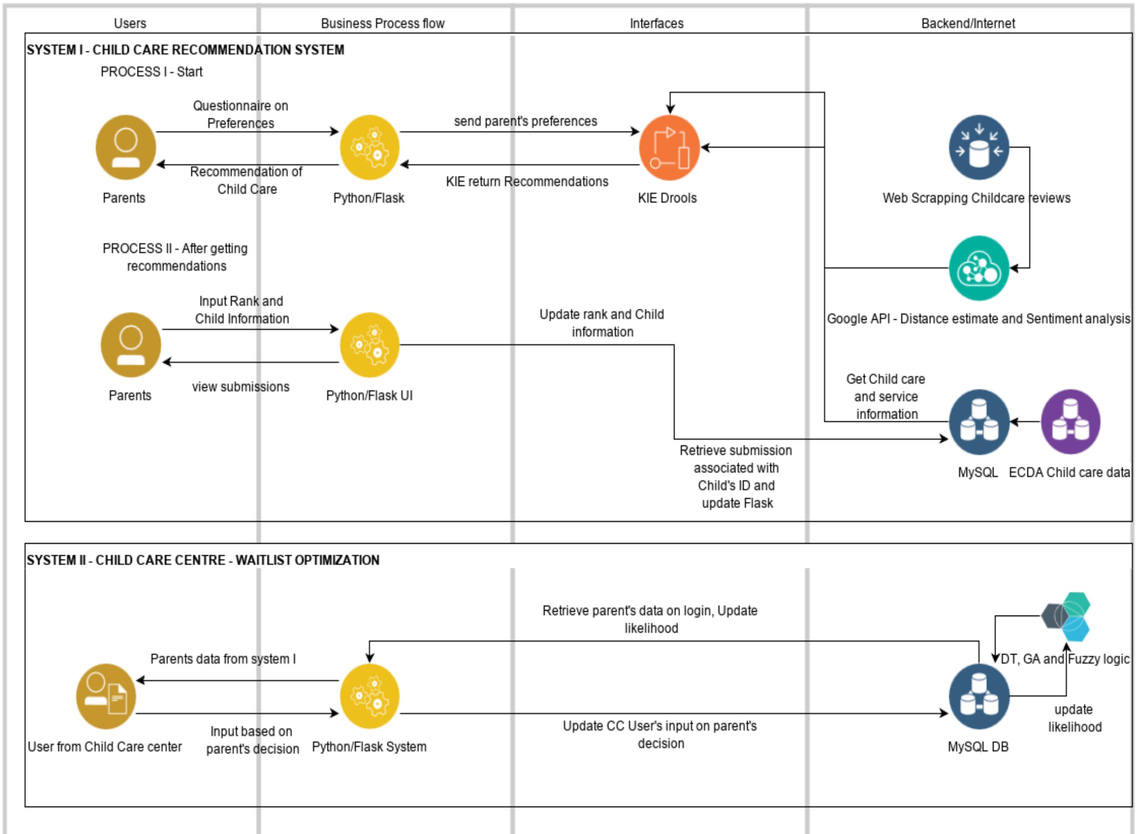


Figure 18: Overall system architecture

Data collected from childcare-matcher system are stored in database (MySQL). These data are accessible by waitlist-optimizer system.

2.4 DESIGN: CHILDCARE-MATCHER FOR PARENTS

Childcare-matcher system have two processes:

1) Parents input their preferences in the form of questionnaire through user interface using Python/Flask. This data is then sent to KIE Drools engine. The engine gets childcare and service information from the database and returns the recommendation to parents via Python/Flask.

2) After receiving recommendation from our system, parents provide a submission which includes information like their ranking of childcare centre and their child's details. Parents can also view past submissions made in the system.

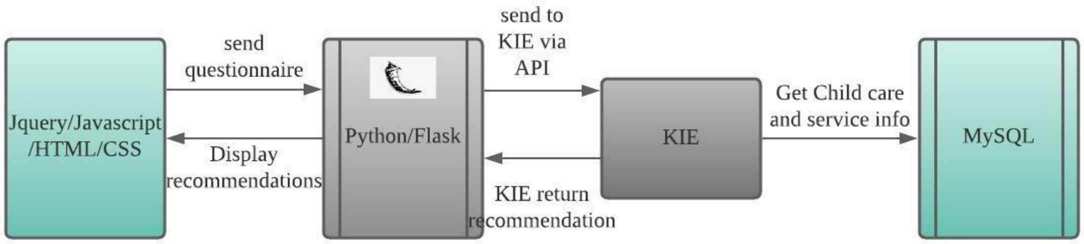


Figure 19: Childcare-matcher design

The following data are extracted and loaded into MySQL:

- Childcare centre data from ECDA (Early Childhood Development Agency)
- Kiasu parents' forums are crawled for reviews and comments

Travelling Distance is one of the key preferences of parents when it comes to selecting a childcare centre.

- We created a distance database using Google API. This is used when we provide parents with a quality recommendation that takes in consideration of the distance factor.
- We have the questionnaire built to know the parent's choice of preference when it comes to traveling distance for childcare centre. (Ex: childcares closer <1, 2 or 3 kms) as a key preference from the questionnaire)

The system has been integrated with rules engine in KIE. Drools BPMN business process together with Drools rules are used in KIE Workbench tool.

Limitations:

- The ideal location field in the questionnaire only allows postal codes inputs.
- Childcare centre receiving these details from Childcare-Matcher system may have duplicates in the ranking input of the users/parents. For instance, a parent may have ranked a childcare as 2nd in a first submission, but ranked the childcare again as 6th in a second submission.

Future Improvements:

- Currently, we manually map scrapped reviews/comments to the childcare centre (represented as codes) as there are many childcares with same/similar names, especially for major childcare centres like PCF Sparkletots. Hence, we need to discern from the reviews which childcare centre the reviewer is referring to e.g., "*I would like to know more about the sparkletots centre in Clementi*". In future, we can use natural language processing to extract childcare centre's name and location from the reviews and automatically map the review to the centre.

- Currently, the format of the comments scrapped from portals (kaisu parents) is like a discussion kind of comments on the quality/maintenance of child care centres and other issues as well being discussed. When it comes to sentiment scores for these discussions, it is not a good practice to get scores for questions asked in the forum while we provide these discussions to parents along with the recommendations that might not be meaningful and useful. Here the improvisation has been done to restrict the sentiment score only to meaningful comments. Also, the future scope for this functionality is to get the sentiment score for entire discussion in a more precise way. And this also accounts for quality comments being scrapped from various websites and forums.

- Allowing anyone to use a child's identification information (NRIC) to register to childcare centre is a risk. We can improve the security of the system by authenticating the user who is submitting registrations for a child.

2.5 DESIGN: WAITLIST OPTIMIZER FOR CHILDCARE CENTRE

The following diagram illustrates the design and interaction between the components of the waitlist-optimizer.

- Python and Flask is used as the main framework for development and deployment.
- The user, who is the childcare centre operator, interface with the system via a HTML frontend throughout the entire process. The frontend displays information on waitlists and collects inputs from the user.

- Childcare centre user, parent's and wait-list information are stored in MySQL database.

- Rules discovery and optimisation functions are set up in waitlist optimizer system. Scikit-Learn and Distributed Evolutionary Algorithms (DEAP) libraries are used for rules discovery and optimization models respectively.

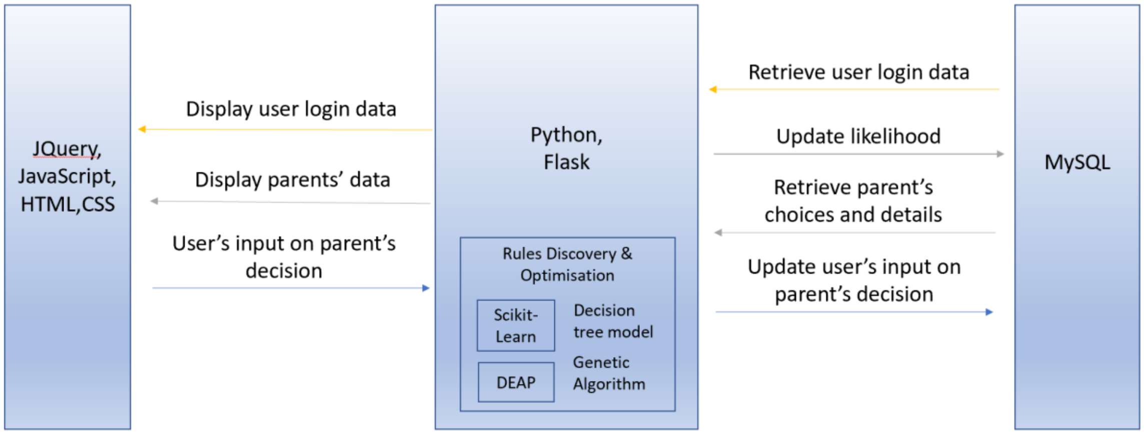


Figure 20: Waitlist optimzer design

3 SYSTEM SOLUTION & IMPLEMENTATION

3.1 CHILDCARE-MATCHER

3.1.1 RULE BASED SYSTEM

The inference diagram is used to represent various factors considered while making recommendations of childcare centres to users (parents).

The root node is the final recommendation. The recommendation decision is divided into various sub goal levels and given in a tree structure. The leaf nodes are data that we gather from the users as they input their preferences through questionnaire.

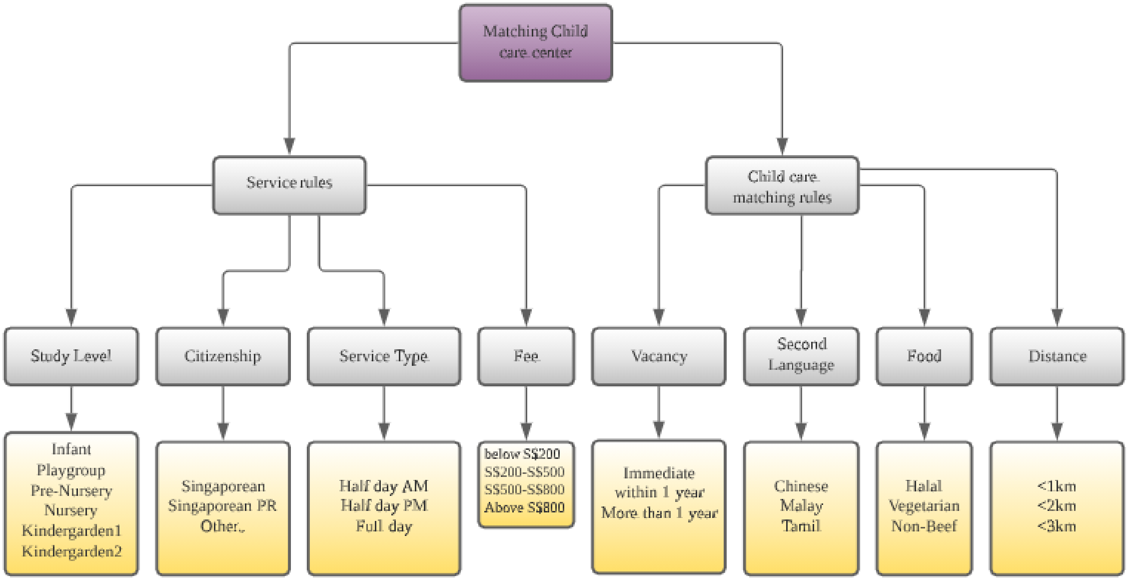


Figure 21: Inference diagram for childcare-matcher

The rule-based system performs filtering based on rules on source of data/input provided by the user(parents). The following factors are used in filtering:

- Study level
- Citizenship
- Type of service (full day, half day AM or half day PM)
- Fee range
- Childcare vacancy
- Second language (Chinese, Malay or Tamil)
- Type of food (Vegetarian, Non-beef, Halal)
- Distance from ideal location

In the event no childcare center match can be found, the rule system will only filter essential factors like study level, vacancies etc.

3.1.2 SENTIMENT ANALYSIS

Reviews/comments in forums and other sources like parent groups in platforms (Facebook, WhatsApp eg.) offers useful information to parents in choosing the right childcare center:

- It help parents decide on the service
- It provides some proof to childcare center reputation
- It provides some insight of childcare center trustworthiness

- It provides other information like facilities, infrastructure and maintenance
- It provides some information on how children's welfare are taken care of

These are important information to a parent. Hence, they have to be provided to the parent as part of childcare center recommendation. In order make this process easy for the parents hunting for childcare, we have:

- Scrapped comments/reviews from various sources like KiasuParent forum
- Analysed comments/reviews for sentiments and provide sentiment magnitude score using Google API. These sentiments are stored in the database.
- Identified a threshold value from overall reviews that we scrapped and divided our comments/reviews based on how positive or negative they are.
- In addition to the list of recommended childcares for the user, our system also gives the reviews for each school recommended, if available.

The parent can then use these comments/reviews available along with recommendations to help them choose childcare centers.

Future Work/Improvements:

- To scrape more reviews/comments and with better accuracy
- Restructure scraped results and organize them in more efficient way
- Gather more of reviews/comments associated with each childcare centre

3.2 WAIT-LIST OPTIMZER

3.2.1 KNOWLEDGE BASED SYSTEM

The waitlist optimizer's objective is to provide childcare centers with waitlist of parent interests.

A key information that childcare centers operators are looking at is the likelihood of prospective parent taking up placements. Hence the core system's function here is to predict this likelihood. There are 3 classes of likelihood to be predicted by this system: 'High', 'Medium' and 'Low'.

For each childcare center user, waitlist-optimizer behaves differently at two phases:

1. **Phase 1**: During this phase, labelled data is scare, especially when the system is early in production stage or when the childcare center operator is a new user of our system. There is small number of parents submitting their interests (to the childcare centers) at this stage. These submissions are labelled data and are scare. At this phase, domain expert knowledge is used, and is represented as fuzzy rules to classify likelihood.

2. **Phase 2**: When there are more labelled data (i.e., due to higher number of parents using our system, data driven approach is adopted. Decision tree models are set up to discover rules leading to likelihood classification. Genetic algorithm is also used to find more optimal rules leading to highest likelihood (i.e. category "High likelihood").

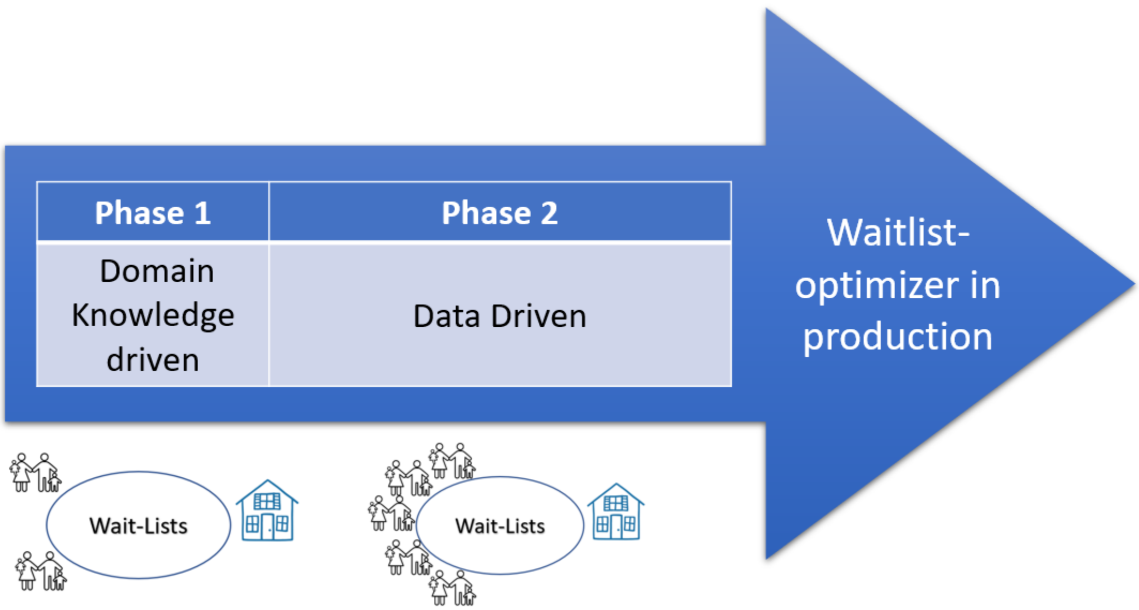


Figure 22: Waitlist optimizer behaves differently at two phases

3.2.2 RULE DEFINITION, DISCOVERY, AND OPTIMISATION

Labelled data are scare in early phase of system. When parents submit their childcare center interests in our childcare-matcher system, these data entries are processed by our rules discovery and optimization models. Parent entries completed with their decision to accept placement in a childcare are used by the models as labelled data.

A threshold is used to determine whether the waitlist-optimizer should behave in Phase 1 or 2. When the number of labelled data falls below the threshold, our system behavior is in Phase 1. In our project, we use a threshold value of 50 (ie. 50 labelled data).

3.2.2.1 PHASE 1: FUZZY RULES

The 3 classes of likelihood are categorized into 'High', 'Medium' and 'Low' here as follows:

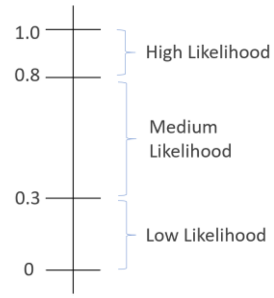


Figure 23: Likelihood classification

Functions are designed to represent domain expert knowledge in fuzzy rule manner, and are used to define likelihood at this stage:

1. <u>Difference (months) between parents' registration of interest and child's enrolment date</u>. When the gap in time between registration and enrolment date is bigger, there is higher the likelihood of parent accepting placement in childcare centre. This is because parents who are sure of their childcare choice tend to register earlier, rather than at the very last minute.

   > Fuzzy logic illustration:
   >
   > IF registration date is 11 months away from enrolment date, THEN parent is most likely to take up offer (100%).
   >
   > For instance, parent register in Jan 2020 for Dec 2020 vacancy. 11 months is used because a current solution, the ECDA's website, only allows for parents to register up to a year before enrolment date.

2. <u>Difference (months) between the date childcare center operator is looking at waitlist and parents' registration date.</u> The smaller the gap in time here means that there is higher likelihood of parent prospect accepting placement in childcare center. If the parent only very recently registered their interest, there is a higher likelihood that the parent will take up the placement offer.

Fuzzy logic illustration: IF parent registered child in Jan 2020 (for later enrolment date) and childcare contacted the parent within Jan 2020, THEN parent is most likely to take up offer (100%).

3. <u>Rank of childcare centre by parent</u>. When parents submit their entries into our system to register their interest in multiple childcare centres, they rank their preference for each childcare centre. A lower number represents a higher rank. A higher rank leads to a higher likelihood of parent prospect accepting placement for that childcare centre.

Fuzzy logic illustration: IF parent indicate a childcare is rank 1, THEN parent is most likely to take up offer (100%).

The following diagram illustrates the design of functions which is used in fuzzy inference. The minimum output among the three functions is used to define the likelihood of a parent accepting placement in a childcare centre.

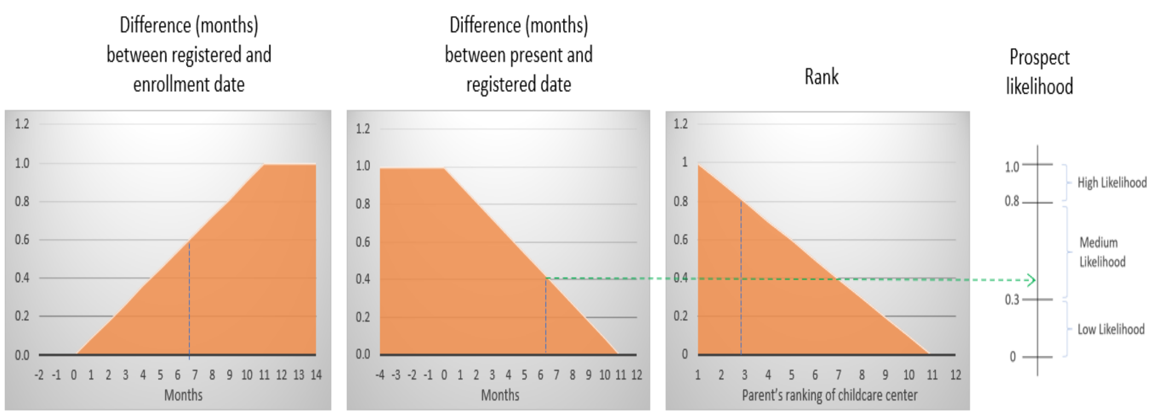


Figure 24: Fuzzy inference of prospect likelihood

Fuzzy logic illustration:

IF parent indicate a childcare is rank 10 (10%) AND childcare contacted parent the same month he/she registered (100%) and parent is registering child to attend only 11 months later (100%), THEN parent is only 10% likely to take up offer from this childcare.

3.2.2.2 PHASE 2: RULES DISCOVERY

When the number of labelled data[6] is above threshold (set as 50 in our system), waitlist-optimizer system becomes data driven and in Phase 2. Prediction of likelihood is done using decision tree (DT) and genetic algorithm (GA) as follows.

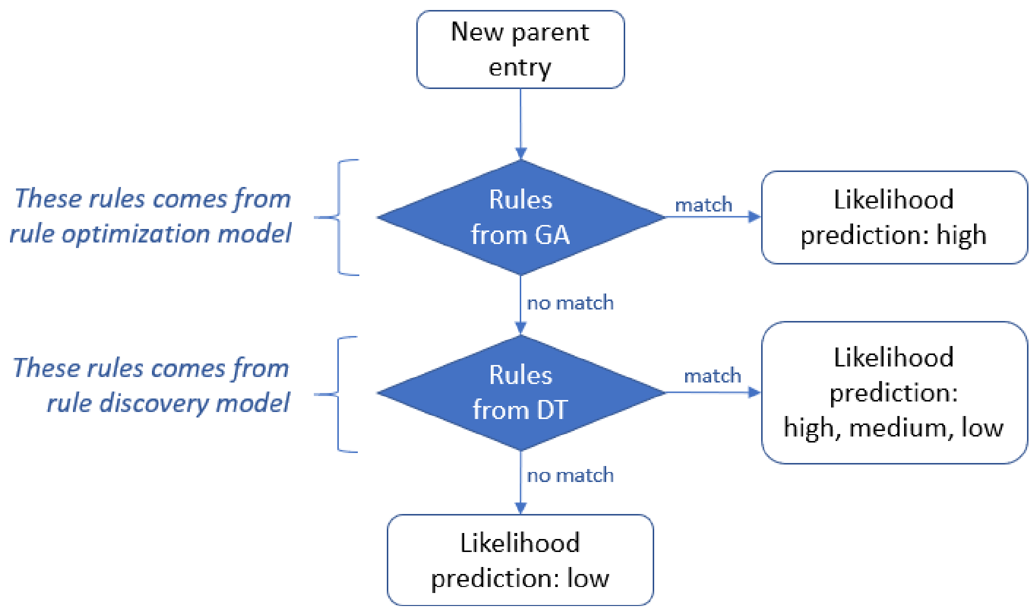


Figure 25: How predication of likelihood is done for each parent

The set of rules from the DT model is the core rule set used to determine likelihood for all likelihood classes (ie. 'High', 'Medium', 'Low'). The GA is an add-on, to find more optimal rules leading to "high" likelihood. What is meant here is that if there is no activation of any rules that comes from GA model, one of rules that come from DT model will be activated definitely.

[6] Labelled data refers to parent entries completed with their decision to accept placement in a childcare.

Phase 2 is data-driven in the sense that when the real world environment changes, or when there are differences in parent's consideration for different childcare centres, these differences are captured from past labelled data, and are used in rule discovery.

Decision tree models are used to discover rules leading to likelihood categories of high, medium and low. GA is also used to find optimal rules leading to high likelihood. The 3 classes of likelihood are categorised into 'High', 'Medium' and 'Low' using ranges:

Likelihood	Range
High	2 and above
Medium	1 to 2
Low	0 to 1

Table 3: Likelihood classification used by rule discovery function

CRISP DM framework is used as framework for the rules discovery process.

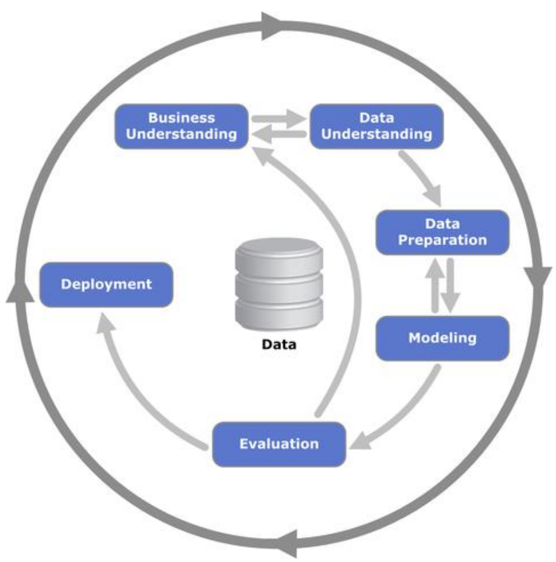


Figure 26: CRISP-DM framework

1. Business Understanding: The business problem is that childcare centre operators need to manage an inflated waitlist of prospective parents which does not reflect true demand. The business objective for the system here is to estimate the likelihood of a prospect accepting placement. The operator may use this information in planning and managing of prospect waitlist.

2. Data Understanding: ECDA website was studied to understand features that parents look for when searching for childcare centre. Knowledge from domain expert (i.e. parents) was gathered. This collective understanding is used to create a sample dataset. The format of dataset matches the format actual data to be collected when waitlist-optimizer is in production.

   Real data about prospects was considered to be collected via surveys, but a sizable dataset for decision tree model development is required. This means a sizable survey is required. Financial cost was a limitation during this project, hence data is simulated and used for model development. When the system is operational, there will be real data on parent prospects when parents provide inputs into the childcare-matcher system.

3. Data Preparation: Decision tree models was set up using Scikit-Learn in Python. Scikit-Learn has a peculiar limitation that only numeric data inputs are accepted for a decision tree models. Hence, there is need for one-hot-encoding of feature data that is categorical.

   For target data, the labelled values for the dataset is made up prospective parent final decisions which are categorised as "Reject offer", "No contact" and "Accept offer". These values are available after parents update their decisions in our childcare-matcher system. These responses represent likelihood of accepting placement and are encoded ordinally.

   A data transformation pipeline had been prepared for data preparation activities. After data transformation, the dataset is partitioned into 2 sets for training and testing.

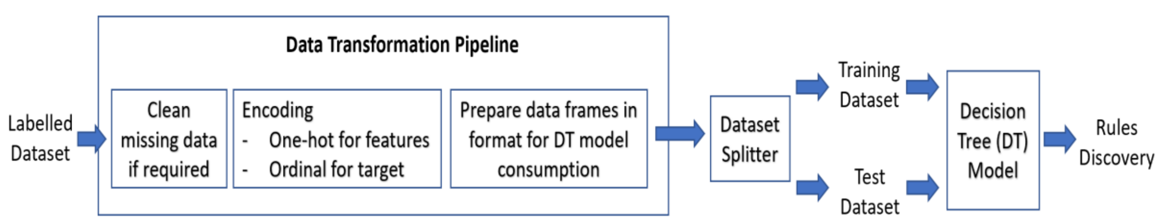
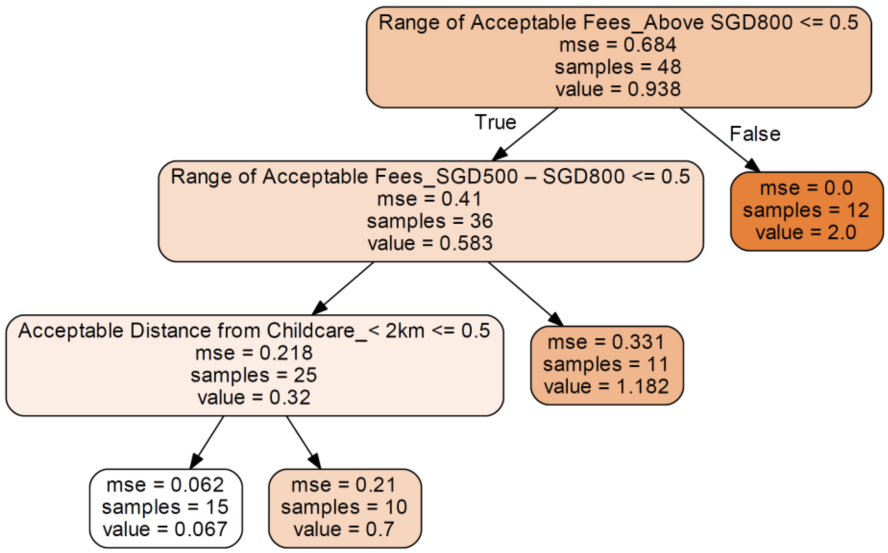


Figure 27: Data pre-processing

4. Modelling: Decision tree Regressor model was set up for the discovery of three set of rules leading to 3 classes of likelihood (i.e. Low, Medium, High) of prospective parent accepting childcare centre placement.

The model produces decision tree diagrams. Rules are also output in text form. These outputs are used for rules interpretation. The rules discovered here are used to determine likelihood of parents accepting placement. The following shows examples of our rule discovery model outputs.




```
|--- Range of Acceptable Fees_Above SGD800 <= 0.50
|   |--- Range of Acceptable Fees_SGD500 - SGD800 <= 0.50
|   |   |--- Acceptable Distance from Childcare_< 2km <= 0.50
|   |   |   |--- value: [0.07]
|   |   |--- Acceptable Distance from Childcare_< 2km >  0.50
|   |   |   |--- value: [0.70]
|   |--- Range of Acceptable Fees_SGD500 - SGD800 >  0.50
|   |   |--- value: [1.18]
|--- Range of Acceptable Fees_Above SGD800 >  0.50
|   |--- value: [2.00]
```

Figure 28: Decision Tree Regressor outputs

5. <u>Evaluation</u>: GridSearchCV from Scikit-Learn library was used to find the optimal set of hyper parameters (ie. tree depth, number of samples per leaf) to form the model. Mean-square error was used as metric in the selection.

   Using test dataset, the decision tree model was also evaluated based on mean-square error between it's prediction and true label values.

6. <u>Deployment</u>: After the evaluation was successful, the code was integrated into the waitlist-optimizer.

3.2.2.3 PHASE 2: RULES OPTIMISATION

When childcare operator manages their waitlist, one task is to allocate staff and resources (financial e.g.) to look into prospective parents with "high" likelihood of accepting placement.

Rules are discovered by decision tree (DT) model for 3 classes of likelihood (i.e. Low, Medium, High). DT model outputs a label of numeric value to indicate the likelihood. Genetic algorithm (GA) is used in our system to find optimal rule sets that leads to "High" likelihood represented as numeric value of 2 or higher.

Likelihood	Range
High	2 and above
Medium	1 to 2
Low	0 to 1

Table 4: Same Likelihood classification used by rule discovery optimisation model

The following describes the design and how GA is used to find more optimal rules.

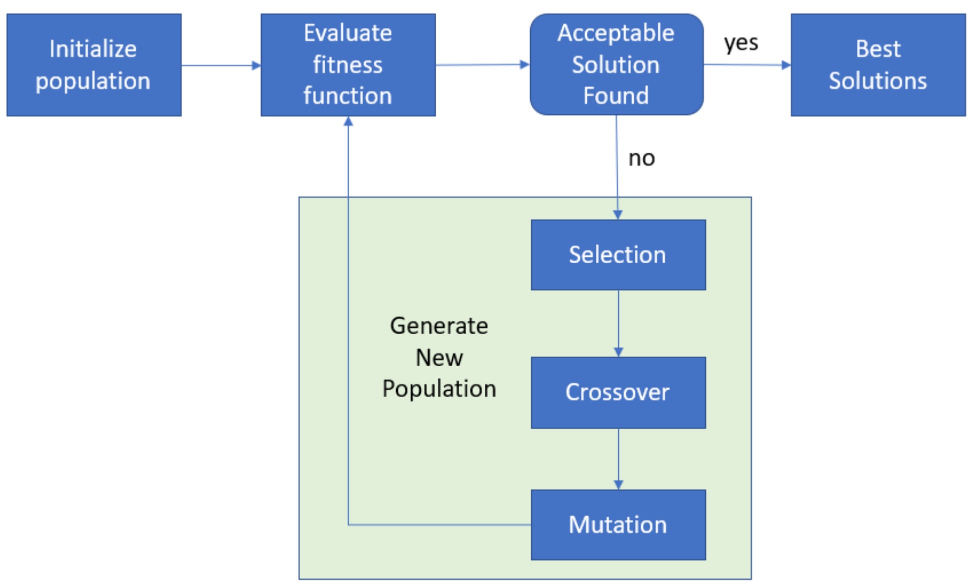


Figure 29: Genetic Algorithm

1. <u>Chromosome design</u>:

   - When the decision tree is being trained using training dataset, the features values of the dataset are one-hot encoded. These feature values represent rule variables. The following shows a partial example:

Description	Values	Encoded values
Rule variable (RV) 1: *Acceptable distance from childcare*	{ <1km, <2km, <3km, Any }	{ 1000, 0100, 0010, 0001 }
Rule variable (RV) 2: *Second Language Taught*	{ Chinese, Malay, Tamil }	{ 100, 010, 001 }

Table 5: Example of rule variable being encoded

   - The encoding format for all rule variables used in both GA and decision tree models matches and are the same. In both GA and DT models, rule variables are represented by sets of binary numbers concatenated together. These rule variable forms the genes of chromosome used in GA. Figure 19 shows a partial example of chromosome genes being represented as sets of binary numbers.

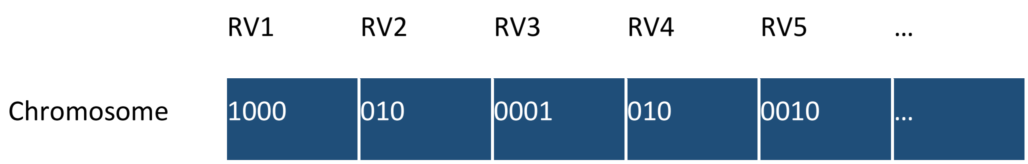


Figure 30: Chromosome genes are rule variables

2. <u>Fitness Value</u>:

   - The trained decision tree regressor model described earlier (Rules Discovery in Section 3.2.2.2) is used here. When a GA generated solution is passed into the decision tree model, one of the decision tree leaf nodes is activated and the model outputs a prediction value.

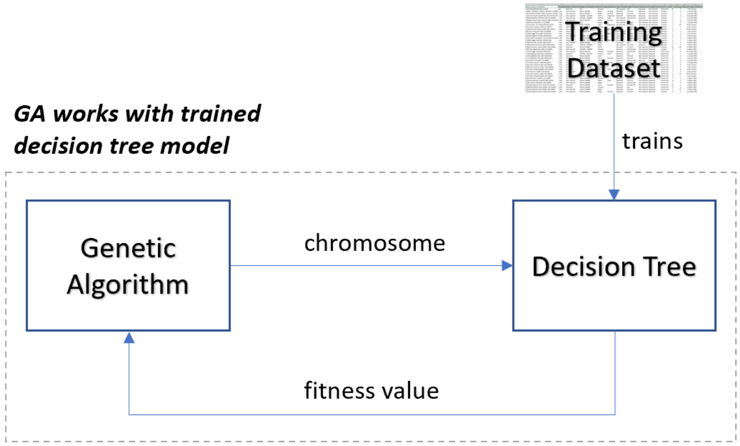


Figure 31: Rule optimisation model

- The fitness value used here is a weighted sum of: **leaf node sample coverage + decision tree regressor prediction value**.

- A specific leaf node in decision tree is activated when the decision tree model is used to provide a likelihood prediction. Sample coverage refers to the proportion of training samples captured by the leaf node relative to total number of training samples used during decision tree training.

- GA is looking for chromosomes (set of rule variables) with maximum fitness values in terms of higher sample coverage and prediction value.

3. Constraint:

- Each rule variable in the chromosome is one-hot encoded. The constraint here that GA qualifies as a valid chromosome/solution is that there can only be one '1' value per rule variable.

4. Selection:

- Tournament selection method is used to select parent chromosomes from the population. The winner of each tournament is used for crossover.

5. Crossover and mutation:

- Random one-point crossover and random flip-bit mutation functions from Distributed Evolutionary Algorithms (DEAP) library is used.

In summary, the following is how GA is used by waitlist-optimizer:

- GA generates a population of solutions. A set of solutions with fitness values equal or above the fitness value of 2.0 is extracted.
- When the system is in live production mode and receiving data inputs from parents, the data input is encoded into chromosome format and checked for matches with this set.
- If there is a match, the data input is classified as “High” likelihood. If there are no matches, rules from decision tree will be used to predict likelihood.

4 SYSTEM FUNCTIONAL DETAILS

4.1 SYSTEM DATA FLOW: CHILDCARE-MATCHER

The following diagram illustrates the details in data flow for childcare-matcher system.

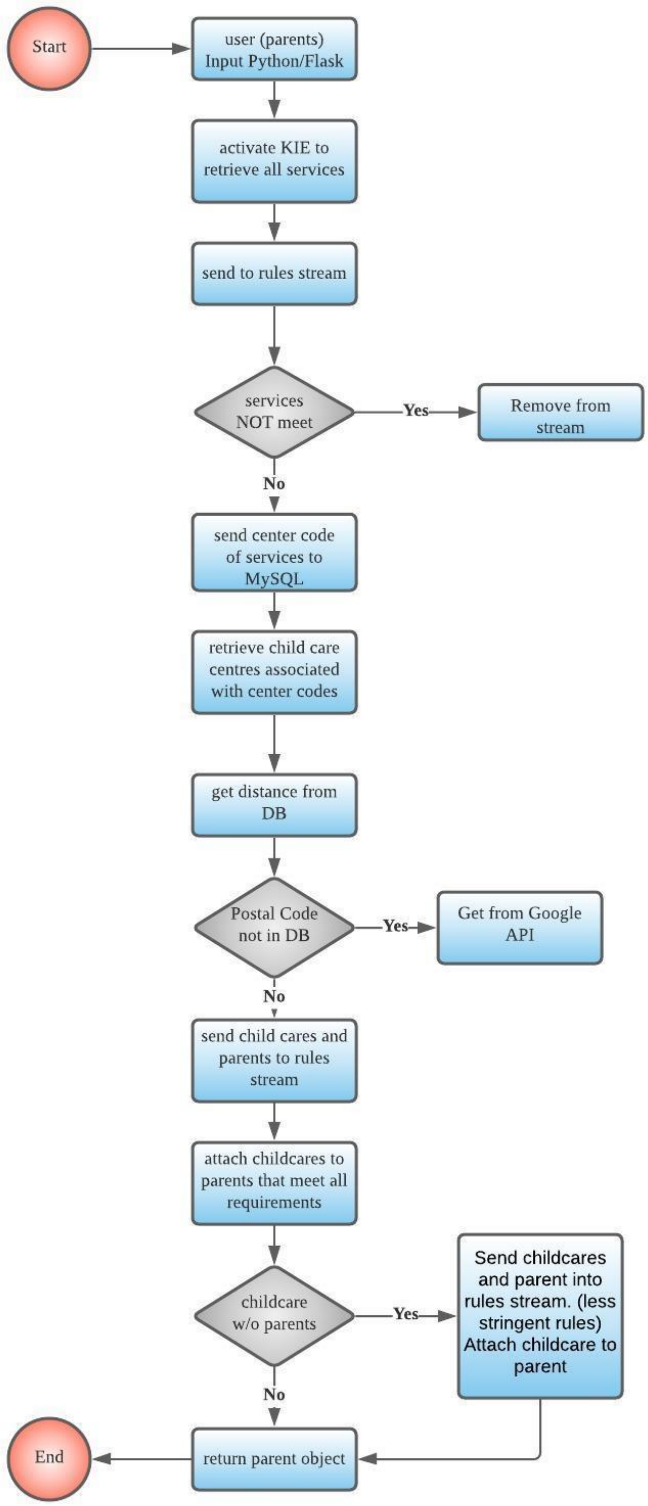


Figure 32: Data flow for childcare-matcher

4.2 SYSTEM DATA FLOW: WAITLIST OPTIMIZER

4.2.1 WAIT-LIST GENERATION AND DISPLAY

The following diagram illustrates the details in data flow for waitlist-optimizer system.

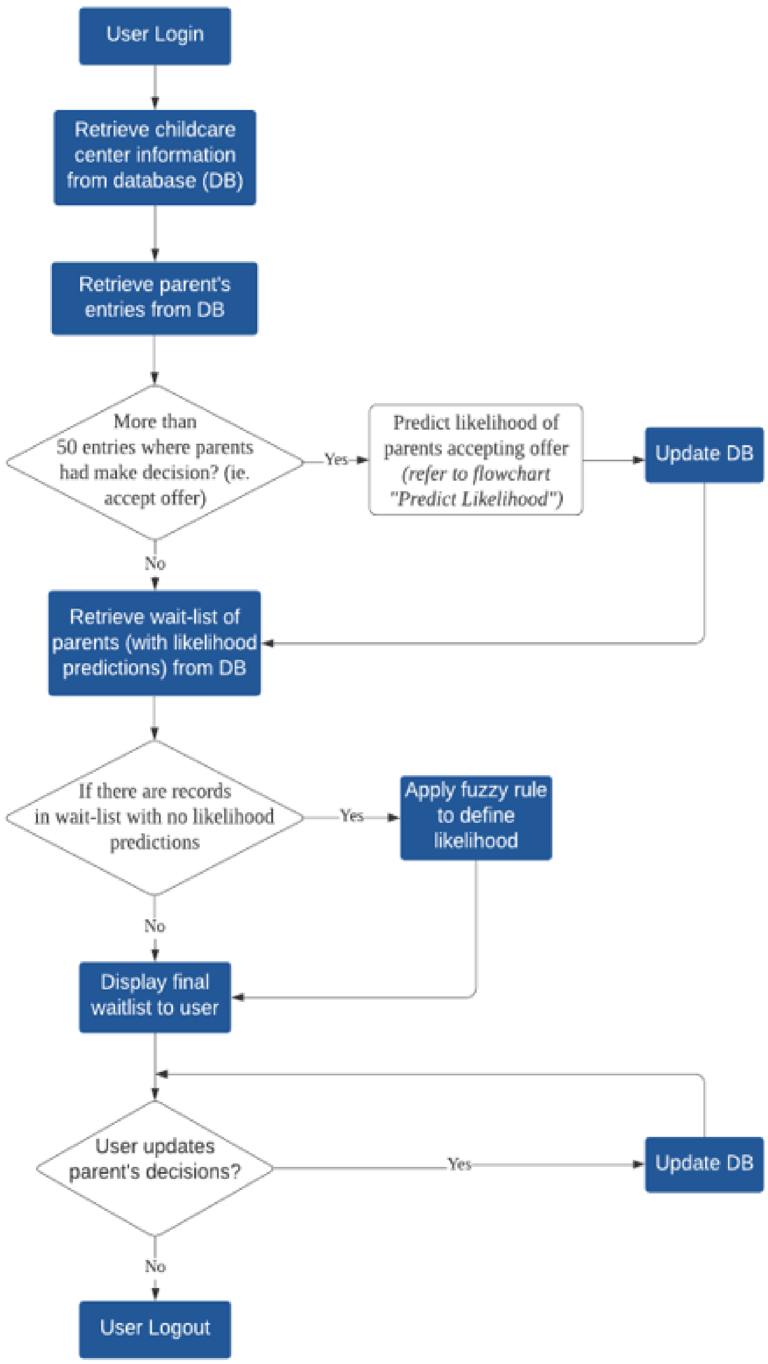


Figure 33: Data flow for waitlist-optimizer

4.2.2 PREDICT LIKELIHOOD

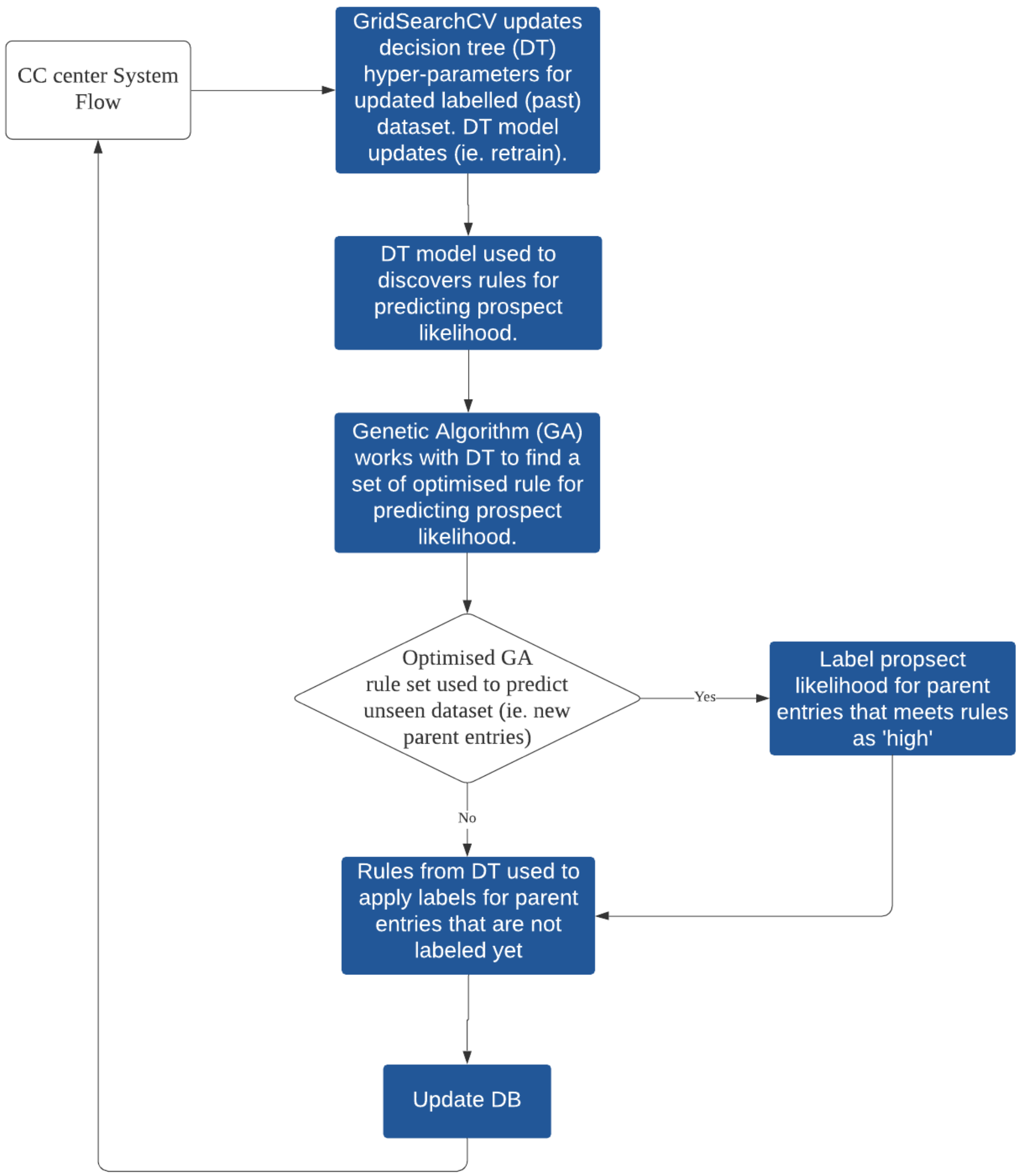


Figure 34: Rule discovery and rule optimisation performed over data

5 CONCLUSION

We have learnt a lot about real world challenges faced by parents and childcare centre operators during knowledge acquisition phase of this project.

Building the system introduced challenges. KIE Drools was new to our team and we need to learn how to integrate and use it with other system components. Working on this project allowed us to learn from each other who are experts in certain areas of development.

Overall, this system that we built is of great value for new parents and is a one stop place to get recommendations. It also provide child care centres with optimised waitlist of parent interests, which they can use to better optimise their resources on managing waitlists.

It was a challenging project, but our team had a very good time working together. We learnt a lot from each other, and we are grateful to each other and our lecturers for this project experience.

6 APPENDICES

APPENDIX A: Project proposal

GRADUATE CERTIFICATE: Intelligent Reasoning Systems (IRS)

PRACTICE MODULE: Project Proposal

Date of proposal: **July 11, 2020**
Project Title: ISS Project – Intelligent Childcare Matcher and Waitlist Optimizer
Sponsor/Client: *(Name, Address, Telephone No. and Contact Name)* Institute of Systems Science (ISS) at 25 Heng Mui Keng Terrace, Singapore NATIONAL UNIVERSITY OF SINGAPORE (NUS) Contact: Mr. GU ZHAN / Lecturer & Consultant Telephone No.: 65-6516 8021 Email: zhan.gu@nus.edu.sg
Background/Aims/Objectives: The proposed intelligent recommendation system aids for both the Parents in choosing the childcare centers also the Childcare center to optimize their high waitlist for admission. **As-Is Process** **Parents**: During the search for a suitable childcare center for their child, Parents do also often conduct a lot of online research on past parents' reviews on potential childcare centers. **Challenges for Parents:** Registering through ECDA website. This website allows parents to register up to 10 childcares, without allowing parents to rank the childcares. It provides only some information on the childcare centres such as levels offered by the childcare, the fees etc.

Childcare Centers: They assign parents to waitlist based on first come first serve basis for the tour of the Childcare center followed by admissions.

Challenges for Childcare Center:

A recurring problem among childcare centre operators today is the need to manage an inflated waitlist of prospective parents which does not reflect true demand. Even if Early Childhood Development Agency (ECDA), the regulatory and developmental agency for the early childhood sector, is able to coordinate the waitlists to reduce inflation issues, the problem remains that some parents may already have found alternative childcare methods and does not withdraw their applications.

To-Be Process

As the number of childcare centres and enrolments in Singapore increase, our system offers solutions toward these problems.

(I) Parents with childcare recommendations, along with sentiment analysis based on actual parents' reviews from our system. Parents can then rank and submit their interest on the childcares through our system.

(II) Childcare centers with an optimised waitlist of parent interests. which include information like likelihood of prospective parent taking up placements. Childcares can then better optimise their resources on managing waitlists.

Requirements Overview:

- Research – Requirements gathering and data gathering to understand the current flow of the process for both parents and child care centers. Interviews to be conducted for knowledge acquisition.
- Programming – Rule based and knowledge-based systems
- System integration – Integration of (JQuery, Javascript, HTML, CSS) with the backend database (MySQL) using Python Flask and Springboot

Resource Requirements (please list Hardware, Software and any other resources)

Hardware: Any typical CPU.

Software:

- Eclipse IDE for programming Springboot app
- PyCharm IDE for programming Python/Flask app
- KIE Drools Workbench for easy programming of KIE logic

Other resources:

- Internet access for linking to Google API

- Google Developer’s account for Google API key
- Google Colab for easy testing and programming of portions of the system including machine learning models e.g., Decision Tree Regressor with Genetic Algorithm

Number of Learner required: (Please specify their tasks if possible)

A team of **four** project members. The tasks completed are specified in their individual contribution reports

Methods and Standards:

Procedures	**Objective**	**Key Activities**
Requirement gathering and analysis	Information on ECDA process flows and ECDA data and how waitlist is being handled by Childcare centre operators	1. Gather & analyze requirements 2. Define internal and external design 3. Prioritize & consolidate requirements 4. Establish functional baseline
Technical Construction	System design - rule based and knowledge-based systems using reasoning systems, e.g. KIE jBPM, Drools, Fuzzy logic, Optimization (GA)	1. Setup development environment 2. Understanding of system context and design 3. Coding
Integration & Testing	To ensure interface compatibility and confirm that the integrated system software meets requirements and is ready to go live	1. Prepare system test document 2. Conduct system integration testing 3. Log exceptions

Delivery	To deploy the system and provide user manual for the execution of the system	1. Software packing and documentation as per ISS guidelines. 2. Submission of user manual for execution of the system

Team Formation & Registration

Team Name: **Team # 01 – Intelligent Childcare Matcher and Waitlist Optimizer**
Project Title (repeated): **Intelligent Childcare Matcher and Waitlist Optimizer**
System Name (if decided): **Intelligent Childcare Matcher and Waitlist Optimizer**
Team Member 1 Name: **ANG PAU HUANG, EDWIN**
Team Member 1 Matriculation Number: **A0195275U**
Team Member 1 Contact (Mobile/Email): **Mobile: 90660958** **Email: e0384906@u.nus.edu**
Team Member 2 Name: **CHEOK MEI LI**
Team Member 2 Matriculation Number: **A0165420N**
Team Member 2 Contact (Mobile/Email): **Mobile: 81219750** **Email: e0167294@u.nus.edu**

Team Member 3 Name: **PADMAPRIYA MATHIVANAN**
Team Member 3 Matriculation Number: **A0215281M**
Team Member 3 Contact (Mobile/Email): **Mobile: 94670442** **Email: e0535371@u.nus.edu**
Team Member 4 Name: **PRIYANSH MISHRA**
Team Member 4 Matriculation Number: **A0215340W**
Team Member 4 Contact (Mobile/Email): **Mobile: 84385004** **Email: priyansh@u.nus.edu**

For ISS Use Only		
Programme Name:	**Project No:**	**Learner Batch:**
Accepted/Rejected/KIV:		
Learners Assigned:		
Advisor Assigned: Contact: Mr. GU ZHAN / Lecturer & Consultant Telephone No.: 65-6516 8021 Email: zhan.gu@nus.edu.sg		

APPENDIX B: Mapped System Functionalities against knowledge, techniques and skills of modular courses: MR, RS, CGS

Courses	System Functionalities
Machine Reasoning	• Rules used for parent-facing recommendation system • KIE Drools built to accommodate the rules
Reasoning System	• Genetic algorithm and decision trees are used for prediction of likelihood of parent taking up a vacancy
Cognitive Systems	This portion is not demonstrated for our project, as virtual assistant and chatbots are not suitable for our scope.

Given the above system functionalities, we can meet these three requirements:

1. Business rule OR Business process OR Knowledge based reasoning techniques
2. Business resource optimization techniques: Search OR Constraint satisfaction OR Evolutionary computing
3. Knowledge Discovery using suitable data mining techniques

APPENDIX C: Installation and User Guide

This is available in this URL or in GitHub project IRS-PM-2020-11-01-ISY5001-GRP1-ChildcareMatcher/Project Report/Childcare Matcher User Guide.pdf

APPENDIX D: Individual reports

This is available in this URL or in GitHub project IRS-PM-2020-11-01-ISY5001-GRP1-ChildcareMatcher/Project Report

APPENDIX E: INTERVIEW OF THE CHILD CARE CENTER

Survey - National University Of Singapore - Project exploration Inbox ×

Priya Parthiban <padmaparthi04@gmail.com> Fri, Aug 7, 11:37 PM
to rosieta.littlegemspreschool

Greetings Ms. Rose
I'm doing a project in Intelligent systems with NUS.
As a part of our Project - I approach you to understand the challenges that puts off when it comes to managing your wait list in order to get parents enrolled in your institution.
We are in process of developing a system for Child Care center wherein they can better manage their wait list. To brief out on our system,

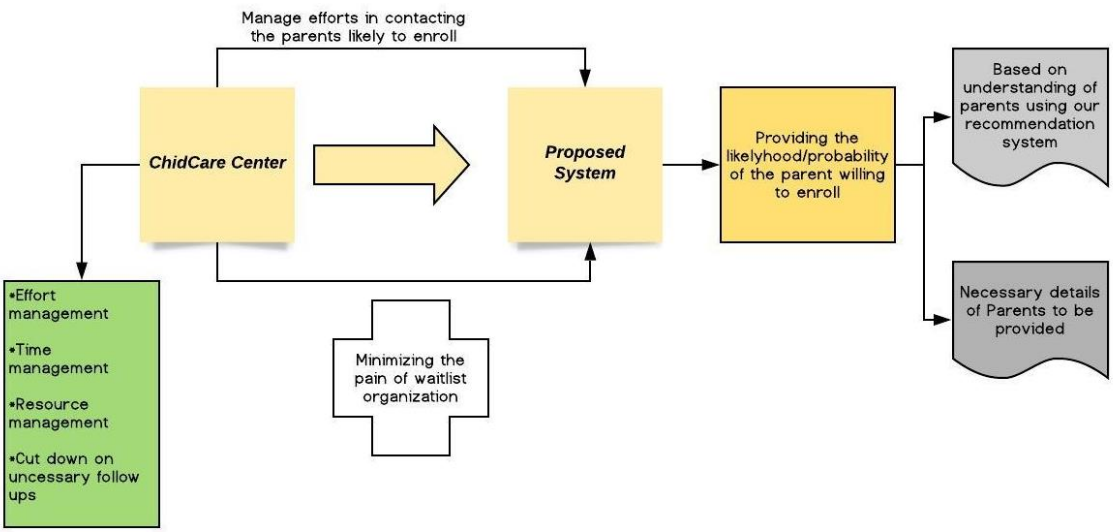
1. We have a initial Child care recommendation system where in the parents register and give their preferences/needs based on their inputs we recommend a list child care suitable for their needs.
2. Based on the inputs provided by parents we come up with a probability/likelihood factor that the parent may/may not be interested to enroll in the particular Child care center.
3. This probability factor and necessary details of the parent to be provided to you which enhances your chances of optimizing time to contact the parents who are more likely to enroll in your center.

Please find below basic flow of the process for your understanding.
Please feel free to let us know

1. If such a process can help reduce your challenges in contacting/follow up of parents ?
2. whether you would be interested to give our model a test try?
3. Challenges that you can share with us in managing wait list based on the current system ?

Thanks for the Time and Regards
Priya
Student - NUS -ISS

***Information in this email would be used for review and audit purposes for the project*



Rosieta Abdul Mutalip <rosieta.littlegemspreschool@gmail.com> Aug 13, 2020, 11:52 AM
to me

Dear Priya,

I would like to try out the new system if it is possible. It can probably help us in the future waiting list for parents. As of now, the challenges that we faced are time consuming as we need to have a virtual tour for parents to view the centre or they can only come down to have a tour when the children are not around in the evening or on Saturdays. They take time to get back to us whether they are interested or not. So it would be good if your system is better and let's see from there.

Looking forward to hearing from you.

Warm Regards,
Rosieta
Principal

LITTLE GEMS PRESCHOOL PTE LTD
3 Ang Mo Kio St 62 #01-09/10 Link@AMK S(569139)
T: 6253 7009 (AMK) HP(92217141)
W: www.littlegemspreschool.sg