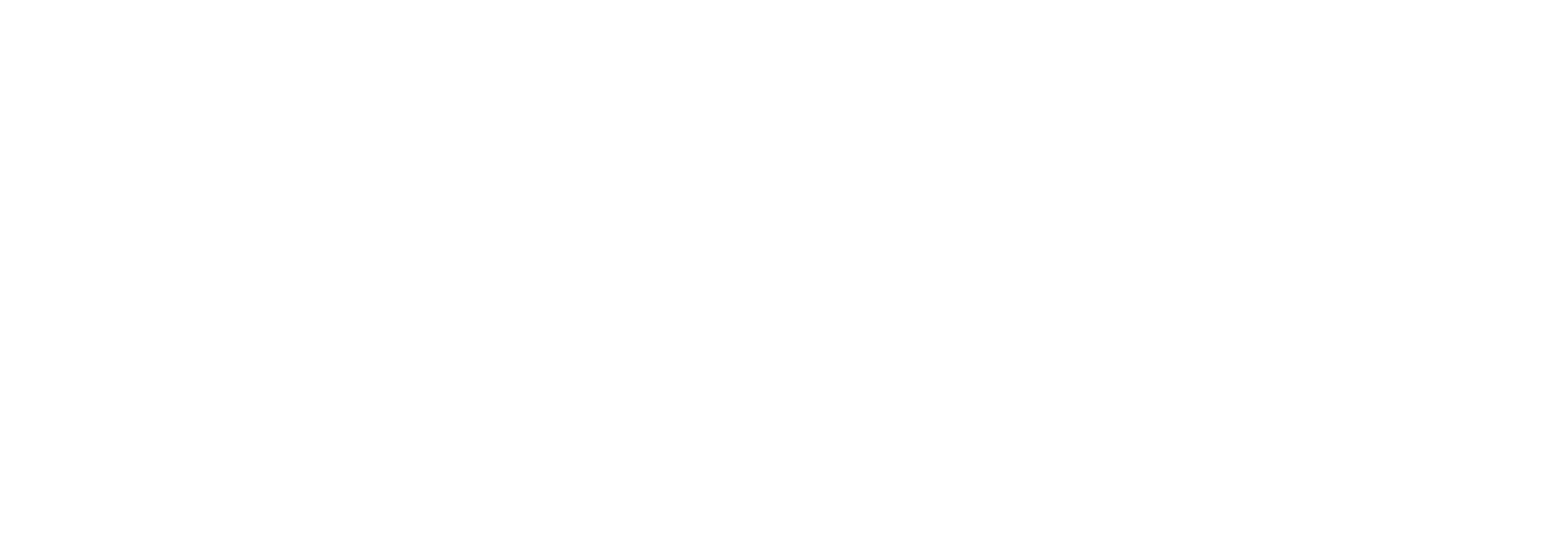
GovHack 2019

NICI Recommendation Engine



Document Information

|  |  |
| --- | --- |
| Document Owner | GovHack 2019 |
| Title | NICI |
| Description | NICI Recommendation Engine |

Table of Contents

[1. NICI Recommendation Engine 4](#_Toc18853028)

[1.1 Objective 4](#_Toc18853029)

[1.1.1 Issues 4](#_Toc18853030)

[1.1.2 Solution Summary 5](#_Toc18853031)

# NICI Recommendation Engine

## Objective

To give the best recommended action to prospective opportunity seekers, where “best” can change based on the requirements of NICI’s owner.

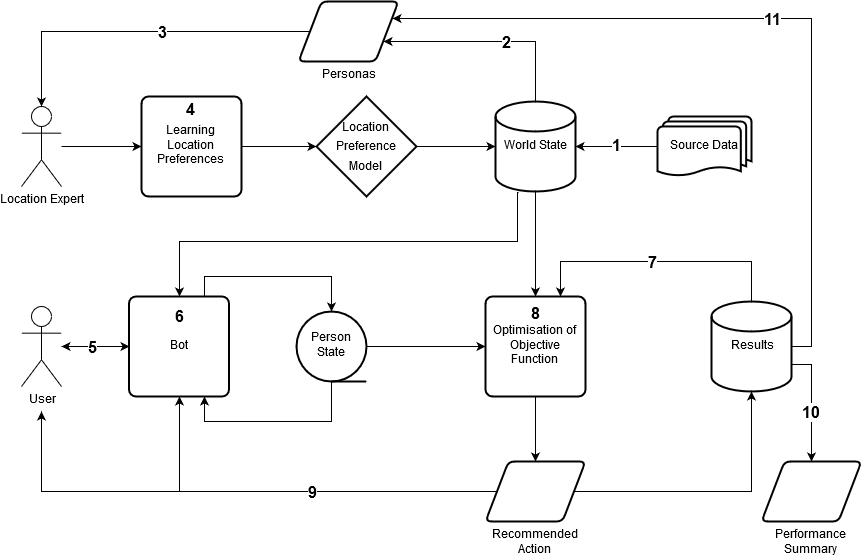
### Issues

1. Training  
   The way things are is not necessarily the way things should be, therefore it would be a mistake to blindly train a recommender based on similarity between the person and existing people in a location. This would discourage diversity (not just cultural diversity, but skill and age diversity as well). Likewise, it would be a mistake to blindly train a recommender based on recommending people most unlike the present population, which ignores community identity and geography/industry-driven demographic differences.   
   Instead NICI needs to give needs-based recommendations. For that purpose, NICI has a component which consults location stakeholders (innovation precinct and council representatives) and uses their needs to guide its training. Over time as more users engage with NICI and more follow-up data is available, NICI can improve its recommendations.
2. Conflicting Objectives  
   This problem naturally has conflicting objectives. Machine learning works by searching for the parametrisation within a given framework (e.g. weights within a neural network, coefficients within a regression-based approach, decision boundaries within a tree-based approach, etc) that best optimises some objective. In this case we have multiple objectives:
   1. Best addressing opportunity seeker needs and preferences in a residential/investment location
   2. Best addressing location needs in opportunity seekers
   3. Optimising success as defined by realised outcomes (be they economic, social, or general well-being measures)
   4. Being fair, in the sense that recommendations are distributed across locations

NICI is designed to allow the owner to specify the relative importance of these different objectives. (These four objectives are weighted into a single objective function, but the relative weighting influences the recommendations. For example, if fairness is most important, then this may result in individuals receiving recommendations that are neither suitable for them nor suitable for the location, simply because no one else has been recommended that location. Conversely, if fairness is not considered, then it may be the case that recommendations skew towards a single location.)

1. (Representation of Locations and Individuals  
   An additional machine learning challenge exists in representing locations and individuals. Because our ultimate recommendation is based on an objective function which comprises of similarity measures between location/personal needs and personal/location preferences, we must represent preferences and states as vectors that can be compared based on some metric in p-space (where p is the number of dimensions). Machine learning allows us to choose a representation that ensures clustering of similar people/locations.   
   For example, if a location is in need of Python data analysts, arguably R data analysts are quite similar people and (without other major differences) should be represented with similar vectors.)

### Solution Summary



1. Demographic, geospatial, economic, social and cultural data for locations are regularly ingested into the World State database, which records the characteristics of each location to which a user might be recommended opportunities. These locations can be regional areas, innovation precincts, or even specific businesses – any locations for which we can automate ETL jobs to gather data. These source data include census data, economic data, and other data as described by this project.
2. Based on census data ingested in step 1, a selection of personas are generated – these represent people who may use NICI to connect to opportunities. Over time these automatically generated personas are progressively supplemented by personas that are representative of users actually using the app.
3. Location experts (representatives of councils or innovation precincts) rate personas based on their need in the location. This generates initial training data for a machine learning element of the recommendation engine.
4. Using the training data developed by location experts, train a set of neural neural networks to score an opportunity seeker based on their need in each location. This is used in step 8 for determining recommended actions.
5. The user and NICI converse about the user’s intentions. NICI is also aware of the world state data and sources from these data to answer immediate enquiries or give additional information to the user when relevant.
6. During the conversation NICI updates a Person State that stores information about the user and guides NICI to ask for more information in different areas. This Person State is used as a set of features for recommending opportunities in step 8.
7. Past person state vectors and corresponding recommendations are read, as well as the outcomes of those recommendations, both specific outcomes (stored in the results database from follow-up contact from NICI) and general outcomes (changes to measures of success in the world state since the historic opportunity was taken). These are used as a set of features for recommending opportunities in step 8.
8. An objective function is optimised across the space of locations, which is a weighted average the following (where weights are based on the custodian of NICI’s business goals: how relatively important each aspect is):
   1. Score of the person/investment in the location’s preferred person/investment model (from step 4)
   2. The similarity of the location’s characteristics to the person’s preferred location (from step 7)
   3. The fairness, which penalises locations that have been overrepresented in past recommendations.
   4. The success of similar recommendations in the past.
9. Based on the best matches found in step 8, NICI returns (either through the chatbot interface or directly to the user) recommended actions: e.g. consider investing in information technology in Whyalla, or consider applying for a much-needed electrician role in Ceduna. The record of the user and NICI’s recommended actions is archived in a Results database.
10. A performance summary extract is able to be produced from the Results database for assessment by NICI’s owner.
11. For subsequent training of location preference models, representative users of NICI are extracted from the results.

The World State is a database connected to source data that describes the locations that NICI can recommend for investment/residence. In order to initially determine

On the back-end NICI maintains vectors to represent:

1. Person State - the person having the conversation, which is updated as the conversation proceeds. This includes three subvectors: the person’s characteristics, the person’s investment preferences (if applicable – the person is considering investing), the person’s location preferences (if applicable – the person is considering moving or other opportunities)
2. The confidence with which NICI knows the person vector, to guide the continuing conversation, which is updated as the conversation proceeds.
3. World State - which is a set of vectors representing the locations where people can move or invest, set before the conversation begins. Each location vector includes three subvectors: the characteristics of the location, the location’s incoming mover preferences, the location’s investment preferences.
4. The confidence with which the world state is known, which is set before the conversation begins.
5. If applicable, the person’s investment preferences.

When the person prompts an end to the conversation, NICI makes the following comparisons:

1. The person’s characteristic vector with the world preference vectors, to find which locations would be most benefited by that person moving/investing,
2. The person’s preference vector with the world characteristic vectors, to find which locations the person would most be benefited moving to/investing in,
3. Historic recommendations, to factor in “fairness” – so that no regional area has a disproportionate representation of recommendations,
4. Success of similar historic recommendations, where success is collected both by follow up from NICI directly and from measurement of additional open data

Problems:

* How does NICI know the possible actions that a person can take?
  + Just limit to move/invest to start with
* How do we create appropriate world state vectors?
  + Thoughts below
* How do we create appropriate person state vectors?
  + Thoughts below
* How do we make recommendations?
  + Use an objective

Ideas:

1. Training of world preferences: creation of personas based on census data and past , representatives of innovation hubs choose the person and investment profiles that best match their objectives.

Models:

1. Person state vectors:
   1. Location Preference Feature Vector:
      1. Interest in moving (trained based on migration statistics and feedback from NICI follow-up)
      2. Interest in diversity (imputed based on NICI questions)
      3. Location preferences (coast/inland quotient, diversity preferences) (imputed based on NICI questions)
         1. Coast/Inland Quotient
         2. Diversity Preferences
         3. Location Size Preferences
      4. Family requirements (imputed based on NICI questions) (Location penalty)
      5. Industry preferences (traversal of ANZSIC based on NICI questions)
      6. etc
   2. Person Vector:
      1. Skill profile (informal traversal of ANZSCO)
      2. Education Profile (informal traversal of ASCED(?))
      3. Cultural Profile (language identification, CAREFUL!)
      4. Disposition Profile (sentiment-style analysis of conversation)
      5. English Proficiency (analysis of conversation)
      6. etc
   3. etc
2. Investor state vectors:
   1. Opportunity Preference Feature Vector:
      1. Industry preferences
3. World state model
   1. Person Preference Feature Vector
      1. Based on training on personas from location representatives (personas derived from census data, augmented over time from users of the app)
   2. Investor Opportunity Preference Feature Vector:
      1. Business profile (based
   3. Location Feature Vector
      1. Cultural profile (from data)
      2. Business profile (from data)
      3. Economic profile (from data)
4. Matching Objective Function made from:
   1. How close does the location align with the person’s preferences?
   2. How close does the person align with the location’s needs?
   3. How fair has the allocation of people to different areas been in the past?
   4. How successful has this recommendation been in the past?

Goal:

* Recommend actions related to:
  + Goals of the individuals
  + Goals of the “world”
  + Objective Function

State of the world <- updated by datasets

State of the person <- updated by directed conversation

Recommended actions <- possible actions are defined (move here, invest here)