

Real Estate Location Analytics and Prototype Design Using Spatial Mining

Prof. Neeraj Bhargava¹, Vaibhav Khanna²

¹Department of Computer Science, School of Engineering and Systems Sciences,
MDS University, Ajmer, Rajasthan, India.

E-mail: profneerajbhargava@gmail.com,

²Department of Computer Science, Dezyne E Col College, Civil Lines,
Ajmer, Rajasthan, India

E-mail: isaacajmer@gmail.com,

ABSTRACT

Finding a perfect residential real estate is a tough and arduous process. This is a multi-criterion decision making situation with several decisive factors. The most important amidst these is the approachability amid the residence location and the points of preference or interest that the user has in his priorities. These points of interests may be service-places, shops, business locations and leisure facilities. For assessing the real estate suitability based on multiple analysis criterion this research paper has incorporated the concept of spatial neighbourhoods into our spatial methods for value analysis.

Key Words : Location analytics, geo app, web maps, real estate valuation, spatial data mining

I INTRODUCTION

Location analytics and Spatial data mining is grounded on the core concept of web maps which are the basis of spatial data science and web GIS. The web map is a platform to display and work with spatial and non-spatial data layers. Using the location analytics toolkit of ArcGIS developers can configure these data layers, style the map for cartography, and can also change the way that users interact with the map.

Web GIS is highly useful in that way that once the application is ready developers can let the users interact with the layers and use the app for location analytics. The final work product is an analytics application which the users can use as a web-map-app. On a simple touch it shall provide an immediate location specific analysis about the data in a visual format. This can be supplemented by visual charts or graphs to assist data related insights.

The web-map is the first building block for Right Valuation Geo App. This application is a Location Analytics Prototype for user assessment of residential Real Estate. The application has been developed using ESRI ArcGIS. Map-Viewer is the platform environment used for development and working on this Location analytics prototype. The Map-Viewer environment sets ESRI base-maps layers for context and subsequently add data layers to the base-maps layers. These base map layers are provided by ArcGIS Online work lab through "Living-atlas" which provides GIS datasets in the form of Map Layers and also demographic datasets compatible with these map layers. We can configure these layers and style the

map for further cartographic analysis. Much part of the analysis depends on configuring the layers and doing the analysis. The development work included pop up configurations using python scripts to facilitate the users in doing their analysis using the Right Valuation app. As a continuous work-product, the app creates informative maps and location specific data for the user, based on multidimensional decision criterion.

The valuation of a residential real estate location largely be contingent to the accessibility between the location and user defined points of interest which we have to access in our everyday life. These could be places of work, malls, grocery stores, dairy facilities, laundry facilities, support services, schools, amusement parks, recreation centres, medical facilities, bus stations, airports and so on.

Each user may have a different criterion for assessing the value of a real estate to judge the suitability or fitment to his needs. Another big decision-making parameter is the budget available or the price concern. The right valuation location analytics prototype uses spatial data mining and intelligent data retrieval for arriving at an informed decision. It has interaction technologies that can be used for cruising through a list of candidate real estate properties in an online internet enabled web app setup. The user can find a competent property through visual interaction with spatial as well as non-spatial data layers. This app can drastically reduce the decision-making time. Location analytics supports in visual comprehension of information, which in turn helps us in informed decision making. Location analytics uses spatial data and quantitative spatial models to support decision-making, including location analysis. [1]. Right decisions require ‘insight beyond the algorithmic analysis’ and right association mining of multidimensional data is highly important for right decision making.

II RELATED WORK

Looking for a real estate is integrally a multidimensional spatial decision. The choice is principally grounded on three high-level measures comprising of domestic needs, estate facilities, and location analytics. Location decision is determined by varied features; including but not limited to eco-friendly factors, access, association to services, and the social and economic demographics of a real estate. Peter Keenan et.al in their research article Gap analysis in decision support systems for real-estate in the era of the digital earth, purposes to recognize the gap amid concept and practice in location analytics. The authors conducted a gap analysis and identified a set of seven contributing factors for location analytics of household real estates. [2]

Residential real estate is very significant division of the economy since it has very noteworthy influence on all areas of exchange transactions in all nations. There is a deficiency of unified and complete data about location analytics related to real estate pricing methods and purchase decision making. [3]

Makariye N. (2017) in IEEE paper titled, “Towards shortest path computation using Dijkstra algorithm” used Dijkstra Algorithm and Spatial clustering of current traffic data to arrive at shortest path computation. In modern car navigation structures, best path discovery is highly important for commuting between two desired points on the road map [4]. The best path may be defined differently by different users of the system. For some the definition of best may be the ‘least time to travel’, for others it may be ‘shortest distance’ or sometimes least ‘traffic density.’

Idri Abdelfattah, MariyamOukarfi, et al (2017) in their paper A new time-dependent shortest path algorithm have presented their work on Dijkstra's Algorithm for time dependent multimodal transport network. [5]

The best route computation and the substitute route are calculated grounded on objective function of current traffic conditions and road traffic density. This is dependent on live data about the traffic on that particular route. The algorithm must compute the alternative paths based on multiple criterion for computation of best route in a certain situation. The dual objectives of saving time and saving cost need to be synchronised with the flow of real time live data [6].

Previous researches illustrated the importance and the usefulness of feature selection. Recursive Feature Exclusion for SVM is utilised and features may be associated to select an finest subset of features. [7]

The “find closest facility problem” requires spatial data mining on multidimensional datasets. Finding what is nearby and then discovering the best route to the desired destination is an everyday requirement. The best of the real estate properties are those that have already conducted this type of analysis and are ranked on the basis of such score. [8]

The Service area algorithms set is required to calculate a logical service buffer around the point of interest. The service area calculation becomes more complex in case of networked location analytics. The computation must include all options of accessible streets that qualify on the computation criterion. For example if we want to computer 20 minutes service area for some desired real estate, then the algorithm should be able to find associations with desired facility set that can be reached in 20 minutes drive time. This will change based on traffic conditions and the rush hour associated with office hous and heavy slow-moving traffic[9].

Calculation of network impedance and ranking of network nodes problem is addressed by this set of algorithms. Sometimes we may need calculations from numerous origins to numerous destinations. The one to one relationship may transform to one to may or many to many relationship. In this case we have to compute the origin destination cost matrix. [10] This kind of analysis is useful if the competing real estate properties are located close to each other and the location analytics prototype is delivering similar computations for several competing sites. In such situations we take a paired approach to computation of best path. For each pair we do this computation and store this in the base attribute table. [11]

Several location analysis models are suggested in network coverage analysis. Network Analyst provides a set of algorithms that can be used to conduct experiments for vehicle routing problem. The alternate solutions can be evaluated for merits on several parameters so that an amenable solution can be identified. [12]

Location-allocation algorithm can be used to conduct experiments for making a useful selection of facilities grounded on their possible interface with desired points of interest. The goal sometimes is to minimize the total distance amongst demand points and amenities. Or the objective function may be to make best use of the number of point of interest covered in the interior zone of a certain distance of amenities, or to discover environment of friendly residential properties. [13]

Time-dependent analysis: Several experiments demand incorporation of live or historical data into spatial analysis. Using this set of algorithms we can integrate live and past traffic dataset into an investigation so as to facilitate discovery of the best route at a specific time of day. [14]

Data Analytics for Intelligent Transportation Systems offer exhaustive exposure to approaches for analysing smart conveyance. [6]

III METHODOLOGY:IMPLEMENTING THE BASE LAYERS

The computation of suitable real estate on the base of competing dimensions and then analysing it with the where component of data is the core functionality of right valuation application. The feature dataset as well as the demographic variables are generally not precise. A fuzzy approach to the location data analytics seems to be the only option. Fuzzification of the Dijkstra algorithm provided the best outputs in the experimental datasets. Q was taken as the bag of all not visited apex nodes which were to be pruned.

In bag Q of this object oriented dataset, u epitomizes the node with the minimum distance from the foundation node.

v was taken as the variant distance from the source and was used to compute the shortest distance, and distance amidst (x, y). This was to compute the detachment amongst two vertices x, y.

previous[v] stored the adjacent vertex to the existing v lengthwise from the source node.

The algorithm used substitute modified distance for assessment.

Similar to the Dijkstra algorithm the preliminary state, for each vertex node v in graph was assumed as infinity.

vertexnode[v]=infinity

previousnode[v]=0 and

Q is the bag of all nodes in graph.

For the source node, dist[source]=0.

Then the following three steps are repeated till completion.

Step 1.

Check if Q is empty.

While Q is not empty, seek x in Q.

If dist[x] is not infinity, then from Q, x is removed

(this suggested that all residual vertices remain unreachable from source which in turn suggested that the analysis is finalized).

Step 2.

After finding v, each neighbor node of u, and calculate $alt = dist[u] + dist_{between}(u, v)$ by using the canonical representation operation

If there exists $alt < dist[v]$,

then replace dist[v] with alt and record the previousnode [v] = u

(if not, the previous record of each neighbor vertex v remains the same).

Step 3.

Choose the one among each v with shortest dist[v],

and replace it with u and go to Step 1.



The algorithm of the proposed FuzzyDijkstra algorithm is presented below:

Method FuzzyDijkstra (Graph, source):

for

```
every vertex v in Graph:           // This initialize sequence for each vertex
    distance[v]:= infinitude;      // Fuzzy time-cost function for computation
    previousnode[v]:= indeterminate; // feature suitability for best path
end for
```

distance[sourcenode]:= 0;

```
Q:= the bag of all nodes           // Complete node set
                                   // Wholly in bag initially
```

While

Q is not unfilled:

x := the vertex in Q corresponding to the minimum distance[];

if distance[x] = infinitude:

```
Break;           // residual vertices are unreachable from source
end if;
```

eliminate x from the bag denoted by Q

```
compute for every nearby node y of x:      // non eliminated nodes only
    alternative_node := distance [x] + distanceBetween (x, y);
    // this provides grouped_graded average combination representation
    // addition of fuzzy numbers results in a crisp number
    if alternative_node< distance [v]:      // Diminish existing nodes
        distance [v] := alternative_node;
        previousnode [v] := x;
        removeyfrom the bag Q;           // eliminating y
    end if
end for;
```

end while;

return to calling routine the final computed distance ();

end FuzzyDijkstra.

To create a new web map, click the New Map button in ArcGIS online cloud based platform. This brings to the map viewer application in ArcGIS. A web map in its simplest form is a reference basemap derived from Living Atlas spatial database with predefined data layers. Next start adding data to the map. The Add button functionality available in cloud

implementation of ArcGIS provides different options. For conducting location analytics experiments the Right Valuation Geo App used Map Layers from ArcGIS Online provided by Living atlas base-map shape-files. To this location data layer was added the Living Atlas demographic content and feature set layer. This is curated, authoritative content dataset provided by ESRI to conduct location analytic experiments. The dataset layers and feature sets are categorized for ease of search. For analysis of the where component the application used shapefile and dataset layers were added as CSV files. In this application we have also created another multidimensional features as spreadsheets in CSV file format. This file shows the available residential real estate locations along with locations of several other related parameters. It also contains the geographic coordinates of the locations. The data layers available in these CSV files can be implemented directly in ArcGIS. Any spreadsheet with geographically related columns will be shown onto the map automatically when added to base map layers. The file includes addresses and latitude and longitude coordinates and also has the price and size of each residential location available for sale.

IV CONCEPTS AND TASK ABSTRACTION

Right Valuation is a location analytics system, premeditated for real estate prospects who intend to buy or rent villas houses, flats, apartments, or condominium cottages in resort like setups. We denote these individuals as the **users** of this prototype- system. The act of discovering an ideal home requires a lot of decision making on multidimensional decision parameters. The alternative set is a list of available residential properties and the users need to make a choice from a list of offered residential properties. These home-properties and the where component associated to them as **contentenders**. The application depicts the fundamental issues in residential real estate selection process.

Reachability defines a bag of points of interest approachability to alternative points of interest on the road network. The application conducts temporal data analysis for mining time dependent patterns in the temporal traffic dataset. The algorithm computes accessibility by some routes under the time constriction.

Accessible section encompasses map locations for which the desired threshold value provided by the user is not exceeded. This kind of edge computation assesses the accessibility.

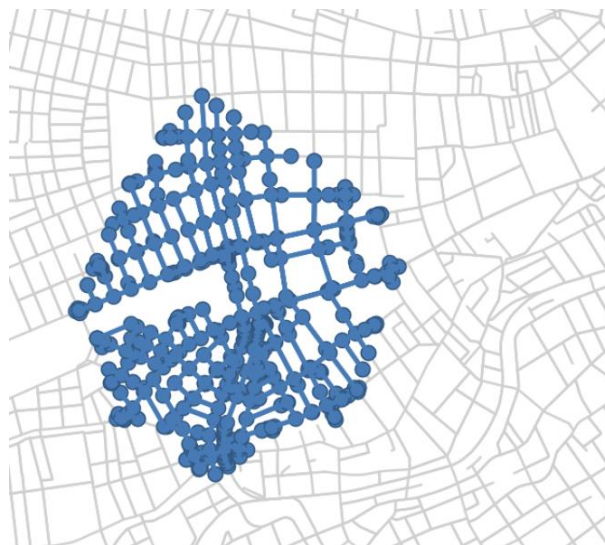


Figure 1 Reachability and road network layer

An activity describes a accessibility request, which encompasses a bag of origins and a set of destinations with computations related to departure time, arrival time and reachability threshold.

The inherent properties for all contending residential properties such as super built-up area, covered area, price, usable carpet area, number of bedrooms, bathrooms, balconies, locality index, neighbourhood index, community services, amenities and play areas etc are decisive factors. A fuzzy decision-making approach is used to accommodate all these and several other decision variable which the user perceives as important. The system uses a neural network approach to assign a weight to a given parameter and increase its priority for customised decision making. Different users may assign different weights to different features associated to the property in the suitability or fitment analysis.

V DATA DESCRIPTION

A web map in its simplest form is a reference base map for projection of other data layers. We have used the ESRI Living Atlas content, which is curated, authoritative content, also categorized for ease of search. For instance, the Living Atlas contains a World Airports layer which we have used to calculate the distance from the airport. This layer is all about the airport locations to be used in computations of suitability of real estate for a user. The treatment is the same for any and all activities on which we would like to conduct computations. The needs and want of application users such as restaurants and errands, shopping centre with GPS coordinates was added as CSV datasets.

Residential properties data include candidates that were available for sale. Each of these real estate embraces numerous spatial and fundamental properties, including geographic positioning, super built-up area, covered area, price, usable carpet area, number of bedrooms, bathrooms, balconies, locality index, neighbourhood index, community services, amenities, play areas etc and price.

Living Atlas Road network data encompass vector dataset in the form of directed graph. When analysis scripts are run this directed graph it can compute values for the whole city.

Once the application has data layers ready, the user can use the analysis tools available within the map viewer. There are different analysis operations available such as drive time analysis, traffic analysis, or zones that can be covered within a 30-minute drive time. The user may also use the Proximity option to get to the Create Drive-Time Areas analysis tool. The system permits the user to set the time to a desired value and see which locations fall within a desired drive time.

VITASK ANALYSIS

User requirement analysis and understanding plays a very important role for suitability and proper usage of the system. We conducted open ended structured discussions starting with the base question, "How would you describe an ideal house." These discussions were conducted with a set of 50 users to develop a better understanding of the user expectations. The user requirements and expectations from the software system are summarized as tasks given below:

Compose reachability limitations interactively: The system should be capable of articulating and unifying reachability constraints to accommodate the proposed daily routines. The system prototype calculates the proximity factor for actions and create a data response of properties that suffice the stated restraints. The features advantages and benefits of the application may be summarized as:

- Visual presentation of computed reachability and interactive interface to conduct analysis.

- Flexibility to compute several “what if” considerations and improve the solutions through visible interface
- Flexibility to change the routine and redefine decision parameters.
- Interactive management of the constraints
- Screen and assign ranking to the contending real estate properties.

The observed research gap is that this type of multi criterion analysis of existing location attributes is not available in the current real estate analytics systems. [15] Different users have different perceptions for what would make a perfect home. For a few real estate buyers price of the property is of utmost importance. For some others the nearness to the desired set of facilities and amenities is having higher priority or weightage.

The Right portrayal of user and analytical requirements is the basis of right valuation for any real estate property. For solving this problem, we designed “Right-Valuation Location Analytics Prototype”. This real estate visualisation system uses location analytics and routing algorithms for visual analysis. The system can help people in discovering the right properties from a list of available properties in the city. The residential real estate decision making is based on multiple criteria and uses MCDM modules of business analytics.

VII VISUAL DESIGN AND DEVELOPMENT

For implementing the Right Valuation Location analytics prototype we start with a basemap. The basemap provides context to the analysis map. There's options such as the Topographic, Imagery, Streets, but there's also two canvas options. The Light Gray Canvas highlights the attribute data and the thematic data by creating a really neutral backdrop.

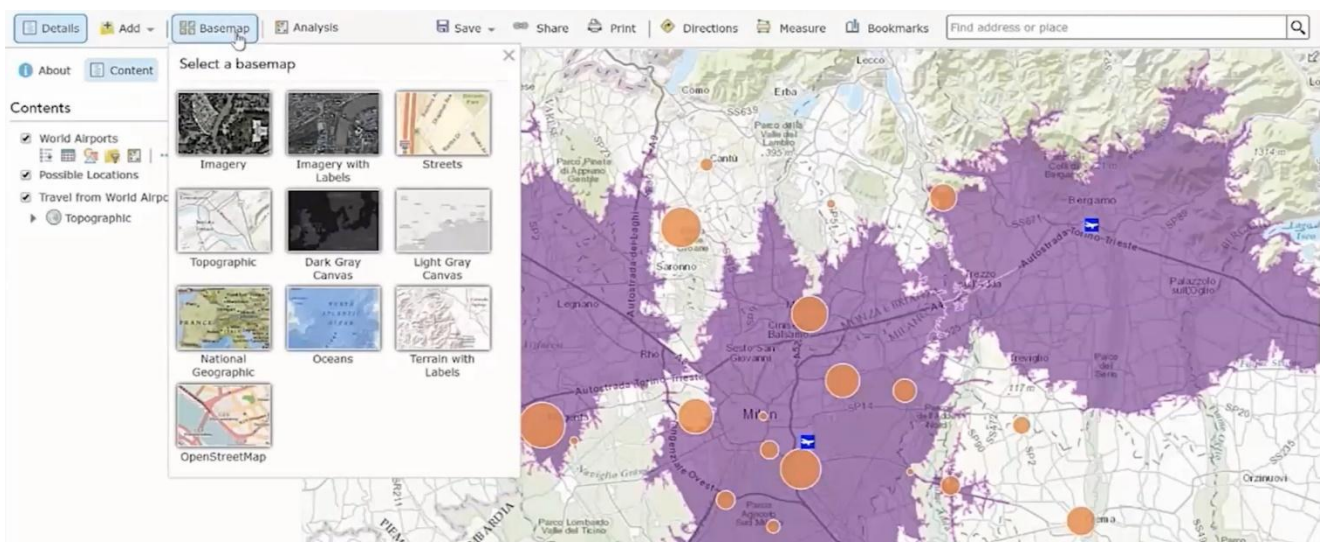


Figure 2 Setting the platform environment and base data layers

Once we have set the context with a basemap, we may now change the way the layers look within the map. The Change Style option brings me to the Smart Mapping interface. Smart Mapping is an interactive tool that intuitively makes suggestions on customisation of the analytics parameters and also as to how a map can be styled based on the characteristics of the data. The data is added in the form of a CSV file and the layers of data are synchronised to match

with the map layers. There is a whole range of options available to organise and present the information and data layers. The user can even visualize comparisons using choropleth maps, charts, bar graphs, and many other similar statistical tools.

The smart map script can then then organise the sites according to the price or any other attribute that the user may require. These things not only help me create a beautiful map, but also help us in exploring patterns within data that might not have seen otherwise. In figure 3 above the Smart Mapping organised the locations on the map and styled my map by price represented through the size of dot.

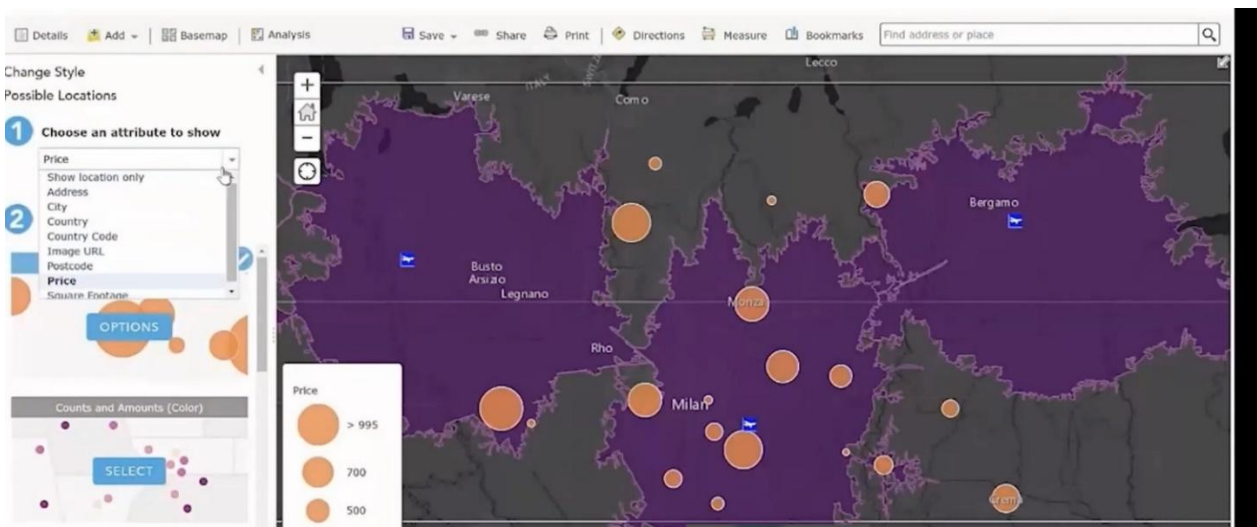


Figure 3 Conducting Analysis

The final thing we will do is to the map interactive. When the users select a location they're considering, they can immediately get a comprehensive pop-up about that location. Within Locations layer, we can select the Configure Pop-up option. I want the title to show the address and the text to show the price and size of each location. I do this by configuring a custom attribute display. I can also include a URL to a picture of each location so that the users can see exactly what each place looks like.

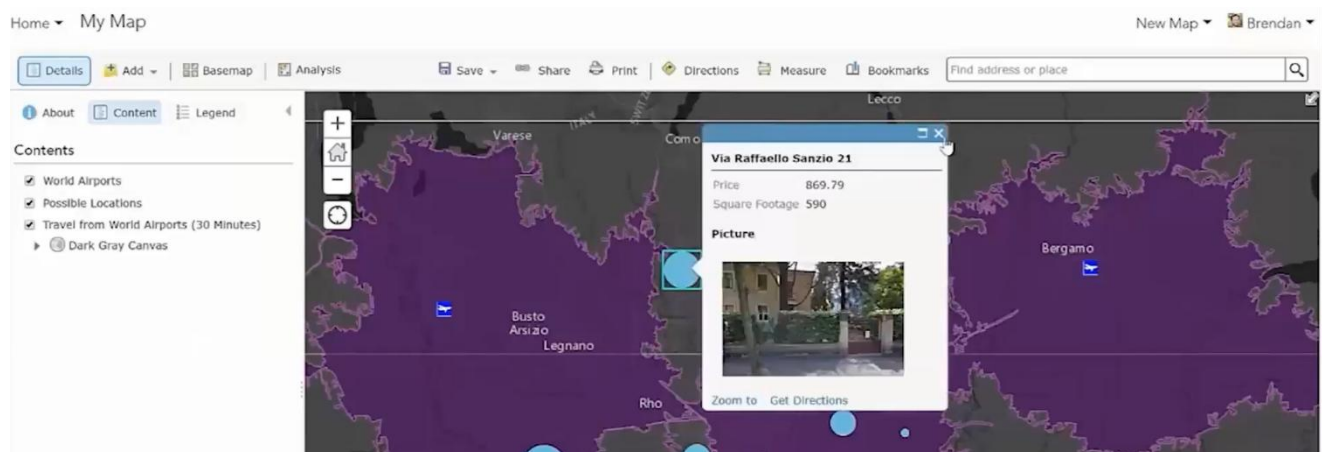


Figure 4 Interactive summary of location which the user is considering to purchase.

Now that we have all my configurations set, we save the system as a web map and put it in a geo app.

The user can not only make use of the existing data but can also use predictive analysis for arriving at the right valuation. The word "prediction" inherently implies something about the future. But it can also be used in a lot of other ways, for example, fill missing values to ensure that you have a complete dataset. Sometimes we want to predict continuous values. In that case, we get into the world of regression analysis. Other times we want to predict categorical variables, and then we're getting into classification. In real estate forecasting, and with all of these use cases, prediction starts to feel like the pinnacle of analytics. We have used linear regression and logistic regression in predicting the real estate valuation.

Also, there are a lot of different ready to use algorithms in ArcGIS to do prediction. We've got inherently spatial methods, like interpolation and geographically weighted regression, and then we've got the more traditional machine learning approaches, like random forest.

VIII CONCLUSION

For assessing the right valuation of a residential real estate, we have developed a spatial data mining and location analytics system. The Association mining of features and attributes is conducted on the basis of this idea, "things that are closer together are more related than things that are farther apart." (Tobler's law). Building that the concept of a spatial relationship into the algorithms has led to huge improvements in decision making time. For assessing the real estate suitability based on multiple analysis criterion we have incorporated the concept of spatial neighbourhoods into our spatial methods for value analysis. The complete system is packaged in the form of a web app that can run on multiple mobile devices and laptops.

REFERENCES

- [1] P. Keenan, "Geographic Information Systems and Location Analytics for Business and Management," 3 December 2020. [Online]. Available: <https://oxfordre.com/business/view/10.1093/acrefore/9780190224851.001.0001/acrefore-9780190224851-e-200..>
- [2] H. R. Dastjerdi, G. McArdle, S. A. Matthews and P. Keenan, "Gap analysis in decision support systems for real-estate in the era of the digital earth," *International Journal of Digital Earth*, p. doi/abs/10.1080/17538947.2020.1808719, 2020.
- [3] McKinsey Global Institute, "The Age of Analytics: Competing In A Data-Driven World," *McKinsey Global Institute Series*, pp. 35-60, 2019.
- [4] N. Makariye, "Towards shortest path computation using Dijkstra algorithm,," pp. vol1, 46-55, 2017.
- [5] A. Idri, "A new time-dependent shortest path algorithm for multimodal transportation network," *Elsevier International Conference on Ambient Systems, Networks and technologies ANT*, pp. CS 109 C, 692-697, 2017.
- [6] M. Chowdhury, A. Amy and D. Kakan, *Data Analytics for Intelligent Transportation Systems*, Elsevier, 2017.
- [7] H. J. Yee, C. Y. Ting and C. C. Ho, "Optimal Geospatial Features for Sales Analytics," in *Proceedings of the 3rd*

- [8] Y. Huang, J. Pei and H. Xiong, "Mining co-location patterns with rare events from spatial data sets," *Geoinformatica*, pp. 10(3), 239–260. , 2006.
- [9] A. M. MacEachren, M. Wachowicz, R. Edsall, D. Haug and R. Masters, "Constructing knowledge from multivariate spatiotemporal data: Integrating spatial visualization with knowledge discovery in database methods," *International Journal of Spatial Information Science*, pp. 13(4), 311–334. , 1999.
- [10] D. Li, S. Wang and H. Yuan, "Software and applications of spatial data mining," *WIREs Data Mining Knowl Discov*, 6: 84–114, p. vol 6: 84–114, 2016.
- [11] S. Shekhar, F. Steven K. and A. Walid G., "Spatial computing: Accomplishments, Opportunities, and Research Needs," *ACM: Communications of the ACM*, pp. Volume 59 Issue 1, 33- 48, 2016.
- [12] H. Momm, G. Easson and J. Kuszmaul, "Evaluation of the use of spectral and textural information by an evolutionary algorithm for multi-spectral imagery classification," *Computers, Environment and Urban Systems*, , 33(6), 463–471, 2009.
- [13] ESRI White Paper, "Location Analytics as a Service," *ESRI Software as a Service*, pp. 12-27, 2017.
- [14] I. E. Putra and K. Rohendi, "Implementation of Geographic Information System with Dijkstra Algorithm Base On Mobile Application: A Model for Disaster Risk Evacuation Route in Padang City Indonesia," *Proceedings of the 2017 International Conference on E-commerce, E-Business and E-Government*, pp. 30-34, 2017.
- [15] Oracle Corporation, "Value of Spatial Analytics in BI," *Oracle Technetwork Whitepaper*, pp. 22-30, 2019.