## Chapter 2: Literature Review

### 2.1 Chapter Overview

The Introduction chapter provided a brief description about the project and its background related with problem domain and solution. This chapter will contain a thorough literature study on apartments, limitations on current apartment prediction applications and different researches based on apartment filtering and searching. The aim is to recognize the technological gap between the current system and the proposed system, and the most optimal machine learning approach to be used for this project to achieve the proposed outcome. It illustrates the problem domain and elaborates about the approaches, techniques and algorithms that can be used to solve the problem. A depth review is conducted to choose the best approaches and techniques. Finally, all existing solutions are compared with the proposed solution.

**2.2 Problem Domain**

This project will be addressing issues regarding finding apartments that is available in the Apartment selling websites. Only the Apartments in Colombo is being considered in proposed project.

One of the most common problem that we found was that most of the people buy/rent an apartment with minimal background knowledge and sometimes they invest a large amount of money for an apartment that might not have good external nor internal factors but it’s the first apartment that they found when they searched. When it comes to advertising apartments for the people to buy/rent the sources people use to choose from are mostly newspaper ads and websites. When it comes to these advertisements, they are basically providing one dimensional information. Most of us desire for a good investment and sometimes we succeed in good investing and most of the times we fail.

According to traditional system in searching for an apartment people investigate newspapers or contact third party brokers visit each site by site then filter out from many choices and select an apartment to live in. This legendary tradition has been followed by many in the world which is more absurd considering the advancement of technology, although the user views the site and select the apartment by first hand still, they get cheated by all other third-party false information without a proper analyzed data. As of the revolution of technology people they started heading internet in search of apartments.

According to a joint study made by Apartment.com and Google around 78% of adults in America look for information online about a service or product they are thinking about buying not only that but around nearly 90% of home buyers search online at some point in the buying process ( Apartments.com and Google , 2018). This trend is just not only found in America but also to be found by almost half of the population on Earth.

Technology have been playing a major role in our day to day life having a major impact, with this technology advancement looking for an apartment via websites trend also have been developing drastically. Studies say that around 72% percent of developed countries turn to internet first when starting a search for an apartment where as others tend to follow the old tradition which leads to unnecessary time consumption and being cheated by third party brokers (The National Association of Realtors® and Google, 2012). The mail goal of our project is to stop such troubles being faced and to make the users readily use our application to find their perfect home in no time but providing them with their desired home according the users feed on our application.

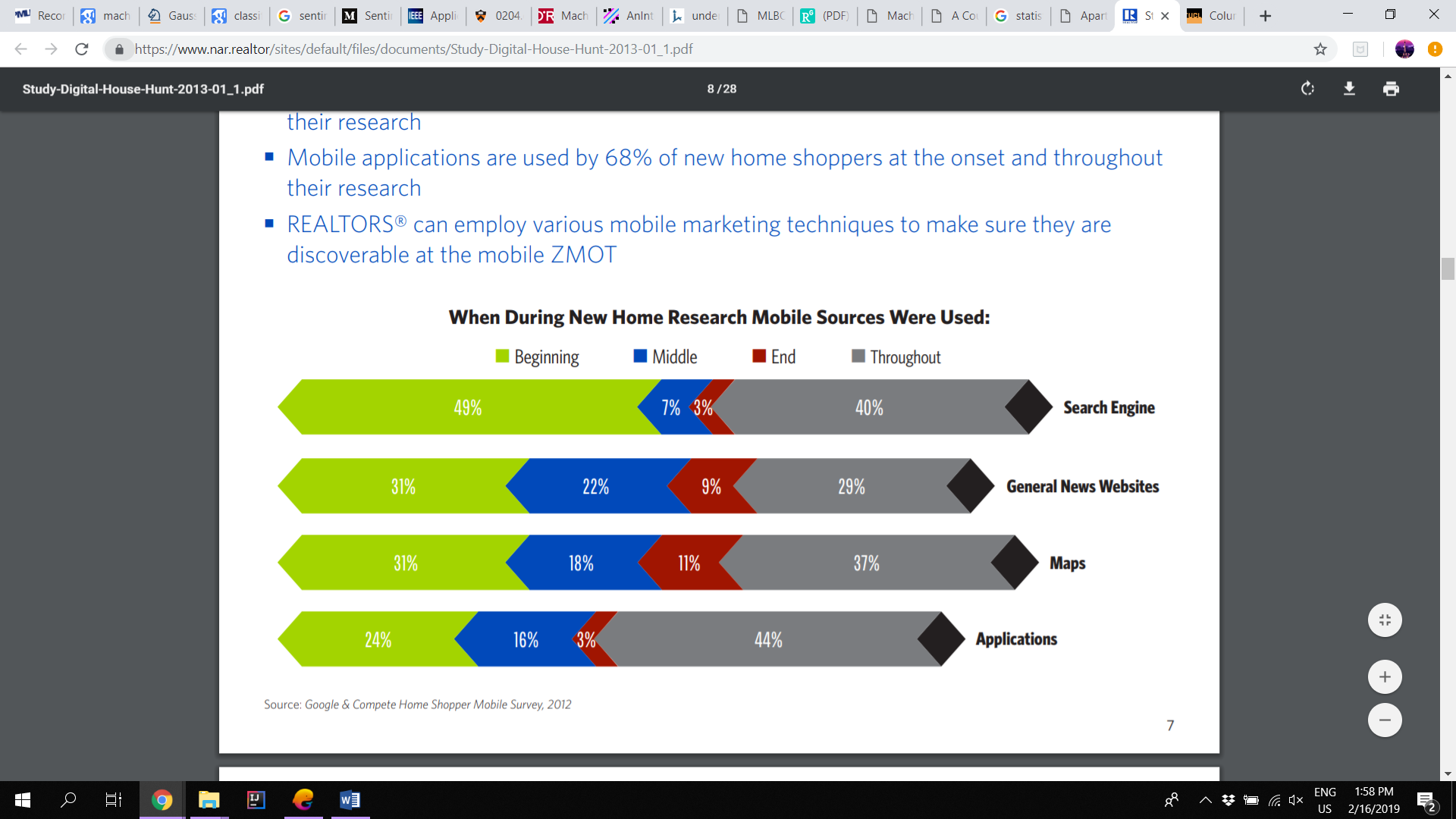


Figure 2.1: Survey on Mobile Application usage in search of homes (The National Association of Realtors® and Google, 2012)

**2.3 Machine Learning techniques for the system**

**2.3.1 What is Machine Learning?**

“The term machine learning refers to the automated detection of meaningful patterns in data” (Shai Shalev-Shwartz and Shai Ben-David, 2014). Also Aurelian Geron says that “Machine learning is the science (and art) of programming computers so they can learn from data,” in [*Hands-on Machine Learning with Scikit-Learn and TensorFlow*](https://www.safaribooksonline.com/library/view/hands-on-machine-learning/9781491962282/?utm_source=oreilly&utm_medium=newsite&utm_campaign=machine-learning-quick-and-simple-definition-body-text-cta)(Geron, 2017)*.*

Machine Learning is sub part of a wide area in Artificial Intelligence that “focuses on teaching the computers how to learn without the need to be programmed for specific tasks” (Chollet, 2017). The main key fact of Machine Learning is that to create algorithms that can learn from and make predictions based on data inputted.

**2.3.2 Machine Learning Categories**

According to Aurelian Geron (Hands on Machine Learning with Scikit Learn and TensorFlow) and Francois Chollet (Deep Learning with Python) the main four broad categories of Machine Learning are Supervised Learning, Unsupervised Learning, Semi- supervised Learning and Reinforcement Learning. (Chollet, 2017) (Geron, 2017)

The below diagram from A Web Developer’s Guide to Machine Learning in JavaScript by Robin Wierruch clearly shows the hierarchical order of Machine Learning with its categories.

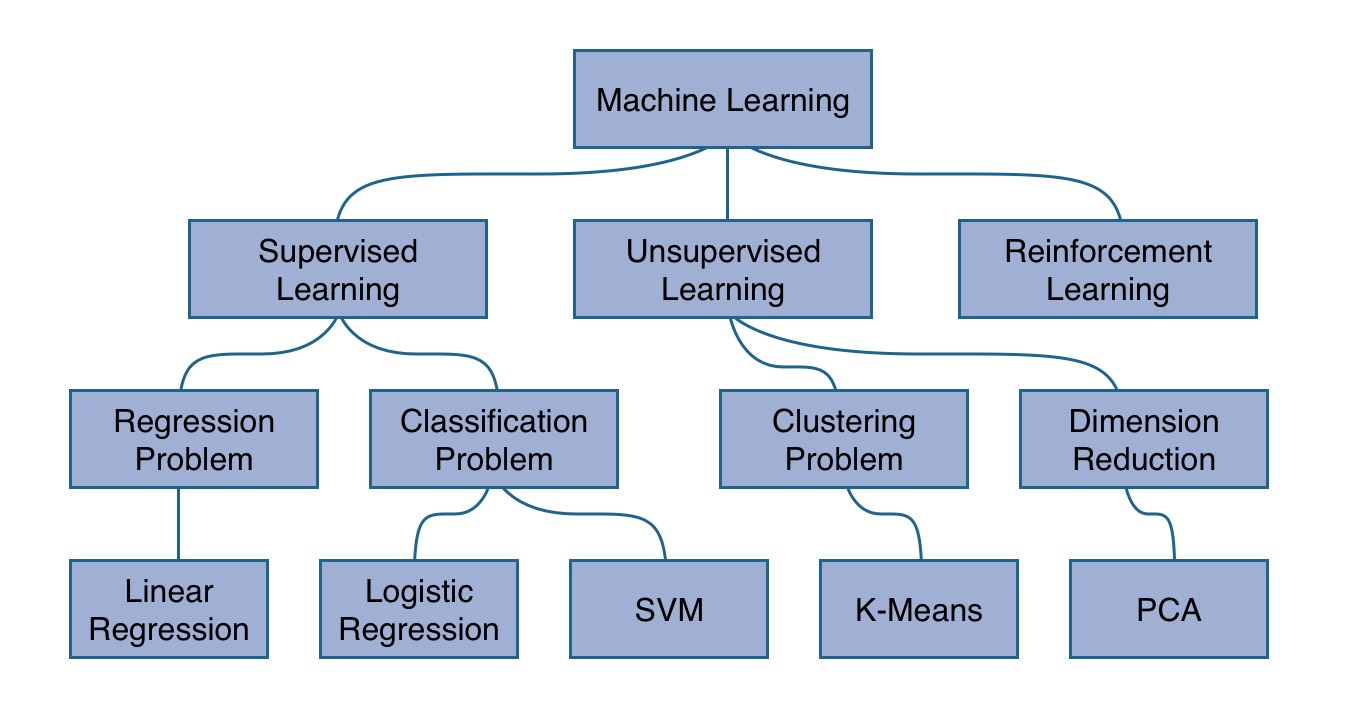


Figure 2.2: picture of Machine Learning hierarchy (Robinwieruch, 2018)

**2.3.2.1 Unsupervised Learning**

As the name itself indicates that unsupervised learning is where we lack supervisors or training data, simply all unlabeled data. In here the inputted data is unlabeled and the system tries to learn the structure from the given data automatically without any guidance of human. This algorithm studies how the systems can infer a function to describe a hidden structure from unlabeled data. The correct output is not figured out by the system since the data are explored and conclusions are based upon drawn datasets that describes hidden structures from unlabeled data. (Alex Smola and S.V.N. Vishwanathan, 2008)

So basically, Unsupervised Learning consists of clustering algorithms which isn’t preferred in our system because the data is sparse and it does more than what we need to be done, which results in poor prediction, therefore this Machine Learning type isn’t used in our project.

* **Clustering**

The functionality of this algorithm is to separate and group a given set of data based on their similarity/ distance measures and the basic principle for doing this grouping is to reduce the intra- group distance while increasing the opportunity for inter- group distance. (Le, 2018)

There are many clustering algorithms available, the following are the ones considered to be most common;

* Connectivity-based algorithms
* Probabilistic
* Density-based algorithms
* Neural networks / Deep Learning
* Centroid-based algorithms
* Dimensionality Reduction

**2.3.2.2 Supervised Learning**

The name Supervised Learning itself tells that it’s an algorithm that requires data scientist or data analyst with machine learning skills to provide both input and desired output, to simply put it the data analyst/ scientist decides on which variables or features the system should be fed such that the algorithm will apply the necessary learned data and provide the required predicted new data. (Furbush, 2018)

After sufficient amount of training the system can provide targets for any input. It starts analyzing from a known training dataset and by applying the learning algorithm it produces an inferred function to make the predictions about the output values. (Chao, 2011)

According to Fig 2.1 Regression and Classifications are two main subset algorithms used under Supervised Learning, for our project we would be focusing on Regression, most specifically Linear Regression because it doesn’t over fit or under fit even if there are hundreds of features and less dataset therefore it predicts 100%. Therefore, we have decided to choose Supervised Learning algorithm to build our project.

The following are considered to be most used Supervised Learning methods;

* Decision Tree
* Linear Regression
* Support Vector Machines

1. **Decision Tree**

Decision tree represents a binary tree model which has a structure of a flowchart, in which each node represents a single input variable say ‘x’. The leaf nodes contain an output variable which is used in making predictions. Decision trees are trained on data for classification and regression problems. (Le, 2016)

Decision trees are nothing but a tool that is used to support in decision making such as; event outcomes, resource cost, and utility. This method allows you to approach any problems in a well-structured and a systematic way in order to achieve a logical conclusion. (Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, Thorsten Joachims, 2016)

1. **Linear Regression**

Linear Regression is one of the most well-known algorithms on statistics and machine learning. This predictive is concerned with minimizing the error of a model and to produce accurate prediction. (Nilsson, 1998)

Linear regressions are basically a representation of an equation that best describes a line that fits the relationship between the input variables ‘x’ and the output variables ‘y’ by finding the coefficients ‘b’ for the input variables and their specific weightings. (Shai Shalev-Shwartz and Shai Ben-David, 2014)

Let’s consider y = b0 + b1\*x,

As mentioned before the goal this algorithm is to find the coefficients of b0 and b1 given that the input is ‘x’ by means we will be able to predict ‘y’.

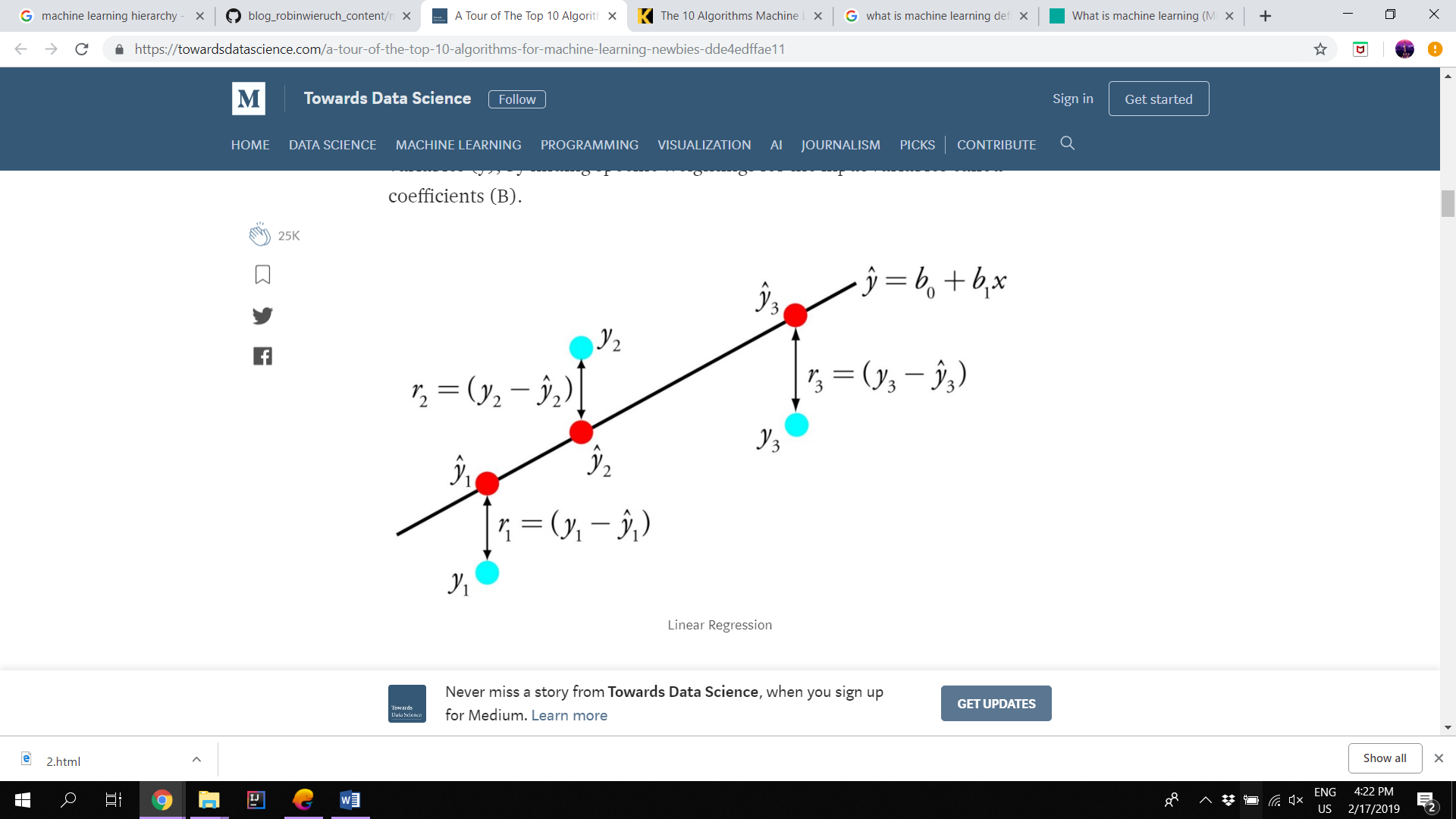


Figure 2.3: Diagram explaining Linear Regression equation. (Le, 2016)

1. **Support Vector Machines**

This learning method is one of the most powerful and accurate among all well-known algorithms. It is a supervised learning model that can be used for both regression and classification. It is also known as a binary classification algorithm. (Nilsson, 1998)

The intension behind this learning algorithm is to find the optimal hyperplane with a maximum distance to the closest point of the two classes. To put it in simple, let’s consider a set of points pf two types in N dimensional place, Support Vector Machines generates a (N-1) dimensional hyperplane in order to separate the given two points. (Shai Shalev-Shwartz and Shai Ben-David, 2014).

This algorithm is capable of handling with both linearly and nonlinearly separable cases by applying kernel function. Since this consists of mathematically complicated algorithms, it’s very complicated to interpret them.

**2.3.2.3 Semi- supervised Learning**

Semi- supervised Learning is the combination of both supervised and unsupervised learning and it is also considered to be widely used machine learning method according to. This method is used when there are more datasets which are unlabeled than labeled since it involves both to produce an accurate learning model. (Muhammad Badruddin Khan,Mohssen M. Z. E. Mohammed and Eihab Bashier Mohammed Bashier, 2016)

**2.3.2.4 Reinforcement Learning**

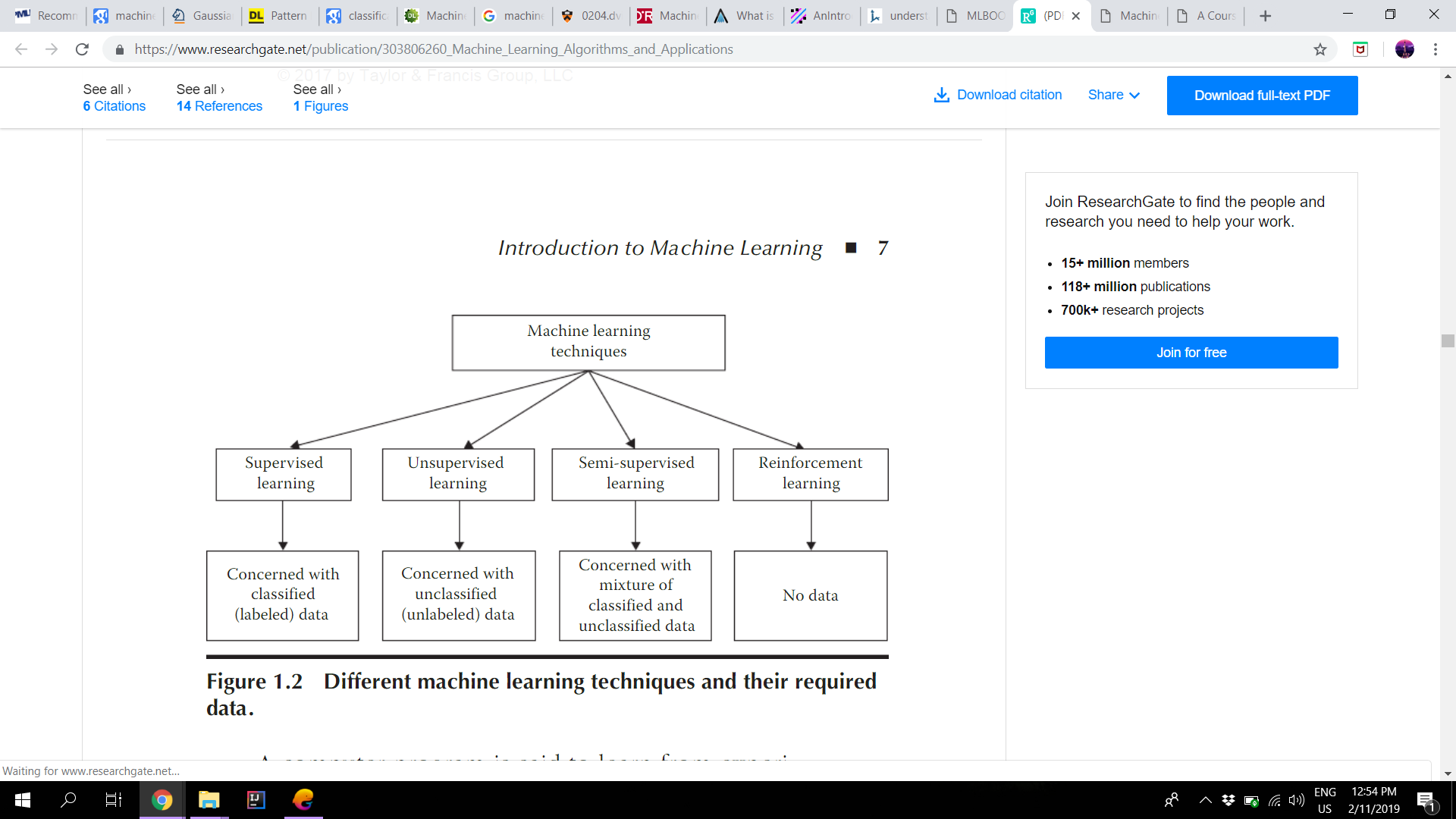
Reinforcement Learning method gathers data via observations and interactions from the environment to take actions that which would reduce the risk and maximize the reward that produces favorable outcomes (Muhammad Badruddin Khan,Mohssen M. Z. E. Mohammed and Eihab Bashier Mohammed Bashier, 2016). The most relevant characteristics of reinforcement learning is trial and error search and delayed reward. In order to maximize its performance this method allows the software to automatically determine the ideal behavior within a given specific context.

Figure 2.4: Short summary of main methods in Machine Learning Techniques. (Muhammad Badruddin Khan,Mohssen M. Z. E. Mohammed and Eihab Bashier Mohammed Bashier, 2016)

### 2.4 Approach

Searching for a perfect apartment to live in with all the necessary factors such as; meeting up the budget, nearby facilities to local supermarkets, schools, public transportation, proximity to workplace, whether the apartment is located near an urbanized area or not, the interior luxury settings of the apartment, crime rate and population of the area, is not a one day/ night work. This selection of apartment with all requirements being met properly would consume time and at times we could be falsely entrusted by an unknown third-party broker. The purpose of this research is to develop an apartment prediction application that will provide accurate results with an analyzed set of data that which would satisfy the buyer and save up his time spent on looking for a perfect home to live.

The proposed prediction system gathers all prerequisite requirements from the user via several questions related to cost, internal and external environmental factors and with providing sample apartment suggestion (depending on the dataset). The responses gathered from the user would be classified with a pre-trained model built utilizing machine learning algorithms and then the apartments list would be suggested.

As people increasingly rely on interactive decision support system to choose products and make decision, to cope up with this recent trend the significance of this project is planned to reduce the amount of days and months spent by individuals in search of an apartment as well as to avoid being cheated by third party brokers, not only these but also by providing a future prediction lasting up to five years, that which would guarantee a happy living and a happy customer using the latest technologies such as Google Cloud, Map APIs etc.

### 2.5 Current Existing System

Majority of the current existing systems are all based on America or Canada or similarly developed countries where as all the applications that are existing so far only provides a recommendation according to the feed that user inputs into the application.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Features** | **Existing Systems** | | | **Our System** |
| **Move In: Apartment Rentals All in One** | **Rent berry: Apartment Rentals Worldwide** | **Realtor.com Rentals: Apartment, Home Rental Search** |
| Location where apartment is situated |  |  |  |  |
| Updates about rentals |  |  |  |  |
| Guidance through all famous apartment companies |  |  |  |  |
| Multiple Language selection |  |  |  |  |
| Selection on Furniture finishing |  |  |  |  |
| Review of real Market demand |  |  |  |  |
| Pet friendly policy checking |  |  |  |  |
| Interactive user-Friendly map provision |  |  |  |  |
| Connection with Landlords |  |  |  |  |
| Find about neighborhood (whether home or apartment) |  |  |  |  |
| Robust filtering according to rent budget |  |  |  |  |
| Crime Rate Prediction |  |  |  |  |
| Estimation of distance between workplace and apartment |  |  |  |  |
| Prediction and suggesting of apartments based on user requirements |  |  |  |  |

Table 2.1: Difference between existing system and ours

## 2.6 Chapter Summary

This chapter was mainly focused on the methodology and techniques that can be used in apartment predications. At start a short review was given on the problem domain for the project. Later, it further discussed about Machine Learning algorithms in brief, in order to give an idea about other algorithms in Machine Learning that can be used for various prediction system as well as discussed about the algorithm that we chose out of many existing algorithms in Machine Learning. Algorithms that were discussed overall were Decision Tree, Support Vector Machines, Linear Regression, Clustering, Reinforcement Learning, Semi-supervise Learning, Unsupervised Learning and Supervised Learning. Then a short review on our approach for the proposed system was also discussed. Finally, a tabular form of comparisons was given between the existing systems and our proposed system.

# References

Apartments.com and Google , 2018. *Online Search Behavior and Trends of Apartment Renters,* s.l.: s.n.

Alex Smola and S.V.N. Vishwanathan, 2008. *Introduction to Machine Learning.* s.l.: university of cambridge.

Chao, W.-L., 2011. *Machine Learning Tutorial,* Taiwan: s.n.

Chollet, F., 2017. *Deep Learning with Python,* s.l.: O'Reilly.

Furbush, J., 2018. *Machine learning: A quick and simple definition,* s.l.: s.n.

Geron, A., 2017. *Hands-On Machine Learning with Scikit-Learn and TensorFlow,* s.l.: O' Reilly.

III, H. D., 2012. *A Course in Machine Learning.* s.l.:s.n.

Le, J., 2016. *The 10 Algorithms Machine Learning Engineers Need to Know,* s.l.: s.n.

Le, J., 2018. *A Tour of The Top 10 Algorithms for Machine Learning Newbies,* s.l.: s.n.

Muhammad Badruddin Khan,Mohssen M. Z. E. Mohammed and Eihab Bashier Mohammed

Bashier, 2016. *Machine Learning: Algorithms and Applications.* s.l.: CRC Press.

Nilsson, N. J., 1998. *INTRODUCTION TO MACHINE LEARNING.* Stanford : Stanford University.

Robinwieruch, 2018. *A Web Developer's Guide to Machine Learning in JavaScript,* s.l.: s.n.

Schapire, R., 2008. *Theoretical Machine Learning,* s.l.: s.n.

Shai Shalev-Shwartz and Shai Ben-David, 2014. *Understanding Machine Learning:.* s.l.:Cambridge University Press.

The National Association of Realtors® and Google, 2012. *The Digital House Hunt: Consumer and Market Trends in Real Estate,* s.l.: s.n.

Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, Thorsten Joachims, 2016. *Recommendations as Treatments: Debiasing Learning and Evaluation,* New York: Cornell University.