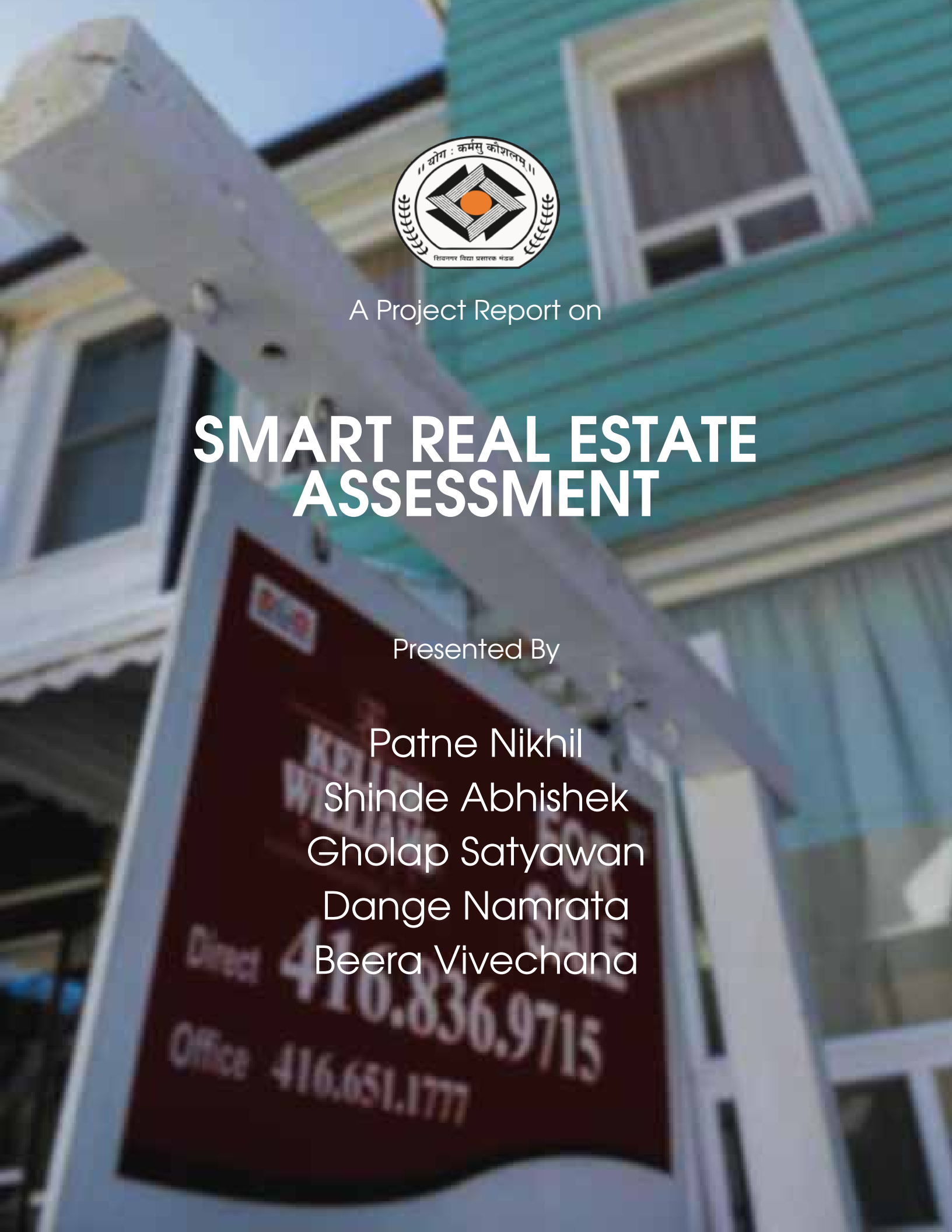


A Project Report on

SMART REAL ESTATE ASSESSMENT

Presented By

Patne Nikhil
Shinde Abhishek
Gholap Satyawan
Dange Namrata
Beera Vivechana

DEPARTMENT OF COMPUTER ENGINEERING

SVPM'S COLLEGE OF ENGINEERING
MALEGAON(Bk)
2018-19

SMART REAL ESTATE ASSESSMENT

This research was done under the guidance of Prof. Nimbalkar S.S. with support of Prof. Kumbhar H.R. (H.O.D. Comp. Dept.) from August 2018.

SVPM's COLLEGE OF ENGINEERING
DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the Project Entitled

Smart Real Estate Assessment

Submitted by

Dange Namrata Milind	Exam No:B150194206
Gholap Satyawan Ashok	Exam No:B150194214
Patne Nikhil Sunil	Exam No:B150194234
Shinde Abhishek Sunil	Exam No:B150194242
Beera Vivechana Ramesh	Exam No:B150194259

is a bonafide work carried out by Students under the supervision of **Prof. S.S. Nimbalkar** and it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Computer Engineering) Project.

Prof. S.S. Nimbalkar
Internal Guide

Prof. H.R. Kumbhar
H.O.D

_____
External Examiner
.

_____
Prof. S.M. Mukane
Principal

Place : SVPM's COE Malegaon(Bk.)
Date :

Acknowledgments

It gives us great pleasure in presenting the preliminary project report on ***'Smart Real Estate Assessment'****.*

I would like to take this opportunity to thank my internal guide ***Prof. Nimbalkar S.S*** *and all teachers for giving me all the help and guidance I needed. I am really grateful to them for their kind support.I appreciate all valuable suggestions given and queries raised by her which really helped us in exploring project in deep.*

I am also grateful to ***Prof. H.R. Kumbhar****, Head of Computer Engineering Department, SVPM COE for his indispensable support, suggestions.*

In the end our special thanks to ***Mr. S.P. Kadam*** *for providing various resources such as laboratory with all needed software platforms, continuous Internet connection, for Our Project.*

Dange Namrata
Gholap Satyawan
Patne Nikhil
Shinde Abhishek
Beera Vivechana

(B.E. Computer Engg.)

Smart Real Estate Assessment

June 6, 2019

Abstract

REAL estate appraisal, which is the process of estimating the price for real estate properties, is crucial for both buys and sellers as the basis for negotiation and transaction. Traditionally, the repeat sales model has been widely adopted to estimate real estate price. However, it depends the design and calculation of a complex economic related index, which is challenging to estimate accurately. Today, real estate brokers provide easy access to detailed online information on real estate properties to their clients. We are interested in estimating the real estate price from these large amounts of easily accessed data. In particular, we analyze the prediction power of online house pictures, which is one of the key factors for online users to make a potential visiting decision. The development of robust computer vision algorithms makes the analysis of visual content possible. In this work, we employ a Recurrent Neural Network (RNN) to predict real estate price using the state-of-the-art visual features.

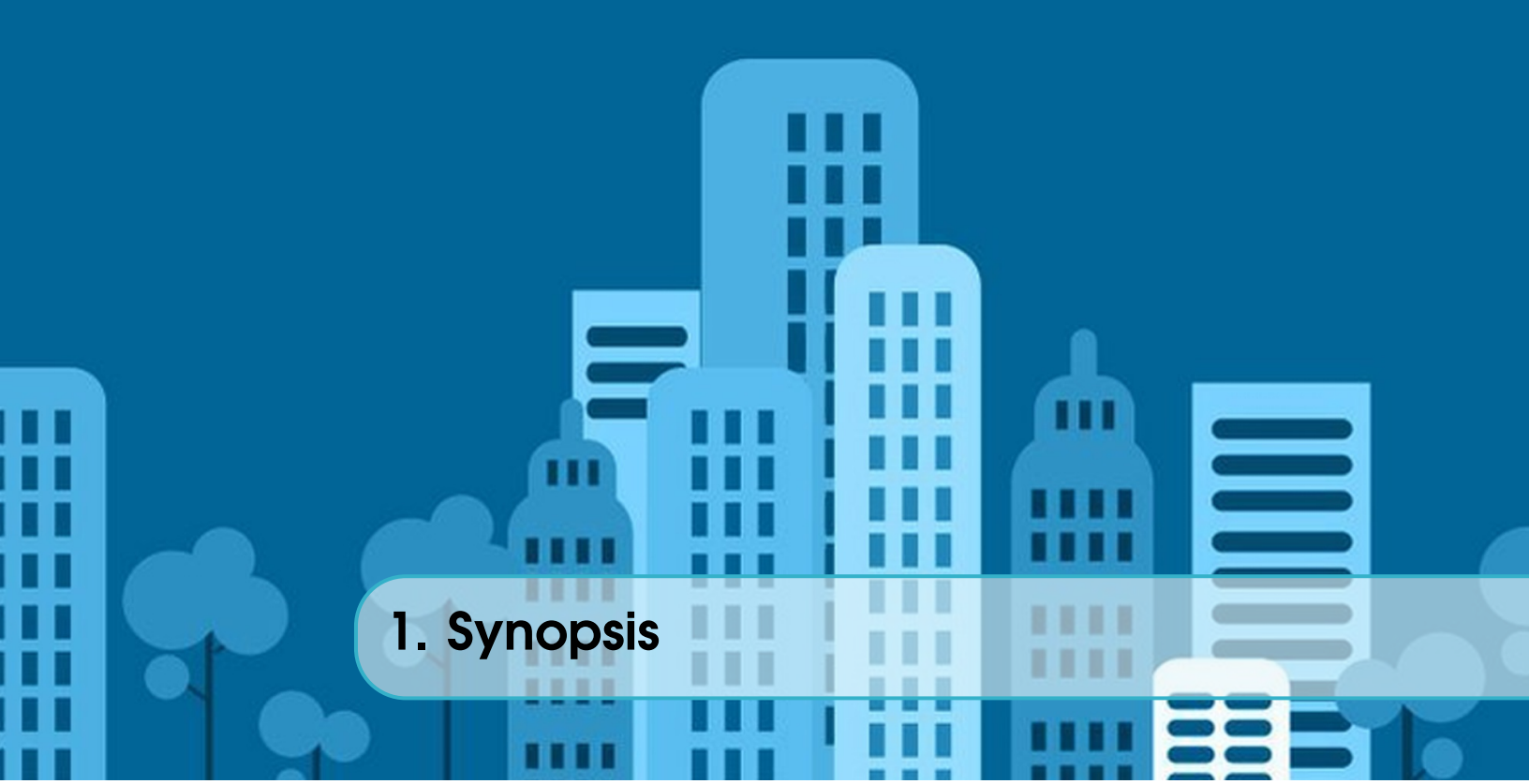
In a smart city, effective and accurate real estate assessments governed by a local government is crucial for determining the property taxes. Such assessments have never been trivial, and inappropriate assessments may result in disputes between property owners and the local government.

Generally for price prediction **Regression** is used (Prediction of continuos valued-function). But here we are going to use **Structured Deep Neural Network** in order to improve efficiency and accuracy. We introduce a deep learning approach to smartly and effectively assessing real estate values. We propose a systematic method to derive a layered knowledge graph and design a structured Deep Neural Network (DNN) based on it. Neurons in a structured DNN are structurally connected, which makes the network time and space efficient; and thus, it requires fewer data points for training. The structured DNN model has been designed to learn from the most recently captured data points; therefore, it allows the model to adapt to the latest market trends.

Contents

List of Figures

List of Tables



1. Synopsis

1.1 Project Title

Smart Real Estate Assessment

1.2 Project Option

Internal project

1.3 Internal Guide

Prof.Nimbalkar S.S

1.4 Problem Statement

To predict the price of the real estate properties which is depends upon various factors like Infrastructure,Neighbourhood environment,House age, Location etc. using Structured Deep Neural Network.

1.5 Goals and Objectives

1.5.1 Goal

To construct an assessment system using Deep Neural Network in a more efficient effective way with, improved accuracy for house-price predic-

tion.

1.5.2 Objectives

- Prediction of prices of real estate property more accurately(Using DNN).
- Elimination of Role of Middle Person(*Agent*).
- Assessments are not only for urban areas but also for Rular areas.
- Real time prediction of can be possible

2. Technical Keywords

2.1 Area of Project

Deep Neural Network

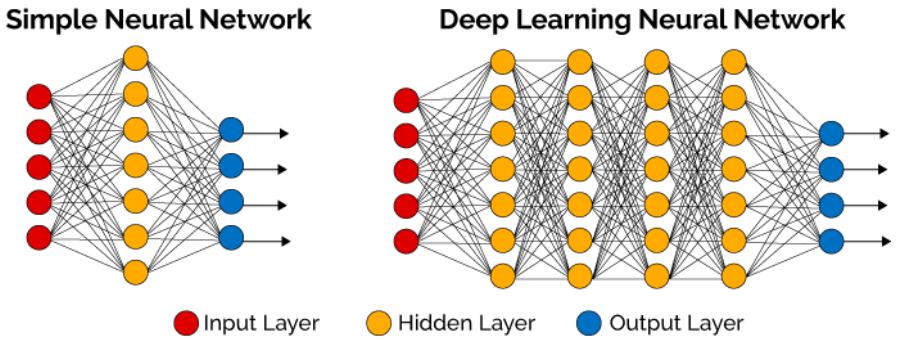


Figure 2.1: Difference between Simple Neural Network and Deep Neural Network, https://becominghuman.ai/deep-learning-made-easy-with-deep-cognition-403fbe445351

2.2 Technical Keywords

1. **Deep Learning**: Deep learning is an aspect of artificial intelligence (AI) that is concerned with emulating the learning approach that human beings use to gain certain types of knowledge. At its simplest, deep learning can be thought of as a way to automate predictive analytics.Deep learning is the branch of machine learning which

having ability to learn multiple level representation from structured data using supervised or unsupervised learning.

2. **Personalization**: Personalization goes more with the user's decisions, this means, the instructions explicitly given. An example could be the background image of twitter could be taken as personalization of your twitter profile.
3. **Online Interest**: Get users interest through social media like twitter,facebook.For that data need to mine.
4. **Social Media**: To mine users point of interest need to get data from usesrs social account.

3. Introduction

3.1 Project Idea

Real estate appraisal, which is the process of estimating the price for real estate properties.
A report published by EPRA(*Europian Public Real Estate Association*) real estate in all its forms accounts approximately 18-20 percent of its economical activity.therefore accurate prediction of real estete properties are crucial. For most of the working classes housing has been one of the largest expense, so to make right decision on the real estate investment is much cruicial.

3.2 Motivation of the Project

As we seen the report published by EPRA and why investment in housing is important the thats why accurate preduction of real estate properties are crucial.Traditionally previous work for prediction of price is based on regression analysis and machine learning.but due to rapid development in Deep Neural Network feild it is much benificial to use DNN instead of Regression.Beacuse of recently developed deep learning,computer becomes smart enough to interpret visual content in similar way that human

can.
peoples can able to estimate price more accurately in this system as compared to previous systems only because of DNN.

3.3 Literature Survey

- J. Frew and G. Jud, "Estimating the Value of Apartment Buildings," Journal of Real Estate Research
  - Techniques : Hedonic modeling techniques to estimate
  - Advantages : Able to estimate prices correctly in proportion of size and number of units
  - Disadvantages : Effect of Aging isn't considered effectively

- R. E. Lowrance, "Predicting the Market Value of Single-Family Residential Real Estate," Technical Report
  - Techniques : Local Linear model and Random Forest model
  - Advantages : lowest expected error on unseen data and model is tailored to zip codes using indicator variables
  - Conclusion : Random forest model may perform better than the local linear model

- X. Hu and M. Zhong, "Applied Research on Real Estate Price Prediction by the Neural Network,"
  - Techniques : Backpropagation neural networks and Elman neural network
  - Advantages : Elman neural network could forecast more accurate and constringe faster than other approaches

- N. Nguyen and A. Cripps, "Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks," Journal of Real Estate Research
  - Techniques : ANN and multiple regression analysis
  - Advantages : when enough data points were available for training, ANNs could perform better than multiple linear regressions

- Y. E. Hamzaoui and J. A. H. Perez, "Application of Artificial Neural Networks to Predict the Selling Price in the Real Estate Valuation Process"
  - Techniques : Feed-forward backpropagation neural network with a single hidden layer
  - Advantages : reliable prediction of house selling prices at that time

- X. Zhang, "Using Fuzzy Neural Network in Real Estate Prices Prediction"
  - Techniques : fuzzy neural network to support fuzzy reasoning and learning
  - Advantages : works better than traditional neural network approaches

S. Chopra, T. Thampy, J. Leahy, A. Caplin, and Y. LeCun, "Discovering the Hidden Structure of House Prices with a Non-Parametric Latent Manifold Model"

- Techniques : Latent Manifold Model
- Advantages : with two trainable components like one parametric component that predicts the "intrinsic" price of a house using its features, and the second one is a non-parametric component that calculates the desirability of the neighborhood; it performs better than pure parametric or non-parametric models
- Disadvantages : Fails in Rular areas

B. J. Ford, H. Xu, and I. Valova, "A Real-Time Self-Adaptive Classifier for Identifying Suspicious Bidders in Online Auctions"

- Techniques : Neural Network model
- Advantages : By training the neural network with newly added structured data, so it can quickly adapt to changing trends in bidding, model is able detect suspicious bidders in online auctions
- Disadvantages : used structured input and learning processes, and the DNNs are fully connected

S. Zhang, Y. Bao, P. Zhou, H. Jiang, and L. Dai, "Improving Deep Neural Networks for LVCSR Using Dropout and Shrinking Structure"

- Techniques : Specialized structured neural network
- Advantages : Shrinking DNN structure with hidden layers decreasing in size from a lower layer to higher layers for the purpose of reducing the model size and making the model time efficient without affecting performance.
- Disadvantages : they failed to justify why shrinking DNN would not affect performance, and also there was a lack of systematic approach for the network reduction

4. Problem Definition and Scope

4.1 Problem Statement

To predict the price of the real estate properties which is depends upon various factors like Infrastructure,Neighbourhood environment,House age,location etc. using *Structured Deep Neural Network*.

4.1.1 Goals and objectives

Goals

The Goal of this System is to predict the price of the real estate property more precisely

Objectives

- Prediction of prices of real estate property more accurately (Using DNN).
- Elimination of Role of Middle Person(Agent).
- Assessments are not only for urban areas but also for rural areas.

4.2 Major Constraints

- Application should be web based.
- Supported by any browser (mobile browser,pc browser).

- The information store will be an PostgreSQL database.

4.3 Outcome

- To get the estimated price of the real estate property.

4.4 Applications

- Housing Investment
- Auctions
- Real time Valuation

4.5 Hardware Resources Required

Developing Environment:

Sr. No.	Parameter	Minimum Requirement
1	CPU Speed	2.5 GHz
2	RAM	8 GB
2	HDD	1 TB

Table 4.1: Hardware Requirements

Operating Environment:

Sr. No.	Parameter	Minimum Requirement
1	CPU Speed	800 MHz
2	RAM	256 MB

Table 4.2: Hardware Requirements

4.6 Software Resources Required

4.6.1 Developing Environment

Platforms :

1. Operating System: Windows 10
2. IDEs: PHP Storm,Visual Studio Code.
3. Programming Language :PHP,Python
4. Database: PostgreSQL .
5. **Tools:**

   (a) Documentation: TexLive(Texworks Editor) and Overleaf(Oniline LaTeX Editor)
   (b) Digram: LucidChart (Online Drawing Tool)

4.6.2 Operating Environment

Platforms :

1. Operating System:Windows/Linux Based version released after 2012,Mac
2. Web Browser: Chrome version released after 2012,Microsoft EDGE,Mozilla

5. Project Plan

5.1 Project Estimates

Agile model:

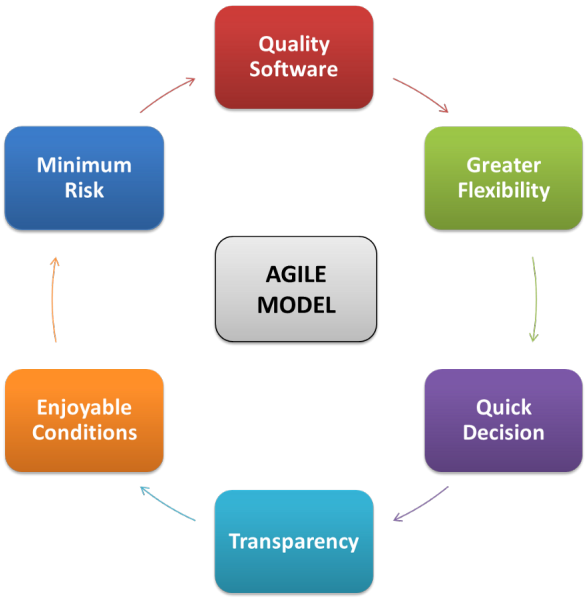


Figure 5.1: Benefits of Agile Model
©https://medium.com/@proximitycrc/6-benefits-of-agile-model-bddf55f976b5

The Agile Method is a particular approach to project management that is utilized in software development. This method assists teams in

responding to the unpredictability of constructing software. It uses incremental, iterative work sequences that are commonly known as sprints.Agile Methods break the product into small incremental builds.**Scrum** is a framework for Agile software development.

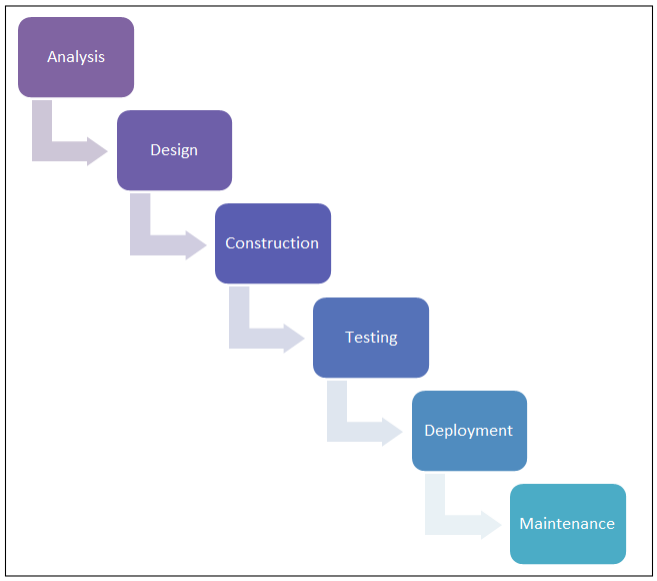


Figure 5.2: Agile Model
©https://www.powerobjects.com/2013/02/11/microsoft-dynamics-crm-and-agile-development

5.2 Spiral Model

We Have used **Spiral Model** for analysis of the system.
Each phase of Spiral Model is divided into four quadrants. The functions of these r quadrants are discussed below with diagram.

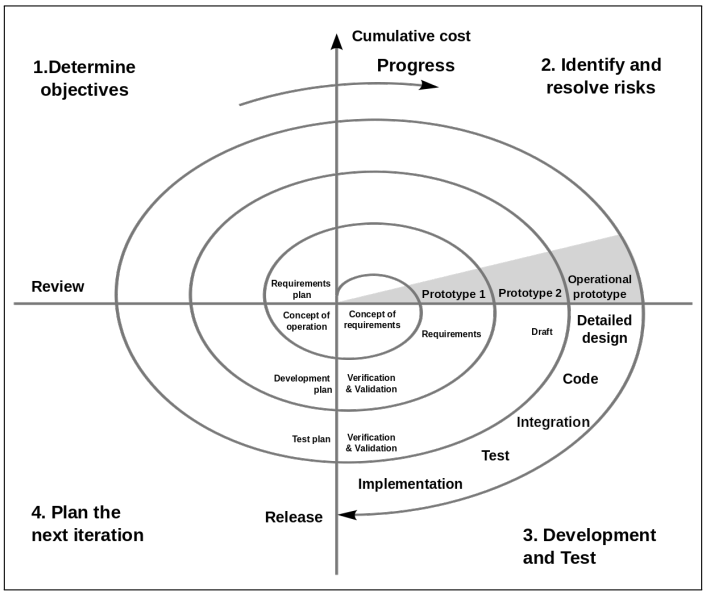


Figure 5.3: spiral Model
©https://upload.wikimedia.org/wikipedia/commons/thumb/e/ec/Spiral model/spiral.png

- **Objectives determination and identify alternative solutions:-**
  Requirements are gathered from the customers and the objectives are identified, elaborated and analyzed at the start of every phase. Then alternative solutions possible for the phase are proposed in this quadrant.
- **Identify and resolve Risks:-**
  During the second quadrant all the possible solutions are evaluated to select the best possible solution. Then the risks associated with that solution is identified and the risks are resolved using the best possible strategy. At the end of this quadrant, Prototype is built for the best possible solution.
- **Develop next version of the Product:-**
  During the third quadrant, the identified features are developed and verified through testing. At the end of the third quadrant, the next version of the software is available.
- **Review and plan for the next Phase:-**
  In the fourth quadrant, the Customers evaluate the so far developed version of the software. In the end, planning for the next phase is started.

5.3 Risk Management

Our project support only NP-complete category problem. There are variety of user who are using our system for the searching the approximated price of the real estate properties. Every user provides personal data consistent with his convenience, his/her request are send to the server. This method is depends upon user sort enclosed therefore this project is below NP-complete category.In our project we have a tendency to use different algorithms i.e. construction of structured DNN Algorithm,And construction of structured knowledge graph which are NP-Complete because it depends on user's parameters.

5.3.1 Risk Identification

"Risk are future unsure events with a chance of prevalence and a possible for loss" Risk identification and management area unit the most issues in each computer code project. Effective analysis of computer code risks can facilitate to effective designing and assignments of work.

Categories of Risk

Risks are known, classified and managed before actual execution of program. These risks are classified in several classes.

Scheduled Risk:

Project schedule get slip once project tasks and schedule unharness risks don't seem to be self-addressed properly. Schedule risks primarily impact on project and at last on economy and should result in project failure. Schedules typically slip because of following reasons:

1. Wrong time estimation
2. Resources are not tracked properly.
3. Failure to identify complex functionalities and time required to develop those functionalities.
4. Unexpected project scope expansions.
5. Facilities are not available on time.

6. Sometime facilities are available but a inadequate
7. Development tools are not in work as expected.

End-User:

1. A delay in one task causes cascading delays in dependent task
2. End user is not solicited.
3. Communication time large required.

Budget Risk:

1. Wrong budget estimation.
2. Project scope expansion.

Operational Risks:

Risks of loss due to improper process implementation, failed system or some external events risks.

1. Database connectivity failure
2. users authentication and building of access tree failed.
3. Communication failure
4. No resource planning
5. Insufficient Network Bandwidth
6. Unavailability/Conguration failure of the Access Point.

Technical risks:

Technical risks generally leads to failure of functionality and performance. Causes of technical risks are:

1. Continuous changing requirements and environment
2. Product is complex to implement.
3. Server failure
4. Security breakdown
5. Scalability of network
6. Difficult project modules integration.

Risk Analysis

ID	Risk Description	Probability	Impact		
			Schedule	Quality	Overall
1	Continuous stream of required changes	Low	High	High	High
2	None of us known how to use technology	Low	Low	Low	Low
3	Real time performance	Low	Low	High	High
4	incorrect key generation	Low	Low	Low	Low
5	Module integration	High	High	High	High
6	System failure like database connectivity	Low	Low	High	High

Table 5.1: Risk Table

5.4 Plan of Project

1	**ANALYSIS PHASE**
1.1	Study of existing system
1.2	Study of discussion and research papers
1.3.1	Problem definition
1.3.2	Scope
1.3.3	Feasibility
1.4	Defining the problem
1.5	Fixing the scope of the project
1.6	Feasibility analysis
1.7	Requirement analysis
1.8	Project estimation
2	**Design Phase**
2.1	Designing GUI
2.2	Developing algorithms of various modules
2.3	Developing data flow diagrams of the system
3	**Coding**
3.1	Coding algorithm
3.2	Coding module
4	**Testing**
4.1	Unit Testing
4.2	Integration Testing
4.3	System Testing
5	**Deployment**
6	**Documentation**

Analysis Phase:

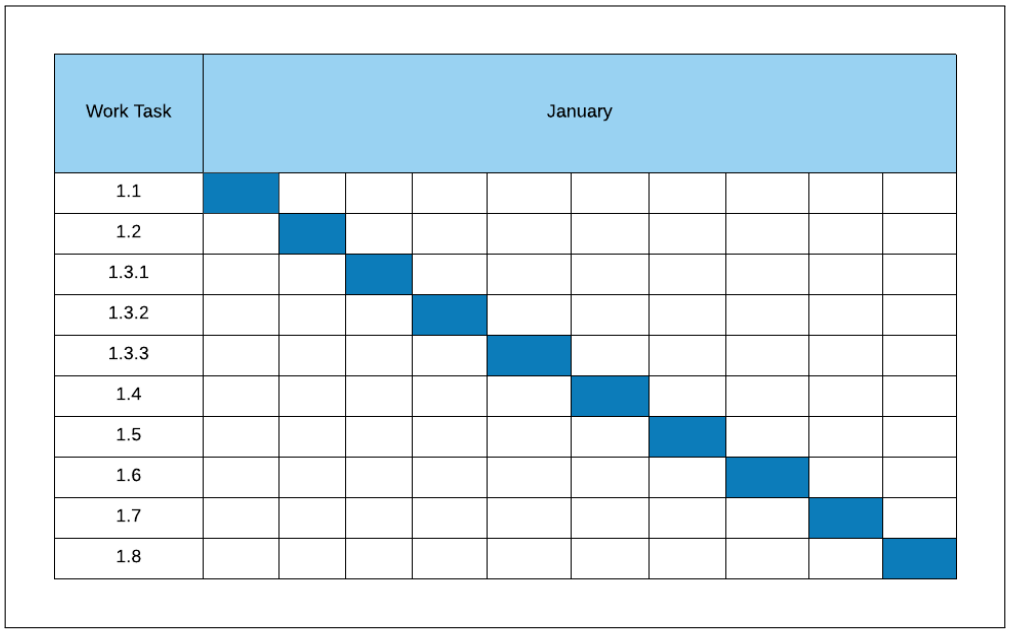


Figure 5.4: Planner1

Design Phase:

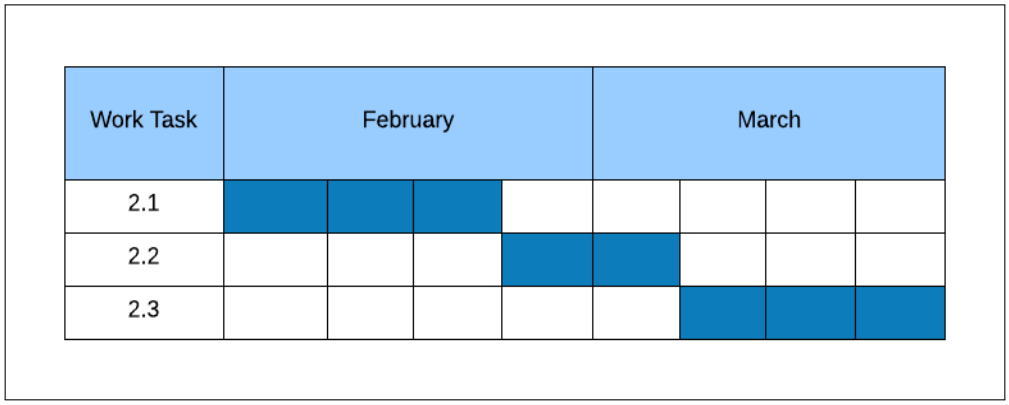


Figure 5.5: Planner2

Coding,Deployment and Documentation Phase:

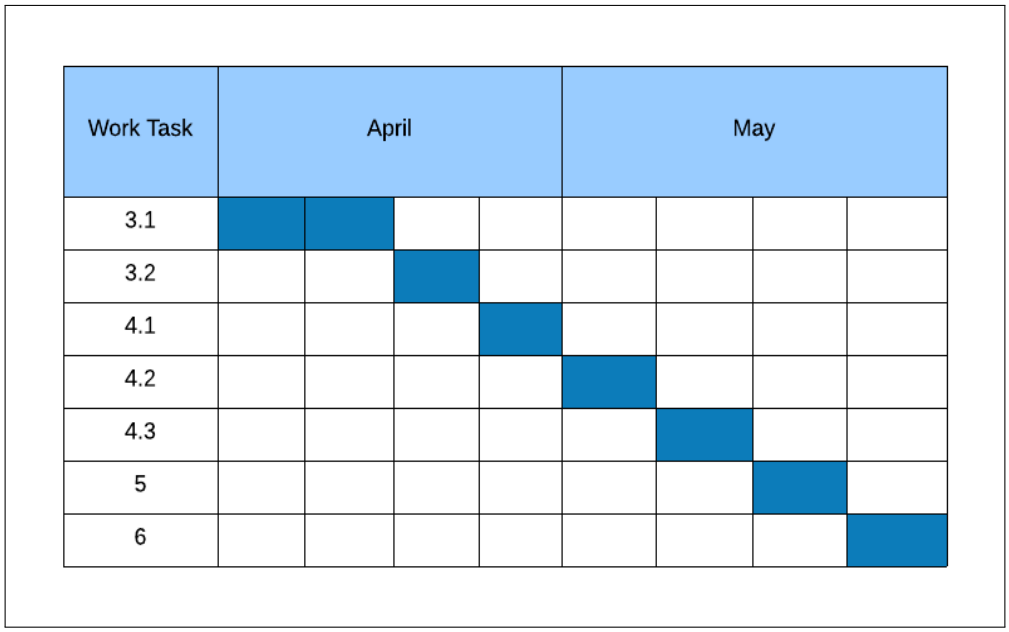


Figure 5.6: Planner3

6. Software Requirement and Specification

6.1 Usage Scenario

6.1.1 User profiles

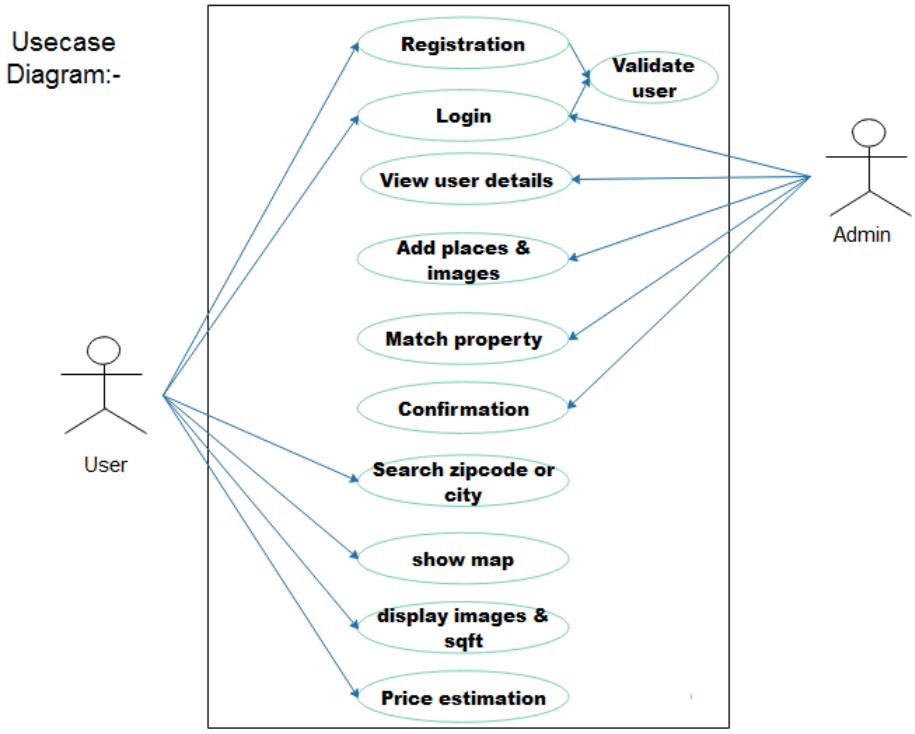
Admin

Admin manage the users and also set policies.It verifies the user and provides the access rights to the System. It is able to add new Data in the model.

User

User can be log in into the system and check for new places, properties and offers. They are able to add new property for sell,rent as well as for bidding.

6.1.2 Usecase Diagram



6.2 Functional Model and Description

6.2.1 Data Flow Diagram

Data Flow Diagram

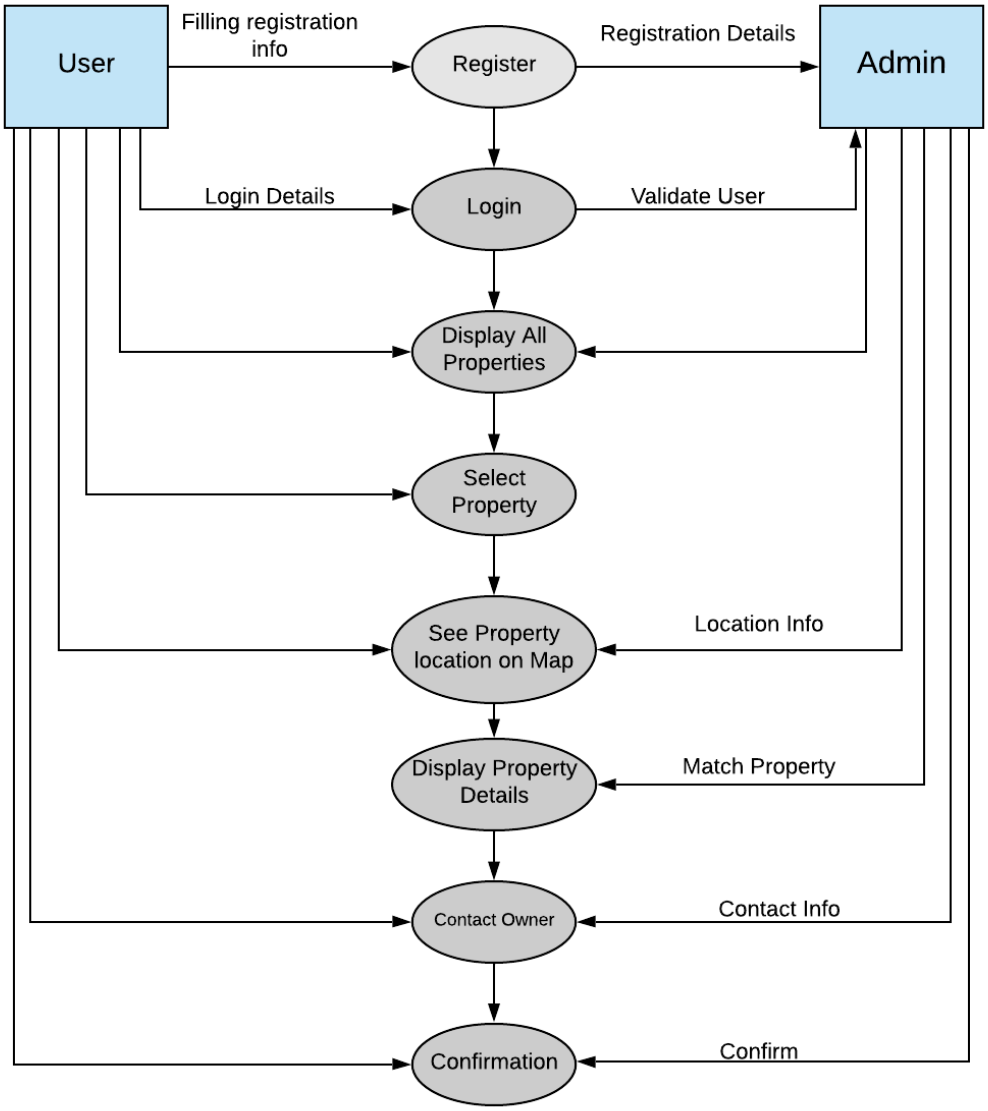


Figure 6.1: Data Flow Diagram

6.2.2 Non Functional Requirements:

Communication interface

It is a one-to-many user interface for communication. The users will use HTTP protocol to send queries to server.For this purpose HTML5 compatible browser is required and user can access the database through the browser by URL.

Performance Requirements

1. Performance should be high i.e. response time should be low in fraction of seconds.
2. Performance is increased by providing one URL to all clients through which they can access the database.
3. Proper updates of records database and some other information.
4. The system does not make any mistake.
5. Results needs to be shown to user in proper manner to make system effective.
6. Appropriate colour combinations should be used.
7. The GUI must be clear,unambiguous and exible for use.

Safety Requirements

As the database on server is centralized one there may be chances of data crash due to virus or other problems of software.So data backup is required.The record les are stored in another server and generated report files are stored on another server.Because of that no any conict occurs.

Security Requirements

Server Application is run only on the Server side, it provides Security and no chances of misuse.Proper authentication must be provided as all the data is on server.So even new admin registration must be provided only under senior admin permission .And only registered users can access the data and makes each other to their friends.

Software Quality Attributes

1. **Adaptability**: Any changes in software or any advanced modification will be possible.This project is reusable for creating new search

engines.The algorithms used in this project can be modified in future to get more effective performance in case of reducing time and space complexity.

2. **Maintainability**: Maintains performance throughout the system.As we are providing results to query instantly within fraction of seconds it improves performance of using this search engine greatly.
3. **Portability**: System is portable to any environment. Java ,ph,python are portable and platform independent language .So main aim of using the PHP for our project is to provide portability to our project.
4. **Reliability**: Having many qualities to make system reliable.We are providing backup and recovery facility in centralize environment. This makes our database more reliable.

Database Requirements

PostgreSQL Server connectors are required for the connectivity between PHP and PostgreSQL server database, and php and PostgreSQL connectivity is used to store and fetch data .

Internalization Requirements

To make our project global extensions required are different country languages e.g. English etc.

Legal Requirements

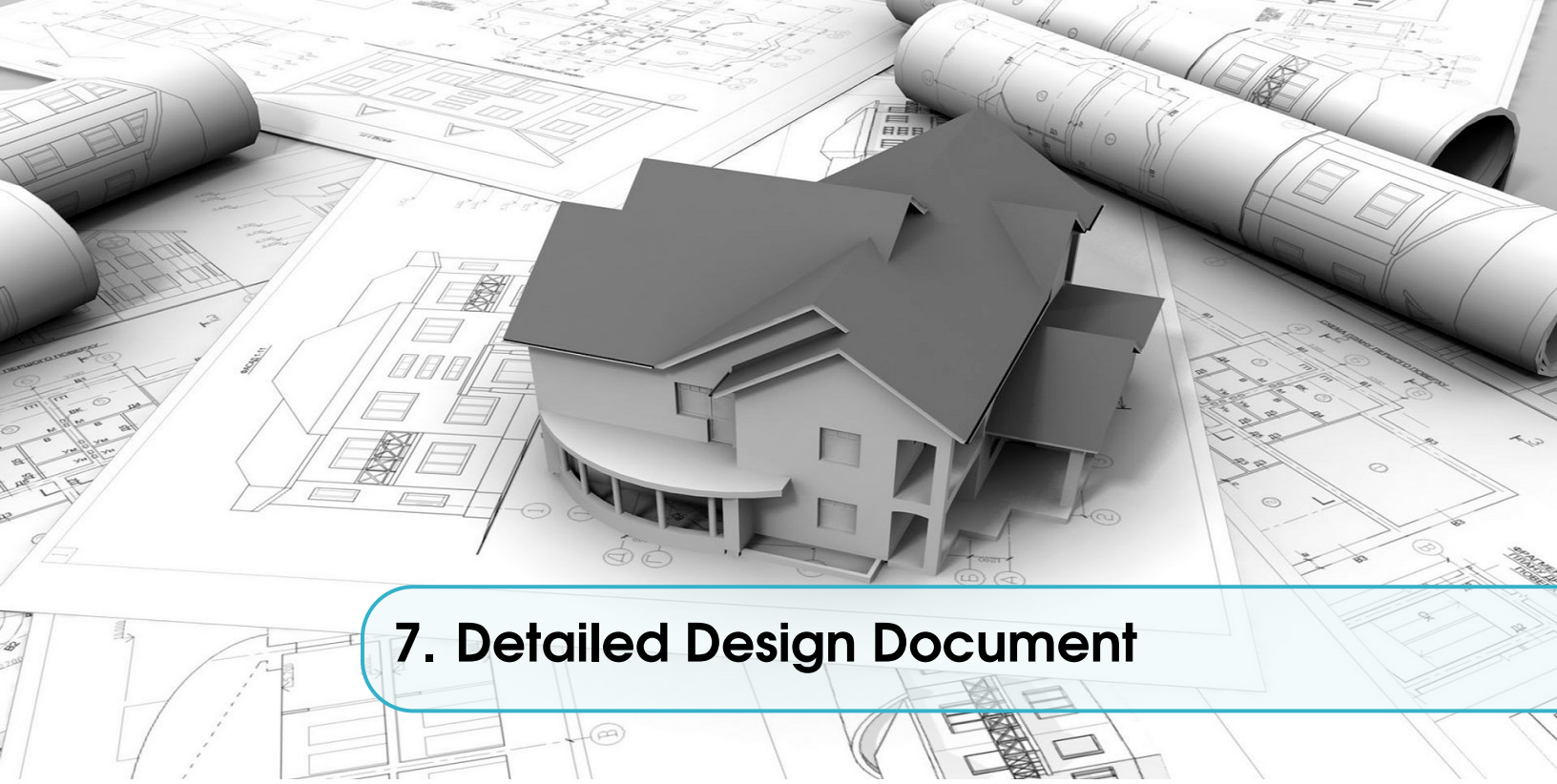
As our project is web based project multiple users may interact with this system simultaneously with different platforms on their PC. Also our project works according to concern of user expectation.

Reuse Objectives for the project

This project can be reused in future for developing new system which provides simultaneously access .Also battery efficiency will be done because of this will be used in future for sensing with least battery.

6.2.3 Software Interface Description

Here we will be using internet connection and HTTP protocol for web based access.



7. Detailed Design Document

7.1 Introduction

This document specifies the design that is used to solve the problem of Product.

7.2 Architectural Design

A description of the program architecture is presented. Subsystem design or Block diagram,Package Diagram,Deployment diagram with description is to be presented.

7.2.1 Data Collection and Preprocessing

We collected real estate data from a real estate website Zillow.com. This website conserves all recent and past house listings data including house features, market features, public records of houses, neighborhood features, etc. An entire number of 15 features are predefined and their associated values are collected. The predefined features include number of beds, number of baths, square footage, lot size, built year, yearly tax, similar houses average sold price, adjacent schools average ratings, fireplace, waterfront, the number of stories, heating, cooling, patio, and park. In

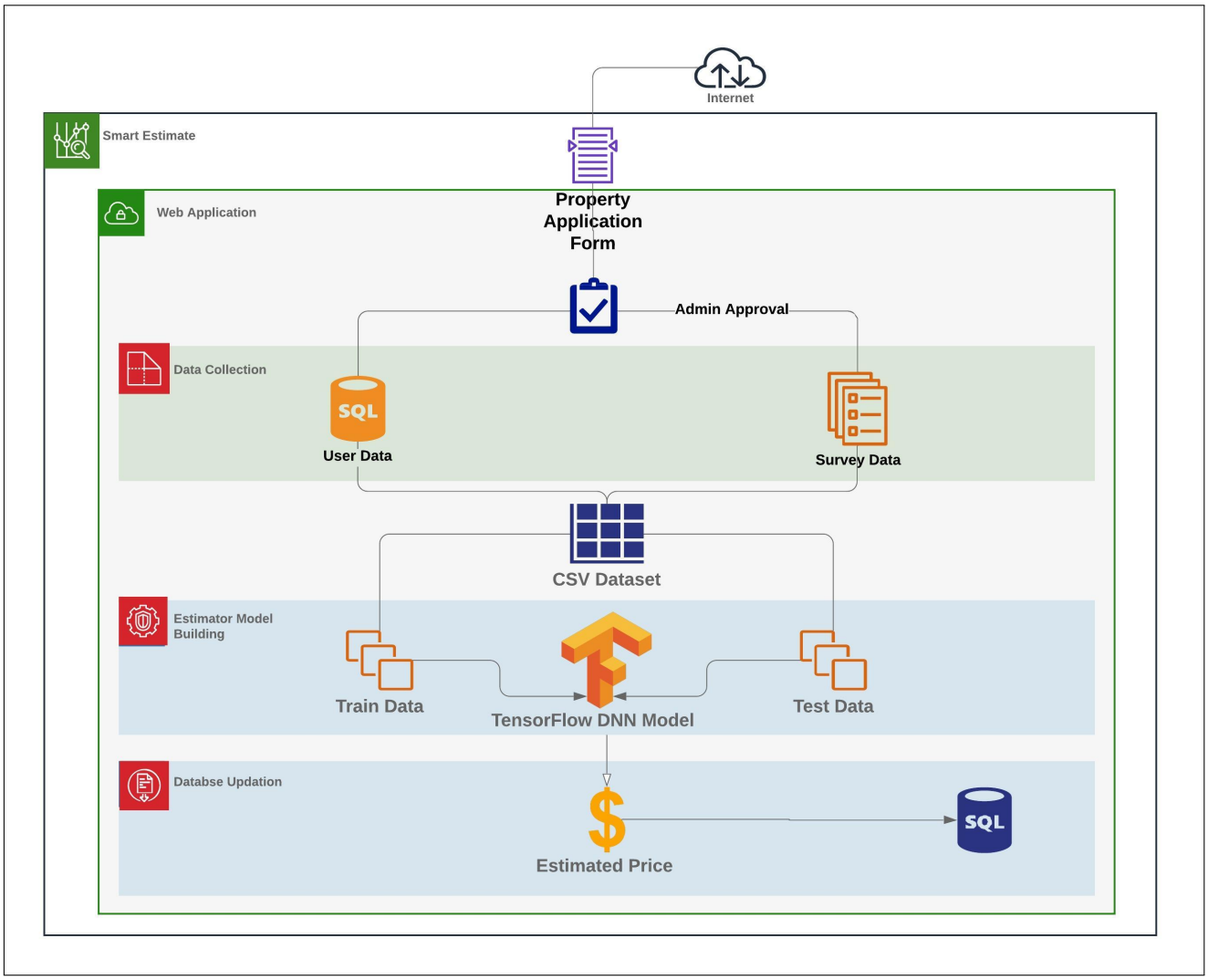


Figure 7.1: System Architecture

our approach, we first remove the following outliers: First is data points with insufficient information, e.g., missing square footage. And second is data points with unreasoning values such as a sold price of 1 as a gift. We additional calculate the average selling price of houses sold in last few months. Houses with too high or too low price-per-square feet are considered all outliers, and thus they are omitted from the training datasets and test datasets.

7.2.2 Layered Knowledge Graph

A knowledge graph is a graph with entities of different types as nodes and several relations between them as edges. Knowledge graph aims to represent entities and the relations between entities. We consider that a Deep Neural Network, which can produce suitable outputs with fewer weak connections between the neurons, would be nearer to a representation of a knowledge graph in a given field. We have selected a fully connected DNN with 2 hidden layers and 12 neurons in each hidden layer for our case study. In this selected neural network, the input layer contains 15 neurons, representing the predefined input features for house cost assessments, and a single output neuron representing the assessed house value or predicted selling price. We derived a layered knowledge graph for the real estate domain. We can see that the first layer contains 15 input nodes indicating that all selected features have important influences on house price assessments. The total number of hidden neurons in the second and third layer has been reduced from 12 to 10 and 7, respectively. This is due to the weak links and weak neurons being removed from the network. Finally, a single neuron in the last layer represents the assessed house value or predicted selling price

7.2.3 Structured DNN

A deep neural network is a type of neural network in which there exist a certain level of complexity. It is a neural network with more than two layers. The structured DNN is considered to match with the knowledge graph. We made experiments on fully-connected DNNs with different

numbers of hidden layers and different numbers of neurons in each hidden layer. The structured Deep Neural Network has four layers such as an input layer, two hidden layers, and an output layer. We set up suitable hyper-parameters for the structured DNN, and trained it using algorithm such as standard feed forward and back propagation algorithm with problem-specific real-time training and fitting techniques. Note that the first layer of the network contains 15 input neurons, which always produce outputs, as there are no biases are connected to the input layer neurons. Even though minor initialized weights create a neural network learn slowly, with sufficient offered data points, adjust a deep neural network with lesser weights which will help to get improved generalization, so to get good performance.

7.3 Component Design

7.3.1 Class Diagram

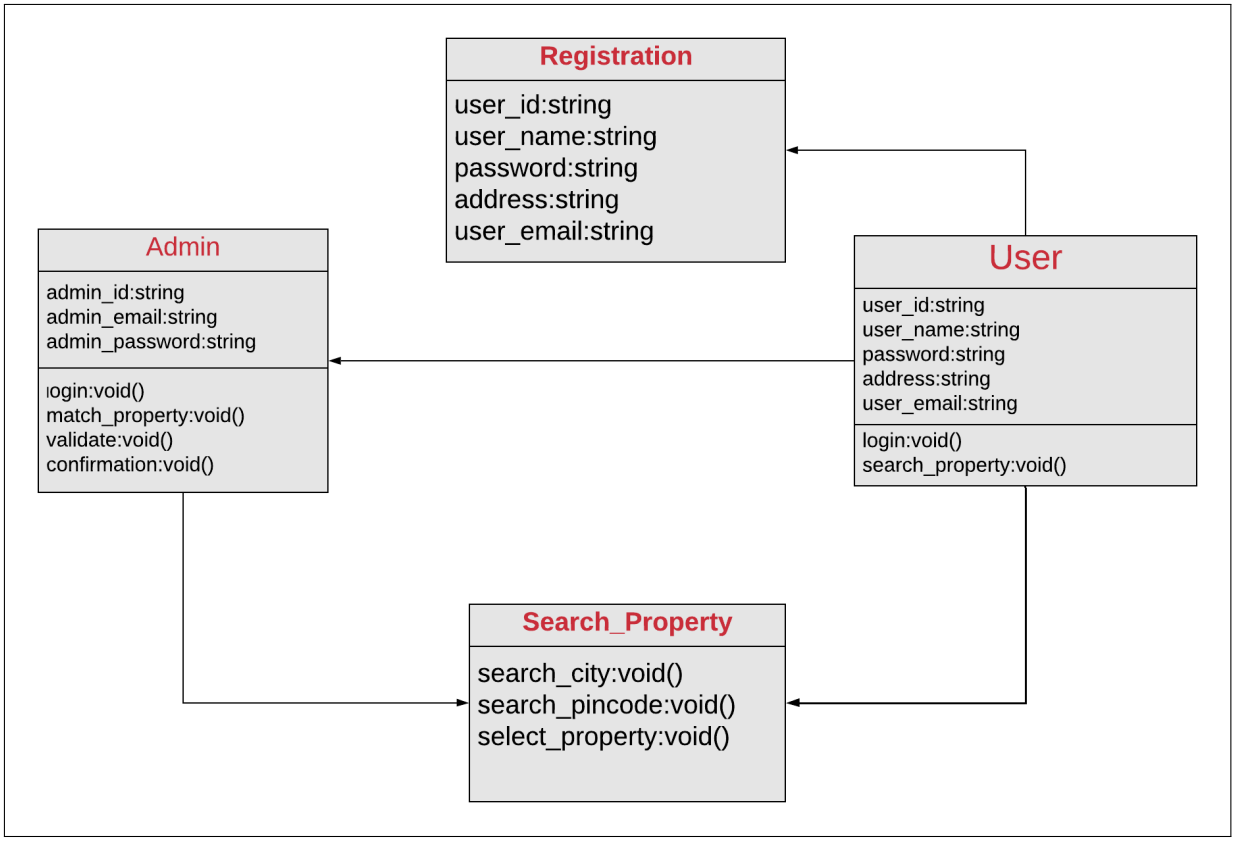


Figure 7.2: Class Diagram

7.3.2 Sequence Diagram

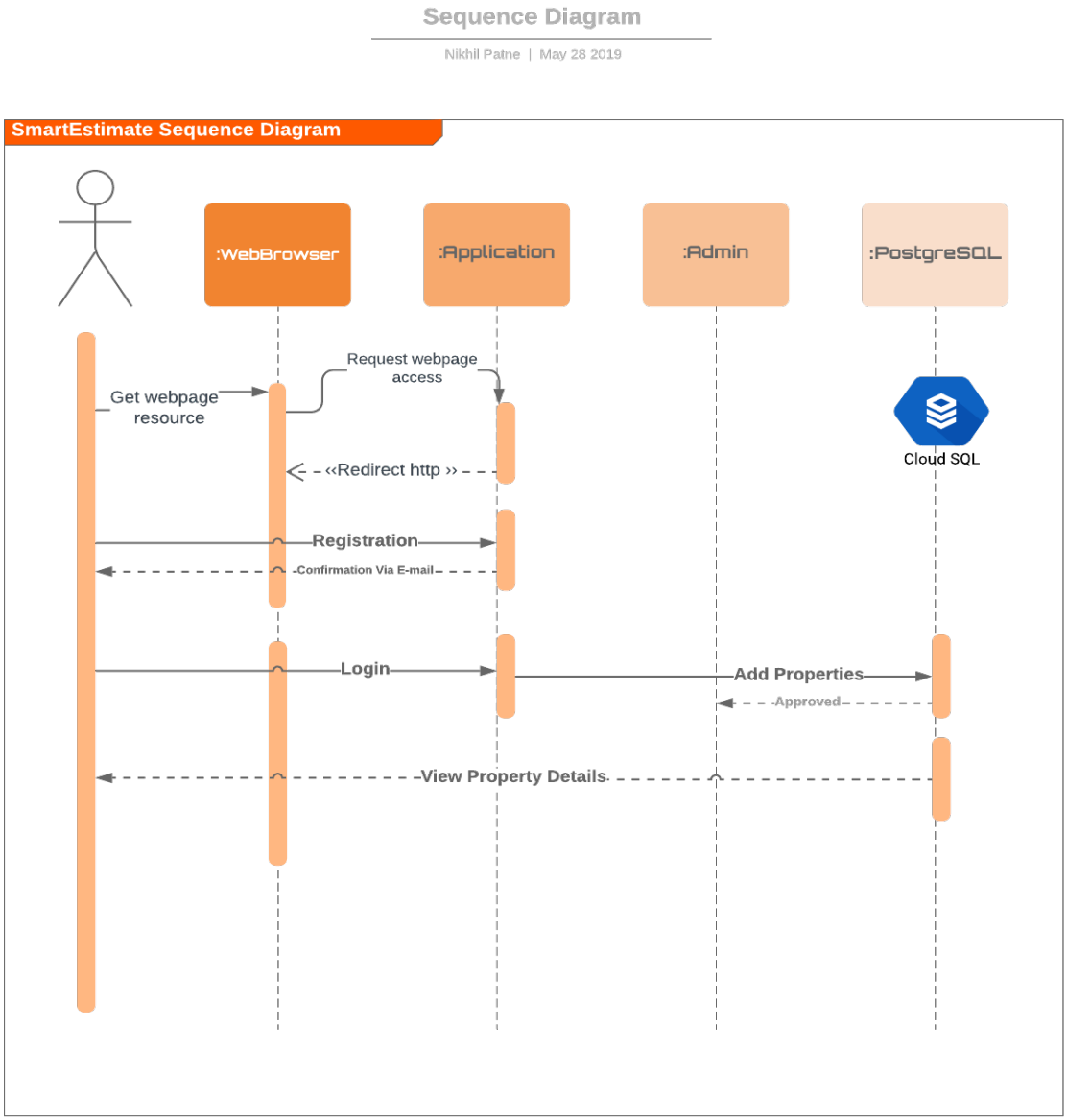


Figure 7.3: Sequence Diagram

7.3.3 Deployment Diagram

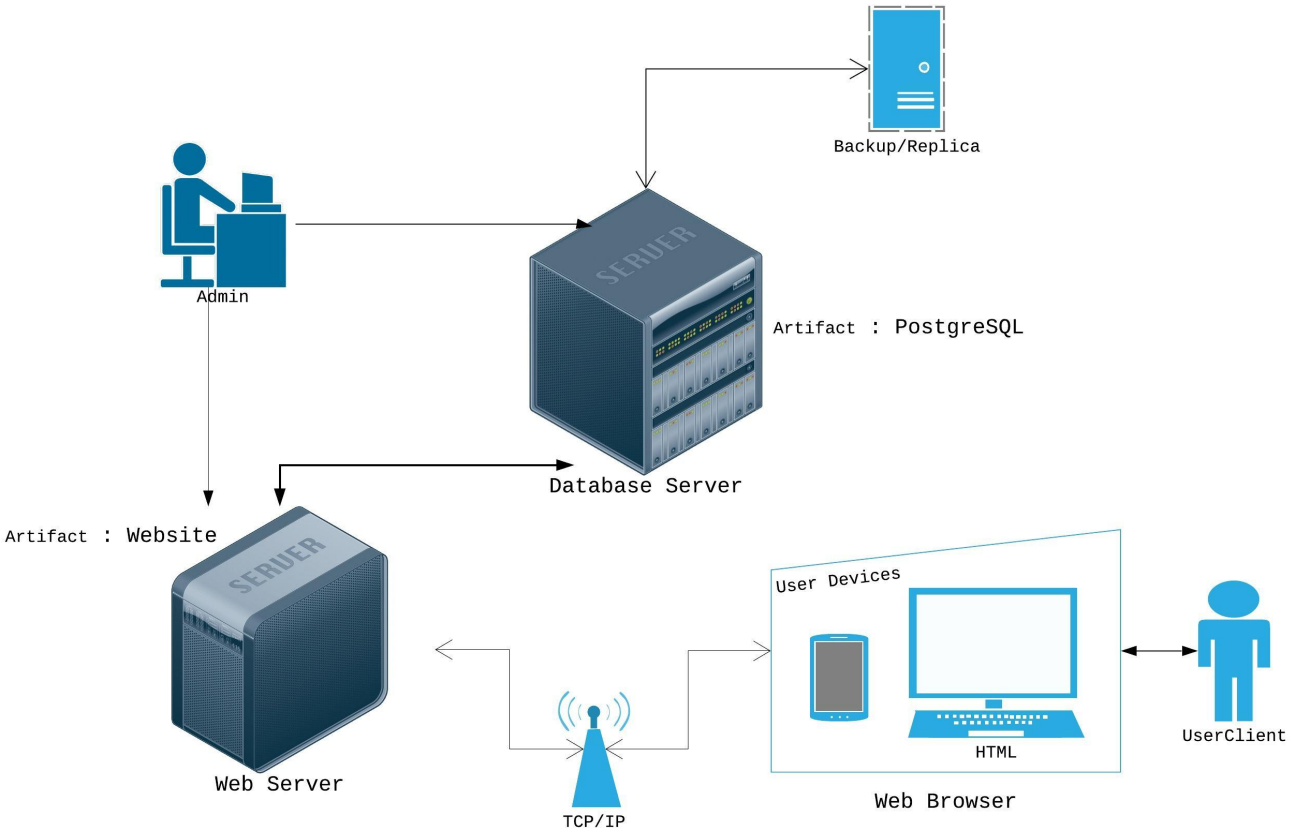


Figure 7.4: Deployment Diagram

8. Algorithms

8.1 Algorithm 1

Name

Construct Structured Knowledge

Input

Training dataset and testing dataset in domain D

Output

Layered Knowledge Graph (G) for domain D

Description

In this algorithm we take the training and testing dataset as the input and construct a Knowledge Graph from it. Firstly in this we initialize the DNN which has k layers and n(i) neurons in each layer. Then we train DNN using the training dataset which we have taken as input. And similarly we test the DNN using the testing dataset. At this point we record its accuracy. After training and testing the DNN, we identify the weak links, weak neurons and the weak layers .So basically the weak links are the weights are the links with weak weights that are below the predefined threshold. We delete these weak links from the DNN. Then we identify weak neurons; these are the neurons with very few links,

1. Initialize DNN Σ with k layers and n_i neurons in each layer.
2. Train Σ using the training dataset.
3. Test Σ using the test dataset and record its accuracy.
4. Repeat the following until there is a significant decrease in accuracy
5. Identify the weak links in Σ, and delete such links.
6. Identify the weak neurons in Σ, and delete or combine them.
7. Identify the weak layers in Σ, and delete or combine them.
8. Train the reduced Σ using the training dataset.
9. Test the reduced Σ using the test dataset and record its accuracy
10. Restore Σ from its previous iteration.
11. Name the hidden nodes and adjust Σ accordingly.
12. Create Φ according to Σ by ignoring all weights.
13. **return** layered knowledge graph Φ.

such links can also be deleted from the DNN. Or such neurons can be combined into a stronger one .After this we examine the hidden layers with very few neurons, which are called the weak layers. In a similar way we can delete the weak layer or combine two adjacent weak layers into a single one. After removal of all the outliers we again train the DNN using the training dataset and test the DNN using the testing dataset and we once again record the accuracy. We restore the resultant DNN which is a structured with no weak links or weak neurons or weak layers and we name the hidden nodes and adjust the DNN accordingly. Then we create a graph which is the layered Knowledge Graph.

8.2 Algorithm 2

Name

Create Structured DNN

Input

from algorithm 1, Layered Knowledge Graph G

Output

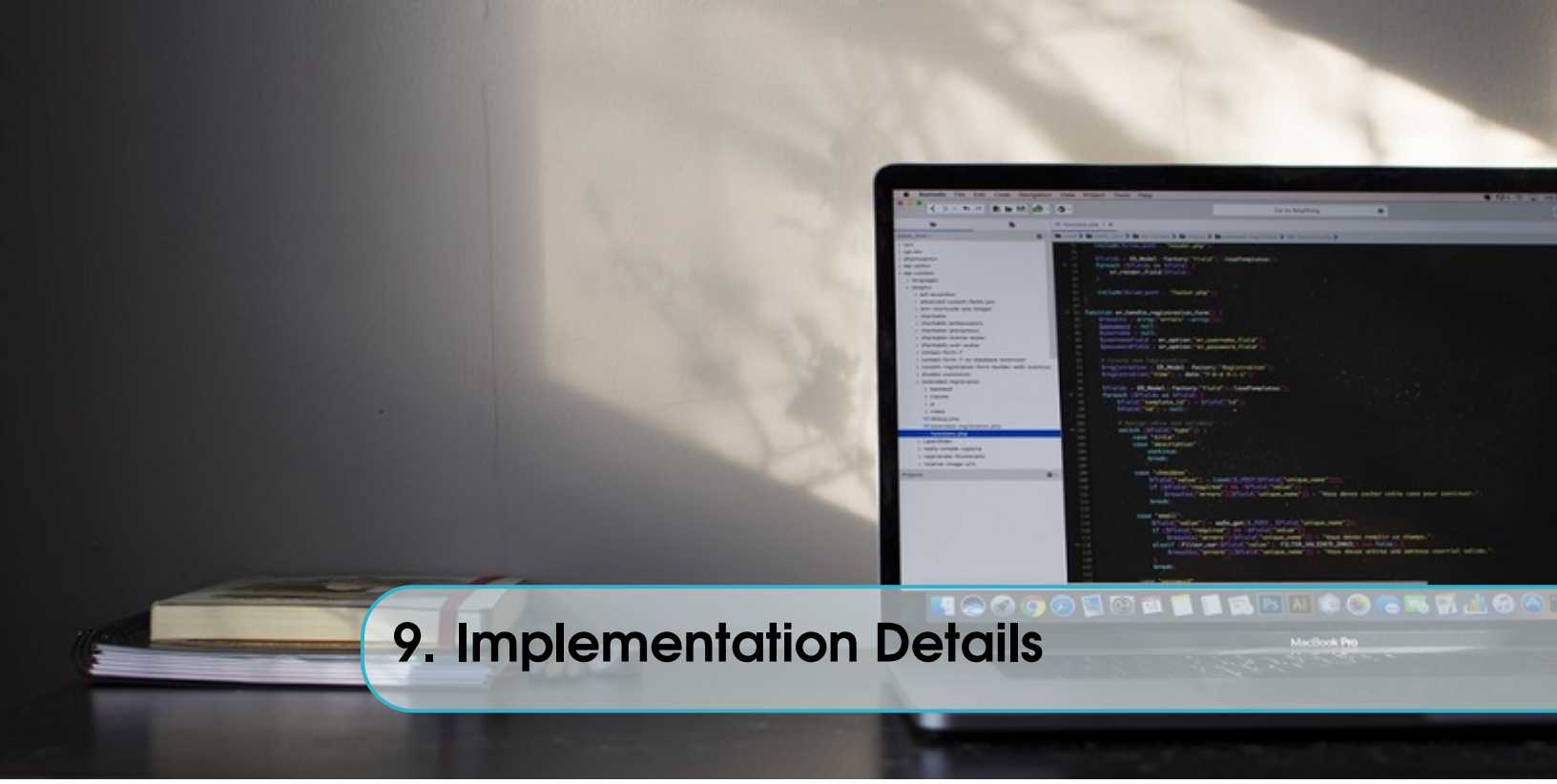
a structered DNN S for Domain D

---

1. Create Ψ with k layers and n_i neurons in each layer as in Φ.
2. Let π(n) be the mapping of node n in Φ to a neuron in Ψ
2. **for** layer l from 1 to k-1 in Φ
3. **for each** node n in layer l
4. **if** node n has an edge to node x in layer l+1 of Φ
   Connect π(n) and π(x) in Ψ
5. **for** layer l from 2 to k in Ψ
6. **for each** neuron n in layer l
7. add a bias node and connect it to neuron n
8. **return** structured DNN Ψ

Description

A layered knowledge graph is used to derive structured DNN it is used for training and price prediction. Constructed DNN has 4 layers :Input layer,2 hidden layer, output layer.structured DNN trained using standard feedforward back-propagation algorithm.We initialize all connection weight with relatively small random value within range [-0.5,0.5].Structured DNN contains only needed connection between neurons,therefore it is Time and space efficient trained using fewer data points



9. Implementation Details

9.1 Dataset

1.Training Set:

Source:
https://www.kaggle.com/apapiu/regularized-linear-models/data

No.of Records: 1461

No.of Features: 81

2.Testing Set:

Source:
https://www.kaggle.com/apapiu/regularized-linear-models/data

No.of Records: 1460

No.of Features: 80

Target Attribute: Estimated Price

1. To add Datasets to collaboratry

```
from google.colab import files

files.upload()

!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle/
!chmod 600 ~/.kaggle/kaggle.json

!kaggle competitions download -c house-prices-advanced-regression-techniques
```

Above Code will Automatically download CSV's from kaggle and added to our local colab Notebook,you may add datasets manually also

2. To Remove Outliers

```
train = pd.read_csv('train.csv', encoding='utf-8')
train_numerical = train.select_dtypes(exclude=['object'])

from sklearn.ensemble import IsolationForest

clf = IsolationForest(max_samples=100, random_state=42)
clf.fit(train_numerical)

y_noano = clf.predict(train_numerical)

train_numerical.reset_index(drop=True, inplace=True)
```

SkLearn Provides IsolationForest class to remove outliers from dataframe

9.2 DNN Model

1. To build and train the model

TenserFlow provides the DNNRegressor class like *LinearRegression()* in SciKit Learn.

```
regressor = tf.contrib.learn.DNNRegressor(feature_columns=engineered_features,
                                          activation_fn=tf.nn.relu, hidden_units=[250, 100, 50])
categorical_cols = {
    k: tf.SparseTensor(indices=[[i, 0] for i in range(training_set[k].size)],
values=training_set[k].values,
                       dense_shape=[training_set[k].size, 1]) for k in FEATURES_CAT}

regressor.fit(input_fn=lambda: input_fn_new(training_set), steps=10000)
```

Here we construct the structured knowledge graph/model framework by calling the DNNRegressor functions and passing in the features, hidden layers, and desired activation function. Here were are using three layers, each with a decreasing number of nodes.

2. Activation Functions

Activation functions are really important for a ANN and DNN also to learn and make sense of something really complicated and Non-linear complex functional mappings between the inputs and output variable.They introduce non-linear properties to our Network.Their main purpose is to convert a input signal of a node in a A-Neural Network to an output signal. That output signal now is used as a input in the next layer in the stack.

Why Activation Function??

If we do not apply a Activation function then the output signal would simply be a simple linear function.A Neural Network without Activation function would simply be a Linear regression Model, which has limited power and does not performs good most of the times.Without Activation function our model will not able to handle complex,non-linear data.

Types Activation Function??

1. Sigmoid or Logistic
2. Tanh (Hyperbolic Tangent)
3. ReLu (Rectified Linear Unit)

Among Them we are using ReLu as a Activation Function.

ReLu:

It has become very popular in the past couple of years. Recently proved that it has 6 times improvement in convergence from Tanh and sigmoid function.ReLu is only be the applied to hidden layers. And if your model suffers form dead neurons during training we should use leaky ReLu or Maxout function.

It's just that Sigmoid and Tanh have not been used nowadays due to the Vanishing Gradient Problem which causes a lots of problems to train,Minimizes the accuracy and performance of a deep Neural Network Model.

Mathematically it can be expressed as

$y = max(0,x)$

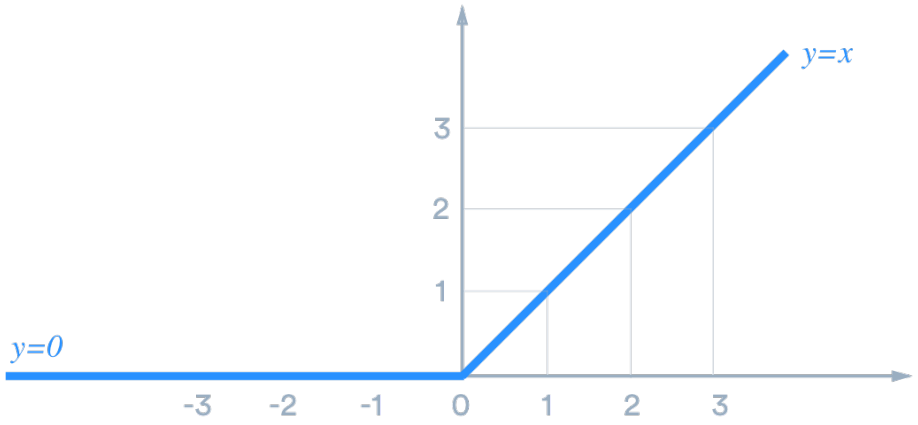


Figure 9.1: Activation function:ReLu

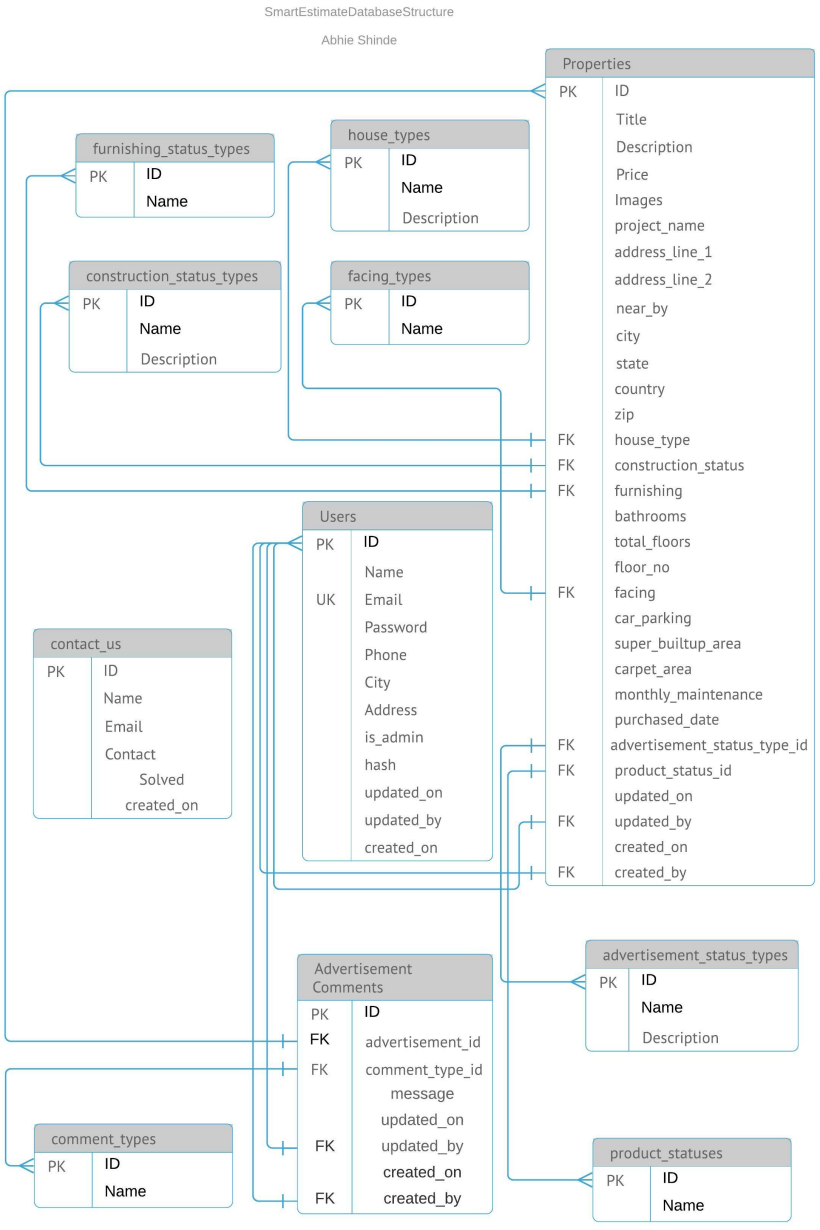
3. Evaluating the model

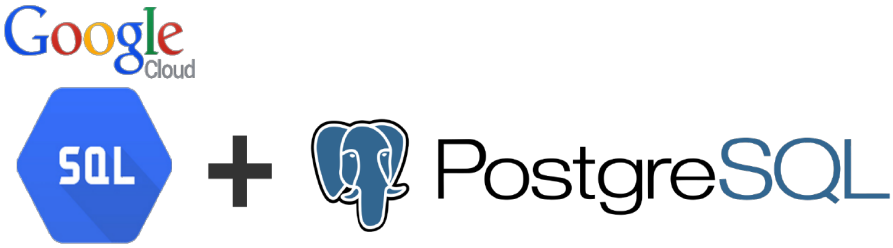
```
ev = regressor.evaluate(input_fn=lambda: input_fn_new(testing_set, training=True), steps=1)
loss_score = ev["loss"]

print("Final Loss on the testing set: ".format(loss_score))
```

4. Visualizing Results

```
import matplotlib.pyplot as plt

plt.style.use('ggplot')
plt.plot(predictions.values, reality.values, 'ro')
plt.xlabel('Predictions', fontsize=30)
plt.ylabel('Reality', fontsize=30)
plt.title('Predictions vs. Reality on dataset Test', fontsize=30)
plt.show()
```

9.3 Database


SmartEstimateDatabaseStructure
Abhie Shinde
Properties
PK ID
Title
Description
Price
Images
project_name
address_line_1
address_line_2
near_by
city
state
country
zip
FK house_type
FK construction_status
FK furnishing
bathrooms
total_floors
floor_no
FK facing
car_parking
super_builtup_area
carpet_area
monthly_maintenance
purchased_date
FK advertisement_status_type_id
FK product_status_id
updated_on
FK updated_by
created_on
FK created_by
furnishing_status_types
PK ID
Name
house_types
PK ID
Name
Description
construction_status_types
PK ID
Name
Description
facing_types
PK ID
Name
Users
PK ID
Name
UK Email
Password
Phone
City
Address
is_admin
hash
updated_on
updated_by
created_on
contact_us
PK ID
Name
Email
Contact
Solved
created_on
Advertisement Comments
PK ID
FK advertisement_id
FK comment_type_id
message
updated_on
FK updated_by
created_on
FK created_by
advertisement_status_types
PK ID
Name
Description
comment_types
PK ID
Name
product_statuses
PK ID
Name



For this project we are using PostgreSQL as Database with the integration of Google Cloud.

Why PostgreSQL ??

1. PostgreSQL supports JSON and other NoSQL features and key-value pairs with HSTORE. It also supports indexing JSON data for faster access.
2. PostgreSQL has native SSL support for connections to encrypt client/server communications.
3. Overall PostgreSQL performance is best in execution of complex queries.
4. PostgreSQL is Object-Relational Database Managment System.
5. PostgreSQL will be ideal for your project if your requirements revolve around complex procedures, integration, High level designs and data integrity.

Why Google Cloud SQL ??

1. Automated Backup.
2. Enhanced Datacenter Availability.
3. More Secure Connectivity.
4. Automated Replicas
5. Easy Collabration with multiple users.
6. Higher Scalibility and Reliability
7. Auto Maintanance

Instance Details:

Instance Name: smartEstimate123
connection id: smart-real-estate-assessment:smartestimate123
public IP Address: 35.240.164.12
Availibilty Zone/Region: asia-southeast1

9.4 Google Maps

JavaScript Maps API 3.37

Throughout project we have used Google Maps javascript API's in-order to show property locations on webpage.

```
function intiMap()  {
  var Wai = {lat: 18.5204, lng: 73.8567};
  var map = new google.maps.Map(document.getElementById('map'),{
    center: pune,
    zoom: 10
  });

  var marker = new google.maps.Marker({
    map: map,
    position: Wai,
    title: 'Nikhil Patne'
  });
}

<script
src="https://maps.googleapis.com/maps/api/js?key=[API_KEY]&callback=initMap"
async defer>
</script>
```

Above block of code will allow you to display map in your webpage.

Note: Billing must be enable to access this service.

10. Tools And Technogies used

PHP Storm is powerful Editor for Building PHP Scripts

Visual Studio Code is a code editor redefined and optimized for building and debugging modern web and cloud applications

PostgreSQL Database is a general-purpose object-relational database management system. It allows you to add custom functions developed using different programming languages such as C/C++, Java, etc. PostgreSQL is designed to be extensible.

Google Meet is a video conferencing app. It is the business-oriented version of Google's Hangouts

Google cloud Platform is a suite of public cloud computing services offered by Google. The platform includes a range of hosted services for compute, storage and application development that run on Google hardware.

Google Colab allows you to execute TensorFlow code in your browser with a single click. Google colab is a cloud based data science work space similar to the jupyter notebook.

GitLab GitLab is a web-based DevOps lifecycle tool that provides a Git-repository manager providing wiki, issue-tracking and CI/CD pipeline features, using an open-source license, developed by GitLab Inc.GitLab offers some similar features for issue tracking and project management as GitHub.

Trello Trello is a task management app that gives you a visual overview of what is being worked on and who is working on it

Google Sheets is a Web-based application that allows users to create, update and modify spreadsheets and share the data live online. The Ajax-based program is compatible with Microsoft Excel and CSV (comma-separated values) files. Spreadsheets can also be saved as HTML. We have maintained our logbook in google sheets

PyCharm is an IDE by Jetbrains. It is used for development in Python and frameworks. You can customize it with themes and plugins. It lets you to enhance productivity while coding by providing some features like suggestions, Local VCS etc.

OverLeaf The aim of Overleaf is to make the process of writing, collaborating, and publishing scientific documents in LaTeX

Lucid-Chart is one of the ideal Drawing collaborative tool which allows you to create,share,manage of UML,wireframes etc

PHP is an open source, interpreted and object-oriented server-side scripting language. It is used to develop dynamic page web applications.

Python For aspiring Data Scientists, Python is probably the most important language to learn because of its rich ecosystem

11. Software Testing

11.1 Introduction

Software testing is the process of evaluation a software item to detect differences between given input and expected output. Also to assess the feature of A software item. Testing assesses the quality of the product. Software testing is a process that should be done during the development process. In other words software testing is a verification and validation process.

11.1.1 Type Of Testing Used

1. Unit Testing
2. Integration Testing
3. Validation Testing
4. System Integration

11.2 Test Cases

11.2.1 Authentication/Login Test Scenario

Test Case ID	001	Test Case Description	Test The Login Functionality		
Created By	Nikhil Patne	Reviewed By	Nimbalkar S.S	Version	1.1

QA Tester's Log	Review comments from Nimbalkar S.S incorporate in version 1.1

Tester's Name	Nikhil Patne	Date Tested	28 May 2019	Test Case (Pass/Fail/Not Executed)	Pass

S#	Prerequisites:
1	Access to Chrome Browser
2	Good Internet Connection
3	Windows/Mac/Linux

S#	Test Data
1	Username = nikhilpatne94@gmail.com
2	Pass = Nick@1234

Test Scenario	Verify on entering valid username and password , the user can login

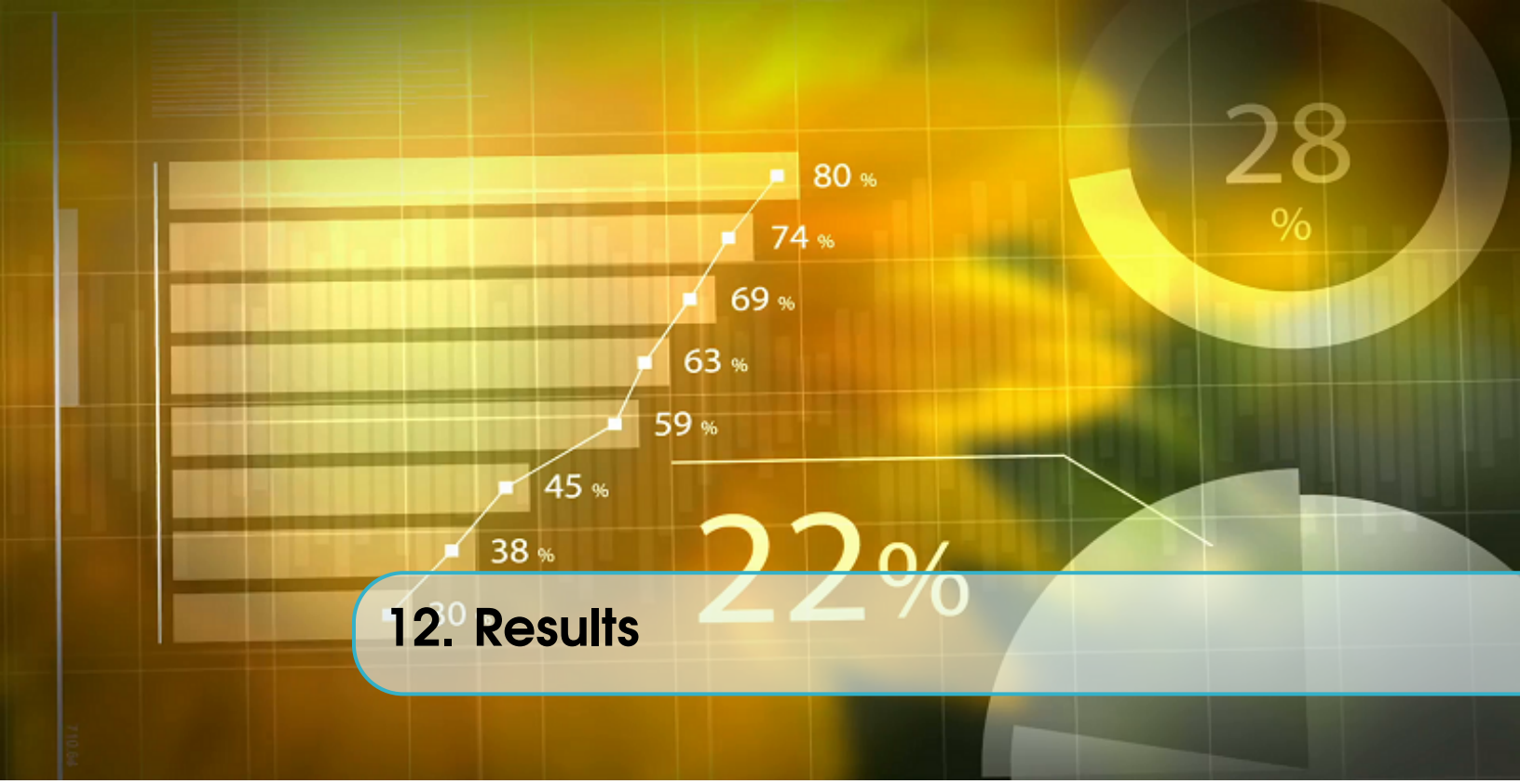
Step #	Step Details	Expected Results	Actual Results	Pass / Fail / Not Executed / Suspended
1	Navigate to http://SmartEstimate.com	Site should open	As Expected	Pass
2	Enter Username & Password	Credentials can be entered	As Expected	Pass
3	Click Submit	Customer gets logged into and dashboard is on display	As Expected	Pass

Figure 11.1: Login Test Cases

11.2.2 Search Bar Testing

Test Cases	Test Cases Objectives	Expected Results	Status (Pass / Fail)
TC-001	Search bar suggestion	When enter in search bar it give suggestions for different properties	Pass
TC-003	Search bar option selection	It should allow the user to select properties	Pass

Figure 11.2: Search Bar Test Cases



12. Results

12.1 Why DNN is better than Regression ??

The main objective of our project is to predict the approximate price of the real estate properties as possible as accurate.As we know Zillow,RedFin etc such organisations are currently working on Price prediction of real estate properties from past few years effectively. Regression Analysis is best suited for prediction of numerical type values i.e in our case Estimated price.

Our Reference author paper stated that the model which is created with DNN algorithm that is much better than model created by Regression Analysis. DNN model will predict more accurate results as compared Regression Model.
Now,

Let see How Finally Regression is Beaten by Deep Neural Network

12.1.1 Results Generated by Regression Model:

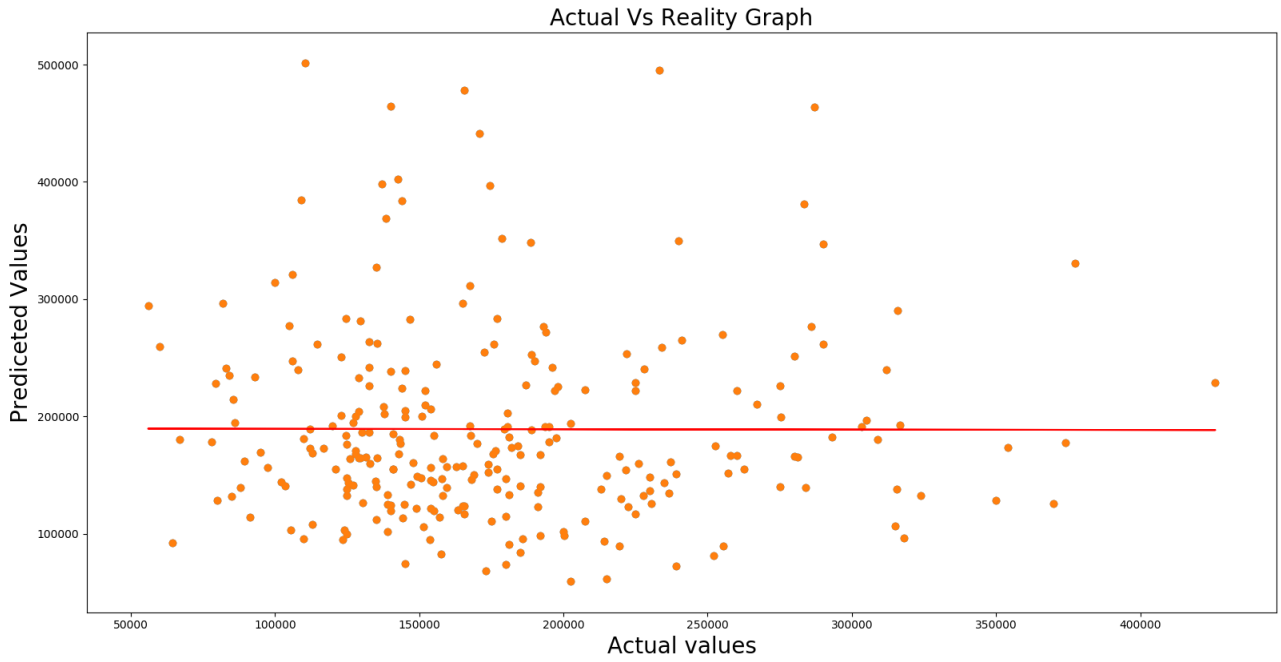


Figure 12.1: Results Generated By Regression Model

Accuracy: 0.069

Error Rate: 0.931

From Above Diagram it has been cleared seen that there is not a great linearity between Actual and Predicted values.

Regression model is not well Fitted.

12.1.2 Results Generated by DNN Model:

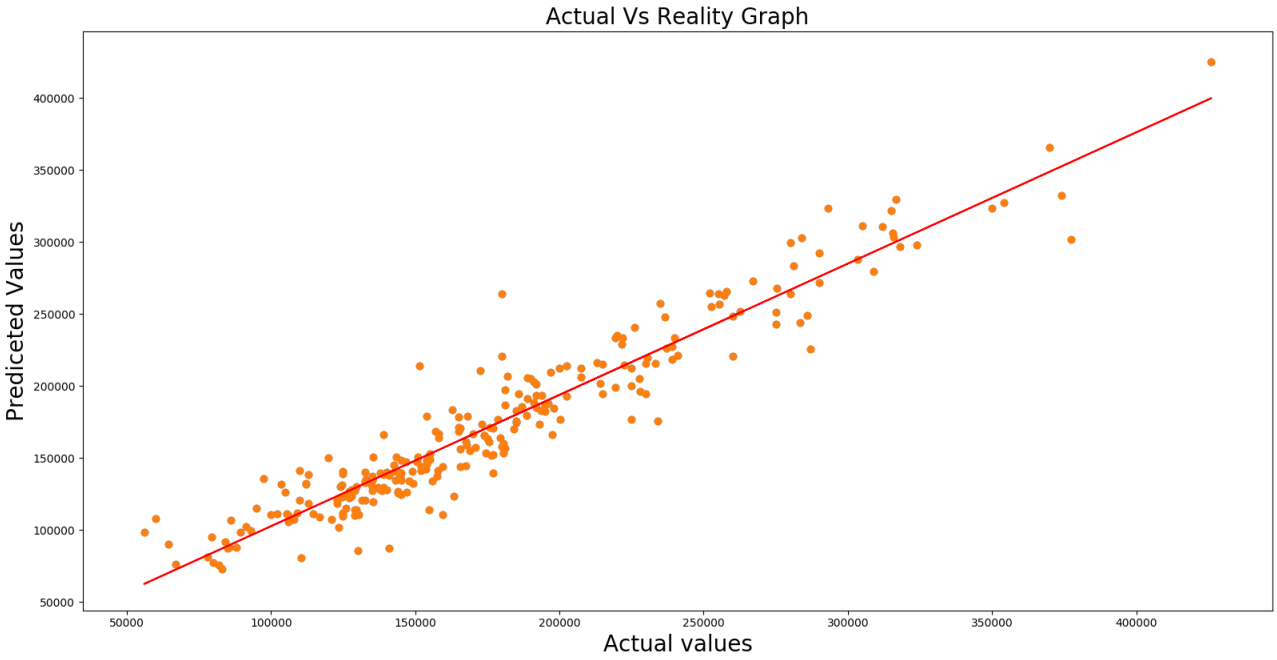


Figure 12.2: Results Generated By DNN Model

Accuracy: 0.9144

Error Rate: 0.085

From Above Diagram it has been cleared seen that the model which is created by DNN algorithm is well fitted as compared to regression analysis.
There is much similarity between Actual Values and Predicted Values.

Hence,Practically it has been proved that DNN model is much better than Regression for prediction of numerical values

12.2 ScreeShots

Home Page:

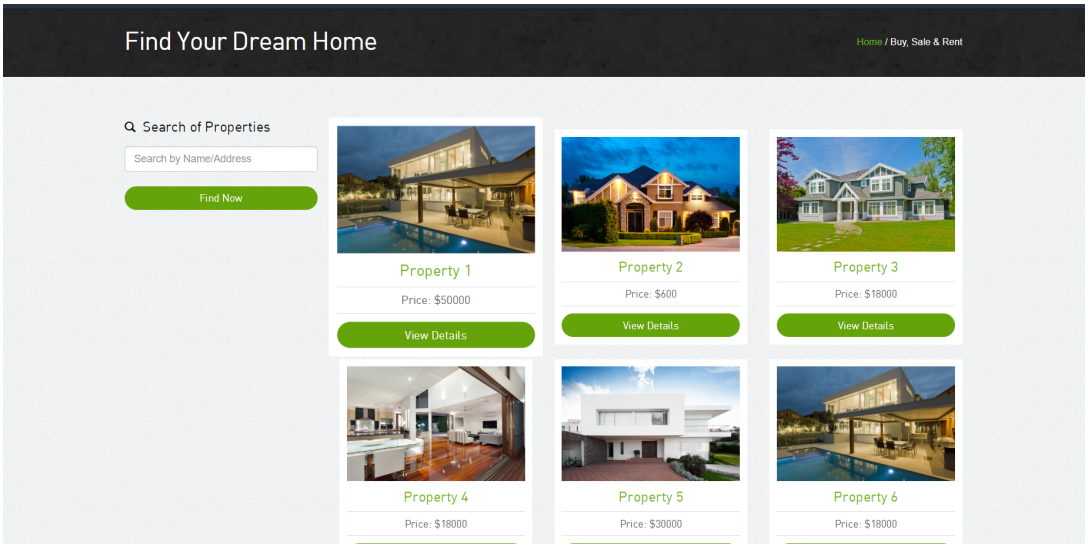


Figure 12.3: Home Page

Login Page:

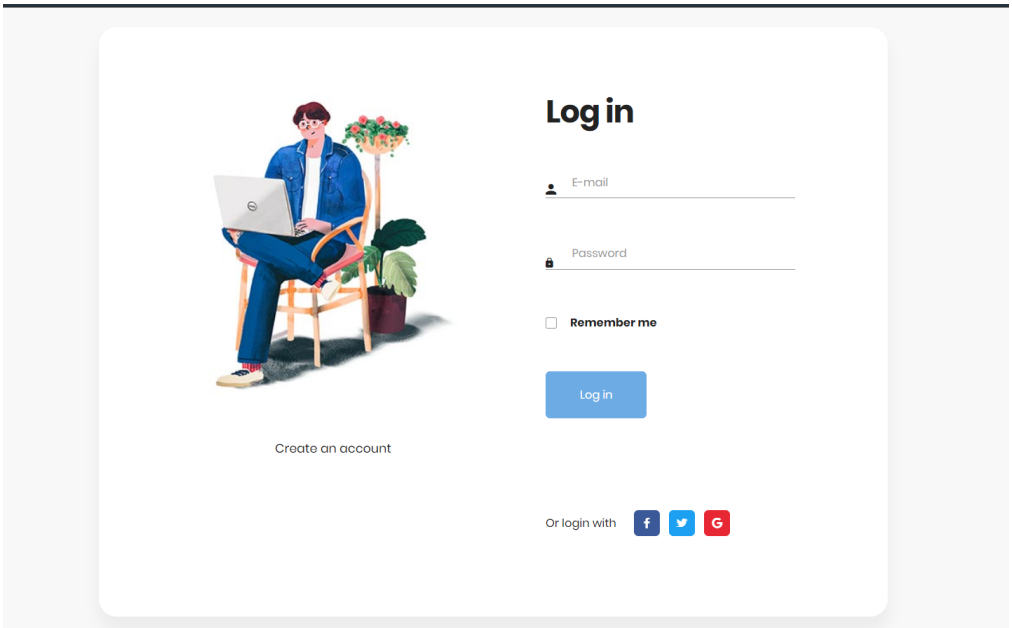


Figure 12.4: Login Page

Contact Us:

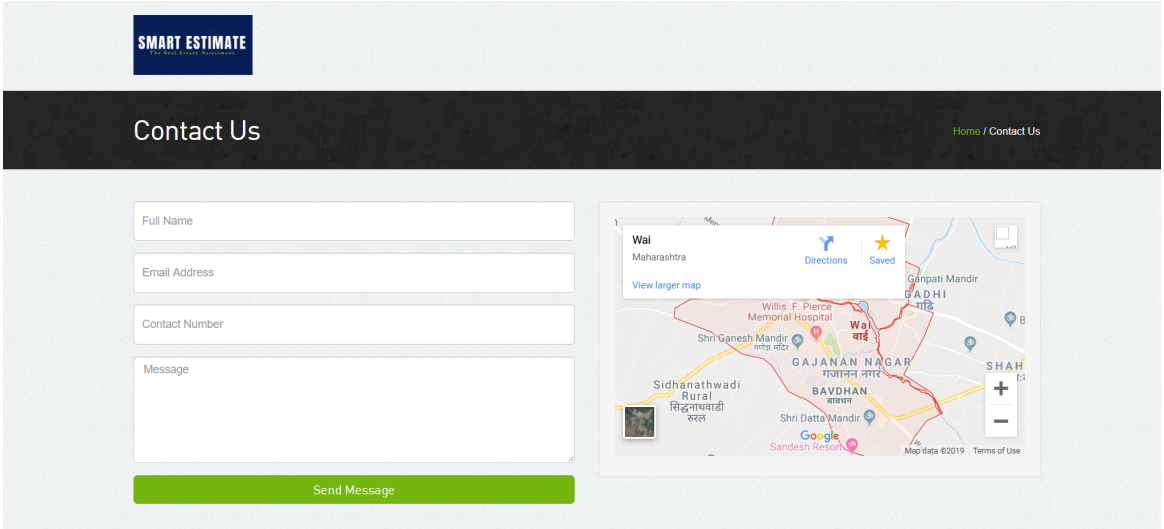


Figure 12.5: Login Page

DataBase:

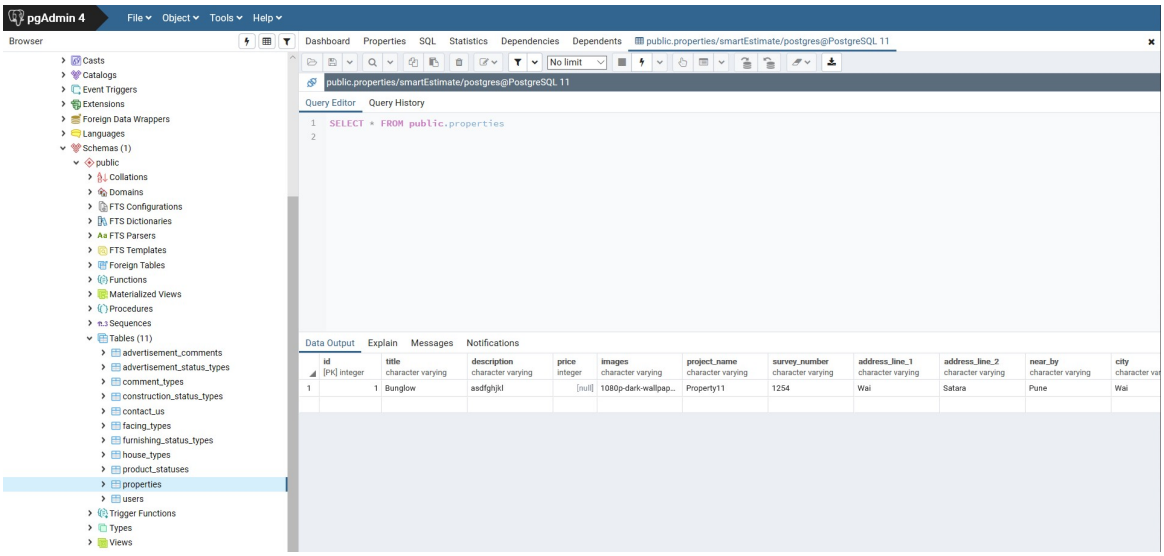


Figure 12.6: Database Design

13. Conclusion and Future Scope

In this way,We have successfully predicted the house price using DNN model and proved DNN Model is much accurate that Linear Regression.but a complex deep architecture may also prone to have the overfitting issue.Our experimental Results proved that Why DNN gives more accurate results.

For future research, We are trying to make our System for Rural Areas also also we will study how to automate the process of extracting layered knowledge graphs from the real estate domain based on historical data, and design structured DNNs using the graphs. We will allow a DNN model to automatically change its network structure along the time, so it can be more scalable and better adapt to new market changes. Furthermore, we plan to implement our approach using mobile cloud computing. that supports assessments of real estate via mobile devices with computation functions deployed in the clouds.

References

Haiping Xu and Amol Gade, *"Smart Real Estate Assessments using Structured Deep Neural Networks"* IEEE 2017, pp.255-266.

Quanzeng You, Ran Pang, Liangliang Cao, and Jiebo Luo, Fellow, IEEE *"Image Based Appraisal of Real Estate Propertie"* in CIKM, 2017, pp. 481–490.

N. Nguyen and A. Cripps Predicting Housing Value: *"A Comparison of Multiple Regression Analysis and Artificial Neural Networks,"* Journal of Real Estate Research, vol. 22, no. 3, pp. 313-336, 2016

Y. E. Hamzaoui and J. A. H. Perez, *"Application of Artificial Neural Networks to Predict the Selling Price in the Real Estate Valuation Process"* in MDM, vol. 2, 2015, pp. 31–36.

J. Frew and G. Jud, *"Estimating the Value of Apartment Buildings"* Journal of Real Estate Research, Vol. 25, No. 1, 2003, pp. 77-86.

X. Zhang, *"Using Fuzzy Neural Network in Real Estate Prices Prediction,"* In Proceedings of the 26th Chinese Control Conference, July 26-31, 2007, Zhangjiajie, Hunan, China, 2006, pp. 399-402..

N. Nguyen and A. Cripps, *Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks,""* Journal of Real Estate Research, vol. 22, no. 3, pp. 313-336, 2001.

N. Nghiep and C. Al,, *Predicting housing value: A comparison of multiple regression analysis and artificial neural networks,"* Journal of Real Estate Research, vol. 22, no. 3, pp. 313–336, 2001.

Information of Project Group Members

1. Name : Patne Nikhil Sunil

2. Date of Birth : 29/10/1997

3. Gender : Male

4. Permanent Address : A/p Dhom, Tal.-Wai, Dist.-Satara

5. E-Mail : nikhilpatne94@gmail.com

6. Mobile : 8605358076

ISSN (Online): 2581-5792

IJRESM

IJRESM

International Journal of Research in Engineering, Science and Management

www.ijresm.com support@ijresm.com

SJIF Impact Factor: 4.308

Certificate

It is here by certified that the manuscript entitled

Real Estate Properties Assessment using Deep Neural Network

by

N. Patane

has been published in

Volume-2, Issue-5, May-2019

in

International Journal of Research in Engineering, Science and Management

All the best for your future endeavors

Editor-in-Chief (IJRESM)

1. Name : Shinde Abhishek Sunil

2. Date of Birth : 14/03/1997

3. Gender : Male

4. Permanent Address : A/p Phaltan, Tal.-Phaltan, Dist.-Satara

5. E-Mail : shindea890@gmail.com

6. Mobile : 7276643616

ISSN (Online): 2581-5792

IJRESM

IJRESM

International Journal of Research in Engineering, Science and Management

www.ijresm.com support@ijresm.com

SJIF Impact Factor: 4.308

Certificate

It is here by certified that the manuscript entitled

Real Estate Properties Assessment using Deep Neural Network

by

A. Shinde

has been published in

Volume-2, Issue-5, May-2019

in

International Journal of Research in Engineering, Science and Management

All the best for your future endeavors

Editor-in-Chief (IJRESM)

1. Name : Gholap Satyawan Ashok
2. Date of Birth : 02/10/1996
3. Gender : Male
4. Permanent Address : A/p Varand, Tal.-Daund, Dist.-Pune
5. E-Mail : satyawanraje7@gmail.com
6. Mobile : 9561522780

ISSN (Online): 2581-5792

IJRESM

IJRESM
International Journal of Research in Engineering, Science and Management
www.ijresm.com support@ijresm.com
SJIF Impact Factor: 4.308

Certificate

It is here by certified that the manuscript entitled

Real Estate Properties Assessment using Deep Neural Network

by

S. Gholap

has been published in

Volume-2, Issue-5, May-2019

in

International Journal of Research in Engineering, Science and Management

Editor-in-Chief (IJRESM)

All the best for your future endeavors

1. Name : Dange Namrata Milind

2. Date of Birth : 08/05/1998

3. Gender : Female

4. Permanent Address : A/p Junction, Tal.-Indapur, Dist.-Pune

5. E-Mail : dangenm@gmail.com

6. Mobile : 8600551630

ISSN (Online): 2581-5792

IJRESM

IJRESM

International Journal of Research in Engineering, Science and Management

www.ijresm.com support@ijresm.com

SJIF Impact Factor: 4.308

Certificate

It is here by certified that the manuscript entitled

Real Estate Properties Assessment using Deep Neural Network

by

N. Dange

has been published in

Volume-2, Issue-5, May-2019

in

International Journal of Research in Engineering, Science and Management

All the best for your future endeavors

Editor-in-Chief (IJRESM)

1. Name : Beera Vivechana Ramesh
2. Date of Birth : 26/01/1995
3. Gender : Female
4. Permanent Address : A/p Baramati, Dist.- Pune
5. E-Mail : vivechanab5@gmail.com
6. Mobile : 9823838097

ISSN (Online): 2581-5792

IJRESM

IJRESM

International Journal of Research in Engineering, Science and Management

www.ijresm.com support@ijresm.com

SJIF Impact Factor: 4.308

Certificate

It is here by certified that the manuscript entitled

Real Estate Properties Assessment using Deep Neural Network

by

V. Beera

has been published in

Volume-2, Issue-5, May-2019

in

International Journal of Research in Engineering, Science and Management

Editor-in-Chief (IJRESM)

All the best for your future endeavors

Details of Paper Publication

Certificate No: MT/3027/2019

MANTECH PUBLICATIONS

Certificate of Publication

This is to certify that the manuscript entitled "***Real Estate Price Estimation Using Deep Neural Network***" submitted by "***A.Shinde, N. Dange, N. Patane, S. Gholap, V. Beera***" has been published in "***Journal of Artificial Intelligence, Machine Learning and Soft Computing"*** ***Volume 4 Issue 1 Year 2019***

Date: 21th May 2019

For Mantech Publications Pvt. Ltd. (OPC)

Authorized Signatory

MANTECH Publications

MANTECH PUBLICATIONS,
203, Plot No-45, Gyankhand-2,
Indirapuram,
Ghaziabad-201012, (U.P.)

link:http://mantechpublications.com/admin/index.php/JoAIMLSC/issue/viewIssue/1945/pdf20

Certificate No: MT/3039/2019

MANTECH PUBLICATIONS

Certificate of Publication

This is to certify that the manuscript entitled "***Real Estate Properties Assessment Using Deep Neural Network***" submitted by "***A.Shinde, N.Dange, N.Patane, S.Gholap, V.Beera***" has been published in "***Journal of Artificial Intelligence, Machine Learning and Soft Computing"*** ***Volume 4 Issue 2 Year 2019***

Date: 30th May 2019

For Mantech Publications Pvt. Ltd. (OPC)

Authorized Signatory

MANTECH Publications

MANTECH PUBLICATIONS,
203, Plot No-45, Gyankhand-2,
Indirapuram,
Ghaziabad-201012, (U.P.)

link:http://mantechpublications.com/admin/index.php/JoAIMLSC/issue/viewIssue/1976/pdf21

Figure 13.1: Documentation
https://tinyurl.com/yyshcnf9

Figure 13.2: Source Code
https://tinyurl.com/yxurybs8