**Project Proposal for SSIP Funding**

|  |  |
| --- | --- |
| **1.** | **Institute Name:** BITS Institute of Technology |
| **2.** | **Department Name:** Computer Science and Engineering |
| **3.** | Title of Project: REAL ESTATE PRICE PREDICTION PROJECT |
| **4.** | **Guiding Faculty Name:** |
| **5.** | **Guiding Faculty Contact Number:** |
| **6.** | **Guiding Faculty E-mail ID:** |
| **7.** | **Team Leader Name:** RAJ R PATEL |
| **8.** | **Team Leader Contact Number:** 9909177985 |
| **9.** | **Team Leader E-mail ID:** RAJPATEL2150@GMAIL.COM |
| **10.** | **Details of Participating Student(S) – (max. 5):**   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Sr. No. | Name of the student | Branch | Enrollment No. | Semester | Contact No. | E-mail ID | | 1 | RAJ R PATEL | CSE | 180050131079 | 5 | 9909177985 | rajpatel2150@gmail.com | | 2 | JAINISH NIKUL SHAH | CSE | 180050131107 | 5 | 7016500675 | jainishnikul2525@gmail.com | | 3 | MIHIR Y PATEL | CSE | 180050131073 | 5 | 9924366651 | mihirpatel3082000@gmail.com | |  |  |  |  |  |  |  | |
| **11.** | **Need for the Project (max. 300 words):**  1) Python  2) Numpy and Pandas for data cleaning  3) Matplotlib for data visualization  4) Sklearn for model building  5) Jupyter notebook, visual studio code and pycharm as IDE  6) Python flask for http server  7) HTML/CSS/Javascript for UI |
| **12.** | **Objective, Methodology of the Project (max. 400 words):**  We have different real eastate-propery websites in the India to buy,sell or rent the particular real estate plot,house or building property.  But we can say they are one sided because their prices are given by the sellers;  for person who is not actually familliar with that location **it is very tough for him/her to judge actual worth price for that property.**  same thing can happen with the seller too. it is also hard for the seller too, to put price for the property which is fair and actual worth price to him and buyer.  As property prices are an important reflection of the economy,and housing price ranges are of great interest for both buyers and sellers.  So, In this project, we are building a data science and machine learning based **model that can Predict the property price** based on certain real life features such as:  -square fits of the property  -location  -flat/tenament  -number of hall,bedroom,kitchen  -number of bathrooms  -with or without furniture  -parking space  **Why we need this ?**  There chances in real life that property both the seller and the buyer are expecting more or less estimate price for the property than its worth actual price.  **Predicting property price based on certain above features give us an exact and worthful amount for particular property regardless of prices which are given by the sellers and expected by the buyers.**  with the help of this predicted price both seller and buyer can agree on amount which is fair for both of them.so no one has to go on loss.  for this kind of situation our data science and machine learning based model is absolutely useful to predict actual worth price for the property.  we call this predicted estimated value "Restimate".  Reatimate home valuation is our estimated market value.  it is not an appraisal.  **it can take on market factors.** Which means when situation like natural calamities or certain market ups and downs due to some specific reasons which can directly or indirectly affect the property values,we can change our prediction model according to that.  For example: pandemic attack like COVID19, all the economy is dropped which is directly affect on real estate prices because property bulding and buyings are going down.  if property is listed for sale,the restimate will change.  A house value is simply more than location and square footage. We are going to take advantage of all of the feature variables available to use and use it to analyze and predict house prices.  We are going to break everything into logical steps that allow us to ensure the cleanest, most realistic data for our model to make accurate predictions from.  1.Gathering Data (Load Data)  2.Analyzing the Test Variable (Sale Price)  3.Multivariable Analysis (Dimensionally Reduction)  4.One Hot Encoding  5.Remove Outlier  6.Modeling and Predictions |
| **13.** | **Current Status of Idea (max. 1000 words):**  **1.Gathering Data (Load Data):**  We are gathering the data – which must be valid. If we say there is no guarantee of validity of data, then no use of analyzing it.  So, we are going to perform Data Mining because the required data is available in variable format across multiple sources. So, we have done Data Loading.  **2.Analyzing the Test Variable (Sale Price):**   * We work on assumption that there are all measurable data that affects a house’s price. * Of course, there may be more parameters that matters as well such as house condition, location, facilities,etc. * But these parameters are more subjective and almost impossible to measure, so we can ignore them.   This all activities are included in one term concept called Data cleaning (i.e. is process of detecting and correcting corrupt or inaccurate records, by modifying or detecting or replacing that data.)  **3.Dimensionalily Reduction (Multivariable Analysis):**  Now we need to put check for a strong correlation among given parameters. And we know that there are too many of the factors on the basis of which the final classification is done. Factors are basically variables called features.  Higher the number of features the harder it gets to visualize the training set and then work on it.Sometimes these features are correlated and hence redundant.  So we are performing Dimensional reduction for reducing the number of random variables under consideration, by obtaining a set of a principal variables. It can be divided into feature selection and feature extraction.  **Feature selection** - try to find a subset of original set of variables or features to get smaller subset which can be used in model.  **Feature Extraction** – reduces the data in a high dimensional space too lower dimension space.  **4) One Hot Encoding:**  We will simply pass dataframes into the functions. Returns a new dataframe with a column for every “level” of rating that exists, along with either 1 or 0 specifying the presence of that rating for a given observation.  **5) Removal outliers:**  Outliers are observations points that are distinct from other observations.  For example:- In my dataset if there was one house with an area of 50 sq. meters for price of Rs 500000 such house may exist on market for various reasons, but they are not statistically meaningful.  We want to make price estimate based on market average so we don’t take such outliers into account.  **6)Selecting a Algorithm:**  Choosing a proper algorithm is very much important task as Machine learning algorithms are programs that can learn from data and improve from experience, without human intervention. Learning tasks may include learning the function that maps the input to the output, learning the hidden structure in unlabeled data.  So we can use Linear Regression, Logistic Regression, Classification and Regression Trees (CART), Naïve Bayes, KNN, K-means, Principal Component Analysis (PCA) , Random Forest regression or XGBoost.  But in our case Linear Regression would be the best as there will be more accuracy as per our input parameters.  So let’s talk about the basic concept that we are going to use and how our fully model would be based on that is Regression in Machine Learning.  Regression is a ML tool helps you make prediction by learning from the existing statistical data – the relationalship between your target (estimate /accurate price prediction for real estates) and a set of parameters (number of bedrooms , living area , plot area , location,etc)  Let us take a simple example of how regression algorithm would work is lets take linear relationalship between area and price  y = K0 + K1 \* x  Where y = price (estimate) and x= area  So simple solving and then getting results as values, plotting, clearing some of the values and the answer is Infront of you i.e. estimate price. So choosing a efficient regression algorithm is very much necessary.  **7) Evaluating model’s performance:**  In final section of the project, we will construct a model and make prediction on the client’s feature set using an optimized model.  GRIDSEARCH – technique that generates candidates from grid of parameters value specified. It performs to create dictionary of all the parameters and their corresponding set of values that you want to test for best performance. |
| **14.** | **Work Plan (including detailed methodologyand time schedule) – (max. 1000 words):**   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Activity / Time | Phase 1  (Duration) | Phase 2  (Duration) | Phase 3  (Duration) | Phase 4  (Duration) | Phase 5  (Duration) | | Data Gathering (gathering relevant data) | 1 month | - | - | - | - | | Analyzing Data (Data Cleaning) | 1 month | - | - | - | - | | Dimensionality Reduction (Check correlation between parameters) | - | 1.5 month | - | - | - | | One Hot Encoding (Reduce the dataframes obtain to 0 and 1) | - | 1 month | - | - | - | | Remove Outliers | - | - | 1 month | - | - | | Choose a proper algorithm for the analysics | - | - | - | 2 months | - | | Evaluating Model’s Performance (GridSearch) | - | - | - | - | 1 month | |
| **15.** | **Output of The Project (max. 200 words):**  After Completing all the above mentioned phases of project, **we will have a dynamic working website** in which ; there will be number of column which will be taking input as certain real life features already mentioned above and will give output as the approximate worth price of real estate. |
| **16.** | **Cost Analysis:**   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Sr. No. | Item/Component | Quantity | Cost per item (Rs.)\* | Total cost (Rs.) | | 1 | AWS Cloud Computing | 1 | $3/month = 224.65 Indian Rupee | $ 24 = 1800 Indian Rupee | | 2 | Domain Name for hosting the website | 1 | Rs 500 per domain | Rs 500 | | **Total** | | | | Rs 2300 | | \* Attach supporting document for cost of item | | | |  | |
|  | **Name and Signature of Students:**  Place: VADODARA  Date: 05/08/2020   |  |  |  | | --- | --- | --- | | Sr. No. | Name of the student | Signature of the student | | 1 | RAJ RAKESHKUMAR PATEL |  | | 2 | JAINISH NIKUL SHAH |  | | 3 | MIHIR YOGESH PATEL |  | | 4 |  |  | |
|  | **Name and Signature of Guiding Faculty:** |
|  | **Name and Signature of Head of Department:** |