Ahsanullah University of Science and Technology (AUST)
Department of Computer Science and Engineering

Project Proposal

Course No.: CSE4126

Course Title: Distributed Database Systems Lab

Submitted To-
Mr. G. M. Shahariar
Lecturer, Department of CSE, AUST.

Ashna Ahmed
Lecturer, Department of CSE, AUST.

Submitted By-
Alphi Shahrin Arpy
180104121
Section-C
Year- 4th
Semester-1st
Department-CSE

Apartment Rent Prediction using K Means Clustering

An Apartment prediction system using K Means Classifier, where almost 2,000 data sample will be used for testing. Data will be clustered into 4 different classes using K Means Clustering algorithm. The machine learning model will be implemented by PL/SQL.

Necessary data about the apartment will be stored in 2 different sites. After a JOIN operation data will be merged and K Means Clustering algorithm will be applied. After clustering data, from the console, a user can enter his/her apartment's feature and will get a predicted clustered class for the apartment.

For the database, there will be one main table House consisting of 23 attributes. House table will be divided into four vertical fragments:

1. HouseSale(SellNo, YearBuilt, YearSold, MonthSold, SalePrice, AparNo) will consist all previous sale details about the apartement.
2. HouseDetails(DetailNo, Size, Floor, HallwayType, AparNo) will have the details about the apartment.
3. HouseFacility(FacNo, HeatingType, AptManagement, ParkingLot, NApart, NManager, NElevator, AparNo) consists of the facilities buyer get from the apartment,
4. HouseFacilitiesNearBy(FactNo, Subway, PublicOffice, Hospital, Department, Mall, ETC, Park, School) which will store the facilities nearby of the apartment.

Based on the YearSold feature, the table HouseSale will be fragmented in two different fragments having apartment sold before and after 2010.

First two fragments will be on Server 1 and the last two fragments will be on server 2. A user will enter his apartments attribute on Server 1 and will get a predicted cluster.

As the table is divided into 2 different servers, the query for each segment in each site will be less complex as it will work on portion of the main table. Features can be implemented for each server based on the fragments. As price varies from time to time, the selling year is an important feature of clustering. Comparatively older apartments will be in one server whereas the new apartments will be on another server to make the table simpler.