

CSE 9099 – Projects

Instructions:

1. You are required to review the project statement and nominate yourselves for three projects based on the preferences. 1: Most preferred 2: Preferred 3: Less preferred. Due date: **28th March 2017**. IF you do not send your nomination we will allocate one project that you are required to work on.
2. CSE 9099 Projects will be created in Piazza for all project related discussion. It is mandatory for you to communicate all the discussions through Piazza.
3. The final list of project allocations will be shared in a couple of days after nominations. The team size cannot exceed 6. Each team should work on a different project. So, the allocations will be done accordingly.
4. You are required to work on one project only. You are required to report to the assigned mentor and update the status. Last moment discussions are not encouraged.
5. No other project other than the one assigned to you will be evaluated for certification.
6. You are required to submit a detailed project document on or before viva scheduled date.
7. The project document weighs 70% and viva weighs 30% of the total marks allotted.
8. You will have to start reviewing the domain and state-of-art to get a better understanding of the problem statement and methodologies used to solve the problem.
9. You are required to conduct the preliminary analysis of the data and discuss with your mentors for any suggestions. Without showing adequate analysis and progress from your end, please do not expect mentors to guide you. Limited help on R will be provided.

1. Route Optimization solution (Mentor: Ramesh Melapu & Yeshwanth Reddy)

Spring-Cleaning offers a wide variety of cleaning services. They provide carpet cleaning, tile and grout cleaning, upholstery cleaning, hardwood floor cleaning and air duct cleaning, etc. Broadly, they are categorized as residential and commercial cleaning. They have presence in 48 cities nationwide with various depots servicing specific regions.

Spring-Cleaning currently has an in house software that collects and maintains the services and schedules and decides vehicle routes. You need to work with Spring-Cleaning to develop an optimal routing solution for their vehicles. This will be accomplished by designing a solution using advanced optimization techniques. Spring-Cleaning strives to service requests by dynamically scheduling their resources under various constraints. You will be given data with routes scheduled on one day for

CSE 9099 – Projects

each of their depots. Every vehicle in a depot would have a set of stops for the day, where stops means customers. In addition, they have to consider constraints like length of the time to complete the job, distance between stops and if the customers have specified a time when they would like to get the service done, etc.

Note: *This project needs advanced mathematical and R skills. Though we cover all the necessary topics to execute this project in CSE 7113c module, you are encouraged to start early to learn the concepts & optimization techniques such as linear programming and Genetic algorithms.*

2. Job recommendation engine (Mentor: Harshit Lamba & Shilpa Kadam)

Quickhire is a US based job portal which is fairly new in the market for job seekers and has data collected for 2014-15 year. The portal is becoming quite popular among seekers for part-time jobs. They are currently shortlisting the jobs based on two relevance factors such as distance between the applicant's location and job location and industry to make recommendations to the applicants. They have collected the job views data which deals with time spent by an applicant on every job view and if they have applied for the job or not, etc. Also, the data has complete profile of the job and applicant. You are required to build a recommendation engine using this data of applicants and jobs and make recommendations accordingly.

Using the job views file alone you are required to build a collaborative filtering based method to recommend jobs. In addition to this, if you can also review the job details and applicant profiles and make more appropriate recommendations it will help the job seekers.

CSE 9099 – Projects

3. Text classification problem. (Mentor: Shilpa Kadam & ShivaPrasad)

LifeLine HealthServices is a top ranked Health care provider in USA with stellar credentials and provides high quality-care with focus on end-to-end Health care services. The Health Care Services range from basic medical diagnostics to critical emergency services. The provider follows a ticketing system for all the telephonic calls received across all the departments. Calls to the provider can be for New Appointment, Cancellation, Lab Queries, Medical Refills, Insurance Related, General Doctor Advise etc. The Tickets have the details of Summary of the call and description of the calls written by various staff members with no standard text guidelines. The challenge is, based on the Text in the Summary and Description of the call, the ticket is to be classified to Appropriate Category (out of 5 Categories) and Subcategories (Out of 20 Sub Categories).

4. Segment and Score customers based on the purchase patterns for a retail Client. (Mentor: Archana & Amit)

A retail company has launched multiple stores across the country and wants to understand their customer behavior. The client database consists of two years of historical transactional data for each store and their customer information. The client wants to develop an engine which segments and provides the score for customers based on their behavior and analyze their purchasing pattern which helps in designing various customer specific promotions. The client even wants to understand the product associations to guide in developing promotions.

5. Econometrics Time Series (Mentor: Jagannadha Rao Basa & Harshit Lamba)

This project focuses on the role of Foreign Direct investment in India's economic development and government policies towards FDI. Attracting inward investment is a key component in its economic development

CSE 9099 – Projects

initiatives, for the potential knowledge transfer and cross fertilization to domestic operators which it can provide. The project has the following objectives: To study the trends and pattern of flow of FDI and To assess the determinants of FDI inflows by studying relationships among the econometrics variables such as GDP, exports, and Political stability index, etc.

6. Developing a robust patient adherence model (Mentor: Maheshkumar & Pavan kumar)

In health care industry non-compliance with medication regimens is a huge problem with socio-economic consequences. Due to non-adherence, pharmacy retailers lose money. More importantly, people get sick and there is healthcare expense. Patients who are on chronic diseases are expected to take their drugs periodically but they do not always follow the prescribed medication as directed. An easy way to measure this is whether they are buying the medicines at the supposed intervals. For example, if a patient bought 90 capsules and he is supposed to have 3 capsules a day, he must come back before 30 days for buying next dose. If they do not come, they are non-adherent. A leading largest pharmacy retailer has collected huge data of patient transactions (fills of their medicines at each month). To improve the medication adherence, the client believes that engaging patients is one of the best ways the client plans to proactively target those members who are at risk. The client needs a model that can account for the risk of non-adherence per patient and predict if the patient is likely to non-adhere.

7. AllState Claim Prediction Challenge (Mentor: Pavan Kumar & Maheshkumar)

Allstate claim prediction challenge was hosted by Kaggle. A key challenge for the insurance industry is to charge each customer an appropriate price for the risk they represent. As risk varies from person to person, a deep understanding of different risk factors helps predict the likelihood and cost of insurance claims. Many factors contribute to the frequency and severity

CSE 9099 – Projects

of car accidents including how, where and under what conditions people drive, as well as what they are driving. Bodily Injury Liability Insurance covers other people's bodily injury or death for which the insured is responsible. The goal of this project is to predict Bodily Injury Liability Insurance claim payments based on the characteristics of the insured's vehicle.

8. Human Activity Recognition (Mentor: Vivek & Maheshkumar)

The Human Activity Recognition challenge was hosted by Kaggle. The database was built from the recordings of 30 study participants performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors. The objective is to classify activities into one of the six activities performed.

The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

9. Demand Forecasting of Retail Products (Mentor: Gautham & Ramesh Bagadi)

A popular retailer in USA CMart, wants to forecast demand for their various products in the retail stores. Forecasting of Demand helps retailer to take necessary measures to plan the inventory and at the same time minimizes the lost revenues from unavailability of products. The forecasting is expected to be performed not only based on the past purchase demands, but also on other various demand influencing factors such as Holiday events, weather changes etc. You are expected to build a demand forecasting framework that forecasts the demand of the products for future 2 weeks using set of advanced machine learning models.

10. Supply Chain - Ship From Store Optimization (Mentor: Bhargavi & Manas)

A UK clothing retailer NewLook is looking for a Supply Chain Ship From Store Optimization Solution to successfully fulfill their Online Orders through their various Stores. The objective of the project is to build

an optimization engine that recommends top 3 stores from a set of stores that can deliver the items of each online order. The recommendation should consider various constraints and should meet multiple objectives such as minimization of the impact of inventory, reduction of the Transit distances and times, minimization of the order costs etc.